FAST LINEAR ESTIMATION METHODS FOR VECTOR AUTOREGRESSIVE MOVING-AVERAGE MODELS

BY SERGIO KOREISHA
University of Oregon

TARMO PUKKILA
University of Tampere

First version received January 1989

Abstract. Three linear methods for estimating parameter values of vector autoregressive moving-average (VARMA) models which are in general at least an order of magnitude faster than maximum likelihood estimation are developed in this paper. Simulation results for different model structures with varying numbers of component series and observations suggest that the accuracy of these procedures is in most cases comparable with maximum likelihood estimation. Procedures for estimating parameter standard error are also discussed and used for identification of nonzero elements in the VARMA polynomial structures. These methods can also be used to establish the order of the VARMA structure. We note, however, that the primary purpose of these estimates is to generate initial estimates for the nonzero parameters in order to reduce subsequent computational time of more efficient estimation procedures such as exact maximum likelihood.

Keywords. Multiple time series; sum of squares function; double regression; maximum likelihood estimation; nonzero parameter elements.

1. INTRODUCTION

Procedures for estimating parameter values in vector autoregressive moving-average (VARMA) models vary considerably in computational time and accuracy. In this study we present three *linear* preliminary estimation methods which, when used in conjunction with the exact maximum likelihood (EML) (Hillmer and Tiao, 1979) procedure, can reduce parameter estimation times by more than an order of magnitude. Moreover, because of the great speed and accuracy associated with parameter value calculations they can also be employed in process order identification.

Let $\mathbf{Z}_t$ be a vector of K equally spaced time series containing N observations. A practical class of models which can be used to represent these series, is the VARMA(p, q) class of multiple stochastic difference equations which can be expressed as

$$\underset{K\times K}{\Phi(B)}\ \underset{K\times 1}{\mathbf{Z}_t} = \underset{K\times K}{\Theta(B)}\ \underset{K\times 1}{\mathbf{a}_t} \tag{1}$$

where $\mathbf{a}_t$ is a K-dimensional normal white noise process with zero mean and covariance matrix Σ, and the $K \times K$ matrices $\Phi(B)$ and $\Theta(B)$, which have finite polynomial elements in the lag operator B, defined as $B^j Z_t = Z_{t-j}$, are

0143-9782/89/04 325–15 $02.50/0

simply

$$\Phi(B) = \boldsymbol{I} - \Phi_1(B) - \ldots - \Phi_p(B)^p \tag{2}$$

$$\Theta(B) = \boldsymbol{I} - \Theta_1(B) - \ldots - \Theta_q(B)^q. \tag{3}$$

In the above expressions $\boldsymbol{I}$ is the identity matrix of order K and Φ_i and Θ_j are respectively the ith autoregressive and the jth moving-average coefficient matrices. Process (1) is stationary if the roots of the determinantal equation $|\Phi(B)| = \mathbf{0}$ are outside the unit circle, and is invertible if the roots of the determinantal equation $|\Theta(B)| = \mathbf{0}$ are outside the unit circle.

Any stationary and invertible VARMA(p, q) process can be expressed as an infinite moving-average process

$$\boldsymbol{Z}_t = \Psi(\boldsymbol{B})\boldsymbol{a}_t \qquad \Psi(\boldsymbol{B}) = \sum_{j=0}^{\infty} \Psi_j \boldsymbol{B}^j \qquad \Psi_0 = \boldsymbol{I}, \tag{4}$$

or as an infinite autoregressive process

$$\Pi(\boldsymbol{B})\boldsymbol{Z}_t = \boldsymbol{a}_t \qquad \Pi(\boldsymbol{B}) = \boldsymbol{I} - \sum_{k=0}^{\infty} \Pi_k \boldsymbol{B}^k. \tag{5}$$

These representations, as well shall demonstrate, are useful for deriving preliminary parameter estimates as well as for interpreting higher-order mixed models.

2. LINEAR ESTIMATION METHODS

In this section we present three *linear* procedures for estimating parameter values of VARMA(p, q) models. The three methods, the sum of squares (SS) approach (Koreisha and Pukkila, 1987) and our extensions of Hannan and Rissanen's (1982) double regression methods for multivariate models, are computationally fast, i.e. 30–50 times faster than maximum likelihood procedures (based on time comparisons made on an IBM 4341), and relatively simple to program. For univariate processes some variations of these procedures can be immediately implemented on any computer system, even microcomputers, provided that the user has access to a regression package.

2.1. *Sum of squares approach*

It can easily be shown that, for any stationary and invertible mixed VARMA(p, q) model, the infinite moving-average weights $\Psi(B)$ and infinite autoregressive weights $\Pi(B)$ can be expressed as

$$\begin{aligned} \Pi_1 &= \Theta_1 = \Phi_1 - \Theta_1 \\ \Pi_k &= \Phi_k + \Theta_1 \Pi_{k-1} + \ldots + \Theta_{k-1} \Pi_1 + \Theta_k \Pi_0 \\ \Psi_j &= \Phi_1 \Psi_{j-1} + \ldots + \Phi_{j-1} \Psi + \Phi_j \Psi_0 - \Theta_j, \end{aligned} \tag{6}$$

where $\Theta_k = \mathbf{0}$ when $k > q$, $\Phi_j = \mathbf{0}$ when $j > p$, and $\Pi_0 = -\boldsymbol{I}$. Consequently, if estimates for $\Pi_1, \ldots, \Pi_M$ and $\Psi_1, \ldots, \Psi_M$ were available, then linear estimates for the elements in Φ and Θ could easily and efficiently be derived from (6) by treating these estimates as 'data' and minimizing with respect to Φ and Θ the SS function

$$S(\Psi, \Theta) = \sum_{k=1}^{M_1} \operatorname{tr}(\varepsilon_k^{\mathrm{T}} \varepsilon_k) + \sum_{j=2}^{M_2} \operatorname{tr}(\delta_j^{\mathrm{T}} \delta_j) \tag{7}$$

obtained from the equations

$$\begin{aligned} \hat{\Pi}_k &= \Phi_k + \Theta_1 \hat{\Pi}_{k-1} + \ldots + \Theta_{k-1} \hat{\Pi}_1 + \Theta_k \hat{\Pi}_0 + \varepsilon_k \\ \hat{\Psi}_j &= \Phi_1 \hat{\Psi}_{j-1} + \ldots + \Phi_{j-1} \hat{\Psi}_1 + \Phi_j \hat{\Psi}_0 - \Theta_j + \delta_j, \end{aligned} \tag{8}$$

where $M_1 = \min(M, p + qM)$ and $M_2 = \min(M, q + pM)$, and the circumflex over the symbols represents estimates of the parameters. For more details on the precise nature of the equations associated with the solution of the minimization of the SS function (7), as well as on the properties of these estimates, such as consistency, see Koreisha and Pukkila (1987).

Preliminary estimates for the Π's in (5) can be obtained using several procedures including among others the Yule–Walker equations (Fuller, 1976) and successive autoregressions (Koreisha and Pukkila, 1987). Estimates for $\Psi_1, \Psi_2, \ldots$ can readily be obtained from the relation $\Pi(B)\Psi(B) = \boldsymbol{I}$.

2.2. *Regression approaches*

Our regression approaches for estimating multivariate models consist in deriving estimates for the innovation series associated with the particular VARMA(p, q) process through autoregressions, and then using them as if they were data together with lagged values of $\boldsymbol{Z}_t$ to estimate a multiple regression model with a prescribed structure.

More specifically, we see that approximating any VARMA(p, q) model by a finite autoregression of order M yields estimates for the innovation series $\boldsymbol{a}_t$, namely $\hat{\boldsymbol{a}}_t$. Substituting these and their lagged values into the general VARMA(p, q) form results in the regression structure

$$\boldsymbol{Z}_t = \sum_{i=1}^{p} \Phi_i \boldsymbol{Z}_{t-i} + \hat{\boldsymbol{a}}_t + \sum_{j=1}^{q} \Theta_j \hat{\boldsymbol{a}}_{t-j} + \boldsymbol{e}_t$$

or

$$\boldsymbol{Z}_t - \hat{\boldsymbol{a}}_t = \boldsymbol{Y}_t = \sum_{i=1}^{p} \Phi_i \boldsymbol{Z}_{t-i} + \sum_{j=1}^{q} \Theta_j \hat{\boldsymbol{a}}_{t-j} + \boldsymbol{e}_t \tag{9}$$

which, when estimated using ordinary least squares, generates the desired parameter estimates. Note, however, that the standard errors obtained directly from the multiple regression (9) will *not* be the true estimates of the standard errors of the estimates in Φ_i and Θ_j because the true residual series

for any VARMA process are contained in $\boldsymbol{a}_t$ and not in the least-square residuals $\boldsymbol{e}_t$. However, we shall suggest a better approximation for the standard errors when we demonstrate that these estimates are consistent. Henceforth this procedure will be referred to as the double regression (DR) method.

Now if we equate the derived values for $\Pi_1, \ldots, \Pi_M$ with their theoretical forms in terms of Φ's and Θ's (6), then it is possible to estimate specially transformed regression structures which will yield not only the desired parameter estimates but also their corresponding standard errors. As an example, consider the stationary and invertible univariate ARMA (1, 1) process

$$z_t = \phi z_{t-1} + a_t - \theta a_{t-1} \tag{10}$$

which can also be expressed as an infinite autoregressive process of the form

$$\begin{aligned} z_t = {} & (\phi - \theta)z_{t-1} + \theta(\phi - \theta)z_{t-2} + \theta^2(\phi - \theta)z_{t-3} \\ & + \theta^3(\phi - \theta)z_{t-4} + \ldots + a_t. \end{aligned} \tag{11}$$

An equivalent finite approximation for (11) is simply

$$\begin{aligned} z_t &= \pi_1 z_{t-1} + \theta\pi_1 z_{t-2} + \theta\sum_{j=2}^{M} \pi_j z_{t-j-1} + a_t \\ &= (\phi-\theta)z_{t-1} + \theta\sum_{j=1}^{M} \pi_j z_{t-j-1} + a_t \\ &= \phi z_{t-1} - \theta(z_{t-1} - \sum_{j=1}^{M} \pi_j z_{t-j-1}) + a_t. \end{aligned} \tag{12}$$

Substitution of the estimated values for π_j in (12) yields the regression

$$z_t = \phi z_{t-1} - \theta\left(z_{t-1} - \sum_{j=1}^{M} \hat{\pi}_j z_{t-j-1}\right) + a_t = \phi z_{t-1} - \theta\hat{a}_{t-1} + a_t, \tag{13}$$

which can now be used to estimate the parameter values and their corresponding standard errors since here the regression error is also the true ARMA residual series. The $\hat{a}_{t-1}$ in (13) are the estimated innovations associated with a long autoregression.

In general, therefore, estimates of parameters and standard errors for any K-dimensional VARMA(p, q) process can be obtained from the regression

$$\boldsymbol{Z}_t = \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{Z}_{t-i} - \sum_{j=1}^{q} \boldsymbol{\Theta}_j \hat{\boldsymbol{a}}_{t-j} + \boldsymbol{a}_t. \tag{14}$$

This method can be viewed as the multivariate extension of Hannan and Rissanen's (1982) first-stage estimation procedure for univariate processes. For convenience, we shall refer to this approach as the Hannan and Rissanen (HR) method. Note that the basic difference between the DR and HR methods is the treatment of the current innovation series $\boldsymbol{a}_t$.

Both regression methods generate consistent parameter estimates because the residuals $\hat{\boldsymbol{a}}_t$ from the fitted autoregression approach (in probability) the true innovation series $\boldsymbol{a}_t$ when the $\boldsymbol{\Phi}_i$ in the long autoregressions are estimated consistently. For the HR method this property can easily be verified since the parameter estimates are derived from $p + q$ normal matrix equations obtained by multiplying (14) from the right by $\boldsymbol{Z}_{t-1}^{\mathrm{T}},\ldots,\boldsymbol{Z}_{t-p}^{\mathrm{T}}$, $\hat{\boldsymbol{a}}_t,\ldots,\hat{\boldsymbol{a}}_{t-q}$ and summing over N. Thus, on dividing both sides of the normal equations by N we immediately notice that the solution of these equations tend in probability to the true parameter if $\hat{\boldsymbol{a}}_t \xrightarrow{p} \boldsymbol{a}_t$. The consistency of the DR estimates can be demonstrated in a similar fashion. Note, however, that asymptotically both procedures will yield the same values for the parameters since the right-hand side variables in (9) and (12) are the same and because $\hat{\boldsymbol{a}}_t$ in the DR dependent variable $\boldsymbol{Z}_t - \hat{\boldsymbol{a}}_t$ is asymptotically uncorrelated with any of the right-hand side variables. Consequently, the asympotic variance of the parameter will also be the same. This suggests that the standard error of the DR estimates can be approximated by the same formula employed in calculating the HR estimates provided that the residual variance in (9) is estimated from regression errors constructed in the same manner as for the HR residuals.

3. MODEL IDENTIFICATION

In this section we shall show how these estimation procedures can be used for identification of the nonzero elements of the polynomial matrices of specified VARMA structures, and suggest how they can be used for the identification of the order of VARMA structures.

3.1. *Identification of the nonzero parameter elements*

In practice the number of nonzero elements in the polynomial matrices of VARMA models is relatively small. Estimation of fully parameterized structures, therefore, is an inefficient way of using computing resources. (Overparameterization also reduces estimation efficiency and can increase the residual mean square error (Ledolter and Abraham, 1981).) Preliminary determination of the nonzero elements of $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Theta}_j$ can be made by deriving approximate values for the standard errors of the parameters obtained using the methods described in the previous subsections. In this study we have used our extension of the Hannan and Rissanen regression method as well as a version of Wilson's (1973) 'conditional estimates of β' method to obtain expressions for the standard error of each element position. We approximated the standard error of the SS estimates with Wilson's formula because asymptotic variance–covariance expressions of these estimates have not yet been derived. It should be noted that a byproduct of the estimation of the

parameter covariance matrix using Wilson's method is an updated estimate for the parameter values themselves. These, of course, could subsequently be used to carry out another estimation iteration and the procedure repeated until convergence is achieved. In our simulation study we shall contrast results obtained from linear methods with one iteration of Wilson's algorithm (W1). This is because the quality of these preliminary estimates, as we shall show, is so remarkably good. (Convergence is usually achieved after one or two additional iterations; for other properties of this estimator see Koreisha and Pukkila (1987).) We should add that the SS estimates were used to initialize Wilson's algorithm. This may partly account for the rapid convergence and quality of the estimates.

These approximations for the parameter variance seem to understate the actual variance of the parameter estimates slightly, thus 'conservatively' leading to overidentification of the number of nonzero parameters, i.e. forcing the number of parameters to be subsequently estimated to increase slightly. For preliminary identification purposes this feature is clearly attractive since, as our results have shown, deletion of a few spurious parameters is in general quite simple whereas incorporation of nonzero elements into a VARMA structure is not.

In order to provide some insight into the time taken by each procedure to obtain parameter estimates, we simulated a stationary, invertible and identifiable VARMA(1, 1) structure with five component series, nine nonzero elements and 200 observations. It took approximately 20 min of virtual CPU time on an IBM 4341 on average to estimate the model using the Wisconsin multiple time series maximum likelihood (WMTS-EML) method (Tiao *et al.*, 1979) when initial parameter values were set at 0.1. In contrast, when the initial values were set equal to those obtained by the SS method (50 parameters), the time taken to estimate the structure ranged from 10 to 14 min (the time required to obtain and use conditional likelihood (CL) estimates as initial values for EML estimation was on average 13.5 min). When the initial values were set equal to those obtained by the HR, DR and W1 methods (excluding the insignificant elements, 19 nonzero values on average for the HR and DR methods and 12 for the W1 procedure), the WMTS-EML estimation times were just 5 min 15 s for the DR and HR procedures and 1 min 39 s for the W1 method. The final structure on average had two additional parameters using the DR and HR estimates, and slightly less than one using the W1 estimates. The *combined* time to obtain SS and W1 parameters together with their corresponding standard errors was just 2 min 20 s. The time taken to obtain all DR and HR estimates was under 15 s.

3.2. *Identification of the order of the structure*

By systematically fitting increasing-order VARMA(p, q) structures to the data using these fast estimation methods and then checking whether the resulting residuals behave like white noise, it should be possible tentatively to

identify the order of the process governing the behavior of the observed series. Residual checks might be made using standard cross-correlation Portmanteau tests or order-determination criteria such as the one suggested by Hannan and Rissanen (1982), namely the criterion for which p and q are chosen to minimize the function $\delta(p, q) = \log|\hat{\Sigma}| + (p + q)K^2(\log N)/N$. Residual randomness, as suggested by Pukkila and Krishnaiah (1988), might also be ascertained by testing whether it is possible to fit VAR(p) structure, $p > 0$, to them. If they are white noise any consistent order-determination criteria will show that the residuals follow an AR(0) process. Conversely, if there is a systematic component present in the residuals it will be identified as a VAR(p) process which will immediately trigger the analysis of residuals associated with a higher VARMA structure. Note that by restricting the application of order-determination criteria to only VAR(p) structures we avoid the computational burden of having to estimate all $(p^* + 1)(q^* + 1)$ model possibilities (where p^* and q^* are *a priori* determined order upper bounds) and more importantly circumvent the problems associated with the estimation of unidentified models. More research, however, will be required to determine the reliability of these identification approaches.

4. SIMULATION RESULTS

In order to contrast our linear methods with maximum likelihood procedures we simulated 50 realizations for each of a variety of model structures with varying parameter values, number of component series K and observations N. Each of these models was stationary and/or invertible, and was identifiable in the sense of Dunsmuir and Hannan (1976). Each representation was thought to typify most practical real data applications, e.g. the density of nonzero elements is low, the variation in the magnitude of parameter values is broad, and the feedback/casual mechanisms are complex.

Table I contains the type of information recorded for each of the model structures we examined. Included are the mean value of the parameter and standard error estimates together with their sample standard deviations (the numbers in parentheses), as well as the frequency of significant nonzero values and the mean number of nonzero parameters tentatively identified by each method. However, for brevity, in this paper we shall only report a fraction of the output that we generated. A complete set of results for the structures reported here, and for those excluded, is available from the authors.

As can be seen from this, as well as the tables which follow, our procedures can dramatically reduce the number of parameters which need to be estimated subsequently and provide excellent initial parameter values. In fact, their performance in many cases is similar to that of the EML procedure. Computational times, however, are a fraction of these obtained via the EML method.

TABLE I

SUMMARY OF SIMULATION RESULTS VARMA (0, 1), $K = 5$, $N = 100$

$$\theta_1 = \begin{bmatrix} 0 & 0 & 0 & 1.1 & 0 \\ 0 & 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0 \\ -0.55 & 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 0.6 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & & & & \\ 0.2 & 1 & & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0.7 & 1 & \\ 0 & 0 & 0 & -0.4 & 1 \end{bmatrix}$$

SS					DR					HR					W1					EML				
(a) Mean value of parameter estimates and their sample standard deviations[a]																								
0.09	0.00	−0.16	0.70	0.08	−0.01	0.03	0.06	0.92	−0.03	−0.01	0.03	0.06	0.99	−0.04	0.00	0.05	0.10	0.90	−0.04	0.01	0.02	−0.06	1.18	0.04
(0.17)	(0.15)	(0.28)	(0.21)	(0.18)	(0.13)	(0.11)	(0.17)	(0.16)	(0.13)	(0.14)	(0.12)	(0.20)	(0.19)	(0.14)	(0.10)	(0.10)	(0.16)	(0.18)	(0.11)	(0.10)	(0.08)	(0.17)	(0.18)	(0.12)
0.00	0.07	0.05	−0.04	0.14	0.00	0.03	0.04	−0.01	0.21	0.00	0.02	0.05	−0.02	0.23	0.01	0.01	0.06	−0.02	0.19	−0.03	0.00	−0.04	0.05	=0.25
(0.14)	(0.16)	(0.19)	(0.17)	(0.17)	(0.11)	(0.11)	(0.16)	(0.16)	(0.13)	(0.12)	(0.13)	(0.18)	(0.17)	(0.15)	(0.11)	(0.12)	(0.15)	(0.15)	(0.10)	(0.12)	(0.11)	(0.21)	(0.22)	(0.13)
−0.04	−0.01	0.04	0.15	0.10	0.00	0.00	−0.01	0.05	0.04	−0.01	−0.01	−0.01	0.06	0.05	−0.01	0.00	−0.02	0.09	0.07	0.01	0.00	0.02	0.01	0.02
(0.17)	(0.15)	(0.17)	(0.21)	(0.15)	(0.11)	(0.10)	(0.18)	(0.17)	(0.11)	(0.12)	(0.11)	(0.19)	(0.18)	(0.12)	(0.13)	(0.09)	(0.19)	(0.18)	(0.11)	(0.10)	(0.10)	(0.15)	(0.18)	(0.12)
−0.35	0.03	−0.14	0.64	0.11	−0.50	0.00	0.01	0.76	0.01	−0.54	0.00	0.01	0.82	0.01	−0.50	−0.02	−0.04	0.81	0.00	−0.56	0.01	−0.02	0.83	0.01
(0.20)	(0.14)	(0.19)	(0.23)	(0.18)	(0.12)	(0.11)	(0.17)	(0.15)	(0.13)	(0.12)	(0.11)	(0.19)	(0.17)	(0.14)	(0.09)	(0.08)	(0.14)	(0.13)	(0.07)	(0.06)	(0.06)	(0.10)	(0.12)	(0.07)
−0.04	−0.01	0.00	−0.01	0.44	(0.00	0.00	0.02	−0.02	0.56	(0.00	0.00	0.02	−0.02	0.60	0.01	0.02	0.03	−0.03	0.60	−0.03	0.00	0.01	−0.02	0.63
(0.16)	(0.13)	(0.21)	(0.20)	(0.21)	(0.11)	(0.12)	(0.14)	(0.16)	(0.15)	0.13)	(0.14)	(0.16)	(0.18)	(0.18)	(0.09)	(0.08)	(0.13)	(0.15)	(0.19)	(0.09)	(0.10)	(0.15)	(0.16)	(0.10)
(b) Mean values of the estimated standard errors and their sample standard deviations[a]																								
0.09	0.11	0.14	0.15	0.12	0.13	0.13	0.19	0.20	0.15	0.14	0.14	0.21	0.22	=0.17	0.09	0.11	0.14	0.15	0.12	0.06	0.06	0.09	0.10	0.07
(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
0.09	0.10	0.13	0.13	0.11	0.11	0.11	0.16	0.17	0.13	0.12	0.12	0.18	0.19	0.14	0.09	0.10	0.13	0.13	0.11	0.10	0.10	0.16	0.17	0.12
(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
0.08	0.09	0.12	0.12	0.10	0.11	0.11	0.16	0.17	0.13	0.11	0.12	0.17	0.18	0.14	0.08	0.09	0.12	0.12	0.10	0.07	0.08	0.12	0.13	0.09
(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
0.07	0.09	0.12	0.12	0.10	0.12	0.13	0.19	0.19	0.15	0.13	0.14	0.20	0.21	0.16	0.07	0.09	0.12	0.12	0.10	0.04	0.04	0.07	0.07	0.05
(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
0.08	0.09	0.12	0.12	0.10	0.11	0.12	0.17	0.18	0.14	0.12	0.13	0.19	0.20	0.15	0.08	0.09	0.12	0.12	0.10	0.07	0.07	0.11	0.12	0.08
(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)

TABLE I (*cont.*)

cy of significant nonzero values (95% confidence) (%)

28	98	16	4	2	2	100	4	4	2	4	100	2	14	2	6	100	2	14	18	24	100	24
20	20	34	6	4	4	2	32	4	6	10	4	28	10	16	22	22	44	10	12	14	12	50
28	22	12	6	4	6	4	2	10	4	8	4	2	12	8	16	18	2	18	18	14	14	10
26	98	22	98	0	4	98	2	98	0	4	98	4	100	4	10	100	4	100	18	22	100	18
6	20	94	4	4	2	4	98	8	8	2	4	98	8	12	10	12	100	16	16	14	18	100

umber of nonzero parameters tentatively identified by each method (95% confidence level)

s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
2.04	4.96	1.09	5.16	1.27	6.48	1.93	7.60	1.85

rom 50 sample runs.

procedure. Computational times, however, are a fraction of these obtained via the EML method.

More specifically for the VMA(1) structure with five component series in Table I, we observe that, regardless of the method used, the parameter values are in close agreement with those simulated. The only minor exception is the SS method estimate for θ_{114} (θ_{ijk} is the moving-average parameter of order i in row j and column k). In addition, we also note that the sample standard deviations of the parameter estimates closely match their estimated values. All large-valued simulated parameters are correctly identified as being nonzero irrespective of the method used. All procedures, however, had difficulty in establishing significance for the low-valued θ_{125} parameter (0.2). Given the magnitude of the value this should not be too surprising; examination of the residual cross-correlations of parameterizations which fixed this element position to be zero, seldom showed a need for incorporation of an additional autoregressive (AR) or moving-average (MA) parameter in the (2,5) position. Substantial improvement on the identification of this parameter, however, was obtained when the sample size was increased. With 200 observations the θ_{125} parameter was correctly identified as being nonzero 52%, 60%, 60%, 76% and 74% of the time using the SS, DR, HR, W1 and EML methods respectively. The percentage of overidentified parameters (Table I (d)) was just under 6% for the DR, HR and W1 methods, and 10% and 14% respectively for the EML and SS methods. With 200 observations the percentage of overidentified parameters using the DR, HR, W1 and EML procedures was respectively 7%, 3%, 7% and 4%. The SS procedure overidentified substantially more parameters (24%). Re-estimation (EML) of structures containing these 'spurious' parameters in the majority of cases yielded values which were insignificant.

The average CPU times taken to estimate this VMA(1) structure are given in Table II. As can be seen the linear methods are dramatically faster than maximum likelihood procedures; when used to generate initial estimates for EML the time taken to obtain final estimates is reduced by a factor of 5.

Tables III and IV contain an abbreviated summary of simulation results for two other model structures, VMA(2) and VARMA(1, 1) respectively, each with three component series. The estimates associated with the SS method are excluded from these tables. This is because the performance of the SS approach was found to be consistently below that of the other two linear

TABLE II

Method	$N = 100$	$N = 200$
HR or DR	2.44 s	13 s
EML	2 min 58 s	4 min 21 s
CL	2 min 30 s	4 min 13 s
SS + W1	1 min 5 s	2 min 3 s
HR + EML or DR + EML	36 s	59 s

methods (for these and other structures tested). Since the estimates generated from the DR and HR were not too different from each other (even for small samples), we alternated reporting results of the simulations associated with these methods. The estimated standard errors, not shown here, closely matched the reported sample standard deviations of the parameter values.

In analyzing the results of the VMA(2) structure with 100 observations (Table III), we note that the estimated parameter values as well as the frequency of significant nonzero values obtained from all methods are quite similar and in agreement with the simulated values. As in the VMA(1) case all procedures had some difficulty in identifying the low-valued θ_{232} parameter (0.3) as being nonzero. (This elusive low-valued parameter was not always recoverable from examination of residual cross-correlations.) Regardless of sample size all other nonzero elements in the polynomial matrices of this structure were virtually always identified. Increasing the number of observations, however, significantly improves the frequency with which the θ_{232} parameter is correctly identified as being nonzero. The increase in the number of observations also substantially reduces the sample variation of the parameter estimates.

The results for the VARMA(1, 1) case with 100 observation (Table IV) serve to confirm the conclusions already expressed about the previous structures: accuracy and speed. For 100 observations the DR procedure, however, had a tendency sometimes to 'misclassify' the elements positions (1, 2) and (2, 2), i.e. AR or MA, and *not* 'misidentify' the element position as being nonzero. (Classification refers to the type of parameter categorization, AR or MA, that a nonzero element is ascertained to be.) Note that, even though the θ_{112} parameter value for the DR method is low, the sum of the AR and MA components (with the appropriate sign modification) approaches the simulated value. The W1 and EML procedures had more difficulty in identifying the ϕ_{132} element position than the DR method. This parameter was correctly identified as being nonzero 70%, 44% and 52% of the time for the DR, W1 and EML methods respectively. The θ_{133} parameter also presented problems for all methods. Correct identification ranged from 60% to 80% across methods. In examining the residual cross-correlations for a subset of trials for which either the ϕ_{132} or θ_{133} parameter was unidentified, we found that in the majority of cases the residuals did not substantiate the need for incorporation of an additional parameter in that location. For the cases in which a significant correlation was found at the appropriate element position, incorporation of either an AR or MA parameter produced white noise residuals. The increase in the number of observations improves the quality of the parameter estimates for all methods, and ameliorates the classification problems that the DR method had with the element positions (1, 2) and (2, 2). The W1 and EML methods, even with this increase are still only able to spot the ϕ_{132} parameter around 50% of the time. With 200 observations correct identification of the θ_{133} parameter jumps to almost 100% for all methods.

TABLE III

SIMULATED MODEL

$$\theta_1 = \begin{bmatrix} 0.7 & 0 & 0 \\ 0 & 1.25 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \theta_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -0.75 & 0 \\ 0 & 0.30 & 0.6 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & & \\ -0.7 & 1 & \\ 0.4 & 0 & 1 \end{bmatrix}$$

	100 Observations									200 Observations								
	HR			W1			EML			HR			W1			EML		
θ_1	0.75	0.00	−0.03	0.75	0.01	−0.04	0.74	0.00	0.01	0.73	0.01	−0.01	0.72	0.00	−0.01	0.72	0.00	0.00
	(0.19)	(0.20)	(0.13)	(0.15)	(0.15)	(0.11)	(0.15)	(0.13)	(0.11)	(0.12)	(0.12)	(0.09)	(0.09)	(0.10)	(0.08)	(0.09)	(0.09)	(0.07)
	−0.11	1.19	0.04	−0.14	1.12	0.07	0.01	1.29	0.00	−0.07	1.20	0.03	0.07	1.18	0.05	0.01	1.27	−0.01
	(0.20)	(0.18)	(0.14)	(0.12)	(0.13)	(0.09)	(0.13)	(0.13)	(0.10)	(0.14)	(0.12)	(0.08)	(0.10)	(0.10)	(0.06)	(0.07)	(0.07)	(0.05)
	0.01	0.01	−0.02	0.01	−0.01	0.00	0.01	0.00	0.00	0.01	−0.01	−0.01	0.14	−0.01	0.00	0.02	0.01	0.00
	(0.20)	(0.17)	(0.12)	(0.15)	(0.13)	(0.12)	(0.16)	(0.14)	(0.10)	(0.12)	(0.11)	(0.10)	(0.09)	(0.10)	(0.07)	(0.09)	(0.09)	(0.07)
θ_2	−0.06	−0.06	0.05	−0.02	−0.03	0.06	−0.02	0.00	0.03	−0.01	−0.02	0.02	0.00	−0.01	0.02	0.01	0.00	0.00
	(0.22)	(0.21)	(0.16)	(0.15)	(0.16)	(0.10)	(0.15)	(0.15)	(0.10)	(0.15)	(0.16)	(0.11)	(0.11)	(0.11)	(0.07)	(0.10)	(0.10)	(0.07)
	0.14	−0.64	−0.05	0.11	−0.57	−0.07	−0.02	−0.78	0.01	0.09	−0.67	−0.05	0.06	−0.66	−0.04	−0.02	−0.75	0.01
	(0.31)	(0.29)	(0.22)	(0.12)	(0.16)	(0.11)	(0.12)	(0.13)	(0.08)	(0.20)	(0.21)	(0.13)	(0.10)	(0.12)	(0.07)	(0.05)	(0.23)	(0.05)
	−0.03	0.27	0.62	−0.03	0.26	0.65	0.02	0.28	0.69	0.00	0.29	0.61	−0.01	0.27	0.63	−0.01	0.29	0.63
	(0.18)	(0.18)	(0.12)	(0.15)	(0.13)	(0.09)	(0.15)	(0.17)	(0.12)	(0.11)	(0.12)	(0.07)	(0.09)	(0.10)	(0.07)	(0.08)	(0.08)	(0.07)
(b) Frequency of significany nonzero values (%)																		
θ_1	100	8	4	100	8	8	100	10	12	100	4	6	100	10	4	100	8	4
	10	100	10	22	100	18	28	100	20	8	100	2	22	100	14	10	100	8
	6	8	0	6	10	8	6	12	10	2	4	8	4	8	6	6	6	4
θ_2	16	14	10	8	14	8	14	14	8	6	20	14	12	12	6	8	10	6
	20	80	20	26	96	8	20	100	14	24	98	16	14	100	18	16	100	10
	4	38	100	10	54	100	14	60	100	0	66	100	4	94	100	6	96	100

TABLE IV

SIMULATED MODEL

$$\phi_1 = \begin{bmatrix} 0.7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0.4 & 0 \end{bmatrix} \quad \theta_1 = \begin{bmatrix} 0 & -1.1 & 0 \\ 0 & .6 & 0 \\ 0 & 0 & -0.5 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & & \\ -0.7 & 1 & \\ 0.4 & 0 & 1 \end{bmatrix}$$

	100 Observations									200 Observations								
	DR			W1			EML			DR			W1			EML		
(a) *Mean value of parameter estimates and their sample standard deviations*																		
ϕ_1	0.70	0.42	0.20	0.66	0.17	0.09	0.69	0.15	0.12	0.70	0.29	0.14	0.65	0.03	0.05	0.69	0.03	0.05
	(0.13)	(0.28)	(0.16)	(0.15)	(0.39)	(0.24)	(0.10)	(0.30)	(0.25)	(0.07)	(0.19)	(0.14)	(0.13)	(0.26)	(0.19)	(0.07)	(0.24)	(0.19)
	−0.03	−0.25	−0.12	−0.01	−0.12	−0.05	0.00	−0.08	−0.06	−0.02	−0.16	−0.08	0.01	−0.03	−0.03	0.01	−0.02	−0.04
	(0.11)	(0.25)	(0.17)	(0.12)	(0.34)	(0.21)	(0.07)	(0.21)	(0.18)	(0.06)	(0.14)	(0.12)	(0.06)	(0.20)	(0.13)	(0.04)	(0.14)	(0.12)
	−0.01	0.46	0.09	0.02	0.37	−0.02	−0.01	0.41	0.02	0.00	0.41	0.06	−0.03	0.31	−0.02	−0.01	0.37	0.00
	(0.13)	(0.20)	(0.12)	(0.18)	(0.23)	(0.18)	(0.16)	(0.24)	(0.18)	(0.10)	(0.15)	(0.10)	(0.10)	(0.19)	(0.14)	(0.10)	(0.20)	(0.12)
θ_1	0.04	−0.62	0.17	−0.11	−1.02	0.12	0.02	−1.00	0.12	0.03	−0.75	0.12	−0.09	−1.14	0.09	−0.01	−1.10	0.05
	(0.27)	(0.30)	(0.23)	(0.29)	(0.39)	(0.29)	(0.24)	(0.28)	(0.33)	(0.16)	(0.19)	(0.18)	(0.17)	(0.26)	(0.22)	(0.13)	(0.22)	(0.22)
	−0.08	−0.31	−0.11	−0.01	0.49	−0.08	−0.02	0.56	−0.09	−0.05	0.39	0.06	0.04	0.61	−0.05	0.00	0.60	−0.05
	(0.25)	(0.31)	(0.21)	(0.28)	(0.46)	(0.27)	(0.21)	(0.27)	(0.25)	(0.16)	(0.20)	(0.15)	(0.15)	(0.26)	(0.17)	(0.11)	(0.16)	(0.17)
	0.01	0.05	−0.38	−0.02	−0.05	−0.47	−0.01	0.00	−0.48	0.00	0.01	0.42	−0.04	−0.11	−0.50	−0.01	−0.03	−0.51
	(0.22)	(0.23)	(0.14)	(0.27)	(0.27)	(0.28)	(0.22)	(0.30)	(0.22)	(0.15)	(0.18)	(0.10)	(0.15)	(0.22)	(0.21)	(0.13)	(0.24)	(0.12)
(b) *frequency of significant values* (%)																		
ϕ_1	96	56	22	98	26	14	100	14	12	100	56	30	100	10	8	100	10	4
	4	30	8	0	22	8	2	8	16	0	22	19	0	12	4	0	4	4
	6	70	2	4	44	2	8	52	6	14	78	4	2	46	2	6	56	6
θ_1	12	76	10	20	88	10	12	96	12	4	98	24	14	100	6	4	100	4
	14	30	8	16	52	10	20	82	14	4	60	10	6	84	8	8	100	6
	2	6	60	8	6	74	12	6	20	6	10	92	8	16	94	4	4	100

We also studied the behavior of these preliminary parameter estimation procedures when the model was not invertible. We simulated 50 realizations for sample sizes of 100 and 200 observations of the same noninvertible VMA(1) bivariate model that Hillmer and Tiao (1979) used in their studies (500 trials). The results were very encouraging. Although the estimates, regardless of the procedure used, were in general agreement with the simulated values, the effect of sample size was more pronounced than in the invertible models examined.

We also checked whether the values of the parameter estimates were affected by the number of trial replications. We doubled the number of trials for a subset of the structures considered, and found that the values of the parameters and their corresponding standard error did not change noticeably by the increase.

Finally we note that the covariance matrices associated with the simulated structures studied were also estimated with great precision by all methods. Although for brevity we have excluded the actual results we would like to mention that estimates obtained via the SS procedure were in general consistently lower than those derived with the other methods.

5. CONCLUDING REMARKS

The accuracy and the computational speed of the procedures discussed here, particularly of the DR and HR methods, in obtaining preliminary parameter estimates and in detecting the nonzero element positions in the ARMA polynomial matrices (and also perhaps for establishing the order of the model itself) may provide the catalyst for making multiple time series a viable alternative methodology for studying economic as well as engineering and scientific problems. These techniques should prove particularly useful in the early model-building stage since they permit analysts to compare the characteristics of various competing models inexpensively. Furthermore, because of their programming simplicity, implementation of these methods on any computer system should be rapid and smooth. For univariate processes some of our methods can be immediately operationalized even on microcomputers provided that a basic regression package is available.

REFERENCES

DUNSMUIR, W. and HANNAN, E. (1976) Vector linear time series models. *Adv. Appl. Probab.* 8, 339–64.

FULLER, W. A. (1976) *Time Series* New York: Wiley.

HANNAN, E. and RISSANEN, J. (1982) Recursive estimation of mixed autoregressive-moving average order. *Biometrika* 69, 81–94.

HILLMER, S. and TIAO, G. (1979), Likelihood function of stationary multiple autregressive moving average models. *J. Am. Statist. Assoc.* 74, 652–60.

KOREISHA, S. G. and PUKKILA, T. M. (1987) Identification of nonzero elements in the polynomial matrices of mixed VARMA processes. *J. R. Statist. Soc. Ser. B* 49, 112–26.

—— and —— (1988) Estimation of the polynomial matrices of vector moving average process. *J. Statist. Comput. Simul.* 28, 313–43.

LEDOLTER, J. and ABRAHAM, B. (1981) Parsimony and its importance in time series forecasting. *Technometrics* 23, 411–14.

PUKKILA, T. and KRISHNAIAH, P. R. (1988) On the use of autoregressive order determination criteria in multivariate white noise tests. *IEEE Trans. Acoust. Speech Signal Process.* 36, 1396–403.

TIAO, G. C, BOX, G. E. P., GRUPE, M. R., BELL, W. R. and CHANG, I. (1979) The Wisconsin Multiple Time Series Program. Department of Statistics, University of Wisconsin, Madison, WI.

WILSON, G. T. (1973) The estimation of parameters in multivariate time series models. *J. R. Statist. Soc. Ser. B* 35, 76–85.