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

First draft July 2007

Revised May 2009

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

18 May 2009

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 performances of the two methodologies in identifying some prespecified 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 components, 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) 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). 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), 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.

> "*While VARMA models involve additional estimation and identification issues, these complications do not justify systematically ignoring these moving average components, as in the SVAR approach.*"
>
> Cooley and Dwyer (1998)

The complexities of identifying and estimating unique VARMA models, and, in sharp contrast, the ease of specifying and estimating vector autoregressions (VARs) have resulted in VARs, dominating the macroeconomic literature, despite ubiquitous warnings about their many practical and theoretical shortcomings. For example, in contrast to VARMA models, VARs are not invariant to aggregation, marginalisation or measurement error. Hence, to avoid misspecification, any modelling of macroeconomic aggregates (such as gross domestic product, industrial production, etc.) should include moving average dynamics, even if the components of these aggregates are assumed to follow finite autoregressive processes. Furthermore, even if we assume a finite order VAR representation for a set of macroeconomic aggregates, modelling a subset of these should again include moving average dynamics (see for example Zellner and Palm, 1974; Fry and Pagan, 2005). Ravenna (2007) warns that caution should be used by researchers using finite order VARs to build dynamic stochastic general equilibrium (DSGE) models, and Fernández-Villaverde et al. (2005) show that linearised versions of DSGE models generally imply a finite order VARMA structure.

The first methodology we consider that returns unique VARMA representations is the Athanasopoulos and Vahid (2008a) extension of 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) (Durbin, 1963). We present the scalar component methodology in Section 2.

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. 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), Poskitt (1992) and Lütkepohl and Poskitt (1996), among others. We present the Echelon form methodology in Section 3.

> "*We see that dealing with VARMA models in Echelon form is not as easy as dealing with univariate ARMA models .... This might be a reason why practitioners are reluctant to employ VARMA models. Who could blame them for sticking with VAR models when they probably need to refer to a textbook to simply write down an identified VARMA representation?*"
>
> Dufour and Pelletier (2008)

Specifying a unique Echelon form VARMA representation involves applying a set of mathematical rules. The advocates of the Echelon form portray this as its major advantage. However, the complexities of the formulae and the apparent lack of intuition behind these formulae have earned this methodology the reputation of being a very complicated method, and have not helped to promote the application of VARMA models in the empirical literature. In Section 4 we theoretically connect the Echelon form to SCMs. This connection provides an intuition behind the complicated Echelon form formulae and shows that understanding the Athanasopoulos and Vahid (2008a) scalar methodology demystifies the Echelon form and eliminates the need for a textbook.

Although many studies have contributed to the Echelon form methodology, no investigation has been undertaken into the finite sample performance of this methodology when attempting to identify VARMA models. In Section 4.1 we conduct Monte-Carlo experiments, and evaluate the ability of both the Echelon form and the SCM methodology to identify some pre-specified VARMA data generating processes (DGPs).

Using real data, Athanasopoulos and Vahid (2008b) conclude that VARMA models specified by the scalar component methodology forecast macroeconomic variables more accurately than VARs. In Section 5 we compile 70 trivariate data sets and perform a similar forecasting exercise. We evaluate the forecasting performance of VARMA models specified by the SCMs versus VARMA models specified by the Echelon form methodology and VAR models with lag lengths chosen by AIC and BIC.

2 A VARMA modelling methodology based on scalar components

The scalar component methodology we employ is the Athanasopoulos and Vahid (2008a) extension of 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.

Stage I: Identification of the scalar components

The aim of identifying scalar components is to examine whether there are any simplifying embedded structures underlying a VARMA(p,q) process. These simple structures (or scalar components) are linear combinations of variables that depend on fewer than p autoregressive lags and fewer than q lags of innovations. Formally, for a given $K-$dimensional VARMA(p,q) process

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

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:

$$\begin{aligned}
\boldsymbol{\alpha}' \boldsymbol{\Phi}_{p_1} &\neq \mathbf{0} \text{ where } 0 \leq p_1 \leq p; \\
\boldsymbol{\alpha}' \boldsymbol{\Phi}_{l} &= \mathbf{0} \text{ for } l = p_1+1, \ldots, p; \\
\boldsymbol{\alpha}' \boldsymbol{\Theta}_{q_1} &\neq \mathbf{0} \text{ where } 0 \leq q_1 \leq q; \\
\boldsymbol{\alpha}' \boldsymbol{\Theta}_{l} &= \mathbf{0} \text{ for } l = q_1+1, \ldots, q.
\end{aligned}$$

The SCM methodology uses a sequence of canonical correlations tests until it discovers K such linear combinations, starting from the most parsimonious SCM$(0,0)$. Denoting 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

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

where $\boldsymbol{\Phi}_i^* = \boldsymbol{A}\boldsymbol{\Phi}_i$, $\boldsymbol{\varepsilon}_t^* = \boldsymbol{A}\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\Theta}_i^* = \boldsymbol{A}\boldsymbol{\Theta}_i\boldsymbol{A}^{-1}$, in which the right hand side coefficient matrices may have many rows of zeros. However, if there are scalar components SCM(p_r, q_r) and SCM(p_s, q_s) which are strongly nested, i.e. when $p_r > p_s$ and $q_r > q_s$, then even if we know $\boldsymbol{A}$, the system will not be identified. This is because SCM(p_s, q_s) implies an exact linear relationship between the lagged variables on the right hand side of SCM(p_r, q_r). 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. We set the moving average parameters to zero in such situations. This is often referred to as the "general rule of elimination".

Stage II: Placing identification Restrictions on Matrix $\boldsymbol{A}$

Not all parameters in $\boldsymbol{A}$ are free parameters. We can multiply each row of $\boldsymbol{A}$ by a constant without changing the structure of the system. We can also linearly combine an SCM with any other SCM with weakly smaller p and q and not change its order. These simple implications of the definition of scalar components leads to the following identification rules that, as Athanasopoulos and Vahid (2008a) show, lead to a uniquely identified $\boldsymbol{A}$. We refer to this system as a canonical SCM representation. These rules are:

1. Normalize one parameter in each row of $\boldsymbol{A}$ to one. Athanasopoulos and Vahid (2008a) suggest a procedure to safeguard against the possibility of normalising on a zero parameter; we do not repeat it here to save space.

2. In all cases where there are two embedded scalar components with weakly nested orders, i.e., $p_1 \geq p_2$ and $q_1 \geq q_2$, if the parameter in the ith column of the row of $\boldsymbol{A}$ corresponding to the SCM(p_2, q_2) 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.

Stage III: Estimation of the Uniquely Identified System

Estimate the parameters of the system using FIML. 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

$$\boldsymbol{\Psi}(L)\boldsymbol{y}_t = \boldsymbol{\Xi}(L)\varepsilon_t, \tag{5}$$

where $\boldsymbol{\Psi}(L) = \boldsymbol{\Psi}_0 - \boldsymbol{\Psi}_1 L - \ldots - \boldsymbol{\Psi}_p L^p$ and $\boldsymbol{\Xi}(L) = \boldsymbol{\Xi}_0 - \boldsymbol{\Xi}_1 L - \ldots - \boldsymbol{\Xi}_q L^q$ is said to be in reverse Echelon form (Lütkepohl and Claessen, 1997) if the pair of polynomials in the lag operators $\boldsymbol{\Psi}(L) = \left[\psi_{rc}(L)\right]_{r,c=1,\ldots,K}$ and $\boldsymbol{\Xi}(L) = \left[\xi_{rc}(L)\right]_{r,c=1,\ldots,K}$, $[\boldsymbol{\Psi}(L) : \boldsymbol{\Xi}(L)]$, are left coprime and possess the following properties:

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

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

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

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

   where $\psi_{rc}^{(j)}$ specifies the element of $\boldsymbol{\Psi}_j$ in row r and column c, and $\xi_{rc}^{(j)}$ specifies the element of $\boldsymbol{\Xi}_j$ in row r and column c.

The maximum row degrees $\boldsymbol{k} = (k_1, \ldots, k_K)'$ are called the **Kronecker Indices** and 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 $\psi_{rc}(L)$ 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 $d(\boldsymbol{k}) = 2mK$.

The theory and examples of the Echelon form representation of VARMA models are given by Solo (1986), Hannan and Kavalieris (1984), Hannan and Deistler (1988) and Tsay (1989) and Lütkepohl

(1993), among others. The "reverse Echelon form" defined above is a variant of the Echelon form in which whenever identification can be achieved by placing a zero restriction on either an autoregressive parameter or on a moving average parameter, the moving average parameter is set to zero. Note that just as in the SCM representation, a reverse Echelon form is a rotation of the VARMA model, here by the matrix Ψ_0 that turns cross equation restrictions into zero restrictions and makes the system identifiable.

The complicated looking definitions of the canonical Echelon form and reverse Echelon form have baffled practitioners and led to comments such as the one quoted above from Dufour and Pelletier (2008). However, as we show below, by understanding the relationship between Kronecker indices and orders of scalar components, one can see that the above definition is nothing but the symbolic representation of identification rules in a special sub-class of scalar component models. To understand that, we need to explain the relationship between Kronecker indices and Kronecker invariants.

An Echelon form or a reverse Echelon form representation is not invariant with respect to an arbitrary reordering of the Kronecker indices. A reordering of the Kronecker indices may change the structure of the left hand side matrix, which contains the contemporaneous relationships. However, the variables in $\boldsymbol{y}_t$ can be permuted such that the Kronecker indices are arranged in descending order (see Poskitt, 2005).

Definition 1 *When the Kronecker indices of $\boldsymbol{y}_t$ are such that $k_1 \geq k_2 \ldots \geq k_K$, these are referred to as Kronecker invariants.*

When a VARMA system is expressed in terms of Kronecker invariants it not only has a canonical form, but it also has a unique representation for each row of the system; i.e., even if further order preserving permutations are possible (by changing the order of two indices that are equal to each other), the structure of the system will not change.

Example 2 *Consider a trivariate stable and invertible VARMA process with Kronecker invariants $\boldsymbol{k} = (k_1, k_2, k_3)' = (1, 1, 0)'$. The total number of freely varying parameters is $d(\boldsymbol{k}) = 2mK = 2\times2\times3 = 12$. The reverse Echelon form representation of the process is*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \psi_{31}^{(0)} & \psi_{32}^{(0)} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \boldsymbol{\Xi}_0 \varepsilon_t - \begin{bmatrix} \xi_{11}^{(1)} & \xi_{12}^{(1)} & 0 \\ \xi_{21}^{(1)} & \xi_{22}^{(1)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}. \quad (6)$$

It is obvious from the example that if we change the order of the first two variables, Kronecker invariants will not change and the structure of the system (i.e. the position of zeros and ones in the system) remains unchanged. Poskitt's (1992) search process is a simple and efficient procedure for the practical specification of Echelon form VARMA models, and is based on searching for Kronecker invariants. We use Poskitt's procedure in the empirical section of this paper. A brief summary of this

procedure is as follows.

Stage I: Obtaining approximate residuals

A long order VAR(h) is fitted and the estimated residuals $\widehat{\varepsilon}_t(h)$ 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: Searching for Kronecker invariants

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

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

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

We employ Poskitt's (1992) search procedure coupled with the BIC as the model selection criterion. From extensive Monte-Carlo experiments we have concluded that 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.[1]

Poskitt's (1992) search 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. If we consider Kronecker invariants, this means that the structure of each equation depends on the structure of equations in the block with the same Kronecker index and other equations below that block.

Using this property, the search starts with all Kronecker invariants being set to zero. We compute the BIC for each equation of the model; i.e., we compute $\text{BIC}_r(k_r)\ \forall\ k_r = 0$, and compare this to $\text{BIC}_r(k_r)\ \forall\ k_r = 1$, for $r = 1, \ldots, K$. For any $\text{BIC}_r(0) \leq \text{BIC}_r(1)$ we fix $k_r = 0$. All other invariants are incremented, and we fix $k_r = 1$ for any $\text{BIC}_r(1) \leq \text{BIC}_r(2)$. This process is repeated until all Kronecker invariants are fixed.

[1] These Monte-Carlo simulation results come from the unpublished PhD dissertation of Athanasopoulos (2007) and are available upon request.

Stage III: Estimation of the identified system

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

4 Scalar Components vis-à-vis Echelon Form

Tsay (1991) explored the scalar components implications of an Echelon form with a set of Kronecker indices. However, at that time it was not possible to establish a direct correspondence between the Echelon form and the existing scalar component methodologies, because the scalar component methodology of Tiao and Tsay (1989) did not specify the structure of the left hand side parameter matrix $\boldsymbol{A}$. However, the scalar component methodology of Athanasopoulos and Vahid (2008a) that we describe above specifies a complete structure, and in the following theorem we establish the relationship between a VARMA model identified using the order of its embedded scalar components and a VARMA model in Echelon form identified via its Kronecker invariants.

Theorem 3 *Suppose that $\boldsymbol{y}_t$ is a stable and invertible VARMA process represented in reverse canonical Echelon form with Kronecker invariants $\boldsymbol{k} = (k_1, \ldots, k_K)'$, where $k_1 \geq \ldots \geq k_K$ and the McMillan degree is $m = \sum_{r=1}^{K} k_r < \infty$. Now suppose that $\boldsymbol{y}_t$ is also represented in a canonical SCM form that consists of $K-$SCMs of orders $s_r = (p_r, q_r)$ for $r = 1, \ldots, K$. The set of Kronecker invariants $\boldsymbol{k}$ is equivalent to a set of SCM orders $\boldsymbol{s}^{\max} = (s_1^{\max}, \ldots, s_K^{\max})'$, where $s_r^{\max} = \max(p_r, q_r)$ for $r = 1, \ldots, K$. Furthermore, if $p_r = q_r \ \forall \ r = 1, \ldots, K$ then the reverse canonical Echelon form and the canonical SCM form are identical if the same permutation of variables with equal indices are chosen, and after innovations are rewritten in the same way.*

Proof. *The first part of the theorem is the same as Theorem 5 of Tsay (1991). Here we show that if $p_r = q_r \ \forall \ r = 1, \ldots, K$ then the reverse canonical Echelon form and the canonical SCM form are equivalent. Since Kronecker invariants are in descending order, the reverse Echelon form rules imply a VARMA(k_1, k_1) model in which the Ψ_{k_1-j} and Ξ_{k_1-j} matrices have rows of zeros in all rows with Kronecker invariants k_r such that $k_1 - k_r > j$ for $j = 0, \ldots, k_1 - 1$. Since Kronecker invariants are in descending order, these rows of zeros are the bottom rows of these matrices. In addition, in any row of the moving average matrices where a zero appears at position c, all elements of that row to the right of c will be zero. Finally, the Ψ_0 and Ξ_0 matrices are lower triangular with unit diagonals, are equal to each other, and have identity submatrices that start from position $\psi_{rr}^{(0)}$ and end at position $\psi_{ss}^{(0)}$ whenever $k_r = k_{r+1} = \cdots = k_s$. In the SCM representation, when $p_r = q_r \ \forall \ r = 1, \ldots, K$ and we arrange these components in descending order, the first part of this theorem ensures that $p_r = k_r$. The definition of scalar components of order $p_r = q_r \ \forall \ r = 1, \ldots, K$ and the "general rule of elimination" described above imply an SCM representation with autoregressive and moving average parameter matrices with zeros in exactly the same positions as those for the reverse Echelon form described above. Also, since the SCMs*

are arranged in descending order, the identification rules in Section 2 imply that the $\boldsymbol{A}$ *matrix is lower triangular, with identity blocks as described above whenever the SCMs are of the exact same order. This means that the matrices* $[\Psi_0, \Psi_1, \ldots, \Psi_p, \Xi_1, \ldots, \Xi_q]$ *in (5) and* $[\boldsymbol{A}, \Phi_1^*, \ldots, \Psi_p^*, \Theta_1^*, \ldots, \Theta_q^*]$ *have the exact same structure. The only difference there can be between the reverse Echelon form implied by the Kronecker invariants and the structure implied by scalar components is that the order of variables with the exact same indices can be permuted, which is inconsequential, and that the former is stated in terms of the innovations in each variable, while the latter is in terms of innovations in the scalar components. However, if we rewrite the innovations of the scalar components model in terms of the innovation in each variable using the relationship* $\varepsilon_t^* = \boldsymbol{A}\varepsilon_t$*, the structure of the moving average matrices will not change because when any lower diagonal matrix is pre-multiplied by a row vector which has zeros at position c and everywhere to the right of c, the outcome will be a row vector with the exact same structure. This completes the proof.* ■

Example 4 *Consider the VARMA*$(1,1)$ *process*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \phi_{11}^* & \phi_{12}^* & \phi_{13}^* \\ \phi_{21}^* & \phi_{22}^* & \phi_{23}^* \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t^* - \begin{bmatrix} \theta_{11}^* & \theta_{12}^* & 0 \\ \theta_{21}^* & \theta_{22}^* & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}^*. \tag{7}$$

This process is a canonical SCM representation and consists of three SCMs of orders $(1,1)$*,* $(1,1)$ *and* $(0,0)$*. Obviously, as Theorem (3) predicts, this model has Kronecker invariants* $\boldsymbol{k} = \boldsymbol{s}^{\max} = (\max(1,1), \max(1,1), \max(0,0))' = (1,1,0)$*. If we substitute* $\boldsymbol{A}\varepsilon_t$ *and* $\boldsymbol{A}\varepsilon_{t-1}$ *for* ε_t^* *and* ε_{t-1}^**, the structure of the moving average parameter matrix does not change, and the resulting system is the reverse Echelon form of a system with Kronecker invariants* $(1,1,0)$*.*

Having considered a situation where the canonical SCM and Echelon forms are identical, we now present an example where this is not the case.

Example 5 *Consider the VARMA process consisting of three SCMs of orders* $(1,1)$*,* $(1,0)$ *and* $(0,0)$*,*

$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \phi_{11}^* & \phi_{12}^* & \phi_{13}^* \\ \phi_{21}^* & \phi_{22}^* & \phi_{23}^* \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t^* - \begin{bmatrix} \theta_{11}^* & \theta_{12}^* & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}^*. \tag{8}$$

Notice now that for the second SCM, the "autoregressive" order is different from the "moving average" order, i.e., $p_r \neq q_r$ *for* $r = 2$*. According to Theorem (3), the corresponding 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 canonical reverse Echelon form representation is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \psi_{31}^{(0)} & \psi_{32}^{(0)} & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} & \psi_{13}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} & \psi_{23}^{(1)} \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \Psi_0 \varepsilon_t - \begin{bmatrix} \xi_{11}^{(1)} & \xi_{12}^{(1)} & 0 \\ \xi_{21}^{(1)} & \xi_{22}^{(1)} & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1}, \quad (9)$$

as in equation (7) with $\xi_{21}^{(1)} = -a_{21}\xi_{11}^{(1)}$ *and* $\xi_{22}^{(1)} = -a_{21}\xi_{12}^{(1)}$. *The Echelon form specification does not impose this restriction, whilst the SCM methodology discovers it and transforms the system to translate this restriction into a row of zeros in the moving average parameter matrix. This leads to a system with 11 free parameters rather than 12. This shows that VARMA models with SCMs with* $p_r \neq q_r$ *for some* r *are rank restricted versions of reverse Echelon forms with Kronecker indices* $k_r = \max(p_r, q_r)$ *for* $r = 1, \ldots, K$.

The above example shows that the SCM methodology discovers some additional restrictions compared to the Echelon form methodology. Since Hannan's Theorem (Hannan and Deistler, 1988) proves that the restrictions in the Echelon form are necessary and sufficient restrictions for the 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.

Theorem 3 shows that the Athanasopoulos and Vahid (2008a) SCM methodology complements the Echelon form methodology and helps us avoid the otherwise necessary reference to the complicated formulae involved with the specification of Echelon form VARMA models. Given a set of Kronecker invariants, applying Stage II of the scalar component methodology can identify a parameter space for a unique VARMA representation which is identical to the parameter space specified by the Echelon form formulae. However, it is important to highlight that these formulae are what makes the Echelon form very attractive and applicable when programming an identification process for VARMA models.

4.1 A Monte Carlo Evaluation

In this section we perform Monte Carlo experiments in order to evaluate the performance of the identification procedures when identifying some pre-specified VARMA data generating processes (DGPs). We consider the DGPs presented in Appendix A, for sample sizes $N = 100, 150, 200$ and 400 observations. Due to the long, manual and challenging process of identifying SCMs, only 50 iterations were 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 were performed for each model and for each sample size. The results are presented in Table 1.[2]

[2]We should note that these results are a summary of the more elaborate tables presented for each individual DGP in the unpublished PhD dissertation of Athanasopoulos (2007). These individual results are available upon request.

In comparing these results, extra attention is required as canonical SCMs and Echelon form models are identical only when $p_r = q_r \ \forall \ r = 1, \ldots, K$ as shown by Theorem 3.

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. (p_{SCM}, q_{SCM}) in the SCM case is the maximum "autoregressive", $p_{SCM} = \max(p_1, \ldots, p_K)$, and "moving average", $q_{SCM} = \max(q_1, \ldots q_K)$, order 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 identify the p and q orders separately, 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 specified correctly. 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 identified correctly.

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 correctly identifies the Kronecker indices of the model. This is then directly comparable to the E.O. of the Echelon form. To clarify how this information is extracted from the simulation results, we present the following example.

Example 6 *Consider the processes of equations (16),*

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

and (15)

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

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 using 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, 2007). 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 (15), 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 the 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, 2007). 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 (14), where the success rate is 83 percent of the Echelon form methodology.

5 Empirical Results

5.1 Data

The data we employ are 40 monthly macroeconomic time series from March 1959 to December 1998 (i.e., $N = 478$ observations), extracted from the Stock and Watson (1999) data set (see Appendix B). These come from eight general categories of economic activity and are transformed in exactly the same way as 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 using the SCM methodology, which we label VARMA(SCM), and using the Echelon form methodology, which we label VARMA(Echelon). We also consider two sets of VAR models: (i) VAR models selected by AIC and (ii) VAR models selected by BIC. We label these VAR(AIC) and VAR(BIC) respectively. We consider 12 as the maximum lag length for the VARs.

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

PANEL A: *DGP of equation* (10)

N	SCM			Echelon	
	M.O.	E.O.	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* (11)

N	SCM			Echelon	
	M.O.	E.O.	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* (12)

N	SCM			Echelon	
	M.O.	E.O.	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* (13)

N	SCM			Echelon	
	M.O.	E.O.	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* (14)

N	SCM			Echelon	
	M.O.	E.O.	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* (15)

N	SCM			Echelon	
	M.O.	E.O.	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* (16)

N	SCM			Echelon	
	M.O.	E.O.	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* (17)

N	SCM			Echelon	
	M.O.	E.O.	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)

5.2 Forecast Evaluation Method

We divide the data into 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). Each model is estimated once in the estimation sample. 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, TMSFE. 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 TMSFE 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 TMSFE, 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 TMSFE 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 best (PB) measure which has been used in the past in forecasting competitions (see Makridakis and Hibon, 2000). This measure shows the percentage of times each model forecasts 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 model specified by the scalar component methodology. For each forecast horizon h, the average relative ratio for the |MSFE| is defined as

$$\overline{|\text{MSFE}_h|} = \frac{1}{M}\sum_{i=1}^{M}\frac{|\text{MSFE(X)}_i|}{|\text{MSFE(VARMA(SCM))}_i|},$$

and the average relative ratio for the TMSFE is defined as

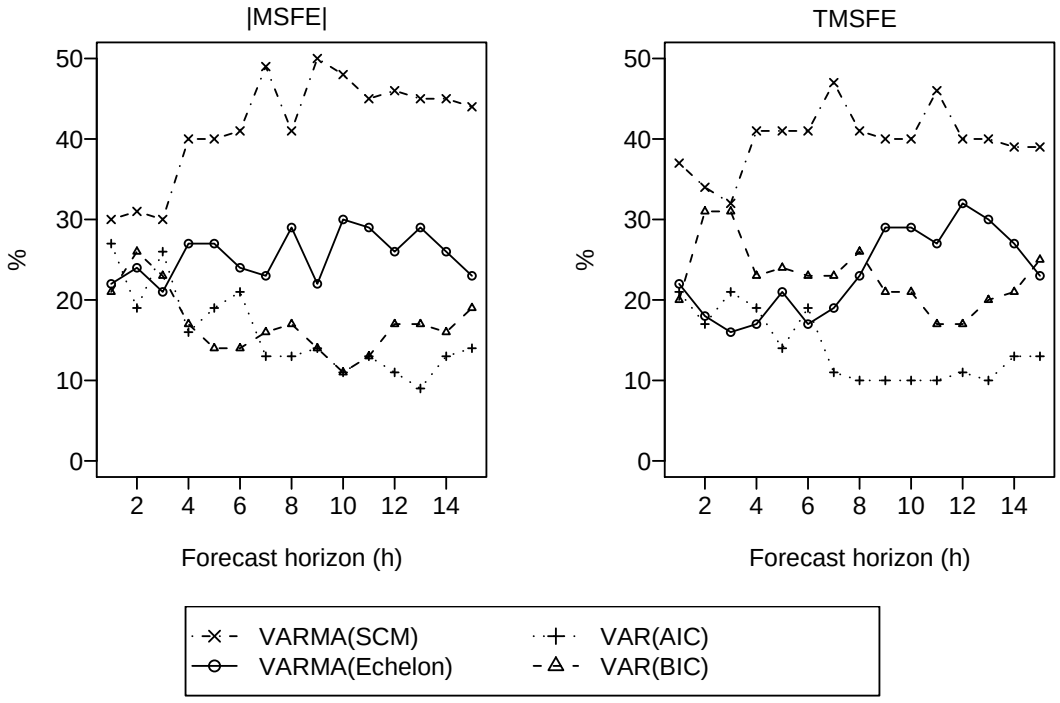
$$\overline{\text{TMSFE}_h} = \frac{1}{M}\sum_{i=1}^{M}\frac{\text{TMSFE(X)}_i}{\text{TMSFE(VARMA(SCM))}_i},$$

where X = {VARMA(Echelon), VAR(AIC), VAR(BIC)} are the alternative models we consider and $M = 70$ is the number of data sets. The reason 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.

5.3 PB Results

The PB counts have been plotted in Figure 1 (we present the actual counts in Table 2 in Appendix C). In each plot there are four lines, each one representing the alternative models we consider. It can be seen clearly from the plots that for both the |MSFE| and the TMSFE, and for all forecast horizons, VARMA models specified by the scalar component methodology forecast better more times than all other competing models.

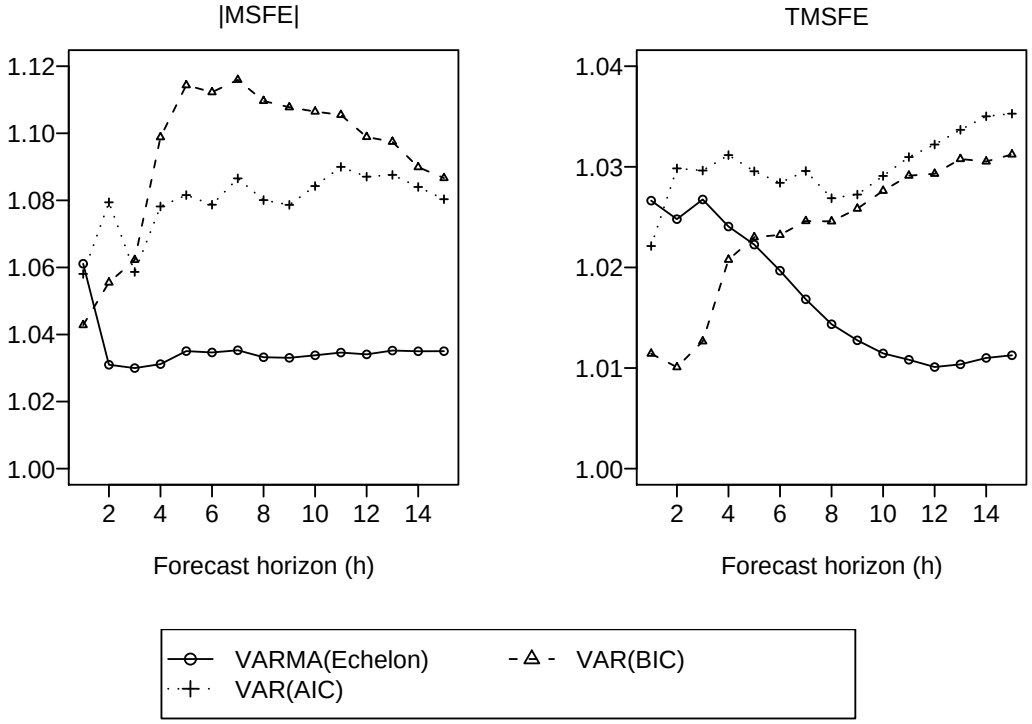
Figure 1: *Percentage better counts for canonical* SCM VARMA *models versus canonical Echelon form* VARMA *models and* VARs *with the lag length chosen by* AIC *and* BIC



5.4 Relative Ratios Results

The results for the relative ratios have been plotted in Figure 2 (we present the actual values in Table 3 in Appendix C). A first look at the two plots indicates that for all forecast horizons, and for both the |MSFE| and the TMSFE, the relative ratio measures are constantly greater than one. A

Figure 2: *Average relative ratios for canonical Echelon form* VARMA *models and* VARs *with the lag length chosen by* AIC *and* BIC *over canonical* SCM VARMA *models*



relative ratio greater than one shows that for that forecast horizon, the scalar component VARMA models forecast better on average than the competing models. For example, for forecast horizon $h = 6-$steps ahead, the SCM VARMA models improve on the |MSFE| (i.e., produce a lower |MSFE|) than the Echelon form VARMA models and the VARs selected by AIC and BIC by 3.5, 7.9 and 11.2 percent, respectively. The Echelon form VARMA models forecast better on average than VARs for $h \geq 2$ when considering the |MSFE| and for $h \geq 5$ when considering the TMSFE.

In Section 4 we conclude that a major difference between the two specifications of VARMA models 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 also have been the reason for the superior performance of the SCMs in the forecast evaluation exercise. In fact, the Echelon form methodology as presented by its various advocates (see for example Lütkepohl and Poskitt, 1996) includes an extra 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, which would be very time-intensive in such an extensive forecasting exercise. Furthermore, each reduction of the parameter space must be monitored, as the Kronecker indices have to be maintained. 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 directions for future research

This paper provides an in-depth comparison of canonical VARMA models specified by scalar components with VARMA models specified by the Echelon form methodology. We perform this comparison at the theoretical, experimental and empirical levels. At the theoretical level we show the connection between these two forms. This has revealed the missing intuition behind the complex formulae used for specifying Echelon form models – which now eliminates using these complexities as the reason for avoiding the identification and estimation of VARMA models. Furthermore, we show that scalar component VARMA models are more flexible in the sense that their maximum "autoregressive" order 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 some pre-specified VARMA data generating processes.

Finally, at the empirical level, the out-of-sample forecast evaluation shows that VARMA models specified by scalar components forecast better than Echelon form VARMA models, which in turn forecast better than VAR models. In the discussion of these forecast results we have acknowledged that our experimental design may have favoured 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 Echelon form VARMA models.

In line with the advocates of the Echelon form, during this research we have found that its greatest advantage is its practicality 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 which still requires a great deal of judgement and intervention 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 as it has happened with automatic univariate ARIMA modelling (see for example Mélard and Pasteels, 2000; Gómez and Maravall, 2001; Hyndman and Khandakar, 2008) 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 will be of great interest and is the subject of our current research.

References

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

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

Athanasopoulos, G. (2007) Essays on alternative methods of identification and estimation of vector autoregressive moving average models, *unpublished PhD dissertation,* Monash University, Department of Econometrics and Business Statistics.

Athanasopoulos, G. and F. Vahid (2008a) A complete VARMA modelling methodology based on scalar components, *Journal of Time Series Analysis*, **29**, 533–554.

Athanasopoulos, G. and F. Vahid (2008b) VARMA versus VAR for macroeconomic forecasting, *Journal of Business and Economic Statistics*, **26**, 237–252.

Box, G. E. P. and G. M. 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.

Cooley, T. and M. Dwyer (1998) Business cycle analysis without much theory. A look at structural VARs, *Journal of Econometrics*, **83**, 57–88.

Dufour, J.-M. and D. Pelletier (2008) Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application, *Discussion Paper,* Department of Economics, McGill University, CIREQ and CIRANO.

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.

Fernández-Villaverde, J., J. F. Rubio-Ramírez and T. J. Sargent (2005) A,B,C's (and D's) for understanding VARs, *NBER Technical Report 308*.

Fry, R. and A. Pagan (2005) Some issues in using VARs for macroeconometric research, *CAMA Working Paper Series 19/2005, Australian National University*.

Gómez, V. and A. Maravall (2001) Automatic modeling methods for univariate series, in D. Peña, G. Tiao and R. S. Tsay (eds.) *A Course in Time Series Analysis*, John Wiley and Sons, New York., pp. 365–407.

Hannan, E. J. (1969) The identification of vector mixed autoregressive-moving average systems,

Biometrica, **56**, 223–225.

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

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

Hannan, E. J. and M. Deistler (1988) *The statistical theory of linear systems*, John Wiley & Sons, New York.

Hannan, E. J. and L. Kavalieris (1984) Multivariate linear time series models, *Advances in Applied Probability*, **16**, 492–561.

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

Hyndman, R. J. and Y. Khandakar (2008) Automatic time series forecasting: The forecast package for R, *Journal of Statistical Software*, **26**, 1–22.

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

Lütkepohl, H. (1993) *Introduction to multiple time series analysis*, Springer-Verlag, Berlin-Heidelberg, 2nd ed.

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

Lütkepohl, H. and D. S. 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. Pasteels (2000) Automatic ARIMA modelling including interventions, using time series expert software, *International Journal of Forecasting*, **16**, 497–508.

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

Poskitt, D. S. (2005) A note on the specification and estimation of ARMAX systems, *Journal of Time Series Analysis*, **26**, 157–183.

Quenouille, M. H. (1957) *The analysis of multiple time series*, London: Charles Griffin & Company.

Ravenna, F. (2007) Vector autoregressions and reduced form representations of DSGE models, *Journal of Monetary Economics*, **54**, 2048–2064.

Reinsel, G. C. (1997) *Elements of multivariate time series*, New York: Springer-Verlag, 2nd ed.

Solo, V. (1986) Topics in advanced time series analysis, in G. Del Pino and R. Rebolledo (eds.) *Lectures in Probability and Statistics*, Springer-Verlag, New York.

Stock, J. H. and M. W. Watson (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*, New York: Oxford University Press.

Tiao, G. C. (2001) Vector ARMA, in D. Peña, G. C. Tiao and R. S. Tsay (eds.) *A Course in Time Series Analysis*, John Wiley and Sons, New York, pp. 365–407.

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

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

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.

Zellner, A. and F. Palm (1974) Time series analysis and simultaneous equation econometric models, *Journal of Econometrics*, **2**, 17–54.

A Data Generating Processes considered in Section 4.1

$$\boldsymbol{y}_t = \begin{bmatrix} 0.5 & -0.6 & 0.7 \\ 0.6 & 0.7 & -0.4 \\ 0.3 & 0.6 & 0.4 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t \tag{10}$$

$$\boldsymbol{y}_t = \varepsilon_t - \begin{bmatrix} 0.5 & -0.6 & 0.7 \\ 0.6 & 0.7 & -0.4 \\ 0.3 & 0.6 & 0.4 \end{bmatrix} \varepsilon_{t-1} \tag{11}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 1 & 0 \\ -0.6 & 0 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.7 & 0.6 & 0.4 \\ 0. & 0. & 0. \\ 0. & 0. & 0. \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t - \begin{bmatrix} -0.7 & 0. & 0. \\ 0. & 0. & 0. \\ 0. & 0. & 0. \end{bmatrix} \varepsilon_{t-1} \tag{12}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 1 & 0 \\ 0.4 & 0.7 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1} \tag{13}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 1 & 0 \\ 0.4 & 0.7 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.5 & 0.6 & -0.4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t - \begin{bmatrix} 0.5 & 0.7 & 0 \\ 0.2 & 0.7 & 0.5 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1} \tag{14}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5 & -0.7 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.7 & -0.5 & 0.7 \\ 0.6 & 0.3 & 0.6 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t - \begin{bmatrix} 0.5 & -0.6 & 0 \\ 0.6 & 0.7 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1} \tag{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 1 & 0 \\ 0 & -0.6 & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.7 & -0.6 & 0.4 \\ 0.6 & -0.5 & -0.4 \\ 0.3 & -0.6 & 0.4 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t - \begin{bmatrix} 0.7 & 0.4 & -0.6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \varepsilon_{t-1} \tag{16}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0. & 1 & 0 \\ 0. & 0. & 1 \end{bmatrix} \boldsymbol{y}_t = \begin{bmatrix} 0.6 & -0.7 & 0.4 \\ 0.7 & 0.5 & -0.4 \\ 0.3 & -0.7 & 0.4 \end{bmatrix} \boldsymbol{y}_{t-1} + \varepsilon_t - \begin{bmatrix} 0.7 & -0.3 & 0.4 \\ 0.2 & 0.6 & 0.5 \\ -0.3 & 0.4 & 0.4 \end{bmatrix} \varepsilon_{t-1} \tag{17}$$

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 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 $(\Delta y_t = y_t - y_{t-1})$ and (3) the first difference of the logarithm, i.e., series transformed to growth rates $(100 * \Delta \ln y_t)$. For complete descriptions of the series refer to Watson (2001).

(i) Output and income

IP[3]	IPP[3]	IPF[3]	IPC[3]	IPUT[3]	PMP[1]	GMPYQ[3]

(ii) Employment and hours

LHUR[1]	LPHRM[1]	LPMOSA[1]	PMEMP[1]

(iii) Consumption, manufacturing and retail

MSMTQ[3]	MSMQ[3]	MSDQ[3]	MSNQ[3]	WTQ[3]	WTDQ[3]	WTNQ[3]
RTQ[3]	RTNQ[3]	CMCQ[3]				

(iv) Real inventories and inventory-sales ratios

IVMFGQ[3]	IVMFDQ[3]	IVMFNQ[3]	IVSRQ[2]	IVSRMQ[2]	IVSRWQ[2]	IVSRRQ[2]
MOCMQ[3]	MDOQ[3]					

(v) Prices and wages

PMCP[1]

(vi) Money and credit quantity aggregates

FM2DQ[3]	FCLNQ[3]

(vii) Interest rates

FYGM3[2]	FYGM6[2]	FYGT1[2]	FYGT10[2]	TBSPR[1]

(viii) Exchange rates, stock prices and volume

FSNCOM[3]	FSPCOM[3]

C Tables

Table 2: *Percentage better counts for canonical* SCM VARMA *models versus canonical Echelon form* VARMA *models and* VARs *with the lag length chosen by* AIC *and* BIC

	Forecast horizon (h)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Average
	PB for the |MSFE|															
VARMA(SCM)	30	31	30	40	40	41	49	41	50	48	45	46	45	45	44	42
VARMA(Echelon)	22	24	21	27	27	24	23	29	22	30	29	26	29	26	23	25
VAR(AIC)	27	19	26	16	19	21	13	13	14	11	13	11	9	13	14	16
VAR(BIC)	21	26	23	17	14	14	16	17	14	11	13	17	17	16	19	17
	PB for the TMSFE															
VARMA(SCM)	37	34	32	41	41	41	47	41	40	40	46	40	40	39	39	40
VARMA(Echelon)	22	18	16	17	21	17	19	23	29	29	27	32	30	27	23	23
VAR(AIC)	21	17	21	19	14	19	11	10	10	10	10	11	10	13	13	14
VAR(BIC)	20	31	31	23	24	23	23	26	21	21	17	17	20	21	25	23

Note: all figures have been rounded to the nearest integer

Table 3: *Average relative ratios for canonical Echelon form* VARMA *models and* VARs *with lag length chosen by* AIC *and* BIC *over canonical* SCM VARMA *models*

	Forecast horizon (h)							Average over forecast horizon			
	1	2	3	4	6	12	15	1–3	1–6	1–12	1–15
	Average relative ratios for the |MSFE|										
VARMA(Echelon)	1.061	1.031	1.030	1.031	1.035	1.034	1.035	1.041	1.037	1.036	1.035
VAR(AIC)	1.058	1.079	1.059	1.078	1.079	1.087	1.080	1.065	1.072	1.078	1.080
VAR(BIC)	1.043	1.055	1.062	1.099	1.112	1.099	1.087	1.054	1.081	1.094	1.094
	Average relative ratios for the TMSFE										
VARMA(Echelon)	1.027	1.025	1.027	1.024	1.020	1.010	1.011	1.026	1.024	1.018	1.017
VAR(AIC)	1.022	1.030	1.030	1.031	1.028	1.032	1.035	1.027	1.028	1.029	1.030
VAR(BIC)	1.011	1.010	1.013	1.021	1.023	1.029	1.031	1.011	1.017	1.022	1.024