18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Graphical models for structural VARMA representations

Oxley, L.[1], **M. Reale**[2], **and G. Tunnicliffe Wilson**[3]

[1] *Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand*
[2] *Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*
[3] *Department of Mathematics and Statistics, University of Lancaster, Lancaster, UK*
Email: marco.reale@canterbury.ac.nz

Abstract: Sparse structural VAR representation can effectively be identified by using graphical modeling. In this paper we extend this approach to the the identification of sparse structural VARMA representations. We illustrate our methods with an application to a set of three monthly flour price series that has been the subject of previous approaches to structural VARMA modeling. We compare and contrast structural VARMA(1,1) and VAR(2) models for this data.
Keywords: *Conditional independence, moralization, multivariate time series.*

1 INTRODUCTION

In Reale and Tunnicliffe Wilson (2001) (hereafter referred to as RTW) we showed how conditional independence graphs could be used for the identification of vector AR models with recursive structural errors. In this paper we extend this approach to the identification of sparse structural VARMA representations. To describe the method we focus on a particular illustration, the modelling of a set of monthly flour price series in three cities, Buffalo, Minneapolis and Kansas City, over the period from August 1972 to November 1980. These are taken from Tiao and Tsay (1989), who used them to illustrate a new approach to VARMA model specification. They have since been used also by Grubb (1992) in a comparative investigation of multiple time series modeling, and more recently by Athanasopoulos and Vahid (2008), who developed the procedure of Tiao and Tsay. In terms of simplicity of procedure and interpretation, goodness of fit and parsimony of parameterisation, our model compares very well with the models developed in these earlier approaches. These previous studies have developed VARMA(1,1) models that can be called *structural*, because they include relationships between current variables. This is in contrast to what we shall call the canonical VARMA(1,1) model, in which current variables are related only to past variables. Our own approach is to identify a model with such structrual relationships using conditional independence graphs based upon partial correlations between current and lagged variables. In this short paper we will take the orders of the VARMA process for the flour price data set as given, to be first order in both autoregressive and moving average parts. We will comment in future work on the process of order selection, which was, of course, a central and important aspect of the work of Tiao and Tsay. We shall also compare and contrast our results from constructing a structural VARMA(1,1) (or SVARMA(1,1)) model for these series, with the results of fitting a structural VAR(2) (or SVAR(2)) model. The series are plotted in Figure 1 with the first and last series offset, respectively, lower and higher than the second series, by 20 units. This is so as to reveal the very similar movements in these series, which makes the construction of a model representing their dependence particularly demanding.

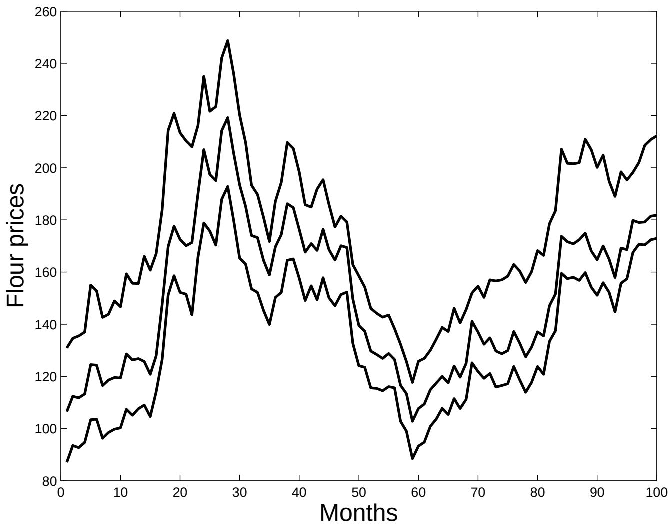

Figure 1: The flour price series.

2 CONSTRUCTION OF AN SVAR MODEL

We write the canonical VAR(p) model for a stationary, zero mean, m-dimensional time series $x_t = (x_{1,t}, x_{2,t}, \dots, x_{m,t})'$ as:

$$x_t = \Phi_1^* x_{t-1} + \Phi_2^* x_{t-2} + \cdots + \Phi_p^* x_{t-p} + e_t. \quad (1)$$

This model represents linear dependence of current values of the series upon past values alone. The error term e_t is the linear innovation, or vector of *canonical* innovations (or residuals) of the series; a multivariate white noise process having variance matrix Σ.

The SVAR(p) model for x_t, proposed in RTW, is of the form

$$\Phi_0 x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + a_t \quad (2)$$

where again a_t is multivariate white noise with variance matrix D. We make two requirements of this model

(i) the components of x_t may be re-ordered so that the coefficient matrix Φ_0 is upper triangular with unit diagonals;

(ii) the variance matrix D is diagonal.

Such a model specification arises naturally under the causal sufficiency assumption, (Spirtes, Glymour and Scheines, 2000), where, in the ordered form, the model describes how each current component of x_t depends on one or more of the current components which are subsequent in the ordering, and upon past values of any of the components. The components of a_t are then known as the *orthogonal* innovations (or residuals). In proposing the SVAR model, our aspiration is that a representation may be found which is attractive in terms of parsimony of parameterisation and simplicity of interpretation.

Any structural model of the form (2) may be transformed, on dividing through by Φ_0, to the canonical form (1) in which

$$\Sigma^{-1} = \Phi_0' D^{-1} \Phi_0, \tag{3}$$

but in general this transformation does not have a unique inverse - there may be more than one SVAR model that is plausibly consistent with the canonical VAR. However, the canonical form is useful for identifying the model order by use of an information criterion such as the AIC (Akaike, 1973). For the flour price series, this order was found to be $p = 2$. As described in RTW, the first step in determining a plausible SVAR representation is to draw the sample conditional independence graph (CIG) between the variables x_t, x_{t-1}, $\ldots x_{t-p}$ based upon the sample partial correlations between the components of these variables. An efficient method for doing this, taken from Whittaker (1990), may be found in RTW. The resulting graph is shown in Figure 2. The links between the variables represented by *nodes* of this graph indicate partial correlations between the variables that are significant at the 1% level. Where there is no link between a pair of nodes, the partial correlation has, in this example, been found to be *not* significant at the 10% level.

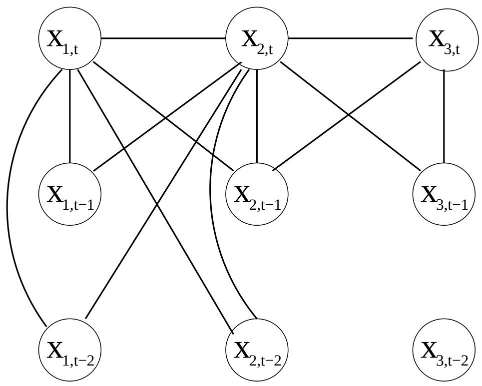

Figure 2: The estimated CIG constructed for the flour price series.

From this CIG we may hypothesize a small number of SVAR models that have the potential to represent the relationships between the series in an efficient and meaningful manner. The model upon which we finally fixed is represented in Figure 3 by a directed acyclic graph (DAG), in which the links naturally represent the predictive effects of contemporary and past variables upon each contemporary variable. The magnitude and significance of the coefficients in the corresponding SVAR model are shown adjacent to the links.
The structure of this model, i.e. the links shown in Figure 3, was selected as one which was consistent with the CIG in Figure 2. This is checked by applying the moralization rule of Lauritzen and Spiegelhalter (1988), by which we can form the CIG that is implied by the model. This rule is to insert an undirected link between any two nodes a and b for which there is a node c with directed links both $a \to c$ and $b \to c$. In this case c is known as a common child of a and b, and the insertion of a new, moral, link as *marrying* a and b, which are known as the parents of c. After doing this for the whole graph the directions are removed from the original links, which remain, together with the new moral links. It is readily checked that on applying this rule Figure 2 follows from Figure 3. In the model represented by Figure 3, the links from x_{t-1} to x_t and from $x_{3,t-1}$ to x_t were initially included, as they appear in Figure 2, but they were found not to be significant, and can be explained as arising by moralization. There is one other consistent model in which the link between $x_{1,t}$ and $x_{2,t}$ is reversed, but this requires the introduction of a link from $x_{1,t-1}$ to $x_{1,t}$.

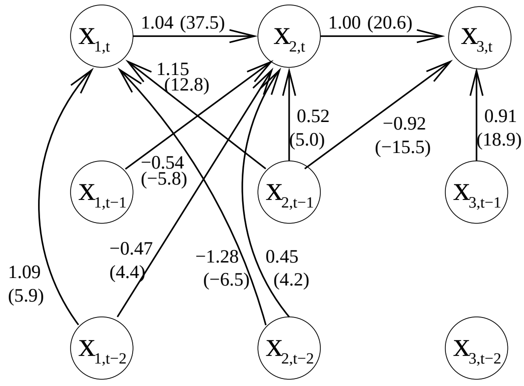

Figure 3: The DAG representing a parsimonious SVAR(2) for the flour price series, with estimated model coefficients, and t values in brackets, adjacent to the links.

3 CONSTRUCTION OF AN SVARMA MODEL

We now consider how a structural VARMA(1,1) model may be formulated and constructed for these series, following ideas that were first explored in Tunnicliffe Wilson, Reale and Morton (2001). We will write the canonical VARMA(1,1) model as

$$x_t = \Phi_1^* x_{t-1} + e_t - \Theta_1^* e_{t-1}, \tag{4}$$

and the structural VARMA(1,1) model as

$$\Phi_0 x_t = \Phi_1 x_{t-1} + a_t - \Theta_1 a_{t-1}, \tag{5}$$

where the white noise series e_t and a_t have the same properties as before, with variance matrices again related, in terms of Φ_0, by (3). The same conditions also apply to Φ_0, but when we transform (5) to (4), we need also to note that

$$\Theta_1^* = \Phi_0^{-1}\,\Theta_1\,\Phi_0. \tag{6}$$

For comparison of (5) with (2), we set out the latter taking $p = 2$, as

$$\Phi_0 x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-1} + a_t, \tag{7}$$

and note two further points:

1. Although the coefficient Φ_0 should be comparable in the two models, the interpretation of Φ_1 may be quite different. To see this, substitute $a_{t-1} = \Phi_0\, x_{t-1} - \Phi_1\, x_{t-2} + \Theta_1\, a_{t-2}$, to write (5) as

$$\Phi_0 x_t = (\Phi_1 - \Theta_1\,\Phi_0)\, x_{t-1} + \Theta_1\,(\Phi_1 x_{t-2} - \Theta_1 a_{t-2}) + a_t. \tag{8}$$

The second term on the right of (8) is an implicit combination of x_{t-2}, x_{t-3}, ... that is comparable with the second term on the right of (7), so that Φ_1 in (7) can be expected to be comparable with $\Phi_1 - \Theta_1\,\Phi_0$ in (5). The elements of Φ_1 in the two models can therefore be expected to be distinct.

2. The variables x_t, x_{t-1} and x_{t-2}, which are needed to construct the CIG appropriate to the SVAR model (7), are directly observed. However, the correspond variables for the SVARMA (5) are x_t, x_{t-1} and a_{t-1}, of which only the first two are directly observed. Moreover a_{t-1} can only be constructed following identification of the coefficient Φ_0 in the model.

This second point is handled by the procedure that we now describe. We note that a_{t-1} is a linear function of the innovations e_{t-1}, which may be substituted in their place when constructing the CIG, *for the purpose of identifying the links between x_t and x_{t-1} in the SVARMA*. Conditioning on a_{t-1} is equivalent to conditioning on e_{t-1}. Consistent estimates of a_t may be obtained by fitting the canonical VARMA(1,1) to the series, or even by fitting an approximating VAR(p) model of fairly low order. We used the former of these two, to construct the sample CIG shown in Figure 4.

This is of exactly the same form as in the CIG for the SVAR(2) model, shown in Figure 2, and leads to our initially formulating a DAG representation that includes the same links between x_{t-1} and x_{t-2}, as for the SVAR(2)

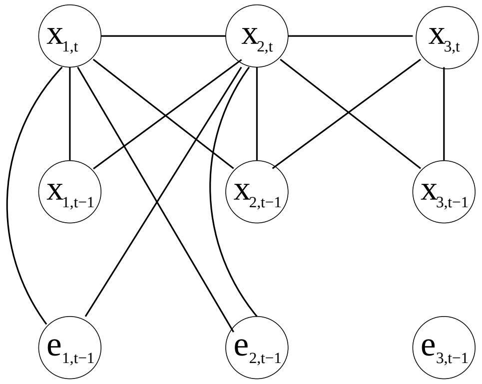

Figure 4: The estimated CIG constructed for the flour price series and innovations.

model. The next step was to estimate the SVARMA(1,1) model corresponding to this DAG, but including *all* links between the components of a_{t-1} and x_t, i.e. estimating all the elements of Θ_1. This is because links identified between e_{t-1} and x_t in the CIG of Figure 4 cannot, at this stage, inform us of the presence or absence of the links between a_{t-1} and x_t in the SVARMA(1,1) model. However, on estimation of this somewhat over-parameterized model, we obtain consistent estimates of a_t and have the choice of constructing a further sample CIG between x_t, x_{t-1} and a_{t-1}, or, as we did in this example, simply remove sequentially those elements of Θ_1 which proved not to be significant following estimation. This resulted in the fitted SVARMA(1,1) represented by the DAG in Figure 5.

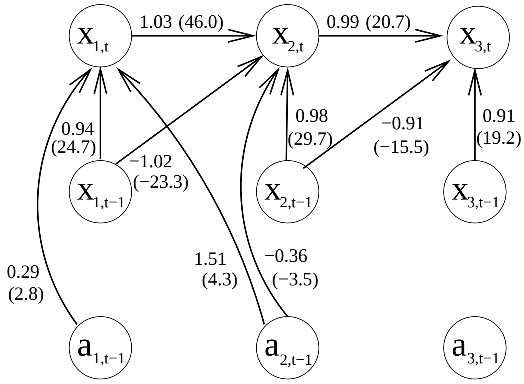

Figure 5: The DAG representing a parsimonious SVARMA for the flour price series, with estimated model coefficients, and t values in brackets, adjacent to the links.

The main point to note about this model is that the parameters are strongly determined and the structure is simple and readily interpretable:

- Series $x_{3,t}$ is essentially equal to $x_{2,t}$, but adjusted for a slightly discounted factor (0.91) of the difference $x_{3,t-1} - x_{2,t-1}$ at the previous month.
- Series $x_{2,t}$ might be considered to be similarly dependent upon $x_{1,t}$, but can also be interpreted as equal to its past value $x_{2,t-1}$ adjusted by the latest change $x_{1,t} - x_{1,t-1}$. Additionally, it is adjusted by its own previous (orthogonal) innovation $a_{2,t-1}$, with a factor of -0.36. An interpretation of this, is that the difference $x_{2,t} - x_{1,t}$ is predicted as an exponentially weighted moving average of its past values, with discount factor 0.36.
- According to this model, it is series $x_{1,t}$ that drives the other two. It follows a highly correlated univariate ARMA(1,1) process, but with the addition of a substantial adjustment by $a_{2,t-1}$, the *previous* orthogonal innovation, or *shock*, in $x_{2,t}$.

The orthogonal innovation series $a_{1,t}$, $a_{2,t}$ and $a_{3,t}$ have very low contemporaneous correlations and their lagged cross-correlations are consistent with multivariate white noise. The standard errors of these series are respectively 6.97, 1.89 and 3.56, with the first of these being, understandably, the largest, because, unlike the other two series, $x_{1,t}$ is predicted only from past values.

Our final model is parameterized by 10 prediction coefficients and three innovation variances. The model of Tiao and Tsay (1989) has 7 significant estimated coefficients plus 9 empirically determined coefficients of a contemporaneous transformation, besides 5 non-zero innovation variance elements. The model of Athanasopoulos and Vahid (2008) has 11 significant estimated coefficients, including just 1 in the contemporaneous transformation matrix, besides 6 non-zero innovation variance elements. We therefore claim that our model compares very favorably, in terms of efficiency of parameterization and simplicity of interpretation, with these previously published models.

REFERENCES

Akaike, H. (1973), A new look at statistical model identification, *IEEE Transactions on Automatic Control, AC-19, 716-723.*

Athanasopoulos, G., and F. Vahid (2008), A complete VARMA modelling methodology based on scalar components. Journal of Time Series Analysis, 29, 533-554.

Grubb, H. (1992) A multivariate time series analysis of some flour price data. Applied Statistics, 41, 95107.

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.

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.

Spirtes, P., C. Glymour, and R. Scheines (2000), Causation, prediction and search, MIT Press, Cambridge, MA.

Tunnicliffe Wilson, G., M. Reale, and A.S. Morton (2001), Developments in multivariate time series modeling, Estadistica, 53, 353-395.

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

Whittaker, J.C. (1990), Graphical models in applied multivariate statistics, Wiley, Chichester.