MONASH University

Department of Econometrics and Business Statistics

A Complete *VARMA* Modelling Methodology Based on Scalar Components

George Athanasopoulos and Farshid Vahid

January 2006

A complete *VARMA* modelling methodology based on scalar components

George Athanasopoulos

Monash University, Melbourne, Australia.

Farshid Vahid

Australian National University, Canberra, Australia.

Summary. This paper proposes an extension to scalar component methodology for the identification and estimation of *VARMA* models. The complete methodology determines the exact positions of all free parameters in any *VARMA* model with a predetermined embedded scalar component structure. This leads to an exactly identified system of equations that is estimated using full information maximum likelihood.

Keywords: Identification, Multivariate time series, Scalar components, *VARMA* models.

Address of correspondence: George Athanasopoulos, Monash University, Monash University, VIC 3800, Australia.
E-mail: George.Athanasopoulos@BusEco.monash.edu.au

1. Introduction

In 1989, George Tiao and Ruey Tsay presented their scalar component model (SCM) methodology for developing vector autoregressive moving average ($VARMA$) models to the Royal Statistical Society. Following the presentation, commentators including Chatfield, Hannan, Reinsel and Tunnicliffe-Wilson, acknowledged the importance of this methodology for multivariate time series analysis, and at the same time pointed out some of its shortcomings. Since then, with the exception of a small subsection in Tsay (1991) which addresses some of these concerns, there have not been any other efforts to advance this methodology. In this paper we extend the Tiao and Tsay (1989) (hereafter referred to as T&T) methodology to address the concerns of the commentators and we present a complete methodology for identifying and estimating $VARMA$ models.

The T&T methodology identifies scalar components embedded in a $VARMA$ model through a series of tests based on canonical correlations between judiciously chosen sets of variables, and then estimates these scalar components. The methodology is based on searching for linear combinations of variables that have simple dynamic structures and provides consistent but not efficient estimates of these linear combinations. The remaining parameters of the identified structure are then estimated conditional on the estimates of such linear combinations. Amongst other things, the commentators highlighted that more attention should be paid on the identification and the estimation of these linear combinations. One reason is that the accuracy of other parameter estimates, depends on how well these linear combinations are estimated. Furthermore, the ultimate purpose is to invert from these linear combinations back to the original variables under study. Hence, a modelling strategy must be able to produce reliable prediction intervals for the original series

rather than for these linear combinations. This is the general direction in which our paper extends the T&T methodology.

In this paper we establish a set of rules which reveal identification restrictions that produce a uniquely identified *VARMA* model with a predetermined scalar component structure. With these restrictions imposed, all parameters of the system are simultaneously estimated using full information maximum likelihood (*FIML*). In sharp contrast to the T&T methodology, the usual consistency and asymptotic efficiency properties of maximum likelihood estimators and prediction intervals do apply here. In the T&T methodology the parameters of the model are estimated in two stages. The complexity of the sampling distribution of the first stage estimators, makes the derivation of the correct standard errors for the second step estimators prohibitive. The users of this methodology provide standard errors for the second stage estimators, treating the first stage estimators as true parameters. This makes these standard errors unreliable for inference.

The structure of the paper is as follows. In Section (2) we present a brief overview of the T&T scalar component methodology. We define scalar component models, and discuss specific identification issues that arise within the scalar component framework. In Section (3) we propose an extension to the T&T procedure. We first present a summary of the concerns of the commentators of the T&T paper and then propose an extension which accounts for these concerns. This section concludes with an outline of all stages of our *VARMA* modelling methodology. In Section (4) we apply our methodology to two multivariate data sets. Section (5) concludes.

2. The scalar component methodology

The aim of exploring scalar component models (*SCM*s) is to examine whether there are any simplifying underlying structures embedded in a $VARMA(p,q)$ process.

DEFINITION 1. *For a given K dimensional $VARMA(p,q)$ process*

$$\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \eta_t - \mathbf{\Theta}_1 \eta_{t-1} - \ldots - \mathbf{\Theta}_q \eta_{t-q}, \tag{1}$$

a non-zero linear combination $z_t = \alpha' \mathbf{y}_t$, follows an $SCM(p_1,q_1)$ if α satisfies the following properties:

$$\alpha' \mathbf{\Phi}_{p_1} \neq \mathbf{0}^T \text{ where } 0 \leq p_1 \leq p, \tag{2}$$

$$\alpha' \mathbf{\Phi}_l = \mathbf{0}^T \text{ for } l = p_1 + 1, \ldots, p, \tag{3}$$

$$\alpha' \mathbf{\Theta}_{q_1} \neq \mathbf{0}^T \text{ where } 0 \leq q_1 \leq q, \tag{4}$$

$$\alpha' \mathbf{\Theta}_{q_l} = \mathbf{0}^T \text{ for } l = q_1 + 1, \ldots, q. \tag{5}$$

The scalar random variable z_t, depends only on lags 1 to p_1 of all variables and lags 1 to q_1 of all innovations in the system. Note that the univariate representation of this random variable is an $ARMA$ process, but of an order different from (p_1, q_1).

The T&T identification process begins from the most parsimonious possibility, i.e., $SCM(0,0)$, which is a system white noise. It then sequentially discovers K linearly independent vectors $(\alpha_1, \ldots, \alpha_K)$ which would rotate the $VARMA(p,q)$ system into a simpler dynamic structure with a significantly lower number of parameters. Hence, if we define matrix $\mathbf{A} = (\alpha_1, \ldots, \alpha_K)'$, then the transformation of a $VARMA(p,q)$ system by this matrix creates another $VARMA(p,q)$ system in terms of the transformed variables $\mathbf{z}_t$,

$$\mathbf{z}_t = \mathbf{\Phi}_1^* \mathbf{z}_{t-1} + \ldots + \mathbf{\Phi}_p^* \mathbf{z}_{t-p} + \varepsilon_t - \mathbf{\Theta}_1^* \varepsilon_{t-1} - \ldots - \mathbf{\Theta}_q^* \varepsilon_{t-q}, \tag{6}$$

where $\mathbf{z}_t = \mathbf{A}\mathbf{y}_t$, $\mathbf{\Phi}_i^* = \mathbf{A}\mathbf{\Phi}_i\mathbf{A}^{-1}$, $\varepsilon_t = \mathbf{A}\eta_t$ and $\mathbf{\Theta}_i^* = \mathbf{A}\mathbf{\Theta}_i\mathbf{A}^{-1}$. This transformed model incorporates whole rows of zero restrictions in the AR and the MA parameter matrices.

EXAMPLE 1. *Consider a trivariate stationary process* $\mathbf{y}_t$ *which has been identified by the T&T methodology to have three scalar components:* $z_{1,t} \sim SCM(1,1)$, $z_{2,t} \sim SCM(1,0)$ *and* $z_{3,t} \sim SCM(0,0)$, *i.e.,*

$$\mathbf{z}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{z}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & \theta_{13}^{(1)} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \tag{7}$$

Notice this system cannot yet be estimated as the parameters $\phi_{13}^{(1)}$ *and* $\theta_{13}^{(1)}$ *are not identified.*

CRITERION 1. *Generalised rule of elimination: Suppose that we have identified two scalar components* $z_{r,t} = SCM(p_r, q_r)$ *and* $z_{s,t} = SCM(p_s, q_s)$, *where* $p_r > p_s$ *and* $q_r > q_s$. *This implies that lags of* $z_{s,t}$ *of order* $1, \ldots, \min\{p_r - p_s, q_r - q_s\}$ *on the right hand side of the dynamic equation for* $z_{r,t}$, *can be written in terms of other variables on the right hand side of* $z_{r,t}$. *Hence, the parameters of the right hand side of the* $z_{r,t}$ *equation are not identified unless we set* $\min\{p_r - p_s, q_r - q_s\}$ *of them equal to zero. In fact, for each lag* $i = 1, ..., \min\{p_r - p_s, q_r - q_s\}$, *either* $\phi_{rs}^{(i)}$ *or* $\theta_{rs}^{(i)}$ *must be set to zero to obtain a uniquely identified system.*

For the system in equation (7) the rule of elimination implies that either one of $\phi_{13}^{(1)}$ or $\theta_{13}^{(1)}$ must be set to zero. In the proposed extension to the T&T methodology presented in Section (3) that follows, we require that the MA parameters are set to zero. Thus, we set $\theta_{13}^{(1)} = 0$, and the system becomes

$$\mathbf{z}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{z}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \tag{8}$$

The determination of embedded scalar component models is achieved through a series of canonical correlation tests. A $SCM(0,0)$ is a linear combination that is unpredictable from the past, and the analysis of canonical correlations between the present and the past to find and estimate such combinations is a direct generalisation of Hotelling (1935) to time series. If we denote the estimated squared canonical correlations between $\mathbf{y}_t$ and $\mathbf{Y}_{h,t-1} \equiv (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-1-h})'$ by $\widehat{\lambda}_1 < \widehat{\lambda}_2 < \ldots < \widehat{\lambda}_K$, then the likelihood ratio test statistic for at least s $SCM(0,0)$, against the alternative of less than s unpredictable components, is given by

$$C(s) = -(n-h)\sum_{i=1}^{s} \ln\left(1 - \widehat{\lambda}_i\right) \overset{a}{\sim} \chi^2_{s \times \{(h-1)K+s\}}. \tag{9}$$

The canonical covariates corresponding to insignificant canonical correlations will be consistent estimates of the scalar components. As shown by Vahid and Engle (1997) and Anderson and Vahid (1998) among others, a generalised method of moment (GMM) based test for the same hypothesis is $(n-h)\sum_{i=1}^{s}\widehat{\lambda}_i$, which is obviously asymptotically equivalent to the above. $SCM(m,0)$ can be found by similar test statistic based on squared canonical

correlations between $\mathbf{Y}_{m,t} \equiv (\mathbf{y}_t', \ldots, \mathbf{y}_{t-m}')'$ and $\mathbf{Y}_{h,t-1} \equiv (\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-h}')'$ where $h \geq m$. $SCM(m,j)$ however, are linear combinations of $\mathbf{Y}_{m,t}$ that cannot be linearly predicted from the history before $t - j$. Hence, the GMM test for this hypothesis estimates a linear combination of $\mathbf{Y}_{m,t}$ that is a moving average of order j and therefore is unpredictable from $\mathbf{Y}_{h,t-1-j} \equiv (\mathbf{y}_{t-1-j}', \ldots, \mathbf{y}_{t-1-j-h}')'$. This imposes a structure on the GMM weighting matrix. The test then is a test of overidentifying restrictions in this system. Alternatively, T&T suggest the statistic

$$C(s) = -(n - h - j) \sum_{i=1}^{s} \ln \left\{ 1 - \frac{\widehat{\lambda}_i}{d_i} \right\} \stackrel{a}{\sim} \chi^2_{s \times \{(h-m)K+s\}} \qquad (10)$$

based on the squared canonical correlations between $\mathbf{Y}_{m,t}$ and $\mathbf{Y}_{h,t-1-j}$. d_i is a correction factor that accounts for the fact that the canonical variates in this case can be moving averages of order j. Specifically,

$$d_i = 1 + 2 \sum_{v=1}^{j} \widehat{\rho}_v \left(\widehat{\mathbf{r}}_i' \mathbf{Y}_{m,t} \right) \widehat{\rho}_v \left(\widehat{\mathbf{g}}_i' \mathbf{Y}_{h,t-1-j} \right) \qquad (11)$$

where $\widehat{\rho}_v(.)$ is the v order autocorrelation of its argument and $\widehat{\mathbf{r}}_i' \mathbf{Y}_{m,t}$ and $\widehat{\mathbf{g}}_i' \mathbf{Y}_{h,t-1-j}$ are the canonical variates corresponding to the ith canonical correlation between $\mathbf{Y}_{m,t}$ and $\mathbf{Y}_{h,t-1-j}$.

Since an $SCM(m,j)$ nests all scalar components of order $(\leq m, \leq j)$, lower order SCMs are also identified when testing for higher orders. T&T arrange these test results in a two way tabulation and also provide a complete set of rules that determines the order of all parsimonious SCMs embodied in a system (see the empirical examples in Section 4).

Furthermore, T&T deduce a consistent estimate of the transformation matrix $\mathbf{A}$ from the estimated canonical coefficients. The eigenvectors corresponding to the statistically insignificant canonical correlations provide consistent estimators for each α in $\mathbf{A} = (\alpha_1, \ldots, \alpha_K)'$. However, since lower order SCMs are also identified, there will be several candidate estimators for the linear space spanned by αs. T&T suggest a procedure to find a consistent estimator of $\mathbf{A}$ from the union of the eigenvectors corresponding to zero eigenvalues in the above tests. Although this is a consistent estimator, its asymptotic variance will be complicated. Using $\widehat{\mathbf{A}}$, they form $\mathbf{z}_t = \widehat{\mathbf{A}} \mathbf{y}_t$ and then estimate $\mathbf{\Phi}_1^*, \ldots, \mathbf{\Phi}_p^*$ and $\mathbf{\Theta}_1^*, \ldots, \mathbf{\Theta}_q^*$. Since the estimates of these coefficient matrices are conditional on $\widehat{\mathbf{A}}$, their asymptotic distribution will also be complex.

3. An extension to the Tiao and Tsay methodology

A major concern of the participants in the discussion of the T&T paper (see the discussion that followed the paper), was the general treatment of the transformation matrix $\mathbf{A}$. The comments by Professors Chatfield, Hannan, Reinsel and Tunnicliffe-Wilson amongst others, can be summarised in the following concerns about the T&T modelling procedure:

(a) The identified $VARMA(p,q)$ model is being stated in terms of the transformed series $\mathbf{z}_t$ and not in terms of the series of interest, the original series $\mathbf{y}_t$.

(b) The number of parameters to be estimated in the $\mathbf{A}$ matrix should be included in the total number of parameters estimated for the model. This makes the reduction in degrees of freedom for the estimation, smaller than what is claimed by T&T.

(c) The T&T procedure calculates the transformation matrix $\mathbf{A}$ via the canonical correlations analysis stated above. This calculation does not produce the most efficient (although consistent) estimates for the parameters, in particular for the case of SCMs with $q > 0$.

(d) Finally, the T&T procedure does not produce standard errors for the parameter estimates in $\mathbf{A}$.

We propose an extension to the T&T modelling procedure to address these concerns. This extension develops identification conditions for the parameters of matrix $\mathbf{A}$, so that the entire system can be estimated efficiently. It should be mentioned that Tsay (1991) discusses the issue of redundant parameters in $\mathbf{A}$ and provides a formula for counting the free parameters in this matrix. Here we establish a set of rules for determining the free parameters in $\mathbf{A}$. Also, since unique identification of $\mathbf{A}$ requires normalising some of its elements to one, we suggest a procedure for ensuring that parameters that are zero are not normalised to one. Finally, we propose full information maximum likelihood estimation for the parameters of the identified $VARMA$ model, rather than relying on the estimated canonical covariates.

Our extension to the T&T process begins after the orders of K scalar components have been determined by the canonical correlations based procedure. T&T proceed by obtaining an estimate of $\mathbf{A}$, forming $\mathbf{z}_t$ and its lags using $\widehat{\mathbf{A}}$, and then estimating the restricted $VARMA$ model given by (6) in which $\mathbf{\Phi}_1^*, \ldots, \mathbf{\Phi}_p^*$ and $\mathbf{\Theta}_1^*, \ldots, \mathbf{\Theta}_q^*$ have many zero restrictions. We instead rewrite the system in terms of the original series $\mathbf{y}_t$, establish further restrictions to identify $\mathbf{A}$, and then estimate the system by full information maximum likelihood.

In our procedure we require that, if the SCM orders are such that the general elimination rule needs to be used, the zero restrictions must be applied to the MA coefficients. This does not entail any loss of generality. As we discussed in Criterion (1) we are free to set either one of the MA coefficient or the AR coefficient to zero. Choosing to set the MA coefficient to zero implies that the only type of restrictions that the SCM orders place on $\mathbf{\Phi}_1^*, \ldots, \mathbf{\Phi}_p^*$, is that entire rows are zero. This means (as is shown in Lemma 1 below) that we can replace $\mathbf{z}_{t-1}, \ldots, \mathbf{z}_{t-p}$ on the right hand side of equation (6) with $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}$ without changing the structure of the system.

LEMMA 1. *In equation* (6) *if we replace* $\mathbf{z}_{t-1}, \ldots, \mathbf{z}_{t-p}$ *by* $\mathbf{A}\mathbf{y}_{t-1}, \ldots, \mathbf{A}\mathbf{y}_{t-p}$ *and obtain the system*

$$\mathbf{z}_t = \mathbf{\Psi}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Psi}_p \mathbf{y}_{t-p} + \varepsilon_t - \mathbf{\Theta}_1^* \varepsilon_{t-1} - \ldots - \mathbf{\Theta}_q^* \varepsilon_{t-q}, \quad (12)$$

then $\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_p$ *will have the same zero restrictions as* $\mathbf{\Phi}_1^*, \ldots, \mathbf{\Phi}_p^*$.

PROOF. $\mathbf{\Psi}_i = \mathbf{\Phi}_i^* \mathbf{A}$ for $i = 1, \ldots, p$, and since the only zero restriction in $\mathbf{\Phi}_i^*$ is that one or more rows are zero, the corresponding rows of $\mathbf{\Psi}_i$ will be zero.

Lemma (1) shows that we can think of the system as

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Psi}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Psi}_p \mathbf{y}_{t-p} + \varepsilon_t - \mathbf{\Theta}_1^* \varepsilon_{t-1} - \ldots - \mathbf{\Theta}_q^* \varepsilon_{t-q}, \quad (13)$$

with $\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_p$ and $\mathbf{\Theta}_1^*, \ldots, \mathbf{\Theta}_q^*$ satisfying the same restrictions as the right hand side parameter matrices of equation (6). However, the system of equation (13) is not yet identified because not all parameters in $\mathbf{A}$ are free. For example, each row of the system can be multiplied by an arbitrary non-zero constant without changing the structure of the system. Hence we need to restrict $\mathbf{A}$ in order for the system to be uniquely identified.

DEFINITION 2. **A** *is identified if, and only if, the only matrix* **H** *such that*

$$\mathbf{H}\mathbf{A}\mathbf{y}_t = \mathbf{H}\boldsymbol{\Psi}_1\mathbf{y}_{t-1} + \ldots + \mathbf{H}\boldsymbol{\Psi}_p\mathbf{y}_{t-p} + \mathbf{H}\varepsilon_t - \mathbf{H}\boldsymbol{\Theta}_1^*\varepsilon_{t-1} - \ldots - \mathbf{H}\boldsymbol{\Theta}_q^*\varepsilon_{t-q} \quad (14)$$

has the same restrictions as (13), *is* $\mathbf{H} = \mathbf{I}_K$.

Let us assume that the ith row of the system is a $SCM(p_i, q_i)$. The zero restrictions on the right and side of the system imply that, the ith row of **H** can only be different from a multiple of a row of an identity matrix, if there are other $SCM(p_j, q_j)$ such that $p_i \geq p_j$ and $q_i \geq q_j$. Therefore, only in such situations we need to impose some restrictions on **A** to eliminate this possibility. Other than this, normalisation of one element in the ith row of **A** to one, ensures that the ith row of **H** will be a row of an identity matrix. The rules for achieving unique identification of the system are therefore the following:

(a) Each row of **A** can be multiplied by a constant without changing the structure of the model. Hence, we are free to normalise one parameter in each row to one. However, as always in such situations, there is a danger of wrongly choosing a parameter whose true value is zero for normalisation, i.e., a zero parameter might be normalised to one. We ignore this possibility for the time being, but we will return to it later.
(b) Any linear combination of a $SCM(p_1, q_1)$ and a $SCM(p_2, q_2)$ is a $SCM(\max\{p_1, p_2\}, \max\{q_1, q_2\})$. In all cases where there are two embedded scalar components with weakly nested orders, i.e., $p_1 \geq p_2$ and $q_1 \geq q_2$ arbitrary multiples of $SCM(p_2, q_2)$ can be added to the $SCM(p_1, q_1)$ without changing the structure. This means that the row of **A** corresponding to the $SCM(p_1, q_1)$ is not identified in this case. To achieve identification, if the parameter in the ith column of the row of **A** corresponding to the $SCM(p_2, q_2)$ is normalised to one, the parameter in the same position in the row of **A** corresponding to $SCM(p_1, q_1)$ should be restricted to zero.

Note that if $p_1 = p_2$ and $q_1 = q_2$ then rule (b) is applied twice because $p_1 \geq p_2$ and $q_1 \geq q_2$ and at the same time $p_2 \geq p_1$ and $q_2 \geq q_1$. This creates an identity submatrix in **A**. Also note that if there is a single SCM whose autoregressive order is the smallest of all other $SCMs$ in the system, the corresponding row of **A** is uniquely identified. This is because any combination of other $SCMs$ with this one produces a SCM with a longer autoregressive order and changes the structure. By the same token, if there is a single SCM with minimal moving average order, the row of **A** corresponding to it is uniquely identified. The following example highlights the above points.

EXAMPLE 2. *Consider the trivariate process in equation* (8). *Lemma* (1) *shows that we can replace* $\mathbf{z}_{t-1}$ *with* $\mathbf{y}_{t-1}$ *without changing the structure of the system, i.e.,*

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \quad (15)$$

Ignoring the **A** *matrix, the parsimony of this model is exaggerated, in the sense that it seems to have ten parameters less than a* $VARMA(1,1)$. *However this is not the case, as the commentators of the T&T paper have pointed out. The parameters in* **A** *have to be*

estimated. However, not all nine of them are free parameters. Firstly, one parameter per row, say diagonal elements, can be normalised to one

$$\begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \quad (16)$$

The third equation is now uniquely identified, because no combination of the other two equations can be added to it and still keep its SCM $(0,0)$ structure. However, since a SCM $(0,0)$ is nested in a SCM $(1,0)$ and SCM $(1,1)$, the third rule above tells us that a_{13} and a_{23} can be set to zero without changing the structure of the system. This leads to

$$\begin{bmatrix} 1 & a_{12} & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \quad (17)$$

The same rule again applies and we can set a_{12} to zero. Thus, the final result is the canonical SCM VARMA representation,

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \quad (18)$$

Notice that the real reduction in the number of parameters is $10 - 3 = 7$.

Imposing restrictions on $\mathbf{A}$ leads to a uniquely identified *VARMA* structure which can be estimated by full information maximum likelihood. We call this structure a canonical *SCM VARMA* representation.

DEFINITION 3. *A canonical SCM VARMA representation is one that:*

(a) *The orders of the SCMs are as small as possible;*
(b) *The zero restrictions needed to account for the redundant parameters have been set on the moving average coefficients rather than the autoregressive ones; and*
(c) *All the redundant parameters in the transformation matrix $\mathbf{A}$ have been restricted to achieve unique identification.*

In the first rule for achieving unique identification, we stated that any *non-zero* element in a row of $\mathbf{A}$ can be normalised to one. A problem can arise in practice if we decide to normalise to one an element whose true value is zero. To safeguard against this, our procedure adds tests of predictability using subsets of variables. Starting from the *SCM* with the smallest order (the *SCM* with minimum $p + q$), exclude one variable, say the Kth variable, and test if a *SCM* of the same order can be found using the $K - 1$ variables alone. If the test is rejected, the coefficient of the Kth variable is then normalised to one and by applying the second rule of unique identification the corresponding coefficients in all other *SCM*s that nest this one are set to zero. If the test concludes that the *SCM* can be formed using the first $K - 1$ variables only, the coefficient of the Kth variable in this *SCM* is zero, and should not be normalised to one. It is worth noting that if the order of this *SCM* is

uniquely minimal, then this extra zero restriction adds to the restrictions discovered before. Continue testing by leaving variable $K - 1$ out and test if the SCM could be formed from the first $K - 2$ variables only, and so on. These tests are all GMM tests (see Hansen 1982) with suitably chosen weighting matrices given by the structure of the system.

To summarise, our complete $VARMA(p, q)$ modelling procedure comprises of the following three stages:

Stage I: Identification of the scalar components.   This stage follows the T&T methodology. The only extra condition we require is that if the rule of elimination is applied, then the zero restrictions are placed on the MA coefficients.

Stage II: Placing identification restrictions on matrix **A**.   Stage II of the identification process applies the identification rules stated above to identify the structure of the transformation matrix **A**. Extensive Monte-Carlo experiments in Athanasopoulos (2005) show that these first two stages of the proposed methodology perform well in identifying some pre-specified data generating processes with various orders of embedded SCMs.

Stage III: Estimation of the uniquely identified system.   This stage uses $FIML$ to estimate the parameters of the uniquely identified structures. The canonical correlations procedure produces good starting values for the parameters, in particular for the SCMs with no moving average components. Alternatively, lagged innovations can be estimated from a long VAR and used for obtaining initial estimates for the parameters as in Hannan and Risannen (1982). The maximum likelihood procedure provides estimates and estimated standard errors for all parameters, including the free parameters of **A**. All usual considerations that ease the estimation of structural forms are also valid here, and should be exploited in estimation.

4.   Empirical Examples

The following two empirical examples illustrate the application of each of the stages of the proposed methodology.

4.1.   US Flour Price Data

This data set has been previously analysed and modelled by Tiao and Tsay (1989), Grubb (1992) and Lütkepohl and Poskitt (1996). The data consists of three monthly series on flour price indices from August 1972 to November 1980, i.e., $N = 100$ observations, for the cities of: Buffalo, Minneapolis, and Kansas. The logarithms of the series are plotted in Figure (1).

In Stage I of the identification process we identify the overall order of the $VARMA$ model and the orders of embedded SCMs in the data. Panel A of Table (1) reports the results of all canonical correlations test statistics divided by their χ^2 critical values. T&T call this table the "Criterion Table". If the entry in the $(m, j)^{th}$ cell is less than one, it shows that there are three SCMs of order (m, j) or lower in this system. Hence, we deduce from this table that the smallest the overall order of the system can be, is either a $VARMA(1, 1)$ or a $VAR(2)$. Given that the primary goal of this research is $VARMA$ modelling, and to be consistent with T&T, we choose a $VARMA(1, 1)$ as the overall order. Conditional on
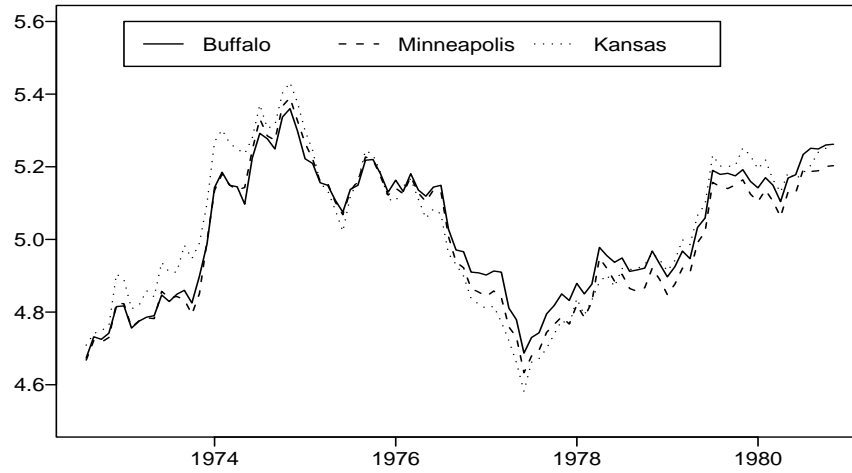
Fig. 1. Logarithms of US flour prices

Table 1. Tables for Stage I of the identification process for the log of US flour price data

	PANEL A: Criterion Table						**PANEL B**: Root Table				
	j						j				
m	0	1	2	3	4	m	0	1	2	3	4
0	34.17^a	5.8	3.0	2.11	1.68	0	**0**	**0**	1	1	1
1	2.38	**0.44**	0.49	0.22	0.34	1	**2**	**3**	3	3	3
2	**0.25**	0.58	0.60	0.49	0.46	2	3	5	6	6	6
3	0.37	0.46	0.67	0.53	0.58	3	3	6	8	9	9
4	0.73	0.62	0.57	0.70	0.77	4	3	6	9	11	12
^{a} *The statistics are normalised by the corresponding 5% χ^2 critical values*											

this overall order, the $C(s)$ tests are performed again to identify the individual orders of embedded SCMs. The number of insignificant canonical correlations found, are tabulated in Panel B of Table (1), referred to as the "Root Table". In the root table, the bold type shows that two scalar components of order $(1,0)$ are initially identified in position $(m,j) = (1,0)$. Then, there are three SCMs of order $(1,1)$ at position $(m,j) = (1,1)$. From these three, the first two are carried through from the previous two $(1,0)$ scalar components, and the other is a new scalar component of order $(1,1)$. Hence, the identified $VARMA(1,1)$ consists of two $SCM(1,0)$ and one $SCM(1,1)$.

Using the identification rules for matrix $\mathbf{A}$, a canonical SCM representation of the $VARMA(1,1)$ model identified would be

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & 0 & 1 \end{bmatrix} \mathbf{y}_t = \mathbf{c} + \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ \psi_{31}^{(1)} & \psi_{32}^{(1)} & \psi_{33}^{(1)} \end{bmatrix} \mathbf{y}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & \theta_{13}^{(1)} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \quad (19)$$

Testing for the implemented normalisations, i.e., normalising the diagonal parameters of $\mathbf{A}$ to one, does not seem to be a problem. Furthermore, $y_{3,t}$ is found to be a $SCM(1,0)$ on its

Table 2. Estimation results for the log of the US flour price data

Estimated $VARMA$ model: $\widehat{\mathbf{A}}\mathbf{y}_t = \widehat{\mathbf{c}} + \widehat{\boldsymbol{\Psi}}_1\mathbf{y}_{t-1} + \varepsilon_t - \widehat{\boldsymbol{\Theta}}_1^*\varepsilon_{t-1}$

$$\widehat{\mathbf{A}} = \begin{bmatrix} 1 & 0 & 0 \\ \underset{(-20.10)^a}{-0.521} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \widehat{\mathbf{c}}' = \begin{bmatrix} \underset{(5.83)}{0.199} & \underset{(3.64)}{0.116} & \underset{(7.35)}{0.336} \end{bmatrix}$$

$$\widehat{\boldsymbol{\Psi}}_1 = \begin{bmatrix} \underset{(33.78)}{1.078} & \underset{(-11.39)}{-0.392} & \underset{(10.64)}{0.273} \\ \underset{(-17.15)}{-0.454} & \underset{(27.61)}{0.741} & \underset{(8.73)}{0.168} \\ \underset{(0.18)}{0.012} & \underset{(-4.98)}{-0.297} & \underset{(45.30)}{1.217} \end{bmatrix} \qquad \widehat{\boldsymbol{\Theta}}_1^* = \begin{bmatrix} \underset{(-4.45)}{-0.616} & \underset{(10.38)}{0.812} & \underset{(1.59)}{0.122} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_\varepsilon = \begin{bmatrix} 1.977 & & \\ 1.012 & 0.648 & \\ 1.998 & 1.119 & 2.571 \end{bmatrix} \times 10^{-3} \qquad \left|\widehat{\boldsymbol{\Sigma}}_\varepsilon\right| = 1.235 \times 10^{-10}$$

Res C.T.[b]

	j			
m	0	1	2	3
0	**0.60**[c]	0.43	0.38	0.42
1	0.44	0.56	0.41	0.41
2	0.38	0.45	0.56	0.51
3	0.57	0.51	0.54	0.56

[a] $t-statistics$ in parentheses
[b] Criterion Table for the residuals from the estimated model
[c] The statistics are normalised by the corresponding 5% χ^2 critical values

own. Therefore, coefficient a_{31} can be set to zero. The resulting estimated $VARMA(1,1)$ model is shown in Table (4.1).

The appropriateness of the model is checked by applying the $C(s)$ test to the residuals. The criterion table, presented at the bottom of Table (4.1), shows that three white noise processes are identified, i.e., the residuals follow a vector white noise process.

Comparison to the T&T Results   T&T also pursue a $VARMA(1,1)$ model of the flour price data. The estimated model is presented in two steps. In the first step the estimated transformation matrix $\widehat{\mathbf{A}}$ is presented and the linearly transformed series,

$$\mathbf{z}_t = \widehat{\mathbf{A}}\mathbf{y}_t, \tag{20}$$

is formed. Then, a standard $VARMA(1,1)$ of the form,

$$\mathbf{z}_t = \mathbf{c}^* + \boldsymbol{\Phi}_1^*\mathbf{z}_{t-1} + \varepsilon_t - \boldsymbol{\Theta}_1^*\varepsilon_{t-1},$$

is estimated. The estimation results are presented in Table (4.1). The contribution of the suggested extension to the T&T methodology becomes clear when comparing the two sets of results. Firstly, the T&T $VARMA(1,1)$ model is based on 24 estimated parameters in comparison to the 16 freely varying parameters of the model identified by our procedure. Also, the $\widehat{\boldsymbol{\Phi}}_1^*$ and $\widehat{\boldsymbol{\Theta}}_1^*$ estimates in T&T are two-step estimates based on the estimates of $\widehat{\mathbf{A}}$.

Table 3. Tiao and Tsay (1989) estimation results for the logarithms of the US flour price data

$$
\text{Estimated model: } \mathbf{z}_t = \widehat{\mathbf{c}}^* + \widehat{\boldsymbol{\Phi}}_1^* \mathbf{z}_{t-1} + \varepsilon_t - \widehat{\boldsymbol{\Theta}}_1^* \varepsilon_{t-1} \text{ where } \mathbf{z}_t = \widehat{\mathbf{A}} \mathbf{y}_t
$$

$$
\widehat{\mathbf{A}} = \begin{bmatrix} -0.40 & 0.83 & -0.40 \\ 0.61 & -0.51 & -0.60 \\ 0.55 & 0.82 & -0.06 \end{bmatrix}
\qquad
\widehat{\mathbf{c}}^{*'} = \begin{bmatrix} \underset{(1.37)}{0.26} & \underset{(-1.78)}{-0.16} & \underset{(-0.67)}{-0.02} \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\Phi}}_1^* = \begin{bmatrix} \underset{(14.67)^a}{0.88} & \underset{(-0.67)}{-0.02} & \underset{(-0.00)}{-0.00} \\ \underset{(1.69)}{0.27} & \underset{(12.75)}{1.02} & \underset{(0.75)}{0.03} \\ \underset{(-2.40)}{-0.84} & \underset{(-0.67)}{-0.12} & \underset{(10.33)}{0.93} \end{bmatrix}
\qquad
\widehat{\boldsymbol{\Theta}}_1^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \underset{(-4.77)}{-1.48} & \underset{(5.53)}{1.05} & \underset{(5.78)}{0.52} \end{bmatrix}
$$

$$
\widehat{\boldsymbol{\Sigma}}_\varepsilon = \begin{bmatrix} 0.12 & & \\ -0.00 & 0.85 & \\ 0.20 & -1.47 & 3.64 \end{bmatrix} \times 10^{-3}
\qquad
\left| \widehat{\boldsymbol{\Sigma}}_\varepsilon \right| = 7.79 \times 10^{-7}
$$

[a] $t-$ *statistics in parentheses. These test statistics are conditional on* $\widehat{\mathbf{A}}$, *which makes them unreliable as they do not incorporate the uncertainty in the estimation of* $\mathbf{A}$.

Therefore, if the uncertainty in $\widehat{\mathbf{A}}$ is not taken into account, the standard errors and hence the $t-$statistics of the second stage estimation are invalid. In contrast, our estimates of the parameter matrices are efficiently estimated using *FIML* and have correct standard errors. Finally, T&T present their estimated model in terms of the transformed series $\mathbf{z}_t$ which is not necessarily the series of interest whereas our model is in terms of $\mathbf{y}_t$.

The results presented from the two studies under consideration are not directly comparable as they stand. To make these results directly comparable we present the reduced form of each of the models in Table (4.1). Panel A presents the reduced form of the model estimated by our proposed methodology. The corresponding reduced form of the model estimated by T&T is presented in Panel B. In Panel C we present the reduced form of the T&T model produced by Grubb (1992). The results between the reduced form of the T&T model calculated here and the one presented by Grubb (1992) vary slightly. This could be due to rounding error. The determinants of the covariance matrices indicate that our model fits the data best, despite requiring less parameters.

4.2. US Macroeconomic Data

The aim of the above example was to illustrate the proposed modelling procedure and to compare the resulting estimated *VARMA* model with previous studies. This highlighted the advantage of our proposed procedure in the efficient estimation of parameters and in calculating standard errors. However, due to the short sample size, it was not feasible to investigate the forecasting performance of estimated models. In our second example, we use three monthly macroeconomic time series that are often modelled jointly in empirical models of the business cycle. The relatively large number of observations in this example allows us to hold some observation for out-of-sample forecast evaluation. The forecast performance of the estimated *VARMA* model is evaluated against two *VAR* models selected by different model selection criteria. Vector autoregressive models are the most popular models in the

Table 4. Reduced form of the estimated models for the log of the US flour price data

---

PANEL A: Athanasopoulos and Vahid

$$\hat{\mathbf{c}}^r = \begin{bmatrix} 0.199 \\ 0.219 \\ 0.336 \end{bmatrix} \qquad \hat{\boldsymbol{\Psi}}_1^r = \begin{bmatrix} 1.078 & -0.392 & 0.273 \\ 0.107 & 0.536 & 0.310 \\ 0.012 & -0.297 & 1.217 \end{bmatrix} \qquad \hat{\boldsymbol{\Theta}}_1^r = \begin{bmatrix} -1.039 & 0.812 & 0.122 \\ -0.541 & 0.423 & 0.064 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\left| \hat{\boldsymbol{\Sigma}}_\eta^{AV} \right| = \left| \hat{\mathbf{A}}^{-1} \right| \left| \hat{\boldsymbol{\Sigma}}_\varepsilon \right| \left| \hat{\mathbf{A}}^{-1^T} \right| = \left| \hat{\boldsymbol{\Sigma}}_\varepsilon \right| = 1.235 \times 10^{-10}$$

PANEL B: Tiao and Tsay (1989)

$$\hat{\mathbf{c}}^r = \begin{bmatrix} -0.254 \\ 0.138 \\ -0.109 \end{bmatrix} \qquad \hat{\boldsymbol{\Phi}}_1^r = \begin{bmatrix} 1.153 & -0.488 & 0.293 \\ 0.181 & 0.450 & 0.322 \\ 0.133 & -0.427 & 1.226 \end{bmatrix} \qquad \hat{\boldsymbol{\Theta}}_1^r = \begin{bmatrix} 1.393 & -1.227 & -0.063 \\ 0.960 & -0.846 & -0.043 \\ 0.600 & -0.528 & -0.027 \end{bmatrix}$$

$$\left| \hat{\boldsymbol{\Sigma}}_\eta^{T\&T} \right| = \left| \hat{\mathbf{A}}^{-1} \right| \left| \hat{\boldsymbol{\Sigma}}_\varepsilon \right| \left| \hat{\mathbf{A}}^{-1^T} \right| = \left| \hat{\boldsymbol{\Sigma}}_\varepsilon \right| \left| \hat{\mathbf{A}} \right|^{-2} = 1.332 \times 10^{-10}$$

PANEL C: Grubb (1992): Reduced form of the T&T model

$$\hat{\boldsymbol{\Phi}}_1 = \begin{bmatrix} 1.24 & -0.56 & 0.27 \\ 0.30 & 0.36 & 0.29 \\ 0.25 & 0.51 & 1.22 \end{bmatrix} \qquad \hat{\boldsymbol{\Theta}}_1 = \begin{bmatrix} 1.22 & -1.12 & -0.04 \\ 0.84 & -0.77 & -0.03 \\ 0.53 & -0.48 & -0.02 \end{bmatrix}$$

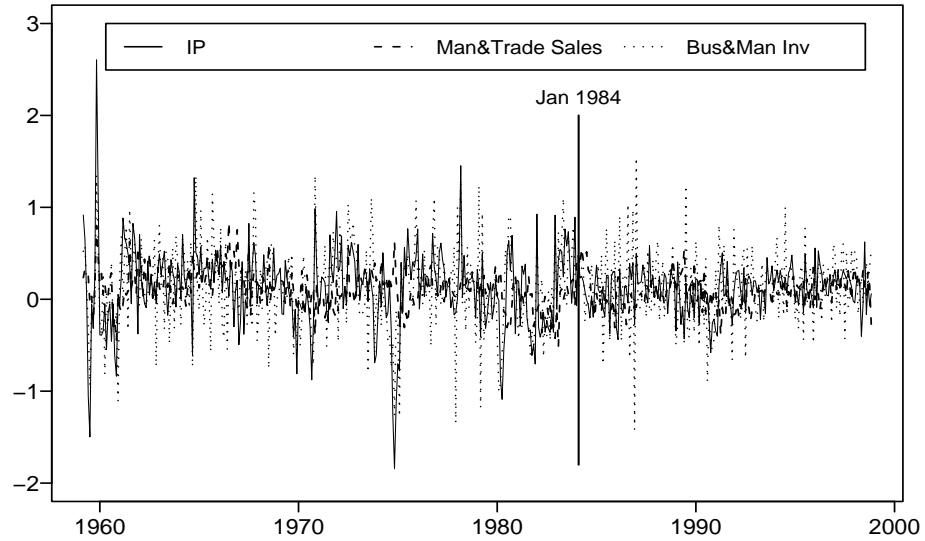$$\left| \hat{\boldsymbol{\Sigma}}_\eta^{GR} \right| = 1.281 \times 10^{-10}$$

Fig. 2. Growth rates of US macroeconomic data

applied time series literature.

The data employed are three monthly macroeconomic US time series extracted from the multivariate data set compiled by Stock and Watson (1999):

- Industrial Production: seasonally adjusted total index (1992=100);

- Manufacturing & Trade Sales: seasonally adjusted total in millions of chained 1992 dollars; and

- Business & Manufacturing Inventories: seasonally adjusted total in millions of chained 1992 dollars.

The series span from March 1959 to December 1998, i.e., $N = 478$ observations. We model the growth rate of these series ($100 \times \Delta \ln (y_{i,t})$ for $i = 1, 2, 3$) jointly. The time series of growth rates are plotted in Figure (2).

We have divided the data into two sub-samples: the estimation sample (March 1959 to December 1983 with $N_1 = 298$ observations) and the hold-out sample (January 1984 to December 1998 with $N_2 = 180$ observations). We estimate each model once and for all using the estimation sample, i.e., all models are estimated using $\mathbf{y}_1$ to $\mathbf{y}_{N_1}$. We then use each estimated model to produce a sequence of h-step-ahead forecasts for $h = 1$ to 15. That is, with $\mathbf{y}_{N_1}$ as the forecast origin, we produce forecasts for $\mathbf{y}_{N_1+1}$ to $\mathbf{y}_{N_1+15}$. The forecast origin is then rolled forward one period, i.e., using observation $\mathbf{y}_{N_1+1}$, we produce forecasts for $\mathbf{y}_{N_1+2}$ to $\mathbf{y}_{N_1+16}$. We repeat this process to the end of the hold-out sample. Therefore, for each model and each forecast horizon h, we have $N_2 - h + 1$ forecasts to use for forecast evaluation purposes.

Table 5. Tables for Stage I of the identification process for the US macroeconomic data

	PANEL A: Criterion Table						**PANEL B**: Root Table				
	j						j				
m	0	1	2	3	4	m	0	1	2	3	4
0	9.11^a	4.17	3.16	2.55	1.85	0	**0**	1	2	2	2
1	2.92	**0.63**	0.82	0.79	1.14	1	**1**	**3**	4	5	5
2	1.39	0.54	1.23	1.14	0.81	2	2	4	6	7	8
3	1.11	1.12	0.89	0.81	1.05	3	2	5	7	9	10
4	1.39	1.01	0.97	1.12	0.99	4	2	5	8	10	12

^{a} *The statistics are normalised by the corresponding 5% χ^2 critical values*

4.2.1.   Model identification and estimation

The Criterion Table in Panel A of Table (4.2.1) shows that the minimum possible overall order of the system is a $VARMA(1,1)$. Reading from the root table in Panel B of Table (4.2.1), we identify one exchangeable $SCM(1,0)$ or $SCM(0,1)$ (see the note on exchangeable models that follows) and two further $SCM(1,1)$.

A note on exchangeable models   In multivariate time series analysis, we sometimes find pathological cases where a process can be represented either as a finite $VAR(p)$ or a finite $VMA(q)$ process. For instance, we might have the bivariate $VARMA(1,0)$ process

$$\mathbf{y}_t = \left[\begin{array}{cc} a & a \\ -a & -a \end{array} \right] \mathbf{y}_{t-1} + \eta_t, \tag{21}$$

which can be equivalently represented as

$$\mathbf{y}_t = \eta_t - \left[\begin{array}{cc} -a & -a \\ a & a \end{array} \right] \eta_{t-1}. \tag{22}$$

For more on such equivalent representations see Lütkepohl (1991). Through the transformation matrix $\mathbf{A} = (\alpha_1, \alpha_2)'$, where $\alpha_1 = (1,0)'$ and $\alpha_2 = (1,1)'$, equation (21) leads to

$$\left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right] \mathbf{y}_t = \left[\begin{array}{cc} a & a \\ 0 & 0 \end{array} \right] \mathbf{y}_{t-1} + \varepsilon_t, \tag{23}$$

i.e., a $SCM(1,0)$ and a $SCM(0,0)$. However through the same transformation matrix $\mathbf{A}$, equation (22) leads to

$$\left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right] \mathbf{y}_t = \varepsilon_t - \left[\begin{array}{cc} -a & -a \\ 0 & 0 \end{array} \right] \varepsilon_{t-1}, \tag{24}$$

i.e., a $SCM(0,1)$ and a $SCM(0,0)$. Such exchangeable models produce a recognisable pattern in the Root Table (as shown in Table (4.2.1)). The choice between the two models is arbitrary. In an extensive forecasting study, Athanasopoulos and Vahid (2005) conclude that the choice between such exchangeable models does not make a significant difference in terms of forecasting accuracy. In this example we proceed with the $SCM(0,1)$ and only report the result for this case. We have checked that the qualitative results reported below are independent of this choice.

Using the identification rules for matrix $\mathbf{A}$, a canonical SCM representation of the $VARMA\,(1,1)$ model would be

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_{31} & a_{21} & 1 \end{bmatrix} \mathbf{y}_t = \mathbf{c} + \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \varepsilon_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & \theta_{13}^{(1)} \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} & \theta_{23}^{(1)} \\ \theta_{31}^{(1)} & \theta_{32}^{(1)} & \theta_{33}^{(1)} \end{bmatrix} \varepsilon_{t-1}. \quad (25)
$$

In the second stage of our procedure, when testing for appropriate normalisations in matrix $\mathbf{A}$, we discover that a linear combination of $y_{1,t}$ and $y_{3,t}$ alone is a $SCM\,(0,1)$, i.e. a_{21} is not significantly different from zero and should not be normalised to 1. Therefore, we set a_{21} to zero. The resulting estimated $VARMA\,(1,1)$ model is shown in Panel A of Table (4.2.1). The Criterion Table for the residuals at the bottom of Panel A shows that the residuals follow a vector white noise process.

4.2.2. Forecast evaluation

In order to evaluate the out-of-sample forecast performance of the estimated $VARMA\,(1,1)$ model, we employ two VAR models as multivariate alternatives, a $VAR\,(12)$ and a $VAR\,(2)$. These models are respectively selected by the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) from all VAR models up to a maximum lag length of 24. Obviously the $VAR\,(12)$ model has the best in-sample fit, as shown by the determinant of the estimated error covariance matrices presented in Panel B of Table (4.2.1). It is important to note that the $VARMA\,(1,1)$ produces a better fit than the $VAR\,(2)$ despite having two less parameters. This means that if the class of models considered had included $VARMA$ models, then at least BIC would have chosen $VARMA\,(1,1)$ in this case. However, in multivariate time series applications, in particular in applied macroeconomics, vector autoregressions are the only class of models that are routinely considered.

For each forecast horizon h, we consider two measures of forecasting accuracy. The first is the determinant of the mean squared forecast error matrix, $|MSFE_h|$, and the second is the trace of the mean squared forecast error matrix, $tr\,(MSFE_h)$. Clements and Hendry (1993) show that the $|MSFE|$ is invariant to elementary operations on the forecasts of different variables at a single horizon, but not invariant to elementary operations on the forecasts across different horizons. The $tr\,(MSFE_h)$ is not invariant to either. In this forecast evaluation exercise, we present both of these measures.

For each forecast horizon h, Table (4.2.2) (see also Figure (3)) reports the percentage improvement (PI_h) in the forecast error measures of the $VARMA\,(1,1)$ model over the VAR alternatives. For example in terms of the $|MSFE|$ the PI for forecast horizon h is calculated as

$$
PI_h = \left(\frac{|MSFE_h^{VAR}|}{|MSFE_h^{VARMA}|} - 1 \right) \times 100.
$$

The results show that in general the $VARMA$ model outperforms the two VAR alternatives. On average over all 15 forecast horizons, the determinants of mean squared forecast error matrices of both VAR models are 14% larger than that of the $VARMA$ model. The magnitude of the PIs are lower for the $tr\,(MSFE)$, but the overall picture is similar. Noticeable from the graphs in Figure (3) is the greater variation observed for the PI of the $VARMA$ against the $VAR\,(12)$ for both the forecast error measures. This is perhaps due to the

Table 6. Estimation results for the US macroeconomic data

PANEL A: Estimated $VARMA$ model: $\widehat{\mathbf{A}}\mathbf{y}_t = \widehat{\mathbf{c}} + \widehat{\mathbf{\Phi}}_1 \mathbf{y}_{t-1} + \varepsilon_t - \widehat{\mathbf{\Theta}}_1 \varepsilon_{t-1}$

$$\widehat{\mathbf{A}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \underset{(-5.744)}{-0.68} & 0 & 1 \end{bmatrix} \qquad \widehat{\mathbf{c}}' = \begin{bmatrix} \underset{(2.478)^a}{0.088} & \underset{(0.217)}{0.001} & \underset{(1.634)}{0.035} \end{bmatrix}$$

$$\widehat{\mathbf{\Phi}}_1 = \begin{bmatrix} \underset{(5.188)}{0.910} & \underset{(-1.451)}{-0.211} & \underset{(-1.877)}{-0.446} \\ \underset{(2.334)}{0.125} & \underset{(62.223)}{0.947} & \underset{(-1.061)}{-0.088} \\ 0 & 0 & 0 \end{bmatrix} \qquad \widehat{\mathbf{\Theta}}_1 = \begin{bmatrix} \underset{(-1.753)}{-0.229} & \underset{(0.375)}{0.085} & \underset{(2.146)}{0.503} \\ \underset{(0.027)}{0.001} & \underset{(-20.225)}{-0.867} & \underset{(0.695)}{0.047} \\ \underset{(0.317)}{0.03} & \underset{(0.855)}{0.107} & \underset{(-4.581)}{-0.289} \end{bmatrix}$$

$$\widehat{\mathbf{\Sigma}}_\varepsilon = \begin{bmatrix} 0.1606 & & \\ 0.0062 & 0.0265 & \\ -0.0154 & -0.0105 & 0.15 \end{bmatrix} \qquad \left| \widehat{\mathbf{\Sigma}}_\varepsilon \right| = 6.09 \times 10^{-4}$$

Res C.T.^{b}

m	j	
	0	1
0	0.26^c	0.26
1	0.29	0.82

PANEL B: Error covariance matrices for VARs selected by AIC and BIC

$$\widehat{\mathbf{\Sigma}}_\varepsilon^{VAR(12)} = \begin{bmatrix} 0.1151 & & \\ -0.001 & 0.0198 & \\ 0.0744 & -0.009 & 0.1704 \end{bmatrix} \qquad \left| \widehat{\mathbf{\Sigma}}_\varepsilon^{VAR(12)} \right| = 2.70 \times 10^{-4}$$

$$\widehat{\mathbf{\Sigma}}_\varepsilon^{VAR(2)} = \begin{bmatrix} 0.1606 & & \\ 0.0072 & 0.0307 & \\ 0.0942 & -0.005 & 0.2027 \end{bmatrix} \qquad \left| \widehat{\mathbf{\Sigma}}_\varepsilon^{VAR(2)} \right| = 7.04 \times 10^{-4}$$

^{a} $t-statistics$ in $parentheses$
^{b} $Criterion$ $Table$ for the $residuals$ $from$ the $estimated$ $model$
^{c} The $statistics$ are $normalised$ by the $corresponding$ 5% χ^2 $critical$ $values$

Table 7. Percentage improvement in —MSFE— and tr(MSFE) of VARMA(1,1) model over VAR models selected by AIC and BIC

	Panel A: PI in terms of $|MSFE|$						
	Forecast Horizon (h)						
	1	*2*	*3*	*6*	*9*	*12*	*15*
$VAR(12)$	18.814	16.704	14.434	13.699	19.327	11.456	-1.885
$VAR(2)$	11.225	16.184	18.209	17.743	11.988	10.859	9.482
	Av. of Forecast Horizon						
	1-3	*1-6*	*1-12*	*1-15*			
$VAR(12)$	16.651	17.318	17.866	15.359			
$VAR(2)$	15.206	18.733	16.070	14.862			

	Panel B: PI in terms of $tr(MSFE)$						
	Forecast Horizon (h)						
	1	*2*	*3*	*6*	*9*	*12*	*15*
$VAR(12)$	3.059	1.070	1.633	1.780	6.177	3.004	-0.533
$VAR(2)$	1.019	0.730	1.475	0.930	0.958	1.183	1.125
	Av. of Forecast Horizon						
	1-3	*1-6*	*1-12*	*1-15*			
$VAR(12)$	1.921	1.685	3.342	2.798			
$VAR(2)$	1.075	1.534	1.315	1.257			

fact that the $VAR(12)$ is likely to include many redundant parameters. Since long-horizon forecasts of each model converges to the unconditional mean implied by the model, it is expected that the forecast performances become closer in longer horizons. The PI values get closer to zero as horizon increases, and in fact the PIs for $h = 15$ relative to the $VAR(12)$ become slightly negative, indicating that the $VAR(12)$ outforecasts the $VARMA(1,1)$ for this forecast horizon.

In empirical business cycle analysis, the purpose of multivariate time series modelling is to forecast a measure of economic activity several periods into the future. In our example, the index of industrial production would be our measure of economic activity. Table 8 shows to what extent the forecasts of the industrial production produced by the $VARMA(1,1)$ model are better than those produced by the VAR alternatives. This shows that for the horizons up to two quarters ahead, that is for the horizons that are very important for business cycle analysts, the $VARMA$ model produces substantially better forecasts than the VAR alternatives.

5.   Conclusion

This paper proposes a $VARMA$ modelling procedure that extends the Tiao and Tsay (1989) scalar component methodology in the directions foreseen by the discussants of that paper. The proposed modelling procedure consists of three stages: (i) identification of the SCMs embedded in a $VARMA$ process; (ii) specification of a canonical scalar component $VARMA$ model by placing identification restrictions on the left hand side matrix that contains the contemporaneous relationships between the dependant variables; and (iii) full information likelihood estimation of the fully specified $VARMA$ model.
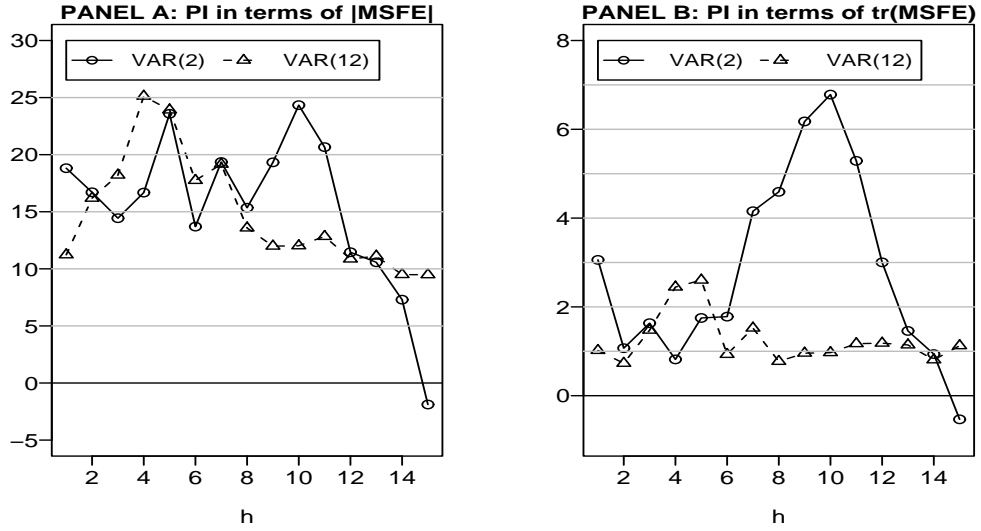
Fig. 3. Percentage improvement in terms of the forecast error measures, the *VARMA* model achieves versus the *VAR* models selcted by *AIC* and *BIC*

Table 8. Percentage improvement in MSFE of the VARMA(1,1) model over VAR models selected by AIC and BIC when forecasting Industrial Production

PI in terms of $MSFE$ forecasting Industrial Production							
	Forecast Horizon (h)						
	1	*2*	*3*	*6*	*9*	*12*	*15*
$VAR(12)$	11.354	18.879	18.313	19.073	6.803	4.260	-3.038
$VAR(2)$	3.095	6.671	5.278	-1.357	-1.061	0.268	0.760
	Av. of Forecast Horizon						
	1-3	*1-6*	*1-12*	*1-15*			
$VAR(12)$	16.182	17.841	12.811	10.249			
$VAR(2)$	5.015	2.889	1.112	0.959			

The complete *VARMA* modelling procedure is applied to two multivariate data sets. The first application highlights the advantage of the proposed extension to the T&T procedure in the estimation of parsimonious *VARMA* models. The second application shows the competitive out-of-sample forecasting performance of the identified *VARMA* model relative to alternative *VAR* models.

References

[1] Anderson, H. M. and Vahid, F. (1998) Testing multiple equation systems for common nonlinear components. *Journal of Econometrics*, **84**, 1-36.

[2] Athanasopoulos, G. (2005) Essays on Alternative Methods of Identification and Estimation of Vector Autoregressive Moving Average Models. *PhD Dissertation*. Monash University, Melbourne.

[3] Athanasopoulos, G. and Vahid, F. (2005) *VARMA* versus *VAR* for macroeconomic forecasting. Working Paper. Monash University, Melbourne.

[4] Clements, M. P. and Hendry, D. F. (1993) On the Limitations of Comparing Mean Squared Forecast Errors (with discussions). *Journal of Forecasting*, **12**, 617-637 .

[5] Durbin, J. (1963) Maximum likelihood estimation of the parameters of a system of simultaneous regression equations. Paper presented to the Copenhagen Meeting of the Econometric Society. reprinted in Econometric Theory, **4**, 159-170, 1988.

[6] Grubb, H. (1992) A Multivariate Time Series Analysis of Some Flour Price Data. *Applied Statistics*, **41**, 95-107.

[7] Hannan, E. J. and Rissanen, J. (1982) Recursive Estimation of Autoregressive-Moving Average Order. *Biometrika*, **69**, 81-94.

[8] Hansen, L. P. (1982) Large Sample Properties of Generalized Method of Moments Estimation. *Econometrica*, 50, 1029-1054.

[9] Hotelling, H. (1935) The most predictable criterion. *Journal of Educational Psychology*, **26**, 139-142.

[10] Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*, 1st edn. Berlin-Heidelberg: Springer-Verlag.

[11] Lütkepohl, H. and Poskitt, D. S. (1996) Specification of Echelon-Form *VARMA* Models. *American Statistical Association: Journal of Business and Economic Statistics*, **14**, 69-79.

[12] Stock, H. J. and Watson, M. W. (1999) A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In *Cointegration, Causality and Forecasting*, (eds. R. F. Engle and H. White), pp. 1-44. A Festschrift in Honour of Clive W.J. Granger, New York: Oxford University Press.

[13] Tiao, G. C. and Tsay, R. S. (1989) Model Specification in Multivariate Time Series (with discussions). *J. R. Statist. Soc. B*, **51**, 157-213 .

[14] Tsay, R. S. (1991) Two Canonical Forms for Vector Arma Processes. *Statistica Sinica*, **1**, 247-269.

[15] Vahid, F. and Engle, R. F. (1997) Codependent Cycles. *Journal of Econometrics*, **80**, 199-221.