doi:10.1111/j.1467-9892.2008.00583.x

THE SAMPLING PROPERTIES OF CONDITIONAL INDEPENDENCE GRAPHS FOR $I(1)$ STRUCTURAL VAR MODELS

By Granville Tunnicliffe Wilson and Marco Reale

Lancaster University and University of Canterbury

First Version received June 2007

Abstract. Structural vector autoregressions allow dependence among contemporaneous variables. If such models have a recursive structure, the relationships among the variables can be represented by directed acyclic graphs. The identification of these relationships for stationary series may be enabled by the examination of the conditional independence graph constructed from sample partial autocorrelations of the observed series. In this article, we extend this approach to the case when the series follows an $I(1)$ vector autoregression. For such a model, estimated regression coefficients may have non-standard asymptotic distributions and in small samples this affects the distribution of sample partial autocorrelations. We show that, nevertheless, in large samples, exactly the same inference procedures may be applied as in the stationary case.

Keywords. Conditional independence; moralization; multivariate time series.

1. INTRODUCTION

In Reale and Tunnicliffe Wilson (2002), we considered the pth-order vector autoregressive model [VAR(p)] of a stationary m-dimensional time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,m})'$ in the form:

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

where d allows for a non-zero mean of x_t and a_t is an independent multivariate (white-noise) process with variance matrix D. We make two requirements for this model:

(i) the elements 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 *et al.*, 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. Then the components of a_t are known as the *orthogonal*

innovations (or residuals), and the model may be described as a structural vector autoregression (SVAR).

Given a sample $x_1, x_2, \ldots, x_T$, our aim is to find, by empirical methods, a model (1), whose form and statistical properties coincide with those described above, and which is also sparse, i.e. has a relatively small number of coefficients.This model should provide a good representation of the observed statistical properties of the data, in particular, its sample second-order moments (autocorrelations). The model (1) may be transformed to the canonical form of a VAR:

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

which does not include dependence among contemporaneous variables, by dividing through by Φ_0. The error term, $e_t = \Phi_0^{-1} a_t$, is the linear innovation, or vector of *canonical innovations (or residuals)* of the series, having variance matrix Σ related to D by

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

The transformation from model (1) to (2) is in general many to one (unless Φ_0 is diagonal). Corresponding to each possible ordering of the series, one may determine, by Choleski factorization, a different matrix Φ_0 which satisfies eqn (3). Given a model in the form (2), inverting the transformation then leads to a set of models having identical distributional properties, each of which is of the form of the SVAR (1). It is not possible to distinguish, by statistical means, which of these models might represent the true dependency between the series. However, identification of a small number of models of this form which are also sparse, may assist in the process of selecting a plausible structural representation which is consistent with prior knowledge of the system generating the observations. For convenience of terminology we will refer to any model of the form (1), and satisfying the conditions (i) and (ii), as an SVAR.If the number of series is small, each of the different orderings of contemporaneous dependence could be explored directly, to find, by subset regression methods, those structural models which can represent the series using sparse coefficient matrices, but for a larger number of series, this would not be very practical.

Statistical procedures, which are part of the methodology of graphical modelling, can potentially enhance this search. The statistical properties of these procedures depend on the assumed structural form of the model used to represent the series; the validity of the causality assumption is not required to establish these properties. The model (1) 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 recursive dependence is indicated by directed edges linking nodes. The nature of the model is that all edges end in nodes representing the contemporaneous variables on the left-hand side of eqn (1). Some edges will start from the past, and some from other contemporaneous variables. As an illustration, Figure 1 shows the DAG representation of a seven-variable SVAR(1) based on a model for daily US dollar term interest rates presented in Tunnicliffe Wilson *et al.* (2001). The

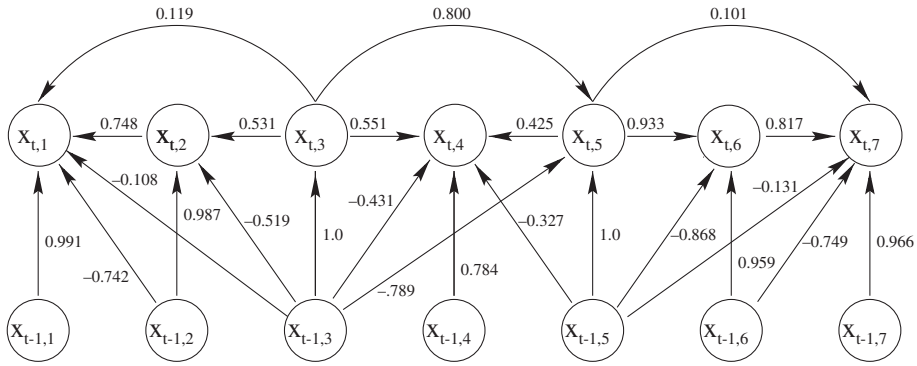

FIGURE 1. The DAG representation of the simulated structural VAR(7) model.

VAR model coefficients are inserted adjacent to the edges describing the dependence. A link is absent where there is no dependence, i.e. the coefficient is zero. The model is completely represented by this diagram, in which no links are shown between past variables, though, strictly,the same pattern of links should be repeated between all lagged (and future) variables.To directly identify such a DAG for a given time series would require estimating a multitude of possible models. To limit this search we first use statistical methods for identifying an undirected *conditional independence graph* (CIG) on the same nodes as the DAG. For a given structural VAR, this CIG may be derived by application of the *moralization* rule of Lauritzen and Spiegelhalter (1988), to the links shown in the corresponding DAG. Figure 2 shows the CIG for the model represented in Figure 1. Although the CIG does not in general uniquely characterize the DAG from which it was derived, it may dramatically constrain the number of possibilities. An important point is that, for Gaussian variables, the existence of a link between two nodes in the CIG may be identified with the existence of a non-zero partial autocorrelation between the corresponding pair of variables, conditional upon all other variables.

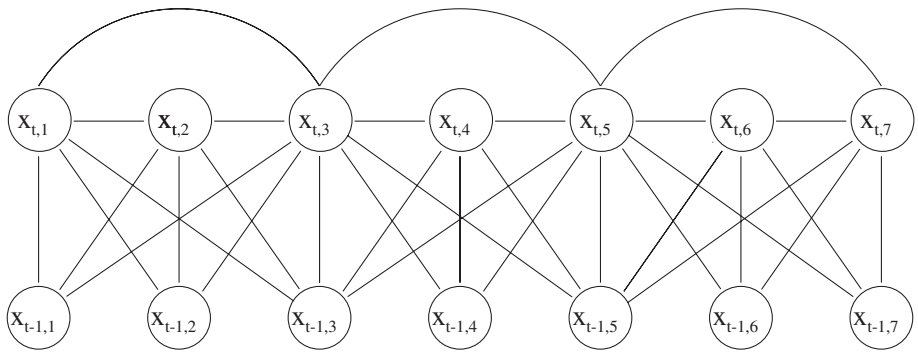

FIGURE 2. The CIG derived from the DAG representation of the simulated structural VAR(7) model.

The value of identifying the structural model (1) lies not only in the advantage of a sparse representation. Swanson and Granger (1997) apply a method of residual orthogonalization to the innovations e_t estimated from fitting a canonical VAR, which is 'meant to help in the specification of a sensible,data-determined ordering of the errors'.They add that 'the ordering of the variables is essential to the interpretation of impulse response functions and forecast error variance decompositions'. The CIG of these innovations is readily identified with the subgraph of the complete VAR model, obtained by restriction to the contemporary variables. By using additional information on links between the contemporaneous and lagged values, our procedure can often uniquely characterize the contemporaneous ordering or reduce the possibilities to a small number, when the CIG constructed solely from canonical residuals permits many possible orderings.

The use of sample lagged partial autocorrelations of the time-series data in testing for the absence or presence of a link between two nodes in the CIG, is set out in Reale and Tunnicliffe Wilson (2001). The distributional theory on which this is based, is presented in Reale and Tunnicliffe Wilson (2002). In this article, we seek to extend our methodology to the case when the vector of series is $I(1)$, so that the first difference of x_t is $I(0)$, but the series may have non-trivial co-integration. In that case the autocorrelations, and therefore the partial autocorrelations, between lagged variables are not strictly defined, and test statistics for relationships between variables may have non-standard asymptotic distributions. Nevertheless, methods based essentially on the screening of sample partial autocorrelations have been applied to series that are generally considered to be integrated or co-integrated when subject to standard tests. For example, Awokuse and Bessler (2003) construct a structural VAR model of U.S. Economic Indicators using a search algorithm to derive the graphical representation. Oxley *et al*. (2004) present an example of modelling New Zealand financial indicators. The main point of this article is to show that exactly the same inference procedures, as are applied in the stationary, $I(0)$ case, are justified for $I(1)$ data in large samples. We will illustrate the results by simulations from the model represented in Figure 1; real examples may be found in the papers previously cited.

2. EXTENSION TO $I(1)$ PROCESSES

Let x_t follow the model (1) with the further condition that the process is Gaussian, and consider the DAG and CIG with nodes corresponding to the elements of $x_t,x_{t-1},\ldots,x_{t-p}$. In the CIG, a link between a contemporaneous element $x_{t,i}$ and any other element $x_{t-h,j}$, $0 \leq h \leq p$, is missing if and only if the partial correlation between $x_{t,i}$ and $x_{t-h,j}$ is zero, given all the other variables in the graph. In the given context, this is in fact equivalent to the condition that the coefficient $\Phi_{h,i,j}$ of $x_{t,i}$ upon $x_{t-h,j}$ is zero in the minimum variance linear predictor of $x_{t,i}$ from all

other contemporaneous and lagged variables. This latter condition provides a more useful definition of the CIG corresponding to eqn (1), when we allow the process to be $I(1)$ rather than stationary. The reason is that autocorrelation functions, in terms of which partial autocorrelations (and their sample values) are naturally defined, only exist in the stationary context, whereas the prediction coefficients are always well defined and well estimated. The moralization rule for the DAG corresponding to eqn (1), can also be directly and readily verified using the CIG based upon this definition, without requiring stationarity.

To determine from a sample $x_1, x_2, \ldots, x_N$, whether the stated link is missing, we can therefore test the hypothesis that $\Phi_{h,i,j} = 0$. The following result provides the required distributional properties for this test. We will assume that the true order p of the process is known, otherwise it may be determined from the saturated canonical VAR in eqn (2), by the use of a model selection criterion.

THEOREM 1. *Let x_t follow an $I(1)$ cointegrated structural VAR(p) process of the form given by eqn* (1), *and form the data matrix X consisting of $m(p+1)$ vectors of length $n = N - p$, composed of elements $x_{t-u,i}$, $t = p+1, \ldots, N$, for each series $i = 1,2, \ldots, m$ and each lag $u = 0, 1, \ldots, p$. Perform the standard OLS regression of the vector corresponding to $x_{t,i}$ in X, upon all other vectors of X and form the usual t value for the estimate of the coefficient $\Phi_{h,i,j}$ of the vector corresponding to $x_{t-h,j}$. Then, under the hypothesis that $\Phi_{h,i,j} = 0$, the t statistic has an asymptotic standard normal distribution.*

The proof is presented in the Appendix. We remark that the regression t values for estimated coefficients in $I(1)$ VAR(p) models do not generally have an asymptotic normal distribution. The proof of the theorem establishes that the estimate of $\Phi_{h,i,j}$ has a component with a non-standard distribution, but that this is always dominated in large samples by a component which is asymptotically normal. Simulations presented in Section 3 illustrate this effect. Our final point in this section is to note that implementation of this test does not require the fitting of regressions with each contemporaneous value $x_{t,i}$, taken in turn as the response. Instead, construct a whole set of empirical partial correlations in exactly the same manner as described for the sample partial correlations in the stationary case, by Reale and Tunnicliffe Wilson (2001):

$$\hat{\rho}(x_{t-u,i}, x_{t-v,j} \,|\{x_{t-w,k}\}) = \frac{-W_{hl}}{\sqrt{(W_{hh}W_{ll})}}; \quad h \neq l, \tag{4}$$

where $W = V^{-1}$ is the inverse of the sample covariance matrix formed from X, and h and l respectively index the lagged variables $x_{t-u,i}$ and $x_{t-v,j}$ in the matrices V and W. The algebraic relationship between regression t-values and empirical partial correlations, expressed as $t = \hat{\Phi}/SE(\hat{\Phi}) = v\hat{\rho}/\sqrt{1-\hat{\rho}^2}$, may then be inverted to yield that $t > z$ if $|\hat{\rho}| > z/\sqrt{z^2 + v} \approx z/\sqrt{N}$, where z is an appropriate critical value of the standard normal distribution and v is the residual degrees of freedom in the regression. The empirical partial autocorrelations may then be used to carry out the tests, exactly as for the stationary case.

3. AN ILLUSTRATIVE SIMULATION

To illustrate the previous result we simulated the $I(1)$ cointegrated sparse structural VAR(1) model of seven series: $\Phi_0 x_t = \Phi_1 x_{t-1} + a_t$, which is represented in Figure 1. The coefficients of Φ_0 and Φ_1 are given in the diagram, and the standard deviations of the components of a_t are, in order, 0.0457, 0.0646, 0.0782, 0.0197, 0.0336, 0.0178 and 0.0164. This was derived from a model for seven term interest rate series presented in Tunnicliffe Wilson *et al.* (2001). It is capable of producing simulated series that are very similar to those of the real-term rates. However, we have slightly adjusted, to the value of unity, the coefficients that relate $x_{t,3}$ and $x_{t,5}$ to their respective lagged values $x_{t-1,3}$ and $x_{t-1,5}$. Each series $x_{t,i}$ is therefore individually $I(1)$, but the vector series has five cointegrating vectors, i.e. there are two $I(1)$ components and five $I(0)$ components of x_t. The CIG derived by moralization of Figure 1 is shown in Figure 2.

We selected 10 pairs of variables that are not linked in Figure 2 and computed the empirical partial correlation coefficient between these pairs for 10,000 replications of simulated series of length both 100 and 600. The proportions of values exceeding the 5% critical threshold defined by our procedure, are shown in Table I, for the length 600 series in the row labelled 'A', and for the length 100 series in the row labelled 'B'. For the results from the length 600 simulations, the proportions lie generally within the range of variability that is expected asymptotically, with the overall proportion being 0.0508. However, the results from the length 100 simulations show a consistent excess proportion, which overall is 0.0604. This is consistent with the fact, given in the proof of the theorem, that in small samples the regression 't' values may contain a component with a non-standard distribution associated with $I(1)$ autoregression. A further length 100 simulation set was carried out, in which Φ_1 is multiplied by 0.5, so that the process was $I(0)$. The results are displayed in the row labelled 'C' of the same table, and show a small excess proportion (overall 0.0523), but much less than that for the $I(1)$ model. These results agree with the large sample theory established in the article.

TABLE I

THE PROPORTIONS OF SELECTED EMPIRICAL PARTIAL CORRELATIONS THAT EXCEED CRITICAL THRESHOLDS IN 10,000 SIMULATIONS

Variable 1	$x_{t,2}$	$x_{t,3}$	$x_{t,4}$	$x_{t,2}$	$x_{t,3}$	$x_{t,4}$	$x_{t,4}$	$x_{t,4}$	$x_{t,5}$	$x_{t,6}$
Variable 2	$x_{t,4}$	$x_{t,6}$	$x_{t,6}$	$x_{t-1,4}$	$x_{t-1,6}$	$x_{t-1,1}$	$x_{t-1,2}$	$x_{t-1,6}$	$x_{t-1,2}$	$x_{t-1,3}$
A	0.044	0.056	0.051	0.050	0.055	0.050	0.046	0.053	0.048	0.055
B	0.061	0.056	0.061	0.059	0.063	0.059	0.062	0.060	0.063	0.060
C	0.054	0.053	0.050	0.050	0.052	0.057	0.054	0.049	0.051	0.053

Rows A and B show results for the $I(1)$ model represented in Figure 1 with simulated series of lengths 600 and 100 respectively. Row C shows results for an $I(0)$ model and series length 100.

4. CONCLUSION

We have shown that the methodology for using conditional independence graphs to identify structural VAR models, may, in large samples, be extended unchanged to the case of $I(1)$ time-series data. The method may be used to identify contemporaneous relationships, which can be of importance for proper interpretation of the impact of contemporaneous shocks on the future behaviour of the series.

APPENDIX

PROOF OF THEOREM. Without loss of generality consider inference for the coefficient $\Phi_{h,1,j}$, taking the response variable in the regression to be $x_{t,1}$. The proof builds on the development in Reinsel (1993, *pp* 165–8), which is based on a result of Ahn and Reinsel (1990). Some modifications and notational changes are made to incorporate the structural form of model. A 1:1 linear transformation can be made from the parameters $\Phi_0, \ldots, \Phi_p$ of eqn (1) to a new set of parameters $C_1, C_2, \varphi_0, \ldots, \varphi_{p-1}$ in an error correction form of the model:

$$w_t = x_t - x_{t-1} = C_1 z_{1,t-1} + C_2 z_{2,t-1} + \varphi_0 w_t + \sum_{j=1}^{p-1} \varphi_j w_{t-j} + a_t, \tag{5}$$

where $z_{1,t}$ is a d-dimensional $I(1)$ series and $z_{2,t}$ is an r-dimensional $I(0)$ series; $r = m - d$ being the co-integrating rank of x_t. In terms of the new parameters, $\Phi_0 = I - \varphi_0$, $\Phi_k = \varphi_k - \varphi_{k-1}$ for $k = 2, \ldots, p-1$ and $\Phi_p = -\varphi_{p-1}$. However, the remaining coefficient, Φ_1, is expressed as

$$\Phi_1 = \varphi_1 + I - \varphi_0 + C_1 Q_1 + C_2 Q_2 \tag{6}$$

where Q_1 and Q_2 are full-rank matrices conforming to the dimensions of C_1 and C_2 in eqn (5). They are defined by the partition $Q' = (Q_1' Q_2')$ of the matrix Q in the Jordan canonical form of

$$\sum_{k=1}^{p} \Phi_0^{-1} \Phi_k = P \begin{pmatrix} I_d & 0 \\ 0 & \Lambda_r \end{pmatrix} Q, \tag{7}$$

where $P = Q^{-1}$. We consider eqn (6) as defining a linear relationship, determined by the matrix functions Q_1 and Q_2 of the true (but unknown) model, between the coefficient Φ_1 (and the previously defined φ_1 and φ_0) and the two new coefficients C_1 and C_2. This relationship also applies to the estimates of these same model coefficients and is used to derive their properties. However, from eqn (7) we may establish that the true value of C_1 is zero, because (following Reinsel, 1993, p. 167), on partitioning $P = (P_1, P_2)$, we can express

$$C_1 = \left(\Phi_0 - \sum_{k=1}^{p} \Phi_k \right) P_1.$$

The properties of the estimated coefficients in eqn (5), are based on the fact that the regressors $z_{1,t-1}$ are purely $I(1)$ and the remainder are $I(0)$. Because the true value of C_1 is zero, we have $\hat{C}_1 = O_p(N^{-1})$ and

$$\left(\hat{C}_2 - C_2, \hat{\varphi}_0 - \varphi_0, \ldots, \hat{\varphi}_{p-1} - \varphi_{p-1}\right) = O_p(N^{-\frac{1}{2}}). \tag{8}$$

Furthermore, the joint distribution of the estimated coefficients in eqn (8) is the same as for an $I(0)$ model, so standard OLS inference is valid in large samples. In particular, regression 't' values are asymptotically standard-normal. Apart from Φ_1, the estimates of Φ_j are linear transformations of those in eqn (8), so are the same as for an $I(0)$ model. For $\hat{\Phi}_1$ we refer to eqn (6). The magnitude of $\hat{C}_1$ is $O_p(N^{-1})$, so provided that the particular element of $\hat{\Phi}_1$ in which we are interested also contains components from $\hat{\varphi}_0$, $\hat{\varphi}_1$ or $\hat{C}_2$, these, having standard error of magnitude $O_p(N^{-\frac{1}{2}})$ will dominate the distribution of the estimate for large N, justifying the use of the $I(0)$ inference.

The only case when the first two of these components are not present arises when $p = 1$, so that φ_1 is not present, and when the parameter of interest is $\Phi_{1,1,1}$, because there is no component $\Phi_{0,1,1}$ in the structural model. We complete the proof by showing that even in this final case, under our null hypothesis (which becomes $\Phi_{1,1,1} = 0$), $\hat{\Phi}_{1,1,1}$ will always contain a component from $\hat{C}_2$. The component of $\hat{\Phi}_{1,1,1}$ contributed by $\hat{C}_2$ is, from eqn (6), $\hat{c}_2 q_2$, where $\hat{c}_2$ is the first row of $\hat{C}_2$ and q_2 is the first column of Q_2. We only need to show that $q_2 \neq 0$ to verify that $\hat{c}_2 q_2$ will contribute a component of magnitude $O(N^{-\frac{1}{2}})$ to $\hat{\Phi}_{1,1,1}$. Now from eqn (7), and using that $p = 1$, we find

$$\left(I - \Phi_0^{-1}\Phi_1\right)_{\text{column1}} = P_2(I_r - \Lambda_r)q_2 \tag{9}$$

Thus $q_2 = 0$ implies that the first column of $\Phi_0^{-1}\Phi_1$ is $(1,0,\ldots,0)'$. But because Φ_0 has unit diagonals, this implies that $\Phi_{1,1,1} = 1$, so contradicting the null hypothesis and demonstrating that in fact q_2 cannot be a zero column. □

NOTE

Corresponding author: Granville Tunnicliffe Wilson, Mathematics and Statistics, Department, Lancaster LA1 4YF, UK. E-mail: g.tunnicliffe-wilson@lancaster.ac.uk

REFERENCES

Ahn, S. K. and Reinsel, G. C. (1990) Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* 85, 813–23.

Awokuse, T. O. and Bessler, D. A. (2003) Vector Autoregressions, policy analysis and directed acyclic graphs: an application to the U.S. economy. *Journal of Applied Economics* 6, 1–24.

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* B 50, 157–224.

Oxley, L., Reale, M. and Tunnicliffe Wilson, G. (2004) Finding directed acyclic graphs for vector autoregressions. In *COMPSTAT Proceedings in Computational Statistics* (ed. J. Antoch), Heidelberg: Physica-Verlag, pp. 1621–28.

Reale, M. and Tunnicliffe Wilson, G. (2001) Identification of vector AR models with recursive structural errors using conditional independence graphs. *Statistical Methods and Applications* 10, 49–65.

Reale, M. and Tunnicliffe Wilson, G. (2002) The sampling properties of conditional independence graphs for structural vector autoregressions. *Biometrika* 89, 457–61.

REINSEL, G. C. (1993) *Elements of Multivariate Time Series Analysis*. New York: Springer-Verlag.
SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000) *Causation, Prediction and Search* (2nd edn). Cambridge, MA: MIT Press.
SWANSON, N. R. and GRANGER, C. W. J. (1997) Impulse response functions based on causal approach to residual orthogonalization in vector autoregression. *Journal of the American Statistical Association* 92, 357–67.
TUNNICLIFFE WILSON, G., REALE, M. and MORTON, A. S. (2001) Developments in multivariate time series modeling. *Estadistica* 53, 353–95.