Ann. Inst. Statist. Math.
34 (1982), Part A, 83-90

PARAMETER ESTIMATION OF AN AUTOREGRESSIVE MOVING AVERAGE MODEL

JUNJI NAKANO

(Received Oct. 9, 1980; revised July 6, 1981)

Summary

An estimator of the set of parameters of an autoregressive moving average model is obtained by applying the method of least squares to the log smoothed periodogram. It is shown to be asymptotically efficient and normally distributed under the normality and the circular condition of the generating process. A computational procedure is constructed by the Newton-Raphson method. Several computer simulation results are given to demonstrate the usefulness of the present procedure.

1. Introduction

In the field of time series analysis, an autoregressive moving average (ARMA) model plays an important role as a parametric model. If a stochastic process $\{y_t\}$ satisfies the equation

$$(1.1) \qquad y_t+b_1y_{t-1}+\cdots+b_py_{t-p}=u_t+a_1u_{t-1}+\cdots+a_qu_{t-q}$$

for all integer t, this process is called an ARMA (p, q) model, where $\{u_t\}$ is a white noise sequence and $b_1, \cdots, b_p, a_1, \cdots, a_q$ are parameters. We assume here that the sequence $\{u_t\}$ consists of independent $N(0, \sigma^2)$ variables.

Several parameter estimation procedures have been introduced by Durbin [6], Walker [12], Hannan [7], Clevenson [4] and Anderson [1]. Some of these procedures used the periodogram $I(\lambda)$,

$$(1.2) \qquad I(\lambda)=\frac{1}{2\pi N}\left|\sum_{t=0}^{N-1} y_t \exp(i\lambda t)\right|^2$$

where $y_0, \cdots, y_{N-1}$ are observations. The logarithm of the periodogram or that of the smoothed periodogram was used for estimating some parameters by Davis and Jones [5], Bloomfield [2] and Hannan and

Nicholls [8]. Wahba [11] also used the log periodogram to obtain an estimate of the log spectral density by a spline approach.

In this article, we present a procedure of estimating parameters of an ARMA model by an application of the principle of least squares to the log spectral density using the log smoothed periodogram. In Section 2, we propose an estimation method under the normality of the sequence $\{u_t\}$ and the circular condition. We note that these restrictions are imposed merely for the simplicity of theoretical treatment. Then, in Section 3, the estimator is shown to be asymptotically as efficient as the maximum likelihood estimator. In Section 4, we provide an iterative procedure based on the Newton-Raphson method as our method needs to minimize a nonlinear function, and one way of obtaining initial estimates is suggested. In Section 5, some computer simulation results are shown to ascertain the asymptotic theory and to make comparison with Anderson's procedure. Comparing with that, our procedure has almost the same efficiency and reduces the computational cost.

2. Notation and method

Let us define two polynomials

$$B(z|b)=\sum_{j=0}^{p} b_j z^j , \quad b_0=1 , \quad b_p\neq 0 , \tag{2.1}$$

$$A(z|a)=\sum_{j=0}^{q} a_j z^j , \quad a_0=1 , \quad a_q\neq 0 . \tag{2.2}$$

We assume that the roots of $B(z|b)=0$ and $A(z|a)=0$ are greater than one in absolute values and the two equations have no roots in common. Then the spectral density of $\{y_t\}$ is given by

$$f(\lambda|\phi)=\frac{\sigma^2}{2\pi}\left|\frac{A(\exp(i\lambda)|a)}{B(\exp(i\lambda)|b)}\right|^2 , \tag{2.3}$$

where $\phi=[b', a', \sigma^2]'$, $b=[b_1, \cdots, b_p]'$, $a=[a_1, \cdots, a_q]'$. For further discussions we introduce so called the circular condition

$$y_{-k}=y_{N-k} , \quad u_{-k'}=u_{N-k'} , \quad (k=1, \cdots, p,\ k'=1, \cdots, q) , \tag{2.4}$$

which is rather artificial but is known to have little effects on the asymptotic inferences. Suppose $\{y_t ; t=0, \cdots, N-1\}$ are available. Then the likelihood function is given by

$$p(y_0, \cdots, y_{N-1}|\phi)=(2\pi)^{-N/2}\prod_{k=0}^{N-1}(2\pi f_k(\phi))^{-1/2} \times\exp\left\{(-1/2)\sum_{k=0}^{N-1}(z_k/f_k(\phi))\right\} , \tag{2.5}$$

where $f_k(\phi)$ and z_k denote the values of $f(\lambda|\phi)$ and $I(\lambda)$ at $\lambda=2\pi k/N$, respectively (see Anderson [1]). This implies that z_k is independently distributed for $k=1, \cdots, [(N-1)/2]$, each as $\Gamma(1, f_k(\phi))$, where $\Gamma(\alpha, \beta)$ denotes the gamma distribution whose probability density function is given by $\{\Gamma(\alpha)\}^{-1}\beta^{-\alpha}x^{\alpha-1}\exp(-x/\beta)$.

Single z_k is an unbiased estimator of $f_k(\phi)$, but it is not consistent. So an estimator based on M_N $(=2m_N+1)$ averages of z_k, i.e.,

$$x_k=\frac{1}{M_N}\sum_{j=-m_N}^{m_N} z_{k+j}\,, \tag{2.6}$$

has been introduced, where we define that $\{M_N\}$ is a sequence of integers such that $M_N\to\infty$ and $M_N^{3/2}/N\to 0$ as $N\to\infty$. We often observe that the logarithm of an estimated spectral density is a fairly well-behaved function. Hence we utilize a statistic $\log x_k$ for the parameter estimation. Applying the results of Davis and Jones [5], it is not difficult to show that

$$\mathrm{E}\,[\log x_k]=\log f_k(\phi)+\psi(M_N)-\log M_N+O\left(\frac{M_N}{N}\right), \tag{2.7}$$

$$\mathrm{V}\,[\log x_k]=\psi'(M_N)+O\left(\left(\frac{M_N}{N}\right)^2\right) \tag{2.8}$$

and

$$\mathrm{V}\,[\sqrt{M_N}\log x_k]=1+O\left(\frac{1}{M_N}\right)+O\left(\frac{M_N^3}{N^2}\right), \tag{2.8$'$}$$

where $\psi(x)$ is the digamma function defined by $\psi(x)=d\log\Gamma(x)/dx$. Since x_k and $x_{k'}$ are dependent if $|k-k'|<M_N$, we have only K_N-1 mutually independent x_k's where $K_N=[N/2M_N]$, for example, $x_{[k]}=x_{M_N k}$ $(k=1, \cdots, K_N-1)$. Applying the method of least squares to $\log x_{[k]}$, we estimate the true parameter ϕ_0 by $\hat{\phi}$ which minimizes

$$\begin{aligned} L(\phi)&=\sum_{k=1}^{K_N-1}\{\log x_{[k]}-\log f_{[k]}(\phi)-\psi(M_N)+\log M_N\}^2 \\ &=\sum_{k=1}^{K_N-1}\left[\log x_{[k]}-\log f_{[k]}(\phi_0)-\psi(M_N)+\log M_N+\log\left\{\frac{f_{[k]}(\phi_0)}{f_{[k]}(\phi)}\right\}\right]^2 \end{aligned} \tag{2.9}$$

with respect to ϕ, where $f_{[k]}(\phi)$ is defined similarly to $x_{[k]}$.

3. Asymptotic theory

In this section we examine the asymptotic properties of the estimator $\hat{\phi}$ defined above. Noting the results of previous section, and applying the law of large numbers, we can show that $\frac{1}{K_N}L(\phi)$ converges to $\frac{1}{\pi}\int_0^\pi \left[\log\left\{\frac{f(\lambda|\phi_0)}{f(\lambda|\phi)}\right\}\right]^2 d\lambda$ in probability as $N\to\infty$. This implies that $\hat{\phi}$ is a consistent estimator of ϕ_0.

Next we see that the estimator is asymptotically normally distributed. From the definition of $\hat{\phi}$, we have

$$(3.1) \qquad \frac{\partial}{\partial\phi}L(\phi)\bigg|_{\phi=\hat{\phi}}=0\,.$$

The Taylor expansion of the left-hand side of (3.1) about $\phi=\phi_0$ is $\frac{\partial}{\partial\phi}L(\phi_0)+\frac{\partial^2}{\partial\phi\,\partial\phi'}L(\phi^*)(\hat{\phi}-\phi_0)$ where ϕ^* is on the line segment joining $\hat{\phi}$ and ϕ_0. Then we obtain

$$(3.2) \qquad \sqrt{N}(\hat{\phi}-\phi_0)=\sqrt{\frac{N}{K_N}}\left\{-\frac{1}{K_N}\frac{\partial^2}{\partial\phi\,\partial\phi'}L(\phi^*)\right\}^{-1}\frac{1}{\sqrt{K_N}}\frac{\partial}{\partial\phi}L(\phi_0)\,.$$

By (2.8)′ and the central limit theorem, we can show that $\sqrt{\frac{N}{K_N}}\frac{1}{\sqrt{K_N}}\times\frac{\partial}{\partial\phi}L(\phi_0)$ converges to $(p+q+1)$-dimensional normal random variable with zero mean and the variance matrix $16I(\phi_0)$ where $I(\phi)$ denotes the Fisher Information defined by

$$(3.3) \qquad I(\phi)=\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\phi}\log f(\lambda|\phi)\frac{\partial}{\partial\phi'}\log f(\lambda|\phi)\right\}d\lambda\,.$$

The term $-\frac{1}{K_N}\frac{\partial^2}{\partial\phi\,\partial\phi'}L(\phi^*)$ converges to $4I(\phi_0)$ in probability by the law of large numbers and the consistency of $\hat{\phi}$. Then $\sqrt{N}(\hat{\phi}-\phi_0)$ converges to $(p+q+1)$-dimensional normal random variable with zero mean and the variance matrix $I(\phi_0)^{-1}$ as $N\to\infty$. Thus we have proved the following theorem.

THEOREM. *Let $y_0,\cdots,y_{N-1}$ be generated by the* ARMA (p,q) *model with the circular condition* (2.4) *and the normality of* $\{u_t\}$. *If the sequence $\{M_N\}$ satisfies $M_N\to\infty$, $M_N^{3/2}/N\to 0$ as $N\to\infty$ and $\hat{\phi}$ is determined by minimizing the function $L(\phi)$ defined by* (2.9), *then*

i) *$\hat{\phi}$ is consistent,*

ii) *$\sqrt{N}(\hat{\phi}-\phi_0)\to N(0, I(\phi_0)^{-1})$ in law.*

That is, $\hat{\phi}$ is asymptotically efficient.

4. Iterative procedure

To obtain an estimate of ϕ_0 from the observations, we must minimize $L(\phi)$ which is a complicated nonlinear function of ϕ. We use the iterative procedure based on the Newton-Raphson method. In our procedure, the $(i+1)$th estimate $\hat{\phi}_{i+1}$ is given by the equations

$$\hat{\theta}_{i+1}=\hat{\theta}_i+\begin{bmatrix}\Psi(\hat{\theta}_i) & \Omega(\hat{\theta}_i)\\ \Omega(\hat{\theta}_i)' & \Phi(\hat{\theta}_i)\end{bmatrix}^{-1}\begin{bmatrix}\beta(\hat{\phi}_i)\\ \alpha(\hat{\phi}_i)\end{bmatrix}, \tag{4.1}$$

$$\hat{\sigma}^2_{i+1}=2\pi\exp\left[\frac{1}{K_N}\sum_{k=1}^{K_N-1}\left\{\log x_{[k]}-\log\left|\frac{A\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|\hat{a}_i\right)}{B\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|\hat{b}_i\right)}\right|^2+\log M_N-\psi(M_N)\right\}\right], \tag{4.2}$$

with $\theta=[b', a']'$, where the rth components of the vectors on the right-hand side of (4.1) are given by

$$[\beta(\phi)]_r=-\sum_{k=1}^{K_N-1}\left[\{\log x_{[k]}-\log f_{[k]}(\phi)+\log M_N-\psi(M_N)\}\times\left\{\sum_{j=0}^{p}b_j\cos\left(\frac{2\pi}{N}M_Nk(r-j)\right)\right\}\Big/\left|B\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|b\right)\right|^2\right], \tag{4.3}$$

$$[\alpha(\phi)]_r=\sum_{k=1}^{K_N-1}\left[\{\log x_{[k]}-\log f_{[k]}(\phi)+\log M_N-\psi(M_N)\}\times\left\{\sum_{j=0}^{q}a_j\cos\left(\frac{2\pi}{N}M_Nk(r-j)\right)\right\}\Big/\left|A\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|a\right)\right|^2\right], \tag{4.4}$$

and the (r, s)th components of matrices are given by

$$[\Psi(\theta)]_{rs}=\sum_{k=1}^{K_N-1}\left[\left\{\cos\left(\frac{2\pi}{N}M_Nk(r-s)\right)\right\}\Big/\left|B\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|b\right)\right|^2\right], \tag{4.5}$$

$$[\Omega(\theta)]_{rs}=-2\sum_{k=1}^{K_N-1}\left[\left\{\sum_{j=0}^{p}b_j\cos\left(\frac{2\pi}{N}M_Nk(r-j)\right)\right\}\times\left\{\sum_{j=0}^{q}a_j\cos\left(\frac{2\pi}{N}M_Nk(s-j)\right)\right\}\Big/\left|B\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|b\right)A\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|a\right)\right|^2\right], \tag{4.6}$$

$$[\Phi(\theta)]_{rs}=\sum_{k=1}^{K_N-1}\left[\left\{\cos\left(\frac{2\pi}{N}M_Nk(r-s)\right)\right\}\Big/\left|A\left(\exp\left(i\frac{2\pi}{N}M_Nk\right)\Big|a\right)\right|^2\right]. \tag{4.7}$$

The terms β and α are the first order derivatives of $L(\phi)$, and Ψ, Ω

and Φ are the second order derivatives, while Ψ and Φ are slightly modified in the same way as the scoring method (see Anderson [1]). When K_N is large enough, we can substitute M_N/N by $1/2K_N$ and the summation from $k=1$ to K_N-1 by the summation from $k=0$ to K_N-1. The effect of these modifications is negligibly small. If we use this modified version, calculation amount can be reduced by applying the FFT algorithm, for example, given by Singleton [10]. Note that this procedure has the same form as Anderson's or the scoring procedure, in which, however, all the components corresponding to (4.3)–(4.7) need summations of N terms instead of our K_N.

As this procedure is iterative, we must have an initial estimate at the first stage. Here we consider the following two important equations for the ARMA (p, q) model (1.1):

$$\sum_{j=0}^{p} b_j\sigma(s-j)=0\,, \qquad s\geqq q+1\,, \tag{4.8}$$

$$\sum_{j=0}^{q} a_j\tau(s-j)=0\,, \qquad s\geqq p+1\,, \tag{4.9}$$

where

$$\sigma(r)=\int_{-\pi}^{\pi} f(\lambda|\phi)\exp(i\lambda r)d\lambda$$

and

$$\tau(r)=\int_{-\pi}^{\pi} \frac{1}{f(\lambda|\phi)}\exp(i\lambda r)d\lambda\,.$$

Then the statistics

$$c_r=\frac{1}{N}\sum_{t=0}^{N-|r|-1} y_t y_{t+|r|}$$

and

$$d_r=\frac{2\pi}{K_N}\sum_{k=0}^{K_N-1}\frac{1}{x_{[k]}}\cos\frac{\pi}{K_N}kr$$

are eligible for consistent estimates of $\sigma(r)$ and $\tau(r)$ respectively. If we substitute these estimates for $\{\sigma(r)\}$ and $\{\tau(r)\}$ of equations (4.8) and (4.9), the solutions of these linear equations can be used as initial consistent estimates.

5. Simulation results

We carried out some computer simulations to investigate the behaviour of the estimate outlined above. Observations were generated from the ARMA (1, 1) model

$$y_t + b_1 y_{t-1} = u_t + a_1 u_{t-1} \tag{5.1}$$

with $b_1 = -0.8$ and $a_1 = 0.5$, where $\{u_t\}$ consists of independent $N(0, 1)$ variables. We made one hundred independent sets of sequences for the length of 90, 180 and 360. To each of these sets we applied the estimation procedures for $M_N = 3$, 5 and 9. At the same time, Anderson's and the scoring procedures were applied for comparison. For all the procedures, initial estimates were obtained by the procedure stated in Section 4 and the results were given after five iterations. Here we concern the behaviour of $\hat{b}_1$ and $\hat{a}_1$, they are summarized in the table below.

Mean, Variance and Mean square error

N	90		180		360	
	b_1	a_1	b_1	a_1	b_1	a_1
$M_N=3$	−0.772 0.647 0.725	0.461 1.74 1.89	−0.783 0.350 0.378	0.473 0.707 0.778	−0.789 0.165 0.176	0.490 0.312 0.322
5	−0.761 0.507 0.660	0.468 1.39 1.50	−0.778 0.298 0.347	0.481 0.636 0.673	−0.788 0.171 0.186	0.492 0.299 0.305
9	−0.734 0.405 0.843	0.469 1.03 1.13	−0.774 0.210 0.276	0.483 0.590 0.621	−0.789 0.159 0.172	0.490 0.310 0.320
Anderson	−0.776 0.577 0.634	0.450 1.29 1.54	−0.784 0.246 0.271	0.474 0.557 0.626	−0.790 0.150 0.161	0.489 0.282 0.295
Scoring	−0.780 0.552 0.594	0.440 1.50 1.86	−0.785 0.228 0.252	0.471 0.619 0.751	−0.790 0.150 0.161	0.487 0.347 0.365
	−0.8 0.464	0.5 0.966	−0.8 0.232	0.5 0.486	−0.8 0.116	0.5 0.242

In this table, three numbers in a group denote respectively, the mean, a hundred times of the variance and a hundred times of the mean square error, from a hundred repetitions. The two rows at the bottom of this table show the asymptotic mean and a hundred times of the asymptotic variance of the maximum likelihood estimator. There are a little differences between Anderson's and the scoring procedures, but we do not discuss this point further, for the main purpose of this section is to examine our procedure. All the variances and the mean square errors of Anderson's procedure are smaller than those of ours except the cases of $(N, M_N) = (90, 5)$, $(90, 9)$ and $(180, 9)$. But the differences become smaller as both N and M_N increase. For example, in the case of $N = 360$ and $M_N = 9$, our procedure can supply almost the same estimates as Anderson's or the scoring procedure, while the computational time required for our procedure was about 1/4 of that for Anderson's or the scoring procedure in our experiment. Considering these simulation results, it may be concluded that K_N needs to be larger than 10 for our estimate to have the asymptotic properties discussed above.

Acknowledgment

The auther wishes to express his thanks to Professor T. Okuno and Associate Professor C. Hirotsu of the University of Tokyo for their advice and encouragement. He also thanks the referees whose comments greatly improved the earlier version of this paper.

TECHNICAL COLLEGE OF THE UNIVERSITY OF TOKUSHIMA

REFERENCES

[1] Anderson, T. W. (1977). Estimation for autoregressive moving average models in the time and frequency domains, *Ann. Statist.*, **5**, 842-865.

[2] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series, *Biometrika*, **60**, 217-226.

[3] Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications, *Technometrics*, **14**, 277-298.

[4] Clevenson, M. L. (1970). Asymptotically efficient estimates of the parameters of a moving average time series, Ph.D. Dissertation, Department of Statistics, Stanford University.

[5] Davis, H. T. and Jones, R. H. (1968). Estimation of the innovation variance of a stationary time series, *J. Amer. Statist. Ass.*, **63**, 141-149.

[6] Durbin, J. (1960). The fitting of time-series models, *Rev. Inst. Internat. Statist.*, **28**, 233-244.

[7] Hannan, E. J. (1969). The estimation of mixed moving average autoregressive systems, *Biometrika*, **56**, 579-593.

[8] Hannan, E. J. and Nicholls, D. F. (1977). The estimation of the prediction error variance, *J. Amer. Statist. Ass.*, **72**, 834-840.

[9] McClave, J. T. (1974). A comparison of moving average estimation procedures, *Commun. Statist.*, **3**, 865-883.

[10] Singleton, R. C. (1969). An algorithm for computing the mixed radix fast Fourier transform, *IEEE Trans. Audio and Electro.*, AU-**17**, 93-103.

[11] Wahba, G. (1980). Automatic smoothing of the log periodogram, *J. Amer. Statist. Ass.*, **75**, 122-132.

[12] Walker, A. M. (1962). Large-sample estimation of parametes for autoregressive processes with moving-average residuals, *Biometrika*, **49**, 117-131.