VARMA Models Identification Using the Infinite VAR Form

CELINA PESTANO-GABINO
Department of Applied Economic
University of La Laguna
Campus de Guajara. La Laguna 38071
SPAIN

Abstract: **-** The aim of this paper is to present a new software package for use in multivariate time series. It is altenative to the software in [8], although they can both be used complementary. Each one provides a practical statistical algorithm to identify VARMA models. The code is written in FORTRAN 90. These methods allows us to answer difficult questions about exchangeable and identifiable models with minimum orders. Both have the advantage of showing the results in easily interpretable tables. The main difference between them is that, following [10], the new software uses coefficients of the infinite VAR form whereas [8] uses autocovariance matrices of the process. This paper provides a demo-version for bivariate stationary and invertible process.

Key-Words: **-** VARMA Models; Specification Stage; Minimum Orders; Identifiability; Infinite VAR Form.

1 Introduction

Analysis of economic and business time series requires that we consider modeling several variables jointly. Vector Autoregressive Moving Average (VARMA) models are effective characterizations for many multivariate time series.

Box and Jenkins (1970) provided a cohesive framework for modeling a univariate time series. The iterative modeling approach consists of three steps: (a) tentative model identification, (b) model estimation and (c) diagnostic checking. This iterative strategy could also be applied to VARMA modeling but new problems arise.

Methods for the identification of simplified VARMA models remain a fundamental area of research at this time. The reduction of the number of parameters involved in VARMA models is an important consideration. Our software[1] is centered on the identification stage of the model. In this stage, the specific problems are: exchangeable models (when several pairs of minimum orders do exist) and non identifiable models (a unique representation given a pair of minimum orders does not exist). The approach to exchangeable models using scalar component models (Tiao and Tsay, 1989) stands out. However, Tiao and Tsay (1989) do not consider the identifiability problem. Identifiable models have been discussed from different points of view by Hannan (1969), Lütkepohl and Poskitt (1996), etc. Other important references in VARMA models can be found in Reinsel (1993), Peña et al. (2001), among others.

The available software for the identification of a mixed VARMA model is not as effective as in the univariate case. We are unware of any proposals in the literature that investigate which pairs of orders are minimum and which pairs of minimum orders have associated identifiable representations. SCA Statistical System (Liu, 1997) provides the available software for forecasting and time series analysis using VARMA models. SCA only uses relatively simple and effective tools to determine the order of a pure vector moving average (VMA) or a pure vector autoregressive (VAR) model (Tiao and Box, 1981).

Pestano and González (2004) provide new answers to difficult questions on identifiability, minimality and exchangeability. They mention several possible ways to analyse the problems through the use of the autocovariance matrices, the coefficients of the infinite VAR form, the coefficients of the infinite VMA form... Pestano and Cruz (2004) presented the software for the first approach (using autocovariance matrices) and this paper offers the software for the second (using coefficients of the infinite VAR form).

Mareschal and Melard (1988) offer the code for *The Corner Method* in the univariate case. The basic tool that they use (the determinant of certain Hankel matrices) is not suitable in the multivariate case. Our software packages are new corner methods

[1] A demo-version for a bivariate process can be requested from the author by e-mail.

to identify VARMA models. The basic tool is the rank of certain Hankel matrices.

This paper in structured as follows. In Section 2 we introduce definitions, questions and results in VARMA models which are needed to understand the utility of the software. In Section 3, we include the instructions to use the program. Moreover, we give suggestions in the art of identifying VARMA models with this software. Finally, Section 4 contains the conclusions.

2 VARMA Models: Identifiability and Minimality Using the Infinite VAR Form

Let X_t, $t \in Z$, be a vector of m random variables. Without loss of generality we assume that $E(X_t)=0$ for any $t \in Z$. We say that X_t admits a VARMA(p,q) representation if

$$X_t+A_1X_{t-1} + ... +A_pX_{t-p} = \varepsilon_t+B_1\varepsilon_{t-1}+...+B_q\varepsilon_{t-q}, \quad (1)$$

for $t \in Z$, where A_i (i=1,...,p) and B_j (j=1,...,q) are mxm matrices, ε_t is a vector white noise process such that $E(\varepsilon_t)=0$, $E(\varepsilon_t\varepsilon'_t)=\Sigma>0$, and $E(\varepsilon_t\varepsilon'_{t+f})=0$ if $f\neq 0$. In particular, if q=0 we have a pure vector autoregressive model, VAR(p); if p=0 we have a pure vector moving average model, VMA(q). We assume that the process is stationary and invertible and we denote its infinite VAR form in the following way (see, for instance, Reinsel,1993):

$$\sum_{i=0}^{\infty} \Pi_i X_{t-i} = \varepsilon_t$$

New concepts arise in the multivariate case.

Definition 1. The orders (p,q) of a VARMA(p,q) model are *minimum orders* (m.o.) if and only if, X_t does not admit a VARMA(p*,q*) representation with $p^*\leq p$ and $q^*<q$ nor $p^*<p$ and $q^*\leq q$.

Definition 2. If a process X_t admits different VARMA representations with m.o., we say that such representations are *exchangeable*.

Definition 3. The VARMA(p,q) representation (1) is *identifiable* if the p+q matrix coefficients A_i (i=1,...,p) and B_j (j=1,...,q) are uniquely determined, i.e., X_t does not admit exchangeable representations with m.o. (p,q).

It must be emphasized that given a pair of m.o., an identifiable representation does not always exist; moreover there may be one for some m.o. but not for others. A non identifiable VARMA(p,q) representation is always a difficult problem in the model estimation stage.

Table 1. Pestano and González (2004) propose the construction of tables with interesting properties. If we use the infinite VAR form, the value

$$T1(i,j)\equiv\text{rank}\left((\Pi_{i-j+h+k-1})_{h,k=1}^{j} \right)$$

is placed in each box (i,j) of *Table 1*, i.e., at the intersection of the i-th column with the j-th row. By convention, T1(i,0)=0 for any $i \in \mathbb{N}$.

Definition 4: We say that

R1={(i,j) / T1(i,j)=T1(i+k,j+k) for any integer k>0}

is the *staired block* of the *Table 1.*

The boxes in a particular diagonal of R1 do have the same value, however boxes in different diagonals can have different values.

Definition 5: (i,j) is the *corner* of a step in R1 iff $T1(i,j)\in R1$, $T1(i\text{-}1,j)\notin R1$ and $T1(i,j\text{-}1)\notin$ *R1*.

2.1 Properties of *Table 1* Using the Infinite VAR Form

Slight variations are apparent when comparing the properties of *Table 1* with those in Pestano and Cruz (2004). In Pestano and Cruz (2004) columns of *Table 1* indicate VMA orders and rows indicate VAR orders. Here columns indicate VAR orders and rows indicate VMA orders.

Property 1. X_t admits a VARMA representation iff $R1\neq\emptyset$.

In particular situations, the following properties could guarantee that a pair of orders associated with a corner of R1 is (or is not) a pair of m.o. (Properties 2-7) and that a VARMA representation is (or is not) identifiable for a given pair of m.o. (Properties 8-9).

Property 2. $(i,0)\in R1$ and $(i\text{-}1,0)\notin R1$ iff (i,0) is a pair of m.o.

Property 3. If T1(i-1,j)<T1(i,j) then (i-u,j-v) is not a pair of m.o., with $1\le u\le i$ and $0\le v\le j$.

Property 4. If T1(i,j)=jm and $(i,j)\notin R1$ then (i-u,j-v) is not a pair of m.o., with $0\le u\le i$ and $0\le v\le j$.

Property 5. If $(i,j)\in R1$, $(i\text{-}1,j)\notin R1$ and T1(i,j)=jm then (i-u,j-v) is not a pair of m.o., with $1\le u\le i$ and $0\le v\le j$.

Property 6. $(0,j)\in R1$ and $(0,j\text{-}1)\notin R1$ iff (0,j) is a pair of m.o.

Property 7. If $(i,j)\in R1$, $(i\text{-}u,j)\notin R1$, $(i,j\text{-}v)\notin R1$ and (i-u,j) and (i,j-v) are not pairs of m.o., with $1\le u\le i$ and $1\le v\le j$, then (i,j) is a pair of m.o.

Property 8. If $(i,j)\in R1$, T1(i,j)=jm iff the VARMA(i,j) representation is identifiable.

Property 9. If $(0,j)\in R1$ then T1(0,j)=jm and the VARMA(0,j) representation is identifiable.

3 Software Demonstration

We have chosen the following stationary and invertible models in order to illustrate the software package VARINF.EXE.

DATOS1.DAT contains data of one simulation with 347 observations of the model

$$X_t = \varepsilon_t + \begin{pmatrix} 1/2 & 15/4 \\ 3/4 & -1/2 \end{pmatrix}\varepsilon_{t-1} + \begin{pmatrix} -7/16 & -7/4 \\ 7/8 & 7/2 \end{pmatrix}\varepsilon_{t-2}$$

DATOS2.DAT contains data of one simulation with 347 observations of the model

$$X_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\varepsilon_t + \begin{pmatrix} 1 & 0.5 \\ 0 & 0 \end{pmatrix}\varepsilon_{t-1} + \begin{pmatrix} 0 & 0 \\ 0.5 & 0.25 \end{pmatrix}\varepsilon_{t-2}$$

In both examples, $\Sigma = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$ and X_t admits a VARMA(0,2) and a VARMA(1,1) representation, both of which are identifiable and with m.o. Note that the VARMA(1,1) representation for the first example is

$$X_t + \begin{pmatrix} 1/2 & 1/4 \\ -1 & -1/2 \end{pmatrix}X_{t-1} = \varepsilon_t + \begin{pmatrix} 1 & 4 \\ -1/4 & -1 \end{pmatrix}\varepsilon_{t-1}$$

and for the second is

$$X_t + \begin{pmatrix} 0 & 0 \\ -1/2 & 0 \end{pmatrix}X_{t-1} = \varepsilon_t + \begin{pmatrix} 1 & 1/2 \\ -1/2 & 0 \end{pmatrix}\varepsilon_{t-1}.$$

Moreover, the first example admits the following VARMA(2,0) representation:

$$X_t + \begin{pmatrix} -1/2 & -15/4 \\ -3/4 & 1/2 \end{pmatrix}X_{t-1}\begin{pmatrix} 7/2 & 7/4 \\ -7/8 & -7/16 \end{pmatrix}X_{t-2} = \varepsilon_t.$$

3.1 Instructions to Obtain *Table 1* Using the Infinite VAR Form

Two files are needed: VARINF.EXE and DATOS. DAT.

DATOS.DAT must contain data from one or several examples with the same sample size (in the demo-version the maximum sample size is 350). The data for each variable must be located in a different column of 14 possible positions. The data for each example must be located following the last one.

Run VARINF.EXE. It reads the data (DATOS.DAT) and saves two files (COEVAR.DAT and SALIDA.DAT). The program asks for:

- The number of examples in DATOS.DAT.
- The sample size (in this demonstration, 347).
- The significance level: normally we choose 95%. For the statistical procedure see Pestano and González (2004).

VARINF estimates:

a) P_1, P_2, ... , P_{11}, the covariance matrix of the estimated parameters and the covariance matrix of ε_t. They are saved in COEVAR.DAT. (It estimates the r+c+3 coefficients of a VAR(r+c+3) by Least Squares Estimation, where r+1 and c+1 are the number of rows and columns, respectively, in *Table 1*. In this demostration r=c=4).

b) Table 1. It is shown on the screen and saved in the file SALIDA.DAT for all the examples

In this demonstration, SALIDA.DAT for DATOS1.DAT must contain the following estimated *Table 1*

	0	1	2	3	4
0	0	0	0	0	0
1	2	2	1	0	0
2	4	3	2	1	0
3	6	4	3	2	1
4	8	6	4	3	2

and SALIDA.DAT for DATOS2.DAT the following

	0	1	2	3	4
0	0	0	0	0	0
1	2	1	1	1	1
2	4	3	2	1	1
3	6	4	3	2	1
4	8	6	4	3	2

We have outlined the border of the staired block R1 in each one.

Table 1 for DATOS1.DAT indicates that the data admits the exchangeable models VARMA(0,2), (VARMA(1,1) and VARMA(2,0). From properties of *Table 1*, they are identifiable and with m.o.

Table 1 for DATOS2.DAT indicates that the data admits a VARMA(0,2) model exchangeable with a VARMA(2,1). From Property 8 the VARMA(2,1) is not identifiable. Properties of *Table 1* could not guarantee that (2,1) is a pair of m.o. Then, in this example *Table 2* is useful (see Pestano and González, 2004).

3.2 The Art of Identifying VARMA Models

As we know, in Statistics, there is certain distance between theory and practice since tables show estimated values. We comment certain slight error which may arise.

From Proposition 7[2] or 8[3], we could decide that (i,j)∈ R1 and (i+1,j+1)∈ R1 although T1(i,j)≠T1(i+1,j+1) in an estimated *Table 1*. For instance, from Property 7, (1,3) and (2,4) must belongs to R1 if T1(2,2)=T1(3,3) and T1(1,2)=T1(2,3) although T1(1,3) ≠T1(2,4) in an estimated table.

Note that, from Proposition 8, for instance T1(1,1)=T1(4,4) implies that:

T1(1,1)=T1(2,2)=T1(3,3),
T1(2,1)=T1(3,2)=T1(4,3),
T1(3,1)=T1(4,2)=T1(5,3),
T1(4,1)=T1(5,2)=T1(6,3),
T1(1,2)=T1(2,3)=T1(3,4),
T1(1,3)=T1(2,4)=T1(3,5),
T1(1,4)=T1(2,5) and
T1(1,5)=T1(2,6).

Graphically:

	0	1	2	3	4	5	6
0							
1		a	b	c	d	e	
2		f	a	b	c	d	e
3		g	f	a	b	c	
4		h	g	f	a		
5		i	h	g			
6			i				

Therefore, it is advisable to ignore this type of error and to outline the border of R1.

4 Conclusions

This paper has presented a software package that can be applied in multivariate time series modeling. It allows us to *characterize* VARMA models, to recognize the *exchangeable* models with m.o. that might exist and to detect *identifiable* and non identifiable representations. The advantages of this package, as compared to available ones, are similar to those in Pestano and Cruz (2004), namely:

- The procedure is relatively more direct than the methods used by other authors.

[2] **Proposition 7 (Pestano and González, 1997):**
If rank(M1(i,j))=rank(M1(i+1,j+1)), then rank(M1(i+1,j))=rank(M1(i+2,j+1)), if and only if rank((M1(i,j+1))=rank(M1(i+1,j+2)).

[3] **Proposition 8 (Pestano and González, 1997):**
If rank(M1(i,j))=rank(M1(i+2,j+2)), then:

a) rank(M1(i,j))=rank(M1(i+1,j+1))
b) rank(M1(i+1,j))=rank(M1(i+2,j+1))
c) rank(M1(i,j+1))=rank(M1(i+1,j+2))
d) rank(M1(i+2,j))=rank(M1(i+3,j+1))
e) rank(M1(i,j+2))=rank(M1(i+1,j+3))

- The algorithm does not require knowledge of the matrix coefficients that appear in the model (1), only the data.

- The results are presented in an easyly interpretable table.

- We give further information about identifiable and exchangeable representations.

This package and the one presented in Pestano and Cruz (2004) are complementary because the model influences which algorithm is better than the orther. Nevertheless, both packages provide satisfactory empirical results.

Acknowledgments

We are grateful to Concepción González. This paper had not been possible without her.

References:

[1] Box, G.E.P. and Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970 (Revised edition published 1976).

[2] Hannan, E.J., The Identification of Vector Mixed Autoregressive-Moving Average Systems, *Biometrica* 56, 1969, pp. 223-225.

[3] Liu, L.-M., *Forecasting and Time Series Analysis Using the SCA Satatistical System*, Vol. 2, Oak Brook, IL: Scientific Computing Associates Corp, 1997.

[4] Lütkepohl, H., *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag, 1991.

[5] Lütkepohl, H. & Poskitt, D. S., Specification of Echelon-Form VARMA Models, *Journal of Business & Economic Statistics* 14 (1), 1996, pp. 69-79.

[6] Mareschal, B. and Melard, G., The Corner Method for Identifying Autoregressive Moving Average Models, *Applied Statistics*, Vol. 37, No. 2, 1988, pp. 301-316.

[7] Peña, D., Tiao, G.O. and Tsay, R.S., *A Course in Time Series Analysis*, New York: John Wiley & Sons, Inc., 2001.

[8] Pestano, C. and Cruz, D.I. A Practical Algorithm to Identify VARMA Models for Economic Data, *WSEAS Transactions on Business and Economics*, Issue 1, Vol. 1, 2004, pp. 13-17.

[9] Pestano, C. and González, C., Matrix Padé Approximation of Rational Functions, *Numerical Algorithms* 15, 1997, pp. 1-26.

[10] ___________________________ Identifiability and Minimality in Rational Models, 2004 (submitted).

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

[12] Tiao, G. C. and Box, G. E. P. Modeling Multiple Time Series with Applications. *Journal of the American Statistical Association* 76, 1981, pp. 802-816.

[13] Tiao, G. C.; Tsay, R. S., Model Specification in Multivariate Time Series, *Journal of the Royal Statistical Society* B 51, No. 2, 1989, pp. 157-213.