

## Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

FACULTAD DE ECONOMICAS

Campus de Somosaguas

**28223 MADRID** 

Teléfono 394 26 11 - FAX 294 26 13







# Documento de trabajo

A Generalized Least Squares Estimation Method for Vector Moving Average Models

Rafael Flores de Frutos Gregorio R. Serrano

No.9611

Septiembre 1996



Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

## A GENERALIZED LEAST SQUARES ESTIMATION METHOD FOR VECTOR MOVING AVERAGE MODELS

Rafael Flores de Frutos Gregorio R. Serrano

Instituto Complutense de Análisis Económico Universidad Complutense de Madrid Campus de Somosaguas 28223 Madrid

## ABSTRACT

A new GLS procedure for estimating VMA models is proposed. Its main feature is to consider explicitly the stochastic structure of the approximation errors arising when lagged VMA innovations are replaced with lagged residuals from a long VAR.

## RESUMEN

Se propone un nuevo método lineal para la estimación de modelos VMA. Este método tiene como característica principal, la de considerar explicitamente la estructura estocástica de los errores de aproximación que se cometen al sustituir las innovaciones del VMA por residuos obtenidos a partir de la estimación de un VAR de orden elevado.

JEL classification: C32

N°-E: 5306620197 N°-C: x-53-231150-6

#### 1. INTRODUCTION

Moving average terms in multivariate models complicate specification and estimation procedures. Thus, while it is recognized that in some situations pure VMA or mixed VARMA models can produce more efficient forecasts than an appropriate finite order VAR approximation, the fact is that VAR models have dominated the empirical work.

Simplifying the task of elaborating VMA and VARMA models has been the goal of many authors, see for instance: Hannan and Rissanen (1982), Spliid (1983) and Koreisha and Pukkila (1989) among others. These authors have developed linear estimation procedures with some desirable features:

- i) They are easy to implement. Most of them only require a standard least squares (LS) routine.
- ii) They are fast. Either no iterations or just a few are needed for obtaining accurate estimates, comparable with that of maximum likelihood (ML) methods. Further, for the univariate case Koreisha and Pukkila (1990) have found that their generalized least squares (GLS) procedure: (1) yields accurate estimations even when short samples are used, (2) seldom generates non-invertible or non-stationary situations, and (3) performs better than ML when a pure moving average (MA) processes is needed to fit a short sample.
- iii) Fast estimation procedures have proved to be quite effective in detecting non-zero parameters.
- iv) The use of these estimates in order to initialize more efficient estimation procedures as exact ML, reduces the number of iterations needed for convergence.

In this paper a new linear method for estimating pure VMA models is proposed¹. We use an idea introduced by Koreisha and Pukkila (1990) in the univariate context, i.e. that there is an approximation error when replacing, in the original VARMA model, lagged innovations with lagged residuals from a long VAR. But instead of using their white noise assumption for the approximation error, we derive its exact stochastic structure, which depends on "L", the order of the long VAR, as well as on "q", the order of the VMA model. We show how the VMA(q+L-1) structure of the epproximation errors will induce a VMA(2q+L) structure in the noise of the transformed model, instead of a VMA(q), as implied by Koreisha and Pukkila's (1990) assumption. Our method not only generalized those proposed in Koreisha and Pukkila (1989) but also that in Koreisha and Pukkila (1990) for univariate models.

The paper is organized as follows. Section 2 describes our proposed GLS approach to estimating VMA processes. Section 3 presents our simulation results. Finally, Section 4 concludes.

## 2. A NEW GLS APPROACH TO ESTIMATING VMA MODELS

Consider the  $k{\times}1$  vector  $z_t$  of time series following the invertible VMA process:

$$z_{t} = \theta_{g}(B) a_{t}$$
 (1)

 $t=1,\,2,\,\dots\,N,$  where  $\theta_q(B)=I-\theta_1B-\dots-\theta_qB^q$  is a k×k finite order (q) polynomial matrix

in the lag operator B, with the roots of  $|\theta(B)| = 0$  lying all outside the unit circle. The  $k \times 1$  vector  $a_t$  is assumed to follow a white noise process with covariance matrix  $\Sigma_a$ .

Two useful alternative representations of process (1) for the whole sample are:

## Alternative Representation 1:

$$vec(Z_N) = D_{\theta N} vec(A_N) + G_{\theta N} vec(A^*)$$
 (2)

## Alternative Representation 2:

$$vec(Z_{N}) = [A' \otimes I_{N}] vec(\Theta) + vec(A_{N})$$
 (3)

where:

$$Z_{N} = [z_{1} \ z_{2} \ ... \ z_{N}]_{(k \times N)}$$
 (4)

$$A_{N} = [a_{1} \ a_{2} \ \dots \ a_{N}]_{(k \times N)}$$
 (5)

$$A^* = [a_{-q+1} \ a_{-q+2} \ \dots \ a_0]_{(k \times q)}$$
 (6)

$$\Theta = [-\theta_1 - \theta_2 \dots - \theta_n]_{(k \times nk)}$$
 (8)

Provided that a finite order VAR(L) is an appropriate approximation to (1), consistent estimates of the elements in A can be obtained from residuals (û,) in the regression:

These estimates can be used for estimating  $vec(\Theta)$  in (3). Using either the first iteration of Spliid's (1983) method or the extension of Hannan and Rissanen (1982) procedure, proposed by Koreisha and Pukkila (1989), consistent estimates of parameters and

<sup>&</sup>lt;sup>1</sup> The extension to mixed VARMA models is in progress.

$$A = \begin{cases} a_0 & a_1 & \dots & a_{N-1} \\ a_{-1} & a_0 & \dots & a_{N-2} \\ \dots & \dots & \dots & \dots \\ a_{-q+1} & a_{-q+2} & \dots & a_{N-q} \end{cases}_{(qk \times N)}$$
(9)

$$G_{\theta,N} = \begin{bmatrix} -\theta_{q} & -\theta_{q-1} & \dots & -\theta_{2} & -\theta_{1} \\ 0 & -\theta_{q} & \dots & -\theta_{3} & -\theta_{2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\theta_{q} & -\theta_{q-1} \\ 0 & 0 & \dots & 0 & -\theta_{q} \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}_{\text{(KN)ode)}}$$

$$(10)$$

$$Z_{t} = \sum_{j=1}^{L} \pi_{j} Z_{t-j} + u_{t}$$
 (11)

their associated standard errors can be obtained by applying LS to the model:

$$vec(Z_N) = [\hat{U}' \otimes I_k] vec(\theta) + vec(A_N)$$
 (12)

where

$$\hat{\mathbf{U}} = \begin{pmatrix} \hat{\mathbf{u}}_0 & \hat{\mathbf{u}}_1 & \dots & \hat{\mathbf{u}}_{N-1} \\ \hat{\mathbf{u}}_{-1} & \hat{\mathbf{u}}_0 & \dots & \hat{\mathbf{u}}_{N-2} \\ \dots & \dots & \dots & \dots \\ \hat{\mathbf{u}}_{-q+1} & \hat{\mathbf{u}}_{-q+2} & \dots & \hat{\mathbf{u}}_{N-q} \end{pmatrix}_{(qk \times N)}$$
(13)

Alternatively, Koreisha and Pukkila (1989) propose the Double Regression (DR) method, as LS to the model:

$$\operatorname{vec}(Z_{N}) - \operatorname{vec}(\hat{U}_{N}) = [\hat{U}' \otimes I_{k}] \operatorname{vec}(\Theta) + \operatorname{vec}(E_{N})$$
 (14)

where

$$\hat{\mathbf{U}}_{\mathbf{N}} = [\hat{\mathbf{u}}_{1} \ \hat{\mathbf{u}}_{2} \ \dots \ \hat{\mathbf{u}}_{\mathbf{N}}]_{\mathbf{u} \times \mathbf{N}} \tag{15}$$

Approximate standard errors for DR estimates can be calculated with the same formula used in calculating standard errors in (12), see Koreisha and Pukkila (1989):

$$\widehat{\text{Cov}}(\text{vec}(\hat{\theta})) = [(\widehat{\mathbf{U}}\widehat{\mathbf{U}}')^{-1} \otimes \widehat{\boldsymbol{\Sigma}}_{\hat{\alpha}}]$$
 (16)

Note that none of above methods take into account the fact that an approximation error will be committed when replacing lagged innovations  $a_{ij}$  with lagged residuals  $\hat{u}_{ij}$ . Noises in transformed models (12) and (14) are implicitly assumed to be white noise processes. We will show that approximation errors [noise in (14)] will follow a VMA(L+q-1) structure, implying a VMA(2q+L) process for noise in (12).

Process (1) will have an infinite VAR representation:

$$z_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t$$
 (17)

LS to (11) gives:

$$Z_{t} = \sum_{j=1}^{L} \hat{\pi}_{j} Z_{t-j} + \hat{u}_{t}$$
 (18)

By subtracting (17) from (18):

$$a_1 = \hat{u}_1 + S_{11} - S_{22} = \hat{u}_1 + \epsilon_2$$
 (19)

where

$$S_{1t} = \sum_{i=1}^{L} \hat{\pi}_i Z_{t-i}$$

$$S_{2t} = \sum_{i=1}^{\infty} \pi_i Z_{t-i}$$

$$\epsilon_i = S_{1t} - S_{2t}$$
(20)

In Koreisha and Pukkila (1989),  $\epsilon_i$  is assumed to be zero. In Koreisha and Pukkila (1990),  $\epsilon_i$  is assumed to be a white noise vector.

S, can be expressed as:

$$S_{R} = \hat{\Pi} R_{\theta} A_{t-1,t-(n+1)}$$
 (21)

where

$$\hat{\Pi} = [\hat{\pi}_1 \ \hat{\pi}_2 \ \dots \ \hat{\pi}_L] \tag{22}$$

$$A_{t-1,t-(q^{*}1.)} = \begin{pmatrix} a_{t-1} \\ a_{t-2} \\ \vdots \\ a_{t-(q^{*}1.)} \end{pmatrix}_{(k(q^{*}1.)\times 1)}$$
(24)

or

$$S_{11} = \psi_1^{s1} a_{t-1} + \psi_2^{s1} a_{t-2} + \dots + \psi_{L+q}^{s1} a_{t-(L+q)}$$
 (25)

where

$$\hat{\Pi} R_a = [ \psi_1^{s1} \psi_2^{s1} \dots \psi_{1+n}^{s1} ]_{n \sim kn \text{ (26)}}$$

From (1)  $S_{21}$  can be expressed as:

$$S_{2i} = E_{i-1}(z_i) = -\theta_1 a_{i-1} - \theta_2 a_{i-2} - \dots - \theta_n a_{i-n}$$
 (27)

Then

$$\epsilon_{t} = \psi_{1}^{\epsilon} a_{t-1} + \psi_{2}^{\epsilon} a_{t-2} + \dots + \psi_{L+q}^{\epsilon} a_{t-(L+q)}$$
(28)

where

$$[\psi_1^{\epsilon} \ \psi_2^{\epsilon} \ \dots \ \psi_{L+q}^{\epsilon}] = [\ \psi_1^{s1} \ \psi_2^{s1} \ \dots \ \psi_{L+q}^{s1}] - [-\theta_1 \ \dots \ -\theta_q \ 0 \ \dots \ 0]$$
 (29)

Note that:

$$\psi_{\mathbf{k}}^{\varepsilon} = 0 \quad \forall \ \mathbf{k}, \quad \text{if } \pi_{\mathbf{i}} = \hat{\pi}_{\mathbf{i}} \quad \forall \ \mathbf{j}$$

In that case  $\epsilon_t = 0$  and  $a_t = \hat{u}_t$ .

In general, finite samples estimation errors will make weights  $\pi_i$  to differ from their LS estimates, and therefore they will induce in  $\epsilon_i$  the VMA(L+q-1) structure (28).

Taking into account (28), process (1) becomes:

$$z_{t} = -\theta_{1} \hat{u}_{t-1} - \theta_{2} \hat{u}_{t-2} - \dots - \theta_{q} \hat{u}_{t-q} + \eta_{t}$$

$$\eta_{t} = a_{t} - \psi_{1} a_{t-1} - \psi_{2} a_{t-2} - \dots - \psi_{2a+L} a_{t-(2a+L)}$$
(31)

where

$$(-\psi_1 \mathbf{B} - \dots - \psi_{2n+L} \mathbf{B}^{2n+L}) = (-\theta_1 \mathbf{B} - \dots - \theta_n \mathbf{B}^n)(\psi_1^{\epsilon} \mathbf{B} + \dots + \psi_{n+L}^{\epsilon} \mathbf{B}^{n+L})$$
(32)

Representation 2 for (31) is:

$$\operatorname{vec}(Z_{N}) = (\hat{U}' \otimes I_{k}) \operatorname{vec}(\Theta) + \operatorname{vec}(H_{N})$$

$$\operatorname{vec}(H_{N}) = D_{\psi N} \operatorname{vec}(A_{N}) + G_{\psi N} \operatorname{vec}(A^{**})$$
(33)

with

$$H_{N} = [\eta_{1} \ \eta_{2} \ \dots \ \eta_{N} \ ]_{(k \times N)}$$

$$A^{**} = [a_{-2a+L+1} \ \dots \ a_{0} \ ]_{(k \times 2a+L)}$$
(34)

Model (33) suggests the feasible GLS estimator for  $vec(\theta)$ :

$$\operatorname{vec}(\hat{\Theta}) = [(\hat{\mathbf{U}} \otimes \mathbf{I}_{k}) \hat{\Omega}^{-1} (\hat{\mathbf{U}}' \otimes \mathbf{I}_{k})]^{-1} (\hat{\mathbf{U}} \otimes \mathbf{I}_{k}) \hat{\Omega}^{-1} \operatorname{vec}(\mathbf{Z}_{N})$$
(35)

whose covariance matrix can be computed as:

$$\widehat{\text{Cov}}(\text{vec}(\hat{\Theta})) = [(\hat{\mathbf{U}} \otimes \mathbf{I}_{\nu}) \hat{\Omega}^{-1} (\hat{\mathbf{U}}' \otimes \mathbf{I}_{\nu})]^{-1}$$
(36)

where

$$\hat{\Omega} = (\hat{G}_{\psi,N} \hat{D}_{\psi,N}) (I_{N+2q+L} \otimes \hat{\Sigma}_a) (\hat{G}_{\psi,N} \hat{D}_{\psi,N})'$$
(37)

An initial estimation of  $vec(\Theta)$ , needed to evaluate D and G, can be obtained from LS to (33). An estimation of  $\Sigma_a$  can be obtained as:

$$\hat{\Sigma}_{a} = \frac{\hat{U}_{N} \hat{U}_{N}^{\prime}}{N} \tag{38}$$

or from:

$$\operatorname{vec}(\hat{\Sigma}_{H_{a}}) = (I + \hat{\Psi}_{1} \otimes \hat{\Psi}_{1}' + \dots + \hat{\Psi}_{2a+1} \otimes \hat{\Psi}_{2a+1}') \operatorname{vec}(\hat{\Sigma}_{a})$$
(39)

once LS has been applied to model (33).

### 3. SIMULATION EXERCISE

Tables 1 and 2 show the simulation results for two vector processes: a VMA(1) and a VMA(2). These models are the same used in Koreisha and Pukkila (1989) for illustrating the properties of their method. As these authors, we simulated 50 realizations for each model. The sample size N was set equal to 100 and the order "L" of the long VAR was set equal to  $\sqrt{N}$ =10.

Both tables have the same structure, the first pannel shows the mean value of parameter estimates obtained with three different estimation procedures: Our Generalized Hannan Risanen (GHR) procedure, Koreisha and Pukkila's (1989) Double Regression (DR) procedure and Hillmer and Tiao's (1979) Exact Maximum Likelihood estimation procedure. The second pannel shows the mean values of the estimated standard errors associated to each parameter. Finally, the third pannel shows the frequency of significant non-zero parameters (95% confidence) tentatively identified by each method.

Comparative results are similar for both models and can be summarized as follows:

- All estimation procedures yield similar parameters estimates. This result is the same obtained by Koreisha and Pukkila (1989).
- 2) Estimated standard errors associated to DR estimates are greater than those associated to GHR or EML methods. This precision gain is behind the differences in the frequencies with which significant non-zero values are identified with the three estimation methods. Those differences are more evident for parameter values under .3. For instance, in the case of the VMA(1) model, the DR method identifies, as being different from zero, the parameter .2 only in 28% of cases while EML does it in 50% and GHR in 62%. On the other hand, both GHR and EML methods lead to overparametrize the model more often than DR.
- 3) All estimation procedures perform poorly in detecting low parameter values, showing a tendency to conclude that they are not significant, even with moderate sample sizes. This result suggests that removing non significant parameters by blindly using the 95% standard rule could lead to important misspecification errors. We leave for future research the evaluation of the possible consequences of such an identification bias.

[INTRODUCE TABLES 1 AND 2]

In this paper we generalize, for VMA models, the extension of Hannan Rissanen estimation method proposed in Koreisha and Pukkila (1989). The idea is that the differece between innovations associated to a VMA model and residuals from a long VAR(L) will not follow a white noise vector. We derive its exact stochastic structure that is a VMA(q+L-1). Taking into account this result we propose our Generalized Hannan Rissanen (GHR) estimator.

Simulations results indicate that GHR performs better than DR. It increases the precision of parameters estimates and helps to better identify significant non-zero parameters. This feature is particularly important in the case of low parameters values, very difficult to detect even using ML estimation procedures.

#### REFERENCES

- Hannan, E.J. and J. Rissanen. 1982. Recursive estimation of mixed autoregressive-moving average order, *Biometrica* 69, 81-94.
- Hillmer, S. and G. Tiao. 1979. Likelihood function of stationary multiple autoregressive moving average models, Journal of the American Statiatical Association, 74, 652-60.
- Koreisha, S.G. and T.H. Pukkila. 1989. Fast linear estimation methods for vector autoregressive moving-average models. *Journal of Time Series Analysis*, 10, 325-39.
- Koreisha, S.G. and T.H. Pukkila. 1990. A generalized least-squares approach for estimation of autoregressive moving-average models, *Journal of Time Series Analysis*, 11, 139-51.
- Spliid, H. 1983. A fast estimation method for the vector autoregressive moving average model with exogenous variables, *Journal of the Amarican Statistical Association*, 78, 843-49.

$$\theta_1 = \begin{bmatrix} 0 & 0 & 0 & 1.1 & 0 \\ 0 & 0 & 0 & 0 & .2 \\ 0 & 0 & 0 & 0 & 0 \\ -.55 & 0 & 0 & .8 & 0 \\ 0 & 0 & 0 & 0 & .6 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 1 \\ .2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & .7 & 1 \\ 0 & 0 & 0 & -.4 & 1 \end{bmatrix}$$

Mean values of the estimated parameters

					Mean	values of	the estim	ated para	meters					
	GHR				DR					EML'				
-0.09	0.03	80.0	1.05	-0.02	-0.05	0.02	0.03	1.06	-0.00	0.01	0.02	-0.06	1.18	0.04
0.01	-0.05	-0.01	-0.01	0.23	-0.00	-0.03	-0.01	-0.01	0.18	-0.03	0.00	-0.04	0.05	0.25
0.00	0.08	-0.07	0.00	0.00	0.03	0.04	-0.03	0.01	0.015	0.01	0.00	0.02	0.01	0.02
-0.56	0.06	0.03	0.73	0.00	-0.50	0.02	0.01	0.77	0.016	-0.56	0.01	-0.02	0.83	0.01
-0.00	0.03	-0.01	0.02	0.51	-0.01	0.03	-0.00	0.02	0.56	-0.03	0.00	0.01	-0.02	0.63
Mean values of the estimated standard errors														
(0.09)	(0.08)	(0.13)	(0.14)	(0.10)	(0.11)	(0.10)	(0.16)	(0.18)	(0.13)	(0.06)	(0.06)	(0.09)	(0.10)	(0.07)
(0.08)	(80.0)	(0.12)	(0.13)	(0.10)	(0.11)	(0.11)	(0.17)	(0.18)	(0.13)	(0.10)	(0.10)	(0.16)	(0.17)	(0.12)
(0.08)	(0.08)	(0.12)	(0.13)	(0.09)	(0.11)	(0.11)	(0.16)	(0.18)	(0.13)	(0.07)	(0.08)	(0.12)	(0.13)	(0.09)
(0.08)	(0.08)	(0.13)	(0.14)	(0.10)	(0.11)	(0.10)	(0.16)	(0.18)	(0.13)	(0.04)	(0.04)	(0.07)	(0.07)	(0.05)
(0.08)	(80.0)	(0.12)	(0.13)	(0.09)	(0.11)	(0.10)	(0.16)	(0.18)	(0.13)	(0.07)	(0.07)	(0.11)	(0.12)	(0.08)
<u> </u>					Frequenc	y of signi	ificative n	on-zero v	alues (%)				,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	
38	36	38	100	26	20	12	18	100	10	14	18	24	100	24
36	44	42	38	62	10	24	22	16	28	10	12	14	12	50
40	38	48	44	44	22	12	14	18	14	18	18	14	14	10
100	32	40	98	32	94	16	12	94	22	100	18	22	100	18
30	46	42	40	92	10	18	16	12	88	16	16	14	18	100

<sup>(\*)</sup> From Koreisha and Pukkila (1989)

 $TABLE\ 2 \\ Summary\ of\ simulation\ results\ VMA(2),\ K=3,\ N=100,\ 50\ replications$ 

$$\theta_1 = \begin{bmatrix} .7 & 0 & 0 \\ 0 & 1.25 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \theta_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -.75 & 0 \\ 0 & .3 & .6 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 1 \\ -.7 & 1 \\ .4 & 0 & 1 \end{bmatrix}$$

Mean values of the estimated parameters

			Me	an values o	f the estima	ted param	eters			
		GHR			DR		BML* ·			
	0.64	0.02	-0.03	0.67	-0.01	-0.02	0.74	0.00	0.01	
θ,	0.00	1.14	-0.00	-0.01	1.23	0.01	0.01	1.29	0.00	
	0.06	0.07	-0.12	0.03	0.02	-0.06	0.01	0.00	0.00	
						1				
	0.03	0.02	-0.02	0.04	0.01	-0.03	0.02	0.00	0.03	
$\Theta_2$	-0.09	-0.83	0.03	-0.05	-0.78	0.02	0.02	0.78	0.01	
	0.00	0.31	0.57	0.01	0.30	0.54	0.02	0.28	0.69	
	_		Mea	values of	the estimate	d standard	i errors			
	(0.14)	(0.13)	(0.10)	(0.17)	(0.16)	(0.12)	(0.15)	(0.13)	(0.11)	
	(0.15)	(0.14)	(0.11)	(0.18)	(0.16)	(0.12)	(0.13)	(0.13)	(0.10)	
	(0.14) (0.13)		(0.10)	(0.18)	(0.16)	(0.12)	(0.16)	(0.14)	(0.10)	
	(0.14)	(0.12)	(0.10)	(0.17)	(0.16)	(0.12)	(0.15)	(0.15)	(0.10)	
	(0.15) (0.		(0.10)	(0.18)	(0.16)	(0.12)	(0.12)	(0.13)	(0.08)	
	(0.14)	(0.13)	(0.10)	(0.18)	(0.16)	(0.12)	(0.15)	(0.17)	(0.12)	
			Frequ	ency of sign	uficative no	n-zero va	lues (%)		-	
	92	36	40	96	14	6	100	10	12	
	40	100	28	6	100	10	28	100	20	
	34	40	36	12	2	8	6	12	10	
	18	22	32	10	16	16	14	14	8	
	34 96 :		38	28	92	20	20	100	14	
Ì	44 64 96		96	18	54	98	14	60	100	

<sup>(\*)</sup> From Koreisha and Pukkila (1989)