Biometrika (2002), **89,** 2, *pp.* 457–461

Printed in Great Britain

The sampling properties of conditional independence graphs for structural vector autoregressions

By MARCO REALE

Department of Mathematics and Statistics, University of Canterbury, PB 4800 Christchurch, New Zealand

m.reale@math.canterbury.ac.nz

AND GRANVILLE TUNNICLIFFE WILSON

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.

g.tunnicliffe-wilson@lancaster.ac.uk

Summary

Structural vector autoregressions allow contemporaneous series dependence and assume errors with no contemporaneous correlation. Models of this form, that also have a recursive structure, can be described by a directed acyclic graph. An important tool for identification of these models is the conditional independence graph constructed from the contemporaneous and lagged values of the process. We determine the large-sample properties of statistics used to test for the presence of links in this graph. A simple example illustrates how these results may be applied.

Some key words: Causality; Moralisation; Partial correlation.

1. Introduction

We consider the structural pth-order vector autoregressive model, SVAR(p), of a stationary, m-dimensional time series $x_t = (x_{t,1}, \ldots, x_{t,m})'$, of the form

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

We require that the variance matrix A of a_t be diagonal and that Φ_0 represent a recursive, i.e. causal, dependence of each component of x_t on other contemporaneous components. This is equivalent to the existence of a reordering of the elements of x_t such that Φ_0 is triangular with unit diagonal. We also require that Φ_p has at least one nonzero element.

A given process x_t generated by such a model has in general many statistically equivalent representations of the same form, corresponding to different causal orderings of the contemporaneous dependence. It may also be represented uniquely by a standard VAR(p) model, and we shall assume that the order p has been previously determined by fitting such a model.

We are interested in the structural form (1) because it is reasonable to suppose that in the true process generating the data each contemporaneous variable may depend only on a smaller number of lagged or contemporaneous variables. This will be reflected in one representation which is sparse in the sense that many of the elements of the coefficient matrices are zero. This sparseness may be represented by a directed acyclic graph with nodes corresponding to the $m(p+1)$ series elements which appear in (1). The only links in this graph are to, and between, contemporaneous values of x_t, in the direction of causality, corresponding to the nonzero elements of Φ_i.

In a previous paper (Reale & Tunnicliffe Wilson, 2001), we advocated a way of identifying a sparse form of the model (1) using a conditional independence graph on the same nodes, estimated using a sample of x_t of length N. In that paper we only considered tests for links to, and between, contemporaneous values of x_t. We made brief reference to a standard result of Anderson (1971, p. 211) to justify these tests. This is that ordinary least squares regression can be used in large samples for inference about autoregressive coefficients. In § 2 of the present paper we derive the sampling properties of the test for a link in the conditional independence graph between any pair of nodes $x_{t-r,i}$ and $x_{t-s,j}$ for lags r and s from zero to some maximum lag $K \geqslant p$. In general this test is nonstandard, in that the variance of the statistic is not the same as that for the conditional independence graph of independent samples from a multivariate normal distribution. The exception is that in large samples the variance is the same if one node has zero lag. In § 3 we present a simple illustrative example.

2. Sampling properties of the test statistic

Following Whittaker (1990) the conditional independence graph of the Gaussian variables $x_{t-k,i}$, for lags $k = 0, 1, \ldots, K$ and series $i = 1, 2, \ldots, m$, is determined by the matrix of pairwise partial correlations of the variables conditional upon all the remaining variables. The nodes corresponding to a pair of variables are linked if and only if their partial correlation is nonzero. We will develop a test for whether or not each single partial correlation is zero, and use this to estimate the conditional independence graph. This estimate will be an exploratory tool, used to indicate sparse forms of (1) which can be efficiently estimated and compared. We will not consider multiple tests, although the theory we present could be extended in that direction.

The starting point for estimation of these partial correlations is the data matrix X consisting of the collection of contemporaneous and lagged data vectors $(x_{K+1-k,i}, \ldots, x_{N-k,i})'$. The sample values of the partial correlations may then be derived from the estimated covariance matrix of the variables, $\hat{V} = X'X/T$, where $T = N - K$. We assume here that the time series or data vectors have been mean-corrected. Our test is to reject $\rho = 0$, for the partial autocorrelation ρ between a particular pair of variables, if $|r| > c$, where r is the sample value of ρ and c is some critical value. To determine c we use the relationship between r and the t coefficient on v degrees of freedom of one member of the pair of variables, in the regression of the other member of the pair upon that and all the remaining variables. This relationship is $r = t/\sqrt{(t^2 + v)}$; see Greene (1993, p. 180). We determine the large-sample properties of t, and in practice we shall apply the test of significance to the t value rather than the partial correlation. The important point here is that ordinary least squares regression does not in general furnish the correct standard error of the regression coefficient in this time series context. This is because the regression will, in general, be upon variables which are in both the past and the future, and consequently the regression errors are not white noise.

Let y, in the set $x_{t-k,i}$, be the variable selected as the regressor, and let w be the remaining variables in this set. Let Y and W be the corresponding subsets of the data matrix X. Let $M = \text{var}(w)$, $P = \text{cov}(w, y)$, $\hat{M} = T^{-1}W'W$ and $\hat{P} = T^{-1}W'Y$, the elements of $\hat{M}$ and $\hat{P}$ being sample covariances of x_t. We assume that x_t does follow the model (1) so that the necessary conditions apply (Priestley, 1994, pp. 324–30) for consistency of these sample values: plim $\hat{M} = M$ and plim $\hat{P} = P$. The solution of the least squares equations $W'W\hat{\beta} = W'Y$ for the vector of regression coefficients therefore satisfies plim $\hat{\beta} = M^{-1}P = \beta$. We use this to define the asymptotic error series of the regression, $e_t = y - w\beta$, which may be expressed in the form of the finite sum

$$e_t = \sum_{i,r} \psi_{i,r} x_{t-r,i}. \tag{2}$$

By construction, $\text{cov}(w, e) = 0$ and we may reformulate the least squares equations as

$$\hat{M}\{T^{1/2}(\hat{\beta} - \beta)\} = T^{1/2}(T^{-1}W'E), \tag{3}$$

where E is the data vector corresponding to e_t. The elements of $D = T^{-1}W'E$ are the sample estimates of $\text{cov}(w, e)$. Being finite linear combinations of the sample covariances of x_t, they therefore satisfy (Priestley, 1994, pp. 337–9)

$$T^{1/2}D \to N(0, Q), \tag{4}$$

where we use $E(D)=0$ and Q is $\lim_{T\to\infty} \mathrm{var}(T^{1/2}D)$. The elements of Q are

$$\begin{aligned} &\lim_{T\to\infty} T^{-1} \operatorname{cov}\left(\sum_t x_{j,t-u}e_t,\ \sum_s x_{l,s-w}e_s\right) \\ &\qquad = \sum_{i,r}\sum_{k,v} \psi_{i,r}\psi_{k,v} \lim_{T\to\infty} T^{-1} \operatorname{cov}\left(\sum_t x_{j,t-u}x_{i,t-r},\ \sum_s x_{l,s-w}x_{k,s-v}\right) \\ &\qquad = \sum_{i,r}\sum_{k,v} \psi_{i,r}\psi_{k,v} \sum_m (\gamma_{i,k,m-r+v}\gamma_{j,l,m-u+w} + \gamma_{i,l,m-r+w}\gamma_{j,k,m-u+v}), \end{aligned} \tag{5}$$

where $\gamma_{i,j,k} = \operatorname{cov}(x_{i,t}, x_{j,t-k})$.

We require a consistent estimator $\hat{Q}$ of Q, and the central point of this paper is to propose one that is practicable. The last expression in (5) is an infinite sum over m that may not be directly estimated because it contains covariances for which no sample value is available. We suggest that these covariances are instead derived from the estimation of any one of the equivalent saturated forms of model (1), represented as $\Phi(B)x_t = a_t$. It is convenient to express (5), using the components of the spectrum S of x_t, as

$$\int_{f=-1/2}^{1/2} [\psi(B)S\psi(B^{-1})'B^{w-u}S_{l,j} + \{\psi(B)S\}_l\{\psi(B)S\}_j B^{-w-u}]\, df. \tag{6}$$

Here we use the backward shift operator B to aid interpretation, substituting $B=\exp(2\pi i f)$ to evaluate the integral. The operator $\psi(B)$ is defined by $e_t = \psi(B)x_t$ and is consistently estimated using the regression coefficients $\hat{\beta}$. The spectrum S is given by

$$S = \Phi(B)^{-1}A\{\Phi(B^{-1})^{-1}\}', \tag{7}$$

which is consistently estimated by ordinary least squares estimation of (1). It is straightforward to evaluate the integral numerically. An important practical point is that the estimated model must be stationary. This may be ensured by padding the observed series x_t, following mean correction, with K zeros at the start and finish. The regression estimators then become the multivariate Yule–Walker estimators, which are known to be stationary (Reinsel, 1993, pp. 89–91).

Again using plim $\hat{M} = M$ we have from (3) that

$$T^{1/2}(\hat{\beta}-\beta) \to N(0, M^{-1}QM^{-1}), \tag{8}$$

where M is consistently estimated by $\hat{M}$.

For comparison, the limiting variance matrix from ordinary least squares regression is $M^{-1}\sigma_e^2$, which will in general be different from that in (8). However, in the particular case that the chosen regressor is a contemporaneous variable, they are the same, so that the regression standard errors may be used. This may be seen from (6), in which $\psi(B)S\psi(B^{-1})'$ reduces to σ_e^2 and $\psi(B)S$ to an operator with no positive power of B, so that the second component of the integral becomes zero. Then Q reduces to $M\sigma_e^2$. In practice, when the chosen regressor is a contemporaneous variable, the estimator of the variance matrix in (8) is also identical in finite samples to the estimator from ordinary least squares regression if the order p used to estimate (1) is set equal to the maximum lag K. This is a consequence of Yule–Walker estimation, which ensures that the model autocovariances, derived here via the model spectrum, are equal to the sample autocovariances up to lag K. This finite-sample coincidence is a reassuring aspect of the method we have proposed.

Now let $\hat{\beta}$ be the estimator of the coefficient relating to a particular link in the conditional independence graph, and let $\hat{\sigma}_\beta$ be the consistent estimator of its true standard error. We therefore reject the presence of a link if $|z| = |\hat{\beta}|/\hat{\sigma}_\beta$ exceeds the appropriate critical value of a standard normal variable. We evaluate the performance of this 'correct' test in § 3 for a simple illustrative example and compare it with the 'incorrect' t test using the wrong standard error from the ordinary least squares regression.

3. APPLICATION TO A SIMPLE EXAMPLE

Consider the bivariate structural autoregression of order two given by

$$x_{t,1} = 0{\cdot}7x_{t-1,1} + a_{t,1}, \quad x_{t,2} = 1{\cdot}5x_{t,1} - 0{\cdot}5x_{t-2,2} + a_{t,2}. \tag{9}$$

This model is purposely chosen to be simple, in order to illustrate the point of the paper. It is represented in Fig. 1(a) by a directed acyclic graph linking nodes up to lag 2. Figure 1(b) shows the conditional independence graph which may be deduced from this. We show no link between the nodes at lags 1 and 2. For the general SVAR(p) model, many, if not all, of the nodes corresponding to the earliest p lags will be linked in both the directed acyclic graph and the conditional independence graph. These links could be determined under the assumption of stationarity. For a univariate AR(p) the conditional independence graph is symmetric under time reversal, but that is not true in the multivariate case.

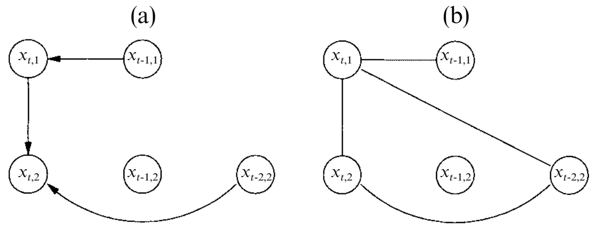

Fig. 1. (a) shows the directed acyclic graph representation of model (9) to lag 2, and (b) its corresponding conditional independence graph.

For this example, with the links as shown, the conditional independence graph contains an extra link corresponding to the moralisation rules (Whittaker, 1990, pp. 75–7). From this conditional independence graph alone we are not able to rule out the possibility that the original model contained a directed link, $x_{t,1} \leftarrow x_{t-2,2}$, in the place of this new link.

Consideration of the conditional independence graph with nodes up to lag 3 does rule this out. Figure 2(a) shows the directed acyclic graph extended to this lag by simply shifting the structure backwards. The solid lines of Fig. 2(b) show the conditional independence graph derived from this; the broken line on Fig. 2(b) is that of a link $x_{t-1,1} - x_{t-2,2}$ which would not arise if the link $x_{t,1} \leftarrow x_{t-2,2}$ were not present in the original model. A test which confirms the absence of the broken line therefore identifies the correct model.

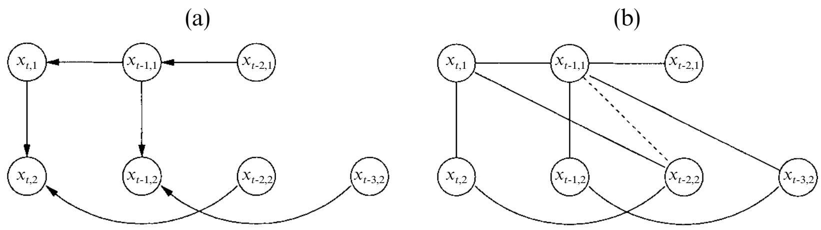

Fig. 2. (a) shows the directed acyclic graph representation of model (9) to lag 3, and (b) its corresponding conditional independence graph, with the broken line showing the link to be tested.

We simulated samples of length 200 from (9) and applied the tests of § 2 for the presence of the link $x_{t-1,1} - x_{t-2,2}$. Using 1000 samples and a nominal size of 5%, we found the correct test rejected the null hypothesis, that no link is present, for 4·8% of samples. The incorrect test rejected the null hypothesis for 10·9% of them. In this example the variance of the z statistic based on the correct value for $\hat{\sigma}_\beta$ was 0·96, and that of the t statistic based on the incorrect variance was 1·47. The importance of using the correct test is clearly demonstrated.

References

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

Greene, W. H. (1993). *Econometric Analysis*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Priestley, M. B. (1994). *Spectral Analysis and Time Series*. London: Academic Press.

Reale, M. & Tunnicliffe Wilson, G. (2001). Identification of vector AR models with recursive structural errors using conditional independence graphs. *Statist. Meth. Applic*. **10**. To appear.

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

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

[*Received August* 2000. *Revised November* 2001]