ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Two canonical VARMA forms: Scalar component models vis-à-vis the Echelon form

George Athanasopoulos, D. S. Poskitt and

Farshid Vahid

July 2007

Working Paper 10/07

Two canonical VARMA forms: Scalar component models vis-à-vis the Echelon form

George Athanasopoulos
Department of Econometrics and Business Statistics,
Monash University, VIC 3800, Australia.
Email: George.Athanasopoulos@buseco.monash.edu.au

D. S. Poskitt
Department of Econometrics and Business Statistics,
Monash University, VIC 3800, Australia.
Email: Don.Poskitt@buseco.monash.edu.au

Farshid Vahid
School of Economics,
Australian National University, ACT 0200, Australia.
Email: Farshid.Vahid@anu.edu.au

24 July 2007

JEL classification: C32, C51

Two canonical VARMA forms: Scalar component models vis-à-vis the Echelon form

Abstract: In this paper we study two methodologies which identify and specify canonical form VARMA models. The two methodologies are: (i) an extension of the scalar component methodology which specifies canonical VARMA models by identifying scalar components through canonical correlations analysis and (ii) the Echelon form methodology which specifies canonical VARMA models through the estimation of Kronecker indices. We compare the actual forms and the methodologies on three levels. Firstly we present a theoretical comparison. Secondly, we present a Monte-Carlo simulation study that compares the performance of the two methodologies in identifying some pre-specified data generating processes. Lastly we compare the out-of-sample forecast performance of the two forms when models are fitted to real macroeconomic data.

Keywords: Echelon form, Identification, Multivariate time series, Scalar component, VARMA model.

1 Introduction

Macroeconomists analyse and forecast aggregate economic activity by studying the dynamics of economic variables such as GDP growth, unemployment and inflation. Univariate autoregressive integrated moving average (ARIMA) processes are a useful class of models for capturing and describing the dynamics of such series. Box and Jenkins (1970) Box and Jenkins (1970) popularised this useful univariate methodology making it arguably the most well known time series tool. However, ARIMA modelling is limited by its inability to capture and model important dynamic inter-relationships between variables of interest.

The direct generalisation of the stationary ARMA model to the multivariate form, leads to the vector ARMA or VARMA model (see amongst others, Quenouille, 1957; Tunnicliffe-Wilson, 1973; Tiao and Box, 1981; Tsay, 1989; Tiao, 2001; Athanasopoulos and Vahid, 2007). This generalisation has been proven to be far from trivial. One of the major issues faced by researchers in the multivariate time series field of VARMA modelling, relates to the identification of unique representations. The issues of identification have been discussed over the years by many researchers including Hannan (1969, 1970, 1976, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988) Lütkepohl (1993) and Reinsel (1997). In this paper we study and compare two methodologies that overcome this issue and achieve unique canonical VARMA representations.

The first methodology is the Athanasopoulos and Vahid (2006) extension to Tiao and Tsay (1989). This methodology comprises three stages. In the first stage, “scalar component models” (SCMs) embedded in the VARMA model are identified using a series of tests based on canonical correlations analysis between judiciously chosen sets of variables. In the second stage, a fully identified structural form is developed through a series of logical deductions and additional canonical correlations tests. Then in the final stage, the identified model is estimated using full information maximum likelihood (FIML).

The second methodology we consider is the Echelon form methodology which involves specifying canonical Echelon form models through the estimation of Kronecker indices. Kronecker indices are simply the maximal row degrees of each individual equation of a VARMA model and are estimated through a series of least squares regressions. Hence, the advocates of this methodology portray its simplicity as a great attribute. This methodology has been developed by many time series analysts such as Akaike (1974, 1976), Kailath (1980), Hannan and Kavalieris (1984), Solo (1986), Hannan and Deistler (1988), Tsay (1991), Lütkepohl (1993), Nsiri and Roy (1992, 1996), Poskitt (1992), Lütkepohl and Poskitt (1996) among others. However, no investigation has been

undertaken into the finite sample performance of this methodology when attempting to identify VARMA models. In this paper we conduct Monte-Carlo experiments and we evaluate the ability of the Echelon form methodology and the SCM methodology to identify some pre-specified VARMA data generating processes (DGPs).

Furthermore, to compare the forecasting performance of the VARMA models specified by the two methodologies we use real data. We compile seventy trivariate sets of monthly macroeconomic variables, and fit canonical SCM VARMA models and canonical Echelon form VARMA models to them, using only one portion of the available sample for estimation and holding the rest of the sample for forecast comparison. Using the estimated models, we forecast these variables 1 to 15 steps into the future throughout the forecast period. We then use several measures of forecast accuracy to compare the performance of the two forms of VARMA models.

The structure of this paper is as follows. Section 2 outlines the scalar component modelling methodology for VARMA models. Section 3 presents the methodology we have implemented in order to identify canonical reverse Echelon form VARMA models. In section 4 we present a theoretical comparison of the two forms and also some experimental results. Section 5 describes the data and the forecast evaluation method and presents the empirical results. Section 6 provides some conclusions and some directions for future research.

2 A VARMA modelling methodology based on scalar components

The scalar component methodology we employ in this paper is the Athanasopoulos and Vahid (2006) extension to the Tiao and Tsay (1989) methodology. In this section we present a brief overview of the methodology. For more details, readers should refer to the above mentioned papers.

The aim of identifying scalar components is to examine whether there are any simplifying embedded structures underlying a VARMA(p,q) process.

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

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

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

$$\begin{aligned}
\boldsymbol{\alpha}' \boldsymbol{\Phi}_{p_1} &\neq \mathbf{0}^T \text{ where } 0 \leq p_1 \leq p; \\
\boldsymbol{\alpha}' \boldsymbol{\Phi}_{l} &= \mathbf{0}^T \text{ for } l = p_1 + 1, \ldots, p; \\
\boldsymbol{\alpha}' \boldsymbol{\Theta}_{q_1} &\neq \mathbf{0}^T \text{ where } 0 \leq q_1 \leq q; \\
\boldsymbol{\alpha}' \boldsymbol{\Theta}_{q_l} &= \mathbf{0}^T \text{ for } l = q_1 + 1, \ldots, q.
\end{aligned}$$

Notice that 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. Tiao and Tsay (1989) employ a sequence of canonical correlations tests to discover K such linear combinations.

Denote the squared sample canonical correlations between $\boldsymbol{Y}_{m,t} \equiv (\boldsymbol{y}_t', \ldots, \boldsymbol{y}_{t-m}')$ and $\boldsymbol{Y}_{h,t-1-j} \equiv (\boldsymbol{y}_{t-1-j}', \ldots, \boldsymbol{y}_{t-1-j-h}')'$ by $\widehat{\lambda}_1 < \widehat{\lambda}_2 < \ldots < \widehat{\lambda}_K$. The test statistic suggested by Tiao and Tsay (1989) for testing for the null of at least s SCM(p_i, q_i) against the alternative of fewer than s scalar components is

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

where 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,

$$d_i = 1 + 2 \sum_{v=1}^{j} \widehat{\rho}_v \left(\widehat{\boldsymbol{r}}_i' \boldsymbol{Y}_{m,t} \right) \widehat{\rho}_v \left(\widehat{\boldsymbol{g}}_i' \boldsymbol{Y}_{h,t-1-j} \right) \tag{3}$$

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

Suppose we have K linearly independent scalar components characterized by the transformation matrix $\boldsymbol{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K)'$. If we rotate the system in equation (1) by $\boldsymbol{A}$, we obtain

$$\boldsymbol{A}\boldsymbol{y}_t = \boldsymbol{\Psi}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Psi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}_1^* \boldsymbol{\varepsilon}_{t-1} - \ldots - \boldsymbol{\Theta}_q^* \boldsymbol{\varepsilon}_{t-q}, \tag{4}$$

where $\boldsymbol{\Psi}_i = \boldsymbol{A}\boldsymbol{\Phi}_i$, $\boldsymbol{\varepsilon}_t = \boldsymbol{A}\boldsymbol{\eta}_t$ and $\boldsymbol{\Theta}_i^* = \boldsymbol{A}\boldsymbol{\Theta}_i \boldsymbol{A}^{-1}$, in which the right hand side coefficient matrices have many rows of zeros. However, as the following simple example shows, even if $\boldsymbol{A}$ is known

there are still situations where the system is not identified.

Example 2 *Consider the bivariate VARMA*$(1,1)$ *system with two scalar components SCM*$(1,1)$ *and SCM*$(0,0)$*, i.e.,*

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} \\ 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} \\ 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}.$$

The second row of the system implies that

$$a_{21} y_{1,t-1} + a_{22} y_{2,t-1} = \varepsilon_{2,t-1}.$$

$y_{1,t-1}$, $y_{2,t-1}$ *and* $\varepsilon_{2,t-1}$ *all appear in the right hand side of the first equation of the system and therefore their coefficients are not identified. We set* $\theta_{12}^{(1)} = 0$ *to achieve identification.*

In general if there exist two scalar components SCM(p_r, q_r) and SCM(p_s, q_s), where $p_r > p_s$ and $q_r > q_s$, the system will not be identified. In such cases $\min\{p_r - p_s, q_r - q_s\}$, autoregressive or moving average parameters must be set to zero for the system to be identified. This is referred to as the "general rule of elimination" (Tiao and Tsay, 1989). The methodology we employ here requires us to set the moving average parameters to zero in these situations (see Athanasopoulos and Vahid, 2006, for a more detailed explanation).

Tiao and Tsay (1989) construct a consistent estimator for $\boldsymbol{A}$ using the estimated canonical covariates corresponding to insignificant canonical correlations. Conditional on these estimates, they estimate the row sparse parameter matrices on the right hand side of equation (1). The lack of proper attention to efficiency in the estimation of $\boldsymbol{A}$, which affects the accuracy of the second stage estimates, was a major criticism of the Tiao and Tsay methodology raised by many eminent time series analysts (see the discussion by Chatfield, Hannan, Reinsel, Tunnicliffe-Wilson that followed Tiao and Tsay, 1989).

The Athanasopoulos and Vahid (2006) extension to the Tiao and Tsay (1989) methodology, concentrates on establishing necessary and sufficient conditions for the identification of $\boldsymbol{A}$ such that all parameters of the system can be estimated simultaneously using FIML (Durbin, 1963). We refer to this system as a canonical SCM representation. These rules are:

1. Each row of $\boldsymbol{A}$ can be multiplied by a constant without changing the structure of the model. Hence, we are free to normalize one parameter in each row to one. However, as always

in such situations, there is a danger of choosing a parameter whose true value is zero for normalization, i.e., a zero parameter might be normalized to one. To safeguard against this, the 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 whether 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 normalized to one and the corresponding coefficients in all other SCMs 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 normalized 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 omitting variable $K-1$ and test whether the SCM could be formed from the first $K-2$ variables only, and so on.

2. 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 of the system. This means that the row of $\boldsymbol{A}$ 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 $\boldsymbol{A}$ corresponding to the SCM$\big(p_2,q_2\big)$ is normalized to one, the parameter in the same position in the row of $\boldsymbol{A}$ corresponding to SCM(p_1,q_1) should be restricted to zero.

A brief summary of our complete VARMA methodology is as follows.

Stage I: Identification of the scalar components

This stage follows the Tiao and Tsay (1989) methodology and comprises two steps:

Step 1: Determining an overall tentative Order

Starting from $m=0$, $j=0$ and incrementing sequentially one at a time, find all zero sample canonical correlations between $\boldsymbol{Y}_{m,t}$ and $\boldsymbol{Y}_{m,t-1-j}$. Organize the results in a two way table. Starting from the upper left corner and considering the diagonals perpendicular to the main diagonal, search for the first time $s+K$ zero eigenvalues are found, given that there were s zero eigenvalues

in position $(p-1, q-1)$ (when either $p=1$ or $q=1$, $s=0$). This (p,q) is taken as the overall order of the system. Note that it is possible to find more than one such (p,q) and therefore more than one possible overall order. In such cases one should pursue all of these possibilities and choose between competing models using a model selection criterion. (This procedure produces exactly the same results as those implied by the "Criterion Table" in Tiao and Tsay (1989).

***Step 2*: Identifying orders of SCMs**

Conditional on (p,q), test for zero canonical correlations between $\boldsymbol{Y}_{m,t}$ and $\boldsymbol{Y}_{m+(q-j),t-1-j}$ for $m = 0,\ldots,p$ and $j = 0,\ldots,q$. Note that since an SCM(m,j) nests all scalar components of order $(\leq m, \leq j)$, for every one SCM$(p_1 < p, q_1 < q)$ there will be $s = \min\{m-p_1+1, j-q_1+1,\}$ zero canonical correlations at position $(m \geq p_1, j \geq q_1)$. Therefore, for every increment above s, a new SCM(m,j) is found. This procedure does not necessarily lead to a unique decision about the embedded SCMs. In all such cases all possibilities should be pursued and the final models can be selected based on a model selection criterion. (The tabulation of all zero eigenvalues produces the "Root Table" of Tiao and Tsay (1989).

Stage II: Placing identification Restrictions on Matrix A

Apply the identification rules stated above to identify the structure of the transformation matrix $\boldsymbol{A}$.

Stage III: Estimation of the Uniquely Identified System

Estimate the parameters of the identified structure using FIML (Durbin, 1963). 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 Rissanen (1982). The maximum likelihood procedure provides estimates and estimated standard errors for all parameters, including the free parameters in $\boldsymbol{A}$. All usual considerations that ease the estimation of structural forms are also applicable here, and should definitely be exploited in estimation.

3 Canonical Reverse Echelon Form

A $K-$dimensional VARMA representation, such as

$$\boldsymbol{\Phi}(L)\boldsymbol{y}_t = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t, \tag{5}$$

where $\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}_0 - \boldsymbol{\Phi}_1 L - \ldots - \boldsymbol{\Phi}_p L^p$ and $\boldsymbol{\Theta}(L) = \boldsymbol{\Theta}_0 - \boldsymbol{\Theta}_1 L - \ldots - \boldsymbol{\Theta}_q L^q$ is said to be in Echelon form if the pair of polynomials, in the lag operators $\boldsymbol{\Phi}(L) = \left[\phi_{rc}(L)\right]_{r,c=1,\ldots,K}$ and $\boldsymbol{\Theta}(L) = \left[\theta_{rc}(L)\right]_{r,c=1,\ldots,K}$, $[\boldsymbol{\Phi}(L) : \boldsymbol{\Theta}(L)]$, is left coprime and possess the following properties:

1. $\boldsymbol{\Phi}_0 = \boldsymbol{\Theta}_0$ is lower triangular with unit diagonal elements,
2. row r of the polynomial operators $[\boldsymbol{\Phi}(L) : \boldsymbol{\Theta}(L)]$ is of maximum degree k_r,
3. the operators have the form of

$$\phi_{rr}(L) = 1 - \sum_{j=1}^{k_r} \phi_{rr}^{(j)} L^j \quad \text{for} \quad r = 1, \ldots, K,$$

$$\phi_{rc}(L) = - \sum_{j=k_r - k_{rc}+1}^{k_r} \phi_{rc}^{(j)} L^j \quad \text{for} \quad r \neq c,$$

$$\theta_{rc}(L) = \theta_{rc}^{(0)} - \sum_{j=1}^{k_r} \theta_{rc}^{(j)} L^j \ \text{ with } \theta_{rc}^{(0)} = \phi_{rc}^{(0)} \text{ for } r, c = 1, \ldots, K,$$

where $\phi_{rc}^{(j)}$ specifies the element of $\boldsymbol{\Phi}_j$ in row r and column c, and $\theta_{rc}^{(j)}$ specifies the element of $\boldsymbol{\Theta}_j$ in row r and column c.

The maximum row degrees $\boldsymbol{k} = (k_1, \ldots, k_K)'$ are called the **Kronecker Indices** which define the structure of the system, and

$$k_{rc} = \begin{cases} \min(k_r + 1, k_c) & \text{for} \quad r \geq c \\ \min(k_r, k_c) & \text{for} \quad r < c \end{cases},$$

for $r, c = 1, \ldots, K$, specifies the number of free parameters in the operator $\phi_{rc}(L)$ for $r \neq c$. The sum of the Kronecker indices $m = \sum_{r=1}^{K} k_r$ is called the **McMillan degree**. The maximum number of freely varying parameters is $d(\boldsymbol{k}) = 2mK$.

Example 3 *Consider a trivariate stable and invertible VARMA process with Kronecker indices,*

$\boldsymbol{k} = (k_1, k_2, k_3) = (1, 1, 0)$. *The total number of freely varying parameters is* $d(\boldsymbol{k}) = 2mK = 2 \times 2 \times 3 = 12$. *The Echelon form representation of the process is,*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \phi_{31}^{(0)} & \phi_{32}^{(0)} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & 0 \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\Theta}_0 \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & \theta_{13}^{(1)} \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} & \theta_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{6}$$

This example demonstrates how Echelon form imposes zero/one restrictions on the parameters of the polynomial lag operators $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ making the VARMA process uniquely identifiable by guaranteeing that the only (possible) unimodular common operator between the resulting left coprime coefficient matrices $\boldsymbol{\Phi}(L)$ and $\boldsymbol{\Theta}(L)$ is $\boldsymbol{\Delta}(L) = \boldsymbol{I}_K$ (see Hannan and Deistler 1988 or Lütkepohl 1993 for detailed discussions on left coprime matrices). Note that if all Kronecker indices are equal, that is $k_1 = k_2 = \ldots = k_K = k$, then all $k_{rc} = k$ and $\boldsymbol{\Phi}_0 = \boldsymbol{\Theta}_0 = \boldsymbol{I}_K$. This type of Echelon form representation is a standard VARMA(p, p) model (for further examples of Kronecker indices and the M^{c}Millan degree of VARMA processes refer to Solo 1986, Hannan and Kavalieris 1984, Hannan and Deistler 1988 and Tsay 1989).

There are two types of restrictions we impose on the right hand side of equation (6). Firstly the zero restrictions we impose on the third row of the coefficients of both $\boldsymbol{y}_{t-1}$ and $\boldsymbol{\varepsilon}_{t-1}$. These restrictions are imposed from the identified order of the model specified by the Kronecker indices. That is, the maximum order of the third row is zero, i.e., $k_3 = 0$.

Secondly, we impose two extra zero restrictions on the first two rows of the coefficients of $\boldsymbol{y}_{t-1}$, i.e., $\phi_{13}^{(1)} = \phi_{23}^{(1)} = 0$. These restrictions are due to the fact that the pairs of parameters $\phi_{13}^{(1)}, \theta_{13,}^{(1)}$ and $\phi_{23}^{(1)}, \theta_{23}^{(1)}$ cannot be uniquely identified. Recall that in the canonical SCM specification of VARMA models, the "general rule of elimination" was applied in such cases. In order to be consistent with the scalar component canonical form specified is Section 2, we can alternatively impose these zero restrictions on the moving average parameters, i.e., $\theta_{13}^{(1)} = \theta_{23}^{(1)} = 0$, and hence specify the canonical reverse Echelon form proposed by Lutkepohl and Claessen (1997),

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \phi_{31}^{(0)} & \phi_{32}^{(0)} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\Theta}_0 \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{7}$$

This amounts to replacing property 3 by

3′. the operators have the form of

$$\theta_{rr}(L) = 1 - \sum_{j=1}^{k_r} \theta_{rr}^{(j)} L^j \quad \text{for} \quad r = 1, \ldots, K,$$

$$\theta_{rc}(L) = - \sum_{j=k_r - k_{rc}+1}^{k_r} \theta_{rc}^{(j)} L^j \quad \text{for} \quad r \neq c,$$

$$\phi_{rc}(L) = \phi_{rc}^{(0)} - \sum_{j=1}^{k_r} \phi_{rc}^{(j)} L^j \quad \text{with } \phi_{rc}^{(0)} = \theta_{rc}^{(0)} \text{ for } r, c = 1, \ldots, K,$$

i.e., interchanging the coefficient restrictions on $\Phi(L)$ and $\Theta(L)$.

Finally, note that the coefficients that appear in (6) and (7) are freely varying in $\mathbb{R}^{12}$. This means that some of these coefficients can also be zero, provided that the row degrees, i.e., the Kronecker indices, are maintained. Thus, in (7), at least one of $\phi_{11}^{(1)}, \phi_{12}^{(1)}, \phi_{13}^{(1)}, \theta_{11}^{(1)}$ and $\theta_{12}^{(1)}$ must be non-zero. Respectively, at least one of the second row freely varying parameters must be non-zero, i.e., at least one of $\phi_{2c}^{(1)}$ for $c = 1, 2, 3$ or $\theta_{2c}^{(1)}$ for $c = 1, 2$, must be non-zero.

A brief summary of the Echelon form methodology implemented in this paper follows.

Stage I A long order VAR(h) is fitted and the estimated residuals $\widehat{\boldsymbol{\varepsilon}}_t(h)$ are obtained. These are used as estimates of the lagged innovations in subsequent stages. As suggested by Lütkepohl and Poskitt (1996) we take $h = \ln(T)$. The general idea is that h has to be greater than the largest Kronecker index.

Stage II Using the estimated residuals from Stage I, Echelon form VARMA models of the form

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \left(\boldsymbol{\Phi}_0 - \boldsymbol{I}_K\right)\left(\widehat{\boldsymbol{\varepsilon}}_t(h) - \boldsymbol{y}_t\right) + \boldsymbol{\Theta}_1 \widehat{\boldsymbol{\varepsilon}}_{t-1}(h) + \ldots + \boldsymbol{\Theta}_q \widehat{\boldsymbol{\varepsilon}}_{t-p}(h) + \boldsymbol{\varepsilon}_t$$

are fitted for a range of Kronecker indices and the optimum model based on model selection criteria, conditional on the selected set of Kronecker indices is selected. There are two issues that need to be addressed here. These are: (i) which model selection criterion should be used and (ii) which efficient procedure for searching for the optimal set of Kronecker indices should be employed.

We employ Poskitt's (1992) search procedure coupled with the BIC as the model selection criterion. Extensive Monte-Carlo experiments in Athanasopoulos (2005) show that with Poskitt's procedure the BIC outperforms the AIC and the HQ, especially for sample sizes of 200 observations or more. For smaller samples the HQ may also be considered.

Poskitt's (1992) procedure explores a significant property of Echelon forms. The restrictions of the r^{th} equation imposed by a set of Kronecker indices $\mathbf{k} = (k_1, \ldots, k_K)$ depend on the Kronecker indices $k_i \leq k_r$. They do not depend on indices greater than k_r. Using this property the search starts from all Kronecker indices being set to zero and increments them. The following example demonstrates the search procedure.

Example 4 *Consider the canonical reverse Echelon form model of equation (7),*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \phi_{31}^{(0)} & \phi_{32}^{(0)} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\Theta}_0 \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \qquad (8)$$

Poskitt's search procedure starts from Kronecker indices $\boldsymbol{k} = (0,0,0)$ and the model selection criterion, in our case $BIC_r(0,0,0)$, is calculated for each row, i.e., $r = 1,2,3$. The Kronecker indices are incremented to $\boldsymbol{k} = (1,1,1)$, and again the model selection criterion $BIC_r(1,1,1)$ is calculated for $r = 1,2,3$. The model selection criteria for each row are compared and for the DGP we have considered here, we should find that $BIC_3(0,0,0) < BIC_3(1,1,1)$ but $BIC_r(0,0,0) > BIC_r(1,1,1)$ for $r = 1,2$. The Kronecker index for the third row is now fixed to zero and the others are incremented. Hence the new set of Kronecker indices are $\boldsymbol{k} = (2,2,0)$. The model selection criteria are again compared and we should now find that $BIC_{1,2}(1,1,0) < BIC_{1,2}(2,2,0)$ and therefore our optimal Kronecker indices are $\boldsymbol{k} = (1,1,0)$.

Given that the Echelon form assumes a lower triangular matrix for the contemporaneous relationships in the model ($\boldsymbol{\Phi}_0 = \boldsymbol{\Theta}_0$ are assumed to be lower triangular), the ordering of the variables in the search procedure is important. Depending on the set $\boldsymbol{k}$ of Kronecker indices considered, the contemporaneous relations in the model change. Therefore, in the search procedure each time a different $\boldsymbol{k}$ is considered the design of $\boldsymbol{\Phi}_0$ changes. To avoid the difficulty of generalising this in a programming sense, all contemporaneous relations where included. For example, in the first row where $y_{1,t}$ is the regressand $y_{2,t}$ and $y_{3,t}$ where included in the regressors and so on.

Stage III Efficient parameter estimates of the uniquely identified Echelon form VARMA model with Kronecker indices $\boldsymbol{k}$ are obtained using FIML.

4 Scalar Components vis-à-vis Echelon Form

The following Theorem, due to Tsay (1991), shows the relationship between the two canonical forms we have presented so far.

Theorem 5 *Suppose that $\boldsymbol{y}_t$ is a stable and invertible VARMA process such as 5, represented in canonical Echelon form with Kronecker indices $\boldsymbol{k} = (k_1, \ldots, k_K)'$ and McMillan degree $m = \sum_{r=1}^{K} k_r < \infty$. Now suppose that $\boldsymbol{y}_t$ is also represented in a canonical SCM representation that consists of $K-$SCMs of orders $s_r = (p_r, q_r)$ for $r = 1, \ldots, K$. The set of Kronecker indices $\boldsymbol{k}$ is equivalent to a set of the SCM orders $\boldsymbol{s}^{\max} = (s_1^{\max}, \ldots, s_K^{\max})$ where $s_r^{\max} = \max(p_r, q_r)$ for $r = 1, \ldots, K$.*

Proof. *This theorem is based on theorem 5, page 266 in Tsay (1991).* ■

Example 6 *Consider the VARMA$(1,1)$ process,*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{9}$$

This model is a canonical SCM representation and consists of three SCMs of orders $(1,1)$, $(1,1)$ and $(0,0)$. According to theorem (5) the equivalent Echelon form model has Kronecker indices $\boldsymbol{k} = \boldsymbol{s}^{\max} = (\max(1,1), \max(1,1), \max(0,0))' = (1,1,0)$. Thus, the equivalent canonical reverse Echelon form representation is exactly as in equation (9).

Having presented a situation that the canonical SCM and Echelon form representations are exactly identical we now present an example where this is not the case.

Example 7 *Consider the VARMA process,*

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{10}$$

Again this model is in canonical SCM representation and consists of three SCMs of orders $(1,1)$,

$(1,0)$ *and* $(0,0)$. *Notice now, that for the second SCM, the "autoregressive" order is different than the "moving average" order i.e.,* $p_r \neq q_r$ *for* $r = 2$. *According to theorem (5), the equivalent Echelon form model has Kronecker indices,*

$$\boldsymbol{k} = \boldsymbol{s}^{\max} = (\max(1,1), \max(1,0), \max(0,0))' = (1,1,0).$$

Thus, the equivalent canonical reverse Echelon form representation is,

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

as in equation (9) with $\theta_{21}^{(1)} = -a_{21}\theta_{11}^{(1)}$ *and* $\theta_{22}^{(1)} = -a_{21}\theta_{12}^{(1)}$. *The Echelon form specification does not impose this restriction, whilst the SCM methodology discovers it and includes it. So now, the two representations, although related in some way, are not exactly identical.*

The above example shows that the SCM methodology discovers some additional restrictions compared to the the Echelon form methodology. Since Hannan's Theorem (Hannan and Deistler, 1988) proves that the restrictions in the Echelon form are the necessary and sufficient restrictions for unique identification of the VARMA models, we can conclude that the extra restrictions discovered by the SCM methodology are restrictions that are supported by the data over and above the necessary conditions for identification

Finally, one should note the difference in the specification procedure between the two methodologies. SCMs are identified by observing the correlation structures within the data via sequential canonical correlations testing. The Echelon form is specified via evaluating model selection criteria for a number of chosen models. Although we could reverse these where SCMs could be chosen using model selection criteria, and Echelon forms could be chosen through a sequence of hypothesis tests, both of these procedures have issues that require the practitioner to make objective choices about, as it has already been discussed. In what follows, comparisons between the two modelling procedures are presented. The comparisons are performed both via simulations and via an extensive empirical forecasting exercise.

4.1 A Monte Carlo Evaluation

Athanasopoulos (2005) has conducted Monte Carlo experiments and has evaluated the identification procedures of the two methodologies. The general conclusion from that study is that both procedures perform quite well in identifying some pre-specified VARMA DGPs. In this section we combine those results and we directly compare the performance of the two methodologies. It should be noted that in combining these results some valuable information about the each procedure has been lost. The individual results are available from the authors upon request or alternatively refer to Athanasopoulos (2005). The DGPs considered are presented in Appendix A. The results are presented in Table 1. The sample sizes considered are $N = 100, 150, 200$ and 400 observations. Due to the long, manual and challenging process of identifying SCMs, only 50 iterations where performed for each process and for each sample size. In contrast we managed to automate Poskitt's search procedure for the Echelon form methodology and therefore 1000 iterations where performed for each model and for each sample size.

In comparing these results, extra attention is required as canonical SCMs and Echelon form models are somewhat, but not exactly equivalent, as shown by Theorem 5.

The first two columns under SCM in each panel in Table 1 show the percentage of times the SCM methodology correctly specifies the maximal order (M.O.) and the exact order (E.O.) of the DGP. The two columns in each panel under "Echelon" show these figures for the Echelon form methodology. However, maximal order and exact order are not the same concept in the two model forms. The M.O. $\left(p_{SCM}, q_{SCM}\right)$ in the SCM case is the maximum "autoregressive", $p_{SCM} = \max\left(p_1, \ldots, p_K\right)$, and "moving average", $q_{SCM} = \max\left(q_1, \ldots q_K\right)$, orders of all the scalar components identified. This corresponds to the order of the identified VARMA(p_{SCM}, q_{SCM}) model. In the Echelon form, the maximum order corresponds to the maximum Kronecker index identified, i.e., $\max(k_1, \ldots, k_K)$. This yields a VARMA(p_{ECH}, q_{ECH}) where $p_{ECH} = q_{ECH} = \max(k_1, \ldots, k_K)$. Therefore, if the DGP is a VARMA(p, q) with $p = q$, the maximum orders are exactly equivalent, however if $p \neq q$ they are not equivalent. The SCM methodology attempts to separately identify the p and q orders, but the Echelon form attempts to identify the maximum of p and q, i.e., $\max(p, q)$.

As with the maximum order, the exact order (E.O.) results are not exactly equivalent either. The exact order being specified correctly by the SCM procedures implies that all "autoregressive" and "moving average" components of the model under consideration have been correctly specified. That is, the procedure identified exactly the SCMs specified below each section of the table. In

contrast, the exact order being specified correctly by the Echelon form methodology means that the Kronecker indices, i.e., the maximum row degrees k_r for $r = 1, \ldots, K$, of the model have been correctly identified.

To make these results comparable the third column of each panel under SCM, labeled $\boldsymbol{k}_{SCM}$ shows the percentage of times the scalar component methodology identifies correctly the Kronecker indices of the model. This is then directly comparable with E.O. of the Echelon form. To clarify how this information is extracted from the simulation results, we present the following example.

Example 8 *Consider the processes of equations (18),*

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

and (17)

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

For the first model the scalar component methodology attempts to identify three scalar components of orders SCM$(1,1)$*, SCM*$(1,0)$ *and SCM*$(1,0)$*. The percentage of times the Kronecker indices are correctly identified by the scalar component procedure is set by the minimum between the percentage of times the maximum order is correctly identified and the percentage of times the procedure identifies no SCM*$(0,0)$*. For example, for* $N = 200$*, the maximum order has been correctly identified* 98 *percent of the time, i.e., the upper bound for identifying the correct Kronecker indices by the scalar component methodology is set to* 98 *percent. Moreover the SCM process has identified zero SCM*$(0,0)$ 100 *percent of the time (these figures are extracted from Table 3.11 in Athanasopoulos 2005). This means that the scalar component methodology identifies the exact Kronecker indices* $\boldsymbol{k}_{SCM} = 98$ *percent of the time. For the model of equation (17), the SCMs are of orders SCM*$(1,1)$*, SCM*$(1,1)$ *and SCM*$(0,0)$*. Looking again at the case of* $N = 200$*, the upper bound for correct identification of the Kronecker indices is set by the maximum order to* 92 *percent. The other bound is* 94 *percent which is the number of times the process identified one SCM*$(0,0)$ *(These figures are extracted from Table 3.13 in Athanasopoulos 2005). Therefore, the Kronecker indices have been identified correctly by the scalar component*

methodology $\boldsymbol{k}_{SCM} = 92$ *percent of the time.*

The results of Table 1, show that both methodologies perform quite well in identifying both the maximum order and the exact order of the Kronecker indices. For sample sizes of 200 or more, for all DGPs (with only a single exception) both methodologies discover the correct Kronecker indices more than 90 percent of the time. The only exception is for the DGP of equation (16) where the success rate is 83 percent of the Echelon form methodology.

Table 1: *Monte Carlo simulation results for SCM versus Echelon form*

PANEL A: *DGP of equation* (12)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	-	-	-	-	-
150	-	-	-	-	-
200	100	96	100	100	100
400	-	-	-	-	-

SCMs - (1,0)(1,0)(1,0)

PANEL B: *DGP of equation* (13)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	96	36	84	88	47
150	96	40	92	90	82
200	94	50	94	90	90
400	98	88	98	90	90

SCMs - (0,1)(0,1)(0,1)

PANEL C: *DGP of equation* (14)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	94	54	90	100	64
150	92	72	92	100	94
200	94	88	94	100	100
400	94	90	94	100	100

SCMs - (1,1)(0,0)(0,0)

PANEL D: *DGP of equation* (15)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	88	52	88	97	49
150	94	78	94	99	82
200	96	94	96	100	95
400	100	86	100	100	100

SCMs - (1,1)(1,0)(0,0)

PANEL E: *DGP of equation* (16)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	68	12	68	94	23
150	76	8	76	95	56
200	92	22	92	96	83
400	96	52	96	92	96

SCMs - (1,1)(0,1)(0,0)

PANEL F: *DGP of equation* (17)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	88	10	88	95	94
150	94	44	94	97	97
200	92	48	92	98	98
400	94	72	94	99	99

SCMs - (1,1)(1,1)(0,0)

PANEL G: *DGP of equation* (18)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	96	10	96	93	88
150	92	18	92	94	94
200	98	20	98	97	97
400	94	62	94	97	97

SCMs - (1,1)(1,0)(1,0)

PANEL H: *DGP of equation* (19)

N	**SCM**			**Echelon**	
	M.O.	E.O.	$\boldsymbol{k}_{SCM}$	M.O.	E.O.
100	80	2	80	86	86
150	94	2	94	91	91
200	96	-	96	93	93
400	98	2	98	97	97

SCMs - (1,1)(1,1)(1,1)

5 Empirical Results

5.1 Data

The data we employ in this paper are 40 monthly macroeconomic time series from March 1959 to December 1998 (i.e., $N = 480$ observations). These are extracted from the Stock and Watson (1999) data set (see Appendix B). 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 are transformed in various ways as indicated in Appendix B. These transformations are exactly the same as those in Stock and Watson (1999) and Watson (2001). We have selected seventy trivariate systems which include at least one combination from each of the eight categories. For example, at least one system from categories $(i),(ii)$ and (iii), one system from $(i),(ii)$ and (iv) and so on. For each of the seventy data sets we identify and estimate VARMA models both via the SCM methodology and the Echelon form methodology.

5.2 Forecast Evaluation Method

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 using the estimation sample, i.e., all models are estimated using $\boldsymbol{y}_1$ to $\boldsymbol{y}_{N_1}$. We then use each estimated model to produce a sequence of h-step-ahead forecasts for $h = 1$ to 15. That is, with $\boldsymbol{y}_{N_1}$ as the forecast origin, we produce forecasts for $\boldsymbol{y}_{N_1+1}$ to $\boldsymbol{y}_{N_1+15}$. The forecast origin is then rolled forward one period, i.e., using observation $\boldsymbol{y}_{N_1+1}$, we produce forecasts for $\boldsymbol{y}_{N_1+2}$ to $\boldsymbol{y}_{N_1+16}$. 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.

For each forecast horizon h, we consider two measures of forecasting accuracy. The first is the determinant of the mean squared forecast error matrix, $|MSFE|$, and the second is the trace of the mean squared forecast error matrix, $tr(MSFE)$. 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 $tr(MSFE)$ is not invariant to either. In this forecast evaluation exercise, both of these measures are informative in their own right, as no elementary operations take place. The only apparent drawback would be with the $tr(MSFE)$, as the rankings of the models using this measure would be affected by the different scales across the variables of the system. Therefore, we have standardized all variables by their estimated standard deviation that is derived from the estimation sample, making the variances of the forecast errors of the three series directly comparable. This makes the $tr(MSFE)$ a useful measure of forecast accuracy.

In order to evaluate the overall forecasting performance of the models over the seventy data sets, we calculate two measures. Firstly, we calculate the percentage better (PB) measure which has been used in forecasting competitions (see Makridakis and Hibon, 2000). This measure is the percentage of times each model performs best in a set of competing models.

The second measure we compute is the average (over the seventy data sets) of the ratios of the forecast accuracy measures for each model, relative to the VARMA. The reason that we compute these ratios, as well as the PB counts, is that it is possible that one class of models is best more than 50 percent of the time, say 80 percent, but that in all those cases other alternatives are close to it. However, in the 20 percent of cases that this model is not the best, it may make huge forecast errors. In such a case, a user who is risk averse would not use this model, as the preferred option would be a less risky alternative. The average of the relative ratios provides us with this additional information.

The relative ratios considered are the average of the relative ratios of the determinants of the mean squared forecast error matrices defined as

$$\overline{RdMSFE_h} = \frac{1}{M}\sum_{i=1}^{M}\frac{\left|MSFE(\text{VARMA}_{Echelon})_i\right|}{\left|MSFE\left(\text{VARMA}_{SCM}\right)_i\right|},$$

and the average of the relative ratios of the traces of the mean squared forecast error matrices defined as

$$\overline{RtMSFE_h} = \frac{1}{M}\sum_{i=1}^{M}\frac{tr\left(MSFE\left(\text{VARMA}_{Echelon}\right)_i\right)}{tr\left(MSFE\left(\text{VARMA}_{SCM}\right)_i\right)},$$

where h is the forecast horizon, and M is the number of data sets considered.

5.3 PB Results

The PB counts have been plotted in Figures 1 and 2 (the actual counts for all measures are presented in Appendix C). In these figures there are two lines, one representing the VARMA models specified by the scalar component methodology (labeled SCM) and the other representing the VARMA models specified by the Echelon form methodology (labeled Echelon form). The marked points on each line depict the percentage of times for which that class of models produces the best forecast for that horizon between the two classes of models. For example, consider the 7 -step-ahead forecast performance. Figure 1 shows that the scalar component VARMA models outperform the Echelon form models as approximately 70 percent of the time they produce lower values of $|MSFE|$. In general both Figures 1 and 2 show that the scalar component VARMA models produce the highest PB counts for all h=1 to 15-step-ahead forecast horizons for both $|MSFE|$ and $tr(MSFE)$.

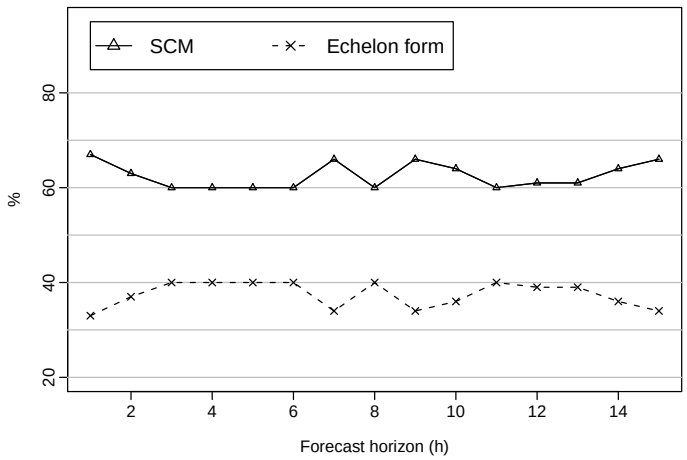

Figure 1: *PB counts for $|MSFE|$ for canonical SCM VARMA models versus canonical Echelon form VARMA models*

5.4 Relative Ratios Results

The results for the relative ratios have been tabulated, in Table 2. A first look at both panels indicates that for all forecast horizons, for both the determinant and the trace of the $MSFE$, the relative ratio measures are constantly greater than one. A relative ratio grater than one shows that for that forecast horizon, the scalar component VARMA models perform better than the Echelon form VARMA as the base model is the SCM, i.e., the Echelon forms produce a larger error than the SCMs. For example, in Panel A of Table 2, for $h = 4 - steps$ ahead forecast horizon, the SCMs improve on the forecasting error of the Echelon form models by 3.5 percent. This means that on average, the SCMs produce approximately a 3 percent lower $|MSFE|$ in comparison to

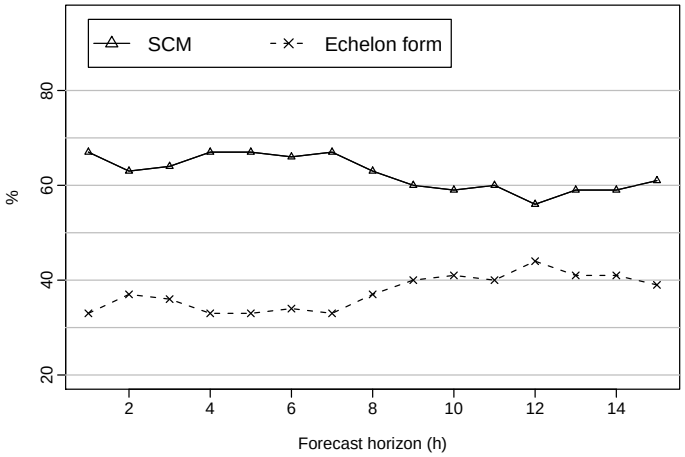

Figure 2: *PB counts for tr$(MSFE)$ for canonical SCM VARMA models versus canonical Echelon form VARMA models*

the Echelon forms.

Table 2: *Average relative ratios for the determinant and the trace of the MSFE matrices for Echelon form VARMA models over SCM VARMA models*

Panel A: $\overline{RdMSFE}$ of Echelon form over SCMs

	Forecast Horizon (h)							Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
Echelon	1.061	1.031	1.030	1.031	1.033	1.034	1.035	1.038	1.036	1.036	1.035

Panel B: $\overline{RtMSFE}$ of Echelon form over *SCM*s

	Forecast Horizon (h)							Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
Echelon	1.027	1.025	1.027	1.024	1.014	1.01	1.011	1.026	1.022	1.018	1.017

In Section 4 we conclude that a major difference between these two specifications is that the SCM methodology potentially identifies restrictions over and above the necessary and sufficient restrictions of the Echelon form. This can make SCMs more parsimonious than Echelon forms which could be an advantage when it comes to out-of-sample forecasting. This could as well have been the reason behind the superior performance of the SCMs in the forecast evaluation exercise. In fact the Echelon form methodology presented by its various advocates (see for example Lütkepohl and Poskitt 1996) includes a 4th step which involves the elimination of any insignificant coefficients from the model via t-tests or χ^2-tests to obtain optimal parsimony on the model.

We do not consider any further reduction of models here because each stage of such reductions would require a FIML estimation and would be very time-intensive in such an extensive forecasting exercise. The study of other reduction strategies that are more compatible with the procedure of identification of Kronecker indices and are more amenable to automation, is the subject of our

current research.

6 Conclusion and Direction for Future Research

This paper provides an in depth comparison of canonical VARMA models specified by scalar components to VARMA models specified by the Echelon form methodology. We perform this comparison on a theoretical, experimental and empirical level. At the theoretical level we show that the canonical Echelon form is somewhat equivalent to the canonical SCM representation, with the latter depicting greater flexibility as the maximum "autoregressive" order of the VARMA model does not have to be the same as the order of the "moving average" component. These orders have to be the same when specifying models via Kronecker indices in the canonical Echelon form. At the experimental level, we show via Monte-Carlo experiments that both of these procedures work very well in identifying the Kronecker indices of VARMA models.

Finally at the empirical level, the out-of-sample forecast evaluation between the two forms shows that the SCMs outperformed the Echelon form models. In the discussion of these forecast results we have acknowledged that our experimental design may have favored the scalar component models as there is a sense in which the Echelon form models are over-parameterised and therefore need to be further refined. It is of interest to note that our results are consistent with the principle of parsimony which favours models with fewer parameters as they tend to forecast more accurately than over-parameterised representations. This highlights the need for further research on refining the Echelon form VARMA models. During this research we have found that the advantage of the Echelon form identification process is its simplicity in application, as we have managed to fully automate this process. This is impossible to do with the scalar component identification process which we have managed to partly automate but still needs some judgement from its user. Therefore if we could find refinement processes for the Echelon form models that we are able to automate, it could lead to bringing VARMA models to the applied econometrician it as has happened with automatic univariate ARIMA modelling (see Mélard and Pasteels 2000, and Gómez and Maravall 2001) and multivariate VAR modelling. Thus, a study examining alternative methods for refining the Echelon form and the effects of the refinement on the forecasting performance of VARMA models would be of great interest.

References

Akaike, H. (1974) A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **19**, 667–674.

Akaike, H. (1976) *"Canonical correlation analysis of time series and the use of information criterion"* in *System Identification*, Mehar, R. and D. Lainiotis Eds., Academic Press, New York, 27–96.

Athanasopoulos, G. (2005) Essays on alternative methods of identification and estimation of vector autoregressive moving average models, *unpublished Ph.D. dissertation,* Monash University, Department of Econometrics and Business Statistics.

Athanasopoulos, G. and F. Vahid (2006) A complete *VARMA* modelling methodology based on scalar components, *Working Paper*, Monash University, Department of Econometrics and Business Statistics.

Athanasopoulos, G. and F. Vahid (2007) VARMA versus VAR for macroeconomic forecasting, *Journal of Business and Economic Statistics*, **forthcoming**.

Box, G. and G. Jenkins (1970) *Time Series Analysis: Forecasting and Control*, San Francisco, California, Holden Day.

Clements, M. P. and D. F. Hendry (1993) On the limitations of comparing mean squared forecast errors (with discussions), *Journal of Forecasting*, **12**, 617–637.

Deistler, M. and E. J. Hannan (1981) Some properties of the parametrisation of ARMA systems with unknown order, *Journal of Multivariate Analysis*, **11**, 474–484.

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.

Gómez, V. and A. Maravall (2001) *"Automatic Modeling Methods for Univariate Series"* in *A Course in Time Series Analysis*, Peña, D., G. C. Tiao, and R. S. Tsay, John Wiley and Sons, New York.

Hannan, E. (1970) *Multiple time series*, John Wiley & Sons, New York.

Hannan, E. and L. Kavalieris (1984) Multivariate linear time series models, *Advances in Applied*

Probability, **16**, 492–561.

Hannan, E. J. (1969) The identification of vector mixed autoregressive-moving average systems, *Biometrica*, **56**, 223–225.

Hannan, E. J. (1976) The identification and parametrisation of ARMAX and state space forms, *Econometrica*, **44**, 713–723.

Hannan, E. J. (1979) *"The statistical theory of linear systems"* in *Developments in Statistics*, Krishnaiah, R. P. Ed., Academic Press, New York, 83–121.

Hannan, E. J. and M. Deistler (1988) *The Statistical Theory of Linear Systems*, John Wiley & Sons, New York.

Hannan, E. J. and J. Rissanen (1982) Recursive estimation of autoregressive-moving average order, *Biometrica*, **69**, 81–94.

Kailath, T. (1980) *Linear Systems*, Prentice Hall, New Jersey.

Lütkepohl, H. (1993) *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin-Heidelberg, 2nd ed.

Lutkepohl, H. and H. Claessen (1997) Analysis of cointegrated VARMA processes, *Journal of Econometrics*, **80**, 223–239.

Lütkepohl, H. and D. Poskitt (1996) Specification of Echelon-form *VARMA* models, *Journal of Business and Economic Statistics*, **14**, 69–79.

Makridakis, S. and M. Hibon (2000) The M3-competition: Results, conclusions and implications, *International Journal of Forecasting*, **16**, 451–476.

Mélard, G. and J.-M. Pasteels (2000) Automatic ARIMA modeling including interventions, using time series expert software, *International Journal of Forecasting*, **16**, 497–508.

Nsiri, S. and R. Roy (1992) On the identification of ARMA Echelon-form models, *Canadian Journal of Statistics*, **56**, 207–231.

Nsiri, S. and R. Roy (1996) Identification of refined ARMA Echelon form models for multivariate time series, *Journal of Multivariate Analysis*, **56**, 207–231.

Poskitt, D. S. (1992) Identification of Echelon canonical forms for vector linear processes using least squares, *The Annals of Statistics*, **20**, 195–215.

Quenouille, M. H. (1957) *The Analysis of Multiple Time Series*, London: Charles Griffin & Company.

Reinsel, G. (1997) *Elements of Multivariate Time Series*, New York: Springer-Verlag, 2nd ed.

Solo, V. (1986) *"Topics in Advanced Time Series Analysis"* in *Lectures in Probability and Statistics*, Del Pino, G. abd R. Rebolledo Eds., Springer-Verlag, New York.

Stock, J. and M. W. Watson (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, A Festschrift in Honour of Clive W. J. Granger, New York: Oxford University Press, pp. 1-44.

Tiao, G. and R. Tsay (1989) Model specification in multivariate time series (with discussions), *Journal of the Royal Statistical Society B*, **51**, 157–213.

Tiao, G. C. (2001) *"Vector ARMA"* in *A Course in Time Series Analysis*, Peña, D., G. C. Tiao, and R. S. Tsay, John Wiley and Sons, New York.

Tiao, G. C. and G. E. P. Box (1981) Modelling multiple time series with applications, *Journal of the American Statistical Association*, **76**, 802–816.

Tsay, R. S. (1989) Parsimonious parametrisation of vector autoregressive moving average models, *Journal of Business and Economic Statistics*, **7**, 327–341.

Tsay, R. S. (1991) Two canonical forms for Vector ARMA processes, *Statistica Sinica*, **1**, 247–269.

Tunnicliffe-Wilson, G. (1973) The estimation of parameters in multivariate time series models, *Journal of the Royal Statistical Society B*, **35**, 76–85.

Watson, M. W. (2001) Macroeconomic forecasting using many predictors, in *Advances in Economics and Econometrics, Theory and Applications*, eds. M. Dewatripont, L. Hansen and S. Turnovsky, Eighth World Congress of the Econometric Society, III, 87-115.

A Data Generating Processes considered in Section 4.1

$$\boldsymbol{y}_t = \begin{bmatrix} 0.5 & -0.6 & 0.7 \\ 0.6 & 0.7 & -0.4 \\ 0.3 & 0.6 & 0.4 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{12}$$

$$\boldsymbol{y}_t = \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & -0.6 & 0.7 \\ 0.6 & 0.7 & -0.4 \\ 0.3 & 0.6 & 0.4 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{13}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 1 & 0 \\ -0.6 & 0 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.7 & 0.6 & 0.4 \\ 0. & 0. & 0. \\ 0. & 0. & 0. \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} -0.7 & 0. & 0. \\ 0. & 0. & 0. \\ 0. & 0. & 0. \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{14}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 1 & 0 \\ 0.4 & 0.7 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}, \tag{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 1 & 0 \\ 0.4 & 0.7 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}, \tag{16}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5 & -0.7 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.7 & -0.5 & 0.7 \\ 0.6 & 0.3 & 0.6 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.5 & -0.6 & 0 \\ 0.6 & 0.7 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}. \tag{17}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 1 & 0 \\ 0 & -0.6 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.7 & -0.6 & 0.4 \\ 0.6 & -0.5 & -0.4 \\ 0.3 & -0.6 & 0.4 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.7 & 0.4 & -0.6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}, \tag{18}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0. & 1 & 0 \\ 0. & 0. & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.6 & -0.7 & 0.4 \\ 0.7 & 0.5 & -0.4 \\ 0.3 & -0.7 & 0.4 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t - \begin{bmatrix} 0.7 & -0.3 & 0.4 \\ 0.2 & 0.6 & 0.5 \\ -0.3 & 0.4 & 0.4 \end{bmatrix} \boldsymbol{\varepsilon}_{t-1}, \qquad (19)$$

B Data Summary

This appendix lists the time series that are used in this paper. The series have been directly downloaded from Mark Watson's web page (http://www.wws.princeton.edu/mwatson/). The names (mnemonics) given to each series and the brief description following each series name have been reproduced from Watson (2001). The superscript index on the series name is the transformation code which corresponds to: (1) the level of the series, (2) the first difference $\left(\Delta y_t = y_t - y_{t-1}\right)$ and (3) the first difference of the logarithm, i.e., series transformed to growth rates $\left(100 * \Delta \ln y_t\right)$. The following abbreviations also 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 SA)
13. MSMQ[3] Manufacturing & trade: manufacturing, total (mil of chained $1992 SA)
14. MSDQ[3] Manufacturing & trade: manufacturing, durable goods (mil of chained $92 SA)
15. MSNQ[3] Manufacturing & trade: manufacturing, nondurable goods (mil of chd. $92 SA)
16. WTQ[3] Merchant wholesalers: total (mil of chained $1992 SA)
17. WTDQ[3] Merchant wholesalers: durable goods total (mil of chained $1992 SA)
18. WTNQ[3] Merchant wholesalers: nondurable goods total (mil of chained $1992 SA)
19. RTQ[3] Retail trade: total (mil of chained $1992 SA)
20. RTNQ[3] Retail trade: nondurable goods (mil of chained $1992 SA)
21. CMCQ[3] Personal consumption expend - total (bil of chained $1992 SAAR)

(iv) *Real inventories and inventory-sales ratios*

22. IVMFGQ[3] Inventories, business, manufacturing (mil of chained $1992 SA)
23. IVMFDQ[3] Inventories, business durables (mil of chained $1992 SA)
24. IVMFNQ[3] Inventories, business nondurables (mil of chained $1992 SA)
25. IVSRQ[2] Ratio for manufacturing & trade: inventory/sales (chained $1992 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 BCI)
30. MDOQ[3] New orders, durable goods industries ($1992 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 BCI)

33. FCLNQ[3] Commercial & industrial loans outstanding in ($1992 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)

C Percentage Better Counts (Raw Data)

Table 3: *PB Counts for of VARMA models specified by the* SCM *methodology versus* VARMA *models specified by the Echelon form methodology*

PANEL A: Counts for the $|MSFE|$ for *SCM* versus Echelon form

	Forecast Horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
SCM	67	63	60	60	60	60	66	60	66	64	60	61	61	64	66
Echelon	33[a]	37	40	40	40	40	34	40	34	36	40	39	39	36	34

PANEL B: Counts for the $tr(MSFE)$ for *SCM* versus Echelon form

	Forecast Horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
SCM	67	63	64	67	67	66	67	63	60	59	60	56	59	59	61
Echelon	33[a]	37	36	33	33	34	33	37	40	41	40	44	41	41	39

[a]Figures are rounded to the nearest integer