

Constructing structural VAR models with conditional independence graphs

Les Oxley^a, Marco Reale^{b,*}, Granville Tunnicliffe Wilson^c

^a Economics Department, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

^b Mathematics and Statistics Department, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

^c Mathematics and Statistics Department, University of Lancaster, Lancaster LA1 4YF, United Kingdom

Received 16 July 2008; received in revised form 4 November 2008; accepted 9 November 2008

Available online 3 December 2008

Abstract

In this paper graphical modelling is used to select a sparse structure for a multivariate time series model of New Zealand interest rates. In particular, we consider a recursive structural vector autoregressions that can subsequently be described parsimoniously by a directed acyclic graph, which could be given a causal interpretation. A comparison between competing models is then made by considering likelihood and economic theory.

© 2008 IMACS. Published by Elsevier B.V. All rights reserved.

Keywords: Graphical models; Directed acyclic graphs; Term structure; Causality

1. Introduction

Technology has impacted extensively on the operations of financial markets which are inhabited by a rich array of fixed-income securities, each bearing a particular rate of interest. The relationship between the yields on these various securities is the province of the term structure of interest rates literature which has a long history and can be traced-back formally to [19]. The more recent literature has been reviewed in [23].

With the popularity of cointegration and VAR/SVAR approaches to estimation in econometrics, a separate literature using these approaches to estimate and test term structure models and implications can be identified and includes [2,5,9,10,15,16,24]. The papers are typically motivated by a concern to understand the term structure for the “related monetary policy control” issues and focus either upon technical estimation issues and often the validity of inferences derived including, importantly, causal inference, the effects of structural change or the testing of various hypotheses. Causality is a particularly important and popular issue given the role of monetary policy intervention.

In this paper we wish to add a significant extra dimension to the debate by using graphical modelling to identify causal mechanisms within multivariate time series models. There is a small, but growing literature which uses directed acyclic graphs in economics and includes [28], which examines the impact of monetary policy on agricultural prices in four Asian economies [7], which examines relationship between money, income, nominal prices and wheat prices

* Corresponding author. Tel: +64 3 364 2987x7681; fax: +64 3 364 2587.

E-mail addresses: les.oxley@canterbury.ac.nz (L. Oxley), marco.reale@canterbury.ac.nz (M. Reale), g.tunnicliffe-wilson@lancaster.ac.uk (G.T. Wilson).

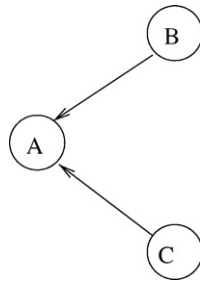


Fig. 1. Directed acyclic graph.

[8], which examines causal relationship between trade, government savings, natural resource exports and GDP in 79 world economies and [6], which examines the dynamics of regional, export-wheat, railroad rates linking several central US regions to Texas Gulf ports. This paper considers for the first time, an application to the term structure of interest rates where little consensus seems to exist on the causal nexus and direction between long and short rates of interest. In particular, there are three alternative views on causality; short rates “cause” long rates (broadly the traditional “Expectations Hypothesis” view); long rates “cause” short rates (here rational inflation expectations have a role); or the “market segmentation” [11], or “preferred habitat” approaches, where causality is discontinuous across maturity periods. The outcome in an empirical sense will be crucial for the efficacy of monetary policy design and implementation.

2. Graphical modelling

Graphical modelling (GM) is a relatively new statistical approach, whose initial ideas were proposed by Dempster [14] and later developed by Darroch [12]. The major attraction of the approach in empirical research is its convenient way to present pairwise relationships between random variables taken from a multivariate context.

The initial step in the approach is the computation of the partial correlations between the variables in the particular multivariate system under study. Once the numerical values are known we can test their significance by using an opportune statistic. Finally the results are presented as a graph, where the random variables are represented by nodes and a significant partial correlation between two random variables is denoted by a line that links them named *edge*. If the variables in the graph are jointly distributed as a multivariate Gaussian distribution, a significant partial correlation implies the presence of conditional dependence. For this reason the graph is called a conditional independence graph (CIG).

A more informative object in GM is the directed acyclic graph (DAG). This is a directed graph where there are arrows linking the nodes and where the joint distribution of the variables can be expressed as a sequence of marginal conditional distributions. By way of example, consider the graph in Fig. 1: its joint density function can be defined as $f(a, b, c) = f(a|b, c)f(b)f(c)$.

Although the DAG and the CIG represent a different definition of the joint probability, there is a correspondence between the two which is embodied by the moralization rule [22]: because of this result we can obtain the CIG from the DAG by transforming the arrows into lines and linking unlinked parents with *moral* edges. By way of example consider the graph in Fig. 2: *A* and *B* are the *parents* of *C*. The moralization of the DAG on the left is obtained by transforming the existing arrows into edges and by adding a *moral* edge which links the parents.

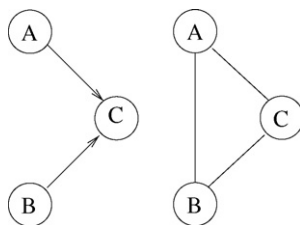


Fig. 2. Moralization of a directed acyclic graph.

While the CIG represents the associations among the variables either in terms of conditional dependence or simply in terms of partial correlation if the joint distribution is not Gaussian, the DAG has a natural interpretation in terms of causality. As it is not the aim of this paper to enter into philosophical debate about the definition of causality, we refer to the main contributions on the causality implied by directed acyclic graphs [30,17,20,25,31,21].

The DAG is very attractive because of its causal implications but in practice all we can observe is the CIG constructed by the sample partial correlations. In order to obtain the DAG from the CIG we have to apply the inverse operation of the moralization, we name it *demoralization*. Unfortunately while the transformation of a DAG into a CIG is unique, the inverse operation of identification and removal of moral edges is not. To this end we need to use all the information we have about the relationships among the random variables in the system.

In this paper we apply this process within the context of multivariate structural VAR models considering first its *saturated* specification, where there are links between every pair of variables (including the contemporaneous variables), with the aim of finding a parsimonious model of the transmission mechanism between interest rates in New Zealand.

3. Multivariate time series

The relationship between several autoregressions can be modelled via the vector autoregression

$$x_t = c + \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_k x_{t-k} + e_t \quad (1)$$

of order k , VAR(k), where x_t, \dots, x_{t-k} are n -dimensional vectors of observations with coefficient vectors Φ_1, \dots, Φ_k , c is the constant and e_t is the error vector, which is assumed IID. If the covariance matrix, H , of e_t is not diagonal, the set of linear Eq. (1) corresponds to a system of seemingly unrelated regressions [33] where the relations among the components of x_t are hidden in H . To highlight such relations we can represent the canonical VAR(k) in (1) in its structural form (SVAR):

$$\Theta_0 x_t = d + \Theta_1 x_{t-1} + \Theta_2 x_{t-2} + \cdots + \Theta_k x_{t-k} + u_t \quad (2)$$

where $\Theta_i = \Theta_0 \Phi_i$ for $i = 0, \dots, k$, $d = \Theta_0 c$ and $u_t = \Theta_0 e_t$ with covariance matrix $\Theta_0 H \Theta_0' = D$, which is diagonal.

If there are no zeros in the coefficient vectors, the SVAR is saturated, but in many cases some variables do not play any significant role in explaining the current variables, x_t . In this case the value of the corresponding coefficient is zero and hence the SVAR is sparse. An examination of the covariance matrix of the variables involved, both current and lagged, can assist in identifying the sparse structure by the computation of the partial correlations. Their significance can be tested using the appropriate sampling properties [27,26,32]. The model (2) may be represented by a DAG in which the components of $x_t, x_{t-1}, \dots, x_{t-p}$ form the nodes, and causal dependence is indicated by arrows linking nodes. The nature of the model is that all arrows end in nodes representing the contemporaneous variables on the left hand side of (2). Some arrows will start from past values, and some from other contemporaneous variables.

The coefficients can be estimated by single equation ordinary least squares (OLS) regression which is fully efficient under the assumption that the vector series is Gaussian but is also applicable and the properties of the estimates reliable, under wider conditions, such as e_t being I.I.D.

Next consider the exploratory tools used to identify the model. The first step is to identify the overall order p of a VAR model for the series. The second and central step is to construct a sample conditional independence graph (CIG) for the variables $x_t, x_{t-1}, \dots, x_{t-p}$ which form the nodes of the graph. At this stage the only causality we can assume is the one indicated by the arrow of time. Nevertheless, it may serve well to suggest the direction of dependence between contemporaneous variables. Such dependence can be simply interpreted in terms of predictability. If we wish to go beyond the concurrence of realizations of different time series, from our point of view of observers of aggregated data, we can use the statistical evidence to support a causal belief. This leads to the connection between causality and probability discussed by De Moivre [13]. A discussion of this argument and of the debate around it, although at the very foundation of this work as well as many others in statistics and econometrics, cannot be contained in an article and we point the interested reader to two recent monographs by Shafer [30] and Pearl [25]. Causal beliefs can be implemented in our procedure when we hypothesize possible directed subgraphs for the edges between contemporaneous variables. The corresponding structural VAR models are then fitted and refined by regression and a model selection criterion such as AIC [1], used to select the best in terms of likelihood.

The statistical procedures are based on a data matrix X which in the general case consists of $m(P + 1)$ vectors of length $n = N - P$, composed of elements $x_{i,t-u}$, $t = P + 1 - u, \dots, N - u$, for each series $i = 1, 2, \dots, m$, and each lag $u = 0, 1, \dots, P$, for some chosen maximum lag P . In the first stage of overall order selection, for each order p we fit, by OLS, the saturated structural VAR regressions of the m contemporaneous (lag 0) vectors on all the vectors up to lag p . Using the sums of squares S_i from these regressions we form the AIC as $n \sum \log S_i + 2k$, where $k = pm^2 + m(m - 1)/2$ is the total number of regression coefficients estimated in the regressions. For the saturated model the causal order of the contemporaneous variables does not affect the result, each one is included only as a regression variable for a subsequent variable in the chosen ordering. Then select the order p which minimizes the AIC.

The next step is to construct the sample CIG for the chosen model order p . In general a CIG is an undirected graph, defined by the *absence* of a link between two nodes if they are independent, conditional upon *all* the remaining variables. Otherwise the nodes are linked. In a Gaussian context this conditional independence is indicated by a zero partial autocorrelation:

$$\rho(x_{i,t-u}, x_{j,t-v} | \{x_{k,t-w}\}) = 0, \quad (3)$$

where the set of conditioning variables is the whole set up to lag p , excluding the variables on the left.

The set of all such partial correlations required to construct the CIG is conveniently calculated from the inverse W , of the covariance matrix V of the whole set of variables, as

$$\rho(x_{i,t-u}, x_{j,t-v} | \{x_{k,t-w}\}) = -\frac{W_{rs}}{\sqrt{(W_{rr}W_{ss})}} \quad (4)$$

where r and s respectively index the lagged variables $x_{i,t-u}$ and $x_{j,t-v}$ in the matrices V and W .

In the wider linear least squares context, defining linear partial autocorrelations as the same function of linear unconditional correlations as in the Gaussian context, the absence of a link still usefully indicates a lack of linear predictability of one variable by the other given the inclusion of all remaining variables.

To estimate the CIG we replace V with the sample covariance matrix \hat{V} formed from the data matrix X , but including only lags up to p . From here we need a statistical test to decide which links are absent in the graph. We are only concerned with links between contemporaneous variables and between contemporaneous and lagged variables, because these are the only ones that appear in the structural model DAG. The test we use is to retain a link when $|\rho| > z/\sqrt{(z^2 + v)} \approx z/\sqrt{n - p}$, where z is an appropriate critical value of the standard normal distribution. This derives from two results. The first is the standard, algebraic, relationship between a sample partial correlation $\hat{\rho}$ and a regression t value given by $\hat{\rho} = t/\sqrt{(t^2 + v)}$ (see Greene [18, p. 180]). The second is the asymptotic normal distribution of the t value for time series regression coefficients, given for example by Anderson [3, p. 211]. Generally, we might wish to apply multiple testing procedures when applying the test simultaneously to all sample partial autocorrelations, but that is not a practical option. The application of GM to VAR systems has been extended by demonstrating that the sampling properties of GM's for stationary VAR's are still valid for for I(1) VAR processes [32].

We then specify the DAG's as recursive equation systems which can be estimated by ordinary least squares.

The next stage in the process is to establish which DAG representations are consistent with the CIG or are nearly so, allowing for statistical uncertainty, considering *demoralization*.

As we mentioned above by this term we mean the inverse operation of moralization which allows to construct a CIG from a given DAG by inserting an undirected link between any two nodes a and b when there is another node c with incoming directed edges $a \rightarrow c$ and $b \rightarrow c$. In this case c is known as a common child of a and b , and the insertion of a new, moral, link will marry the parents. After this operation for the whole graph, the directions are removed from the original links.

Of course we attach the arrow of time to links from the past to the present, so the challenge is to clarify the directions of the recursive ordering of contemporaneous variables. Normally there are alternative competitive models and eventually we compare them by using likelihood-based methods.

4. Interest rate transmission in New Zealand

We apply the methodology explained in the previous sections to the interest rate mechanism in New Zealand after the implementation of the Reserve Bank Act in February 1990. To this aim we consider the model proposed by Oxley [24]

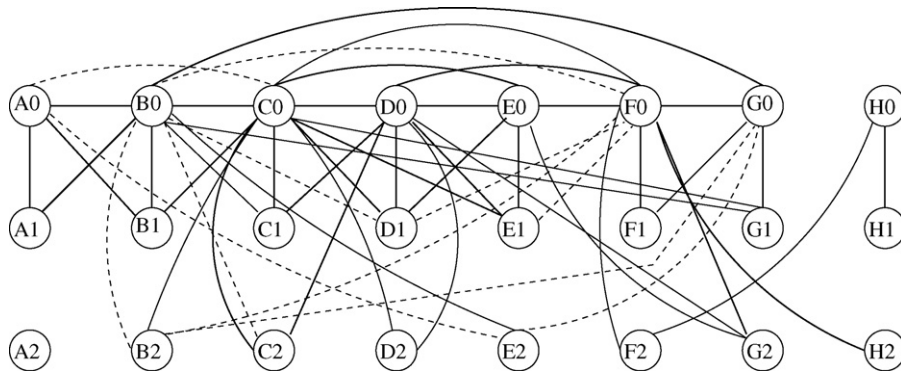


Fig. 3. Conditional independence graph.

who identified a structural VAR using standard procedures. The paper by Oxley also provides a thorough discussion of the economic background for the interested reader.

The application is of interest to the economists as the issues involved for this New Zealand case are multi-faceted and involves the presence of indirect effects.

The data used are monthly, seasonally unadjusted interest rates taken from the Reserve Bank of New Zealand Financial Statistics database for the period February 1990–April 2002. The individual series considered are the rates on money at call (denoted A); 90 day bank bills (B); the yield on 1, 3 and 5 year Government stock (C–E, respectively); base lending rate (F) first mortgage housing rate (G) and the uncovered interest parity with the US (H).

We identified a VAR(2) and hence considered all the variables up to the second lag. Once the sample partial correlation matrix was computed we tested with the appropriate procedures explained above the significance of its elements and constructed the CIG in Fig. 3, where the variables at time t are indicated as A_0, B_0, \dots, H_0 , the variables at time $t - 1$ as A_1, B_1, \dots, H_1 and the variables at time $t - 2$ as A_2, B_2, \dots, H_2 .

We then considered several DAG models consistent with the CIG. Such initial models were obtained by substituting undirected edges linking lagged variables to contemporaneous variables with directed edges. For the edges between contemporaneous variables, several models with different directions suggested by the alternative economic theories were considered. The different models thus constructed were refined by using subset regression to eliminate the moral links.

Finally we used likelihood based measures to compare the different models. In particular we considered the Akaike information criterion and the Bayesian information criterion (BIC) [29].

Here we present the two best models together with Table 1 providing their values in terms of parameters, deviance and the different information criteria. They are represented in Figs. 4 and 5.

For both of them we can observe some common features as the lack of relevance of the uncovered interest parity and the central role of the 90 day bank bills interest rate. Both models would support a market segmentation hypothesis confirming what the Reserve Bank of New Zealand recognised: “The transmission mechanism is multi-faceted, and the ultimate effect of central bank operations on the end objective of price stability is very indirect. Partly because the linkages are so indirect, public and market expectations of future policy changes play a very important role in monetary policy. In effect, they can short-cut, bringing the beginning and end of the transmission mechanism closer together.” Archer et al. [4].

Table 1
Information criteria.

Model	k	AIC	BIC
Best	42	−97.85	−424.75
Alternative	37	−96.87	−438.11

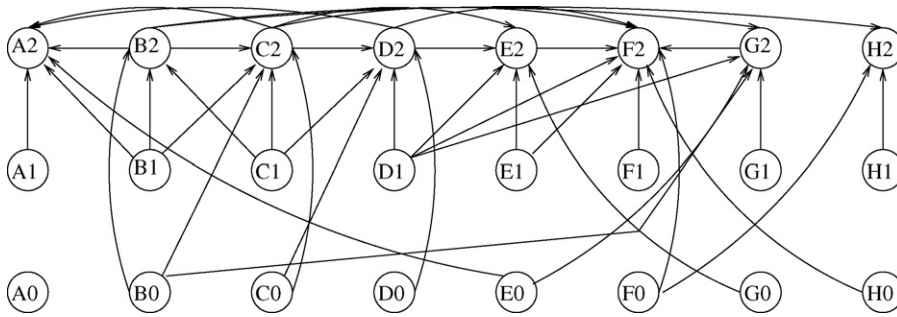


Fig. 4. Best model.

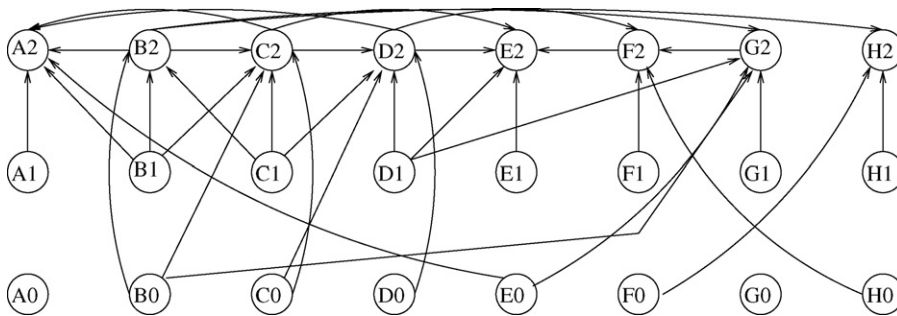


Fig. 5. Alternative model.

Acknowledgements

The authors wish to thank two anonymous referees for their helpful comments.

References

- [1] H. Akaike, A new look at statistical model identification, *IEEE Transactions on Automatic Control* AC-19 (1973) 716–723.
- [2] G. Amisano, M. Cesura, C. Giannini, M. Seghelini, The transmission mechanism among Italian interest rates, *Statistica* 58 (1997) 25–50.
- [3] T.W. Anderson, *The Statistical Analysis of Time Series*, John Wiley & Sons, New York, 1971.
- [4] D. Archer, A. Brookes, M. Reddell, A cash rate system for implementing monetary policy, *Reserve Bank of New Zealand Bulletin* 62 (1999) 51–61.
- [5] F.-C. Baglioni, C.A. Favero, Measuring monetary policy with VAR models: an evaluation, *European Economic Review* 42 (1998) 1069–1112.
- [6] D.A. Bessler, S.W. Fuller, Railroad wheat transportation markets in the Central Plains: modeling with error correction and directed graphs, *Transportation Research Part E* 36 (2000) 21–39.
- [7] D.A. Bessler, S. Lee, Money and prices: U.S. data 1869–1914 (a study with directed graphs), *Empirical Economics* 27 (2002) 427–446.
- [8] D.A. Bessler, N. Loper, Economic development: evidence from directed acyclic graphs, *The Manchester School* 69 (2001) 457–476.
- [9] C. Bruneau, E. Jondeau, Long-run causality with an application to international links between long-term interest rates, *Oxford Bulletin of Economics and Statistics* 61 (1999) 545–568.
- [10] M. Camarero, J. Ordóñez, C. Tamarit, An analysis of the transmission mechanism of monetary policy in Spain using a structural cointegrated VAR approach, *Applied Economics* 34 (2002) 2201–2212.
- [11] K. Cuthbertson, The Expectations Hypothesis of the Term Structure: The UK Interbank Market, *The Economic Journal* 106 (1996) 578–592.
- [12] J.N. Darroch, S.L. Lauritzen, T.P. Speed, Markov fields and log-linear interaction models for contingency tables, *Annals of Statistics* 8 (1980) 522–539.
- [13] A. De Moivre, *The Doctrine of Chances*, W. Pearson, London, 1718.
- [14] A.P. Dempster, Covariance selection, *Biometrics* 28 (1972) 157–175.
- [15] K. Drakos, The term structure of deviations from the interest parity, *Journal of International Financial Markets, Institutions and Money* 13 (2003) 57–67.
- [16] T. Engsted, K. Nyholm, Regime shifts in the Danish term structure of interest rates, *Empirical Economics* 25 (2000) 1–13.
- [17] C. Glymour, G.F. Cooper, *Computation, Causation and Discovery*, MIT Press, Cambridge, MA, 1999.
- [18] W.H. Greene, *Econometric Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1993.
- [19] J.M. Keynes, *A Treatise on Money*, Harcourt, Brace and Co., New York, 1930.

- [20] S.L. Lauritzen, Causal inference from graphical models, in: D.R. Cox, C. Klüppelberg (Eds.), *Complex Stochastic Systems*, Chapman & Hall, London, 2001.
- [21] S.L. Lauritzen, T.S. Richardson, Chain graph models and their causal interpretations, *Journal of the Royal Statistical Society Series B* 64 (2002) 321–361.
- [22] S.L. Lauritzen, D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their applications to expert systems, *Journal of the Royal Statistical Society B* 50 (1988) 157–224.
- [23] A. Melino, The term Structure of interest rates: evidence and theory, *Journal of Economic Surveys* 2 (1988) 335–366.
- [24] L. Oxley, Identifying an interest rate transmission mechanism for New Zealand, mimeo, 2000.
- [25] J. Pearl, *Causality*, Cambridge University Press, Cambridge, England, 2000.
- [26] M. Reale, G. Tunnicliffe Wilson, The sampling properties of conditional graphs for structural vector autoregressions, *Biometrika* 89 (2002) 457–461.
- [27] M. Reale, G. Tunnicliffe Wilson, Identification of vector AR models with recursive structural errors using conditional independence graphs, *Statistical Methods and Applications* 10 (2001) 49–65.
- [28] S. Saghaian, M. Hasan, M. Reed, Monetary impacts and overshooting of agricultural prices: the case of four asian economies, *Journal of Agricultural and Applied Economics* 34 (2002) 95–109.
- [29] G. Schwarz, Estimating the dimension of a model, *Annals of Statistics* 6 (1978) 461–464.
- [30] G. Shafer, *The Art of Causal Conjecture*, MIT Press, Cambridge, MA, 1996.
- [31] P. Spirtes, C. Glymour, R. Scheines, *Causation, Prediction and Search*, MIT Press, Cambridge, MA, 2000.
- [32] G. Tunnicliffe Wilson, M. Reale, The sampling properties of conditional independence graphs for I(1) structural VAR models, *Journal of Time Series Analysis* 29 (2008) 802–810.
- [33] A. Zellner, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association* 57 (1962) 348–368.