Constructing Structural VAR Models with Conditional Independence Graphs

[1]**Les Oxley** , [2]**Marco Reale and** [3]**Granville Tunnicliffe Wilson**

[1]Department of Economics, University of Canterbury,
Private Bag 4800, Christchurch, New Zealand. E-Mail: les.oxley@canterbury.ac.nz
[2]Department of Mathematics and Statistics, University of Canterbury,
Private Bag 4800, Christchurch, New Zealand
[3]Department of Mathematics and Statistics, University of Lancaster,
Lancaster LA1 4YF, United Kingdom

ABSTRACT

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 Keynes.

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 has developed.

Here 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.

This paper considers 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*, 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.

1 GRAPHICAL MODELLING

Graphical modelling (GM) is a relatively new statistical approach, whose initial ideas were proposed by Dempster (1972) and later developed by Darroch *et al.* (1980). 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 or (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.

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 (Lauritzen and Spiegelhalter, 1988): 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.

While the CIG represents the associations among the variables either in terms of conditional dependence or simply in terms of partial correlation, the DAG has a natural interpretation in terms of causality (Pearl, 2000).

The DAG is a very attractive because of its causal interpretation but in practice all we can observe is the CIG implied 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.

2 MULTIVARIATE TIME SERIES

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

$$
\begin{aligned}
x_t &= c + \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \\
&\quad + \ldots + \Phi_k x_{t-k} + e_t
\end{aligned} \tag{1}
$$

of order k, VAR(k), where $x_t, \ldots, x_{t-k}$ are n-dimensional vectors with coefficient vectors $\Phi_1, \ldots, \Phi_k$, 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 equations (1) corresponds to a system of seemingly unrelated regressions (Zellner, 1962) 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):

$$
\begin{aligned}
\Theta_0 x_t &= d + \Theta_1 x_{t-1} + \Theta_2 x_{t-2} + \\
&\quad + ldots + \Theta_k x_{t-k} + u_t
\end{aligned} \tag{2}
$$

where $\Theta_i = \Theta_0 \Phi_i$ for $i = 0, \ldots, k$, $d = \Theta_0 c$ 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 lagged variables on the RHS in (2) do not play any 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 (Reale and Tunnicliffe Wilson, 2002 and 2001). The model (2) may be represented by a directed acyclic graph (DAG) in which the components of x_t, x_{t-1}, $\ldots$, 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, $x_{t-1}, \ldots, x_{t-p}$ 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. The corresponding structural VAR models are then fitted and refined by regression and a model selection criterion such as AIC (Akaike, 1973), 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 $x_{i,t-u}$, $t = P+1-u, \ldots N-u$, for each series $i = 1, 2, \ldots, m$, and each lag $u = 0, 1, \ldots, P$, 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:

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

where the set of conditioning variables on the right 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

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

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 $|\rho| > z/\sqrt{(z^2 + \nu))} \approx z/\sqrt{n-p}$, 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 $\hat{\rho} = t/\sqrt{(t^2 + \nu)}$ (see Greene, 1993 p 180). The second is the asymptotic normal distribution of the t value for time series regression coefficients, given for example by Anderson (1971, 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 (Tunnicliffe Wilson and Reale, 2002).

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.

3 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 (2000) 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, D and 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
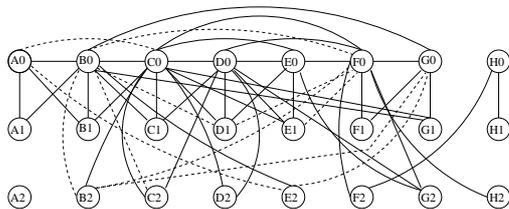explained above the significance of its elements and constructed the CIG in figure 1.

Figure 1. Conditional independence graph.

We then considered all the models consistent with the CIG and used subset regression to eliminate the moral links.

The final step was to used likelihood based measures to compare the different models. In particular we considered the Akaike information criterion and the Bayesian information criterion (BIC) (Schwarz, 1978).

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

Table 1. Information criteria.

Here we present the two best models together with a table providing their values in terms of parameters, deviance and the different information criteria. They are represented in figure 2 and 3.
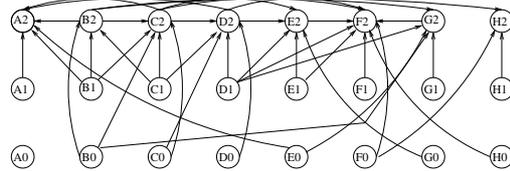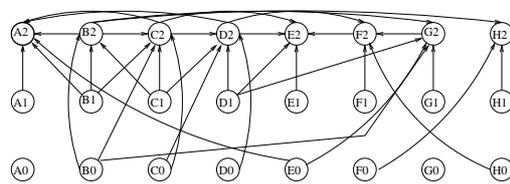

Figure 2. Best model.

Figure 3. Alternative model.

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.

4 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. John Wiley & Sons Ltd, New York.

Darroch J.N., S.L. Lauritzen and T.P. Speed (1980), Markov fields and log-linear interaction models for contingency tables, *Annals of Statistics*, **8**, 522-539.

Dempster, A.P. (1972), Covariance selection, *Biometrics*, **28**, 157-175.

Greene, W.H. (1993), Econometric Analysis. Prentice-Hall, Englewood Cliffs New Jersey.

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

Oxley, L. (2000), Identifying an interest rate transmission mechanism for New Zealand, mimeo.

Pearl, J. (2000), Causality. Cambridge University Press, Cambridge England.

Reale, M. and G. Tunnicliffe Wilson (2002), The sampling properties of conditional graphs for structural vector autoregressions, *Biometrika*, **89**, 457-461.

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**, 49-65.

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

Tunnicliffe Wilson, G. and M. Reale (2002), Causal diagrams for I(1) structural VAR models, University of Canterbury Mathematics and Statistics Department Research Reports, **UCDMS2002/6**.

Zellner, A. (1962), An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association*, **57**, 348-368.