Are VAR Models Good Enough for Forecasting Macroeconomic Variables?

George Athanasopoulos
Monash University Clayton,
Victoria 3800 Australia

and

Farshid Vahid
Australian National University
Canberra, ACT 2601 Australia

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

In this paper:

- Present and complete the “Scalar Component Methodology” of Tiao and Tsay (1989) for “developing” VARMA models
- Study the properties of this methodology through simulations
- Extensive application to Macro-Economic Data
- Overall we find evidence of superiority of VARMA models for forecasting

VAR v VARMA

- VAR(p)
  - Dominate the field of macro-econometric modelling
  - Good forecasting record
- MOTIVATION for VARMA(p,q)
  - More general
  - Parsimonious representations
    - any invertible VARMA can be represented by an infinite order VAR
    - effects on forecasting
  - Aggregation:
    - induces MA dynamics
  - See Lütkepohl (1987) where a linearly transformed:
    - VARMA process has a finite VARMA(p,q) representation
    - However not necessarily the case with VAR

- VARMA difficulties
  - Chatfield

“If univariate ARIMA modelling is difficult then VARMA modelling is even more difficult - some might say impossible”

- General VARMA(p,q) model

$$\mathbf{y}_t = (\mathbf{y}_{1t},\ldots\ldots,\mathbf{y}_{kt})'$$

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

where $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0},\mathbf{\Sigma}_\varepsilon)$ and $\mathbf{\Sigma}_\varepsilon$ is positive definite

- Identification problem
  - Echelon form
    - Lütkepohl and Poskitt (1996)
  - Scalar Components
    - Simplifying Underlying Structures

Identification Problem

- Consider a k = 2 y_t ~ VARMA(1,1)

$$\mathbf{y}_t = \mathbf{\Phi y}_{t-1} + \boldsymbol{\eta}_t - \mathbf{\Theta}\boldsymbol{\eta}_{t-1}$$

- $\phi_{11} = \phi_{12} = \theta_{11} = \theta_{12} = 0$

$$y_{1,t} = \eta_{1,t}$$

$$y_{2,t} = \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} - \theta_{21}\eta_{1,t-1} - \theta_{22}\eta_{2,t-1} + \eta_{2,t}$$

- ϕ_{21} and θ_{21} are not separately identified
- "Rule of Elimination"

Scalar Component Methodology

(Tiao and Tsay, JRSS, 1989)

- Definition:

$$\mathbf{y}_t \sim VARMA(p,q)$$

$$z_t = \boldsymbol{\alpha}'\mathbf{y}_t \sim SCM(p_1, q_1)$$

$$\textit{if } \boldsymbol{\alpha} \textit{ satisfies} \quad \begin{aligned} \boldsymbol{\alpha}'\mathbf{\Phi}_{p_1} &\neq \mathbf{0}^T \textit{ where } 0 \le p_1 \le p, \\ \boldsymbol{\alpha}'\mathbf{\Phi}_l &= \mathbf{0}^T \textit{ for } l = p_1 + 1, \ldots, p, \\ \boldsymbol{\alpha}'\mathbf{\Theta}_{q_1} &\neq \mathbf{0}^T \textit{ where } 0 \le q_1 \le q, \\ \boldsymbol{\alpha}'\mathbf{\Theta}_{q_1} &= \mathbf{0}^T \textit{ for } l = q_1 + 1, \ldots, q. \end{aligned}$$

- Find k-linearly independent vectors $\mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k)'$ which transform

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

into

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

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

- A series of C/C tests:

Let $\hat{\lambda}_1 < \hat{\lambda}_2 < \ldots < \hat{\lambda}_k$ be the squared C/C

between $\mathbf{Y}_{m,t} \equiv (\mathbf{y}_t', \ldots \mathbf{y}_{t-m}')'$ and $\mathbf{Y}_{h,t-j-1} \equiv \left(\mathbf{y}_{t-j-1}', \ldots \mathbf{y}_{t-j-1-h}'\right)'$

then the LR test statistic:

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

tests for *s SCM*s of order (m,j) versus the alternative of less than *s SCM*s of this order.

d_i is a correction factor accounting for the cases that the canonical covariates can be MA(j)

- EXAMPLE: $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{k,t})^T$
- sequence of C/C tests
- Underlying WN process $z_{i,t}$ ~ **SCM(0,0)**
  - Is there a linear combination of $\mathbf{y_t}$ which has zero correlation with $\mathbf{y_{t-1}}, \ldots$?
- MA(1) process $z_{i,t}$ ~ **SCM(0,1)**
  - Is there a linear combination of $\mathbf{y_t}$ which has zero correlation with $\mathbf{y_{t-2}}, \ldots$ given that it has one period serial correlation?
- AR(1) process $z_{i,t}$ ~ **SCM(1,0)**
  - Is there a linear combination of $\mathbf{y_t}$ and $\mathbf{y_{t-1}}$ which has zero correlation with $\mathbf{y_{t-1}}, \ldots$?
- ARMA(1,1) process $z_{i,t}$ ~ **SCM(1,1)**
  - Is there a linear combination of $\mathbf{y_t}$ and $\mathbf{y_{t-1}}$ which has zero correlation with $\mathbf{y_{t-2}}, \ldots$?
  - and so on …
- Up to i = 1 ,…, k such combinations
- “Criterion” and “Root” tables

Identified Model (for k = 3)

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

$$z_{3,t} = u_{3,t} \Rightarrow z_{3,t-1} = u_{3,t-1}$$

$$z_{1,t} = \phi_{13}^{(1)*} z_{3,t-1} + \theta_{13}^{(1)*} u_{3,t-1} + \sum_{j=1}^{2} \phi_{1j}^{(1)*} z_{j,t-1} + u_{1,t} - \sum_{j=1}^{2} \theta_{1j}^{(1)*} u_{j,t-1}$$

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

or

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

Concerns:

(Hannan, Reinsel, Chatfield, Tunnicliffe-Wilson, Ord etc.)

- Transformed series $\mathbf{z}_t$
- # of parameters in **A**
  - Real reduction in parameters is less than 10
- C/C estimates not most efficient
- No standard errors in **A**

Extension to Tiao and Tsay

- Keep the model in terms of $\mathbf{y}_t$

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

- Normalise

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

- Set of rules for determining the free parameters in **A**
  - 3rd equation uniquely identified
  - SCM(0,0) is nested in SCM(1,0) and SCM(1,1)

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

  - SCM(1,0) is nested in SCM(1,1)

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

- Number of parameters reduced by: 10-3 = 7

Checking for Correct Normalisations

- Safeguard against normalising a zero parameter to 1.
- Apply the C/C test to subsets of variables

  Example (continued)

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

Check for individual $y_{1,t}$, $y_{2,t}$ or $y_{3,t}$ ~ SCM(0,0)

- ➢ If yes set that coefficient to one
- ➢ If not check for combinations of two and so on…

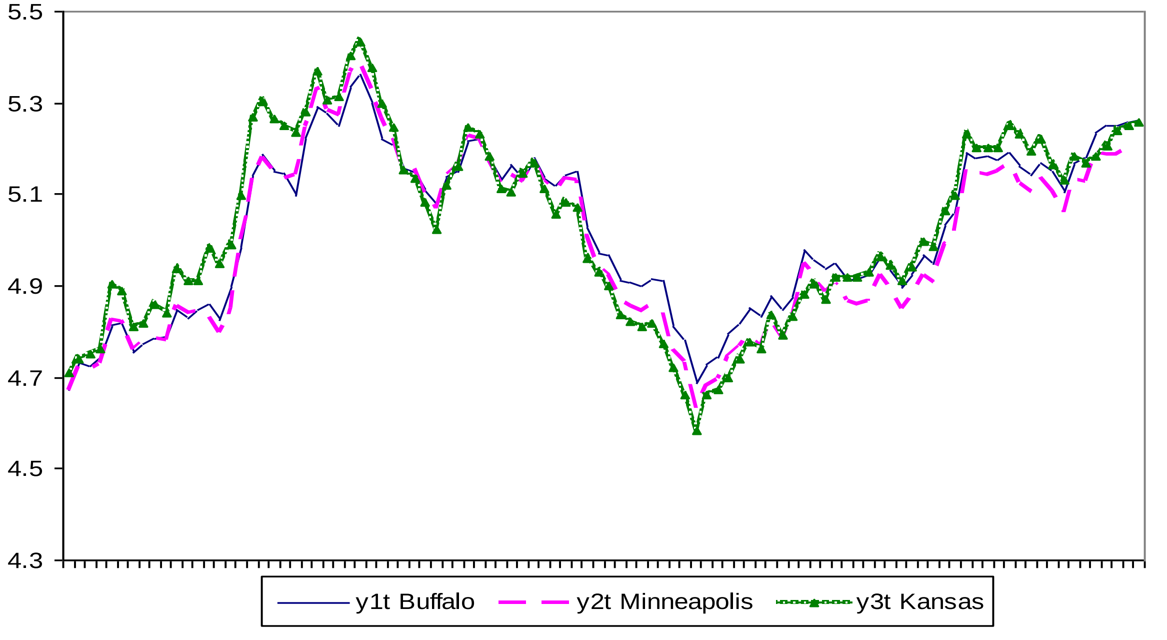
- The number of parameters to be estimated can potentially be further reduced

Summary of Extension

- Keep model in terms of y_t
- Provide set of rules for determining the free parameters parameters in **A**
- Safeguard against normalising a zero parameter to 1
- Estimate the model by FIML

- # Example: US flour price data

Tiao and Tsay (1989), Grubb (1992), Lutkepohl and Poskitt (1996)



- # **Stage I**: Overall Tentative Order

Criterion Table

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

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

VAR(2) or VARMA(1,1)

- **Stage II**: Individual SCMs

Root Table

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

2 ~ SCM(1,0) and 1 ~ SCM(1,1)

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

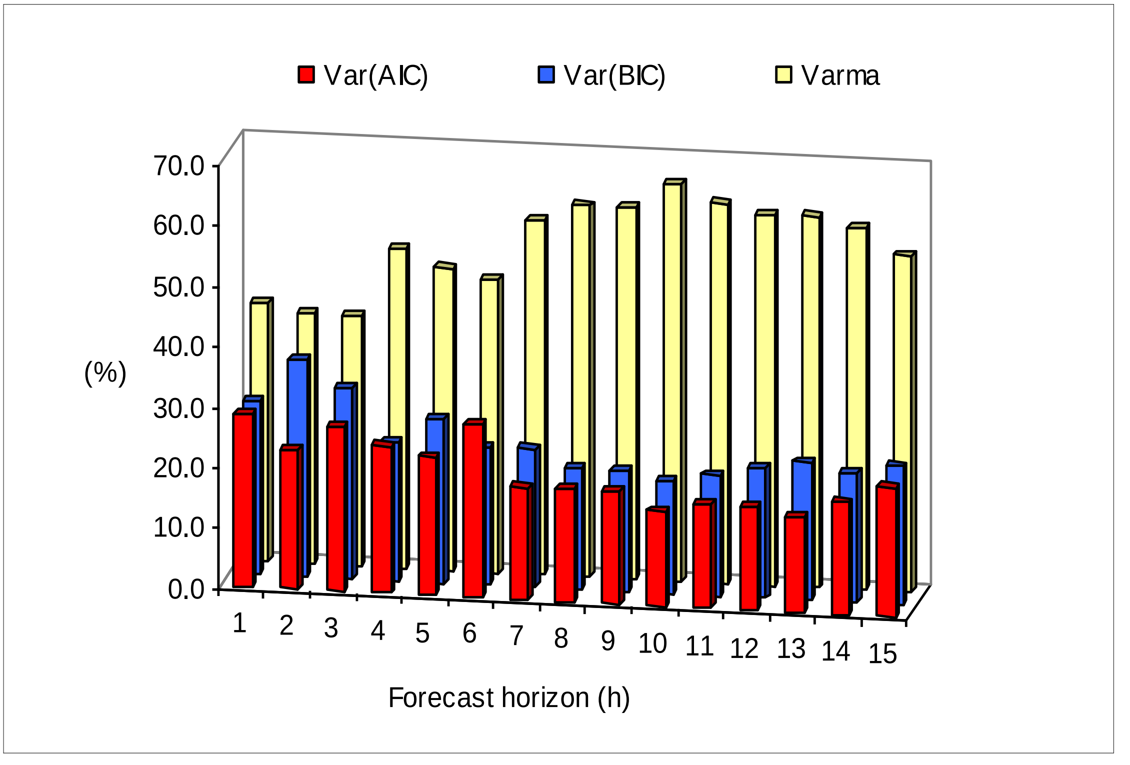
$$\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ \phi_{31}^{(1)} & \phi_{32}^{(1)} & \phi_{33}^{(1)} \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t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

- Normalisations check finds $y_{3,t}$ ~ SCM(1,0)

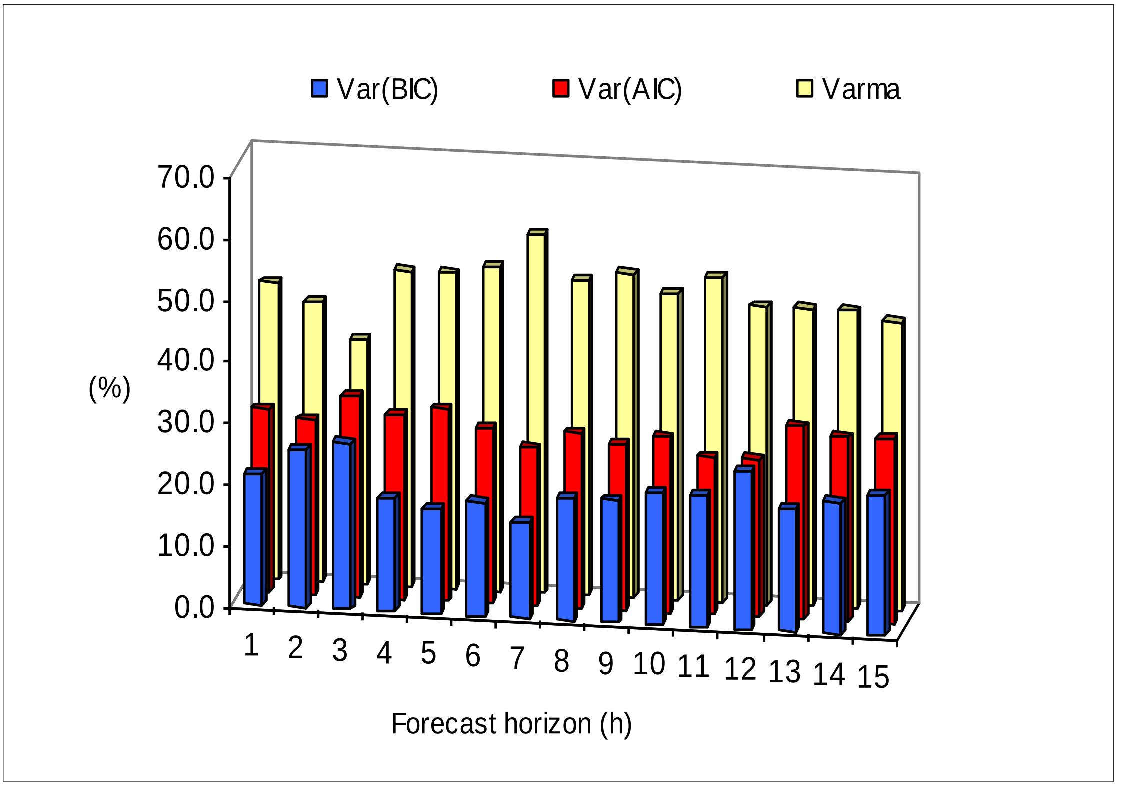
$$\begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{y}_t = \begin{bmatrix} \phi_{11}^{(1)} & \phi_{12}^{(1)} & \phi_{13}^{(1)} \\ \phi_{21}^{(1)} & \phi_{22}^{(1)} & \phi_{23}^{(1)} \\ \phi_{31}^{(1)} & \phi_{32}^{(1)} & \phi_{33}^{(1)} \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t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

- Application to Macro Data
- Data
  - Stock and Watson (1999) & (2001)
  - 40 monthly series 1959:1 – 1998:12 (N=480)
  - 8 major categories:
    - Output and Real Income
    - Employment and Unemployment
    - Consumption, Retail Sales and Housing
    - Real Inventories and Sales
    - Prices and Wages
    - Money and Credit
    - Interest Rates
    - Exchange Rates
  - 70   3 variable systems
    - VARMA
    - VAR selected by AIC and SC
    - Restricted and Unrestricted
- Forecasting and Forecast Evaluation
  - Test sample: $N_1 = 300$
  - Hold-out: $N_2 = 180$
  - h = 1 to 15 step ahead forecasts
  - PB of |FMSE| and tr(FMSE)
  - $Ratio_h = \frac{1}{M} \sum_{i=1}^{M} \frac{|FMSE(VAR)_i|}{|FMSE(VARMA)_i|}$

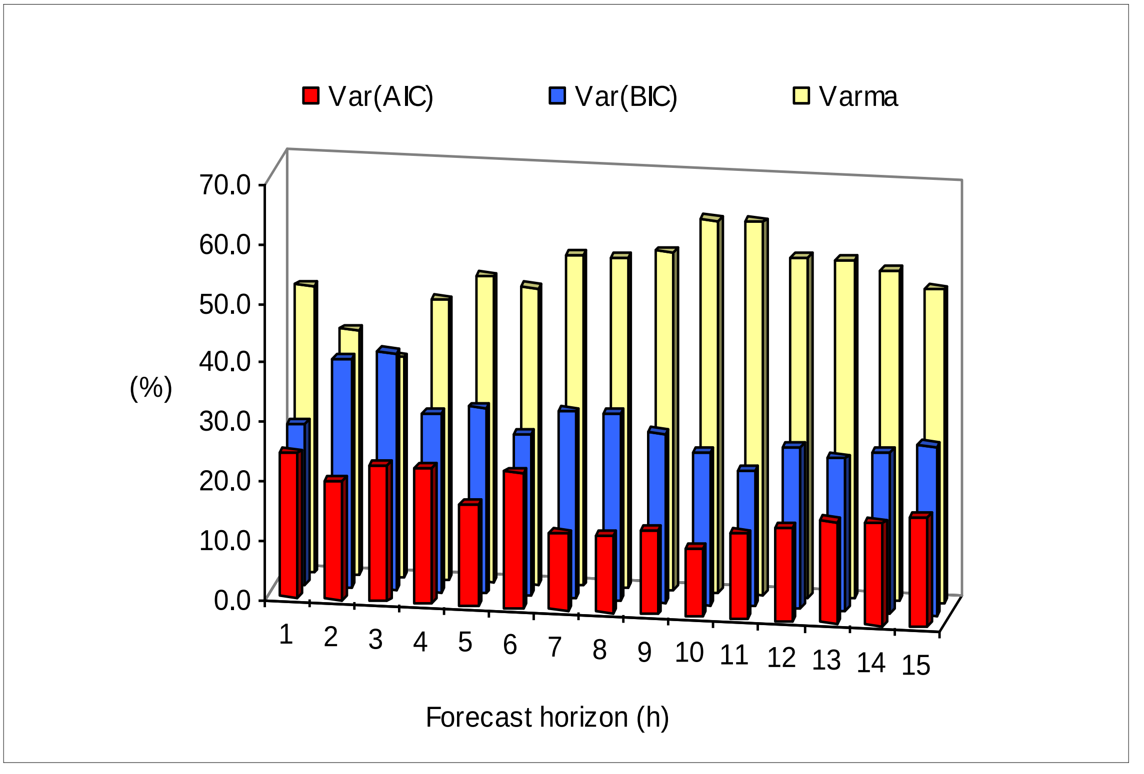
PB counts for |*FMSE*| for the Unrestricted Models



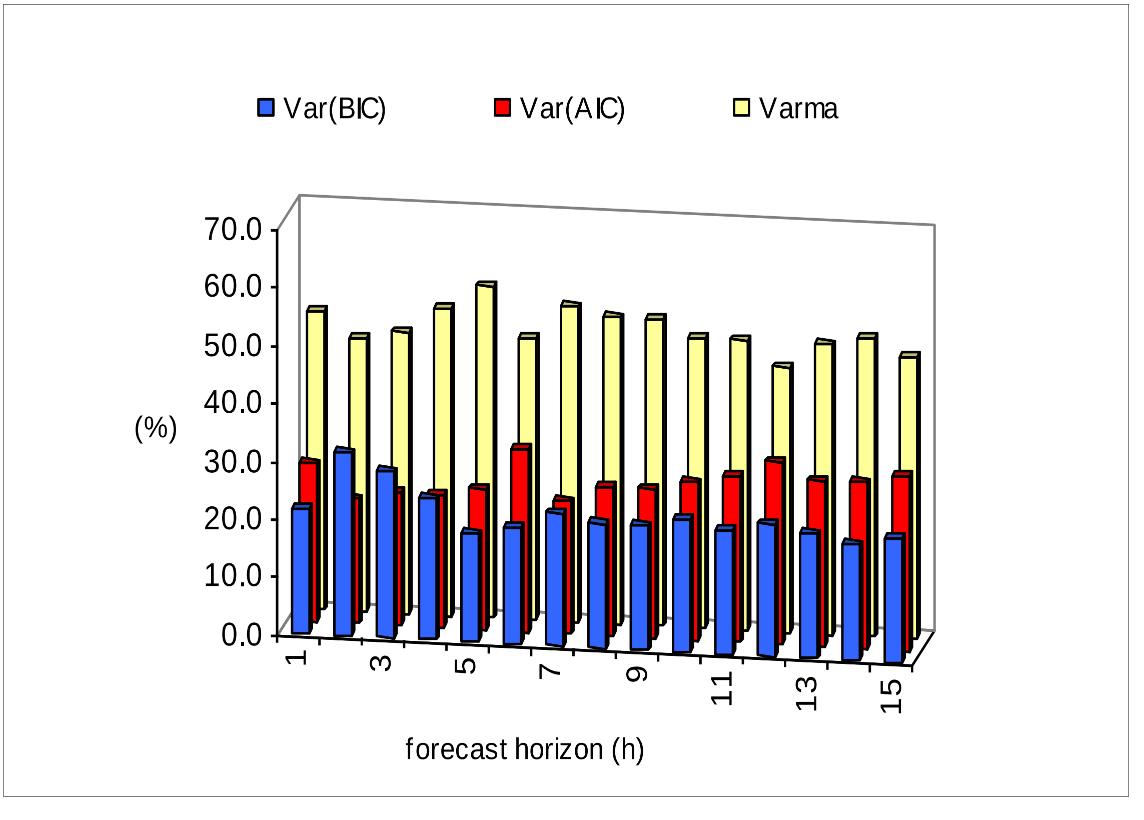
PB counts for |*FMSE*| for the Restricted Models



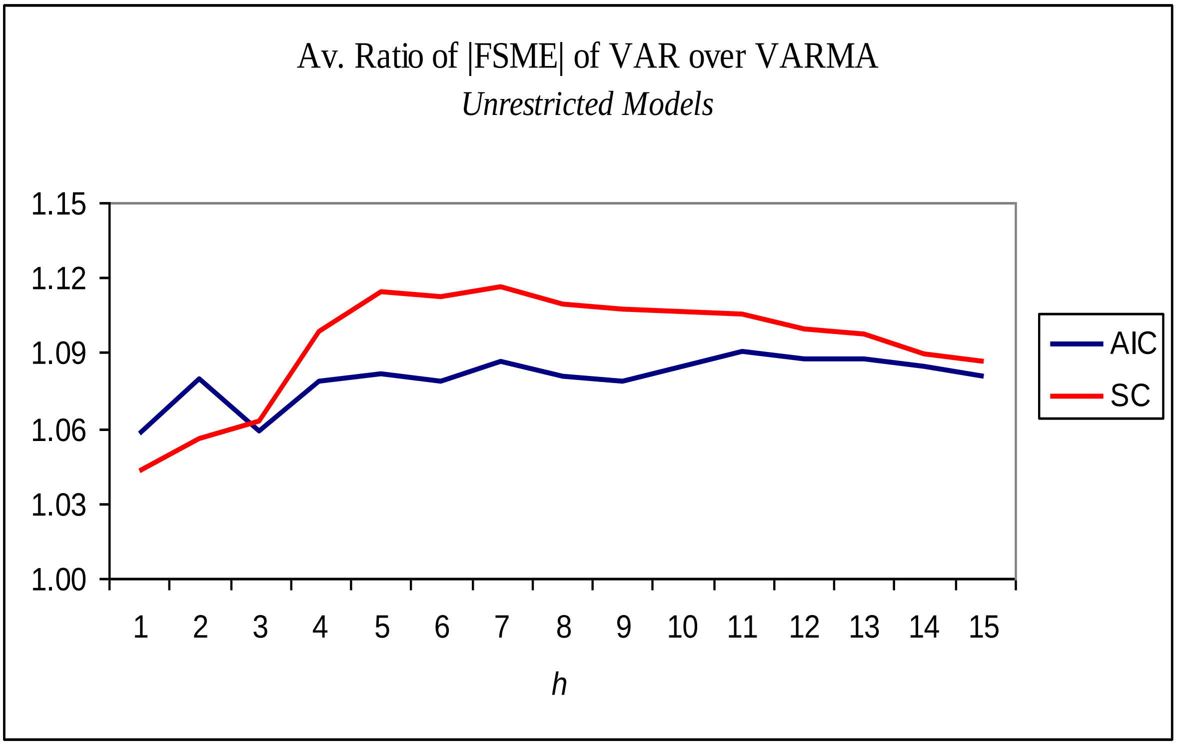
PB counts for $tr(FMSE)$ for the Unrestricted Models



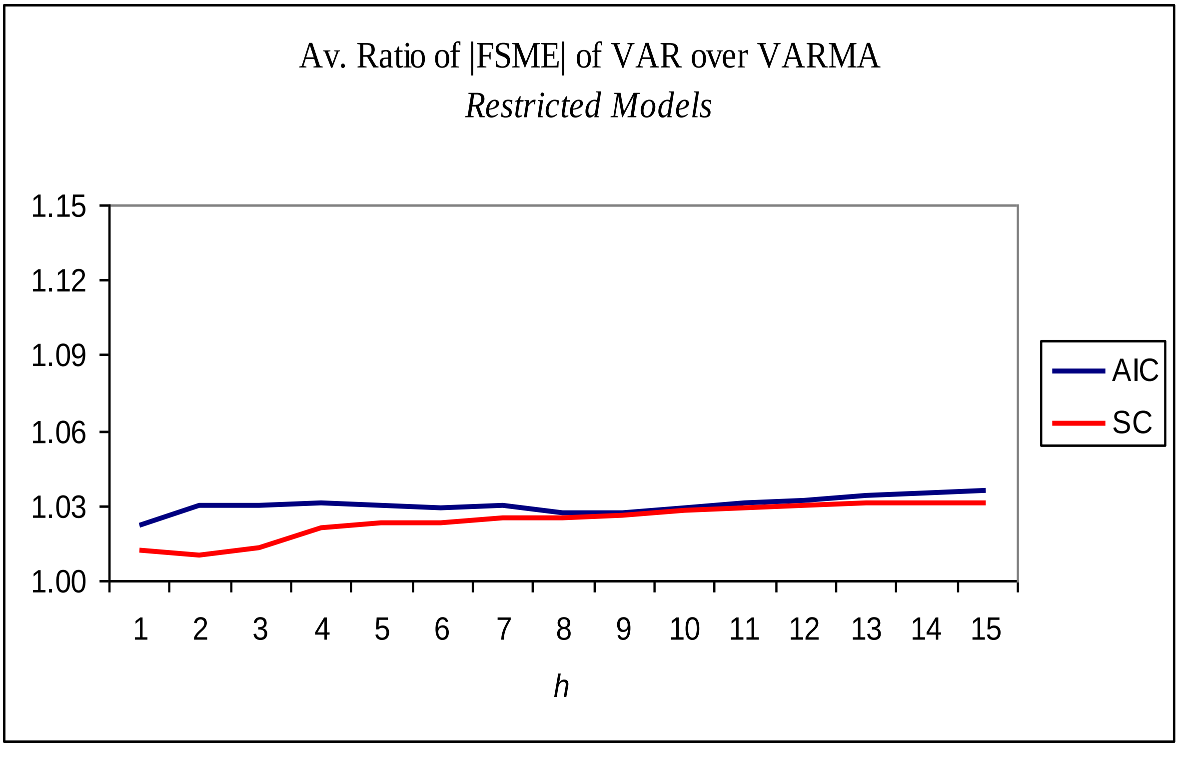
PB counts for $tr(FMSE)$ for the Restricted Models



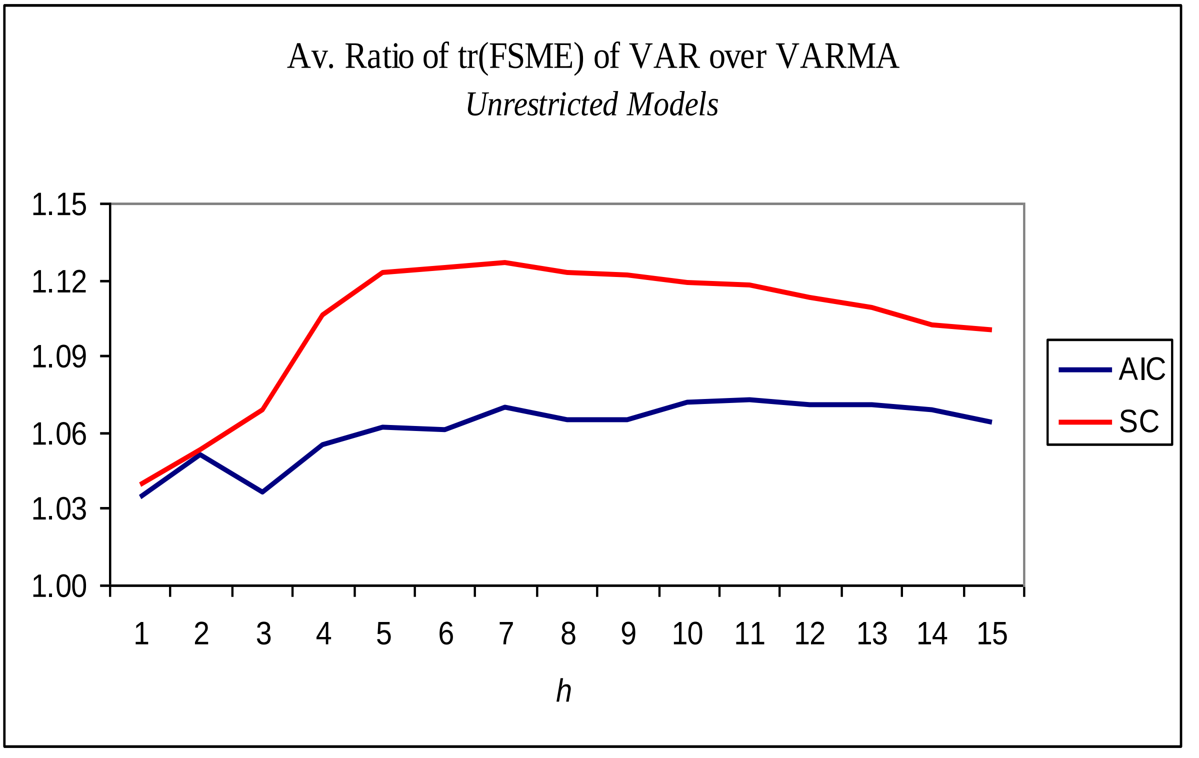
- Summary
- Completed the methodology
    - Addressed issues raised in the discussion of Tiao and Tsay (1989) and
- Simulations
    - Small number of repetitions
    - Good insight
- Empirical Application
    - VARMA in many cases outperform their VAR counterparts
- Future direction
    - Consider larger dimensions
    - Look at Echelon form
    - Study the implications for cointegration analysis

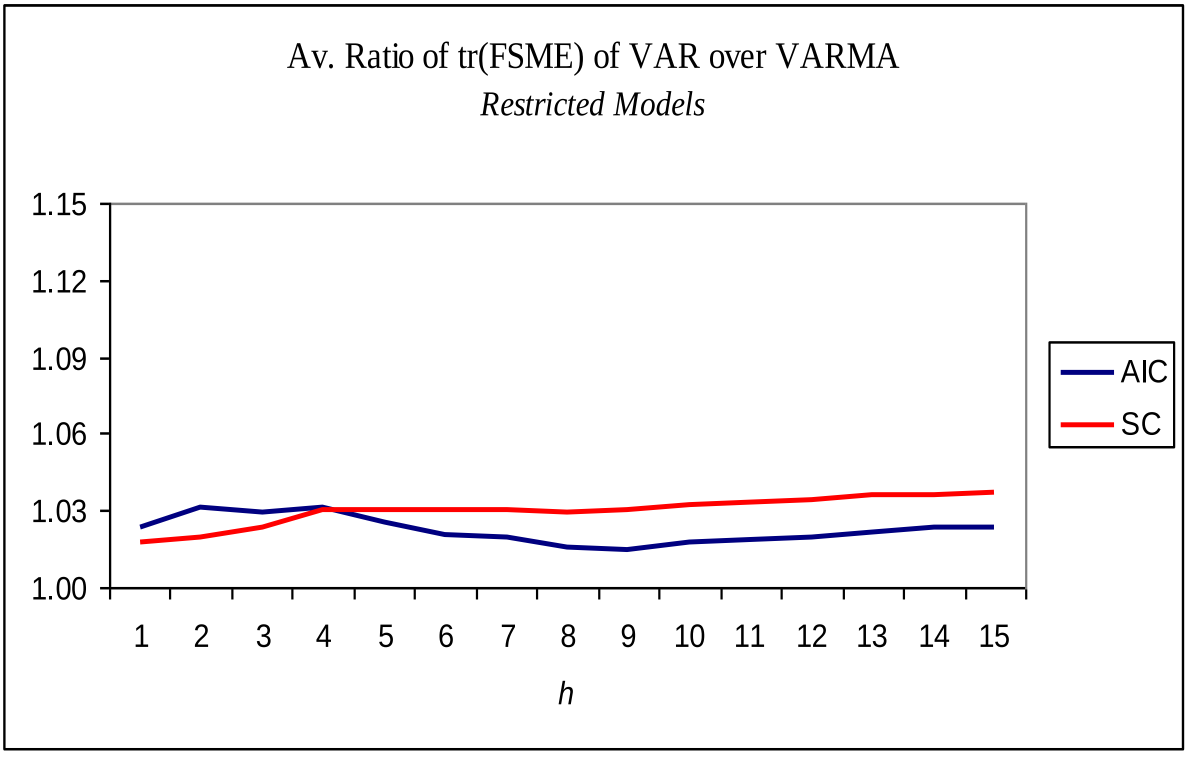


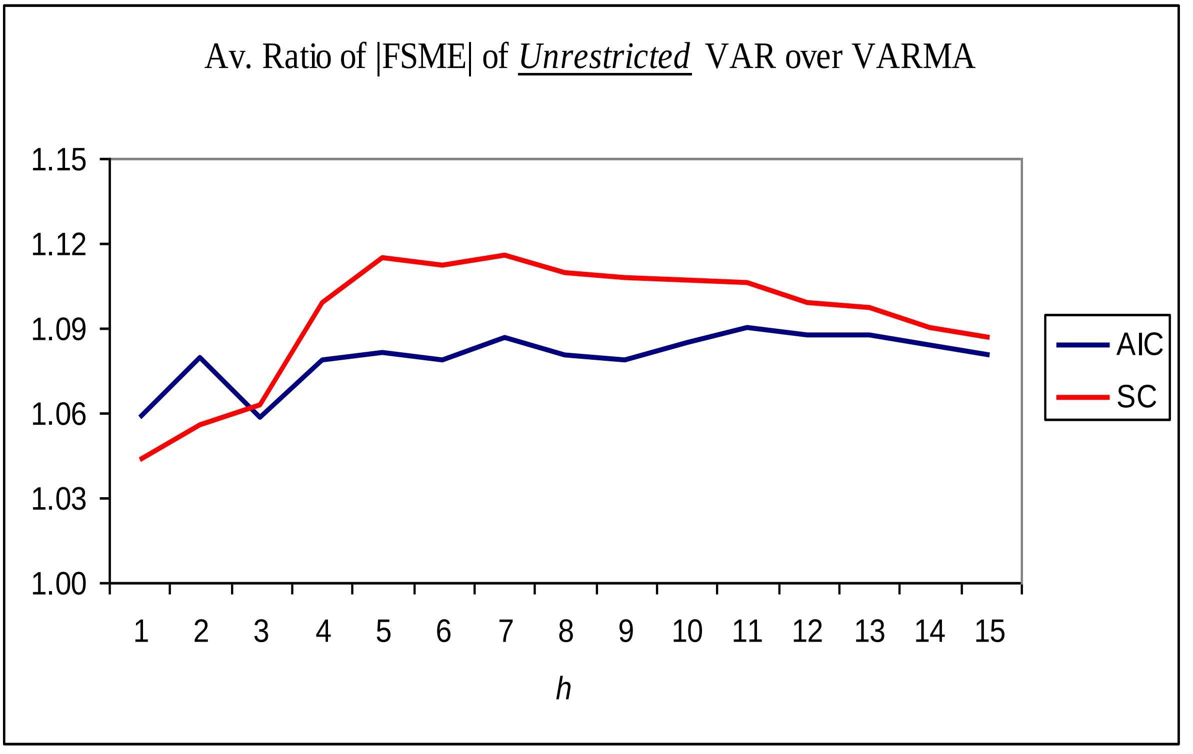
Av. Ratio of |FSME| of VAR over VARMA
Unrestricted Models
1.15
1.12
1.09
1.06
1.03
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
h
AIC
SC

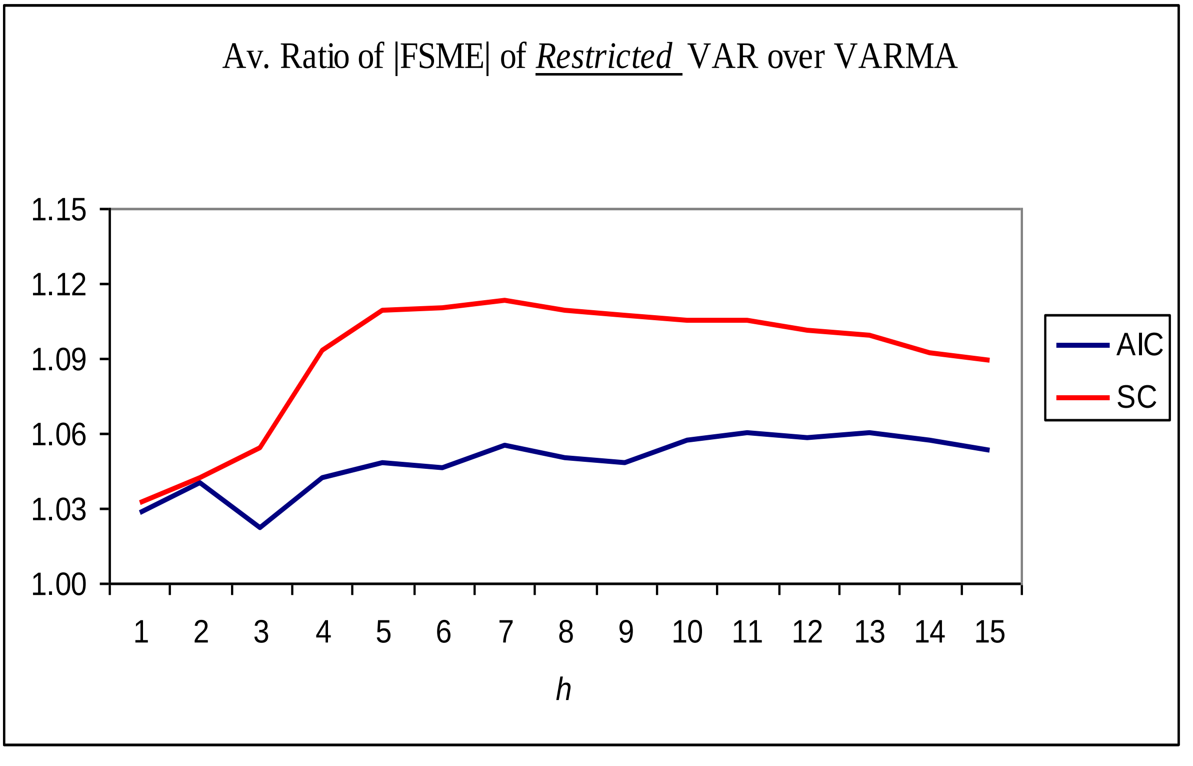


Av. Ratio of |FSME| of VAR over VARMA
Restricted Models
1.15
1.12
1.09
1.06
1.03
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
h
AIC
SC


Av. Ratio of tr(FSME) of VAR over VARMA
Unrestricted Models
1.15
1.12
1.09
1.06
1.03
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
h
AIC
SC


Av. Ratio of tr(FSME) of VAR over VARMA
Restricted Models
1.15
1.12
1.09
1.06
1.03
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
h
AIC
SC


Av. Ratio of |FSME| of Unrestricted VAR over VARMA
1.15
1.12
1.09
1.06
1.03
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
h
AIC
SC


Av. Ratio of |FSME| of Restricted VAR over VARMA
1.15
1.12
1.09
1.06
1.03
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
h
AIC
SC