Identification of vector AR models with recursive structural errors using conditional independence graphs

Marco Reale* **and Granville Tunnicliffe Wilson**†

** Department of Mathematics & Statistics,*
University of Canterbury,
Private Bag 4800, Christchurch, New Zealand.

Report Number: UCDMS2000/4 May 2000

Keywords: Partial correlation, Moralization, Causality, Graphical modelling, Lending channel

† *Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.*

Identification of vector AR models with recursive structural errors using conditional independence graphs

Marco Reale[*] and Granville Tunnicliffe Wilson[†]

May 2000

Abstract

In canonical vector time series autoregressions, which permit dependence only on past values, the errors generally show contemporaneous correlation. By contrast structural vector autoregressions allow contemporaneous series dependence and assume errors with no contemporaneous correlation. Such models having a recursive structure can be described by a directed acyclic graph. We show, with the use of a real example, how the identification of these models may be assisted by examination of the conditional independence graph of contemporaneous and lagged variables. In this example we identify the causal dependence of monthly Italian bank loan interest rates on government bond and repurchase agreement rates. When the number of series is larger, the structural modelling of the canonical errors alone is a useful initial step, and we first present such an example to demonstrate the general approach to identifying a directed graphical model.

KEY WORDS: Partial correlation, moralization, causality, graphical modelling, lending channel.

[*]Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800 Christchurch, New Zealand

[†]Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.

1 Introduction

The canonical pth order vector autoregressive model, VAR(p), of a stationary, m dimensional time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,m})'$ is of the form:

$$x_t = c + \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + e_t \tag{1}$$

where c allows for a non-zero mean of x_t and e_t is multivariate white noise with general covariance matrix V. Our working assumption is that the series is Gaussian but our methods should be applicable under wider conditions, such as e_t being I.I.D., presented for example in Anderson (1971). This model is attractive because its estimation from a sample $x_1, x_2, \ldots, x_n$, by least squares applied separately to each component of x_t, is straightforward. For large sample length n it is also fully efficient provided there are no subset constraints on these separate regressions, (Judge *et al* 1985), the properties of the estimates given by the regression are reliable, and the estimate of V is independent of the estimates of Φ_k. The order p of the regression may be determined by various methods including inspection of a multivariate partial autocorrelation sequence, see Reinsel (1993) pp 69-70, or minimization of an order selection criterion such as AIC, Akaike (1973), HIC, Hannan and Quinn (1979), or SIC, Schwarz (1978). Such criteria have been extensively used in time series contexts. Recently Swanson and White (1995) have used them for linear and non-linear modeling of multiple time series and in our examples we also tabulate their values as an aid to model selection.

There are various approaches to multiple time series modeling which seek either to transform models such as (1) to a form which includes contemporaneous relationships among the variables, or to identify directly such a form, see for example Box and Tiao (1977) and Tiao and Tsay (1989). Our aim in this paper is similar; our approach is to consider the structural autoregressive model of the same form as (1) but with the addition of contemporaneous dependence through a matrix coefficient Φ_0:

$$\Phi_0 x_t = d + \Phi_1^* x_{t-1} + \Phi_2^* x_{t-2} + \cdots + \Phi_p^* x_{t-p} + a_t. \tag{2}$$

The general relationship between time series and structural econometric models was developed by Zellner and Palm (1974). In our present context the equivalence between (1) and (2) is given by $\Phi_i^* = \Phi_0 \Phi_i$ and $\Phi_0 e_t = a_t$. A requirement of (2) is that the variance matrix D of a_t is diagonal. We require a further condition on Φ_0, that it represents a recursive (causal) dependence of each component of x_t on other contemporaneous components. This is equivalent to the existence of a re-ordering of the elements of x_t such that Φ_0 is triangular with unit diagonal. Each possible ordering of x_t therefore gives a potentially distinct form of (2), but these are all statistically equivalent, corresponding to factorizations of

$$V^{-1} = \Phi_0' D^{-1} \Phi_0. \tag{3}$$

This contrasts with the unique form of (1), which is the attractive feature of that model from a time-series modeling viewpoint.

The value of (2) therefore lies in the possibility that there is one particular form which, as a consequence of its representing a true simple mechanism, is more parsimonious in its parameterization than either (1) or the other forms of (2). This would be reflected in the ability to exclude many of the elements of Φ_0 and Φ_i^* from the model without penalizing the fit in comparison with the saturated forms of either (1) or (2). Identification of such a model may then provide added insight into the true mechanisms which generate the data.

That is what we seek to achieve by the method described in this paper. However we introduce this method in section 2 without reference to dynamic structure. The relationship (3) shows that some, though not necessarily complete, information on the structure of Φ_0 is available from the variance matrix of the innovations. We therefore first illustrate the method, in section 2, without reference to any lagged structure, by its application to an innovation series e_t. This arises from a canonical autoregressive-moving average (ARMA) model fitted to a series of seven daily dollar term interest rates. In section 3 we consider how the method may then be extended to identify structural autoregressions. Section 4 contains an an example which illustrates this approach using three monthly monetary time series..

2 Recursive structure and partial correlations

Neglecting, for the present treatment, any effects of time series model estimation, we suppose that we have observations on the vector Gaussian white noise innovations process e_t with the usual sample covariance matrix $\hat{V}$. We wish to determine *from the data* the form of possible sparse structural matrices Φ_0 which are compatible with $\hat{V}$. There may be no unique such form without imposing further constraint using insight from the modeling context. Swanson and Granger (1997) consider an almost identical problem, but focus more on testing for the constraints which are implied by a particular structural form of Φ_0 which has commonly occurred in practice. Their tests are expressed in terms of partial autocorrelations which, as they remark, are not directional and would therefore appear less appropriate for recursive (causal) models. We also use pair-wise partial autocorrelations, but conditioning on all remaining variables (i.e. components of e_t) rather just one other variable at a time. This is because such partial correlations are used to construct the conditional independence graph (CIG) of the variables, following procedures presented for example in Whittaker (1990). As Swanson and Granger also remark, the structural form of dependence between the variables is naturally expressed by (and is equivalent to) a directed acyclic graph (DAG), in which nodes representing variables are linked with arrows indicating the direction of any causal dependence. A DAG implies a single CIG for the variables, but the possible DAGs which might explain a particular CIG may be several or none. The point is that, subject to sampling variability, the CIG is a constructible quantity and a useful one for expressing the data determined constraints on permissible DAG interpretations.

The CIG consists of nodes representing the variables, two nodes being *without* a link if and only if they are independent conditional upon *all* the remaining variables. In a Gaussian context this conditional independence is indicated by a zero partial autocorrelation:

$$\rho\left(e_{it}, e_{jt} | \{e_{kt}, k \neq i, j\}\right) = 0. \tag{4}$$

In the wider linear least squares context, defining linear partial autocorrelations as the same function of linear unconditional correlations as in the Gaussian context, (4) still usefully indicates lack of linear predictability of one variable by the other given the inclusion of all remaining variables. The link with Granger causality is quite evident. The set of all such partial correlations required to construct the CIG is conveniently calculated as

$$\rho\left(e_{it}, e_{jt} | \{e_{kt}, k \neq i, j\}\right) = -W_{ij}/\sqrt{(W_{ii}W_{jj})} \tag{5}$$

where $W = V^{-1}$. The sample values are obtained by substituting the sample value $\hat{V}$ of V.

Our example uses the innovations from the series of daily dollar term rates over the period from 30th November 1987 to 12th April 1990, excluding non-trading days. The maturity terms are 6 month, 1, 2, 3, 5, 7 and 10 years. Figure 1 illustrates just the six month, two year and ten year rates. The movements in the series are clearly highly correlated as supported by Table 1 which shows the correlation matrix of the linear innovations from a well-fitting canonical ARMA(1,1) model estimated for these seven series. Table 2 shows the corresponding matrix of partial autocorrelations as defined by (5).

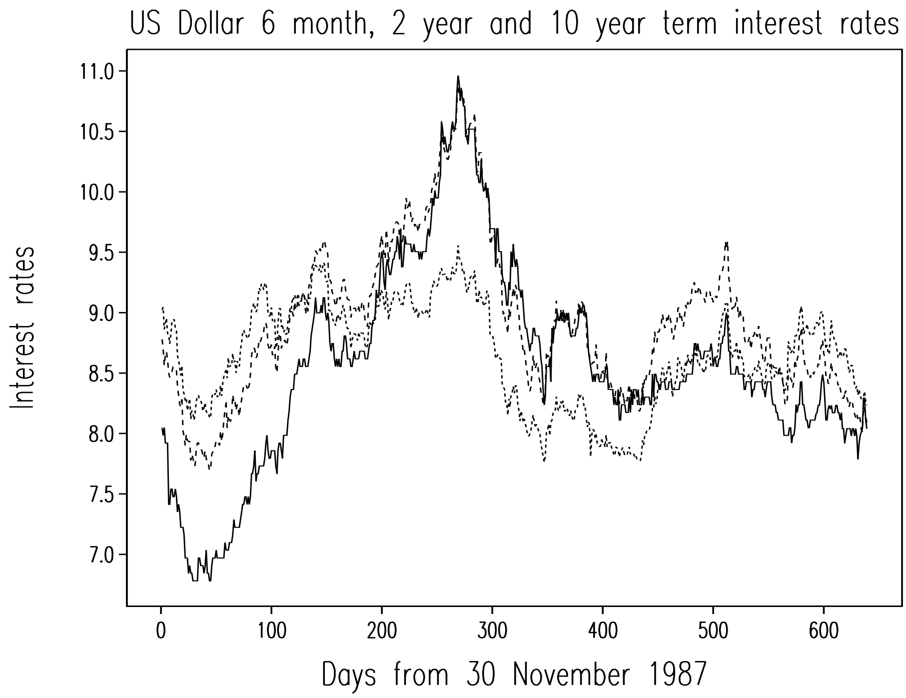


Figure 1: Six-month (solid line), two year (broken line) and ten year (dotted line) dollar term rate series.

Some of these are marked to show significance at the 5% level. The significance levels are obtained by using the relationship between a regression t value and the sample partial correlation $\hat{\rho}$ given by $\hat{\rho} = t/\sqrt{(}t^2 + \nu)$ (see Greene, 1993,p 180). Here ν is the residual degrees of freedom in the regression of one of the variables in the partial autocorrelation, upon all the other variables. The t value is that attached, in this regression, to the other variable in the partial autocorrelation. This is a relationship deriving from the linear algebra of least squares, and is not reliant upon statistical assumptions. Standard assumptions *are* needed to support the usual distribution of t under the null hypothesis that the true value of the relevant variable is zero, which is equivalent to $\rho = 0$. There are of course statistical pitfalls in applying the test simultaneously to all sample partial autocorrelations. Our attitude is similar to that advocated by Box and Jenkins (1976) for the identification, for example, of autoregressive models using time series partial autocorrelations. We use

Table 1: Correlation coefficients of dollar term rate innovations.

$e_{t,1}$	1.000						
$e_{t,2}$	0.799	1.000					
$e_{t,3}$	0.516	0.538	1.000				
$e_{t,4}$	0.502	0.515	0.944	1.000			
$e_{t,5}$	0.452	0.458	0.883	0.924	1.000		
$e_{t,6}$	0.418	0.425	0.838	0.887	0.965	1.000	
$e_{t,7}$	0.420	0.423	0.812	0.863	0.941	0.967	1.000
	$e_{t,1}$	$e_{t,2}$	$e_{t,3}$	$e_{t,4}$	$e_{t,5}$	$e_{t,6}$	$e_{t,7}$

Table 2: Partial correlation coefficients of dollar term rate innovations with * indicating significance at the 5% level.

$e_{t,1}$	1.00						
$e_{t,2}$	*0.720	1.00					
$e_{t,3}$	0.022	*0.113	1.00				
$e_{t,4}$	0.037	0.017	*0.689	1.00			
$e_{t,5}$	0.019	-0.021	*0.154	*0.270	1.00		
$e_{t,6}$	-0.045	-0.012	-0.050	0.042	*0.543	1.00	
$e_{t,7}$	0.038	0.006	-0.055	0.010	*0.115	*0.658	1.00
	$e_{t,1}$	$e_{t,2}$	$e_{t,3}$	$e_{t,4}$	$e_{t,5}$	$e_{t,6}$	$e_{t,7}$

these values to suggest possible models; after fitting these we apply more formal tests and diagnostic checks to converge on an acceptable model.

In this example the critical value for significance at the 5% level is an absolute partial correlation exceeding 0.081. In fact all the marked values exceed the 1% critical value of 0.106, but we indicate the three lower valued significant partial correlations with broken lines in Figure 2 which represents the tentative conclusions from Table 2 as a CIG. It's form is almost that of the linear structure which was investigated by Swanson and Granger. This figure was presented and discussed by Tunnicliffe Wilson (1992) but without any further modeling. Central to the interpretation of a CIG is the separation theorem. The CIG is *constructed* by *pairwise* separation of variables which are independent conditional on the remainder. The separation theorem states that if two *blocks* of variables are separated, i.e. there is no link between any member of the first block and any member of the second, then the two blocks are completely independent conditional on the remaining variables.

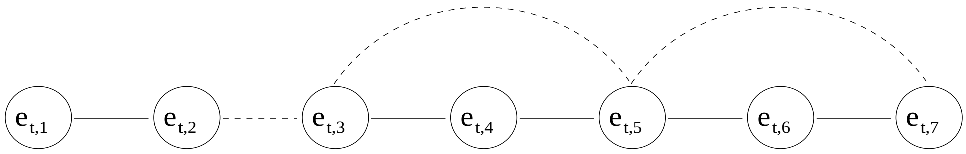


Figure 2: Conditional Independence Graph derived from Table 2 for the dollar term rate innovation series.

See for example Whittaker (1990, pp 64-67) for a general proof and references to more straightforward proofs in the Gaussian case, where the result can be read directly from the joint density. To illustrate the theorem, Figure 2 implies that {1,2} are independent of {6,7} given {3,4,5}. Using a straightforward notation for joint and conditional densities, the first part of (6) is a consequence of this CIG. Similar arguments lead to the remaining parts of (6).

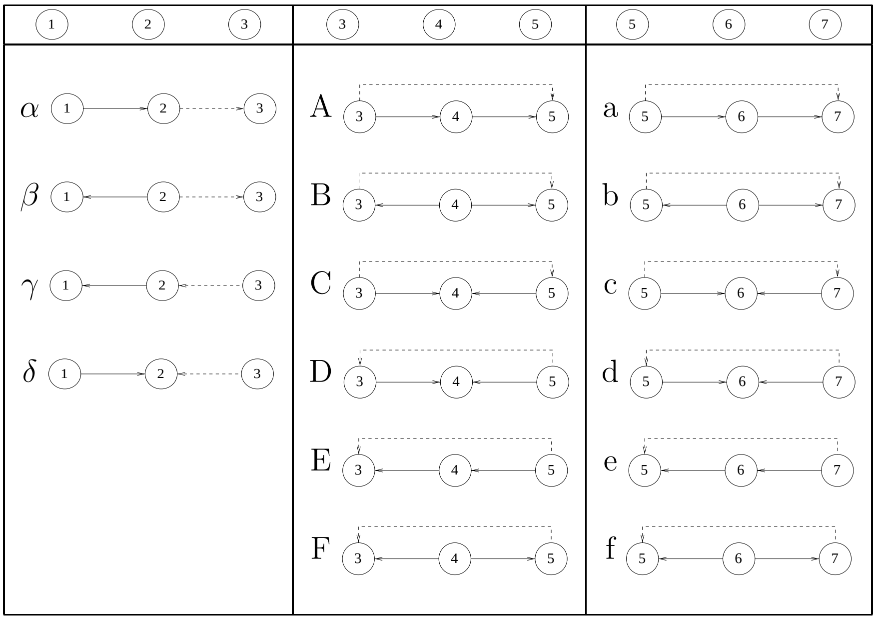
$$f_{1\ldots7} = f_{1,2|3,4,5} f_{3,4,5} f_{6,7|3,4,5} = f_{1,2|3} f_{3,4,5} f_{6,7|5} = f_{1|2} f_{2|3} f_{3,4,5} f_{6,7|5}. \tag{6}$$

This brings us to the next step which is to determine what DAG structures can explain this CIG, and to estimate them. This is part of a much wider problem of the search for causal structure, covered for example by Spirtes, Glymour and Scheines (1993). The procedure to determine the CIG implied by a given DAG has become known as *moralization*, following Lauritzen and Spiegelhalter (1988). A node B in a DAG is a *parent* of a node A if there is an arrow *from* B directly linking *to* A. Moralization is the construction of a CIG by linking (marrying), for each node of a DAG, all of its parents. The original links are retained with their directional arrows deleted. In the Gaussian context it can be seen from (3) that moralization corresponds to the creation of non-zero entries in V^{-1}, which characterizes the CIG, from the non-zero entries in Φ_0, which characterizes the DAG.

If the graph in Figure 2 were linear, with no links $e_{t,3} - e_{t,5}$ or $e_{t,5} - e_{t,7}$, there would be just seven possible DAG interpretations. These are obtained by choosing one node as a *root* or *pivot* and directing all arrows away from it. Arrows could not come together as that would create a new, moral, link. The links $e_{t,3} - e_{t,5}$ and $e_{t,5} - e_{t,7}$ introduce more possibilities. These can be listed by first considering the possible DAG interpretations of sub-graphs of Figure 2 as shown in Table 3.

Selected sub-graphs, one from each column, can be re-assembled to construct an overall DAG. This can only be done however, without creating a new moral link, provided neither nodes 3 or 5 have parents in two components. This results in 28 possible combinations:

Table 3: Possible directed subgraphs.



1)αAa; 2)αAc; 3)αCa; 4)αCc; 5)βAa; 6)βAc; 7)βCa; 8)βCc; 9)γAa; 10)γAc; 11)γBa; 12)γBc; 13)γCa; 14)γCc; 15)γDa; 16)γDb; 17)γDc; 18)γDd; 19)γDe; 20)γDf; 21)γEa; 22)γEb; 23)γEc; 24)γEd; 25)γEe; 26)γEf; 27)γFa; 28)γFc. All these models are in fact statistically equivalent, as factorizations such as 6 readily confirm. They may be estimated with full efficiency by separate regressions. For example Figure 3 shows one possible DAG, γAa, with values attached to the links which correspond to the regressions coefficients of $e_{t,1}$ on $e_{t,2}$, $e_{t,2}$ on $e_{t,3}$, $e_{t,3}$ on *none*, $e_{t,4}$ on $e_{t,3}$, $e_{t,5}$ on $\{e_{t,3},e_{t,4}\}$, $e_{t,6}$ on $e_{t,5}$ and $e_{t,7}$ on $\{e_{t,5},e_{t,6}\}$. To assess this model we use minus twice the log-likelihood, which we call the deviance. This is given by $n\sum\log\hat{\sigma}_r^2$ where $\hat{\sigma}_r^2$ are the MLEs of the residual variances from these regressions (not the bias corrected mean square errors). This may be compared with the deviance of the saturated model in which each variable is regressed upon all previous variables, or equivalently with $n\log\det\hat{V}$. It is possible that some of the CIG links might be explained by moralization, for example the link $e_{t,3}\rightarrow e_{t,5}$ in subgraph C. This was checked by fitting the model without this link and a significant increase in the deviance indicated that it should not be removed. The link between $e_{t,5}$ and $e_{t,7}$ was similarly retained. The assumption that the error covariance matrix D is diagonal is the basis of a diagnostic test applied to the sample correlations of the residuals from the model

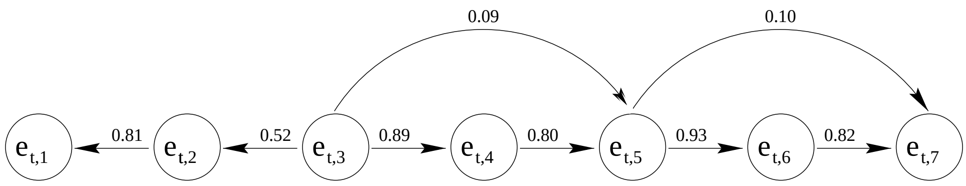


Figure 3: Graph γAa with link regression coefficients.

Table 4: Residual correlation matrix for the innovations model.

1	1.000						
2	-0.092	1.000					
3	0.143	0.000	1.000				
4	-0.044	0.025	0.000	1.000			
5	0.010	-0.071	0.000	0.000	1.000		
6	-0.025	-0.043	-0.056	0.110	0.050	1.000	
7	0.032	0.047	-0.024	0.038	0.029	0.001	1.000
	1	2	3	4	5	6	7

regressions. These are shown in Table 4 for the fitted model γCa and the correlation of 0.143 between residuals for $e_{t,1}$ and $e_{t,3}$ is significant. Consequently we introduced a further term into the model as shown in Figure 4. The correlations between the residual pairs $e_{t,1}$,$e_{t,2}$ and $e_{t,1}$,$e_{t,3}$ were both reduced to less than 0.01 with very little change to the other correlations. There remains one correlation between the residuals for $e_{t,4}$ and $e_{t,6}$ with a value of 0.11 which is just significant. We could find no simple model extension that would account for this. Table 5 provides a comparison of these two models with the

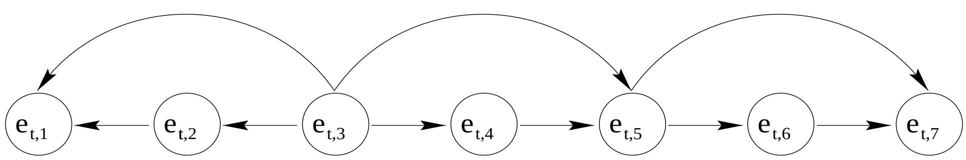


Figure 4: Graph γAa with an added link.

saturated model in terms of the deviance and three model selection criteria. For ease of interpretation the table shows the deviance and criteria with the values for the saturated model subtracted. Selective elimination of coefficients with small t values in a model may cause the deviance increase to be larger than that typical of the chi-squared distribution having degrees of freedom given by the reduction in number of parameters. The selection criteria AIC, HIC and SIC provide protection against this. They are obtained by adding to the deviance the respective *penalties* $2k$, $2k \log \log n$ and $k \log n$, where n is the series length and k is the number of fitted parameters. A negative value of a criterion favours selection of the indicated model, in comparison with the saturated model. The criteria are ordered by increasing penalty on the number of parameters. On the basis of this table and the diagnostic residual correlations, we conclude that the final fitted DAG is an acceptable parsimonious model for the structure of the innovations series. It is not unique, but limits the range of possible structures considered when the model is extended to include lagged variables.

Table 5: Comparisons of the different models.

model	reduction in no. of pars.	increase in deviance	relative AIC	relative HIC	relative SIC
γAa	13	38.63	12.63	-9.62	-44.53
γAa improved	12	21.04	-2.95	-23.50	-55.72

3 Identifying structural autoregressions

It will be usual that the order p of a canonical autoregression will have been, at least tentatively, identified for the series. Structural autoregressive model identification then proceeds by construction of the CIG using the data matrix X consisting of the collection of contemporaneous and lagged data vectors $(x_{p+1-k,i}, \ldots, x_{n-k,i})'$ for each series $i = 1, \ldots, m$ and each lag $k = 0, \ldots, p$. Assuming that the time series or data vectors have been mean corrected we use the covariance matrix estimate $\hat{V} = X'X/(n-p)$. A CIG is constructed from $\hat{V}$ in a similar manner as before, but with two differences:

1. the significance levels used are $z/\sqrt(z^2 + \nu) \approx z/\sqrt{n-p}$, where z is a critical value of the standard normal distribution.

2. we retain only those links which are significant and are either *between* contemporaneous variables or *attach to* contemporaneous variables from lagged variables.

These differences arise from the time series context, where the usual properties of regression estimation hold only in large samples and for regression on lagged values. See for example

Anderson (1971, p211). The main consequence is that we can assume only a large sample Normal distribution for the (so-called) t values in the autoregression. Also, these properties do not hold for a time series regression equation which includes both future and past regressors, because the errors are not then in general uncorrelated. A sample partial autocorrelation between $x_{t-h,i}$ and $x_{t-k,j}$ for some $0 < h, k < p$ can correspond only to a t value in the regression of one of these, say $x_{t-h,i}$ on all the other values, including at least one past and one future value. The t value for $x_{t-k,j}$ and therefore the sample partial correlation between $x_{t-h,i}$ and $x_{t-k,j}$ will not have the required properties. See however Reale and Tunnicliffe Wilson (2000) for an account of how the sampling properties may be determined in this case.

In summary, the significance levels specified in 1 can only be applied to the links specified in 2. These are however the only links we consider for selection of a structural autoregression which, viewed as a DAG, only contains such links. For a stationary VAR(p) model, the subgraph of the CIG that consists of just the links specified in this way will be unchanged if the maximum lag used in its construction is greater than the true order p, but there may be some loss of efficiency in the statistical inference.

As an example consider the structural VAR(1) model expressed in (7) and represented by the DAG in Figure 5(a), where for convenience we now refer to the three series as x_t, y_t and z_t.

$$\begin{pmatrix} 1 & 0 & 0 \\ -\phi_{210} & 1 & 0 \\ 0 & -\phi_{321} & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \phi^*_{111} & 0 & \phi^*_{131} \\ 0 & \phi^*_{221} & 0 \\ 0 & 0 & \phi^*_{331} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \\ w_t \end{pmatrix} \tag{7}$$

On division by Φ_0 model (7) is brought to the canonical form in which, it may readily be determined, the only sparseness is the zero value of ϕ_{121}. Thus, including the innovation correlations, but not their variances, 11 parameters would be required in the canonical model, whereas only 6 are required in the structural form shown. Furthermore, the two other possible structural forms, corresponding to different orderings of the contemporaneous variables, would each be completely saturated, with 12 parameters. Where a sparse structural form might exist, it is therefore worthwhile investigating. In this example, with only three possible structural orderings, each could be fitted and regression testing used to discover the most parsimonious. The graphical modeling approach does however provide immediate insight into the model selection.

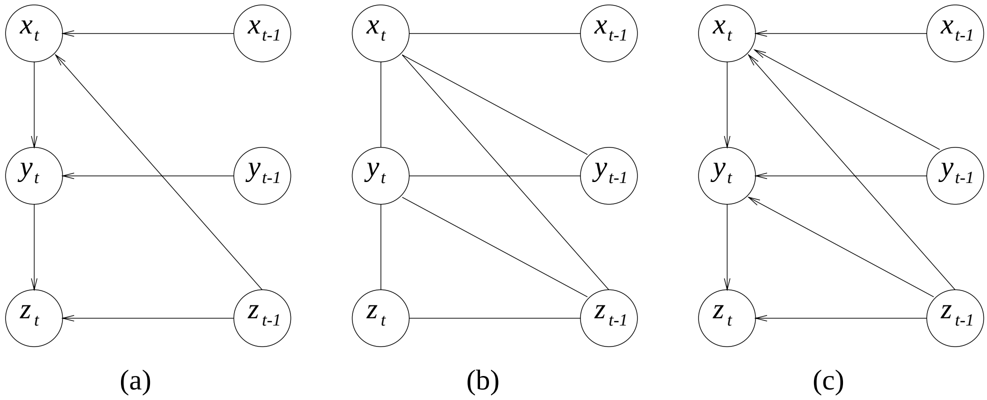


Figure 5: (a) The DAG representation of the model 7, (b) the subgraph of the CIG derived from (a), and (c) the DAG representations compatible with (b).

The CIG obtained by appropriate moralization of this time series DAG is shown in Figure 5(b). Note that, following the points made above, the moral link between x_{t-1} and z_{t-1} implied by Figure 5(a) is not shown in the CIG.

We point out now that the CIG of the innovations e_t from the canonical form of a VAR(p) model, is the same as the subgraph associated with the contemporaneous values in the CIG constructed for the series x_t. This is because the distribution of x_t conditional upon p or more past values is by definition the same as the distribution of e_t, apart from the conditional mean. For the example model (7) a linear CIG would therefore be found for the innovations from a canonical AR(1). This would allow three DAG interpretations, $x_t \rightarrow y_t \rightarrow z_t$, $x_t \leftarrow y_t \rightarrow z_t$ and $x_t \leftarrow y_t \leftarrow z_t$, with the fourth possibility, $x_t \rightarrow y_t \leftarrow z_t$, excluded because it would imply a moral link between x_t and z_t in the CIG. Now consider the possible structural VAR(1) models compatible with Figure 5(b). The possible directions attachable to links between contemporaneous variables are those just listed. We then attach the direction of the arrow of time to the remaining links, but consider the possibility that some of them may be moral links. Note first that the link $x_{t-1} \rightarrow x_t$ must be true, it cannot be created as a moral link. That implies the direction $x_t \rightarrow y_t$, else a moral link would form between x_{t-1} and y_t. Then $x_t \rightarrow y_t \rightarrow z_t$ is the only possibility for contemporaneous dependence. The position is now that shown in Figure 5(c).

The direct links $x_{t-1} \rightarrow x_t$, $y_{t-1} \rightarrow y_t$ and $z_{t-1} \rightarrow z_t$ cannot be explained as moral links. Each of the remaining links $y_{t-1} \rightarrow x_t$, $z_{t-1} \rightarrow x_t$ and $z_{t-1} \rightarrow y_t$ might be so explained but it is also compatible with Figure 5(b) that they are all causal links. Regression estimation of the model represented in Figure 5(c) will establish if they are real.

To summarize, the aims and strategy of CIG based identification are:

1. to clarify the recursive ordering of contemporaneous variables, i.e. the direction of the links between these variables. An ordering (or possibly more than one) is selected which is generally consistent with the evidence presented in the CIG, taking into account the effects of moralization.

2. for any such ordering, to determine the selection of links from lagged to contemporaneous variables. An initial model including all such links which appear in the CIG is a useful starting point. Regression estimation of the selected DAG model can be used to remove any unnecessary links and establish which might be explained by moralization.

Efficient estimation of the selected model is done by separate regressions of each contemporaneous variable on those causal variables indicated in the DAG. The overall model is again assesed by a deviance $n' \sum \log \hat{\sigma}_r^2$, where $n' = n - p$ is here the length of data vectors used in the regression, and $\hat{\sigma}_r^2$ are the MLEs of the residual variances from these regressions (not the bias corrected mean square estimates).

Progress can only be made if the CIG is relatively sparse; no discrimination of structural models can be made if it saturated. Much depends on noting the absence of links which would be present if certain contemporanous directions were not avoided. One must be wary though of building up long chains of logic based upon the statistical evidence in the CIG. As we know from section 2, model fitting reveals a link $x_{t,1} - x_{t,3}$ which Figure 2 fails to reveal. Likelihood based comparisons with a saturated model will indicate which models are plausible and may discriminate clearly between some competing models. Checks should be applied to confirm that the residuals are orthogonal innovations, and used as a possible guide to model improvement.

4 Structural Autoregressive modeling - an example

We now consider a real example of three monthly Italian monetary times series to assess the existence of the *lending channel* of the monetary transmission.

Bernanke and Blinder in 1988 proposed a model for aggregate demand which allows the existence of another monetary transmission mechanism together with the existing *money channel*.

According to the latter we can distinguish only two different assets in the market: money and bonds; every other asset is a perfect substitute of one of them. In this case an open market sale of bonds by the central bank would force the bond interest rate up because the household sector, as a matter of accounting, must hold less money and more bonds. If there is not full and instantaneous adjustment of the prices there will be a loss of money in real terms for households; this will cause eventually an increase in real interest rates which in turn can have effects on investments and real economy.

The other view is called *lending channel* or *credit channel*. According to this theory bank loans and bonds are not perfect substitutes so that we can distinguish three different assets in the market: money, bonds and bank loans. Under these conditions monetary policy can work on either bond interest rate or loan interest rate or both so that an impact on the latter can be independent from an effect on the former. An example (see Kashyap and Stein, 1993) is given by an open market sale, which, reducing banks' reserves, as a matter of accounting will make banks release less loans. If money and bonds are close substitutes there will be a minimal impact on bonds interest rate. Nevertheless the cut on loan supply will push up their cost with an influence on the real economy. In this case we have a weak money channel but a strong lending channel.

In Bernanke and Blinder's model there are three necessary conditions for the existence of the lending channel:

1. from the firms point of view intermediated loans and open market bonds must not be perfect substitutes;

2. the central bank must be able, by changing the amount of reserves of the banking system, to affect the supply of intermediated loans;

3. there must be an imperfect price adjustment to monetary policy shocks.

There is a clear evidence that the Italian economy matches at least two of the above mentioned conditions and hence provides a suitable environment to verify the existence of the lending channel. In fact as Buttiglione and Ferri (1994) pointed out:

1. Italian firms, as their balance sheets show, are funded far more by banking credit than issued bonds or commercial paper, so that they can unlikely be seen as perfect substitutes. There isn't any commercial paper market indeed;

2. during the eighties Italian banks reduced the amount of securities in their portfolios and now the adjustment is completed. Hence they can't neutralise a shock on reserves by asset management anymore.

   Moreover the lack of a secondary market for CDs has prevented banks to using liability management in response to monetary restrictions.

The third condition, concerning the speed of price adjustment, although central for any monetary economics theory, is normally less apparent and more difficult to assess.

Recently Bagliano and Favero (1998) carried out an empirical analysis to test whether the lending channel has worked in Italy. They estimated two different VAR models to investigate the transmission from the monetary policy impulse to the government bond and loan interest rate and hence to their difference. A widening of the latter imply the existence of the credit channel (see Bernanke and Blinder, 1988). It involves three variables: 1) the repurchase agreement interest rate (a), whose innovations may be viewed as monetary policy innovations; 2) the average interest rate on government bonds with residual life longer than one year (b); 3) the average interest rate on bank loans (c). With the second VAR system, of order five, they assessed the impulses of monetary policy to real economy. It includes four different variables: 1) the difference between bank loan and government bond; 2) the loan interest rate; 3) the industrial production; 4) the inflation.

The results of their analysis supported the existence of the lending channel in the Italian monetary market.

We partially used Bagliano and Favero's framework to apply our VAR model identification strategy, investigating the relationships among the variables of the first VAR system they estimated, to verify if there is a direct causal effect from a monetary policy impulse (the repurchase agreement interest rate) to the loan interest rate. To pursue our analysis we used the same monthly time series taken from the same sources (Bank of Italy) over the period January 1986-December 1993; they are shown in Figure 6.

Bagliano and Favero found that a canonical autoregressive model of order 2 adequately described the series. We follow the procedure of the previous section, refering to our three series as x_t, y_t and z_t, by first constructing the lagged data vectors for lags 0, 1, and 2 for each series. The resulting data matrix X is then used to construct the covariance matrix $\hat{V}$ from which the sample partial correlations shown in Table 6 are derived. The critical value for significance at the 5% level is 0.207. Figure 7 shows the appropriate subgraph of the CIG of the lagged variables constructed using this threshold, with the addition of

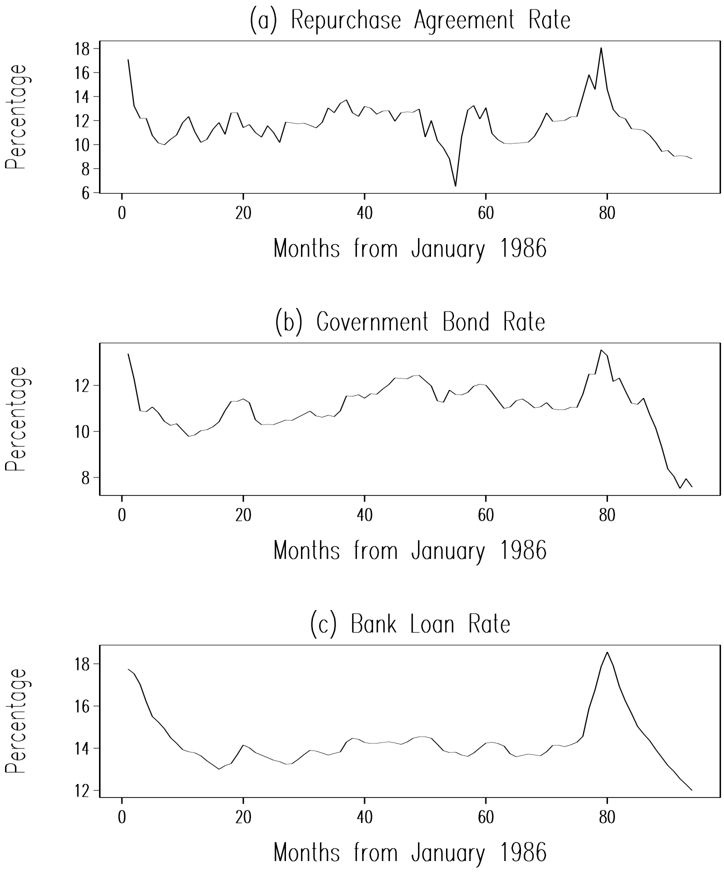


Figure 6: Three monthly Italian monetary time series: (i) the repurchase agreement interest rate, (ii) the average interest rate held on government bonds, (iii) the average interest rate on bank loans, shown over the period January 1986 to December 1993.

Table 6: Partial autocorrelations of the lending policy series.

	x_t	y_t	z_t	x_{t-1}	y_{t-1}	z_{t-1}	x_{t-2}	y_{t-2}	z_{t-2}
x_t	1.000	0.387	0.214	0.449	-0.353	-0.096	0.016	0.167	-0.033
y_t		1.000	0.364	-0.224	0.742	-0.393	0.015	-0.140	0.309
z_t			1.000	0.200	-0.080	0.876	-0.154	-0.199	-0.598

two links, $z_t - x_{t-1}$ and $z_t - y_{t-2}$ shown by broken lines. These are included because their partial autocorelations are very close to the threshold. The series are only of moderate length so that some additional power for detecting non-zero partial correlations is justified. The main point to note is the clear absence of a link $y_{t-1} \to z_t$. A moral link would be

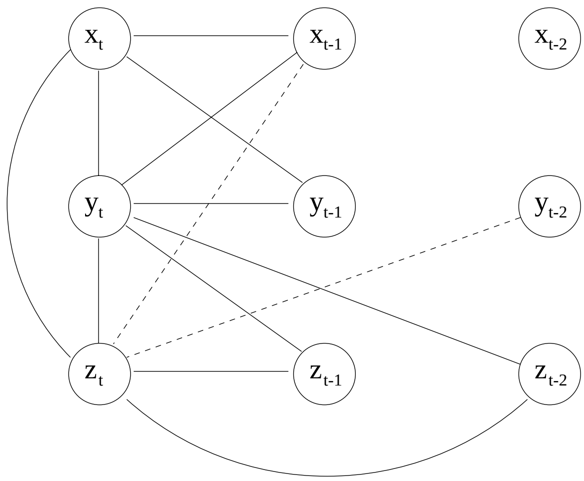


Figure 7: The subgraph of the CIG for the monetary policy series derived from table 6.

expected here unless we assign the direction between contemporaneous variables: $y_t \to z_t$. There is no such clear indication of the remaining choice of contemporaneous links. Of the three possibilities we consider first that with $x_t \to y_t$ and $x_t \to z_t$. We then fitted the DAG derived from Figure 7 by assigning these contemporaneous directions and with all the other links directed from the past to the present. The results indicated that three links could be removed: $y_{t-1} \to x_t$, $z_{t-2} \to y_t$ and $x_{t-1} \to z_t$, the occurrence of the first two of these in Figure 7 is explained by moralization. The DAG representing this as model A is shown in Figure 8 and Table 7 shows the likelihood criteria relating to this model. According to these the model appears to be quite acceptable in comparison with the saturated VAR(2) model. A further possibility, model B, was investigated by reversing the link between x_t and y_t as also shown in Figure 8. It was only possible to remove two links in this case, so the model has one more link than model A. The likelihood criteria in Table 7 for this model show that it is also acceptable. The model coefficients are shown on the graphs; their t values are all in excess of 3.0, except that for the link $x_{t-1} \to y_t$ the values for models A and B are respectively -1.87 and 2.03. Statistical criteria do not show a clear preference for one model, although economic considerations would strongly favour model A. Statistical evidence was however strongly against the reversal also of the link $x_t \to z_t$.

The main point of economic interest is the influence on the average bank loan rate of the two other series, and that is the same for both models.

Moralization of both graphs in Figure 8 yields the same CIG, but one which differs from Figure 7 by having the extra links $z_{t-1} - x_t$, $z_{t-2} - x_t$, $y_{t-2} - y_t$ and $y_{t-2} - x_t$. Only a careful study, possibly by simulation, would indicate whether we should have expected to detect these. Our general conclusion though is that study of Figure 7 lead us swiftly to the specification of a good structural model for these series.

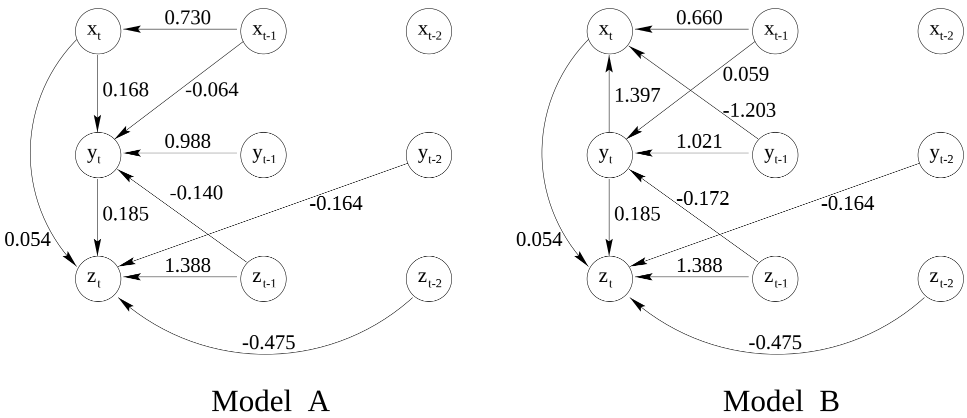


Figure 8: The subgraphs of the models A and B for the monetary policy series.

Table 7: Comparisons of structural AR models for the monetary policy series.

model	reduction in no. of pars.	increase in deviance	relative AIC	relative HIC	relative SIC
A	11	17.14	-4.86	-16.16	-32.83
B	10	14.49	-5.50	-15.78	-30.94

5 Conclusion

We have investigated how conditional independence modeling may be used in the selection of structural AR models. The example we have presented is the only one which we have so far investigated, not the pick of the best from a range. The aim has been to identify parsimonious structure, which may be valuable in various applications of the model. Our practical experiences suggest that the approach is of considerable value in achieving our stated aim. In our example the model supports and quantifies a particular lending channel hypothesis which is important for monetary policy. The methods we have used are accessible and visually appealing and we hope this work will encourage their wider application in this context.

References

Akaike, H. (1973), "A new look at Statistical Model Identification", *IEEE Transactions on Automatic Control*, **AC-19**, 716-723,

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*. Wiley: New York.

Bagliano, F.C. and Favero, A.C. (1998) "Il canale del credito della politica monetaria. Il caso Italia". in (ed. G. Vaciago) *Moneta e Finanza*, Il Mulino: Bologna.

Bernanke, B.S. and Blinder, A.S. (1988) "Credit, money and aggregate demand", *American Economic Review: Papers and Proceedings*, **78**, 435-439.

Box, G.E.P., and Jenkins, G.M. (1976), *Time Series Analysis, Forecasting and Control*, revised edition, Holden-Day: Oakland.

Box, G.E.P. and Tiao, G.C. (1977), "A canonical analysis of multiple time series", *Biometrika*, **64**, 355-365.

Buttiglione, L. and Ferri, G. (1994), "Monetary policy transmission via lending rates in Italy: any lessons from recent experience?", *Temi di Discussione*, **n. 224**.

Greene, W.H. (1993), *Econometric Analysis*, second edition, Prentice-Hall: Englewood Cliffs.

Hannan, E.J. and Quinn, B.G. (1979), "The determination of the order of an autoregression", *Journal of the Royal Statistical Society Series B*, **41**, 190-195.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H. and Lee, T.C. (1985), *The Theory and Practice of Econometrics*, second edition, Wiley: New York.

Kashyap, A.K. and Stein, J.C. (1993), "Monetary policy and bank lending", *NBER Working Paper*, **n. 4317**.

Lauritzen, S.L. and Spiegelhalter, D.J. (1988), "Local computations with probabilities on graphical structures and their applications to expert systems". *Journal of the Royal Statistical Society Series B*, **50**, 157-224.

Reinsel, G.C. (1993), *Elements of Multivariate Time Series Analysis*, Springer-Verlag: New York.

Reale, M. and Tunnicliffe Wilson, G. (2000), "The sampling properties of conditional graphs for structural vector autoregressions", Research Report n. 2000/3, Mathematics and Statistics Dept., University of Canterbury.

Spirtes, P., Glymour, C. and Scheines, R (1993), *Causation, Prediction and Search.* Springer-Verlag: New York.

Swanson, N.R. and White, H. (1995)," A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks", *Journal of Business and Economic Statistics*, **13**, 265-275

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

Schwarz, G. (1978), "Estimating the dimension of a model", *The Annals of Statistics*, **6**, 461-464.

Tiao, G.C. and Tsay, R.S. (1989), "Model specification in multivariate time series", *Journal of the Royal Statistical Society Series B*, **51**, 157-213.

Tunnicliffe Wilson, G. (1992), "Structural models for structural change", *Quaderni di Statistica e Econometria*, **14**, 63 - 77.

Whittaker, J.C. (1990), *Graphical Models in Applied Multivariate Statistics*, Wiley: Chichester.

Zellner, A. and Palm, F. (1974), Time series analysis and simultaneous equation econometric models, *Journal of Econometrics*, **2**, 17-54.