ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

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

VARMA versus VAR for Macroeconomic Forecasting

George Athanasopoulos and Farshid Vahid

January 2006

Working Paper 4/06

VARMA versus *VAR* for Macroeconomic Forecasting

George Athanasopoulos*
Monash University
Clayton, Victoria 3800
AUSTRALIA

Farshid Vahid
Australian National University
Canberra, ACT, 0200
AUSTRALIA

January 30, 2006

Abstract

In this paper, we argue that there is no compelling reason for restricting the class of multivariate models considered for macroeconomic forecasting to *VAR*s given the recent advances in *VARMA* modelling methodology and improvements in computing power. To support this claim, we use real macroeconomic data and show that *VARMA* models forecast macroeconomic variables more accurately than *VAR* models.

KEY WORDS: Identification, Multivariate time series, Scalar component, *VARMA* model.

JEL classification: C32, C51

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

1 INTRODUCTION

Finite order vector autoregressive moving average *VARMA* models are motivated by the Wold decomposition theorem (Wold 1938) applied in a multivariate setting, as an appropriate class of models for stationary time series. Hence, the study of *VARMA* models has been an important area of time series analysis for a long time (see, amongst others, Quenouille 1957; Hannan 1969; Tunnicliffe-Wilson 1973; Hillmer and Tiao 1979; Tiao and Box 1981; Tiao and Tsay 1989; Tsay 1991; Poskitt 1992; Lütkepohl 1993; Lütkepohl and Poskitt 1996; Reinsel 1997; Tiao 2001). However, macroeconomists have yet to be convinced of the advantages of employing such models. One of the reasons might be that, to date, no applied macroeconometric research paper has considered the *VARMA* model as an alternative to the finite order vector autoregressive *VAR* model.

Ever since the publication of the seminal paper by Christopher Sims (Sims 1980), the finite order *VAR* model has become the cornerstone of macro-econometric modelling. The reason for this cannot be that economic theory implies finite order *VAR* dynamics for economic variables. Economic theory rarely has any sharp implications about the short-run dynamics of economic variables. In the rare situations where theoretical models include a dynamic adjustment equation, one has to work hard to exclude moving average terms from appearing in the implied dynamics of the variables of interest. Even if we believe that a finite order *VAR* is a good dynamic model for a particular set of variables, for any subset of these variables, a *VARMA* model rather than a *VAR* model would be appropriate. Therefore, the apparent lack of interest in multivariate models with moving average errors could be either because they are too difficult to implement, or because pure autoregressive models can perform just as well.

Any invertible *VARMA* process can be approximated by a finite long order *VAR*. However, this does not imply that forecasts based on an estimated long order *VAR* will be as good as those based on a parsimonious *VARMA* model, since a long order *VAR* has many estimated parameters. It appears that the main reason for the lack of enthusiasm for using models with moving average errors is that they are too difficult. In particular, in a multivariate setting, the identification and estimation of *VARMA* models are quite involved. This is in sharp contrast to the ease of identification and estimation of *VAR* models. This difficulty in model identification and estimation has thus far prevented any comprehensive assessment of whether *VARMA* models outperform *VAR* models in forecasting macroeconomic variables in finite samples.

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

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

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

This model is not identified if $\phi_{21} = \phi_{22} = \theta_{21} = \theta_{22} = 0$. *In this case the second equation implies that* $y_{2,t-1} = \eta_{2,t-1}$ *and therefore* ϕ_{12} *and* θ_{12} *in the first equation cannot be identified separately.*

There are several methods for identifying *VARMA* models. The method that we consider in this paper is the Athanasopoulos and Vahid (2005) extension to Tiao and Tsay (1989). This methodology consists of three stages. In the first stage, "scalar component models" (*SCM*s) 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*). An overview of this methodology is presented in Section 2. In this paper we employ this methodology in order to answer the question of whether *VARMA* models outperform *VAR* models in forecasting macroeconomic variables.

To compare the forecasting performance of *VARMA* and *VAR* models we use real data. We compile seventy trivariate sets of monthly macroeconomic variables, and fit *VAR* and *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 *VARMA* and *VAR* models.

In practice, *VAR* models are used not only for forecasting, but also for impulse response and variance decomposition analysis. However, since the true model is not known, it is not possible to use real data to compare the performance of *VARMA* versus *VAR* models in tasks other than forecasting. Comparisons of impulse responses and variance decompositions can only be performed when the data generating process is known. However, because of the large dimension of the parameter space in multivariate time series models, designing a Monte Carlo study that is sufficiently rich to lead to convincing general conclusions would be difficult, if not impossible. For this reason, we use real macroeconomic data for this investigation and compare the out-of-sample forecast performance of fitted *VARMA* and *VAR* models. The advantage of this method is that the results will be of direct relevance for macroeconomic forecasting. The drawback is the inability to assess or comment on which model produces a better impulse response function or a better decomposition of forecast error variance, since these objects of interest are not observable.

The structure of the paper is as follows. Section 2 outlines our *VARMA* modelling methodology. Section 3 describes the data, the forecast evaluation method and the empirical results. Section 4 provides some conclusions.

2 A *VARMA* MODELLING METHODOLOGY BASED ON SCALAR COMPONENTS

The *VARMA* modelling methodology we employ in this paper is the Athanasopoulos and Vahid (2005) extension to Tiao and Tsay (1989). 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 2 *For a given $K-$dimensional* VARMA(p, q) *process*

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

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:*

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

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 $\boldsymbol{Y}_{m,t} \equiv \left(\boldsymbol{y}'_t, \ldots, \boldsymbol{y}'_{t-m}\right)$ 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

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

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,

$$d_i = 1 + 2\sum_{v=1}^{j} \widehat{\rho}_v\left(\widehat{\boldsymbol{r}}'_i\boldsymbol{Y}_{m,t}\right)\widehat{\rho}_v\left(\widehat{\boldsymbol{g}}'_i\boldsymbol{Y}_{h,t-1-j}\right) \qquad (8)$$

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 (2) by $\boldsymbol{A}$, we obtain

$$\boldsymbol{A}\boldsymbol{y}_t = \boldsymbol{\Psi}_1\boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Psi}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t - \boldsymbol{\Theta}^*_1\boldsymbol{\varepsilon}_{t-1} - \ldots - \boldsymbol{\Theta}^*_q\boldsymbol{\varepsilon}_{t-q}, \qquad (9)$$

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

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

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 2005 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 (2). 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 (2005) 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). 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 *SCM*s that nest this one are set to zero. If the test concludes that the *SCM* can be formed using the first $K - 1$ variables only, the coefficient of the Kth variable in this *SCM* is zero, and should not be 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 $SCM(p_2, q_2)$ 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.

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 *SCM*s

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 *SCM*s. 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 $\boldsymbol{A}$

Apply the identification rules stated above to identify the structure of the transformation matrix $\boldsymbol{A}$. Extensive Monte-Carlo experiments in Athanasopoulos (2005) show that these two stages perform well in identifying some pre-specified data generating processes with various orders of embedded *SCM*s.

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 *SCM*s with no moving average components. Alternatively, lagged innovations can be estimated from a long *VAR* and used for obtaining initial estimates for the parameters as in Hannan and Risannen (1982). The maximum likelihood procedure provides estimates and estimated standard errors for all parameters, including the free parameters 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 EMPIRICAL RESULTS

3.1 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 4). 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 4. 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.

3.2 Models Considered

For each of the seventy data sets we estimate five models: (i) a *VARMA* model developed employing the *SCM* methodology of Section 2, (ii) a *VAR* model chosen by the *AIC*, (iii) a *VAR* model chosen by the *BIC* (iv) a restricted version of (ii) with all insignificant coefficients restricted to zero, and (v) a restricted version of (iii) with all insignificant coefficients restricted to zero. We consider the restricted *VAR* models to ensure that unfavourable results of *VAR* models are not due to redundant parameters in the unrestricted *VAR*s. In both (iv) and (v)

restrictions are imposed one at a time by eliminating the parameter with the highest p-value among all insignificant parameters at the 5 percent level of significance. Restricted models are estimated using the seemingly unrelated regression estimation method (Zellner 1963) as not all equations include the exact same regressors.

3.3 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

$$\overline{RdMSFE_h} = \frac{1}{M}\sum_{i=1}^{M}\frac{|MSFE\,(VAR)_i|}{|MSFE\,(\ VARMA)_i|},$$

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

$$\overline{RtMSFE_h} = \frac{1}{M}\sum_{i=1}^{M}\frac{tr\,(MSFE\,(VAR)_i)}{tr\,(MSFE\,(VARMA\ )_i)},$$

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

3.4 *PB* Results

The *PB* counts have been plotted in Figures 1 to 5 (the actual counts for all measures are presented in Appendix 4). In these figures there are three lines, each one representing a class of models. The marked points on each line depict the percentage of times for which that class of models produces the best forecast for that horizon amongst all models. For example, consider the 7 -step-ahead forecast performance for *VARMA* models versus the unrestricted *VAR* models selected by the *AIC*, i.e., *VAR*(*AIC*) , and those selected by the *BIC*, i.e., *VAR*(*BIC*). Figure 1 shows that the *VARMA* models outperform both sets of *VAR* models, as approximately 60 percent of the time they produce lower values of $|MSFE|$. In general, Figure 1 shows that the *VARMA* models produce the highest *PB* counts for the $|MSFE|$ for all $h = 1$ to 15-step-ahead forecast horizons when compared to their *VAR* counterparts.

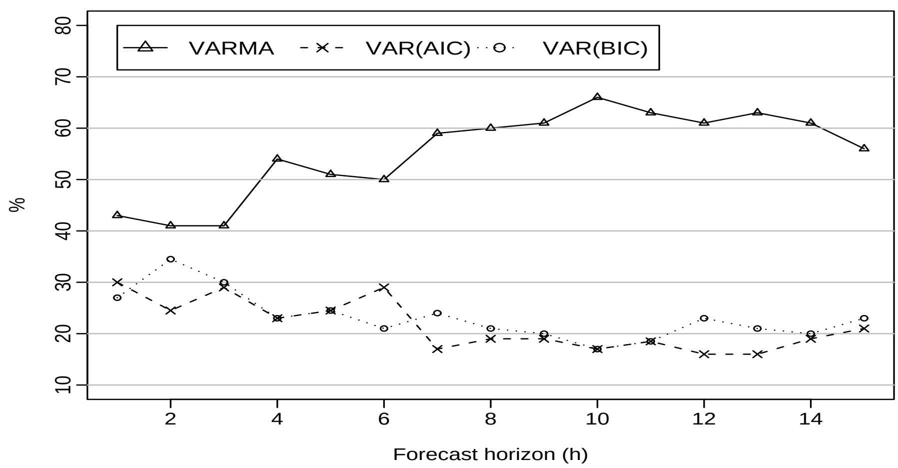

Figure 1: *PB counts for* $|MSFE|$ *for VARMA versus unrestricted VAR models selected by AIC and BIC*

Figure 2 shows the PB counts again for the $|MSFE|$, however now the VAR counterparts of the $VARMA$ models are VAR models whose insignificant lags have been omitted as described in Section 3.2. This figure shows that the only forecast horizon for which $VARMA$ models are outperformed by VARs is for $h = 3$. For this forecast horizon the restricted VARs selected by the AIC perform slightly better than their $VARMA$ counterparts.

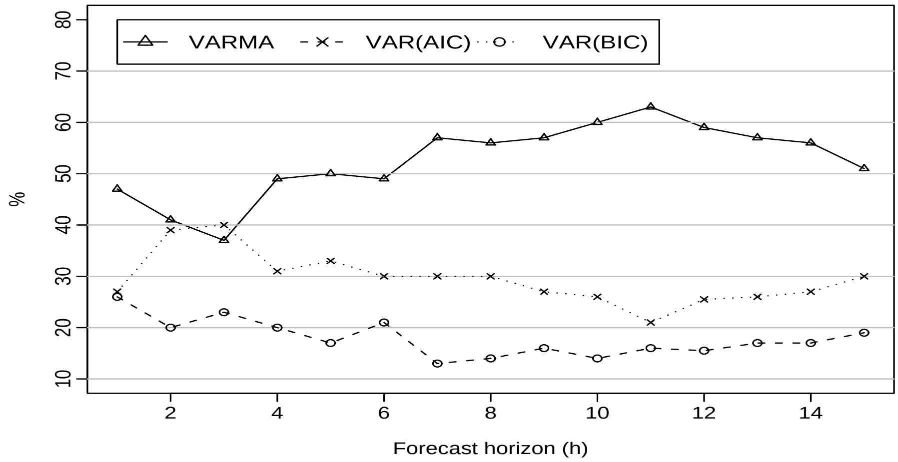

Figure 2: *PB counts for |MSFE| for VARMA versus restricted VAR models selected by AIC and BIC*

Figures 3 and 4 present the PB counts of $VARMA$ models versus unrestricted and restricted VARs respectively, when the forecast accuracy measure is $tr\,(MSFE)$. Again the results are overwhelmingly in favour of the $VARMA$ models.

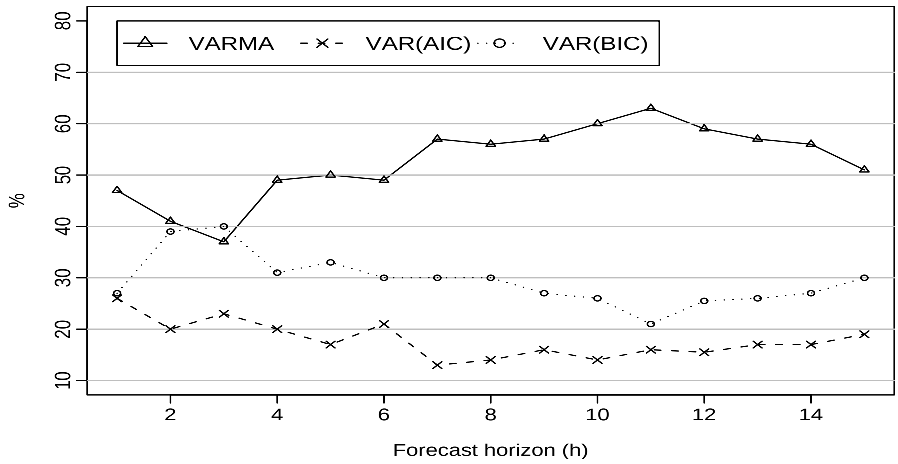

Figure 3: *PB counts for tr(MSFE) for VARMA versus unrestricted VAR models selected by AIC and BIC*

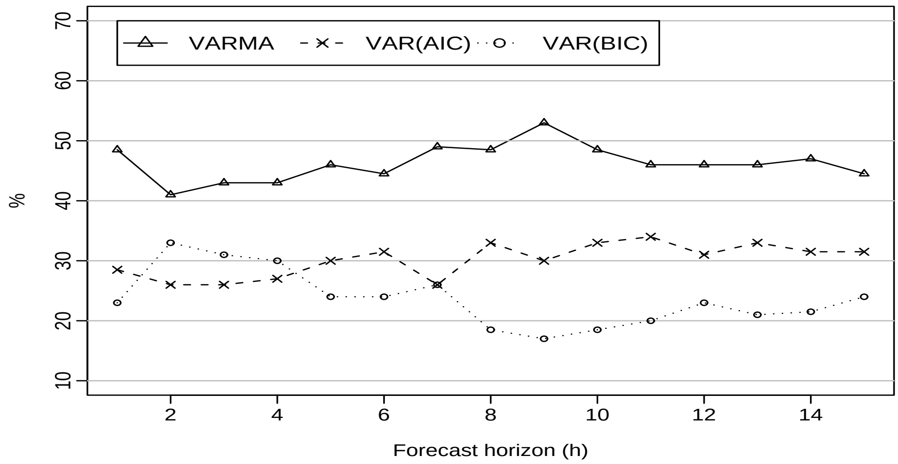

Figure 4: *PB counts for tr(MSFE) for VARMA versus restricted VAR models selected by AIC and BIC*

As a general observation for both accuracy measures and for both restricted and unrestricted *VAR* models, *VARMA* models perform better at least 50 percent of the time for forecast horizons of more than six or seven steps ahead. The significance of the 50 percent figure is that if we could somehow choose the best of the *VAR* models selected by either the *AIC* or the *BIC*, the *VARMA* models would still "out-forecast" them.

Furthermore, Figure 5 shows a head-to-head comparison of the *PB* counts for the $|MSFE|$ of *VARMA* models versus *VAR*s, selected by either the *AIC* or the *BIC*. The results based on the $tr\,(MSFE)$ are qualitatively similar and the raw figures for both measures are presented in Appendix 4. This comparison further supports the dominance of the *VARMA* models. Now that the comparison is between *VARMA* models and one class of *VAR* models at a time, the *VARMA* models outperform their *VAR* counterparts for each and every forecast horizon. The lowest *PB* count that the *VARMA* models achieve is 51 percent, for the $h = 3$ forecast horizon when compared to the restricted $VAR(AIC)$. Every other count is above 53 percent with a maximum of 79 percent for $h = 10$ and 12 when compared to unrestricted $VAR(AIC)$. Furthermore, for $h > 3$ at least 60 percent of the time *VARMA* models outperform their *VAR* counterparts.

3.5 Relative Ratios Results

The results for the relative ratios are tabulated in Table 1, and plotted in Figure 6. All panels of Table 1 and Figure 6 indicate that for all forecast horizons and for both restricted and unrestricted *VAR*s, all relative ratio measures are consistently greater than one. A relative ratio greater than one shows that for that forecast horizon *VARMA* models perform better than *VAR*s. For example, in Panel A for $h = 4$, the determinant of the $MSFE$ matrix for

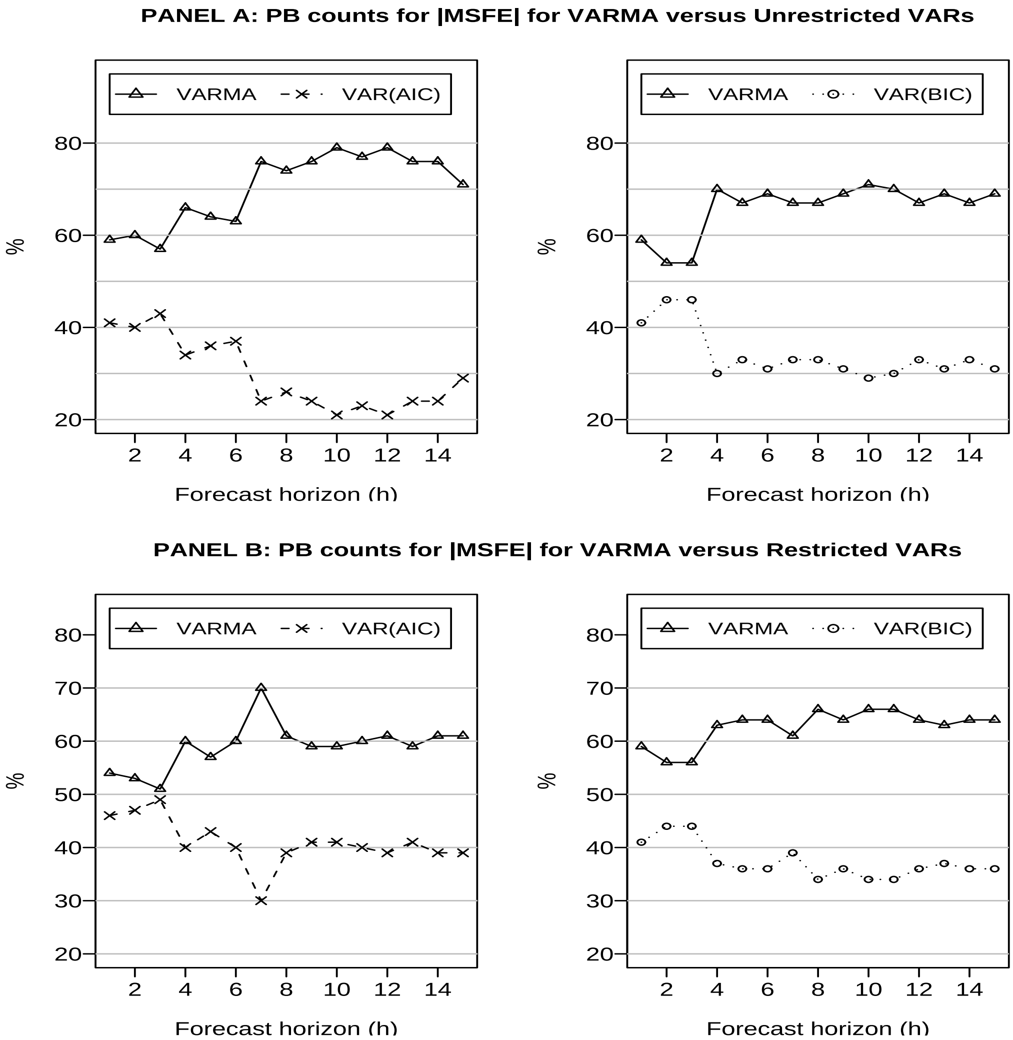

Figure 5: *A head to head comparison of PB counts for* $|MSFE|$ *between VARMA and unrestricted and restricted VAR models selected by either AIC or BIC*

VAR(AIC) is on average 8 percent larger than that of *VARMA*s. For the same horizon, this improvement jumps to approximately 10 percent when compared to the *VAR*(BIC) models.

Table 1: *Average relative ratios for the determinant and the trace of the MSFE matrices for VAR models selected by AIC and BIC over VARMA*

Panel A: $\overline{RdMSFE}$ of *Unrestricted VAR* over *VARMA*

	Forecast Horizon (h)							Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
VAR(AIC)	1.058	1.079	1.059	1.078	1.080	1.087	1.080	1.069	1.075	1.078	1.080
VAR(BIC)	1.043	1.055	1.062	1.099	1.110	1.099	1.087	1.065	1.089	1.094	1.094

Panel B: $\overline{RdMSFE}$ of *Restricted VAR* over *VARMA*

	Forecast Horizon (h)							Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
VAR(AIC)	1.028	1.040	1.020	1.042	1.050	1.058	1.053	1.033	1.042	1.046	1.048
VAR(BIC)	1.032	1.042	1.054	1.093	1.109	1.101	1.089	1.055	1.083	1.090	1.091

Panel C: $\overline{RtMSFE}$ of *Unrestricted VAR* over *VARMA*

	Forecast Horizon (h)							Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
VAR(AIC)	1.022	1.030	1.030	1.031	1.027	1.032	1.035	1.028	1.028	1.029	1.030
VAR(BIC)	1.011	1.010	1.013	1.021	1.025	1.029	1.031	1.014	1.019	1.022	1.024

Panel D: $\overline{RtMSFE}$ of *Restricted VAR* over *VARMA*

	Forecast Horizon (h)							Av. of Forecast Horizon			
	1	*2*	*3*	*4*	*8*	*12*	*15*	*1-4*	*1-8*	*1-12*	*1-15*
VAR(AIC)	1.021	1.025	1.023	1.025	1.013	1.017	1.021	1.023	1.020	1.018	1.019
VAR(BIC)	1.014	1.013	1.016	1.023	1.026	1.032	1.034	1.017	1.021	1.024	1.026

The last four columns of each panel of Table 1 show the average relative ratio over several forecast horizons. These also highlight the improvement *VARMA* models bring to out-of-sample forecasting in comparison to *VAR*s. The last column shows the overall average improvement over all fifteen forecast horizons. This overall improvement ranges between approximately 2 percent, achieved for the average of the relative ratio for the $tr\,(MSFE)$ for restricted *VAR*(AIC) models (see Panel D), and 9.5 percent, achieved for the average of the relative ratio for the $|MSFE|$ for restricted *VAR*(BIC) models (see Panel A).

While the main objective of this forecasting exercise is to compare the forecast performance of *VARMA* models with that of *VAR*s, one can also compare the forecast performance between the *VAR* models selected by the two model selection criteria. Figure 6 allows for a head-to-head comparison between the *VAR* models selected by the AIC those selected by the *BIC*. Panels A and B compare the determinants of the mean squared forecast error matrices for the unrestricted and restricted *VAR*s. Panel A shows that for $h \leq 2$, *VAR*(BIC) perform better, but for all $h > 2$ *VAR*(AIC) dominate. The performance of the two seems to converge beyond $h \geq 7$.

Panel B compares the performance of the restricted *VAR* models. Recall that the restricted

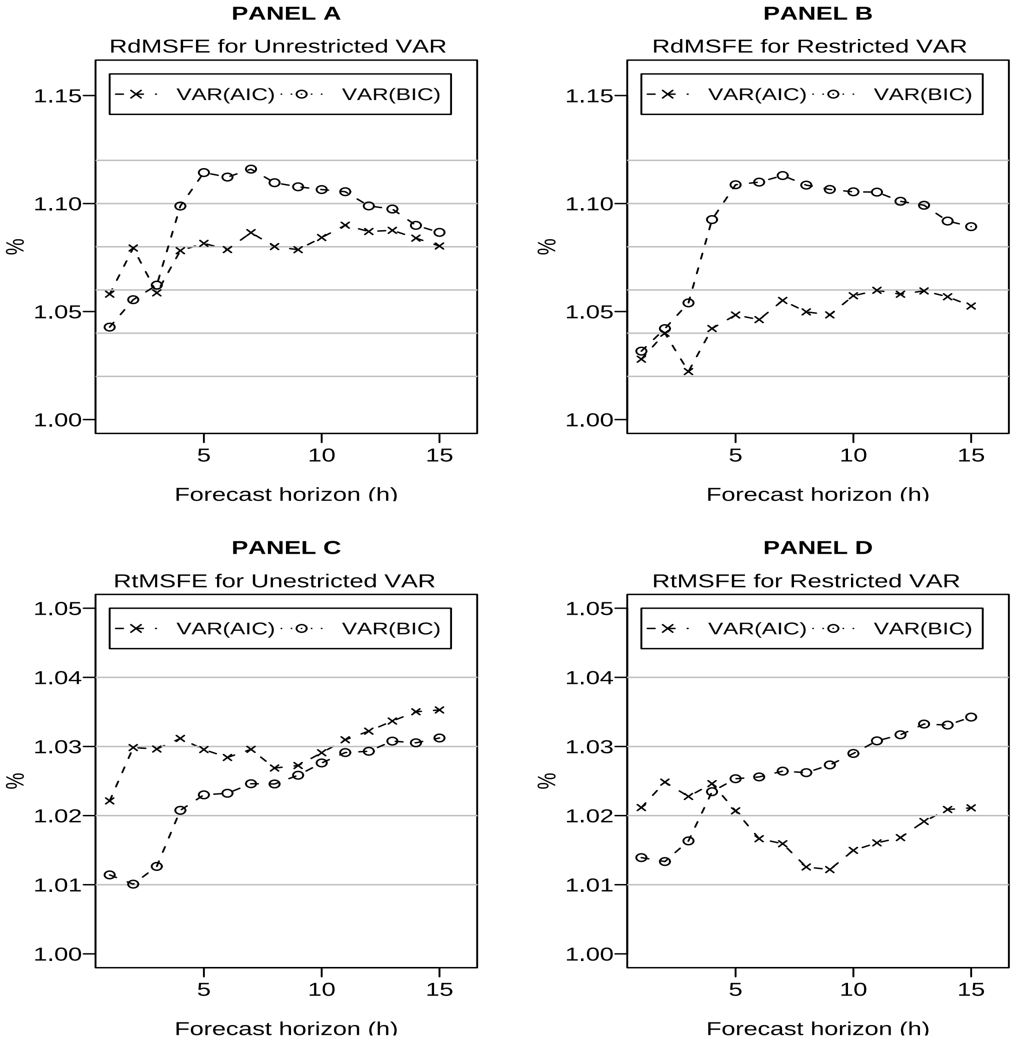

Figure 6: *Average relative ratios of the determinant and the trace of the MSFE matrices for VAR models selected by AIC and BIC over VARMA*

VAR models are VARs with their lag length selected by a model selection criterion which then have their insignificant right hand side variables omitted. This figure shows that the restricted $VAR(AIC)$ models are at least as good as the restricted $VAR(BIC)$ models. Comparing panels A and B shows that eliminating insignificant variables improves the performance of $VAR(AIC)$ models considerably, but it does not improve (in fact it slightly worsens) the performance of $VAR(BIC)$ models. Panels C and D present the performance of the VAR models based on the $tr\,(MSFE)$. The strong message from these figures is that $VAR(AIC)$ models must be used for forecasting only after their insignificant variables are eliminated.

4 CONCLUSION

The message of this paper is that we can obtain better forecasts for macroeconomic variables by considering *VARMA* models rather than restricting ourselves to VAR models. With recent methodological advances in the identification and estimation of *VARMA* models, and with the improvement in computing power and econometrics software, there is no compelling reason to restrict the class of models to VARs only. Our empirical results show that *VARMA* models developed by a scalar components methodology outperform VAR models in forecasting macroeconomic variables. Are these favourable results specific to *VARMA* models developed by the scalar component methodology that we adopt in this paper? The answer is negative. Athanasopoulos (2005) shows that the same conclusion emerges when one uses an "echelon form" approach (Hannan and Deistler 1988; Lütkepohl and Poskitt 1996) to develop *VARMA* models. Therefore the improvements seem to be a result of expanding the set of possible models to *VARMA* models rather than any specific approach for developing *VARMA* models.

APPENDIX A: 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 ($\Delta y_t = y_t - y_{t-1}$) and (3) the first difference of the logarithm, i.e., series transformed to growth rates ($100 * \Delta \ln y_t$). 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)

APPENDIX B: RAW DATA FOR THE PERCENTAGE BETTER (*PB*) COUNTS CONSIDERING ALL MODELS

Table 2: *PB counts for $|MSFE|$ for VARMA versus unrestricted VAR models selected by AIC and BIC*

	Forecast horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
VARMA	43.0[a]	41.0	41.0	54.0	51.0	50.0	59.0	60.0	61.0	66.0	63.0	61.0	63.0	61.0	56.0
VAR(BIC)	27.0	34.5	30.0	23.0	24.5	21.0	24.0	21.0	20.0	17.0	18.5	23.0	21.0	20.0	23.0
VAR(AIC)	30.0	24.5	29.0	23.0	24.5	29.0	17.0	19.0	19.0	17.0	18.5	16.0	16.0	19.0	21.0

Table 3: *PB counts for $|MSFE|$ for VARMA versus restricted VAR models selected by AIC and BIC*

	Forecast horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
VARMA	44.0[a]	42.0	37.0	46.0	46.0	49.0	54.0	47.0	49.0	47.0	49.0	48.0	46.0	47.0	46.0
VAR(BIC)	23.0	24.0	24.0	21.0	20.0	17.0	20.0	20.0	20.0	21.5	23.0	26.0	23.0	23.0	23.0
VAR(AIC)	33.0	34.0	39.0	33.0	34.0	34.0	26.0	33.0	31.0	31.5	29.0	26.0	31.0	30.0	31.0

Table 4: *PB counts for tr(MSFE) for VARMA versus unrestricted VAR models selected by AIC and BIC*

	Forecast horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
VARMA	47.0[a]	41.0	37.0	49.0	50.0	49.0	57.0	56.0	57.0	60.0	63.0	59.0	57.0	56.0	51.0
VAR(BIC)	27.0	39.0	40.0	31.0	33.0	30.0	30.0	30.0	27.0	26.0	21.0	25.5	26.0	27.0	30.0
VAR(AIC)	26.0	20.0	23.0	20.0	17.0	21.0	13.0	14.0	16.0	14.0	16.0	15.5	17.0	17.0	19.0

Table 5: *PB counts for tr(MSFE) for VARMA versus restricted VAR models selected by AIC and BIC*

	Forecast horizon (h)														
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
VARMA	48.5[a]	41.0	43.0	43.0	46.0	44.5	49.0	48.5	53.0	48.5	46.0	46.0	46.0	47.0	44.5
VAR(BIC)	23.0	33.0	31.0	30.0	24.0	24.0	26.0	18.5	17.0	18.5	20.0	23.0	21.0	21.5	24.0
VAR(AIC)	28.5	26.0	26.0	27.0	30.0	31.5	26.0	33.0	30.0	33.0	34.0	31.0	33.0	31.5	31.5

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

APPENDIX C: RAW DATA FOR THE HEAD TO HEAD COMPARISON OF THE PERCENTAGE BETTER (*PB*) COUNTS

Table 6: Percentage of times VARMA models forecast more accurately than unrestricted VAR models selected by AIC

Accuracy	Forecast horizon (h)														
Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$\|MSFE\|$	59	60	57	66	64	63	76	74	76	79	77	79	76	76	71
$tr\,(MSFE)$	61	57	57	67	69	71	76	71	76	77	76	73	74	74	71

Table 7: Percentage of times VARMA models forecast more accurately than unrestricted VAR models selected by BIC

Accuracy	Forecast horizon (h)														
Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$\|MSFE\|$	59	54	54	70	67	69	67	67	69	71	70	67	69	67	69
$tr\,(MSFE)$	59	54	50	61	63	61	66	69	70	70	70	67	66	64	63

Table 8: Percentage of times VARMA models forecast more accurately than restricted VAR models selected by AIC

Accuracy	Forecast horizon (h)														
Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$\|MSFE\|$	54	53	51	60	57	60	70	61	59	59	60	61	59	61	61
$tr\,(MSFE)$	57	59	57	63	61	59	64	61	63	59	59	57	56	59	57

Table 9: Percentage of times VARMA models forecast more accurately than restricted VAR models selected by BIC

Accuracy	Forecast horizon (h)														
Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$\|MSFE\|$	59	56	56	63	64	64	61	66	64	66	66	64	63	64	64
$tr\,(MSFE)$	61	54	56	59	60	63	61	69	69	67	63	63	61	63	63

References

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 Vahid, F. (2005), "A Complete *VARMA* Modelling Methodology Based on Scalar Components," working paper, Monash University, Department of Econometrics and Business Statistics.

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

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.

Hannan, E. J. (1969), "The Identification of Vector Mixed Autoregressive-Moving Average Systems," *Biometrika*, 56, 223-225.

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

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

Hillmer, S. C., and Tiao, G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," *Journal of the American Statistical Association*, 74, 652-660.

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

Lütkepohl, H., and Poskitt, D.S. (1996), "Specification of Echelon-Form *VARMA* Models," *Journal of Business and Economic Statistics*, 14, 69-79.

Makridakis, S., and Hibon, M. (2000), "The M3-Competition: Results, Conclusions and Implications," *International Journal of Forecasting*, 16, 451-476.

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

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1-48.

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

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

Tiao, G. C., and Box, G. E. P. (1981), "Modeling Multiple Time Series with Applications," *Journal of the American Statistical Association*, 76, 802-816.

Tiao, G. C., and Tsay, R. S. (1989), "Model Specification in Multivariate Time Series" (with discussions), *Journal of the Royal Statistical Society, Ser. B*, 51, 157-213 .

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, Ser. 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, Eight World Congress of the Econometric Society, III, 87-115.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, Stockholm: Almqvist and Wiksell.

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