〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 16, No. 1, 1987

A Newton-Raphson Solution for MA Parameters of Mixed Autoregressive Moving-Average Process+

B.S. CHOI*

ABSTRACT

Recently a new form of the extended Yule-Walker equations for a mixed autoregressive moving-average process of orders p and q has been proposed. It can be used to obtain $p+q+1$ parameter values from the first $p+q+1$ autocovariance terms. The autoregressive part of the equations is linear and can be easily solved. In contrast the moving-average part is composed of nonlinear simultaneous equations. Thus some iterative algorithms are necessary to solve them. The iterative algorithm presented by Choi(1986) is very simple but its convergence has not been proved yet. In this paper a Newton-Raphson solution for the moving-average parameters is presented and its convergence is shown. Also numerical examples illustrate the performance of the algorithm.

1. Introduction

Consider the stationary autoregressive moving-average (ARMA) process of orders p and q,

$$\beta(L)y_t=\alpha(L)v_t, \tag{1}$$

where $\beta(L)=\beta_0+\beta_1L+\cdots+\beta_pL^p$, $\alpha(L)=\alpha_0+\alpha_1L+\cdots+\alpha_qL^q$, $\beta_0=\alpha_0=1$, L is the backshift operator and $\{v_t\}$ is a sequence of independent and identically distributed random variables with mean 0 and variance $\sigma^2(>0)$. The stationarity means that the charac-

* Department of Applied Statistics, Yonsei University Seoul, 120, Korea

+ This work is supported in part by Korea Science and Engineering Foundation under Grant 871-0105-004-1.

teristic equation, $\beta(z)=0$, of the autoregressive (AR) part has all the roots outside the unit circle. The relation between the parameters and the autocovariances is given by Choi (1986) as follows. If γ_i is defined by

$$\beta_0\gamma_i+\beta_1\gamma_{i-1}+\cdots+\beta_i\gamma_0=\begin{cases}\alpha_i & \text{if } i=0,\cdots,\ q,\\ 0 & \text{if } i>q,\end{cases} \tag{2}$$

where $\beta_l=0$ for $l>p$, then the autocovariance function, $\sigma(l)=\mathrm{Cov}(y_t, y_{t+l})$, satisfies that

$$\beta_0\sigma(i)+\beta_1\sigma(i-1)+\cdots+\beta_p\sigma(i-p)$$

$$=\begin{cases}(\gamma_0\alpha_i+\gamma_1\alpha_{i+1}+\cdots+\gamma_{q-i}\alpha_q)\sigma^2 & \text{if } i=0,\cdots,q, \quad (3)\\ 0 & \text{if } i>q. \quad (4)\end{cases}$$

Equations (3) and (4) are called the extended Yule-Walker equations for the mixed ARMA process. The problem considered in this paper is, given $p+q+1$ autocovariances $\sigma(0),\sigma(1),\cdots,\sigma(p+q)$, to solve these equations for $p+q+1$ parameters $\beta_1,\cdots,\beta_p$, $\alpha_1,\cdots,\alpha_q$ and σ. We can easily obtain $\beta_1,\cdots,\beta_p$ from the given autocovariances by solving the simultaneous linear equations (4) with $i=q+1,\cdots,q+p$ using an algorithm proposed by Zohar (1974). Hereafter, we assume that not only $\{\sigma(l)\ ;\ l=0,\cdots,p+q\}$ but also $\{\beta_1,\cdots,\beta_p\}$ are known and devote ourselves to solving the nonlinear equations (3) for $\alpha_1,\cdots,\alpha_q$ and σ.

If the covariance generating function is defined by $g(z)=\sum_{l=-\infty}^{\infty}\sigma(l)z^l$, then Equations (2), (3) and (4) are equivalent to

$$\beta(z)\beta(z^{-1})g(z)=\sigma^2\alpha(z)\alpha(z^{-1}),\quad z\neq 0. \tag{5}$$

If the R.H.S. of (5) is denoted by $A(z)$, then we can show the multiplicity of the solutions for $\alpha_1,\cdots,\alpha_q$ and σ through it. If $z_1,\cdots,z_q$ denote the reciprocals of the roots of the characteristic equation, $\alpha(z)=0$, of the moving-average (MA) part, then $A(z)$ can be written as

$$A(z)=\sigma^2\prod_{i=1}^{q}(1-z_iz)(1-z_iz^{-1}).$$

If any real value z_i is replaced by $1/z_i$, or if any pair $(z_i,\bar{z}_i)$ of complex conjugate values are by their reciprocals, then $A(z)$ is unaltered up to the constant σ. Therefore any of the different forms of $\alpha(z)=\prod_{i=1}^{q}(1-w_iz)$ where $(w_i,\bar{w}_i)$ is either $(z_i,\bar{z}_i)$ or $(1/z_i, 1/\bar{z}_i)$ results in the same autocovariance function. Consequently the simultaneous system of the nonlinear equations (3) has at most 2^q different solution sets for $\alpha_1,\cdots,\alpha_q$ and σ.

To express y_t in terms only of previous history time series analysts usually impose the invertibility on the ARMA process, i.e., the equation $\alpha(z)=0$ has all the roots outside the unit circle. (See, e.g., Anderson [1971].) Then, there is a unique solution set for $\alpha_1, \cdots, \alpha_q$ and σ of the simultaneous equations (3).

Choi(1986) proposed an iterative algorithm to obtain the unique solution of the MA parameters and the innovation variance, which is based on a special property of triangular Toeplitz matrices. Even though it has a very simple form and is easy to programme, its convergence has not yet proved. The purpose of this paper is to present a convergent algorithm. The algorithm proposed in Section 2 is simply a Newton-Raphson solution of the nonlinear equations. Its convergence, which is of second order, will be shown in Section 3. In Section 4 appropriate sets of initial values that lead to the invertible solution will be presented. In Section 5 some numerical examples will be presented to illustrate the usefulness of the algorithm.

2. A Newton-Raphson solution

Let Σ be a $(q+1)\times(p+1)$ matrix whose (i, j) element is $\sigma(i-j)$, and let Γ be a $(q+1)\times(p+1)$ matrix whose (i, j) element is 0 if $i<j$, γ_{i-j} otherwise. Also, define a $(q+1)\times(q+1)$ Toeplitz matrix Γ_q whose (i, j) element is 0 if $i>j$, γ_{j-i} otherwise. If $\boldsymbol{\beta}$ denotes $(\beta_0, \beta_1, \cdots, \beta_p)'$, then it is known (Choi [1986]) that (2) and (3) imply that

$$\Sigma\boldsymbol{\beta}=\sigma^2\Gamma_q\Gamma\boldsymbol{\beta}. \tag{6}$$

If we let $C_q=\sigma\Gamma_q$, $C=\sigma\Gamma$ and

$$c_i=\sigma\gamma_i,\ i=0, \cdots, q, \tag{7}$$

then (6) becomes

$$C_qC\boldsymbol{\beta}=\Sigma\boldsymbol{\beta}. \tag{8}$$

We first solve (8) for $c_0, \cdots, c_q$, and then calculate $\alpha_1, \cdots, \alpha_q$ and σ through (2) and (7).

Let $\boldsymbol{c}=(c_0, c_1, \cdots, c_q)'$ and $\boldsymbol{f}(\boldsymbol{c})=(\Sigma-C_qC)\boldsymbol{\beta}$. We are going to solve $\boldsymbol{f}(\boldsymbol{c})=\boldsymbol{0}$ by the Newton-Raphson method. Let T_i be a $(q+1)\times(p+1)$ matrix whose (r, s) element is c_{i+r-s} if $1\leq s\leq i+r$ and $1\leq r\leq q-i+1$, 0 otherwise. Also, let S_i be another $(q+1)\times(p+1)$ matrix whose (r, s) element is c_{i-r+s} if $1\leq r\leq q+1$ and $\max(1, r-i)\leq s\leq q-i+1$, 0 otherwise. If U_i is defined by T_i+S_i, it follows that

$$\frac{\partial\boldsymbol{f}}{\partial c_i}=-(T_i+S_i)\boldsymbol{\beta}=-U_i\boldsymbol{\beta},\ i=0, \cdots, q. \tag{9}$$

If W is defined by $(U_0\beta, \cdots, U_q\beta)$, and if the superscript (n) means the value at the n-th iteration, then the Newton-Raphson method gives that

$$\boldsymbol{c}^{(n+1)}=\boldsymbol{c}^{(n)}+(W^{(n)})^{-1}(\sum-C_q^{(n)}C^{(n)})\beta. \tag{10}$$

A direct calculation shows that

$$\sum_{i=0}^{q} c_i T_i=\sum_{i=0}^{q} c_i S_i=C_q C. \tag{11}$$

It follows that

$$W\boldsymbol{c}=2C_qC\beta. \tag{12}$$

Thus the iterative equation (10) can be simplified as

$$\boldsymbol{c}^{(n+1)}=\frac{1}{2}\boldsymbol{c}^{(n)}+(W^{(n)})^{-1}\sum\beta. \tag{13}$$

It is worth mentioning that Wilson(1969) derived the Newton-Raphson algorithm for a pure MA process, i.e., a case of $p=0$. The algorithm is naturally a special case of (10). Hence, the iterative formula (13) implies that Equation (4) of Wilson's paper can be simplified as

$$\boldsymbol{\theta}^{(t+1)}=\frac{1}{2}\boldsymbol{\theta}^{(t)}+(T^{(t)})^{-1}\boldsymbol{c}^{*}.$$

3. Convergence of the algorithm

Equation (13) equals

$$W^{(n)}\boldsymbol{c}^{(n+1)}=\frac{1}{2}W^{(n)}\boldsymbol{c}^{(n)}+\sum\beta. \tag{14}$$

A similar way to derive (12) shows that

$$W^{(n)}\boldsymbol{c}^{(n+1)}=(C_q^{(n)}C^{(n+1)}+C_q^{(n+1)}C^{(n)})\beta$$

and $W^{(n)}\boldsymbol{c}^{(n)}=2C_q^{(n)}C^{(n)}\beta$.

These equations and Equation (8) imply that (14) equals

$$(C_q^{(n)}C^{(n+1)}+C_q^{(n+1)}C^{(n)})\beta=(C_q^{(n)}C^{(n)}+C_qC)\beta. \tag{15}$$

We define the error $e_i^{(n)}=c_i^{(n)}-c_i$ and its related vector and matrices $\boldsymbol{e}^{(n)}=\boldsymbol{c}^{(n)}-\boldsymbol{c}$, $E^{(n)}=C^{(n)}-C$, $E_q^{(n)}=C_q^{(n)}-C_q$. Substracting $(C_qC^{(n)}+C_q^{(n)}C)\beta$ from the both sides of (15) gives that the R.H.S. becomes $E_q^{(n)}E^{(n)}\beta$ and the L.H.S. becomes that

$$\begin{aligned}&\{C_q^{(n)}(C^{(n+1)}-C)+(C_q^{(n+1)}-C_q)C^{(n)}\}\beta\\&=(C_qE^{(n+1)}+E_q^{(n+1)}C)\beta+(E_q^{(n)}E^{(n+1)}+E_q^{(n+1)}E^{(n)})\beta\\&=A(\boldsymbol{c})\boldsymbol{e}^{(n+1)}+A(\boldsymbol{e}^{(n)})\boldsymbol{e}^{(n+1)},\end{aligned}$$

where $A(\boldsymbol{x})$ is a $(q+1)\times(q+1)$ matrix each element of which is a linear combination of elements of the vector $\boldsymbol{x}=(x_0, x_1, \cdots, x_q)'$. It follows that

$$\boldsymbol{e}^{(n+1)}+A(\boldsymbol{c})^{-1}A(\boldsymbol{e}^{(n)})\boldsymbol{e}^{(n+1)}=A(\boldsymbol{c})^{-1}E_q^{(n)}E^{(n)}\boldsymbol{\beta}. \tag{16}$$

Let $\|\boldsymbol{x}\|$ denote the norm of a vector $\boldsymbol{x}=(x_1, \cdots, x_n)'$ defined by $\max_i|x_i|$, and let $\|A\|$ be its subordinate norm of a matrix $A=(a_{ik})$ defined by $\max_i \sum_{k=1}^{n}|a_{ik}|$. If d_n is defined by $\|\boldsymbol{e}^{(n)}\|$, then the definition of $A(\boldsymbol{e}^{(n)})$ implies that there exists a positive constant h_1 satisfying $\|A(\boldsymbol{e}^{(n)})\|\leq h_1 d_n$. It follows that

$$\|\boldsymbol{e}^{(n+1)}+A(\boldsymbol{c})^{-1}A(\boldsymbol{e}^{(n)})\boldsymbol{e}^{(n+1)}\|\geq(1-h_1h_2d_n)d_{n+1}$$

and

$$\|A(\boldsymbol{c})^{-1}E_q^{(n)}E^{(n)}\boldsymbol{\beta}\|\leq h_2(q+1)d_n(p+1)d_n\|\boldsymbol{\beta}\|=h_3d_n^2,$$

where $h_2=\|A(\boldsymbol{c})^{-1}\|$ and $h_3=(p+1)(q+1)h_2\|\boldsymbol{\beta}\|$. Therefore, Equation (16) implies that

$$(1-h_1h_2d_n)d_{n+1}\leq h_3d_n^2. \tag{18}$$

If an initial vector $\boldsymbol{c}^{(0)}$ is chosen so that $d_0<1/(h_1h_2+h_3)$, then (18) implies that $\{d_n\}$ is strictly decreasing and then $d_n<d_0$ for any n. Thus, we know that

$$d_{n+1}\leq h_3d_n^2/(1-h_1h_2d_n)<(h_1h_2+h_3)d_n^2, \tag{19}$$

or equivalently,

$$d_n<\{(h_1h_2+h_3)d_0\}^{2^n}/(h_1h_2+h_3),\quad n=1, 2, \cdots. \tag{20}$$

Since d_0 is chosen to be less than $1/(h_1h_2+h_3)$, d_n converges to zero. Equation (19) shows that the convergence is of second order.

4. Choice of starting values

Equation (20) shows that a sufficient condition for the convergence of Algorithm (13) is to choose a starting vector $\boldsymbol{c}^{(0)}$ satisfying $\max|c_i^{(0)}-c_i|<1/(h_1h_2+h_3)$. Since neither the true vector $\boldsymbol{c}$ nor $h_1h_2+h_3$ is known, it isn't possible to find $\boldsymbol{c}^{(0)}$ satisfying the sufficient condition. Instead we propose a method to select a starting vector sufficiently close to the true vector as follows. The definition of γ_i in (2) implies that the ARMA (p, q) process $\{y_t\}$ can be represented by the $MA(\infty)$ process satisfying

$$y_t=\sum_{l=0}^{\infty}\gamma_l v_{t-l}. \tag{21}$$

Following Durbin's method (1959), i.e., to use an $MA(s)$ model with a fairly large s to estimate the parameters, we consider an invertible $MA(s)$ model

$$y_t^{(s)}=\sum_{l=0}^{s}\gamma_l^{(s)}v_{t-l}\,, \tag{22}$$

where the parameters $\gamma_0^{(s)},\cdots,\gamma_s^{(s)}$ are determined so that $\mathrm{Cov}(y_t^{(s)},\ y_{t-l}^{(s)})$ equals $\sigma(l)$ for $l=0,\cdots,\ s$. Then it is known (see, e.g., Anderson[1971]) that $y_t^{(s)}$ converges to y_t in L^2 as s tends to infinity. Thus we can choose an integer s so that $\gamma_0^{(s)},\cdots,\gamma_q^{(s)}$ are as close to $\gamma_0,\cdots,\gamma_q$ respectively as we wish. We apply Wilson's algorithm to this suitable $MA(s)$ process with starting values

$$c_0^{(0)}=\{\sigma(0)+2\sum_{l=1}^{s}\sigma(l)\}^{1/2}$$

$$c_i^{(0)}=-(\beta_1 c_{i-1}^{(0)}+\cdots+\beta_i c_0^{(0)}),\quad i=1,\cdots,s.$$

It is known that Wilson's algorithm with the starting values has a convergence property of second order. More precisely the resulting values converge to the true parameter values satisfying the invertible condition of the $MA(s)$ process. Therefore the final values of Wilson's will be suitable as starting values to Algorithm (13). Since the former is a special case of the latter, we do not need a separate computer program for the former. However, since s is large, we need a lot of computing time and storage for inversion of W.

We consider another possible choice of starting values for Algorithm (13), which needs less computation. Let

$$c_0^{(0)}=(\sigma(0))^{1/2}$$

$$c_i^{(0)}=-(\beta_1 c_{i-1}^{(0)}+\cdots+\beta_i c_0^{(0)}),\quad i=1,\cdots,q. \tag{23}$$

These values are simply chosen so that $\alpha^{(0)}(z)=0$ has no roots inside the unit circle. Even if we don't know whether d_0 less than $1/(h_1h_2+h_3)$ or not, numerical examples show that the set of starting values (23) gives the same result as that using Wilson's algorithm. When the Gauss-Jordan method is applied to calculate the inverse of a matrix, the latter needs computing operations $\{(s+1)/(q+1)\}^3$ times and memory locations $\{(s+1)/(q+1)\}^2$ times as many as the former. In particular, if the characteristic equation, $\beta(z)=0$, has a root near the unit circle, then s should be sufficiently large, and so are the ratios. Furthermore, the larger s is, the larger the total number of iterations becomes. Thus, it is recommended to use the starting values in (23) rather than the ones based on Wilson's algorithm.

5. Numerical examples

(a) Let $p=1$, $q=2$, $\beta_1=0.3$, $\alpha_1=-0.7$, $\alpha_2=-0.18$ and $\sigma^2=1.0$. Then the autocovariances are $\sigma(0)=2.0158241758$, $\sigma(1)=-1.1247472527$, $\sigma(2)=0.15742417582$, $\sigma(3)=-0.047227252747$. Also, Equation (2) gives that $\gamma_0=1.0$, $\gamma_1=-1.0$ and $\gamma_2=0.12$.

The computing result with the starting values in (23) is in Table 1, where $d_n^*=\max|c_i^{(n)}-c_i^{(n-1)}|$.

Table 1

n	$\sigma^{(n)}$	$\gamma_1^{(n)}$	$\gamma_2^{(n)}$	d_n^*
0	1.419797231	−.3000000000	.0900000000	
1	1.246031511	−.6244565473	.0855910085	3.5E-1
2	1.102810459	−.8225054120	.1025466176	1.3E-1
3	1.034721385	−.9362260934	.1136362117	6.2E-2
4	1.007119446	−.9865683560	.1186570911	2.7E-2
5	1.000443719	−.9991573065	.1199157318	6.6E-3
6	1.000001952	−.9999962920	.1199996292	4.4E-4
7	1.000000000	−.9999999999	.1200000000	2.0E-6
8	1.000000000	−1.000000000	.1200000000	3.8E-11
9	1.000000000	−1.000000000	.1200000000	1.5E-20

If Wilson's algorithm is applied to the autocovariances with $s=19$, then after 7 iterations the values are $\sigma^{(7)}=1.000000000$, $\gamma_1^{(7)}=-1.000000000$, $\gamma_2^{(7)}=0.1200000000$, $d_7^*=2.0E-6$ and satisfy $\max_{0\le i\le 2}|c_i^{(7)}-c_i|<1.0E-11$. If we let them be the starting values for Algorithm (13), then the result is as follows.

Table 2

n	$\sigma^{(n)}$	$\gamma_1^{(n)}$	$\gamma_2^{(n)}$	d_n^*
0	1.000000000	−1.000000000	.1200000000	
1	1.000000000	−1.000000000	.1200000000	8.2E-11
2	1.000000000	−1.000000000	.1200000000	1.5E-22

(b) The second example illustrates a case when a root of characteristic equation, $\beta(z)=0$, is near the unit circle. Let $p=1$, $q=2$, $\beta_1=0.95$, $\alpha_1=-0.2$, $\alpha_2=-0.15$ and

$\sigma^2=1$. Then, $\gamma_1=-1.15$ and $\gamma_2=0.9425$. Also, the autocovariances are $\sigma(0)=11.433333333$, $\sigma(1)=-10.889166667$, $\sigma(2)=10.194708333$, and $\sigma(3)=-9.6849729167$. With the starting values specified in (23) Algorithm (13) gives the following result.

Table 3

n	$\sigma^{(n)}$	$\gamma_1^{(n)}$	$\gamma_2^{(n)}$	d_n^*
5	1.000001879	−1.149997099	.9424981743	9.2E-4
6	1.000000000	−1.150000000	.9425000000	1.9E-6
7	1.000000000	−1.150000000	.9425000000	1.9E-12

When we apply Wilson's algorithm to the autocovariances with $s=19$, we can find that d_n^* does not decrease monotonically but oscillates. After 16 iterations the values are $\gamma_1^{(16)}=-.0025145178$, $\gamma_2^{(16)}=.4897638886$, $\sigma^{(16)}=1.318407178$ and $d_{16}^*=7.9$. Table 4 shows the result when these values are used as the starting values for Algorithm (13).

Table 4

n	$\sigma^{(n)}$	$\gamma_1^{(n)}$	$\gamma_2^{(n)}$	d_n^*
5	1.000001038	−1.149856810	.9424666852	8.1E-4
6	1.000000000	−1.149999710	.9424999626	1.0E-6
7	1.000000000	−1.150000000	.9425000000	2.0E-12
8	1.000000000	−1.150000000	.9425000000	7.5E-24

6. Application

Algorithm (13) can be used to obtain estimates of the MA parameters of an ARMA process through sample autocovariances. To calculate AR parameters of the process we may use Zohar's algorithm mentioned before. Although these are not the maximum likelihood estimates, they will be good starting values for the maximum likelihood estimation.

Acknowledgement

The author would like to thank a referee for his helpful suggestion.

References

(1) Anderson, T.W. (1971). *The Statistical Analysis of Time Series*, Wiely, New York.

(2) Choi, B.S. (1986). An algorithm for solving the extended Yule-Walker equations of an auto-regressive moving-average time series, *IEEE Transaction on Information Theory*, Vol. IT-32, 417~419.

(3) Durbin, J. (1959). Efficient estimation of parameters in moving-average models, *Biometrika*, Vol. 46, 306-316.

(4) Wilson, G. (1969). Factorization of the covariance generating function of a pure moving-average process, *SIAM Journal of Numerical Analysis*, Vol. 6, 1~7.

(5) Zohar, S. (1974). The solution of a Toeplitz set of linear equations, *Journal of Association for Computing Machinery*, Vol. 21, 272~276.