Are VAR Models Good Enough for Macro-Econometric Modelling?

George Athanasopoulos*
Monash University
Clayton, Victoria 3800
Australia

Farshid Vahid
Australian National University
Canberra, ACT 2601
Australia

May 28, 2004

Abstract

VAR models are used in practice in preference to $VARMA$ models due to the difficult issues involved in the identification and estimation of $VARMA$ models. This paper examines if VAR models are good enough for forecasting macroeconomic variables. To answer this question, we extend the Tiao and Tsay identification procedure for $VARMA$ models and proposes a complete $VARMA$ modelling procedure. We then examine the properties of this identification procedure through simulation and used it to determine $VARMA$ models for many trivariate sets of macroeconomic variables and compare the out-of-sample forecasting performance of these models against VAR models fitted to the same data.

Keywords: Multivariate time series models, $VARMA$ models, identification, forecasting.

JEL Classification: C32.

1 Introduction

Vector autoregressive models have become the cornerstone of modern applied macro-econometric models. Interestingly, there is no applied macro-econometric research paper that even considers a vector autoregressive moving average, $VARMA$, model as an alternative[1]. The reason for this cannot be that economic

*Corresponding author. E-mail: George.Athanasopoulos@BusEco.monash.edu.au

[1] Although the study of $VARMA$ models has been on the statistical agenda for a long time, see Hannan (1969), Tunicliffe-Wilson (1973).

theory implies VAR dynamics for economic variables because economic theory rarely has any sharp implication about short-run dynamics of economic variables. In rare situations where the theoretical model includes a dynamic adjustment equation, one has to work hard to exclude moving average terms appearing in the implied dynamics of the variable of interest. Even if the theoretical model implies an autoregressive model for a variable, aggregation, say of monthly variables to quarterly frequency, induces moving average dynamics. Also, even if we believe that VAR is a good dynamic model for a set of variables, it implies that for any subset of these variables, a $VARMA$ model, rather than a VAR model, would be appropriate. Therefore, the apparent lack of interest in models with moving average errors can be the result of either that they are too hard, or that pure autoregressive models can do just as well.

Any invertible $VARMA$ process can be approximated by a finite order VAR. However, this does not imply that forecasts based on estimated long order VARs will be as good as those based on a $VARMA$ model because long order VARs have many parameters to estimate. We believe that the main reason for the lack of enthusiasm for using models with moving average errors is that they are too difficult. In particular in a multivariate setting, the identification and estimation of $VARMA$ models are quite involved, and this is in sharp contrast to the ease of identification and estimation of VAR models. This difficulty in model identification and estimation has so far prevented any comprehensive assessment of the question of whether VAR models are good enough for modelling $VARMA$ processes in finite samples. One of the objective of the present paper is to do just that.

Theoretically, there is no subtlety involved in estimation of an identified $VARMA(p,q)$ model. Based on the assumption of normality, the likelihood function conditional on the first p observations fixed and the q errors before time $p+1$ set to zero, is a well-defined function that can be calculated recursively. The exact likelihood function can also be calculated via the Kalman filter after the model is written in its state space form. However, it is not possible to fit a "general" $VARMA(p,q)$ to any set of observations and then try to reduce the system to a more parsimonious one by eliminating the insignificant parameters. The reason is that if the parameters of a $VARMA(p,q)$ satisfy certain conditions, then it will not be identified. The following simple example illustrates this point.

Consider the following bivariate $VARMA(1,1)$ process

$$\begin{aligned} y_{1,t} &= \phi_{11}y_{1,t-1} + \phi_{12}y_{2,t-1} + \theta_{11}\eta_{1,t-1} + \theta_{12}\eta_{2,t-1} + \eta_{1,t} \\ y_{2,t} &= \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + \theta_{21}\eta_{1,t-1} + \theta_{22}\eta_{2,t-1} + \eta_{2,t} \end{aligned} \tag{1}$$

The first equation (and therefore the model) is not identified if all parameters of the second equation are zero, i.e., if $\phi_{21} = \phi_{22} = \theta_{21} = \theta_{22} = 0$. This is because in this case $y_{2,t-1} = \eta_{2,t-1}$ and therefore ϕ_{12} and θ_{12} cannot be identified separately.

This problem appears in other areas of econometrics as well. In simultaneous equations, the identification of a structural equation often requires that the parameters in other equations of the system not lie in specific restricted subspaces. For example, exclusion of an exogenous variable from a structural equation only helps its identification if that exogenous variable appears with a non-zero coefficient elsewhere in the system. Another example is in nonlinear models were identification of some of the parameters depends on the true model not being linear. Clearly then, it is important in these cases that we understand the identification issues, and if possible, we pre-test the required conditions for identification and determine a uniquely identified structure that is likely to be estimable from the particular data set under study.

There are several methods for identification of $VARMA$ models. One method is through the determination of the "Kronecker indices" and the identification of the "echelon form" of the $VARMA$ model (see Lütkepohl, 1991, or Reinsel, 1993, for definition of the echelon form and the Kronecker index). In this method, estimates of lag innovations are derived from a first stage long VAR as suggested by Hannan and Rissanen (1982), and the Kronecker indices are determined in a second stage using a model selection criteria (see, for example, Hannan and Diestler, 1988, and Lütkepohl and Poskitt, 1996). Another method, suggested by Tiao and Tsay (1989), is to identify the "scalar components" imbedded in a $VARMA$ model through a series of tests based on canonical correlations between judiciously chosen sets of variables, and then estimate these scalar components. Our present paper extends this methodology[2], and uses it to answer the question that whether VAR models are good enough for forecasting macroeconomic variables.

The methodology of Tiao and Tsay (1989) is based on searching for linear combinations of variables that have simple dynamic structures. Their method,

[2] There are several other classes of multivariate dynamic models in the literature. These include the dynamic factor model of Peña and Box (1987) and the multivariate version of structural time series model of Harvey (1991).

however, provides consistent but not efficient estimates of these linear combinations. Then, the remaining parameters of the identified structure are estimated based on these possibly inefficient estimates of such linear combinations. Many eminent time series analysts, such as Chatfield, Hannan, Reinsel and Tunnicliffe-Wilson among others, commented on the seminal paper of Tiao and Tsay (1989), and their comments were published in the same issue of the Journal of Royal Statistical Society. In their comments they point out that more attention should be paid to the estimation of these linear combinations, since the ultimate purpose is to invert from these linear combination back to the original variables under study. This is the direction that we extend Tiao and Tsay's methodology.

Using artificially generated data from simple data generating processes, we are able to show that our extension of Tiao and Tsay works well. However, it would be difficult, if not impossible to provide a convincing answer for the question of whether VAR models are sufficient for macroeconometric modelling by means of Monte Carlo simulations. There are many difficulties involved in designing Monte Carlo experiments in multivariate time series, in particular in such a way that their results would be relevant for macroeconometric analysis (see, for example, Vahid and Issler, 2002). Therefore, we use real macroeconomic data for our investigation and compare the out of sample performance of fitted VAR and $VARMA$ models. The advantage of this method is that our results will be of direct relevance for macro-economic forecasting. The drawback is that we are not able to comment on which model produces a better impulse response function or a better decomposition of forecast error variance because these objects are interest are not observable.

The structure of the paper is as follows.

Section (2) provides a brief description of the scalar component methodology of Tiao and Tsay (1989). Section (3) explains our extension of Tiao and Tsay. Section (5) reports the accuracy of the identification procedure in a small scale simulation exercise. As an applied demonstration of the proposed identification procedure for $VARMA$ models, section 5 reports estimated $VARMA$ models for two widely analysed data sets. Section 6 reports the results of an extensive empirical application for out of sample forecasts for competing $VARMA$ and VAR models for many three variable systems of macroeconomic aggregates. The data used was extracted from the data set employed by Stock and Watson (1999) for comparing linear and nonlinear univariate models for forecasting macroeconomic time series. Finally, Section 7 concludes.

2 The Scalar Component Methodology

The aim of exploring scalar component ($SCMs$) is to examine whether there are any simplifying underlying structures for a $VARMA\,(p,q)$ process.

Definition 1 *For a given k-dimensional $VARMA(p,q)$ process*

$$\mathbf{y}_t = \mathbf{\Phi}_1\mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\eta}_t - \mathbf{\Theta}_1\boldsymbol{\eta}_{t-1} - \ldots - \mathbf{\Theta}_q\boldsymbol{\eta}_{t-q}, \tag{2}$$

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

$$\begin{aligned}
\boldsymbol{\alpha}'\mathbf{\Phi}_{p_1} &\neq \mathbf{0}^T \text{ where } 0 \leq p_1 \leq p, &(3)\\
\boldsymbol{\alpha}'\mathbf{\Phi}_{l} &= \mathbf{0}^T \text{ for } l = p_1+1, ..., p, &(4)\\
\boldsymbol{\alpha}'\mathbf{\Theta}_{q1} &\neq \mathbf{0}^T \text{ where } 0 \leq q_1 \leq q, &(5)\\
\boldsymbol{\alpha}'\mathbf{\Theta}_{q1} &= \mathbf{0}^T \text{ for } l = q_1+1, ..., q. &(6)
\end{aligned}$$

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).

Tiao and Tsay's methodology starts from the most parsimonious possibility (i.e., $SCM\,(0,0)$ which is a system white noise) and discovers k linearly independent vectors $(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_k)$ which would rotate the $VARMA(p,q)$ system into a simpler dynamic system with significantly lower number of parameters. That is, if we define matrix $\mathbf{A} = (\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_k)'$, then transformation of a $VARMA\,(p,q)$ system by this matrix creates another $VARMA\,(p,q)$ system in terms of transformed variables $\mathbf{z}_t = \mathbf{A}\mathbf{y}_t$ and transformed innovations $\mathbf{u}_t = \mathbf{A}\boldsymbol{\eta}_t$, but there are many zero restrictions on the AR and the MA parameter matrices of the transformed model.

The determination of embedded scalar component models are achieved through a series of canonical correlation tests. An $SCM\,(0,0)$ 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 $\mathbf{Y}_{h,t} \equiv \left(\mathbf{y}'_{t-1},\ldots\mathbf{y}'_{t-h}\right)'$ 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

$$C(s) = -(n-h)\sum_{i=1}^{s} \ln\left(1-\widehat{\lambda}_i\right) \overset{a}{\sim} \chi^2_{s\times\{(h-1)k+s\}}, \tag{7}$$

and the canonical variates 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 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 $\mathbf{Y}_{m,t} \equiv \left(\mathbf{y}_t', \ldots \mathbf{y}_{t-m}'\right)'$ and $\mathbf{Y}_{m+h,t} \equiv \left(\mathbf{y}_{t-1}', \ldots \mathbf{y}_{t-m-h}'\right)'$. $SCM(m,j)$ however are linear combinations of $\mathbf{Y}_{m,t}$ that cannot be linearly predicted from the history before $t-j$. Therefore, 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 $\mathbf{Y}_{h,t-j} \equiv \left(\mathbf{y}_{t-j-1}', \ldots \mathbf{y}_{t-j-h}'\right)'$. This imposes a structure on the GMM weighting matrix. The test then is a test of overidentifying restrictions in this system. Alternatively, Tiao and Tsay (1989) suggest the statistic

$$C(s) = -(n-h-j)\sum_{i=1}^{s} \ln\left\{1-\frac{\widehat{\lambda}_i}{d_i}\right\} \overset{a}{\sim} \chi^2_{s\times\{(h-m)k+s\}} \tag{8}$$

based on the squared canonical correlations between $\mathbf{Y}_{m,t}$ and $\mathbf{Y}_{h,t-j}$. 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

$$d_i = 1 + 2\sum_{v=1}^{j} \widehat{\rho}_v\left(\widehat{\mathbf{r}}_i'\mathbf{Y}_{m,t}\right)\widehat{\rho}_v\left(\widehat{\mathbf{g}}_i'\mathbf{Y}_{h,t}\right) \tag{9}$$

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

Since an $SCM(m,j)$ nests all scalar components of order $(\leq m, \leq j)$, low order $SCMs$ also show up when testing for higher orders. Tiao and Tsay (1989) provide a complete set of rules that determines the order of all parsimonious $SCMs$ embodied in the system. They also deduce a consistent estimate of the transformation matrix $\mathbf{A}$ from the estimated canonical coefficients, and form the transformed variables $z_t = \boldsymbol{\alpha}'\mathbf{y}_t$. However this transformed system still may not be uniquely identified for the following reason.

Criterion 2 *General Elimination Rule: Suppose we found* $z_{1,t} = SCM(p_1, q_1)$ *and* $z_{2,t} = SCM(p_2, q_2)$ *where* $p_2 > p_1$ *and* $q_2 > q_1$*. This implies that one of the variables of the right hand side of the dynamic equation for* $z_{2,t}$*, for example* $z_{1,t-1}$ *can be written in terms of other variables on the right hand side of* $z_{2,t}$*. Hence, the parameters of the right hand side of* $z_{2,t}$ *equation are not identified unless we set some of them to zero. In fact, for each lag* $i = 1, ..., \min(p_2 - p_1, q_2 - q_1)$*, we need to set either the coefficient of* $z_{1,t-i}$ *or that of* $u_{1,t-i}$ *to zero to obtain a uniquely identified system.*

3 An extension of Tiao and Tsay

A major concern of the participants in the discussion of the Tiao and Tsay (1989) paper (see the discussion that follows the paper), was the general treatment of the transformation matrix $\mathbf{A}$. Tiao and Tsay's general representation of the identified $VARMA(p, q)$ model is

$$\mathbf{z}_t = \mathbf{\Phi}_1^* \mathbf{z}_{t-1} + \ldots + \mathbf{\Phi}_p^* \mathbf{z}_{t-p} + \mathbf{u}_t - \mathbf{\Theta}_1^* \mathbf{u}_{t-1} - \ldots - \mathbf{\Theta}_q^* \mathbf{u}_{t-q}, \tag{10}$$

where $\mathbf{z}_t = \mathbf{A}\mathbf{y}_t$, $\mathbf{\Phi}_i^* = \mathbf{A}\mathbf{\Phi}_i\mathbf{A}^{-1}$, $\mathbf{u}_t = \mathbf{A}\boldsymbol{\eta}_t$ and $\mathbf{\Theta}_i^* = \mathbf{A}\mathbf{\Theta}_i\mathbf{A}^{-1}$.

The comments by Professors Reinsel, Chatfield, Tunnicliffe-Wilson and Hannan amongst others can be summarised in the following concerns about the Tiao and Tsay modelling procedure:

1. 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 $\mathbf{y}_t$.

2. 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, thus making the reduction in degrees of freedom for the estimation smaller than what is claimed by Tiao and Tsay.

3. The Tiao and Tsay 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 SCM's with $q > 0$.

4. Finally, the Tiao and Tsay procedure does not produce standard errors for the parameter estimates in $\mathbf{A}$.

In liu of these concerns we propose the following extension of the Tiao and Tsay modelling procedure that develops identification conditions for the parameters of the matrix $\mathbf{A}$ in order to be able to estimate them efficiently. First, we note that the space spanned by $\mathbf{z}_{t-1}$ to $\mathbf{z}_{t-p}$ is the same as the space spanned by $\mathbf{y}_{t-1}$ to $\mathbf{y}_{t-p}$. Hence the right hand side of (10) can be written in terms of $\mathbf{y}_{t-1}$ to $\mathbf{y}_{t-p}$ without affecting the restrictions implied by the scalar component orders. Hence $\mathbf{A}$ does not play a substantive role on the right hand side of (10).

Definition 3 $\mathbf{A}$ *is identified if and only if the only matrix* $\mathbf{H}$ *such that* $\mathbf{HAy}_t$ *has the same scalar component structure as* $\mathbf{Ay}_t$ *is the identity matrix.*

Important considerations for unique identification of $\mathbf{A}$ are the following:

1. Each row of $\mathbf{A}$ 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 1. However, as always in such situations, there is a danger of wrongly choosing a parameter whose true value is zero for normalisation. We ignore this possibility for the time being, but we will return to it later.

2. Any linear combination of an $SCM\left(p_1, q_1\right)$ and an $SCM\left(p_2, q_2\right)$ is an $SCM\left(\max\left\{p_1, p_2\right\}, \max\left\{q_1, q_2\right\}\right)$. Therefore if there is a single SCM whose autoregressive order is the smallest of all other $SCMs$ in the system, the corresponding row of $\mathbf{A}$ is uniquely identified. This is because any combination of other $SCMs$ with this one produces an 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 $\mathbf{A}$ corresponding to it is uniquely identified.

3. In all cases where there are two embedded scalar components with nested orders, i.e., $p_1 \leq p_2$ and $q_1 \leq q_2$ arbitrary multiples of $SCM\left(p_1, q_1\right)$ can be added to the $SCM\left(p_2, q_2\right)$ without changing the structure. This means that the row of $\mathbf{A}$ corresponding to the $SCM\left(p_2, q_2\right)$ is not identified in this case. To achieve identification, if the parameter in the i^{th} column of the row of $\mathbf{A}$ corresponding to the $SCM\left(p_1, q_1\right)$ is normalised to 1, the parameter in the same position in the row of $\mathbf{A}$ corresponding to $SCM\left(p_2, q_2\right)$ should be restricted to zero.

The following example highlights the above points.

Example 4 *Consider a trivariate stationary process* $\mathbf{y}_t$ *which has been identified by the Tiao and Tsay identification process to have three scalar components:* $z_{1,t} \sim SCM(1,1)$, $z_{2,t} \sim SCM(1,0)$ *and* $z_{3,t} \sim SCM(0,0)$. *Hence the scalar components specification of the* $VARMA(1,1)$ *is*

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \mathbf{\Theta}_1\boldsymbol{\varepsilon}_{t-1},$$

thus,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} \tag{11}$$

The parameter $\theta_{13}^{(1)}$ *is set to zero because of the "general rule of elimination" explained in the criterion* (2) *above.*

Ignoring the $\mathbf{A}$ *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 to Tiao and Tsay's paper have pointed out. The parameters in* $\mathbf{A}$ *have to be estimated. However not all nine parameters in* $\mathbf{A}$ *are free parameters. Firstly, one parameter per row, say diagonal elements, can be normalised to one.*

$$\begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} \tag{12}$$

The third equation is now uniquely identified, because no combination of the other two equations can be added to it and keep its $SCM\,(0,0)$ *structure. However since an* $SCM\,(0,0)$ *in nested in an* $SCM\,(1,0)$ *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, which leads to*

$$\begin{bmatrix} 1 & a_{12} & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{13}$$

The same rule again implies that we can set a_{12} *to zero without changing the*

structure,

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{14}$$

Hence in fact the real reduction in the number of parameters is $10 - 3 = 7$.

Before moving to estimation, we must discuss the possibility of normalising a zero parameter to 1. To safeguard against this possibility, our procedure adds tests of predictability using subsets of variables. Starting from the SCM with the smallest order, we leave one variable out, say the k^{th} one and test if an SCM of the minimal order can be found using the $k-1$ variables alone. If test is rejected, then we normalise the coefficient of the k^{th} variable to 1 and set the corresponding coefficients in all other SCMs that nest this one to zero. If test points out that we can form the SCM using the first $k-1$ variables only, the coefficient of the k^{th} variable in this SCM is zero, and it should not be normalised to 1. It is worth noting that this extra zero restriction adds to the restrictions discovered before. We then leave 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 with suitably chosen weighting matrices given by the structure of the system.

4 A Complete $VARMA(p,q)$ Modelling Process

Having presented in detail the proposed extension of the Tiao and Tsay identification process, we next propose a complete modelling process for $VARMA(p,q)$ models. This process consists of three stages. stage I identifies the scalar component structure of the $VARMA$ model as in Tiao and Tsay (1989). Stage II we applies the proposed extension in dealing with the transformation matrix $\mathbf{A}$ and stage III estimate the identified model.

4.1 Stage I: Identification of the Scalar Components

Stage I consists of two steps.

Step 1: Determining an overall tentative order In step I, an overall tentative order for the $VARMA(p,q)$ process is specified via the canonical correlations based statistic, $C(s)$, of equation (8).

Choice of h In practice h controls the effective sample size in the estimation of the canonical correlations. Thus the lower the h the larger the effective sample size. However we have to keep in mind that h has to be at least as large as[3] m. Therefore as in Tiao and Tsay (1989) we also recommend $h = m$ for this first step of the first stage of the identification procedure.

The test statistic $C(s)$ is arranged in a two-way table, referred to by Tiao and Tsay as the *criterion table*, for $m, j = 0, 1, \ldots$. From the criterion table we can read the overall tentative order of the $VARMA(p, q)$ process $\mathbf{y}_t$, by searching for the lower right rectangular pattern of insignificant $C(s)$ [4].

Step 2: Identifying individual orders of SCMs Given the overall tentative order of the model we now identify individual $z_{it} \sim SCM(p_i, q_i)$ orders, again via the $C(s)$ statistic. The hierarchical process starts from testing for the simplest possible white noise scalar components, i.e., $z_{1t} \sim SCM(0, 0)$ and then the test is applied sequentially for $i = 1, .., k$ for each $m, j = 0, 1, \ldots$.

Choice of h For this section Tiao and Tsay suggest $h = m + (q - j)$ at the position (m, j). Their intuition is, that given the overall tentative order of the model specified in step 1 (say $VARMA(p, q)$) each series is serially uncorrelated after lag q.

These results are again summarised in a two way tabulation in what Tiao and Tsay refer to as the *root table*. The root table presents the number of zero eigenvalues found for each order by applying the above test[5]

A note on exchangeable models In multivariate time series analysis, there are pathological cases where a process can be represented either as a finite $VAR(p)$ or a finite $VMA(q)$ process.

Example 5 *For example the bivariate process*

$$\mathbf{y}_t = \begin{bmatrix} a & a \\ -a & -a \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{15}$$

[3] If $h < m$ then the autocovariance matrix we are considering is by design of reduced rank. See Tiao and Tsay (1989) for further details.

[4] For examples of the criterion table please refer to section (6).

[5] For examples of the root table see the examples of section (6). For further examples see Tiao and Tsay (1989).

can be equivalently represented as

$$\mathbf{y}_t = \begin{bmatrix} a & a \\ -a & -a \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{16}$$

As we have already identified in this case the first presentation, equation (15)*, implies an* $SCM\,(0,0)$ *and a* $SCM\,(1,0)$*. The second presentation, equation* (16)*, implies an* $SCM\,(0,0)$ *and a* $SCM\,(0,1)$*. These exchangeable models produce a recognisable pattern in the root table. Having recognised this, we choose one of the two possibilities (we prefer the autoregressive over the moving average components as it simplifies the estimation procedure) and proceed. This does not pose any potential loss*[6]*.*

Summary of Stage I of the identification process Stage I of the proposed modelling process consists of two steps. In step one we identify the overall maximal order (p, q) of the $VARMA$ model. Then conditional on this, in step 2 we identify the individual orders of the embedded SCMs.

4.2 Stage II: Dealing with the transformation matrix A

In stage II of the identification process we apply the proposed extension of section (3) to identify the structure of the transformation matrix $\mathbf{A}$.

4.3 Stage III: Likelihood Estimation

In stage III the parameters of the uniquely identified structures are then estimated via maximum likelihood. The canonical correlation 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. The maximum likelihood procedure provides estimates and estimated standard errors for all parameters, including the free parameters of $\mathbf{A}$. All usual considerations that ease the estimation of structural forms, are also valid here and should definitely be exploited in estimation

[6] In the extensive empirical application of section (7) in many models we had exchangeable models. Both alternatives were applied and evaluated in a forecast error sense and there did not seem to be much difference between the alternatives.

5 Simulation

The aim of the following simulations is to explore the performance of both stages of the identification process. Also the simulations help the reader develop a better understanding of the proposed representation of the $VARMA(p,q)$ models, and other ideas on these models as developed in this study.

5.1 Performance of Stage I of the Identification procedure

To evaluate the performance of the first stage of the identification process we have generated data from various $VARMA(p,q)$ models, for varying sample sizes and performed the first stage of the identification process on the generated data series. The models have been presented and the identification results have been tabulated in what follows. To assist the reader with reading and analysing the simulation results, we have created table (1), were we briefly describe the role of each column of the tables the simulation results are presented in.

Table 1: Description of tables exploring stage I of the identification procedure

Column	**Label**	**Description**
1	N	Number of observations generated.
2	**M.O.**	Percentage of times the procedure correctly identifies the *maximum* (p,q) *order* of the model.
3-6	$SCM(p_i,q_i)$	Each column shows possible SCMs the identification procedure could identify, as these are nested within the maximum (p,q) order.
	0^* 1 2 3	Each of these sub-columns shows the number of possible SCMs of order (p_i,q_i) the procedure could possibly identify (recall $k=3$, i.e., maximum three SCMs can be identified for each order).
	*	Signifies the correct number of scalar components of order (p_i,q_i) the procedure should identify.
7	**E.O.**	Percentage of times the procedure correctly identifies the *exact order*, i.e., identifies correctly every individual $SCM(p_i,q_i)$ for $i=1,\ldots,k$, of the $VARMA(p,q)$ model.

It should be mentioned that due to the complex and long process of reading off the criterion and root tables only 50 iterations were performed for each of the models. The two most important columns of the tables are the maximum order and the exact order columns although as we have already mentioned the exact order found here is not final as this is reconfirmed in stage II. Having given a general description for each of the tables that follow we next present the simulation results for each of the seven models we have simulated from.

5.1.1 $VAR(1)$ and $VMA(1)$

The simulation process starts from two simple, restricted[7] $VAMRA(p,q)$, models. First the $VAR(1)$ process,

$$\mathbf{y}_t = \begin{bmatrix} 0.5 & -0.6 & 0.7 \\ 0.6 & 0.7 & -0.4 \\ 0.3 & 0.6 & 0.4 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{17}$$

From table (2) one can see that the identification procedure does very well in identifying both the maximum and the exact order of the model even with very small samples of $N = 50$ observations.

Table 2: VARMA(1,0)

N	**M.O.**	$SCM(0,0)$				$SCM(0,1)$				$SCM(1,0)$				$SCM(1,1)$				**E.O.**
		$\mathbf{0}^*$	1	2	3	$\mathbf{0}^*$	1	2	3	0	1	2	$\mathbf{3}^*$	$\mathbf{0}^*$	1	2	3	
50	**100**	**100**	-	-	-	**100**	-	-	-	2	-	6	**92**	**92**	6	-	2	**92**
200	**100**	**100**	-	-	-	**100**	-	-	-	2	-	2	**96**	**96**	2	-	2	**96**

The second simple process we look at is the $VMA(1)$ or $VARMA(0,1)$.

$$\mathbf{y}_t = \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & -0.6 & 0.7 \\ 0.6 & 0.7 & -0.4 \\ 0.3 & 0.6 & 0.4 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{18}$$

Table (3) indicates that the identification procedure does well in identifying the maximum order of the process being a $(0,1)$ and improves with identifying the exact order as the sample size increases. Here we see that the process has difficulty in identifying the exact order as it should show no $SCM(1,0)$ but it tends to find one.

[7] Restricted in the sence that one of either p or q is set to zero.

Remark 6 *This is a regular observation from our simulations (the few recorded and the many non-recorded). The identification process has some difficulty distinguishing between an $AR(1)$ and an $MA(1)$ which also commonly observed in univariate $ARMA(p,q)$ modelling, as this depends on the strength of second order correlations and beyond. This aspect of the identification procedure improves as the sample size increases.*

Table 3: VARMA(0,1)

N	M.O.	$SCM(0,0)$				$SCM(0,1)$				$SCM(1,0)$				$SCM(1,1)$				**E.O.**
		0*	1	2	3	0	1	2	**3***	**0***	1	2	3	**0***	1	2	3	
50	88	**48**	46	6	-	-	14	42	**36**	**24**	38	32	6	**100**	-	-	-	**2**
100	96	**84**	16	-	-	-	-	20	**80**	**36**	46	18	-	**100**	-	-	–	**36**
150	96	**92**	8	-	-	-	-	12	**88**	**40**	54	6	-	**100**	-	-	–	**40**
200	94	**100**	-	-	-	-	-	8	**92**	**50**	50	-	-	**100**	-	-	–	**50**
400	98	**100**	-	-	-	-	-	2	**98**	**88**	10	-	2	**100**	-	-	-	**88**

5.1.2 $VARMA(p,q)$

The third model we simulate from is a $VARMA(1,1)$ which comprises of three scalar components of orders, $(0,0),(1,0)$ and $(1,1)$, as shown in equation (19). From the representation one can see in application the proposed extension of section (3), as we have imposed the necessary restrictions on elements of $\mathbf{A}$. Also since there are SMCs where both $p_1 > p_2$ and $q_1 > q_2$ (namely $SCM(1,1)$ and $SCM(0,0)$) one can also see in application the generalised rule of elimination of criterion (2) as we have set $\theta_{13}^{(1)} = 0$[8].

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 1 & 0 \\ 0.4 & 0.7 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}, \quad (19)$$

The results of the simulations of this process (presented in table (4)) show that the identification process does well in identifying both the maximum order and the exact order even for sample sizes of 150 observations.

Next we have again simulated from a similar $VARMA(1,1)$ as in (19), however the second scalar component is a moving average component rather than an

[8] These rules can also be observed in all the models that follow. However we will not comment any further on these, as we purely revert our attention to the simulation results.

Table 4: VARMA(1,1) - SCM(1,1) SCM(1,0) SCM(0,0)

N	**M.O.**	$SCM\,(0,0)$				$SCM\,(0,1)$				$SCM\,(1,0)$				$SCM\,(1,1)$				**E.O.**
		0	**1***	2	3	**0***	1	2	3	0	**1***	2	3	0	**1***	2	3	
50	**36**	-	**98**	2	-	**12**	52	36	-	6	**56**	38	-	96	**4**	-	-	**4**
100	**88**	2	**98**	-	-	**58**	38	4	-	2	**90**	8	-	46	**52**	2	-	**52**
150	**94**	2	**98**	-	-	**84**	16	-	-	4	**92**	4	-	20	**78**	2	-	**78**
200	**96**	2	**98**	-	-	**98**	2	-	-	-	**98**	2	-	6	**94**	-	-	**94**
400	**100**	6	**94**	-	-	**94**	6	-	-	6	**88**	6	-	8	**86**	6	-	**86**

autoregressive component.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 1 & 0 \\ 0.4 & 0.7 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}, \tag{20}$$

As one can see from table (5) the identification procedure does well in all sample sizes above 150 in identifying the maximal order however it does find it a little harder in pinning down the exact order.

Table 5: VARMA(1,1) - SCM(1,1) SCM(0,1) SCM(0,0)

N	**M.O.**	$SCM\,(0,0)$				$SCM\,(0,1)$				$SCM\,(1,0)$				$SCM\,(1,1)$				**E.O.**
		0	**1***	2	3	0	**1***	2	3	**0***	1	2	3	0	**1***	2	3	
50	**28**	-	**88**	12	-	4	**60**	36	-	**12**	56	32	-	98	**2**	-	-	**2**
100	**68**	2	**96**	2	-	2	**96**	2	-	**16**	70	14	-	88	**12**	-	-	**12**
150	**76**	2	**98**	-	-	8	**88**	4	-	**8**	90	-	-	86	**14**	-	-	**8**
200	**92**	2	**98**	-	-	2	**96**	2	-	**22**	78	-	-	76	**24**	-	-	**22**
400	**96**	2	**98**	-	-	6	**92**	2	-	**62**	38	-	-	42	**52**	6	-	**52**
700	**98**	8	**92**	-	-	4	**90**	6	-	**78**	22	-	-	24	**72**	4	-	**72**

The last two models we simulate from are again $VARMA\,(1,1)$ models however we allow for the nature of them to be a little more complex than before as we increase the orders of the individual p_i and q_i's. The results and the models are shown below.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 1 & 0 \\ 0 & -0.6 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 0.7 & -0.6 & 0.4 \\ 0.6 & -0.5 & -0.4 \\ 0.3 & -0.6 & 0.4 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.7 & 0.4 & -0.6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} \tag{21}$$

Table 6: VARMA(1,1) - SCM(1,1) SCM(1,0) SCM(1,0)

N	**M.O.**	$SCM\,(0,0)$				$SCM\,(0,1)$				$SCM\,(1,0)$				$SCM\,(1,1)$				**E.O.**
		$\mathbf{0}^*$	1	2	3	$\mathbf{0}^*$	1	2	3	0	1	$\mathbf{2}^*$	3	0	$\mathbf{1}^*$	2	3	
50	**52**	**90**	10	-	-	**6**	30	50	14	-	14	**60**	26	98	**2**	-	-	**2**
100	**96**	**100**	-	-	-	**10**	82	8	-	2	-	**96**	2	88	**10**	2	-	**10**
150	**92**	**100**	-	-	-	**18**	80	2	-	-	4	**96**	-	82	**18**	-	-	**18**
200	**98**	**100**	-	-	-	**20**	80	-	-	-	8	**92**	-	72	**28**	-	-	**20**
400	**94**	**100**	-	-	-	**64**	36	-	-	-	-	**100**	-	38	**62**	-	-	**62**
700	**96**	**100**	-	-	-	**78**	22	-	-	-	4	**94**	-	24	**72**	2	2	**72**

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5 & -0.7 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 0.7 & -0.5 & 0.7 \\ 0.6 & 0.3 & 0.6 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & -0.6 & 0 \\ 0.6 & 0.7 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{22}$$

Table 7: VARMA(1,1) - SCM(1,1) SCM(1,1) SCM(0,0)

N	**M.O.**	$SCM\,(0,0)$				$SCM\,(0,1)$				$SCM\,(1,0)$				$SCM\,(1,1)$				**E.O.**
		0	$\mathbf{1}^*$	2	3	$\mathbf{0}^*$	1	2	3	$\mathbf{0}^*$	1	2	3	0	1	$\mathbf{2}^*$	3	
50	**40**	8	**92**	-	-	**8**	42	46	-	**28**	58	14	-	80	18	**2**	-	**2**
100	**88**	6	**94**	-	-	**24**	70	6	-	**66**	32	2	-	28	62	**10**	-	**10**
150	**94**	2	**98**	-	-	**44**	54	2	-	**88**	12	-	-	2	68	**30**	-	**44**
200	**92**	6	**94**	-	-	**58**	36	6	-	**92**	8	-	-	8	44	**48**	-	**48**
400	**94**	6	**94**	-	-	**76**	22	2	-	**94**	6	-	-	4	24	**72**	-	**72**
700	**98**	6	**94**	-	-	**92**	8	-	-	**94**	6	-	-	-	8	**92**	-	**92**

5.2 Performance of Stage II of the Identification procedure.

The following three models are examples where the second stage of the identification process contributes to the identification procedure as a whole. These models have been used in Stage I of the simulation exercise, but now have extra restrictions on the transformation matrix $\mathbf{A}$ to reflect the need of this second stage of the identification process. Lets take for example model $(M1)$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 1 & 0 \\ 0.4 & 0 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \text{ (M1)}$$

Stage I should identify an overall $VARMA\,(1,1)$ with three scalar components

of orders $(1,1)$, $(1,0)$ and $(0,0)$. Applying the restriction process of section (3) we find that there should be a white noise scalar component which is a result of a linear combination of all three series $y_{1,t}$, $y_{2,t}$ and $y_{3,t}$. Thus from there we would proceed by normalising one of the coefficients of any one of the series to one. However observing the data generating mechanism of model $(M1)$, the coefficient of $y_{2,t}$ is zero and if we normalised on this we would set it "incorrectly" to one, thus setting a zero parameter to something non-zero. Table (8), in column two under the heading M1, reflects the results of the test procedure of actually identifying that there is a linear combination of only the two components, $y_{1,t}$ and $y_{3,t}$, that loads as a $SCM\,(0,0)$.

Next model $(M2)$ is a continuation of model three and additional to the initial restriction of two components being a $SCM\,(0,0)$ we have that $y_{2,t}$ on its own is a $SCM\,(1,0)$. Again the test performance, for both tests, is presented in table (8) under the heading $(M2)$.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.4 & 0 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} \quad \text{(M2)}$$

Finally the next model is a $VARMA\,(1,1)$ with three scalar components of orders $(1,1)$, $(1,0)$ and $(1,0)$. The test of the second stage of the identification process performs considerably well (as shown in table (8)) in identifying that $y_{3,t}$ on its own as a $SCM\,(1,0)$ and does not need to be linearly combined with another series as stage I of the identification process would have indicated.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} 0.7 & -0.6 & 0.4 \\ 0.6 & -0.5 & -0.4 \\ 0.3 & -0.6 & 0.4 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.7 & 0.4 & -0.6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} \quad \text{(M3)}$$

5.3 General Comments

From the simulations performed (for both stages of the identification procedure) one can make a couple of general observations. Firstly the identification procedure employed here is based on a sequence of hypothesis tests, and hence the overall size of the procedure is larger than the level of significance of each test. Also, as long as the level of significance is kept constant for all sample sizes the procedure

Table 8: Identification Simulation - Stage II

N	**M1**	**M2**		**M3**
	$y_{1,t} + a_{13}y_{3,t}$ $\sim$ $SCM\,(0,0)$	$y_{1,t} + a_{13}y_{3,t}$ $\sim$ $SCM\,(0,0)$	$y_{2,t}$ $\sim$ $SCM\,(1,0)$	$y_{3,t}$ $\sim$ $SCM\,(1,0)$
50	86	80	64	78
100	86	88	66	72
150	86	90	78	78
200	92	88	84	84
400	84	92	80	72

is not consistent, in the sense that there is always a fixed probability of type I error even if the sample size goes to infinity. Secondly for the order of models investigated here, sample sizes of fewer than $N = 150$ observations seem to be inappropriately small to be modelled using this procedure.

6 Examples

The following two examples fully illustrate in an empirical sense the complete methodology outlined in section (4).

6.1 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 series with $N = 100$ observations each on monthly flour price indices for the cities of, Buffalo, Minneapolis, and Kansas City. As in Tiao and Tsay (1989) we are looking at the logarithms of the series plotted in figure (1).

In the first stage of the identification process we read of the criterion table (table (9)) and find possible maximum orders of $VARMA\,(1,1)$ or $VAR\,(2)$.Since the primary goal of this paper is $VARMA\,(p,q)$ modelling we proceed with the overall order being $VARMA\,(1,1)$. The individual orders of the SMCs identified from the first stage of the identification process when persisting with an overall order of $VARMA\,(1,1)$ can be read of the root table, table (10), in which we find that there are two scalar components of order $(1,0)$ and one of order $(1,1)$.

Figure 1: US Flour Prices

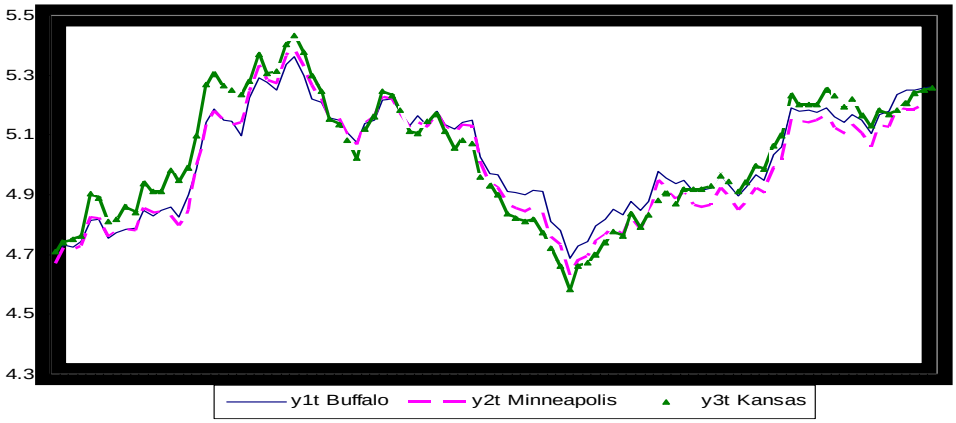

Table 9: Criterion Table

	j				
m	0	1	2	3	4
0	34.17	5.8	3.0	2.11	1.68
1	2.38	**0.44**	0.49	0.22	0.34
2	**0.25**	0.58	0.60	0.49	0.46
3	0.37	0.46	0.67	0.53	0.58
4	0.73	0.62	0.57	0.70	0.77

The statistics are normalised by the corresponding 5% χ^2 critical values

One possible representation of the $VARMA(p,q)$ model identified having applied the rules of section (3) would be

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ 0 & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ \phi_{31}^{(1)} & \phi_{32}^{(1)} & \phi_{33}^{(1)} \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & \theta_{13}^{(1)} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} \tag{23}$$

Testing for the appropriate normalisations we find that normalising the diagonal parameters of $\mathbf{A}$ to one is not a problem. Further than that we find that the y_{3t} on its own loads as an $SCM(1,0)$ thus we can set $a_{32} = 0$. The resulting estimated $VARMA(1,1)$ model is shown in table (11). At the bottom of column 1 in table (11), we check for the appropriateness of the model by applying the test

Table 10: Root Table

	j				
m	0	1	2	3	4
0	**0**	**0**	1	1	1
1	**2**	**3**	3	3	3
2	3	5	6	6	6
3	3	6	8	9	9
4	3	6	9	11	12

statistic $C(s)$ of section (2) shown in equation (8) to the residuals. The criterion table shows that three white noise processes are identified, i.e., the residuals follow a vector white noise process.

6.2 US financial data

The data for this empirical application consists of three US quarterly time series with $N = 173$ observations each:

1. Real US GDP growth rate calculated as.

$$y_{1,t} = 100 \times \Delta \ln (GDP)$$

2. Interest rate spread

$$y_{2,t} = \text{long-term interest rate } - \text{short-term interest rate}$$

3. Money Growth

$$y_{3,t} = 100 \times \Delta \ln (M3)$$

The three series are plotted in figure (2).

In the first stage of the identification process we can read of the criterion table and we find possible maximum orders a $VARMA(1,1)$ or a $VAR(2)$.

Once again we persist with the $VARMA(1,1)$ overall order and reading of the root table, table (13) , we find two scalar components of order $(1,0)$ and one of order $(1,1)$.

Table 11: Estimation results for the log of the Flour Price Data

	Unrestricted $VARMA$	Restricted $VARMA$
$\widehat{\mathbf{A}}$	$\begin{bmatrix} 1 & 0 & 0 \\ -0.521 \\ (-20.1)^a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ -0.523 & 1 & 0 \\ (-20.5) & & \\ 0 & 0 & 1 \end{bmatrix}$
$\widehat{c}^T$	$\begin{bmatrix} 0.199 & 0.116 & 0.336 \\ (5.835) & (3.64) & (7.352) \end{bmatrix}$	$\begin{bmatrix} 0.202 & 0.118 & 0.341 \\ (7.38) & (4.9) & (10.77) \end{bmatrix}$
$\widehat{\mathbf{\Phi}}_1$	$\begin{bmatrix} 1.078 & -0.392 & 0.273 \\ (33.781) & (-11.392) & (10.641) \\ -0.454 & 0.741 & 0.168 \\ (-17.149) & (27.607) & (8.728) \\ 0.012 & -0.297 & 1.217 \\ (0.183) & (-4.979) & (45.304) \end{bmatrix}$	$\begin{bmatrix} 1.069 & -0.382 & 0.272 \\ (43.85) & (-15.2) & (12.43) \\ -0.461 & 0.747 & 0.166 \\ (-17.72) & (30.08) & (8.96) \\ 0 & -0.283 & 1.215 \\ & (-12.4) & (53.5) \end{bmatrix}$
$\widehat{\mathbf{\Theta}}_1$	$\begin{bmatrix} -0.616 & 0.812 & 0.122 \\ (-4.455) & (10.383) & (1.588) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -0.615 & 0.814 & 0.121 \\ (-20.97) & (26.16) & (4.21) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
$\widehat{\mathbf{\Sigma}}$	$\begin{bmatrix} 1.977 & & \\ 1.012 & 0.648 & \\ 1.998 & 1.119 & 2.571 \end{bmatrix} \times 10^{-3}$	$\begin{bmatrix} 1.977 & & \\ 1.009 & 0.645 & \\ 1.998 & 1.116 & 2.572 \end{bmatrix} \times 10^{-3}$

Res C.T.[b]

	j			
m	0	1	2	3
0	0.6[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 unrestricted model

[c] *The statistics are normalised by the corresponding 5% χ^2 critical values*

Table 12: Criterion Table

	j				
m	0	1	2	3	4
0	28.23	5.54	2.61	1.61	1.49
1	1.15	**0.83**	0.28	1.18	1.06
2	**0.64**	0.3	0.91	0.81	0.77
3	0.37	0.94	0.57	0.88	0.81
4	1.51	0.98	0.93	0.83	0.76

The statistics are normalised by the corresponding 5% χ^2 critical values

Figure 2: US Financial Data

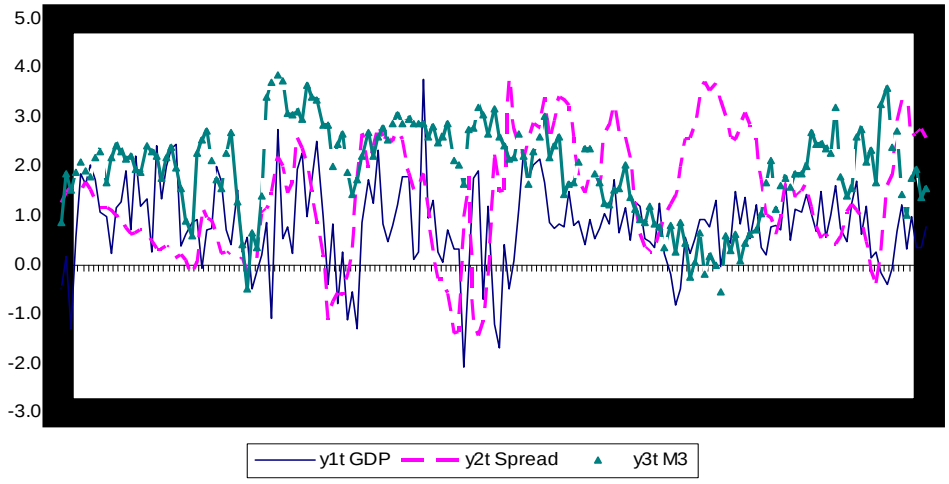

Table 13: Root Table

	j				
m	0	1	2	3	4
0	**0**	**0**	1	1	1
1	**2**	**3**	3	3	3
2	3	5	6	6	6
3	3	6	8	9	9
4	3	6	9	11	12

Again a possible representation of the model is,

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ 0 & a_{32} & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ \phi_{31}^{(1)} & \phi_{32}^{(1)} & \phi_{33}^{(1)} \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & \theta_{13}^{(1)} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{24}$$

Testing for the normalisations applied above we find that $y_{1,t}$ and $y_{3,t}$ can be modelled as an $SCM\ (1,0)$ thus we change the order of the equations for the model and write

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ \phi_{31}^{(1)} & \phi_{32}^{(1)} & \phi_{33}^{(1)} \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0 & 0 & 0 \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} & \theta_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1} \tag{25}$$

Table (14) shows the estimation results.

Table 14: Estimation results for the log of the Flour Price Data

	Unrestricted $VARMA$	Restricted $VARMA$
$\widehat{\mathbf{A}}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
$\widehat{c}^T$	$\begin{bmatrix} 0.105 \ (0.486) & 0.482 \ (1.911) & 0.391 \ (4.151) \end{bmatrix}$	$\begin{bmatrix} 0.356 \ (3.723) & 0.481 \ (4.721) & 0.348 \ (4.188) \end{bmatrix}$
$\widehat{\mathbf{\Phi}}_1$	$\begin{bmatrix} 0.271 \ (3.583) & 0.170 \ (3.038) & 0.129 \ (1.566) \\ -0.392 \ (-2.7261) & 0.881 \ (4.613) & 0.004 \ (0.046) \\ -0.026 \ (-0.516) & -0.021 \ (-0.6251) & 0.825 \ (9.536) \end{bmatrix}$	$\begin{bmatrix} 0.299 \ (4.043) & 0.155 \ (2.802) & 0 \\ -0.389 \ (-3.325) & 0.887 \ (16.763) & 0 \\ 0 & 0 & 0.821 \ (19.846) \end{bmatrix}$
$\widehat{\mathbf{\Theta}}_1$	$\begin{bmatrix} 0 & 0 & 0 \\ 0.346 \ (2.347) & 0.226 \ (2.027) & -0.140 \ (-1.269) \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.329 \ (2.692) & 0.205 \ (1.964) \\ 0 & 0 & 0 \end{bmatrix}$
$\widehat{\mathbf{\Sigma}}$	$\begin{bmatrix} 0.633 & & \\ -4.377 \times 10^{-2} & 0.298 & \\ 4.943 \times 10^{-2} & 2.338 \times 10^{-2} & 0.265 \end{bmatrix}$	$\begin{bmatrix} 0.647 & & \\ -5.009 \times 10^{-2} & 0.302 & \\ 5.002 \times 10^{-2} & 2.28 \times 10^{-2} & 0.266 \end{bmatrix}$

Res C.T.[b]

	j			
m	0	1	2	3
0	0.36[c]	0.63	0.25	0.43
1	0.59	0.46	0.43	0.66
2	0.23	0.42	0.56	0.68
3	0.40	0.67	0.68	0.69

[a] t statistics in parentheses

[b] Criterion table for the residuals from the unrestricted model

[c] *The statistics are normalised by the corresponding 5% χ^2 critical values*

7 Application to Macro-Economic Data

The aim of this section is to actually apply our modelling strategy to more than just one or two data sets, and to analyse how this modelling strategy performs in a forecasting sense. We run a forecasting race for forecasting real macroeconomic data, between $VAR(p)$ models selected via model selection criteria, and $VARMA(p,q)$ models identified via the proposed modelling strategy.

7.1 Data

The data employed in this paper are 40 monthly macroeconomic time series from 1959:1-1998:12 (i.e. $N = 480$ observations) extracted from the Stock and Watson (1999) data set (also used in Watson, 2001). The series fall within eight general categories of economic activity: (i) Output and Real Income; (ii) Employment and Unemployment; (iii) Consumption, Manufacturing, Retail Sales and Housing; (iv) Real Inventories and Sales; (v) Prices and Wages; (vi) Money and Credit; (vii) Interest Rates; $(viii)$ Exchange Rates, Stock Prices and Volume. The data was transformed in various ways replicating the transformations in Stock and Watson (1999) and Watson (2001) (see appendix A).

The initial aim of the modelling process was to randomly select any trivariate ($k = 3$) system of the series from the Stock and Watson data set and model it. Soon enough it was discovered that many series caused problems. For example when we employed series which were transformed to the second order difference of the logarithms, these returned negative roots in the moving average coefficients close to unity, which is a sign of over-differencing. Thus we have systematically chosen series to avoid these types of problems. We have created in total seventy trivariate systems, and what we can claim is that the trivariate systems we have constructed include at least one combination from each of the eight categories (i.e., at least one combination from (i), (ii) and (iii), at least one from (i), (ii) and (iv).and so on).

7.2 VAR models

Two VAR models are selected for each set of the three variables using Akaike Information Criterion (AIC) and Schwarz Criterion (SC).

$$AIC(p) = \ln\left|\widehat{\Sigma}_{\varepsilon}(p)\right| + 2 \times \frac{pk^2}{N},$$

and

$$SC(p) = \ln\left|\widehat{\Sigma}_{\varepsilon}(p)\right| + \ln(N) \times \frac{pk^2}{N},$$

where $\widehat{\Sigma}_{\varepsilon}(p)$ is the estimated variance-covariance matrix of the errors of the $VAR(p)$ model, k is the dimensionality of the system, thus making pk^2 the number of free parameters to be estimated. For $N \geq 8$ the SC is a more strict criterion than the AIC, due to the penalty factor of $\ln(N)$, thus selecting more parsimonious models. The unrestricted $VAR(p)$ models are estimated via equation by equation Ordinary Least Squares (OLS) and the restricted models are efficiently

estimated via Seemingly Unrelated Regressions (Zellner, 1962) estimation procedure.

7.3 Forecasting and Results

Each series has been divided into two subsamples. The first subsample which is often referred to as the test (or initialisation see Makridakis et. al., 1998) sample, contains the first 25 years worth of data (i.e., $N_1 = 300$). The second subsample referred to as the hold-out sample contains the last 15 year worth of data (i.e., $N_2 = 180$). Each model is estimated within the test sample and we perform out-of-sample forecasting in the hold-out sample. The out-of-sample forecast horizon ranges from $h = 1$ to 15 step ahead forecasts[9]. We forecast each horizon in a N_1 moving window manner. That is, once the model is estimated we fit the model in y_1 to y_{N_1} and forecast all out-of-sample horizons beyond this point. Then we move the fitting window along and fit it to y_2 to y_{N_1+1}. As y_{N_1+1} is now observed we forecast all out-of-sample horizons beyond this point and so on. Thus for each forecast horizon h, there are $N_2 - h + 1$ forecasts available for forecast evaluation purposes. We attempt to summarise the performance of the competing models in the next two sections.

For each of the forecast horizons we consider two popular measures of forecasting accuracy. The first is the determinant of the forecast mean squared error matrix ($|FMSE|$), and the second is the trace of the forecast mean squared error matrix ($tr\,(FMSE)$). It should be mentioned that both the test and hold-out sample observations have been standardised by the estimated standard deviation that comes from the test sample making the variances of the forecast errors of the three series directly comparable and therefore making $tr\,(FMSE)$ a legitimate measure for forecast accuracy. These measures have been compared in the two following ways.

7.3.1 Percentage better (PB)

This measure reports counts of the number of times each model has produced a smaller forecast error measure for each of the forecast horizons. The counts have been graphed in three dimensional bar charts in appendix (B), see figures (3) to (7). Also see tables (15) to (18) in appendix (C) for the raw figures. The

[9] We are trying to keep the forecast horizons as compatible as possible with other forecasting competitions such as the M1 and M3 (see Makridakis and Hibbon 1979, Makridakis and Hibbon 2000).

first observation one can make from the graphs is that their is only one instance ($h = 3$ for the $tr\,(FSME)$ of the unrestricted models) that the $VARMA$ models are outperformed by the VAR selected by SC. For all the other forecast horizons and accuracy measures the $VARMA$ models dominate.

Another observation one can make is that the $VARMA$ models seem to improve their performance over the VAR models in the longer term forecast horizons. As advocates of the $VARMA$ model we were hoping that this would be the case. We expected[10] the more parsimonious $VARMA$ models, which can be approximated by long VARs, to outperform in a forecasting sense these long VARs[11].

For both accuracy measures, of both restricted and unrestricted models, the $VARMA$ models perform better at least fifty percent (50%) of the time for forecast horizons of more than six or seven steps ahead. The importance of the 50% figure is that if we had some way of selecting the best of the VAR models (from either AIC or BIC) the $VARMA$ models will still dominate.

7.3.2 Average of the ratios of the forecast errors

Tables (19) to (22) in appendix (D) show the average ratio of the forecast error measures of the VARs over the $VARMA$ models. For example table (19) present the ratio,

$$Ratio_h = \frac{1}{M}\sum_{i=1}^{m}\frac{|FMSE\,(VAR)_i|}{|FMSE\,(VARMA)_i|},$$

for each of the forecast horizons $h = 1, \ldots, 15$, where M is the number of data sets considered. Thus figures grater than one show how much worst the VARs selected by either AIC or SC perform in comparison to the $VARMA$ models for that particular horizon.

The reason we observe these ratios as well as the PB counts above, is that it is possible that one model is best in more than 50% of the times, say 80%, but in all those cases other alternatives are close to it. However, in the 20% of cases that this model is not the best, it may make huge errors. In such a case, a user who is risk averse would not use this model, because when the model is wrong, it is badly wrong. The average of the ratios gives us this extra information.

Moreover, we take the ratios first and then average them afterwards, rather than average the det $(FSME)$ over all M data sets and then find the ratio, because different sets of variables may have different degrees of predictability.

[10] Reference of Tunicliffe-Wilson who states exactly this.

[11] Various forecasting studies (see the references to the M competitions) and experts advertise the moto of parsimony for better forecasting performance.

The figures show that for each of the forecast horizons the $VARMA$ models outperform the VARs. Moreover these tables confirm the observation made above about the $VARMA$ models performing better in the longer horizon forecasts as we see the average ratio measure increases the further out we forecast.

8 Summary and future direction

This paper suggests an extension to Tiao and Tsay's methodology for identification of $VARMA$ models in the direction foreseen by researchers who commented on the original work of Tiao and Tsay. The simulations are based on a very small number of repetitions, due to the complex and time consuming nature of the process of identification. However they do present the reader with some insight to the performance and the usefulness of each stage of the proposed identification process. The empirical section is split in two parts. Firstly, it presents two examples of applying the proposed $VARMA$ modelling process in full. Secondly it includes an extensive empirical application, using many sets of macroeconomic variables from the Stock and Watson(1999) data set, to perform out-of-sample forecasting comparisons between estimated VAR and $VARMA$ models. From the empirical application we find that there are many cases in out-of sample forecasting where the $VARMA$ models outperform their VAR counterparts. This conclusion is stronger when we concentrate on the longer term forecast horizons. Thus from this study we can conclude that there is great scope in investing more time for the development of the $VARMA$ models, in an empirical sense. That is to develop these models in a way that empirical econometrician can consider them as alternatives to other multivariate modelling techniques.

A Data Summary

This appendix lists the time series employed in this paper. The series have been directly downloaded from Mark Watson's web page. The names (mnemonics) given to each series and the brief description following the series name have been replicated as in Watson (2001). The superscript index on the series name is the transformation code which corresponds to: (1) level of the series $\left(y_t^1\right)$, (2) first difference $\left(y_t^2 = \Delta y_t = y_t - y_{t-1}\right)$ and (3) first difference of the logarithm i.e. series transformed to growth rates $\left(y_t^3 = 100 * \Delta \ln y_t\right)$. Also the following abbreviations appear in the brief data descriptions: SA = seasonally adjusted; SAAR = seasonally adjusted at an annual rate; NSA = not seasonally adjusted.

(*i*) *Output and Income*

1.	IP[3]	Industrial Production: Total Index (1992=100,SA)
2.	IPP[3]	Industrial Production: Products, Total (1992=100,SA)
3.	IPF[3]	Industrial Production: Final Products (1992=100,SA)
4.	IPC[3]	Industrial Production: Consumer Goods (1992=100,SA)
5.	IPUT[3]	Industrial Production: Utilities (1992=100,SA)
6.	PMP[1]	NAPM Production Index (Percent)
7.	GMPYQ[3]	Personal Income (Chained) (series #52) (Bil 92$, SAAR)

(*ii*) *Employment and Hours*

8.	LHUR[1]	Unemployment Rate: All Workers, 16 Years & over (%,SA)
9.	LPHRM[1]	Avg. weekly hrs. of Production Wkrs.: Mfg., Manufacturing. (SA)
10.	LPMOSA[1]	Avg. weekly hrs. of Production Wkrs.: Mfg., Overtime Hrs. (SA)
11.	PMEMP[1]	NAPM Employment Index (Percent)

(*iii*) *Consumption, Manufacturing and Retail Sales, and Housing*

12.	MSMTQ[3]	Manufacturing & Trade: Total (Mil of Chained 1992 Dollars) (SA)
13.	MSMQ[3]	Manufacturing & Trade: Manufacturing, Total (Mil of Chained $1992 SA)
14.	MSDQ[3]	Manufacturing & Trade: Manufacturing, Durable Goods (Mil of Chained $1992 SA)
15.	MSNQ[3]	Manufacturing & Trade: Manufacturing, Nondurable Goods (Mil of Ch. $1992 SA)
16.	WTQ[3]	Merchant Wholesalers: Total (Mil of Chained 1992 Dollars) (SA)
17.	WTDQ[3]	Merchant Wholesalers: Durable Goods Total (Mil of Chained 1992 Dollars) (SA)
18.	WTNQ[3]	Merchant Wholesalers: Nondurable Goods Total (Mil of Chained $1992 SA)
19.	RTQ[3]	Retail Trade: Total (Mil of Chained 1992 Dollars) (SA)
20.	RTNQ[3]	Retail Trade: Nondurable Goods (Mil of Chained 1992 Dollars) (SA)
21.	CMCQ[3]	Personal Consumption Expend (Chained) - Total (Bil $92, SAAR)

(*iv*) *Real Inventories and Inventory-Sales Ratios*

22. IVMFGQ[3] Inventories, Business, Manufacturing (Mil of Chained 1992 Dollars) (SA)
23. IVMFDQ[3] Inventories, Business Durables (Mil of Chained 1992 Dollars) (SA)
24. IVMFNQ[3] Inventories, Business Nondurables (Mil of Chained 1992 Dollars) (SA)
25. IVSRQ[2] Ratio for Manufacturing & Trade: Inventory/Sales (Chained 1992 Dollars) (SA)
26. IVSRMQ[2] Ratio for Manufacturing & Trade: Manufacturing Inventory/Sales (87$) (SA)
27. IVSRWQ[2] Ratio for Manufacturing & Trade: Wholesaler; Inventory/Sales (87$) (SA)
28. IVSRRQ[2] Ratio for Manufacturing & Trade: Retail Trade; Inventory/Sales (87$) (SA)
29. MOCMQ[3] New Orders (net) - Consumer Goods & Materials, 1992 Dollars (BCI)
30. MDOQ[3] New Orders, Durable Goods Industries, 1992 Dollars (BCI)

(*v*) *Prices and Wages*

31. PMCP[1] NAPM Commodity Prices Index (Percent)

(*vi*) *Money and Credit Quantity Aggregates*

32. FM2DQ[3] Money Supply - M2 in 1992 Dollars (BCI)
33. FCLNQ[3] Commercial & Industrial Loans Outstanding in 1992 Dollars (BCI)

(*vii*) *Interest Rates*

34. FYGM3[2] Interest Rate: US Treasury Bills, Sec Mkt, 3-MO. (% p.a. NSA)
35. FYGM6[2] Interest Rate: US Treasury Bills, Sec Mkt, 6-MO. (% p.a. NSA)
36. FYGT1[2] Interest Rate: US Treasury Const Maturities, 1-YR. (% p.a. NSA)
37. FYGT10[2] Interest Rate: US Treasury Const Maturities, 10-YR. (% p.a. NSA)
38. TBSPR[1] Term Spread FYGT10-FYGT1

(*viii*) *Exchange Rates, Stock Prices and Volume*

39. FSNCOM[3] NYSE Common Stock Prices Index: Composite (12/31/65=50)
40. FSPCOM[3] S&P's Common Stock Prices Index: Composite (1941-43=10)

B Percentage Better Counts (Graphs)

Figure 3: PB counts for $|FMSE|$ for the Unrestricted Models

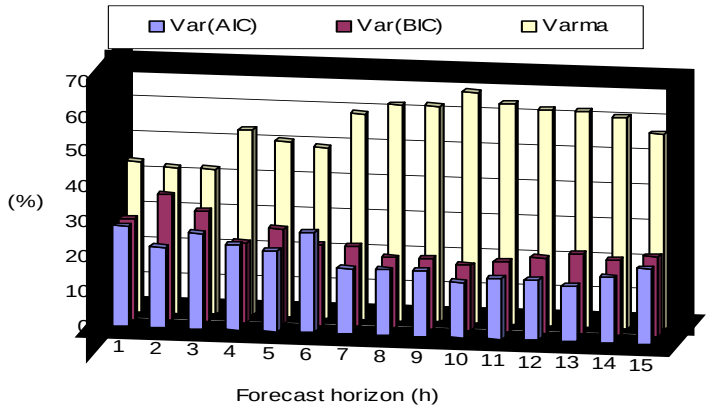

Figure 4: PB counts for $|FMSE|$ for the Restricted Models

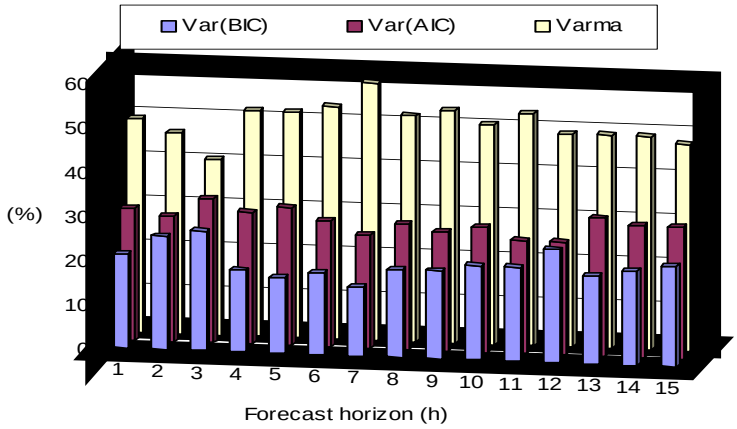

Figure 5: PB counts for $tr\,(FMSE)$ for the Unrestricted Models

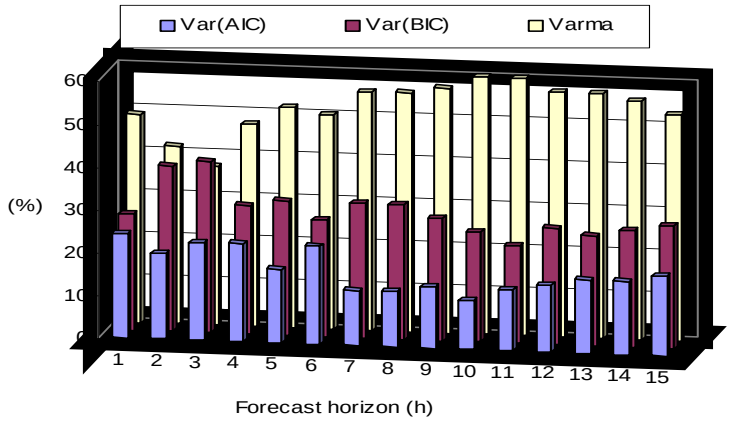

Figure 6: PB counts for $tr\,(FMSE)$ for the Restricted Models

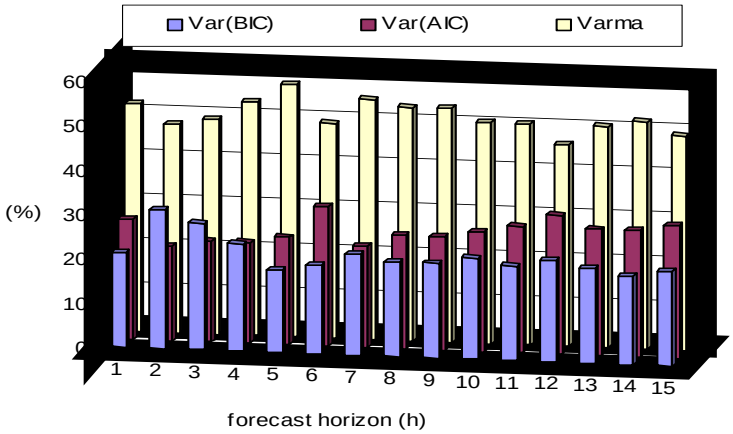

Figure 7:

C Percentage Better Counts (Raw Data)

Table 15: PB Counts for det(FMSE) of Unrestricted models

	Forecast Horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$VARMA$	43[a]	41	41.5	53	50	48.5	58.5	61	61	65.5	63	61.5	61	60	56
$VAR(AIC)$	28.5	23	27	24	23	28.5	18.5	19	19	15.5	17	17	16	19	21
$VAR(SC)$	28.5	36	31.5	23	27	23	23	20	20	19	20	21.5	23	21	23

Table 16: PB Counts for det(FMSE) of Restricted models

	Forecast Horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$VARMA$	49[a]	45.5	40	51	51.5	53	59	51	53	50	53	49	49	49	47
$VAR(AIC)$	30	29	33	30	31.5	28.5	25.5	29	27	29	25.5	25.5	31	30	30
$VAR(SC)$	21	25.5	27	19	17	18.5	15.5	20	20	21	21.5	25.5	20	21	23

Table 17: PB Counts for tr(FMSE) of Unrestricted models

	Forecast Horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$VARMA$	49[a]	41	37	47	51.5	50	56	56	57	63	63	57	57	56	53
$VAR(AIC)$	24	20	23	23	17	23	13	13	14	11	14	16	17	17	18.5
$VAR(SC)$	27	39	40	30	31.5	27	31	31	29	26	23	27	26	27	28.5

Table 18: PB Counts for det(FMSE) of Restricted models

	Forecast Horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$VARMA$	51.5[a]	47	48.5	53	57	49	54	53	53	50	50	46	50	51	49
$VAR(AIC)$	27	21.5	23	23	24	31	23	26	26	27	29	31	29	29	30
$VAR(SC)$	21.5	31.5	28.5	24	19	20	23	21	21	23	21	23	21	20	21

[a] Figures are rounded to the nearest .5 figure

D Average Ratios

Table 19: Average Ratio of det(FTSE) of VAR over VARMA for Unrestricted models

	Forecast Horizon (h)								Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*6*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
$VAR(AIC)$	1.058	1.079	1.059	1.078	1.079	1.08	1.087	1.08	1.069	1.075	1.078	1.08
$VAR(SC)$	1.043	1.055	1.062	1.099	1.112	1.11	1.099	1.087	1.065	1.089	1.094	1.094

Table 20: Average Ratio of det(FTSE) of VAR over VARMA for Restricted models

	Forecast Horizon (h)								Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*6*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
$VAR(AIC)$	1.022	1.03	1.03	1.031	1.028	1.027	1.032	1.035	1.028	1.028	1.029	1.03
$VAR(SC)$	1.011	1.01	1.013	1.021	1.023	1.025	1.029	1.031	1.014	1.019	1.022	1.024

Table 21: Average Ratio of tr(FTSE) of VAR over VARMA for Unrestricted models

	Forecast Horizon (h)								Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*6*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
$VAR(AIC)$	1.034	1.051	1.036	1.055	1.061	1.065	1.071	1.064	1.044	1.054	1.059	1.061
$VAR(SC)$	1.039	1.053	1.069	1.106	1.125	1.123	1.113	1.1	1.067	1.096	1.103	1.103

Table 22: Average Ratio of tr(FTSE) of VAR over VARMA for Restricted models

	Forecast Horizon (h)								Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*6*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
$VAR(AIC)$	1.024	1.031	1.03	1.031	1.021	1.016	1.019	1.024	1.029	1.025	1.022	1.022
$VAR(SC)$	1.017	1.02	1.024	1.03	1.03	1.03	1.034	1.037	1.023	1.026	1.028	1.03

References

[1] Anderson, H.M. and Vahid, F. (1998), 'Testing Multiple Equation Systems for Common Nonlinear Components', *Journal of Econometrics*, 84, 1-36.

[2] Grubb, H. (1992), 'A Multivariate Time Series Analysis of Some Flour Price Data', *Applied Statistics*, 41, 1, 95-107.

[3] Hamilton, J. D. (1994), *Time Series Forecasting*, Princeton, Princeton University Press.

[4] Hannan, E.J. (1969), 'The Identification of vector mixed autoregressive-moving average systems', *Biometrika*, 56, 223-225

[5] Hannan, E.J. and M. Deistler (1988), *The Statistical Theory of Linear Systems*, Wiley.

[6] Hannan, E.J and J. Rissanen (1982), 'Recursive Estimation of Autoregressive-Moving Average Order', *Biometrika*, 69, 1, 81-94

[7] Harvey, A.C. (1991), *Forecasting, Structural Time Series Models and the Kalman Filter*,.Cambridge University Press.

[8] Hotelling, H. (1935), 'Demand Functions with Limited Budgets', *Econometrica*, Vol. 3, p.66-78.

[9] Lutkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin-Heidelberg.

[10] Lutkepohl, H. and Poskitt, D.S. (1996), 'Specification of Echelon-Form $VARMA$ Models', *Americal Statistical Association: Journal of Business and Economic Statistics*, 14, 1, 69-79.

[11] Makridakis, S. and Hibon, M. (1979), 'Accuracy of Forecasting: An Empirical Investigation', (with discussion), *Journal of the Royal Statistical Society, Series A*, Vol. 142, Part 2, 79-145.

[12] Makridakis, S. and Hibon, M. (2000), 'The M3-Competition: Results, conclusions and implications', *International Journal of Forecasting*, 16, 451-476.

[13] Makridakis, S., Wheelwright S. C., Hyndman, R. (1998), *Forecasting Methods and Application*, (3rd Edition), New York, Wiley.

[14] Pena, D. and Box, G.E.P. (1987), 'Identifying a simplifying structure in time series', *Journal of the American Statistical Association*, 82, 836-843.

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

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

[17] Tiao, G.C and Tsay, R.S. (1989), 'Model Specification in Multivariate Time Series', *Journal of the Royal Statistical Society*, 51, 2, 157-213.

[18] Tunnicliffe Wilson, G. (1973), 'The Estimation of Parameters in Multivariate Time Series Models', *Journal of the Royal Statistical Society B*, 35, 1, 76-85.

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

[20] F. Vahid and Issler, J.V. (2002), 'The importance of common cyclical features in VAR analysis: A Monte-Carlo study', *Journal of Econometrics*, 109, 341-363.

[21] Watson, M. W. (2001), *Macroeconomic Forecasting Using Many Predictors*, in M. Dewatripont, L. Hansen and S. Turnovsky (eds), Advances in Economics and Econometrics, Theory and Applications, Eight World Congress of the Econometric Society, Vol. III, page 87-115.

[22] Zellner, A. (1963), 'Estimators for seemingly unrelated regression equations: Some exact finite sample results', *Journal of the American Statistical Society* 58: 977-992.