Statistics

COMPARATIVE STUDY OF ESTIMATION METHODS OF THE PARAMETERS OF VMA(1) VIA SIMULATION AND CONSTRUCTION OF MONTE-CARLO CONFIDENCE INTERVALS

S. M. AQIL BURNEY*

SUMMARY: Four methods of estimation of parameters of two components vector moving average model VMA(1) are compared. These methods are exact maximum likelihood (via Kalman filtering), Yule Walker type, moments estimation and Godolphin type. All estimations are based on simulated data using two different covariance matrices. All methods are asymptotically equivalent well inside the invertibility region. Method of constructing Monte-Carlo confidence intervals for parameters is suggested for the time series of the fitted model.

Key Words: Monte-Carlo confidence, statistics, time series.

INTRODUCTION

The properties of the different estimators of moving average model can be analyzed using one of the following approaches:

i) Through the criteria function used for the estimation (such as exact likelihood function or one of its sum of squares approximations),

ii) Through simulation.

The first approach is suitable for those methods of estimation in which a numerical optimization procedure is needed to optimize a function to obtain estimates. By writing criteria functions for these methods in comparable form, we may establish inequalities between them (for univariate MA(q) model (16). Further analysis may be carried out by examining the expected values of these criteria functions for the true values of parameters of a moving average model. This first approach will not yield measures (such as of bias) relating to the sampling distribution of the estimators although indirect information about such measures may be obtained. The second approach not only provides information on sampling distributions of estimators but also provides opportunities to realize and understand other problems (such as computational problems, failure cases etc.). Orcutt and Winokur (15) studied various aspects of estimation for univariate AR(1) model; Nelson (14) compared methods of estimation for univariate MA(1) model. More recently, Ansley and Newbold (1) analyzed by simulation the properties of estimators frequently used in the analysis of univariate ARMA (p,q) models. However, there are few such studies for vector ARMA models giving the small, moderate and large sample properties of various estimators. Tjstheim and Paulson (22) studied the bias of univariate and multivariate AR models. Hillmer and Tiao (10) compared exact and conditional likelihood methods by simulating bivariate MA(1) model with one of the latent roots of characteristic polynomial $|\lambda I-\theta|$ is unity. Our objective is to analyze and compare (by simulation) the properties of four estimators, to be mentioned in the next section for this model. This paper is organized as follows. In section 3 the design of the simulation study is presented. Section 4 describes the presentation of results of this

* From Department of Statistics, University of Karachi, Pakistan.

experiment. In section 5 results are discussed and in the section 6 conclusions are made. In section 7 concluding remarks are made with the intention to point out the limitation of our studies, possible extension and modifications and any noticeable point.

The estimators and their computations

Consider a stationary and invertible bivariate time series $\{\underline{X}_t\}$ t=0,±1,±2,...., where $\underline{X}_t = (X_{1t}, X_{2t})^T$ generated by the model $\underline{X}_t = \underline{\varepsilon}_t - \theta_1 \underline{\varepsilon}_{t-1}$ where $\{\underline{\varepsilon}_t\}$ with e $=(\varepsilon_{1t}, \varepsilon_{2t})$ is a sequence of random vectors identically and independently normally distributed with zero mean vector and covariance matrix Σ.

The estimators of θ and Σ of bivariate MA(1) model which were considered in this study are:

1. Exact maximum likelihood estimators $\hat{\theta}_e$ and $\hat{\Sigma}_e$ which may be computed via Kalman filtering using algorithm 2 (see Appendix) (3).

2. Yule-Walker type estimator $\hat{\theta}_Y$ and $\hat{\Sigma}_Y$ which have been computed using algorithm 1 of Appendix.

3. Moment estimators $\hat{\theta}_M$ and $\hat{\Sigma}_M$ which may be computed using algorithm 3 of Appendix.

4. The Godolphin estimators $\hat{\theta}_\sigma$ and $\hat{\Sigma}_\sigma$ which can be computed using algorithm 1 of Burney (4).

We have used algorithm 21 (a quasi Newton minimizer) of Nash (13) to maximize the likelihood function to give $\hat{\theta}_e$ and $\hat{\Sigma}_e$ with the Yule-Walker type estimators $\hat{\theta}_Y$ and $\hat{\Sigma}_Y$ as initial estimators. It is now generally well known that even in moderately sized samples (such as n=100) approximate maximum likelihood estimator may yield estimates substantially different from $\hat{\theta}_e$ when the roots of the MA model are close to the unit circle (5,11). Such a situation can lead to no (constrained) maximum in the invertibility region. In such situations, we cannot define the estimate in the invertibility boundary as usually done for univariate MA models (1). Whenever a non-invertible estimate was obtained for θ it was considered as a failure as such estimates is impossible to interpret.

DESIGN OF THE EXPERIMENT

Simulation is experimental mathematics (18) and simulation study needs a design for the experiment (12). The idea underlying the design of the following experiment is to use various different parameter values covering the parameter space. The overall results then indicate the performance of each of the estimators mentioned above in different regions of the parameter space.

For bivariate MA(1) models, we have a two dimensional difference equation with characteristic equation

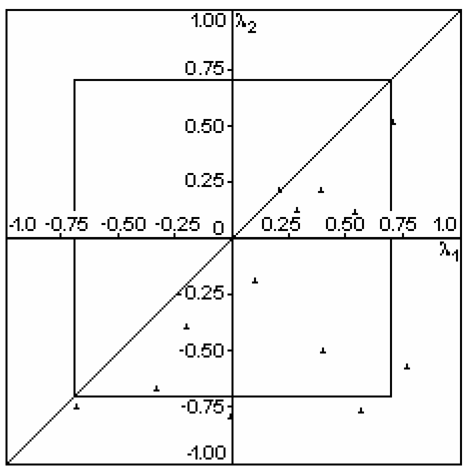
$$0 = |\lambda I - \theta| = \lambda^2 - (\nu_{11} + \nu_{22})I + \nu_{11}\nu_{22} - \nu_{12}\nu_{21}.$$

Hence the characteristic values are:

$$\lambda_{1/2} = \frac{\vartheta_{11} + \vartheta_{12}}{2} \pm \left[\frac{(\vartheta_{11} + \vartheta_{22})^2}{4} - (\vartheta_{11}\vartheta_{22} - \vartheta_{12}\vartheta_{21}) \right]^{1/2}$$

All ν_{11}, ν_{12} and ν_{22} for which $|\lambda_1|,|\lambda_2| <1$ give invertible MA(1) process. In our simulation study MA(1) processes with wide range of characteristic roots in different regions of the unit squares will be used. Figure 1 shows the points (λ_1, λ_2) used in the simulation study and the Table 1a gives the θ_1 which characteristics roots λ_1 and λ_2. We have used one of the methods mentioned in Ripley (18) (the use of Cholesky decomposition of Σ to generate bivariate normal deviates) to produce the bivariate normal deviates ε_t follows bivariate N(O,Σ) with the two covariance structures shown in Table 1b. Thus the properties of the above men-

Figure 1: Latent roots of θ matrix in MA (1).
λ_1= larger latent roots,
λ_2= smaller latent roots.



tioned four estimators for MA(1) model were studied for a total of 28 bivariate MA(1) models. We computed estimates for three different series lengths n=50, 100 and 500. To reduce the effect of starting-up values, 200 pre-sample observations were generated. The number of replications performed for each series length are given in Table 1c.

(Note: we used real values of λ_1 and λ_2 in our simulation study.

However, if the characteristic roots are complex and thus conjugate complex, then λ_1 and λ_2 are given by

$\lambda_1 = r(\cos \nu + i \sin \nu)$, $\lambda_2 = r(\cos \nu - i \sin \nu)$ and it can be shown that

$\nu_{22} = 2r \cos \nu - \nu_{11}$, $\nu_{21} = (2\nu_{11}\cos \nu - \nu^2_{11} - r^2) / \nu_{12}$ provided ν_{12}=0, where r is radius vector. Hence, for given combination of ν_{11}, ν_{21} and angle ν, an MA series can be generated.)

RESULTS

To analyze and compare the small, moderate and large samples performance of the estimators mentioned in section 2, the usual simulation statistics (bias, variance, standard error, mean squared error together with CPU time) are reported in the form of tables. The tables are numbered as follows: For the Table A5.x.y.z, x corresponds to serial number of Table 1a for the choice of θ, y corresponds to the serial number of Table 1b for the choice of Σ and z corresponds to the serial number of Table 1c for the choice of n (series length). Hence, in total we have obtained 84 results tables in the Appendix. For illustration we have presented Tables A5.1.1.3., A5.2.1.1, A5.2.1.2 and A5.2.1.3.

Table 1: (a) parameters values of the MA (1) models used in the simulation study.

S.No.	ν_{11}	ν_{12}	ν_{21}	ν_{22}	λ_1 †	λ_2
1	0.500	-0.470	0.600	-0.600	0.093	-0.193
2	0.200	0.000	0.000	0.200	0.200	0.200
3	0.200	0.010	0.700	0.200	0.284	0.116
4	-0.500	-0.060	0.500	-0.10	-0.200	-0.400
5	0.500	-0.060	0.500	0.10	0.400	0.200
6	0.600	-0.275	0.800	-0.700	0.400	-0.500
7	0.100	0.015	-0.300	0.550	0.540	0.110
8	-0.500	0.060	0.500	-0.500	-0.327	-0.673
9	1.000	-1.000	0.500	-1.000	0.700	-0.700
10	0.200	0.250	-0.600	1.000	0.700	0.500
11	-0.800	0.100	-0.400	0.600	0.571	-0.771
12	0.800	0.100	-0.400	-0.600	0.771	-0.571
13	-0.600	-0.021	0.700	-0.850	-0.695	-0.755
14	-1.000	0.400	-0.500	0.200	0.000	-0.800

† λ1 and λ2 are latent roots of matrix θ.

(b) Covariance matrices Σ

1 $\begin{vmatrix} 1.000 & 0.500 \\ 0.500 & 1.000 \end{vmatrix}$ 2 $\begin{vmatrix} 1.000 & -0.500 \\ -0.500 & 1.000 \end{vmatrix}$

(c)

S.No.	Series length	No. of Replications*
1	50	250
2	100	125
3	500	25

* Note series length x no. of replications= constant

Table 2: CPU time for estimating the parameters of the MA (1) models* using the exact maximum likelihood estimators (θ_e,Σ_e) and Godolphin type estimators (θ_G,Σ_G).

(a) When $\Sigma = \begin{vmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{vmatrix}$

Latent roots †		n=50		n=100		n=500	
S.No. λ_1	λ_2	$\hat{\theta}_e$	$\hat{\theta}_G$	$\hat{\theta}_e$	$\hat{\theta}_G$	$\hat{\theta}_e$	$\hat{\theta}_G$

S.No.	λ_1	λ_2	n=50 $\hat{\theta}_e$	n=50 $\hat{\theta}_G$	n=100 $\hat{\theta}_e$	n=100 $\hat{\theta}_G$	n=500 $\hat{\theta}_e$	n=500 $\hat{\theta}_G$
1	0.093	-0.193	2.85	16.89	5.35	17.05	17.40	16.27
2	0.200	0.200	2.91	6.39	4.95	5.13	17.96	3.90
3	0.284	0.116	2.89	13.47	4.97	13.21	22.56	12.16
4	-0.200	-0.400	3.21	11.83	5.11	11.94	23.36	9.12
5	0.400	0.200	3.06	11.49	5.09	11.20	21.63	9.96
6	0.400	-0.500	3.20	17.36	5.19	17.60	21.71	17.57
7	0.540	0.110	3.35	8.60	5.27	7.26	22.20	6.19
8	-0.327	-0.673	4.15	10.69	5.93	10.59	23.36	9.12

(B) When $\Sigma = \begin{vmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{vmatrix}$

S.No.	λ_1	λ_2	n=50 $\hat{\theta}_e$	n=50 $\hat{\theta}_G$	n=100 $\hat{\theta}_e$	n=100 $\hat{\theta}_G$	n=500 $\hat{\theta}_e$	n=500 $\hat{\theta}_G$
1	0.093	-0.193	3.20	19.24	5.15	19.81	22.99	18.48
2	0.200	0.200	2.74	6.38	4.83	4.93	17.11	3.63
3	0.284	0.116	2.96	13.06	4.97	13.50	21.72	13.08
4	-0.200	-0.400	3.03	11.41	5.05	10.93	21.79	6.69
5	0.400	0.200	3.33	11.54	5.11	12.14	21.91	11.42
6	0.400	-0.500	3.69	19.60	5.27	19.81	22.67	19.03
7	0.540	0.110	3.66	9.40	5.32	7.87	22.80	7.50
8	-0.327	-0.673	4.35	11.99	5.75	11.84	23.59	11.83

+ See Table 1;
† Latent roots of θ matrix;
Note: CPU time in seconds on VAX 785;
Note: The remaining MA (1) models of Table 1 are not given, as for these processes $\hat{\theta}_e$ is preferable (see section 4).

Table A5.2.1.1: Simulation study of MA (1) (Bivariate)
No. of observation in each simulation=50,
No. of lags used for DRPG=30,
No. of simulations=250, seed used=99999763,
Delta=0.010000.

SIMULATED RESULTS FOR MA (1)

Tru Parameters	0.200	0.000	0.000	0.200	1.000	0.500	1.000
Kalman Filtering							
Mean Estimates	0.202	-0.006	-0.017	0.219	0.945	0.482	0.984
Variances	0.041	0.033	0.035	0.035	0.033	0.022	0.036
Bias	0.002	-0.006	-0.017	0.019	-0.055	-0.018	-0.016
Standard Error	0.013	0.012	0.012	0.012	0.011	0.009	0.012
Mean Sqr. Err.	0.041	0.033	0.036	0.036	0.036	0.023	0.036
Average Cpu Time			2.91 seconds				
Yule-Walker							
Mean Estimates	0.178	-0.006	-0.009	0.189	0.914	0.466	0.948
Variances	0.031	0.030	0.032	0.027	0.032	0.022	0.036
Bias	-0.022	-0.006	-0.009	-0.011	-0.086	-0.034	-0.052
Standard Error	0.011	0.011	0.011	0.010	0.011	0.009	0.012
Mean Sqr. Err.	0.031	0.030	0.032	0.027	0.039	0.023	0.039
Average AR Orde	2						
Average Cpu Time			0.06 seconds				
Moment Method							
Mean Estimates	0.151	-0.008	0.008	0.165	1.052	0.511	1.096
Variances	0.018	0.023	0.021	0.015	0.065	0.031	0.067
Bias	-0.049	-0.008	0.008	-0.035	0.052	0.011	0.096
Standard Error	0.009	0.010	0.009	0.008	0.016	0.011	0.017
Mean Sqr. Err.	0.020	0.023	0.021	0.017	0.067	0.031	0.077
Average Cpu Time			0.01 seconds				
Godolphin Type							
Mean Estimates	0.170	0.013	0.012	0.180	0.979	0.476	1.017
Variances	0.044	0.036	0.045	0.043	0.033	0.022	0.037
Bias	-0.030	0.013	0.012	-0.020	-0.021	-0.024	0.017
Standard Error	0.013	0.012	0.013	0.013	0.012	0.009	0.012
Mean Sqr. Err.	0.045	0.036	0.045	0.044	0.034	0.022	0.038
Average Cpu Time			6.39 seconds				

No. of failures for exact likelihood
0
No. of failures for moment method
2
No. of failures for Godolphin type
0
Total CPU Time 2366.07 seconds

Table A5.2.1.2: Simulation study of MA (1) (Bivariate).
No. of observation in each simulation=100,
No. of lags used for DRPG=30,
No. of simulations=125, seed used=99999763,
Delta=0.010000.

SIMULATED RESULTS FOR MA (1)

Tru Parameters	0.200	0.000	0.000	0.200	1.000	0.500	1.000
Kalman Filtering							
Mean Estimates	0.198	-0.022	-0.006	0.193	0.951	0.497	1.007
Variances	0.017	0.016	0.013	0.013	0.018	0.011	0.024
Bias	-0.002	-0.022	-0.006	-0.007	-0.049	-0.003	0.007
Standard Error	0.012	0.011	0.010	0.010	0.012	0.009	0.014
Mean Sqr. Err.	0.017	0.017	0.013	0.013	0.021	0.011	0.024
Average Cpu Time			4.95 seconds				
Yule-Walker							
Mean Estimates	0.188	-0.017	-0.010	0.186	0.937	0.490	0.990
Variances	0.016	0.016	0.014	0.013	0.018	0.011	0.023
Bias	-0.012	-0.017	-0.010	-0.014	-0.063	-0.010	-0.010
Standard Error	0.011	0.011	0.011	0.010	0.012	0.009	0.014
Mean Sqr. Err.	0.016	0.016	0.014	0.014	0.022	0.011	0.024
Average AR Orde	2		0.12 seconds				
Average Cpu Time							
Moment Method							
Mean Estimates	0.158	-0.020	0.001	0.174	1.039	0.529	1.102
Variances	0.008	0.014	0.008	0.008	0.026	0.012	0.034
Bias	-0.042	-0.020	0.001	-0.026	0.039	0.029	0.102
Standard Error	0.008	0.011	0.008	0.008	0.014	0.010	0.016
Mean Sqr. Err.	0.010	0.014	0.008	0.009	0.027	0.013	0.044
Average Cpu Time			0.01 seconds				
Godolphin Type							
Mean Estimates	0.191	-0.017	-0.006	0.191	0.965	0.496	1.021
Variances	0.021	0.020	0.018	0.018	0.019	0.011	0.024
Bias	-0.009	-0.017	-0.006	-0.009	-0.035	-0.004	0.021
Standard Error	0.013	0.013	0.012	0.012	0.012	0.009	0.014
Mean Sqr. Err.	0.021	0.020	0.018	0.018	0.020	0.011	0.024
Average Cpu Time			5.13 seconds				

No. of failures for exact likelihood
0
No. of failures for moment method
0
No. of failures for Godolphin type
0
Total CPU Time 1294.02 seconds

Table A5.1.1.3: Simulation study of MA (1) (Bivariate).
No. of observation in each simulation=500,
No. of lags used for DRPG=30,
No. of simulations=25, Seed used=99999763,
Delta=0.010000.

SIMULATED RESULTS FOR MA (1)

Tru Parameters	0.500	-0.470	0.600	-0.600	1.000	0.500	1.000
Kalman Filtering							
Mean Estimates	0.499	-0.481	0.610	-0.607	0.967	0.485	0.996
Variances	0.002	0.003	0.002	0.003	0.005	0.003	0.004
Bias	-0.001	-0.011	0.010	-0.007	-0.033	-0.015	-0.004
Standard Error	0.010	0.011	0.008	0.011	0.015	0.011	0.013
Mean Sqr. Err.	0.002	0.003	0.002	0.003	0.006	0.003	0.004
Average Cpu Time			17.40 seconds				
Yule-Walker							
Mean Estimates	0.496	-0.471	0.610	-0.611	0.965	0.486	0.993
Variances	0.002	0.003	0.002	0.003	0.005	0.003	0.004
Bias	-0.004	-0.001	0.010	-0.011	-0.035	-0.014	-0.007
Standard Error	0.009	0.011	0.008	0.011	0.015	0.011	0.012
Mean Sqr. Err.	0.002	0.003	0.002	0.003	0.007	0.003	0.004
Average Cpu Time	2		0.65 seconds				
Moment Method							
Mean Estimates	0.112	-0.088	0.141	-0.179	1.168	0.824	1.386
Variances	0.001	0.002	0.002	0.001	0.008	0.005	0.009
Bias	-0.388	0.382	-0.459	0.421	0.168	0.324	0.386
Standard Error	0.006	0.010	0.008	0.008	0.018	0.015	0.019
Mean Sqr. Err.	0.152	0.148	0.212	0.178	0.036	0.110	0.158
Average Cpu Time			0.01 seconds				
Godolphin Type							
Mean Estimates	0.502	-0.492	0.604	-0.591	0.971	0.483	0.999
Variances	0.004	0.004	0.002	0.004	0.005	0.003	0.004
Bias	0.002	-0.022	0.004	0.009	-0.029	-0.017	-0.001
Standard Error	0.013	0.012	0.009	0.013	0.015	0.011	0.013
Mean Sqr. Err.	0.004	0.004	0.002	0.004	0.006	0.003	0.004
Average Cpu Time			16.27seconds				

No. of failures for exact likelihood
0
No. of failures for moment method
0
No. of failures for Godolphin type
0
Total CPU Time 872.72 seconds

Table A5.2.1.3: Simulation study of MA (1) (Bivariate). No. of observation in each simulation=500, No. of lags used for DRPG=30, No. of simulations=25, seed used=99999763, Delta=0.010000.

SIMULATED RESULTS FOR MA (1)

Tru Parameters	0.200	0.000	0.000	0.200	1.000	0.500	1.000
Kalman Filtering							
Mean Estimates	0.200	-0.007	0.012	0.184	0.968	0.487	0.998
Variances	0.002	0.003	0.002	0.002	0.005	0.003	0.004
Bias	0.000	-0.007	0.012	-0.016	-0.032	-0.013	-0.002
Standard Error	0.010	0.011	0.008	0.010	0.015	0.011	0.013
Mean Sqr. Err.	0.002	0.003	0.002	0.003	0.006	0.003	0.004
Average Cpu Time			17.96 seconds				
Yule-Walker							
Mean Estimates	0.194	-0.002	0.011	0.186	0.964	0.486	0.993
Variances	0.002	0.003	0.002	0.003	0.005	0.003	0.004
Bias	-0.006	-0.002	0.011	-0.014	-0.036	-0.014	-0.007
Standard Error	0.009	0.011	0.008	0.010	0.015	0.011	0.013
Mean Sqr. Err.	0.002	0.003	0.002	0.003	0.007	0.003	0.004
Average AR Orde	2						
Average Cpu Time			0.63 seconds				
Momemnt Method							
Mean Estimates	0.178	0.002	0.012	0.181	1.040	0.523	1.078
Variances	0.001	0.002	0.002	0.002	0.008	0.004	0.006
Bias	-0.022	0.002	0.012	-0.019	0.040	0.023	0.078
Standard Error	0.005	0.008	0.008	0.009	0.018	0.012	0.016
Mean Sqr. Err.	0.001	0.002	0.002	0.002	0.009	0.004	0.012
Average Cpu Time			0.01 seconds				
Godolphin Type							
Mean Estimates	0.200	-0.009	0.011	0.184	0.971	0.489	1.002
Variances	0.003	0.005	0.002	0.002	0.005	0.003	0.004
Bias	0.000	-0.009	0.011	-0.016	-0.029	-0.011	0.002
Standard Error	0.011	0.014	0.009	0.009	0.015	0.011	0.013
Mean Sqr. Err.	0.003	0.005	0.002	0.002	0.006	0.003	0.004
Average Cpu Time			3.90 seconds				

No. of failures for exact likelihood=0

No. of failures for moment method=0

No. of failures for Godolphin type=0

Total CPU Time 576.70 seconds

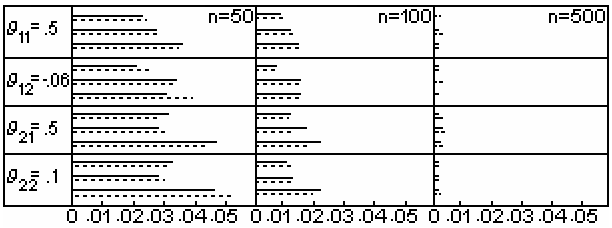


Figure 8: MSE of $\hat{\theta}_e$, $\hat{\theta}_Y$ and θ_G for given values of θ and Σ, ...represents MSE When $\Sigma = \begin{vmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{vmatrix}$ and represents MSE when $\Sigma \begin{vmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{vmatrix}$

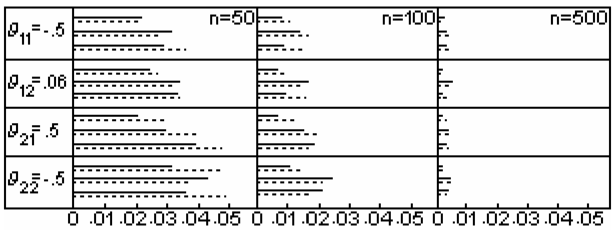


Figure 9: MSE of $\hat{\theta}_e$, $\hat{\theta}_Y$ and θ_G for given values of θ and Σ, ...represents MSE When $\Sigma = \begin{vmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{vmatrix}$ and represents MSE when $\Sigma \begin{vmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{vmatrix}$

DISCUSSION

The results have been analyzed in terms of CPU time, simulated bias, variance and mean square error of estimate. We have also examined the performance of the estimators near the boundary of invertibility region. The following conclusions emerge:

i. Considering the performance in terms of computation time required for the estimators, the CPU time required for $\hat{\theta}_e$ increases with the sample size and the CPU time for $\hat{\theta}_\sigma$ is large for small samples and reduces as the sample size increases, especially well inside the invertibility region. However, CPU time increases for $\hat{\theta}_\sigma$ as spectral radius of θ is near to unity (Table 2). CPU time largely depends on number of iterations required to reach the optimal solution. We have considered a failure if the optimization (in case of obtaining $\hat{\theta}_e$ or $\hat{\theta}_\sigma$) is not completed in 100 seconds or 500 iterations. We came across such failures very rarely. Further, CPU time for computing $\hat{\theta}_e$ and $\hat{\theta}_\sigma$ does not change with the rise of Σ. We noted that the CPU time for Yule-Walker type estimator $\hat{\theta}_Y$ is substantially less than for the ML estimator $\hat{\theta}_e$ via Kalman filtering and the Godolphin type estimator $\hat{\theta}_\sigma$. As $\hat{\theta}_Y$ does not need any optimization routine, the CPU time depends only on the number of observations irrespective of the latent roots of θ. The moment estimator $\hat{\theta}_M$ needs less time as compared to $\hat{\theta}_Y$. Here a failure is marked if convergence is not achieved within 50 iterations.

ii. The performance (in terms of bias and variance) of $\hat{\theta}_e$, $\hat{\theta}_Y$ and $\hat{\theta}_\sigma$ differs from one another, especially for n=50 and n=100. However, as expected, distributional properties are equivalent for the long series such as n=500, especially well inside the invertibility region. The estimator $\hat{\theta}_e$ (ML estimator) has less bias for n=50 as compared to $\hat{\theta}_Y$ and $\hat{\theta}_\sigma$ throughout the invertibility region. In most of the cases for n=50 and n=100, the bias is negative irrespective of any method used. For example Figures 2, 3, 4, 5 show bias $\pm$ 2 S.D. (S.D. = standard deviation) for $\hat{\theta}_e$, $\hat{\theta}_Y$ and $\hat{\theta}_\sigma$ when n=50, 100 and 500. The reason for more bias in $\hat{\theta}_\sigma$ as compared to $\hat{\theta}_e$ is due to the fact that the series length is not large and the auto correlation series is truncated for stability of the estimator. The bias for $\hat{\theta}_Y$ and $\hat{\theta}_\sigma$ is much higher near the non-invertibility region as Figure 5 shows (when λ_1 = 0.700 and λ_2= 0.500, Table 1a). In most of the cases well inside the invertibility region (e.g. the inner square of Figure 1), $\hat{\theta}_\sigma$ has less bias than $\hat{\theta}_Y$. Figures 2, 3, 4 show bias $\pm$2 S.D. for $\hat{\Sigma}_e$, $\hat{\Sigma}_Y$ and $\hat{\Sigma}_\sigma$. It is noticeable that $\hat{\Sigma}_Y$ has more bias compared to the two estimators. Similar remarks hold for most of the models defined well inside the invertibility region.

iii. When comparing mean squared error (MSE) for $\hat{\theta}_e$, $\hat{\theta}_Y$, $\hat{\theta}_\sigma$, the MSE for the exact likelihood estimates is smaller than those for the other estimators $\hat{\theta}_Y$ and $\hat{\theta}_\sigma$. The MSE for Godolphin type estimators is considerably higher, especially for n=50, as compared to the other two estimators, for example see Figures 6, 7, 8, 9. However, as the series length increases, mean square error decreases for the eight values of θ of Table 1a. This situation mainly results from an increased simulated variance for Godolphin type estimates (such a situation usually arises for conditional likelihood estimates as discussed by Hillmer and Tio (10) and for approximate likelihood estimators (5). It was found that MSE is substantially high for Godolphin type estimators for the MA models for the last five values of θ when one of the latent roots of θ is close to the edge of the invertibility region (in our case outside the inner square of Figure 1). However, for series of large length such as 500, $\hat{\theta}_\sigma$ performs satisfactorily in a sense as less noninvertible estimates appear for and θ performance in terms of MSE is comparable with $\hat{\theta}_e$. We have reported results for the last six values of θ of Table 1a when n=500, as for smaller series lengths, for these types of MA(1) models, $\hat{\theta}_\sigma$ shows poor performance. However, $\hat{\theta}_Y$ is preferable as compared to $\hat{\theta}_\sigma$ for smaller series lengths such as n=50, 100.

iv. The Yule-Walker type of estimator $\hat{\theta}_Y$ is stable even near the boundary of the invertibility region and provides a good initial estimate to compute $\hat{\theta}_e$. It was found that almost always AR(7) is sufficient to provide such estimates for the MA(1) models defined in Table 1a-b. Table 3 shows on average the order of AR model for different series lengths as different values of Σ.

v. Moment estimators are inefficient. However, for the models with ν_{12}= ν_{21}= 0, they may provide satisfactory initial estimates well inside invertibility region (14) as results

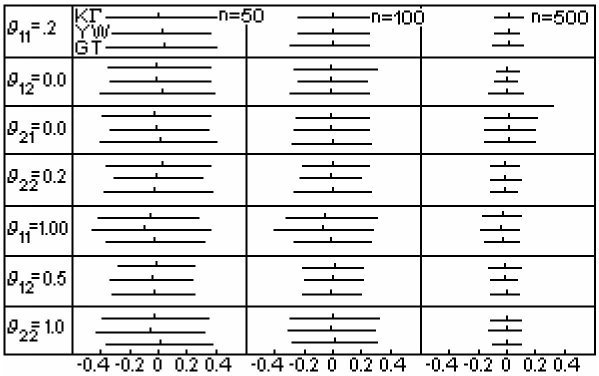


Figure 2: Bias ±2 SD (bias and SD are computed by simulation) for $\hat{\theta}_e$ [via Kawman filtering (KΓ $\hat{\theta}_Y$ (Yule-Walker type estimator-VW))] and Godolphin type estimator (GT)) for given values of θ and Σ.

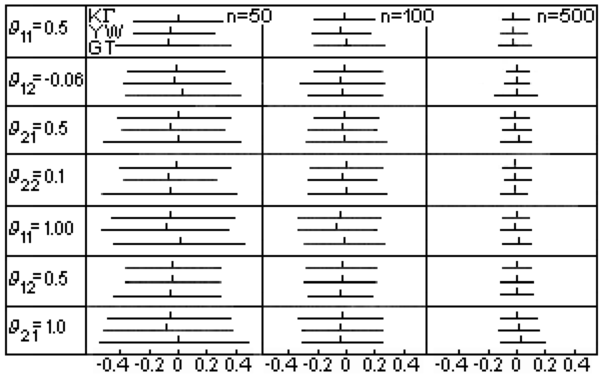


Figure 4: Bias ±2 SD (bias and SD are computed by simulation) for $\hat{\theta}_e$ [via Kawman filtering (KΓ $\hat{\theta}_Y$ (Yule-Walker type estimator (VW))] and Godolphin type estimator-(GT) for given values of θ and Σ.

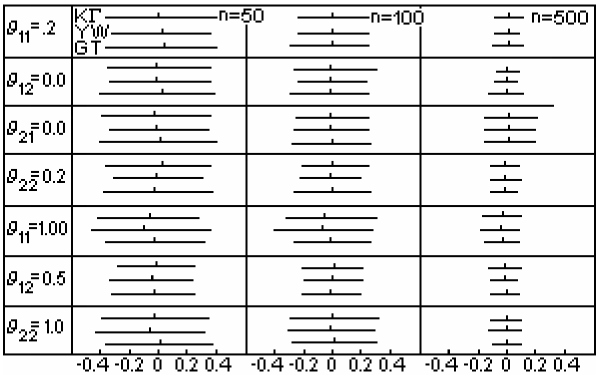


Figure 3: Bias ±2 SD (bias and SD are computed by simulation) for $\hat{\theta}_e$ [via Kawman filtering (KΓ $\hat{\theta}_Y$ (Yule-Walker type estimator-VW))] and $\hat{\theta}_G$ (Godolphin type estimator-GT) for given values of θ and Σ.

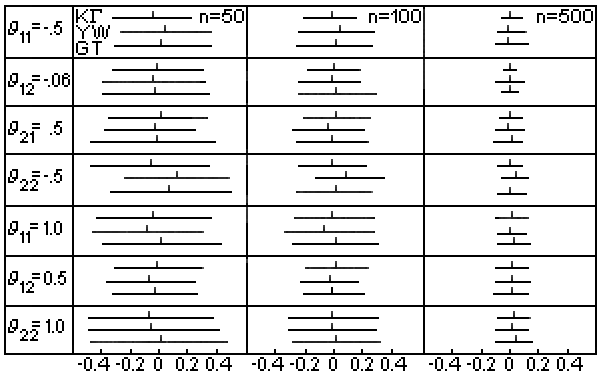


Figure 5: Bias ±2 SD (bias and SD are computed by simulation) for $\hat{\theta}_e$ [via Kawman filtering (KΓ $\hat{\theta}_Y$ (Yule-Walker type estimator-VW))] and $\hat{\theta}_G$ (Godolphin type estimator-GT) for given values of θ and Σ.

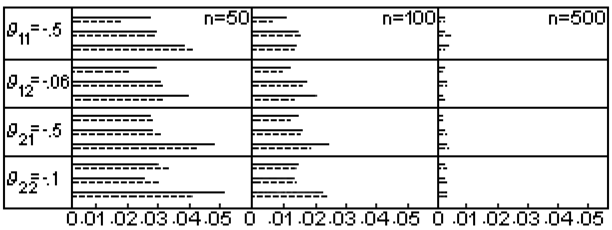


Figure 6: MSE of $\hat{\theta}_e$, $\hat{\theta}_Y$ and θ_G for given values of θ and Σ, ...represents MSE When $\Sigma = \begin{vmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{vmatrix}$ and_represents MSE when $\Sigma \begin{vmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{vmatrix}$

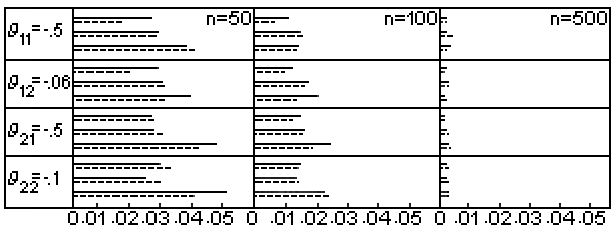


Figure 7: MSE of $\hat{\theta}_e$, $\hat{\theta}_Y$ and θ_G for given values of θ and Σ, ...represents MSE When $\Sigma = \begin{vmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{vmatrix}$ and_represents MSE when $\Sigma \begin{vmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{vmatrix}$

show when ν_{11}= 0.2, ν_{12}= ν_{21}= 0, ν_{22}=0.2.

vi. Monte-Carlo Confidence Intervals. For this section we will be using θ as a scalar quantity and any element of parameter matrix Θ. Now we will discuss the Monte-Carlo method for constructing confidence interval for θ which might be possible while doing extensive simulation experiment or a sample time series is available for the fitted model.

Suppose $\hat{\theta}$ is a consistent estimator of θ with cumulative distribution function F_θ. Let θ^* be a sample from F_θ. Now the variation in θ^* about $\hat{\theta}$ can be used to infer the variation of $\hat{\theta}$ about θ. Suppose $\hat{\theta}$ - θ follows F_O so θ^* - $\hat{\theta}$ follows F_O. Then upper and lower 1/2 α confidence limits for θ^* may be obtained either analytically or via simulation, see Ripley 1987 using empirical cdf of θ^*. Thus (1-)- confidence interval for $\hat{\theta}$ is

$$\theta \in (\hat{\theta} - \mathbf{F_O^-}(1 - 0.5\alpha), \hat{\theta}^* - \mathbf{F_O^-}(0.5\alpha));$$

when $\mathbf{L = F_\theta^-}(0.5\alpha) = \hat{\theta} - \mathbf{F_O^-}(0.5\alpha)$

and $\mathbf{U= F_\theta^-}(1 - 0.5\alpha) = \hat{\theta} - \mathbf{F_O^-}(1-0.5\alpha)$

and F_O is symmetrical about 0 then we can have U - $\hat{\theta}$ = $\hat{\theta}$ - L and θ - $\in$ (L,U). In our case F_O can be assumed as normal asymptotically (as apparent from our extensive simulation experiment which can also be confirmed using any standard Package such as MINITAB). Bunkland (2) calls it a Monte-Carlo confidence interval. For the construction of Monte-Carlo confidence region we extend the work latter on using simulated annealing for efficient computation.

Table 3: Average AR order* required for Yule-Walker type estimator for the MA (1) models of Table 5.1.

$$\Sigma = \begin{vmatrix} 1.0 & 0.5 \\ & 1.0 \end{vmatrix} \qquad \Sigma = \begin{vmatrix} 1.0 & -0.5 \\ - & 1.0 \end{vmatrix}$$

S.No.	Series length			Series length		
	50	100	500	50	100	500
1	2	2	2	2	2	2
2	2	2	3	2	2	2
3	2	2	3	2	2	2
4	2	2	3	2	2	3
5	2	2	3	2	2	3
6	2	2	3	2	2	3
7	2	2	3	2	2	4
8	2	3	5	2	3	5
9	2	3	5	-	-	4
10	2	3	5	3	4	7
11	2	3	5	2	2	5
12	2	3	5	2	3	5
13	2	4	7	-	-	-
14	2	3	5	-	3	-

* See Tiao and Box 1981)

CONCLUSION

In this study we have examined the behavior of four estimators of parameters of bivariate MA(1) model for the series of small, moderate and large lengths. Although the estimators $\hat{\theta}_e$, $\hat{\theta}_Y$ and $\hat{\theta}_\sigma$ are asymptotically equivalent well inside the invertibility region, it was found that, for particular parameter values, their sampling properties for moderate series length can differ substantially. Near the boundary of the invertibility region the exact maximum likelihood estimator $\hat{\theta}_e$ offers substantial gain (in terms of MSE) over the Yule-Walker type estimator $\hat{\theta}_Y$ and Godolphin type estimator $\hat{\theta}_\sigma$. In particular, it seems likely that practical situations arise in which the Godolphin type estimator $\hat{\theta}_\sigma$ would be regarded as undesirable while $\hat{\theta}_Y$ still provides good initial estimates for computing $\hat{\theta}_e$. Godolphin's (8) method may not be as computationally economical as it is for univariate MA(q) model, however, this approach is economical as compared to exact likelihood estimation (via Kalman filtering) at the cost of an increase in MSE. The moment estimator $\hat{\theta}_M$ is consistent and the value of $\hat{\theta}_M$ depends on initial estimates. The moment estimate performs satisfactorily for those values of θ for which latent roots are close to zero.

CONCLUDING REMARKS

1. Our study was limited to the analysis of a few methods of estimation for the bivariate MA(1) model. This Monte Carlo experiment can be extended to vector ARMA (p,q) models to study the performance of methods of estimation (such as conditional sum of squares and unconditional sum of squares estimators, exact likelihood estimators).

2. The Yule-Walker type estimator $\hat{\theta}_Y$ was used as initial estimator to initialize optimization routine to get $\hat{\theta}_e$ or $\hat{\theta}_\sigma$. Bivariate extension of Durbin (6) estimator may also be used. This estimator may have less bias although it is not always guaranteed that this estimator provides invertible model when actually the model is invertible (22).

3. The choice of m, the total number of auto/cross correlations involved in computing $\hat{\theta}_\sigma$ should be proportional to the series length to get stable estimates of θ such that the estimates are unchanged up to, say the third decimal place (for further details on $\hat{\theta}_\sigma$, see (4).

APPENDIX

An initial estimator for VMA(1) model

In practice finite order VAR models are needed to represent MA(q) model (17). We have used a test as suggested by Tiao and Box (21) to determinea l, order of VAR model and found that this test provided useful finite order VAR approximations for VMA(1) model for the purpose of estimation.

Here we are not interested in digressing to order determination of the AR models (12).

An initial estimator for VMA(1) model may be suggested using high order VAR representation given by

$$\sum_{j=0}^{m} \Lambda_j X_{t-j} = \in_t$$

It can be shown that and (1.2) gives $\Lambda_J = \theta_1^J$, **J**=1,2,....,**m**,

$$\sum_{j=0}^{m} \tilde{\theta}_1^j C(s-j) = 0 \qquad s=1,.....,m \qquad (1.1)$$

Thus an estimator θ_1 for θ_1 will be obtained by solving the above system of equations (1.1) for θ_1^J, J=1,....,m. Note we are not using the information present in θ_1^J, J=2,....,m. An estimate of is given by (1.2). We may use the following algorithm to compute the estimate of Σ is given by

$$\Sigma = \sum_{j=0}^{m} \tilde{\theta}^j C(j) \qquad (1,2)$$

We may use the following algorithm to compute the estimate of θ.

Algorithm 1

Step 1. Set l = 1, specify p

Step 2. Solve (1.1) for Λ_j's taking m= l. Compute

$$\Sigma(\ell) = \sum_{j=0}^{\ell} \tilde{\Lambda}_j C_j$$

Step 3. Test $H_o : l_o = l$ using test statistic

$M(l) = (n^* -1/2- lk) \log_e U$,

where , $U = \dfrac{|\Sigma(\ell)|}{|\Sigma(\ell-1)|}$, $n^* = n - p - 1$

and $M(l) \sim X^2{}_k{}^2$ and $M(O) \sim 2X^2{}_k{}^2$ (21).

Step 4. If $M(O) > 2X^2{}_k{}^2$ (0.05) or $M() > X^2{}_k{}^2$ (0.05) then $l = l+1$ and go to step 2.

Step 5. Stop

In Step 2 of the above algorithm can be solved for $\{\Lambda_j\}$ using Whittle (23) algorithm,

Exact likelihood estimation of VMA(1) model using Kalman filter

The transition and measurement equations for k-variate MA(1) model are (9),

$\alpha_t = \mathbf{T}\alpha_{t-1} + \mathbf{R} \in_t$

$\mathbf{X}_t = \mathbf{h}^T \alpha_t$

where $\alpha_t = (\alpha_{1t}, \alpha_{2t},.....,\alpha_{kt})^T$, $X_t = (X_{1t}|X_{2t},.....,X_{nt})^T$, $\in_t = (\in_{1t}, \in_{2t},.....,\in_{kt})^T$, and $\in_t \sim N(O, \Sigma$ The matrices T, R and h are given by

$$T = \begin{bmatrix} 0 & I_k \\ 0 & 0 \end{bmatrix}, R = \begin{bmatrix} I_k \\ -\theta_1 \end{bmatrix}, \text{and } h = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

For Kalman filter, initial conditions are as follows (20).

$\mathbf{a}_{1/o} = \mathbf{a}_o = \mathbf{0}$

$\mathbf{P_o = TP_oT^T + R \Sigma R^T}$

Thus using the above expression for P_o, P_o can be determined as follows:

$$P_o = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix} + \begin{bmatrix} I \\ -\theta_1 \end{bmatrix} \Sigma \begin{bmatrix} I & -\theta_1^T \end{bmatrix},$$

which gives

$\mathbf{P_{11} = \Sigma + \theta_1 \Sigma \theta_1^T}$

$\mathbf{P_{12} = \Sigma \theta_1^T}$

$\mathbf{P_{22} = \theta \Sigma \theta_1^T}$

On applying prediction and updating equations (9), it can be shown that the one step prediction error vector ν_t and the corresponding covariance matrix are given by

$$\nu_t = X_t + \theta_1 \Sigma F^{-1}_{t-1} \nu_{t-1}, \qquad t \geq 1$$
$$F_t = F_1 - \theta_1 \Sigma F^{-1}_{t-1} \theta_1^T, \Sigma \qquad t \geq 2$$

where ν_o=0 and $F_1 = \Sigma + \theta_1 \Sigma \theta_1^T$.
Hence, for given values of θ and Σ, the (-2 log likelihood) function,

$$\sum_{t=1}^{n} \log|F_t| + \sum_{t=1}^{n} \nu_t^T F_t^{-1} \nu_t^T$$

for k-variate MA(1) can be computed using the following algorithm.

Algorithm 2

Step 1. For t=1, set ν1=X1, compute
$F_1 = \Sigma + \theta_1 \Sigma \theta_1^T$. **FLIK** = $\log|F_1| + \nu^T{}_1 F^{-1}{}_1 \nu_1$.
Step 2. t = t + 1
Step 3. $\nu_t = X_t - \theta_1 \Sigma F^{-1}_{t-1} \nu_{t-1}$
Step 4. $F_t = F_1 - \theta_1 \Sigma F^{-1}_{t-1} \theta^T{}_1 \Sigma$; if $|F_t| < (1+\delta)|\Sigma|$, then go to Step 7 for quick recursion, when δ is prefixed number (7, 20).
Step 5. **FLIK** = **FLIK** + $\log F_t + \nu^T{}_t F_t^{-1} \nu$
Step 6. If t<n, go to Step 2.
Step 7. Quick recursion then follow step 5.
Step 8. If t<n, go to Step 7.
Step 9. Stop.

The choice of δ, which is generally a small positive number, say, 0.01 or 0.001, determines the trade-off between accuracy and computational efficiency for the approximation to the likelihood function using quick recursion. Results concerning the trade-off between accuracy and computational efficiency for the approximation are given in (4) for two bivariate MA(1) models with different θ_1 and the same Σ matrix.

METHOD OF MOMENTS FOR VMA(1)

Algorithm 3

Recall K-variate MA(1) model

$X_t = \epsilon_t - \theta_1 \epsilon_{t-1}, \qquad t=0, \pm 1, \ldots..$

where $\epsilon t \sim N(0,\Sigma)$ and $E(\epsilon_i \epsilon_j) = \delta_{ij} \Sigma$ and δ_{ij} is Kronecker delta.

$$\Sigma = \Gamma(0) - \theta_1 \Sigma \theta_1^T$$

We have

$$\theta^T = -\Sigma^{-1} = \Gamma(1) \qquad (3.1)$$

and replacing Γ(O) and Γ(1), the cross-covariance matrix functions at lag O and 1 by sample estimates $C_{(o)}$ and $C_{(1)}$ gives

$$\Sigma = C(0) - \hat{\theta}_1 \Sigma \hat{\theta}_1^T$$
$$\hat{\theta}^T = \Sigma^{-1} C(1) \qquad (3.2)$$

This system equation may be expressed as

$$(I - \hat{\theta}_1 \otimes \hat{\theta}_1)^{-1} \mathrm{Vec}(\hat{\Sigma}) = \mathrm{Vec}(C(0))$$
$$\mathrm{Vev}(\hat{\theta}_1) = -(I \otimes C(1)^T \mathrm{Vec}(\hat{\Sigma}^{-1}) \qquad (3.3)$$

These equations may be solved iteratively for $\hat{\Sigma}$ and $\hat{\theta}_1$ in that order, using the most recent estimates at each step and setting $\hat{\theta}_1$=0 at the start.

ACKNOWLEDGEMENT

I would like to thank Professor B. D. Ripley for his guidance throughout this work and Ministry of Education, Government of Pakistan for financial assistance and faculty of science University of Karachi.

REFERENCES

1. Ansley CF, Newboold P : Finite sample properties of estimators for autoregressive moving average models. Journal Econometrics, 13:159-183, 1980.

2. Buckland ST : Monte-Carlo methods for confidence interval estimation using the Bootstrap technique. BIAS, 10:194-212, 1983.

3. Burney SMA : Computating Likelihood Function of Vector MA(q) Model. (Submitted for possible Publication, Egyptian Statistical Journal), 1989.

4. Burney SMA : A New Method For Large Sample Maximum Likelihood Estimation of Vector Moving Average Models. (Submitted for publication.) PJS, Lahore, Pakistan, 1989.

5. Davidson JEH : Problems with the estimation of moving average processes. Journal of Econometrics, 16:295-310, 1981.

6. Durbin J : The fitting of time series models. Rev Int Stat, 28:233-244, 1959.

7. Gardner G, Harvey AC, Phillips GDA : An algorithm for exact likelihood estimation by means of Kalman filtering. Appl Statist, 29:311-322, 1980.

8. Godolphin EJ : A direct representation for the maximum *likelihood estimator of a Guassian moving average process. Biometrika, 64:375-384, 1977.*

9. Harvey AC : Time Series Models. Philif Allam Publishers Limited. Oxford, 0x5 4SE, 1981.

10. Hillmer SC, Tio GC : Likelihood function of stationary multiple autoregressive moving average models. J Amer Statist Assoc, 74:652-661, 1979.

11. Kang KM : A comparison of estimators of moving average processes, Unpublished paper. (Australian Bureau of Statistics, Canberra), 1975.

12. Lutkepohl H : Comparison of criteria for estimating the order a vector autoregressive process. J Time Series, 6:35-52, 1985.

13. Nash JC : Compact Numerical Methods for computers: Linear Algebra and Function Minimization. Adam Hilger Ltd, Bristol, 1979.

14. Nelson CR : The first order moving average process, Journal of Econometrics, 2:121-141, 1974.

15. Orcutt GH, Winokur HS : First order auto regression inference, estimation and prediction. Econometrics, 37:1-14, 1969.

16. Osborn DR : On the criteria function used for the estimation of moving average process. J Amer Statist Assoc, 77:388-392, 1982.

17. Parzen E : Multiple time series : Determining the order of approximating autoregressive schemes. In Multivariate Analysis, IV, Ed by P Krishnaiah, North-Holland, Amsterdam, pp 283-295, 1977.

18. Ripley BD : Stochastic Simulation. John Wiley and sons, New York, 1987.

19. Sergo, Koreisha, Pukkila T : Fast linear estimation method for vector autoregressive moving-average models J Tim Series Analysis, 10:325-340, 1989.

20. Shea BL : Maximum likelihood estimation of multivariate ARMA process via Kalman filter. In Time Series Analysis: Theory and Practice 5 , Ed by OD Anderson), North-Holland, 1984.

21. Tiao GC, Box GEP : Modeling multivariate Time series with applications. J Amer Statist Assoc, 70:802-816, 1981.

22. Tjstheim D, Paulson J : Bias of some commonly used time series estimates. Biometrika, 70:398-399, 1983.

23. Whittle P : On the fitting of multivariate auto regressions and the approximate canonical factorization of a spectral density matrix. Biometrika, 50:129-134, 1963.

Correspondence:
SM Aqil Burney
Department of Statistics,
University of Karachi,
Karachi 75270,
PAKISTAN.