

The Maximum Likelihood Estimation of  
Parameters in Mixed Autoregressive  
Moving-Average Multivariate Models

Rusdu Saracoglu

April 1977

Staff Report: 20

PACS: #2635

The views expressed herein are solely those of the author and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The material contained is of a preliminary nature, is circulated to stimulate discussion, and is not to be quoted without permission of the author.

## I. INTRODUCTION

The recent interest amongst economists in univariate and, especially, multivariate mixed Auto-Regressive Moving-Average (ARMA) models and related estimation procedures is partly due to the fact that many economic problems are most naturally formulated in terms of such models (for example see Sargent [9]), and partly due to the existence of computationally feasible algorithms for estimating these models. The estimation problem has been studied by a number of authors both in the frequency domain and in the time domain. Among frequency domain procedures one can cite the works of Hannan ([5], [6]) and Hannan and Nicholls [7]. Even though the frequency domain methods can be shown to be asymptotically efficient, in multivariate cases computational difficulties are, more or less, prohibitive. Time domain methods, on the other hand, are not only computationally feasible but lend themselves to easy interpretations in terms of more familiar regression techniques. In this category Marquardt [8] deals with the problem of estimation for univariate models. Wilson [11] has extended Marquardt's method as they apply to multivariate cases.

This paper reformulates Wilson's problem and offers an alternative, though somewhat similar, algorithm for estimating the parameters of vector ARMA processes.

## II. THE MODEL AND SOME PRELIMINARY RESULTS

In this paper we are concerned with the following model:

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.1)$$

where  $X_t = \{x_{it}\}; i=1, \dots, m$  is an  $m$  component vector stationary stochastic process with zero mean, and  $\varepsilon_t = \{\varepsilon_{it}\}; i=1, \dots, m$  is a vector of random variables satisfying:

$$E\varepsilon_{it}\varepsilon_{jt-s} = 0 \quad \forall i, j, s, \quad s \neq 0$$

$$E\varepsilon_t\varepsilon_t' = \Omega \quad < \infty$$

We shall assume that the lengths of AR and MA parts ( $p$  and  $q$ ) are predetermined and henceforth take them as fixed. The unknown parameters of the model are the matrices  $\phi_i, i=1, \dots, p; \theta_i, i=1, \dots, q$  and  $\Omega$ . Following Wilson [11] we shall refer to these parameters, excluding  $\Omega$ , collectively as  $B = \{B_i\} i=1, \dots, K$  where  $K = m^2(p+q)$ .

Throughout the paper it will be assumed that the zeros of the characteristic polynomials of both AR and MA parts lie outside the unit circle so that equation (2.1) has a pure AR representation as well as a pure MA representation.

In equation (2.1) series  $\varepsilon_t$  are not directly observable but their estimates,  $a_t$ , can be generated from  $X_t$  series recursively for each set of parameter estimates  $\beta = \{\beta_i\} i=1, \dots, K$ . The following results are borrowed from Wilson [11] without proof:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T a_t a_t' = \Omega \quad (2.2)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial a_t}{\partial \beta_j} \right) a_t' = 0 \quad (2.3)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial a_t}{\partial \beta_i} \right) \left( \frac{\partial a_t'}{\partial \beta_j} \right) = 0 \quad (2.4)$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial^2 a_t}{\partial \beta_i \partial \beta_j} \right) a_t' = 0 \quad (2.5)$$

$$\frac{1}{T} \sum_{t=1}^T a_t' \tilde{\Omega}^{-1} a_t = m \quad (2.6)$$

where  $\tilde{\Omega}$  is the estimate of  $\Omega$  given by

$$\tilde{\Omega} = \frac{1}{T} \sum_{t=1}^T a_t a_t'$$

If we assume that  $a_t$  has a joint normal distribution, then the log likelihood function for the parameters  $\beta$  and  $\tilde{\Omega}$  can be written as:

$$\log L = -\frac{mT}{2} \log 2\pi - \frac{T}{2} (\log |\tilde{\Omega}| + \frac{1}{T} \sum_{t=1}^T a_t' \tilde{\Omega}^{-1} a_t) \quad (2.7)$$

By virtue of (2.6) we can simplify the log likelihood function as:

$$\log L = -\frac{mT}{2} (1 + \log 2\pi) - \frac{T}{2} \log |\tilde{\Omega}| \quad (2.8)$$

Hence maximizing the likelihood function is equivalent to minimizing  $|\tilde{\Omega}|$ .

Due to highly nonlinear nature of the objective function (2.8), analytical methods for finding the extrema are computationally unfeasible.

Instead, the approach that was taken in the literature was to devise iterative schemes that converged to extrema of the objective function

(for examples, see Amemiya [1], [2]; Berndt, Hall, Hall, Hausman [3];

Wilson [11]). Basically, all these methods involve the following steps:

(1) choose a starting parameter vector  $\beta^{(0)}$ ; (2) generate  $a_t^{(0)}$ ; (3) calculate the matrix  $\tilde{\Omega}^{(0)}$  as an estimate of  $\Omega$ ; (4) using  $\beta^{(0)}$ ,  $a_t^{(0)}$ , and  $\tilde{\Omega}^{(0)}$ , calculate

parameter corrections  $h^{(0)}$  and corrected parameters  $\beta^{(1)} = \beta^{(0)} + h^{(0)}$  and go to step (2). The key problem in nonlinear estimations is to devise a criteria for choosing  $h^{(0)}$  such that not only convergence is assured, but the resulting parameters have some desirable statistical properties.

### III. APPROXIMATE THEORY

Let  $u$  be a vector of length  $mT$  that consists of stacked  $a$ 's so that the first  $T$  elements of  $u_j$  are  $a_{11}$  through  $a_{1T}$ , the second  $T$  elements  $a_{21}$  through  $a_{2T}$  and so on. In what follows we shall include a superscript of the form  $(i)$  to denote that the variable in question is calculated on the basis of parameters as obtained at iteration  $(i)$ . Initial values will correspondingly be denoted with the superscript  $(0)$ .

In order to synthesize an algorithm we start by expanding  $u^{(i+1)}$  into Taylor series:

$$u^{(i+1)} = u^{(i)} - \Gamma^{(i)} h^{(i+1)} + \text{Remainder} \quad (3.1)$$

where  $-\Gamma^{(i)}$  is an  $mT \times K$  matrix of partial derivatives of  $u_j$  with respect to parameters  $\beta$  evaluated at  $\beta = \beta^{(i)}$ , and  $h^{(i+1)} = \beta^{(i+1)} - \beta^{(i)}$ .

Ignoring the remainder and rewriting (3.1) yields the following convenient form:

$$u^{(i)} = \Gamma^{(i)} h^{(i+1)} + \hat{u}^{(i+1)} \quad (3.2)$$

where  $\hat{u}^{(i+1)}$  is hatted to indicate the approximation. An inspection of (3.2) suggests that parameter corrections,  $h^{(i+1)}$ , could be obtained by regressing  $u^{(i)}$  on  $\Gamma^{(i)}$ , and the residuals of the regression will be  $\hat{u}^{(i+1)}$ .

We, therefore, propose the following parameter corrections:

$$h^{(i+1)} = (\Gamma^{(i)'} \Psi^{(i)-1} \Gamma^{(i)})^{-1} (\Gamma^{(i)'} \Psi^{(i)-1} u^{(i)}) \quad (3.3)$$

where  $\Psi^{(i)}$  is a positive definite matrix to be determined.

Recall from (2.6) that

$$\frac{1}{T} \sum_{t=1}^T a_t^{(i)'} \tilde{\Omega}^{(i)-1} a_t^{(i)} = m \quad (3.4)$$

Let  $\tilde{\Omega} = \frac{\Delta}{m} V$  where  $\Delta = |\tilde{\Omega}|$

Then substituting in (3.4) yields

$$\frac{1}{T} \sum_{t=1}^T a_t^{(i)'} V^{(i)-1} a_t^{(i)} = \Delta^{(i)} \quad (3.5)$$

In terms of u's

$$\frac{1}{T} u^{(i)'} (V^{(i)-1} \otimes I_T) u^{(i)} = \Delta^{(i)} \quad (3.6)$$

At this point we propose to take, in equation (3.3),

$$\Psi^{(i)-1} = (V^{(i)-1} \otimes I_T) \quad (3.7)$$

so that

$$h^{(i+1)} = (\Gamma^{(i)'} (V^{(i)-1} \otimes I_T) \Gamma^{(i)})^{-1} \Gamma^{(i)'} (V^{(i)-1} \otimes I_T) u^{(i)} \quad (3.8)$$

It is a well known result that residuals from GLS with covariance matrix given by (3.7) minimize the quadratic form

$$\frac{1}{T} \hat{u}^{(i+1)'} (V^{(i)})^{-1} \otimes I_T \hat{u}^{(i+1)} \quad (3.9)$$

which can be written in terms of a's as

$$\begin{aligned} &= \frac{1}{T} \sum_{t=1}^T \hat{a}_t^{(i+1)'} V^{(i)-1} \hat{a}_t^{(i+1)} \\ &= \frac{\Delta^{(i)}}{m} \frac{1}{T} \sum_{t=1}^T \hat{a}_t^{(i+1)'} \tilde{\Omega}^{(i)-1} \hat{a}_t^{(i+1)} \\ &= \frac{\Delta^{(i)}}{m} \tau^{(i+1)} \end{aligned} \quad (3.10)$$

where  $\tau^{(i+1)}$  is the trace of  $\hat{\Omega}^{(i+1)} \tilde{\Omega}^{(i)-1}$ . As at iteration step (i+1),  $\Delta^{(i)}$  and m are fixed, then minimizing (3.9) is equivalent to minimizing,  $\tau^{(i+1)}$ .

Proposition A: Given  $\tilde{\Omega}^{(i)}$  and  $\tau^{(i+1)}$  there exists an upper bound on  $\hat{\Delta}^{(i+1)} \equiv |\hat{\Omega}^{(i+1)}|$ .

Proof:

$$\text{Let } S = \hat{\Omega}^{(i+1)} \tilde{\Omega}^{(i)-1}$$

$$\text{Then } |S| = |\hat{\Omega}^{(i+1)}| |\tilde{\Omega}^{(i)-1}|$$

$$|\hat{\tilde{\Omega}}^{(i+1)}| = |S| \cdot |\tilde{\Omega}^{(i)}|$$

$$\hat{\Delta}^{(i+1)} = |S| \cdot \Delta^{(i)} \tag{3.11}$$

Given  $\Delta^{(i)}$  then  $\sup \hat{\Delta}^{(i+1)} = \sup |S| \cdot \Delta^{(i)} = \Delta^{(i)} \sup |S|$ .

Let the diagonal matrix  $\Lambda_s$  be the matrix of characteristic roots of S, and  $\tau_s$  be the trace of S. Then we have

$$|S| = |\Lambda_s| = \prod_{i=1}^m \lambda_i$$

$$\tau_s = \sum_{i=1}^m \lambda_i$$

It can easily be shown that given  $\tau_s$ ,  $|S|$  will be maximized when

$\lambda_1 = \lambda_2 = \dots = \lambda_m = \lambda$ . Hence

$$\lambda = \frac{\tau_s}{m} \quad \text{and} \quad \sup |S| = \left[ \frac{\tau_s}{m} \right]^m$$

and therefore

$$\sup \hat{\Delta}^{(i+1)} = \left[ \frac{\tau_s}{m} \right]^m \Delta^{(i)} \tag{3.12}$$

**Proposition B:** Parameter corrections as given by (3.8) minimize the least upper bound of  $\hat{\Delta}^{(i+1)}$ .

**Proof:** Proof follows directly from (3.12) by substituting  $\tau^{(i+1)}$  for  $\tau_s$  and noting that parameter corrections minimize  $\tau^{(i+1)}$  as given by (3.10).



Proposition C:  $\tau^{(i+1)} \leq m$

Proof: It has been established that

$$\frac{1}{T} \hat{u}^{(i+1)'} (V^{(i)})^{-1} \otimes I_T \hat{u}^{(i+1)} = \frac{\Delta^{(i)}}{m} \tau^{(i+1)}$$

Now if we substitute (3.2) into (3.9) and use the definition of  $h^{(i+1)}$  together with (3.6), we get

$$\frac{\Delta^{(i)}}{m} \tau^{(i+1)} = \Delta^{(i)} - u^{(i)'} R u^{(i)} \tag{3.13}$$

where

$$R = (V^{(i)})^{-1} \otimes I_T \Gamma^{(i)'} (\Gamma^{(i)})' (V^{(i)})^{-1} \otimes I_T \Gamma^{(i)} \Gamma^{(i)'} (V^{(i)})^{-1} \otimes I_T$$

Then  $\tau^{(i+1)} \leq m$  follows from noting that R is a nonnegative definite matrix.

Based on propositions A, B, and C we now propose the following algorithm.

1. Given the parameter vector  $\beta^{(i)}$ , generate the vectors  $a_t^{(i)}$  recursively using equation (2.1).
2. Compute the sample covariance matrix  $\tilde{\Omega}^{(i)}$  and the related determinant  $\Delta^{(i)}$  and matrix  $V^{(i)}$ .
3. Form the vector  $u^{(i)}$  by stacking the vectors  $a_t^{(i)}$  and compute the partial derivative matrix  $\Gamma^{(i)}$ . Note that using the formulas that generated  $a_t^{(i)}$ , the partial derivative matrix can be computed analytically in a very simple manner.

4. Check for convergence by the following two methods:
  - a.  $\Delta^{(i)} - \Delta^{(i+1)} < \text{some prescribed tolerance.}$
  - b.  $\max \Gamma^{(i)'} (V^{(i)})^{-1} \otimes I_T u^{(i)} < \text{some prescribed tolerance.}$
5. Compute the candidate parameter corrections  $h^{(i+1)}$  using equation (3.8), and check whether

$$r = \frac{h^{(i+1)'} \Gamma^{(i)'} (V^{(i)})^{-1} \otimes I_T u^{(i)}}{h^{(i+1)'} h^{(i+1)}} > \alpha \quad (3.14)$$

is satisfied, where  $\alpha$  is a constant less than one. If satisfied go to step (7).

6. If (3.14) is not satisfied, increase the diagonal elements of  $(V^{(i)})^{-1} \otimes I_T$  and go to step (5).
7. Form the corrected parameters

$$\beta^{(i+1)} = \beta^{(i)} + \delta h^{(i+1)}$$

where  $\delta$  is chosen in a manner described by Berndt, [3:p 656]; and go to step (1).

As a result of the convergence theorem by Berndt et al.

[3:p 656] the algorithm described above assures that

$$\lim_{i \rightarrow \infty} \Gamma^{(i)'} (V^{(i)})^{-1} \otimes I_T u^{(i)} = 0$$

so that the iterative process is assured of convergence.

IV. VARIANCE-COVARIANCE MATRIX OF ESTIMATES

In this section we derive the asymptotic variance-covariance matrix of the vector  $\beta$  using the well known properties of maximum likelihood estimates, that they are consistent and asymptotically efficient. We start by noting that asymptotic efficiency implies asymptotic variance-covariance matrix of estimates is equal to

$$-E \left[ \frac{\partial^2 \log L}{\partial \beta \partial \beta'} \right]^{-1} \quad (4.1)$$

which is the Cramer-Rao lower bound (see Theil [10]:p 386).

From equation (2.8) we get

$$\log L = -\frac{mT}{2} (1 + \log 2\pi) - \frac{T}{2} \log \Delta \quad (4.2)$$

$$\frac{\partial \log L}{\partial \beta} = -\frac{T}{2} \frac{1}{\Delta} \frac{\partial \Delta}{\partial \beta} \quad (4.3)$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta'} = -\frac{T}{2} \frac{1}{\Delta} \frac{\partial^2 \Delta}{\partial \beta \partial \beta'} \quad (4.4)$$

Now using the definition of  $\Delta$  as given by equation (3.6) we get

$$\frac{\partial^2 \Delta}{\partial \beta \partial \beta'} = \frac{2}{T} \left[ \left( \frac{\partial u}{\partial \beta} \right)' (V^{-1} \otimes I_T) \left( \frac{\partial u}{\partial \beta'} \right) \right] \quad (4.5)$$

Substituting (4.5) into (4.4) yields

$$\frac{\partial \log L}{\partial \beta \partial \beta'} = -\frac{1}{\Delta} \left[ \left( \frac{\partial u}{\partial \beta} \right)' (V^{-1} \otimes I_t) \left( \frac{\partial u}{\partial \beta'} \right) \right] \quad (4.6)$$

Noting that maximum likelihood estimates are consistent, and recalling the definition of the matrix  $\Gamma$ , we can write the asymptotic variance-covariance matrix of the vector  $\beta$  as

$$\text{var}(\beta) = \Delta[\Gamma'(V^{-1} \otimes I_T)\Gamma]^{-1} \quad (4.7)$$

where (4.7) is evaluated at optimal parameter values.

## References

- [1] Amemiya, T. "The Nonlinear Two-Stage Least-Squares Estimator," Journal of Econometrics, (1974), 105-110.
- [2] \_\_\_\_\_ "The Nonlinear Limited-Information Maximum-Likelihood Estimator and the Modified Nonlinear Two-Stage Least-Squares Estimator," Journal of Econometrics, (1975), 375-386.
- [3] Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman. "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement, (1974), 653-665.
- [4] Hannan, E. J. Multiple Time Series, New York: John Wiley and Sons, 1970.
- [5] \_\_\_\_\_ "The Estimation of Mixed Moving-Average Autoregressive Systems," Biometrika, (1969), 579-593.
- [6] \_\_\_\_\_ "The Estimation of Relationships Involving Distributed Lags," Econometrica, (1965), 206-224.
- [7] \_\_\_\_\_ and D. F. Nicholls, "The Estimation of Mixed Regression, Autoregression, Moving-Average, and Distributed Lag Models," Econometrica, (1972), 529-547.
- [8] Marquardt, D. W. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," Journal of the Society for Industrial and Applied Mathematics, (1963), 431-441.
- [9] Sargent, T. J. "The Demand for Money During Hyperinflations Under Rational Expectations," International Economic Review, forthcoming.
- [10] Theil, H. Principles of Econometrics, New York: John Wiley and Sons, 1971.
- [11] Wilson, G. T. "The Estimation of Parameters in Multivariate Time Series Models," Journal of the Royal Statistical Societies (Series B), (1973), 76-85.