ESTADÍSTICA (2001), **53**, 160, 161, pp. 353 - 395.

DEVELOPMENTS IN MULTIVARIATE TIME SERIES MODELING

Granville Tunnicliffe Wilson
Department of Mathematics and Statistics
Lancaster University
Lancaster LA1 4YF, UK.

Marco Reale
Department of Mathematics and Statistics
University of Canterbury
Christchurch, New Zealand

Alex S. Morton
Department of Statistics
University of Warwick
Coventry CV4 7AL, UK.

ABSTRACT

We consider two new modeling procedures for multiple time series which address the challenge of providing both a good representation of the structure, and an efficient parameterization. We first review a method, applied to vector autoregressions of low order, which uses conditional independence graphs to identify a sparse structural autoregressive representation. We show by an example how this may be extended to identify a sparse structural form of an ARMA(1,1) model for a series of seven daily US dollar term rates. The identified structure reveals the pivotal role of the series of two year rates, and highlights sources of heteroscedasticity.

Vector autoregressions of high order are widely used to provide an empirical approximation to multiple time series structure, but the large number of parameters in these models restricts the possible maximum lag when the series is of moderate length. We present, and illustrate by example, a simple extension of the vector autoregression in which the predictors are smoothed functions of the past variables. This allows information from higher lags to be used in a model of relatively low order, and can improve forecasts at higher lead times.

KEY WORDS

Conditional independence, vector autoregression, vector ARMA, term rates, smoothing, heteroscedasticity, forecasting, cycles.

1. INTRODUCTION

The pth order vector autoregressive model, or VAR(p) model, for a stationary, m dimensional, time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,m})'$, is characterized statistically by the fact that the prediction of x_t from all past values x_{t-1}, $x_{t-2}, \ldots$, is a linear combination of the finite set $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$. In some practical applications it is clear that this provides a very good approximation to the structure of the series for some relatively small value, say 1, 2 or 3, of the order p. In other cases it may only provide a good approximation by using a relatively high order, in which case the need to estimate a very large number of coefficients will adversely affect the application of the model to prediction.

In this paper we first consider the case of low order autoregressions and how we might find an efficient, *i.e.* sparse, parameterization of the model. We

have two reasons for seeking an efficient parameterization. Firstly, even in the case of a low order model, reductions in the number of coefficients can reduce prediction error variance, particularly when the dimension m is large. Secondly, a parameterization which is parsimonious in its use of coefficients more strongly supports a causal interpretation of that representation. Used with other relevant background knowledge it may assist in the understanding of the true mechanisms giving rise to the series. We will present an approach to this problem using structural VAR models, that permit contemporaneous dependence between the series. We show how methods based on conditional independence graphs between contemporaneous and lagged values, may be used for identifying parsimonious models of this form.

There are many reasons why a vector autoregressive moving average model, with relatively low autoregressive and moving average orders p and q, the VARMA(p,q) model, may be appropriate for a series. For example a VAR(1) process with added observation error has a VARMA(1,1) representation. Also, if we only observe $m-1$ components of an m dimensional VAR(1) process, they will, in general, be represented by a VARMA(1,1) model. The VARMA process is characterized statistically by the fact that the prediction $\hat{x}_t$ of x_t from all past values $x_{t-1}, x_{t-2}, \ldots$, is in fact a linear combination of the finite set of variables $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ and $e_{t-1}, e_{t-2}, \ldots, e_{t-q}$, where e_t is recursively defined as the prediction error, or linear innovation, $x_t - \hat{x}_t$. We shall show how the methods based on conditional independence graphs may be extended to structural VARMA model, using an application to a VARMA(1,1) model for seven daily US dollar term rates.

There has been much research into methods for finding simplifying structure of multivariate time series, based on VAR and VARMA models forms. For exam-

ple Box and Tiao (1977) consider a linear transformations of the given series to a new set of series that have a particular canonical structure that reflects their predictability. Peña and Box (1987) propose a simplifying structure for time series in terms of a smaller number of factor series. The reduced rank VAR models of Ahn and Reinsel (1988) provide a more parsimonious parameterization by identifying structure in the parameter matrices of VARMA models. Tiao and Tsay (1989) propose a method of identifying a transformation of the series that results in scalar component models. Collectively, these determine a vector ARMA structure, but presented in a form in which any simplifying structure is captured by component models of lower order.

Despite these advances, for empirical applications, and particularly in forecasting, it is still usual to use a high order VAR model to approximate the structure of multiple time series. These models are only asymptotically correct, and for consistent estimation, the order used must increase with the length of series available. The estimation of the large number of parameters in such a higher order model, adversely affects its predictive performance. Doan, Litterman and Sims (1984) address this problem by using a Bayesian method for estimating the VAR. We shall use an approach that retains a low order model, but still incorporates past values at higher lags into the predictor. We do this by constructing a standard set of predictors which are exponentially smoothed combinations of past values. In the final part of this paper we explore this approach; a generalization of the standard VAR(p) model which requires the user to select a single smoothing constant. It has the potential to improve the forecasting ability of the VAR model, particularly at higher lead times.

2. GRAPHICAL MODELING OF STRUCTURAL VAR AND VARMA PROCESSES

2.1. The Structural VAR Model and its Graphical Representation

The VAR model for which we seek an efficient parameterization has a structural form which allows for contemporaneous dependence through a matrix coefficient Φ_0:

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

One requirement of this model is that the variance matrix D of a_t is diagonal. A further condition, on Φ_0, is that it represent a recursive (causal) dependence of each component of x_t on other contemporaneous components. This is equivalent to the existence of a re-ordering of the elements of x_t such that Φ_0 is triangular with unit diagonal. The model may be transformed to canonical form by dividing through by Φ_0, in which form x_t is expressed as a linear combination $x_{t-1}, x_{t-2}, \ldots, x_{t-p}$ with an error term $e_t = \Phi_0^{-1} a_t$. This is the linear innovation, having variance Σ related to D by

$$\Sigma^{-1} = \Phi_0' D^{-1} \Phi_0. \tag{2}$$

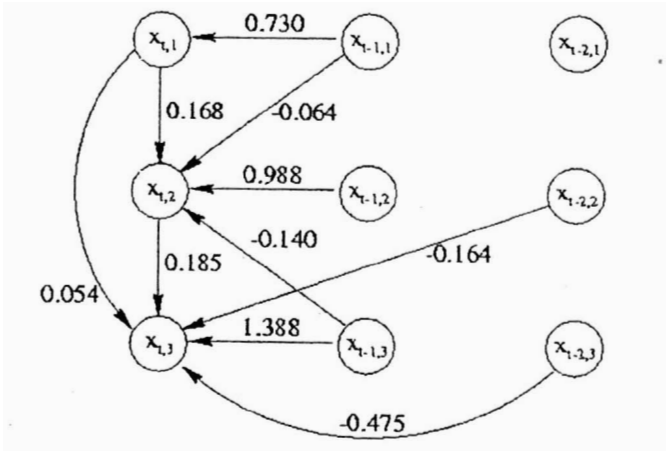
The model (1) is not unique, in that the transformation to the canonical form, which is unique, may be reversed by the choice of any matrix Φ_0 which satisfies (2). Each possible ordering of the series gives a different form of (causal) structural model. Our objective is to discover such an ordering which gives a model representation (1) requiring a relatively small number of coefficients. This will not be possible in all cases. The method we use to explore the possibilities will reveal this to us.

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 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 (1). Some arrows will start from the past, and some from other contemporaneous variables. As an illustration, we reproduce a structural model presented by Reale and Tunnicliffe-Wilson (2001). The data were 9 years of monthly values of three variables of the Italian monetary system; the re-purchase agreement interest rate, the average interest rate on government bonds, and the average interest rate on bank loans. The DAG representing the chosen structural VAR(2) model for these series is shown in Figure 1. The numbers attached to the links are the coefficients in the linear predictor for the corresponding contemporaneous variable.

These coefficients are estimated by single equation ordinary least squares (OLS) regression. This is fully efficient under our working assumption, that the vector series is Gaussian. Our methods are also applicable, and the properties of the estimates given by the regression are reliable, under wider conditions, such as e_t being I.I.D., presented for example in Anderson (1971).

FIGURE 1

The DAG representation of the structural VAR(2) identified for the Italian monetary variables.



2.2. Identification of the Structural VAR Model

We now describe the exploratory tools used to identify the model in this example. 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. Being based upon statistical correlations, the only causality which can be deduced from this is that indicated by the arrow of

time. Nevertheless, it may serve to suggest the direction of dependence between contemporaneous variables by admitting only a small number of possible simple DAG interpretations. 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.

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 $x_{t-u,i}$, $t = P+1-u, \ldots, N-u$, 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 step 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. We then select the order p which minimizes the AIC. In this way we selected the order $p = 2$ for the Italian monetary series, which is in agreement with that found in a previous analysis by Bagliano and Favero (1998).

2.3. The Conditional Independence Graph for a VAR(p) Process

The next step is to construct the sample CIG for the chosen model order p. This CIG consists of the same nodes as those shown in Figure 1, representing the variables up to lag 2. 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:

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

where the set of conditioning variables on the right is the whole set up to lag p, excluding the variables on the left. As shown by Whittaker (1990), 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

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

where $r = (p+1)(i-1)+u+1$ and $s = (p+1)(j-1)+v+1$ respectively index, in the matrices V and W, the $(p+1)m$ variables $x_{t-u,i}$ and $x_{t-v,j}$, for the m series at lags zero to p. In the wider linear least squares context, defining linear partial autocorrelations as the same function of linear unconditional correlations as in the Gaussian context, (3) still usefully indicates lack of linear predictability of one variable by the other given the inclusion of all remaining variables.

2.4. Estimation of the Conditional Independence Graph.

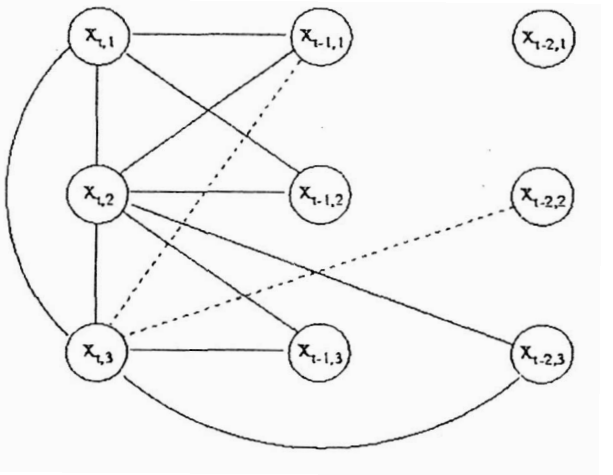
To estimate the CIG we use in place of V the sample covariance matrix $\hat{V}$ formed from the data matrix X, but including only lags up to p. We then need a statistical test to decide which links are absent in the graph. We are only concerned with links between contemporaneous variables and between 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, and $\nu = n - k$ is the residual degrees of freedom in the regression of any one column of X on all the remaining columns. 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). Of course we should properly apply multiple testing procedures when applying the test simultaneously to all sample partial autocorrelations, but that is not a practical option. Our attitude is similar to that advocated by Box and Jenkins (1976) for the identification, for example, of autoregressive models using time series partial autocorrelations. We use these values to suggest possible models; after fitting these we apply more formal tests and diagnostic checks to converge on an acceptable model.

To return to our example, the critical value for significance at the 5% level is 0.207. Figure 2 shows the appropriate subgraph of the CIG of the lagged variables constructed using this threshold, with the addition of two links, $x_{t,3} - x_{t-1,1}$ and $x_{t,3} - x_{t-2,2}$ shown by broken lines. These are included because their partial autocorrelations are very close to the threshold. The series are only of

moderate length so that some additional power for detecting non-zero partial correlations is justified.

FIGURE 2

The CIG estimated for the Italian monetary series.



2.5. Selection of a Directed Graphical Interpretation

We next proceed to determine which DAG representations are consistent with the CIG in (2), or are nearly so, allowing for statistical uncertainty. For this purpose we use, in an inverse manner, the moralization rule of Lauritzen and Spiegelhalter (1988), by which we can form the CIG that would arise from

any hypothesized DAG interpretation. 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 $a \rightarrow c$ 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 as marrying the parents. After doing this 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. In this example the main point to note is the clear absence of a link $x_{t-1,2} - x_{t,3}$. A moral link would be expected here unless we assign the direction between contemporaneous variables: $x_{t,2} \rightarrow x_{t,3}$, which also opens up the possibility that $x_{t,2} - x_{t,3}$ is a moral link. There is no similar clear indication of the remaining choice of contemporaneous links, though the exclusion of directed cycles, such as $x_{t,2} \rightarrow x_{t,3} \rightarrow x_{t,1} \rightarrow x_{t,2}$, limits the possibilities. We do note however that $x_{t-1,2} - x_{t,1}$ might be explained as a moral link, by assuming the direction $x_{t,1} \rightarrow x_{t,2}$. By such considerations we were led to the model represented in Figure 1. By the AIC this model is judged better than the saturated model, with 11 fewer parameters. Details are given in Reale and Tunnicliffe-Wilson (2001) of one other model for these series which shared this property but which required one more parameter. Moralization of Figure 1 in fact yields a CIG which differs from Figure 2 by the inclusion of extra links $x_{t-1,3} - x_{t,1}$, $x_{t-2,3} - x_{t,1}$, $x_{t-2,2} - x_{t,2}$ and $x_{t-2,2} - x_{t,1}$, which were not detectable from the length of data available. Our general conclusion is that inspection of Figure 2 lead us swiftly to the specification of a good structural model for these series; one which on the evidence of the AIC represents the structure as well as the full VAR(2) model with fewer than half the parameters, and with improved predictive ability.

2.6. Structural VARMA Modeling of Term Rates

We here present a novel extension to structural ARMA modeling, of the methods introduced in the previous sections. We do this by use of a single example. We consider a series of seven daily dollar term rates over the period from 30th November 1987 to 12th April 1990, excluding non-trading days. The maturity terms are 6 month, 1, 2, 3, 5, 7 and 10 years. Figure 3 illustrates just the six month, two year and ten year rates. The movements in the series are clearly highly correlated. Our starting point is a saturated canonical multivariate ARMA(1,1) model characterized by 98 ARMA coefficients and 21 correlations between the innovation series. This had been previously estimated for these seven series in order to provide a simulation model for assessing financial trading over a long period. The mean adjusted model is of the form

$$x_t = \Phi_1 x_{t-1} + e_t - \Theta_1 e_{t-1}, \tag{5}$$

where e_t is Gaussian white noise with covariance matrix Σ. The parameters were estimated by maximum likelihood using the method of Tunnicliffe Wilson (1973) to maximize the concentrated likelihood $-\frac{1}{2}n \log\det\hat{\Sigma}$, where $\hat{\Sigma}$ is the sample covariance matrix of the innovations series e_t. The innovations are regenerated from the mean corrected series, for given values of the coefficient matrices Φ_1 and Θ_1, by the recursion for $t = 1, 2, \ldots$:

$$e_t = x_t - \Phi_1 x_{t-1} + \Theta_1 e_{t-1}. \tag{6}$$

FIGURE 3

Six-month (solid line), two year (broken line) and ten year (dotted line) dollar term rate series.

Because data was plentiful the problem of transient non-stationarity in the series e_t due to lack of knowledge of e_0 was dealt with by setting $e_0 = 0$, discarding the first 39 values of e_t and using just the remaining 600. The length n therefore refers to this number.

Our plan is to identify and estimate a structural multivariate ARMA(1,1) model:

$$\Phi_0 x_t = \Phi_1^* x_{t-1} + a_t - \Theta_1^* a_{t-1}, \tag{7}$$

where the variance matrix D of a_t is diagonal. The equivalence between (5) and (7) is given by $\Phi_1^* = \Phi_0\Phi_1$. $\Theta_1^* = \Phi_0\ \Theta_1\Phi_0^{-1}$, $a_t = \Phi_0 e_t$ and $\Phi_0 V \Phi_0' = D$. The same recursive (causal) structure for Φ_0 is assumed as for the structural VAR model (1), and our aim again is to identify a model of the form (7) in which the matrices Φ_0 and Φ_1^* are sparse.

2.7. A Conditional Independence Graph for the VARMA Model

Consider (7) as defining a directed acyclic subgraph in the $3m$ variables consisting of (the components of) x_t, x_{t-1} and a_{t-1} with independent errors a_t. Because a_{t-1} is not defined until we have identified the structural model and know Φ_0, we cannot simply apply our graphical model identification procedure to this set of variables. We do however know the lagged canonical residuals e_{t-1} accurate to $O_p(n^{-\frac{1}{2}})$, as a result of estimating (5). We can then achieve our aim in two steps. In the first we apply the procedure to the data vectors corresponding to the component variables of x_t, x_{t-1} and e_{t-1}. We can replace a_t by e_t because this is a linearly equivalent set of variables. In defining the partial correlations it will not therefore affect the links within components of x_t and between these and the components of x_{t-1}. Once these links are

defined, then so are the structural residuals a_t. This is the main step because it identifies the causal ordering within x_t. In the second step the links between x_t and a_{t-1} can be identified by subset regression if further parsimony is sought.

TABLE 1

Partial Correlations for Structural ARMA Modeling.

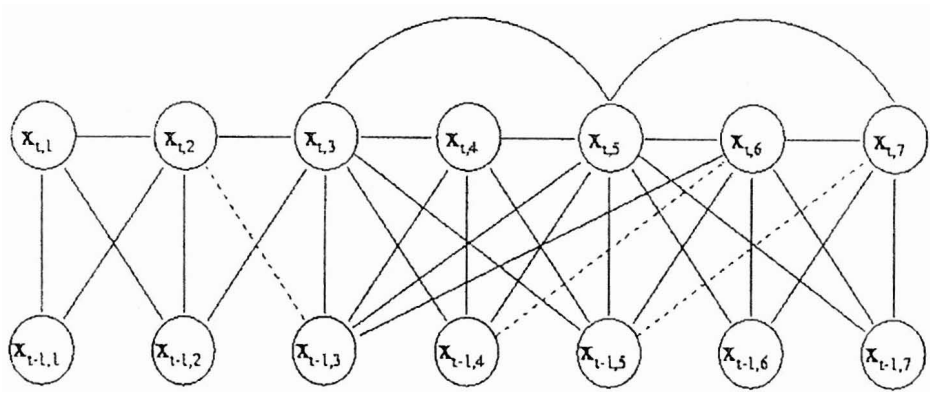
	$x_{t,1}$	$x_{t,2}$	$x_{t,3}$	$x_{t,4}$	$x_{t,5}$	$x_{t,6}$	$x_{t,7}$
$x_{t,1}$	1.000						
$x_{t,2}$	0.720	1.000					
$x_{t,3}$	0.021	0.113	1.000				
$x_{t,4}$	0.038	0.017	0.689	1.000			
$x_{t,5}$	0.019	-0.019	0.154	0.270	1.000		
$x_{t,6}$	-0.044	-0.013	-0.050	0.042	0.544	1.000	
$x_{t,7}$	0.038	0.006	-0.055	0.010	0.115	0.658	1.000
$x_{t-1,1}$	0.892	-0.660	0.009	-0.034	-0.023	0.022	-0.030
$x_{t-1,2}$	-0.619	0.846	-0.110	-0.025	0.047	-0.010	0.008
$x_{t-1,3}$	0.018	-0.086	0.654	-0.392	-0.172	0.108	-0.011
$x_{t-1,4}$	-0.040	-0.002	-0.396	0.570	-0.121	-0.085	0.029
$x_{t-1,5}$	-0.234	-0.020	-0.135	-0.168	0.732	-0.393	-0.104
$x_{t-1,6}$	0.027	-0.022	0.042	-0.040	-0.426	0.820	-0.557
$x_{t-1,7}$	-0.021	-0.002	0.044	-0.003	-0.135	-0.606	0.957

For this example we therefore construct the data matrix X from the 21 data vectors corresponding to x_t, x_{t-1} and e_{t-1}, and from this form the

corresponding sample partial autocorrelation matrix. The elements of this are shown in Table 1 for the links, within x_t and between x_t and x_{t-1}, of interest in identifying the contemporaneous causality. Figure 4 shows the corresponding sub-graph of the CIG, constructed using a significance level of 5% corresponding to a critical value of 0.081. We have used broken lines to indicate the three most marginal links.

FIGURE 4

The CIG derived from Table 1 for structural ARMA model identification

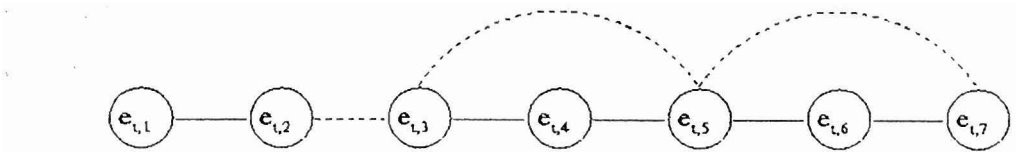


2.8. The Linear Innovations Graph

We now note a property of this graph, that the subgraph linking the components of x_t is identical with the CIG of the linear innovations series e_t from model (5), which is shown in Figure 5. This is because the covariance

FIGURE 5

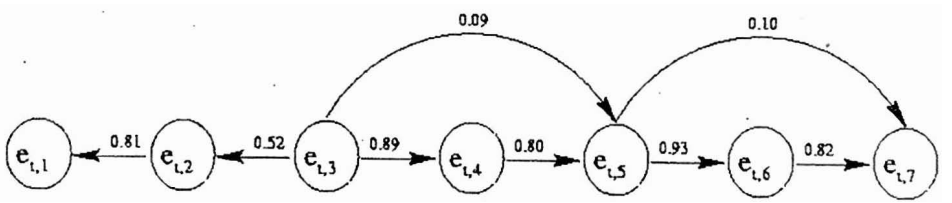
Conditional Independence Graph derived for the dollar term rate innovation series.



structure of e_t is precisely the covariance structure of x_t conditioned on the past variables used to predict it: in this case x_{t-1} and e_{t-1}. The causal structure of e_t can be studied in isolation and this is a theme of Swanson and Granger (1997) who suggest that a linear causal structure may be appropriate in many cases. In Reale and Tunnicliffe Wilson (2001) the possible DAG interpretations of Figure 5 are explored as an illustration of graphical modeling in the standard case of multivariate data without lagged structure. They found in this case 28 models which were statistically indistinguishable interpretations, and estimated one of them, shown in Figure 6. The structural error vector a_t resulting from this estimation should have diagonal covariance matrix, so inspection of its sample correlation matrix provides an initial diagnostic check of the fitted DAG. In this example, Reale and Tunnicliffe Wilson (2001) found clear evidence of model inadequacy in the form of a significant cross-correlation between the structural residuals $a_{t,1}$ and $a_{t,3}$. Introducing a further directed link, from the innovation $e_{t,3}$ to $e_{t,1}$ in Figure 6, rectified this inadequacy.

FIGURE 6

A directed graph which explains the independence graph shown in Figure 5, with the regression coefficients associated with the links.

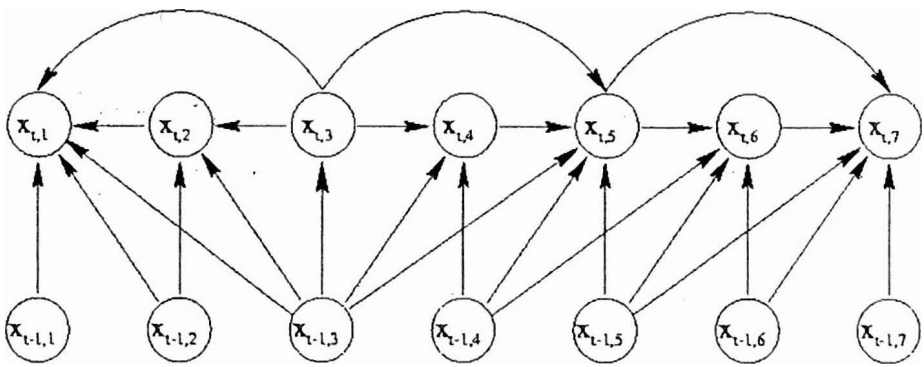


2.9. Selection of the Directed Graphical Interpretation

We can build upon the results of this innovation series modeling as we move to structural model identification based upon the CIG of Figure 4. We shall hope that the extra information in this figure will help to distinguish among, and reduce, the various DAG interpretations possible for Figure 5, considered as an isolated subgraph. We shall assess any tentatively identified model by regressing each component of x_t on other specified elements of x_t and x_{t-1}, and all elements of e_{t-1}, forming the deviance as before. When diagnostic checking suggests that an acceptable structural model of the form (7) has been found, it will then be re-estimated by maximum likelihood for full efficiency, including, at least initially, all coefficients of the lagged terms a_{t-1}. There is not space to

FIGURE 7

The first DAG identified for the structural ARMA model.



give a complete account of our model exploration, but the strategy is to consider which directions of contemporaneous dependence are compatible with the relative sparseness of links from the past, having considered the implications of moral linkage. We first considered the ordering shown previously in Figure 6. One reason for this is that $x_{3,t}$ is a convincing candidate from this Figure as a pivotal variable, the influence from which then spreads out to the other variables. Using this ordering we selected the lagged variables from those indicated by the links in Figure 4. This resulted in the model represented in Figure 7 and referred to as A in Table 2, which presents the evidence for comparing the models, relative to the saturated model, using the HIC, Hannan and Quinn (1979) and SIC, Schwarz (1978), as well as the AIC.

TABLE 2

Comparisons of structural ARMA models using criteria relative to the saturated model.

model	reduction in no. of pars.	increase in deviance	relative AIC	relative HIC	relative SIC
A	44	74.71	-13.29	-88.60	-206.75
B	44	58.55	-29.45	-104.76	-222.91
MLE of B	44	83.17	-4.83	-80.14	-198.30

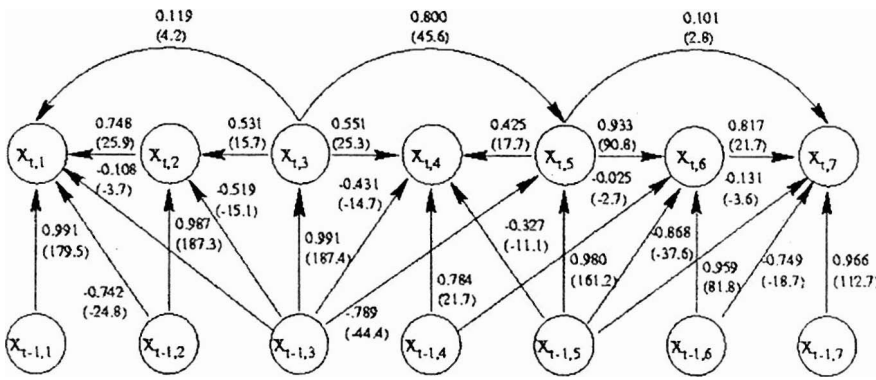
The link $x_{t-1,3} \rightarrow x_{t,1}$ had also been introduced as a possibility along with $x_{t,3} \rightarrow x_{t,1}$ following the finding for the innovations model that $e_{t,3} \rightarrow e_{t,1}$ was required and both are moderately significant. Other than these, 9 of the links from lagged variables which appear in Figure 4 also occur in the model and 8 of the other 9 can be accounted for by moralization. The remaining link, $x_{t-1,3} \rightarrow x_{t,6}$ was found to be quite insignificant. This model is preferable to the saturated model according to all three criteria in Table 2.

2.10. Model Exploration and Checking

Although this is a good model, we explore other possible models which are similarly compatible with the CIG, in an attempt to discover how precisely this structural ordering is determined. We first consider model B with a reversal of ordering between variables four and five as shown in Figure 8. The results

FIGURE 8

The final DAG identified for the structural ARMA model.



in Table 2 show that a deviance improvement of approximately 16 is achieved with no net change in the number of parameters. We have here a clear indication of model preference which was not possible by consideration of the innovations alone. The residual correlation matrix is shown in Table 3. The largest value is 0.1053 between variables 5 and 6. All the values in this table are close to those of the corresponding sub-model fitted to the innovation series, so no improvement can be expected by introducing further lagged variables. A similar reversal of the ordering between variables 6 and 7 drastically

TABLE 3

Residual correlation matrix for the structural ARMA model.

1	1.000						
2	-0.001	1.000					
3	0.000	-0.001	1.000				
4	0.040	0.066	0.000	1.000			
5	0.035	-0.036	0.000	-0.001	1.000		
6	-0.014	-0.034	-0.056	0.065	0.105	1.000	
7	0.036	0.044	-0.024	0.011	0.045	-0.004	1.000
	1	2	3	4	5	6	7

increased the deviance and is not otherwise reported. However, a reversal of order between variables 1 and 2, *i.e.* a replacement of $x_{t.2} \rightarrow x_{t.1}$ by $x_{t.1} \rightarrow x_{t.2}$ and $x_{t-1.2} \rightarrow x_{t.1}$ by $x_{t-1.1} \rightarrow x_{t.2}$, increased the deviance by only 0.5, so the two possible orderings are statistically indistinguishable. Our conclusion is that model B is best, with the MLEs of the coefficients (t values) of the links shown in Figure 8 and the likelihood criteria for the maximum likelihood estimation of this final model shown in Table 2. Series 3, the two year rate, is the pivotal variable, but series 1 and 2 appear to be only weakly linked to the remainder, and their ordering is not critical. The number of autoregressive coefficients shown in Figure 8 is 26, compared with 70 for a saturated model and this reduction is acceptable according to all the criteria shown.

2.11. The Innovation Components of the Structural Model

The contemporaneous recursive ordering defines the residual series a_t as orthogonal components of the residual innovations e_t. The variances of these *orthogonal innovations* are shown in Table 4 in comparison with the *canonical* variances of e_t. Apart from the pivotal series $x_{t,3}$, the recursive modeling accounts for a substantial fraction of the canonical innovation variance, with $x_{t,4}$, $x_{t,6}$ and $x_{t,7}$ being most nearly determined.

TABLE 4

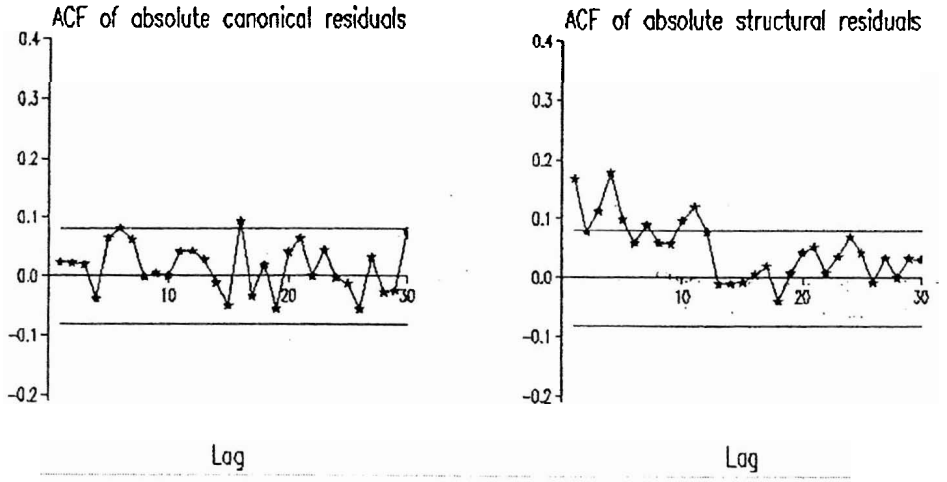
Variances of canonical and orthogonal innovations (times 10000).

Series	1	2	3	4	5	6	7
Canonical variance	58.5	57.6	60.8	53.9	49.5	45.9	41.5
Orthogonal variance	20.9	41.7	61.2	3.9	11.3	3.2	2.7

The components $x_{t,1}$ and $x_{t,2}$ are much less precisely dependent upon the remainder. The orthogonal innovations also appear to reflect different sources of heteroscedasticity. In particular, as shown in Figure 9, the absolute values of the canonical residuals for the five year rate series, show noticeable autocorrelation at low lags which is not evident in the absolute canonical residuals for the same series.

FIGURE 9

The sample autocorrelations of the absolute canonical residuals and the absolute structural residuals for the five year rate, showing evidence of heteroscedasticity in the latter. The plot shows the two standard error limits about zero for the sample autocorrelations of white noise.



In this example we have therefore achieved our aim of identifying a parsimonious model form with no loss of predictive ability. We have also found a convincing representation of the flow of information both from the past to the present, and between contemporaneous values, with the two year rate variable being of greatest importance and the five year rate of secondary importance. The sources of heteroscedasticity are also better identified in this model.

3. AN EXTENSION OF VECTOR AUTOREGRESSION

3.1. Background and Introduction

The models we now present are an extension of the standard VAR model, to be used in a similar manner as an empirical tool, primarily for the purpose of forecasting, and secondly for characterizing the series properties by estimating their joint spectrum.

The models are, formally, a sub-class of the vector ARMA models. However, the moving average part is of a fixed and simple form, requiring the specification of a single smoothing coefficient θ. The models were first motivated by a continuous time model presented in the univariate case by Belcher, Hampton and Tunnicliffe Wilson (1994). The idea is to use a set of smoothed functions of past values of the series as linear prediction variables. If $\theta = 0$ these functions of the past reduce to the usual lagged values of the series and the model simplifies to the standard vector autoregression. The discrete time model is described in the univariate case by Morton and Tunnicliffe Wilson (2001a), and an application of the continuous time version to the bivariate case given in Morton and Tunnicliffe Wilson (2001b). They call the extended autoregressions ZAR models. Here, we present the discrete time form of the model for multiple time series showing how it can be fitted and used for forecasting. These VZAR models are applied in a similar manner to the high order VAR models, with the aim of obtaining a good approximation to the second order structure of the series. Examples in the univariate case show that, for a given order of model, a much better approximation may be obtained by these extended models than by a simple autoregression, and the improvement is not

very sensitive to the choice of θ. Of course, if the series is truly represented by a low order autoregression, then a higher order will be required within the class of extended models.

3.2. Linear Combinations of the Past

The smoothing is based upon powers W^k of an operator W applied to the time series x_t, where W is defined, in terms of the backward shift operator B, by

$$W = \frac{B - \theta}{1 - \theta B} \tag{8}$$

Because W is a unimodular transformation, having unit gain at all frequencies, the series $W^k x_t$ all have the same spectrum and autocovariance structure as x_t, when that is stationary. We form the series $x(k)_t = W^k x_t$ recursively for $k = 1, 2, \ldots$, using

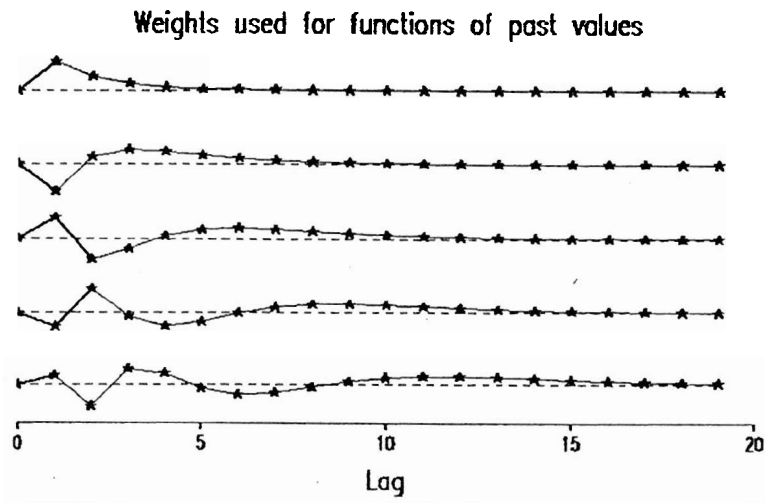
$$x(k)_t = -\theta x(k-1)_t + x(k-1)_{t-1} + \theta x(k)_{t-1} \tag{9}$$

taking $x(0)_t = x_t$. Because $x(k)_t$ includes a component $(-\theta)^k x_t$, we will remove this to form smooth combinations of past values alone, as

$$\tilde{x}(k)_t = x(k)_t - (-\theta^k) x_t = \left[W^k - (-\theta)^k\right] x_t. \tag{10}$$

FIGURE 10

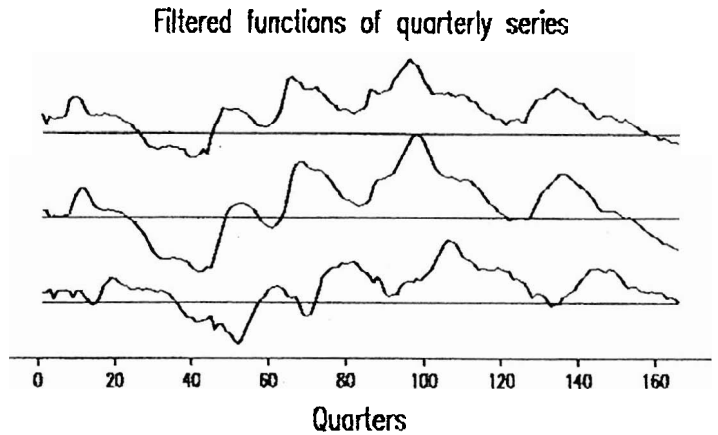
From the top down, the weights attached to past values by the operation defined in (10), using $\theta = 0.5$, *and for orders* $k = 1, 2, \ldots, 5$.



For example $\tilde{x}(1)_t = (1 - \theta^2)(1 - \theta B)^{-1} x_{t-1}$, which is proportional to a lagged exponential weighted moving average (EWMA), and $\tilde{x}(2)_t = (1 - \theta^2)\{(1+\theta^2)B - 2\theta\}(1-\theta B)^{-2} x_{t-1}$. The plots in Figure 10 show the weights which are attached to past values by $\tilde{x}(k)_t$ in the case $\theta = 0.5$ and for $k = 1, 2, \ldots, 5$. Regression upon exponentially weighted combinations of past values was used by Bray (1971) for prediction of unemployment from the rate of growth of GDP. His regressors were obtained by successive application of EWMAs with increasing values of the smoothing parameter so that his weights were all positive.

FIGURE 11

From the top down, a series together with the series resulting from applying the operation defined in (10) for orders $k = 1$ and 4, and using $\theta = 0.5$.



When applying (9) to a time series sample $x_1, x_2, \ldots, x_n$ we will start from $t = 2$ and will need to choose a value of $x(k)_t$ for $t = 1$. Because the gain of the operator W is 1 at all frequencies, a natural choice is $x(k)_1 = x(k-1)_1 = \ldots = x_1$, which is equivalent to assuming that $x(k)_t = x_1$ for all $t \leq 0$. In Figure 11 we illustrate the application of powers of W and W^4 to the seasonally adjusted quarterly series of USA unemployment rates from 1959 Q1 to 2000 Q2. This is one of a set of three time series which we shall use in a subsequent example to illustrate prediction based on these linear combinations of the past.

3.3. The Extended VAR Model

The model we fit to our vector time series x_t is the regression

$$x_t = c + \xi_1 \tilde{x}(1)_t + \xi_2 \tilde{x}(2)_t + \cdots + \xi_p \tilde{x}(p)_t + e_t \qquad (11)$$

in which ξ_k are matrix coefficients. This is a canonical form of the model, in which the errors e_t have covariance matrix Σ. Provided these coefficients are full, *i.e.* there are no elements constrained to be zero, the parameters can be estimated with full efficiency by OLS applied to the separate single equation regressions for each element of x_t. A structural model which includes contemporaneous regressors, and having errors a_t with diagonal covariance matrix D could also be fitted. This is exactly equivalent to (11) if there are no parameter constraints. We shall use the canonical form because prediction from the past is our aim. The models naturally form a nested class as p increases.

We can overcome concern about the effects on the regression of the transient errors at the start of $x(k)_t$ by introducing extra terms in (11) which allow for these effects. We suggest that a maximum model order P is chosen. The data matrix for the regression consists of the $m(P+1)$ vectors $x(k)_{ti}, t = 1, 2, \ldots, n$, for $k = 0, 1, \ldots, P$ and $i = 1, 2, \ldots, m$. We augment this with the P vectors of impulse responses of the operators W^k for $k = 1, 2, \ldots, m$. These are the responses illustrated in (10), and are readily generated using (9), but starting from $t = 0$, and setting $x_t = 0$ for all t except $x_0 = 1$. The values of $x(k)_t$

so generated for $t = 1, 2, \ldots, n$ then comprise the required impulse response vectors. We advise that these vectors are included, whatever order $p \leq P$ of model is fitted, in each of the single equation regressions of (11). It is in theory only necessary to include impulses corresponding to W^k for $k \leq p$. The reason for our advice is for comparability with standard vector autoregression. In the case $\theta = 0$, the inclusion of the impulses is equivalent to dummy variables for the effects in lagged regression of unknown values x_t for $t = 0, -1, \ldots, P-1$. The net effect of this is to use only regression vectors with elements of x_{t-k} for $t = P+1, P+2, \ldots, n$. The advantage of using fixed P when fitting any order p of model up to P, is that the dependent vector in the regressions remains the same, rather than reducing in length with increasing p. This is better for comparing the prediction error variances from models of increasing order. For our model comparisons we shall also use the AIC, HIC and SIC, with the deviance (minus twice the log likelihood) defined as $n \log \hat{\Sigma}$ and penalized by $2pm^2$ for the AIC, $2pm^2 \log \log n$ for the HIC and $pm^2 \log n$ for the SIC. The estimate $\hat{\Sigma}$ is the sample value formed from the residuals $\hat{e}_t$ without correction for parameter degrees of freedom.

3.4. Properties of the Extended VAR Model

Before we show the example we give some further properties of the new model. First, consider the regression model (11) in the case when x_t is a weakly stationary vector process. Because we can invert (8) to get $B = (W + \theta)/(1 + \theta W)$, any combination of past values $B^k x_t$, by expansion of B^k in terms of W^k, can also be expressed as a combination of the regressors $\tilde{x}(k)_t$. The

VZAR models therefore have the same asymptotic predictive potential as the class of VAR models.

Secondly, we write the model in the form.

$$\xi(W)(x_t - \mu) = (\xi_0 - \xi_1 W - \xi_2 W^2 - \cdots - \xi_p W^p)(x_t - \mu) = e_t. \qquad (12)$$

where $\xi_0 = 1 + \sum \xi_k (-\theta)^k$. We have here reversed the corrections of W^k given in (10), and gathered all these into the coefficient ξ_0. The net coefficient of x_t from all the terms in (12) is still therefore 1. which is the value of $\xi(W)$ evaluated at $W = 1$. If we now substitute $W = (B - \theta)/(1 - \theta B)$ into (12) and simplify. we obtain:

$$\phi(B)(x_t - \mu) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(x_t - \mu) = (1 - \theta B)^p e_t. \quad (13)$$

The matrices ϕ_k are the usual autoregressive coefficients of a VARMA model. but the operator $(1 - \theta B)^p$ is a scalar moving average, acting on the error vector e_t which we now assume is white noise. Provided therefore, that $\det \phi(B)$ has no zeros for $B < 1$, model (13), and hence also model (12), represents a weakly stationary VARMA process with mean μ. Now (8) represents a Möebius transformation, see Priestley (1985 p 170), of the unit disk to itself, so the stationarity condition may be equivalently expressed by the requirement

that $\det \xi(W)$ has no zeros for $W < 1$. The coefficients $\xi_0^{-1}\xi_k$ therefore occupy the same parameter space as that of a standard VAR.

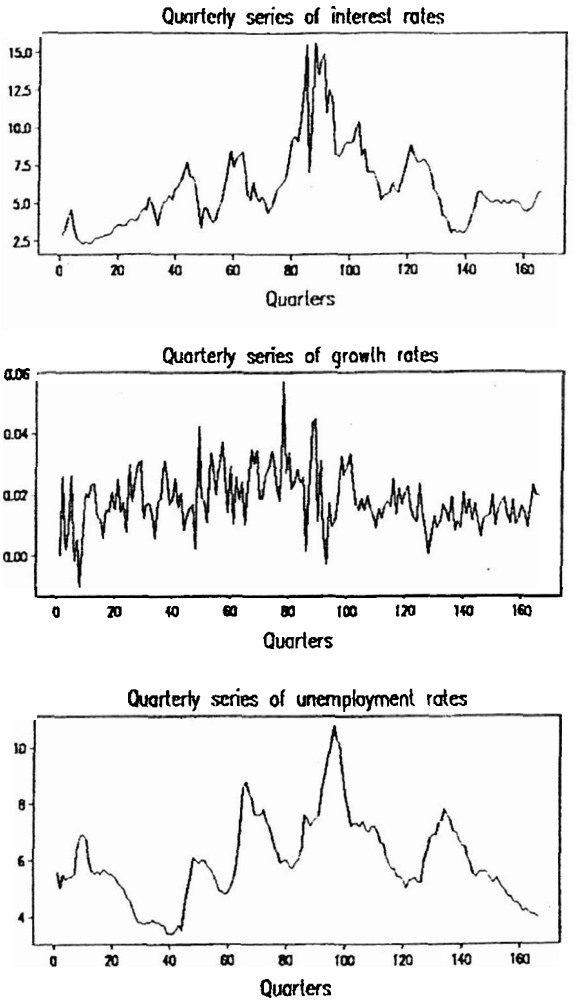
Other. spectral. properties of the model are given for the univariate case by Morton and Tunnicliffe Wilson (2001a), and these readily generalize to the multivariate case. A final property of the model, which is not presented elsewhere, is the state space representation, given for the mean corrected variables by

$$\begin{pmatrix} \Theta & I & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots \\ \cdots & 0 & \Theta & I \\ \xi_0 & -\xi_1 & \cdots & -\xi_p \end{pmatrix} \begin{pmatrix} x(0)_t \\ x(1)_t \\ \vdots \\ x(p)_t \end{pmatrix} = \begin{pmatrix} I & \Theta & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots \\ \cdots & 0 & I & \Theta \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} x(0)_{t-1} \\ x(1)_{t-1} \\ \vdots \\ x(p)_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ e_t \end{pmatrix}. \tag{14}$$

The model (12) appears in the last line of this representation. with the previous lines recursively defining the variables $x(k)_t$. The equation should be divided through by the matrix on the left, to bring it into the standard form. The observation equation is simply the selection of the first m state variables. It is possible to use (14) with the Kalman filter to form one step ahead predictions and derive the exact likelihood of the model, avoiding the estimation of the transients. It is recommended that the square root form of the Kalman filter is used to avoid any problem with state estimation. The estimation by regression is however very straightforward, and the most useful application of (14), again with the Kalman filter, is to multi-step forecasting.

FIGURE 12

Three USA macro-economic indicators used for the model example. All series consist of quarterly values from 1959 Q1 to 2000 Q2. The first is the interest rates. the second is the quarterly change in the logarithm of the GDP and the third is the unemployment rate.



3.5. An Example of Modeling Three Quarterly Series.

The series which we use for illustration are three USA macro-economic indicators, the treasury bill rate, the GDP in current dollars, and the unemployment rate. This is a selection of the variables used by Doan, Litterman and Sims (1984) in their Bayesian autoregressive model. We use the quarterly seasonally adjusted values from 1959 quarter 1 to 2000 quarter 2. For estimation we will exclude the final three years of data, retaining them for illustration of out of sample forecasts. We apply the logarithmic transformation to our three series before fitting the model, and we also difference the second series. Figure 12 shows plots of the series.

We first describe the result of fitting a standard vector autoregression to the three series. The AIC selected a model order 8, although the HIC selected order 3. The HIC will generally select a lower order model, because it is designed to estimate the order consistently, when it is supposed that there is a true finite order model. The AIC tends to overestimate such a true order, but because we are attempting to approximate the structure of the series for the purposes of forecasting, and do not suppose that there is a true finite order, we believe that the AIC is the more appropriate criterion for obtaining a good approximating model. We did examine the results from fitting the model of order 3, but the forecasts did not follow the series at all well. We therefore show the results of fitting the VAR(8) model and extrapolating the final 3 years of data, from 1997 Q3 to 2000 Q2, in Figure 13.

By comparison, the order chosen for the ZAR model, using $\theta = 0.5$, was 5. In fact this was the first minimum and the AIC reduced to a lower value at lag 9

and continued to reduce. The HIC indicated order 2. The forecasts from the ZAR(5) model are shown in Figure 14.

FIGURE 13

Out of sample forecasts of the three USA macro-economic indicators over the period from 1997 Q3 to 2000 Q2, with further forecasts to 2003 Q2, obtained using a standard VAR(8) model.

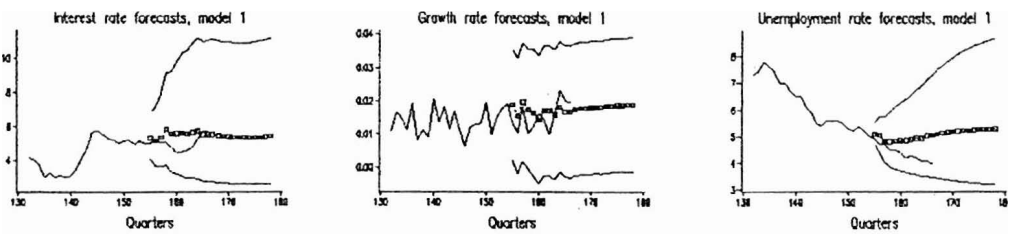
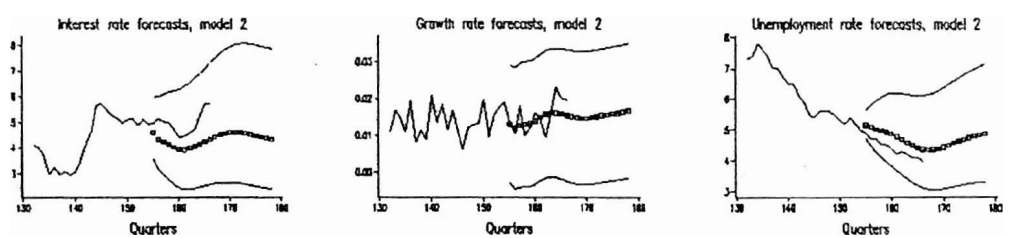


FIGURE 14

Out of sample forecasts of the three USA macro-economic indicators over the period from 1997 Q3 to 2000 Q2, with further forecasts to 2003 Q2, obtained using the vector ZAR(5) model.



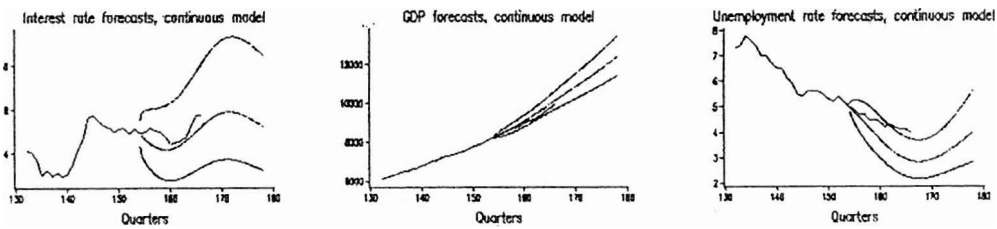
3.6. Model Comparisons

One can only draw limited conclusions from one example, and our aim is mainly to compare and contrast the two models. Although we do not show them, we remark first that the model residuals and their sample autocorrelations were generally similar, showing no clear preference for one model. However, the lower AIC was for the ZAR model, by 17.0, and a within sample forecasting comparison showed that at lead time 6 the ZAR model had lower mean square forecast error, by 35%, 8% and 26% respectively, for all three series. Before commenting on the out of sample forecasts, we draw attention also to Figure 15, which shows forecasts produced by a continuous time ZAR model of order 4. This was estimated by maximum likelihood, using the Kalman filter to evaluate the likelihood. In this case, the logarithms of the GDP series were used without differencing and the forecasts are shown of the original values of this series. The error limits shown on all the forecasts represent 95% confidence.

The main difference in the out of sample forecasts is that both the discrete and continuous ZAR models show the unemployment continuing to decrease over the last three years. The almost continuous decline of unemployment over the last 7 or 8 years, associated with a continuing healthy economy, has appeared to suggest a structural break in the normal pattern of economic cycles. However, the ZAR models manage to forecast this continuing decline, on the basis of estimation from data up to 1997 quarter 2. They do however predict a reversal around the end of 2000. By contrast with the ZAR model, the standard VAR model forecasts show little movement at higher lead times.

FIGURE 15

Out of sample forecasts of the three USA macro-economic indicators over the period from 1997 Q3 to 2000 Q2. obtained using the vector CZAR(4) model.

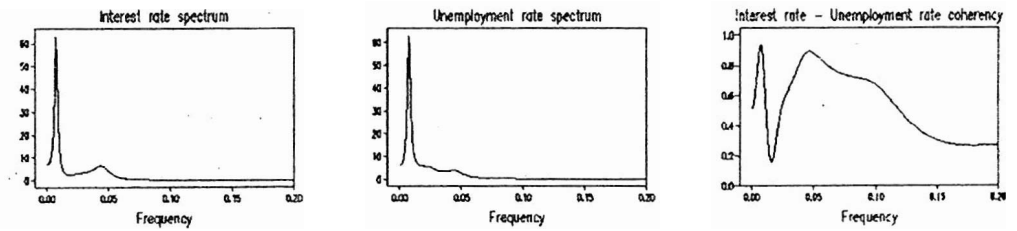


3.7. Spectral Properties of the Model

As Morton and Tunnicliffe Wilson (2001a) point out. the ZAR model can achieve a finer resolution of the spectrum at lower frequencies, and this can result in improvements to forecast accuracy at higher lead times. We show some of the spectral properties of the ZAR(5) and AR(8) models in Figures 16 and 17 respectively. We display the spectral densities of the interest rate and unemployment rate series and the coherency between them. These reflect a strong dependence at very low frequencies and around frequencies associated with the five-year economic cycle. The ZAR model resolves some of the features more clearly. In fact there is evidence of spectral features at frequencies corresponding to periods of 5, 10 and 2.5 years, the last of these being clear also in the growth rate spectrum which we do not show.

FIGURE 16

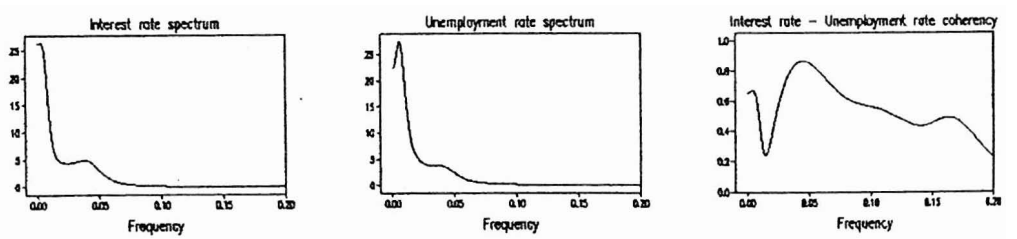
The spectra of the interest rate series and unemployment rate series, together with the coherency between the series, as estimated by the vector ZAR(5) model.



There are still many questions to be resolved regarding the ZAR model, for example the appropriate choice of the parameter θ in any given example. There may be dangers in over-fitting the low frequency structure of the series if θ

FIGURE 17

The spectra of the interest rate series and unemployment rate series, together with the coherency between the series, as estimated by the vector AR(8) model.



is chosen too large. We require order selection criteria which are reliable and also assess the forecasting ability of the model at higher lead times. The example we have shown does however suggest that this new class of models can achieve real improvements in forecast accuracy.

4. CONCLUSION

We have presented two new directions of development in multivariate time series modeling, with real examples to demonstrate their value. We hope they will provide useful additions to the range of modeling methods for multivariate time series which has been built up over the years, and which we have briefly reviewed. It may be possible to combine the different approaches, such as the two methods we have presented here, to achieve further advances in the aims of multivariate time series modeling.

REFERENCES

AHN, S.K. and REINSEL, G.C. (1988). Nested Reduced-Rank Autoregressive Models for Multiple Time Series. *Journal of the American Statistical Association.* **83**: 849-856.

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.* Wiley: New York.

BAGLIANO, F.C. and FAVERO, A.C. (1998). Il Canale del Credito della Politica Monetaria. Il Caso Italia. (ed. G. Vaciago). In *Moneta e Finanza.* Il Mulino: Bologna.

BELCHER, J., HAMPTON, J.S. and TUNNICLIFFE WILSON, G. (1994). Parameterization of Continuous Time Autoregressive Models for Irregularly Sampled Time Series Data. *Journal of the Royal Statistical Society Series B.* **56**: 141-155.

BOX, G.E.P. and JENKINS, G.M. (1976). Time Series Analysis, Forecasting and Control, revised edn. Oakland: Holden-Day.

BOX, G.E.P. and TIAO, G.C. (1977). A Canonical Analysis of Multiple Time Series. *Biometrika.* **64**: 355-365.

BRAY, J. (1971). Dynamic Equations for Economic Forecasting with the G.D.P. - Unemployment Relation and the Growth of G.D.P. in the United Kingdom as an Example. *Journal of the Royal Statistical Society Series A.* **134**: 167-227.

DOAN, T., LITTERMAN, R. and SIMS, C. (1984). Forecasting and Conditional Projections Using Realistic Prior Distributions. *Econometric Reviews.* **3**: 1-100.

GREENE, W.H. (1993). *Econometric Analysis.* Second edition. Prentice-Hall: Englewood Cliffs.

HANNAN, E.J. and QUINN. B.G. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society Series B.* **41**: 190-195.

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 Series B.* **50**: 157-224.

MORTON, A.S. and TUNNICLIFFE WILSON, G. (2001a). A Class of Modified High Order Autoregressive Models with Improved Resolution of Low Frequency Cycles. Submitted to *J. Time Series Analysis.*

MORTON, A.S. and TUNNICLIFFE WILSON, G. (2001b). Forecasting Economic Cycles Using Modified Autoregressions. *Proceedings of the Joint Statistical Meetings, Indianapolis. Aug. 2000. Bus. Econ. Stats. Section.* 207-212.

PEÑA, D. and BOX, G.E.P. (1987). Identifying a Simplifying Structure in Time Series. *Journal of the American Statistical Association.* **82**: 836-843.

PRIESTLEY, H.A. (1985). *Introduction to Complex Analysis*, Clarendon Press, Oxford.

REALE, M. and TUNNICLIFFE WILSON, G. (2001). Identification of Vector AR Models with Recursive Structural Errors Using Conditional Independence Graphs. *J. Ital. Statist. Soc.* (To appear).

SWANSON, N.R. and GRANGER, C.W.J. (1997). Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions. *Journal of the American Statistical Association.* **92**: 357-367.

SCHWARZ, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics.* **6**: 461-464.

TIAO, G.C. and TSAY, R.S. (1989). Model Specification in Multivariate Time Series. *Journal of the Royal Statistical Society Series B.* **51**: 157-213.

TUNNICLIFFE WILSON, G. (1973). The Estimation of Parameters in Multivariate Time Series Models. *Journal of the Royal Statistical Society Series B.* **35**: 76-85.

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