한국통계학회논문집 제2권 2호
1995년 12월 pp. 85-93

An Algorithm for Hannan and Rissanen's ARMA Modeling Method

Chul Eung Kim and Byoung Seon Choi[1)]

Abstract

Hannan and Rissanen proposed an innovation regression method of ARMA modeling, which is composed of three stages. Its second-stage is to choose orders of the ARMA model using the BIC, which needs a lot of calculation to estimate several regression models. We are going to present a simple and efficient algorithm for the second stage using a special property of triangular Toeplitz matrices.

1. Introduction.

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

$$\phi(B)y_t = \theta(B)v_t, \tag{1}$$

where $\phi(B)=-\phi_0-\phi_1B-\cdots-\phi_pB^p$, $\theta(B)=-\theta_0-\theta_1B-\cdots-\theta_qB^q$, $\phi_0=\theta_0=-1$, $\phi_p \neq 0$, $\theta_q \neq 0$, B is the backshift operator, and $\{v_t\}$ is a Gaussian white noise process with means 0 and variances $\sigma^2 (>0)$. We assume that the model is stationary and invertible, *i.e.*, the equations $\phi(z)=0$ and $\theta(z)=0$ have all the roots outside the unit circle. Also we assume that the two equations have no common root. This assumption is sometimes called coprimal. Since the process is assumed to be stationary, the autocovariance function(ACVF) and the autocorrelation function(ACRF) are defined as

$$\sigma(j)=cov(y_t, y_{t+j}), \quad j=0,\pm1,\pm2,\cdots,$$

$$\rho_j=\sigma(j)/\sigma(0), \quad j=0,\pm1,\pm2,\cdots.$$

When a T-realization $\{y_1,\cdots,y_T\}$ of the ARMA model is given, the sample ACVF and the sample ACRF are defined by

$$\hat{\sigma}(j)=\hat{\sigma}(-j)=\frac{1}{T}\sum_{t=1}^{T-j}y_t y_{t+j}, \quad j=0,1,\cdots,$$

$$\hat{\rho}_j=\hat{\sigma}(j)/\hat{\sigma}(0), \quad j=0,\pm1,\pm2,\cdots.$$

1) Department of Applied Statistics, Yonsei University, Seoul 120-749, KOREA.

Hannan and Rissanen (1982) proposed a three-stage ARMA modeling method by fitting the current observation to past observations and estimated values of past innovations. We are going to discriminate the estimates of the three stages by accents. For example, the first-stage, the second-stage, and the third-stage estimates of μ are denoted by $\check{\mu}$, $\tilde{\mu}$, and $\hat{\mu}$, respectively.

1. The innovation $\check{v}_t$ is calculated by fitting a long AR model. Let N_T be an upper bound of the AR orders such that N_T increases as T does, but it does not increase too quickly. It is sufficient to assume

$$N_T < (\ln T)^a, \quad 0<a<\infty.$$

For $n\ (\le N_T)$, fit the AR(n) model

$$y_t = \sum_{l=1}^{n} \beta_l y_{t-1} + v_t$$

to the observations $\{y_1, \cdots, y_T\}$. Let $\check{\beta}_1, \cdots, \check{\beta}_n$ be the YW estimate of $\beta_1, \cdots, \beta_n$. Then, calculate the estimated innovations

$$\check{v}_t = -\sum_{l=0}^{n} \check{\beta}_l y_{t-l}, \quad t = n+1, \cdots, T,$$

where $\check{\beta}_0=-1$. It is recommended to choose the AR order n using the BIC or the AIC. Due to its overparameterization tendency the AIC may be more practical.

2. Let K and I be sufficiently large so that they are greater than the true orders p and q, respectively. For each $(k,i) \in \{(k,i) | k=0, \cdots, K, \ i=0, \cdots, I\}$, calculate the ordinary least square estimates $\tilde{\phi}_{k,i}, \cdots, \tilde{\phi}_{k,k}, \tilde{\theta}_{i,1}, \cdots, \tilde{\theta}_{i,i}$, minimizing

$$\frac{1}{T} \sum_{t=t_0+1}^{T} \left(y_t - \sum_{j=1}^{k} \phi_{k,j} y_{t-j} + \sum_{l=1}^{i} \theta_{i,l} \check{v}_{t-l} \right)^2,$$

where $t_0 = \max(n+k, n+i)$. Denote its minimum by $\tilde{\sigma}^2_{k,i}$. Choose $\tilde{p}$ and $\tilde{q}$ minimizing

$$BIC(k,i) = \ln \tilde{\sigma}^2_{k,i} + (k+i) \frac{\ln T}{T}$$

among $k=0, \cdots, K$ and $i=0, \cdots, I$. The estimated innovations in this stage are defined by

$$\tilde{v}_t = y_t = 0, \qquad t \le 0,$$
$$\tilde{v}_t = y_t - \sum_{j=1}^{\tilde{p}} \tilde{\phi}_{\tilde{p},j} y_{t-l} + \sum_{l=1}^{\tilde{q}} \tilde{\theta}_{\tilde{q},l}\, \tilde{v}_{t-l}, \qquad t = 1, \cdots, T.$$

3. Using the estimates of the second stage as initial values, we apply maximum likelihood techniques. One of them is as follows. Calculate $\{x_t\}$ and $\{z_t\}$ by

$$x_t = z_t = 0, \quad t \le 0,$$
$$x_t = -\sum_{l=1}^{\tilde{q}} \tilde{\theta}_{\tilde{q},l} x_{t-l} + y_t, \quad t = 1, \cdots, T,$$
$$z_t = -\sum_{l=1}^{\tilde{q}} \tilde{\theta}_{\tilde{q},l} z_{t-l} + v_t, \quad t = 1, \cdots, T.$$

Regress $\tilde{v}_t + x_t - z_t$ on

$$x_{t-1}, \cdots, x_{t-\tilde{p}}, -z_{t-1}, \cdots, -z_{t-\tilde{q}}$$

for $t = t_1+1, \cdots, T$, where $t_1 = \max(\tilde{p}, \tilde{q})$, in order to obtain the refined estimates $\hat{\phi}_{\tilde{p},1}, \cdots, \hat{\phi}_{\tilde{p},\tilde{p}}, \hat{\theta}_{\tilde{q},1}, \cdots, \hat{\theta}_{\tilde{q},\tilde{q}}$. The corresponding estimate of the white noise variance is

$$\hat{\sigma}^2_{\tilde{p},\tilde{q}} = \frac{1}{T} \sum_{t=t_1+1}^{T} \left\{ y_t - \sum_{j=1}^{\tilde{p}} \hat{\phi}_{\tilde{p},j} y_{t-j} + \sum_{l=1}^{\tilde{q}} \hat{\theta}_{\tilde{q},l}\, \hat{v}_{t-l} \right\}^2$$

where $\hat{v}_t$ is the re-estimated residuals.

The three-stage method is essentially due to Durbin (1960), who, however, gave no rule for determining the long AR order n and took p and q given. Hannan and Kavalieris (1984) recommended to repeat the second stage to obtain consistent estimates of the orders.

In the second stage it is necessary to calculate

$$\{ \tilde{\sigma}_{k,i} \mid k = 0, \cdots, K, \; i = 0, \cdots, I \} .$$

Since this stage needs a lot of calculation, Hannan and Rissanen proposed to calculate them for only the case $k = i$ using a modified Whittle algorithm. In Section 2, we present a computationally efficient algorithm to calculate $\tilde{\sigma}^2_{k,i}$ for $k = 0, \cdots, K$ and $i = 0, \cdots, I$.

2. Algorithm

It is well-known that the AR parameters satisfy the extended Yule-Walker equations

$$\rho_j = \phi_1 \rho_{j-1} + \cdots + \phi_p \rho_{j-p}, \quad j = q+1, q+2, \cdots. \tag{2}$$

Since the ARMA(p,q) process is stationary, it can be represented by the MA(∞) model

$$y_t = \psi(B)v_t,$$

where

$$\psi(B) = \phi^{-1}(B)\theta(B) = \sum_{j=0}^{\infty}\psi_j B^j.$$

More precisely,

$$\phi_0\psi_j + \phi_1\psi_{j-1} + \cdots + \phi_j\psi_0 = \begin{cases}\theta_j, & j = 0,\cdots,q,\\ 0, & j = q+1, q+2,\cdots,\end{cases}$$

where ϕ_j is assumed to be 0 for $j = p+1, p+2, \cdots$. It is known (Choi [1986]) that the parameters and the ACVF satisfy the relations

$$\begin{aligned}\phi_0\sigma(j) &+ \phi_1\sigma(j-1) + \cdots + \phi_j\sigma(j-p)\\ &= (\psi_0\theta_j + \psi_1\theta_{j+1} + \cdots + \psi_{q-j}\theta_q)\sigma^2, \quad j = 0,\cdots,q.\end{aligned} \tag{3}$$

Since

$$cov(v_t, y_{t+j}) = \sigma^2\psi_j, \quad j = 0,1,\cdots,$$

the first-stage estimates of σ^2 and ψ_j are

$$\check{\sigma}^2 = \frac{1}{T}\sum_t^T \check{v}_t^2$$

$$\check{\psi}_j = \frac{1}{\sigma^2}\left(\frac{1}{T}\sum_t^T \check{v}_t y_{t+j}\right), \quad j = 1,2,\cdots.$$

Our purpose is to obtain $\tilde{\sigma}^2_{k,i}$ using these estimates.

To obtain AR coefficients for several pairs of the orders, it is necessary to generalize the extended Yule-Walker equations. For $k = 1,2,\cdots$ and $i = 0,1,\cdots$, we define a k-dimensional Toeplitz matrix and two vectors as

$$B(k,i) = \begin{pmatrix}\rho_i & \rho_{i-1} & \cdots & \rho_{i-k+1}\\ \rho_{i+1} & \rho_i & \cdots & \rho_{i-k+2}\\ \vdots & \vdots & & \vdots\\ \rho_{i+k-1} & \rho_{i+k-2} & \cdots & \rho_i\end{pmatrix},$$

$$\boldsymbol{\rho}(k,i) = (\rho_{i+1},\cdots,\rho_{i+k})^t,$$

$$\boldsymbol{\gamma}(k,i) = (\rho_{i-k},\cdots,\rho_{i-1})^t.$$

Also, we denote $\tilde{\boldsymbol{x}} = (x_n,\cdots,x_1)^t$ for any vector $\boldsymbol{x} = (x_1,\cdots,x_n)^t$. If $B(k,i)$ is nonsingular, then we let

$$\boldsymbol{\phi}(k,i) = B^{-1}(k,i)\ \boldsymbol{\rho}(k,i),$$

$$\tilde{\boldsymbol{\pi}}(k,i) = B^{-1}(k,i)\ \boldsymbol{\gamma}(k,i),$$

$$\theta(k,i) = \rho_{i+k+1} - \tilde{\boldsymbol{\phi}}(k,i)^t\ \boldsymbol{\rho}(k,i),$$

$$\eta(k,i) = \rho_{i-k-1} - \boldsymbol{\pi}(k,i)^t\ \boldsymbol{\gamma}(k,i),$$

$$\lambda(k,i) = \rho_i - \boldsymbol{\pi}(k,i)^t\ \boldsymbol{\rho}(k,i).$$

For $k = 0$ and $i = 0,1,\cdots$, we let

$$\theta(0,i) = \rho_{i+1},$$

$$\eta(0,i) = \rho_{i-1},$$

$$\lambda(0,i) = \rho_i.$$

We denote the jth elements of $\boldsymbol{\phi}(k,i)$ and $\boldsymbol{\pi}(k,i)$ by $\phi_{k,j}^{(i)}$ and $\pi_{k,j}^{(i)}$ for $j = 1,\cdots,k$. Also, let $\phi_{k,0}^{(i)} = \pi_{k,0}^{(i)} = -1$ for each (k,i).

For $k = 1,2,\cdots$ and $i = 0,1,\cdots$, we can calculate $\{\phi_{k,1}^{(i)},\cdots,\phi_{k,k}^{(i)}\}$ using a simplified Trench-Zohar algorithm as follows.

Initial values for recursion:

$$\phi_{1,1}^{(i)} = \frac{\theta(0,i)}{\lambda(0,i)},$$

$$\pi_{1,1}^{(i)} = \frac{\eta(0,i)}{\lambda(0,i)},$$

$$\lambda(1,i) = \lambda(0,i)(1-\phi_{1,1}^{(i)}\pi_{1,1}^{(i)}).$$

For $k = 1,2,\cdots$,

$$\theta(k,i) = \rho_{i+k+1} - \phi_{k,1}^{(i)}\rho_{i+k} - \cdots - \phi_{k,k}^{(i)}\rho_{i+1},$$

$$\eta(k,i) = \rho_{i-k-1} - \pi_{k,1}^{(i)}\rho_{i-k} - \cdots - \pi_{k,k}^{(i)}\rho_{i-1},$$

$$\phi_{k+1,k+1}^{(i)} = \frac{\theta(k,i)}{\lambda(k,i)},$$

$$\pi_{k+1,k+1}^{(i)} = \frac{\eta(k,i)}{\lambda(k,i)},$$

$$\lambda(k+1,i) = \lambda(k,i)\{1-\phi_{k+1,k+1}^{(i)}\pi_{k,k+1-j}^{(i)}\}.$$

For $j = 1,\cdots,k$,

$$\phi_{k+1,j}^{(i)} = \phi_{k,j}^{(i)} - \phi_{k+1,k+1}^{(i)}\pi_{k,k+1-j}^{(i)},$$

$$\pi_{k+1,j}^{(i)} = \pi_{k,j}^{(i)} - \pi_{k+1,k+1}^{(i)}\phi_{k,k+1-j}^{(i)}.$$

We are going to renovate it so that it is not necessary to calculate the dummy sequence $\{\pi_{k,j}^{(i)}\}$. If $i = 0$, then

$$\boldsymbol{\pi}(k,0) = \boldsymbol{\phi}(k,0).$$

If $i = 1,2,\cdots$, then the definition of $\boldsymbol{\phi}(k,i-1)$ implies

$$\sum_{j=1}^{k}\left(-\frac{\phi_{k,k-j}^{(i-1)}}{\phi_{k,k}^{(i-1)}}\right)\boldsymbol{\rho}(k,i-1-k+j) = \boldsymbol{\rho}(k,i-1-k).$$

If we let

$$\pi_{k,j}^{(i)} = -\frac{\phi_{k,k-j}^{(i-1)}}{\phi_{k,k}^{(i-1)}},$$

then the last equation is equivalent to

$$B(k,i)\,\tilde{\boldsymbol{\pi}}(k,i) = \boldsymbol{r}(k,i).$$

If we assume the nonsingularity of $B(k,i)$, then

$$\pi_{k,j}^{(i)} = -\frac{\phi_{k,k-j}^{(i-1)}}{\phi_{k,k}^{(i-1)}}, \quad j = 1,\cdots,k.$$

In summary we obtain the following algorithm for the AR coefficients. As defined in Trench (1983), a matrix is called strongly nonsingular if its principle submatrices are nonsingular.

Algorithm 1. *An Algorithm for the AR parameters*
Assume that $B(k,i)$ is strongly nonsingular, then we can calculate the AR coefficients as follows.

For $i = 0$, use the Levinson-Durbin Algorithm to calculate

$$\{\phi_{k,j}^{(0)} \mid k = 1,2,\cdots,\ j = 1,\cdots,k\}.$$

For $i = 1,2,\cdots$,

For $k = 0$, let

$$\lambda(0,i) = \rho_i,$$

$$\phi_{1,1}^{(i)} = \frac{\rho_{i+1}}{\rho_i}.$$

For $k = 1,2,\cdots$,

$$\theta(k,i) = \rho_{i+k+1} - \phi_{k,1}^{(i)}\rho_{i+k} \cdots - \phi_{k,k}^{(i)}\rho_{i+1},$$

$$\lambda(k,i) = \lambda(k-1,i)\left\{1 - \frac{\phi_{k,k}^{(i)}}{\phi_{k,k}^{(i-1)}}\right\},$$

$$\phi_{k+1,k+1}^{(i)} = \frac{\theta(k,i)}{\lambda(k,i)}.$$

For $j = 1,2,\cdots,k$,

$$\phi_{k+1,j}^{(i)} = \phi_{k,j}^{(i)} + \phi_{k+1,k+1}^{(i)} \frac{\phi_{k,j-1}^{(i-1)}}{\phi_{k,k}^{(i-1)}}. \qquad \square$$

For a finite sample case, the estimate of $B(k,i)$ is strongly nonsingular with probability 1. Thus, the strong nonsingularity assumption is not so crucial in practical analyses. When a T-realization $\{y_1,\cdots,y_T\}$ is given, we obtain the second-stage estimates $\{\tilde{\phi}_{k,1}^{(i)},\cdots,\tilde{\phi}_{k,k}^{(i)}\}$ by replacing the ACRF with the sample ACRF in Algorithm 1.

Let $\Sigma(i+1,k+1;0)$, $\hat{\Sigma}(i+1,k+1;0)$, and $\Psi_{k,i}$ be $(i+1)\times(k+1)$ matrices, whose (r,s) elements are

$$(\Sigma(i+1,k+1;0))_{r,s} = \sigma(r-s),$$

$$(\hat{\Sigma}(i+1,k+1;0))_{r,s} = \hat{\sigma}(r-s),$$

$$(\Psi_{k,i})_{r,s} = \begin{cases} 0, & r < s, \\ \psi_{r-s}, & otherwise, \end{cases}$$

respectively. Also let Ψ_i be a $(i+1)\times(i+1)$ matrix whose (r,s) element is

$$(\Psi_i)_{r,s} = \begin{cases} 0, & r > s, \\ \psi_{s-r}, & otherwise. \end{cases}$$

If the underlying process is from the ARMA(k,i) model, then Equation (2) becomes

$$\Sigma(I+1,K+1;0)\,\phi_*(k,i) = \sigma^2\Psi_i\Psi_{k,i}\,\phi_*(k,i),$$

where

$$\phi_*(k,i) = (\phi_{k,0}^{(i)},\phi_{k,1}^{(i)},\cdots,\phi_{k,k}^{(i)})^t.$$

Let

$$c_0 = \sigma, c_1 = \sigma\psi_1, \cdots, c_i = \sigma\psi_i, \; C_i = \sigma\Psi_i, \; C_{k,i} = \sigma\Psi_{k,i}.$$

Then

$$C_iC_{k,i}\,\phi_*(k,i) = \Sigma(i+1,k+1;0)\,\phi_*(k,i).$$

It has been shown (Choi [1986]) that the inverse matrix of an upper triangular Toeplitz matrix is also upper triangular Toeplitz. Using this property we can efficiently calculate the inverse matrix of C_i. If we let

$$D_i = C_i^{-1},$$

$$d_0 = \frac{1}{c_0},$$

$$d_j = -\frac{1}{c_0}(c_jd_0 + \cdots + c_1d_{j-1}), \quad j = 1,2,\cdots,$$

then the (r,s) element of D_i is

$$(D_i)_{r,s} = \begin{cases} 0, & r > s, \\ d_{s-r}, & otherwise. \end{cases}$$

It should be noted that each element of D_i does not depend on i. We can calculate the second-stage estimate $\tilde{C}_{k,i}$ by

$$\tilde{C}_{k,i}\ \tilde{\boldsymbol{\Phi}}_*(k,i) = \breve{D}_i\ \breve{\Sigma}(i+1,k+1;0)\ \tilde{\boldsymbol{\Phi}}_*(k,i).$$

In the second-stage of the Hannan and Rissanen algorithm it is sufficient to estimate σ, or equivalently c_0. It can be done by considering only the first element of the vector equation, *i.e.*,

$$\tilde{\sigma}_{k,i} = -\sum_{r=0}^{i}\sum_{s=0}^{k} \breve{d}_r\ \hat{\sigma}(r-s)\ \tilde{\phi}_{k,s}^{(i)}.$$

3. Comments

The algorithm presented in the previous section is simpler than the modified Whittle algorithm, and it makes applications of Hannan and Rissanen's method computationally easier. It may be worth mentioning that the MA coefficients of the ARMA model can be obtained also through the Newton-Raphson algorithm.

Algorithm 2. *A Newton-Raphson algorithm*

For $j = 0,\cdots,q$, let T_j and S_j be $(q+1)\times(p+1)$ matrices, whose (r,s) elements are

$$(T_j)_{r,s} = \begin{cases} c_{j+r-s}, & 1 \le s \le j+r,\ 1 \le r \le q-j+1, \\ 0, & otherwise, \end{cases}$$

$$(S_j)_{r,s} = \begin{cases} c_{j-r+s}, & 1 \le r \le q+1,\ \max\{1,r-j\} \le s \le q-j+1, \\ 0, & otherwise, \end{cases}$$

respectively. Also, let

$$\boldsymbol{c}_q = (c_0, c_1, \cdots, c_q)^t,$$

$$U_j = T_j + S_j,$$

$$W = (U_0\ \boldsymbol{\Phi}_*,\ \cdots,\ U_q\ \boldsymbol{\Phi}_*).$$

Then the Newton-Raphson solution of

$$C_q C_{p,q}\, \phi_*(p,q) = \Sigma(q+1,p+1;0)\, \phi_*(p,q)$$

is obtained by the recursive equation

$$c_q^{(n+1)} = \frac{1}{2} c_q^{(n)} + (W^{(n)})^{-1} \Sigma(q+1,p+1;0)\, \phi_*(p,q),$$

where the superscript (n) means the value at the nth iteration. □

Wilson (1969) has derived the Newton-Raphson algorithm for a pure MA process which is a special case of Algorithm 2. It is worth mentioning that Algorithm 2 shows Equation (4) of Wilson's paper can be simplified as

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

References

[1] Choi, B. S. (1985). An algorithm for solving the extended Yule-Walker equations of an autoregressive moving-average time series, *IEEE Transactions on Information Theory*, IT-32, 417-419.

[2] Durbin, J. (1960). The fitting of time series models, *International Statistical Review*, Vol. 28, 234-244.

[3] Hannan, E. J. and L. Kavalieris (1984). A method for autoregressive-moving average estimation, *Biometrika*, Vol. 72, 273-280.

[4] Hannan, E, J. and J. Rissanen (1982). Recursive estimation of mixed autoregressive moving average order, *Biometrika*, Vol. 69, 81-94. Correction; (1983), Vol. 70, 303.

[5] Trench, W. F. (1964). An algorithm for the inversion of finite Hankel matrices, *SIAM Journal*, Vol. 13, 1102-1107.

[6] Whittle, P. (1963). On the fitting of multivariate autoregression, and the approximate canonical factorization of a special density matrix, *Biometrika*, Vol. 50, 129-134.

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

[8] Zohar, S. (1969). Toeplitz matrix inversion: The algorithm of W. F. Trench, *Journal of the Association for Computing Machinery*, Vol. 16, 592-601.