Journal of the Korean
Statistical Society
Vol. 26, No. 1, 1997

The Asymptotic Distribution of the Extended Yule-Walker Estimates of a Mixed ARMA Process

Chul Eung Kim and ByoungSeon Choi [1]

ABSTRACT

The asymptotic distribution of the extended Yule-Walker estimates of a mixed ARMA process is derived. Also, a recursive algorithm is derived to calculate the extended Yule-Walker estimates recursively, which is a generalization of the Levinson-Durbin algorithm.

Key Words : Autoregressive moving-average; Extended Yule-Walker equations; Asymptotic distribution; Consistency.

1. INTRODUCTION

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

$$\phi(B)y_t = \theta(B)v_t, \tag{1.1}$$

where $\phi(z) = -\phi_0 - \phi_1 z - \cdots - \phi_p z^p$, $\theta(z) = -\theta_0 - \theta_1 z - \cdots - \theta_q z^q$, $\phi_0 = \theta_0 = -1$, $\phi_p \neq 0$, $\theta_q \neq 0$ and B is the backshift operator and $\{v_t\}$ is a sequence of

[1]Department of Applied Statistics Yonsei University, Seoul, 120-749, Korea

independent and identically distributed random variables with mean 0 and finite variance $\sigma^2(> 0)$. 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, assume that the two equations have no common roots. Define the autocovariance function (ACVF) and the autocorrelation function (ACRF) by

$$\sigma(j) = Cov(y_t, y_{t+j}), \quad j = 0, \pm 1, \pm 2, \cdots$$

and

$$\rho_j = \sigma(j)/\sigma(0), \quad j = 0, \pm 1, \pm 2, \cdots,$$

respectively. It is well-known that the ACVF satisfies the extended Yule-Walker (EYW) equations

$$\sigma(j) = \phi_1 \sigma(j-1) + \cdots + \phi_p \sigma(j-p), \quad j = q+1, q+2, \cdots . \tag{1.2}$$

Let $\{y_1, y_2, \cdots, y_T\}$ be a T-realization of a stationary process with mean 0. Then, the ACVF and ACRF are estimated by

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

and

$$\hat{\rho}_j = \hat{\sigma}(j)/\hat{\sigma}(0), \quad j = 0, \pm 1, \pm 2, \cdots,$$

respectively. The parameters $\phi_1, \phi_2, \cdots, \phi_p$ can be estimated by a method of moments, *i.e.*, by solving the sample EYW equations. Since the orders p and q are not known a priori, it is necessary to solve the sample EYW equations for several pairs of orders. Define the EYW estimates $\hat{\phi}_{k,1}^{(i)}, \hat{\phi}_{k,2}^{(i)}, \cdots, \hat{\phi}_{k,k}^{(i)}$ as the solutions of the simultaneous equations

$$\hat{\sigma}(j) = \hat{\phi}_{k,1}^{(i)} \hat{\sigma}(j-1) + \hat{\phi}_{k,2}^{(i)} \hat{\sigma}(j-2) + \cdots + \hat{\phi}_{k,k}^{(i)} \hat{\sigma}(j-k), \quad j = i+1, \cdots, i+k . \tag{1.3}$$

It has been shown (see, *e.g.*, Choi [1992, Chapter 5]) that EYW estimates $\hat{\phi}_{k,1}^{(i)}, \hat{\phi}_{k,2}^{(i)}, \cdots, \hat{\phi}_{k,k}^{(i)}$ are useful to identify an ARMA model. The asymptotic distribution of the EYW estimates should be derived to identify an ARMA model statistically through pattern identification methods. Mann and Wald (1943) derived the asymptotic distribution of the estimates for a pure autoregressive (AR) model, *i.e.*, $q = 0$. Choi (1992, p. 16) has proposed the asymptotic distribution for a mixed ARMA model, *i.e.* $q \neq 0$. However, its derivation has not been presented yet. The main purpose of this paper is to present its derivation.

2. THEOREM AND PROOF

For a positive integer k, define a $k \times k$ Toeplitz matrix $\Sigma(k,i)$ and a vector by

$$\Sigma(k,i) = \begin{bmatrix} \sigma(i) & \sigma(i-1) & \cdots & \sigma(i-k+1) \\ \sigma(i+1) & \sigma(i) & \cdots & \sigma(i-k+2) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(i+k-1) & \sigma(i+k-2) & \cdots & \sigma(i) \end{bmatrix}$$

and

$$\sigma(k,i) = (\sigma(i+1), \sigma(i+2), \cdots, \sigma(i+k))^t,$$

respectively. Let

$$\phi = (\phi_1, \phi_2, \cdots, \phi_p)^t,$$
$$\theta = (\theta_1, \theta_2, \cdots, \theta_q)^t,$$
$$\theta_* = (\theta_0, \theta_1, \cdots, \theta_q)^t.$$

The corresponding estimates are denoted by $\hat{\Sigma}(k,i)$, $\hat{\sigma}(k,i)$, $\hat{\phi}$, $\hat{\theta}$ and $\hat{\theta}_*$, respectively. For $k = p, p+1, \cdots$ and $i = q, q+1, \cdots$, define k-dimensional column vectors by

$$\sigma(k,i) = (\phi_1, \phi_2, \cdots, \phi_p, 0, \cdots, 0)^t$$

and

$$\hat{\phi}(k,i) = (\hat{\phi}_{k,1}^{(i)}, \hat{\phi}_{k,2}^{(i)}, \cdots, \hat{\phi}_{k,k}^{(i)})^t.$$

Then, Equation 1.3 implies

$$\hat{\phi}(k,i) = \hat{\Sigma}(k,i)^{-1}\hat{\sigma}(k,i).$$

It is known (see, *e.g.*, Choi [1992, p. 16]) that $\Sigma(p,q)$, $\Sigma(p+1,q), \cdots$ and $\Sigma(p,q+1)$, $\Sigma(p,q+2), \cdots$ are nonsingular. Thus, the consistency of the sample ACVF implies the consistency of the EYW estimates as follows.

Lemma 1. Let $\{y_1, y_2, \cdots, y_T\}$ be a T-realization from the $ARMA(p,q)$ model in Equation 1.1. If $(k,i) \in \{(k,q) \mid k = p, p+1, \cdots\} \bigcup \{(p,i) \mid i = q, q+1, \cdots\}$, then $\hat{\phi}(k,i)$ is consistent to $\sigma(k,i)$. □

If $k = p+1, p+2, \cdots$ and $i = q+1, q+2, \cdots$, then the consistency of $\hat{\phi}(k,i)$ does not hold due to the singularity of $\Sigma(k,i)$.

Using Lemma 1 and the Cramér-Wold device, one can derive the asymptotic distribution of the EYW estimates as follows.

Theorem 1. Asymptotic distribution of the EYW estimates

Let $\{y_1, \cdots, y_T\}$ be a T-realization from the $ARMA(p,q)$ model in Equation 1.1. If $(k,i) \in \{(k,q) \mid k = p, p+1, \cdots\} \bigcup \{(p,i) \mid i = q, q+1, \cdots\}$, then $\sqrt{T}\{\hat{\phi}(k,i) - \phi(k,i)\}$ is asymptotically normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $\sigma^2 \Sigma(k,i)^{-1} \Xi (\Sigma(k,i)^{-1})^t$, where (r,s) element of Ξ is $\theta_*^t \, \Sigma(q+1, r-s)\theta_*$. □

Proof. If $(k,i) \in \{(k,q) \mid k = p, p+1, \cdots\} \bigcup \{(p,i) \mid i = q, q+1, \cdots\}$, then $\hat{\sigma}(k,i)$ can be written as

$$\frac{1}{T}\begin{pmatrix} \sum_{t=1}^{T-i-1} y_t y_{t+i} & \sum_{t=1}^{T-i-1} y_t y_{t+i-1} & \cdots & \sum_{t=1}^{T-i-1} y_t y_{t+i-p+1} \\ \sum_{t=1}^{T-i-2} y_t y_{t+i+1} & \sum_{t=1}^{T-i-2} y_t y_{t+i} & \cdots & \sum_{t=1}^{T-i-2} y_t y_{t+i-p+2} \\ \vdots & \vdots & & \vdots \\ \sum_{t=1}^{T-i-k} y_t y_{t+i+k-1} & \sum_{t=1}^{T-i-k} y_t y_{t+i+k-2} & \cdots & \sum_{t=1}^{T-i-k} y_t y_{t+i+k-p} \end{pmatrix}\phi$$

$$-\frac{1}{T}\begin{pmatrix} \sum_{t=1}^{T-i-1} y_t v_{t+i+1} & \sum_{t=1}^{T-i-1} y_t v_{t+i} & \cdots & \sum_{t=1}^{T-i-1} y_t v_{t+i+1-q} \\ \sum_{t=1}^{T-i-2} y_t v_{t+i+2} & \sum_{t=1}^{T-i-2} y_t v_{t+i+1} & \cdots & \sum_{t=1}^{T-i-2} y_t v_{t+i+2-q} \\ \vdots & \vdots & & \vdots \\ \sum_{t=1}^{T-i-k} y_t v_{t+i+k} & \sum_{t=1}^{T-i-k} y_t v_{t+i+k-1} & \cdots & \sum_{t=1}^{T-i-k} y_t v_{t+i+k-q} \end{pmatrix}\theta_*.$$

Denote it as

$$\hat{\sigma}(k,i) = \tilde{\Sigma}(k,i)\phi - \frac{1}{T}\tilde{\Delta}(k,i)\theta_*.$$

It follows that

$$\begin{aligned} & \sqrt{T}\{\hat{\phi}(k,i) - \phi(k,i)\} \\ = \; & \sqrt{T}\{\hat{\Sigma}(k,i)^{-1}\hat{\sigma}(k,i) - \phi(k,i)\} \\ = \; & \sqrt{T}\hat{\Sigma}(k,i)^{-1}\{\tilde{\Sigma}(k,i)\phi - \hat{\Sigma}(k,i)\phi(k,i)\} - \frac{1}{\sqrt{T}}\hat{\Sigma}(k,i)^{-1}\tilde{\Delta}(k,i)\theta_*. \end{aligned}$$

Clearly, the first term of the RHS converges to $\mathbf{0}$ in probability and $\hat{\Sigma}(k,i)$ converges to $\Sigma(k,i)$ in probability by Lemma 1. Thus, it is left to derive the asymptotic distribution of $-\tilde{\Delta}(k,i)\theta_*/\sqrt{T}$.

The stationarity implies that the ARMA process can be represented by an $MA(\infty)$ model

$$y_t = \sum_{j=0}^{\infty} \psi_j v_{t-j} \ , \ \psi_0 = 1.$$

Let

$$y_t^{(n)} = \sum_{j=0}^{n} \psi_j v_{t-j},$$

$$\mathbf{v}_t = (v_{t+i+1-q}, \cdots, v_{t+i+k})^t,$$
$$\mathbf{u}_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \mathbf{v}_t,$$
$$\mathbf{u}_T^{(n)} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t^{(n)} \mathbf{v}_t.$$

For a fixed $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_{k+q})^t$, it can be shown

$$y_t^{(n)} \alpha^t \mathbf{v}_t = (\sum_{r=0}^{n} \psi_r v_{t-r})(\sum_{s=1}^{k+q} \alpha_s v_{t+i-q+s}).$$

Thus, two sequences $\{y_{t-r}^{(n)} \alpha^t \mathbf{v}_{t-r} \mid r = 0, 1, \cdots\}$ and $\{y_{t+s}^{(n)} \alpha^t \mathbf{v}_{t+s} \mid s = i+k+n+1, i+k+n+2, \cdots\}$ are independent. It means that $\{y_t^{(n)} \alpha^t \mathbf{v}_t\}$ is strictly stationary $(i+k+n)$-dependent. The central limit theorem for dependent random variables due to Hoeffding and Robbins (1948) implies that there exists a random vector $\mathbf{u}^{(n)}$ satisfying

$$\alpha^t \mathbf{u}_T^{(n)} \xrightarrow{d} \alpha^t \mathbf{u}^{(n)} \quad \text{as } T \to \infty,$$

where '$\xrightarrow{d}$' means 'convergence in distribution' and $\alpha^t \mathbf{u}^{(n)}$ is a normal random variable with mean 0. Since $y_t^{(n)}$ converges to y_t in L^2 as $n \to \infty$, there exists a random vector $\mathbf{u}$ satisfying

$$\alpha^t \mathbf{u}^{(n)} \xrightarrow{d} \alpha^t \mathbf{u} \quad \text{as } n \to \infty,$$

where $\alpha^t \mathbf{u}$ is a normal random variable with mean 0. Chebyshev's inequality implies

$$\begin{aligned}
& Pr\left(\left|\alpha^t \mathbf{u}_T - \alpha^t \mathbf{u}_T^{(n)}\right| > \epsilon\right) \\
\leq\ & \frac{1}{\epsilon^2} Var\left(\alpha^t \mathbf{u}_T - \alpha^t \mathbf{u}_T^{(n)}\right) \\
=\ & \frac{1}{\epsilon^2 T} Var\left\{\sum_{t=1}^{T} \left(y_t^{(n)} - y_t\right) \alpha^t \mathbf{v}_t\right\} \\
=\ & \frac{1}{\epsilon^2 T} Cov\left\{\sum_{t=1}^{T} \left(y_t^{(n)} - y_t\right) \alpha^t \mathbf{v}_t, \sum_{s=1}^{T} \left(y_s^{(n)} - y_s\right) \alpha^t \mathbf{v}_s\right\} \\
=\ & \frac{1}{\epsilon^2 T} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left\{\left(y_t^{(n)} - y_t\right)\left(y_s^{(n)} - y_s\right) \alpha^t \mathbf{v}_t \mathbf{v}_s^t \alpha\right\}.
\end{aligned}$$

Therefore, for large n, the following holds;

$$\begin{aligned} & Pr\left(\left|\alpha^t \mathbf{u}_T - \alpha^t \mathbf{u}_T^{(n)}\right| > \epsilon\right) \\ = & \frac{1}{\epsilon^2 T} \sum_{|t-s| \le k+q-1} E\left\{\left(y_t^{(n)} - y_t\right)\left(y_s^{(n)} - y_s\right)\alpha^t \mathbf{v}_t \mathbf{v}_s^t \alpha\right\} \\ \le & \frac{1}{\epsilon^2 T}(2k+2q-1)T \; Var\left\{\left(y_t^{(n)} - y_t\right)\alpha^t \mathbf{v}_t\right\} \\ = & \frac{2k+2q-1}{\epsilon^2} Var\left\{\left(y_t^{(n)} - y_t\right)\alpha^t \mathbf{v}_t\right\}. \end{aligned}$$

Since $y_t^{(n)}$ converges to y_t in L^2,

$$\lim_{n\to\infty} \limsup_{T\to\infty} \; Pr\left(\left|\alpha^t \mathbf{u}_T - \alpha^t \mathbf{u}_T^{(n)}\right| > \epsilon\right) = 0,$$

for any $\epsilon(> 0)$. Therefore, Theorem 4.2 of Billingsley (1968) implies that

$$\alpha^t \mathbf{u}_T \xrightarrow{d} \alpha^t \mathbf{u} \quad \text{as } T \to \infty. \tag{2.4}$$

Since the convergence in Equation 2.4 holds for any α, the Cramér-Wold device implies (see, *e.g.*, Billingsley [1968, p. 49]) that

$$\mathbf{u}_T \xrightarrow{d} \mathbf{u} \quad \text{as } T \to \infty.$$

Let

$$u_T(j) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t v_{t+j}, \; j = 1, 2, \cdots,$$

and

$$w_T(m) = \sum_{j=0}^{q} \theta_j u_T(i+m-j), \quad m = 1, 2, \cdots, k.$$

As $T \to \infty$, $\mathbf{w}_T$ converges to a normal random vector with mean vector $\mathbf{0}$.

Its asymptotic variance-covariance matrix can be obtained as follows. For $r, s = 1, 2, \cdots$,

$$\begin{aligned} & \lim_{T\to\infty} Cov(u_T(r), u_T(s)) \\ = & \lim_{T\to\infty} Cov\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t v_{t+r}, \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t v_{t+s}\right) \\ = & \sigma^2 \sigma(r-s). \end{aligned}$$

Thus, for $r, s = 1, 2, \cdots, k$,

$$\lim_{T\to\infty} Cov(w_T(r), w_T(s)) = \sigma^2\theta_*^t\Sigma(q+1, r-s)\theta_*.$$

Since $-\tilde{\Delta}(k,i)/\sqrt{T}$ has the same asymptotic distribution as $-\mathbf{w}_T$, the theorem holds. Q.E.D.

If $k = p$ and $q = 0$, then the Ξ becomes $\Sigma(p,0)$, and then, the variance-covariance matrix becomes $\sigma^2\Sigma(p,0)^{-1}$. Thus, Theorem 1 is a generalization of Mann and Wald theorem.

3. CONCLUSION

The asymptotic distribution of the EYW estimates of a mixed ARMA process is derived, which is a generalization of Mann and Wald's distribution.

APPENDIX

When fitting an ARMA model to a realization of time series, the orders are usually unknown. Thus, it is necessary to calculate the EYW estimates for several pairs of orders. If $q = 0$, then they can be calculated recursively through the Levinson-Durbin algorithm. Choi (1992, p. 8) has presented the following algorithm to calculate the EYW estimates for a mixed ARMA model, which is clearly a generalization of the Levinson-Durbin algorithm. However, its derivation has not been published yet. In this appendix its derivation is presented.

An Algorithm for calculating the EYW Estimates

1. *Initial step.* Obtain $\hat{\phi}_{k,j}^{(0)}$ for $k = 1, 2, \cdots$ and $j = 1, \cdots, k$ by the Levinson-Durbin algorithm.

2. *Recursive step.* For $i = 1, 2, \cdots$ and $k = 1, 2, \cdots$, use the following recursion;

$$\hat{\phi}_{k,0}^{(i)} = -1,$$

$$\hat{\phi}_{k,j}^{(i)} = \hat{\phi}_{k+1,j}^{(i-1)} - \frac{\hat{\phi}_{k+1,k+1}^{(i-1)}}{\hat{\phi}_{k,k}^{(i-1)}}\hat{\phi}_{k,j-1}^{(i-1)}, \quad (j = 1, \cdots, k).$$

Derivation. Let

$$\hat{\phi}(k+1, i-1) = \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix}.$$

Using the bordering technique in matrix inversion (see, *e.g.*, Faddeeva [1959, pp. 107-111]), one can show

$$y = \hat{\phi}^{(i-1)}_{k+1,k+1}$$
$$\mathbf{x} = \hat{\phi}(k, i) - \hat{\Sigma}(k, i)^{-1}\hat{\mathbf{s}}(k, i)\hat{\phi}^{(i-1)}_{k+1,k+1},$$

where $\hat{\mathbf{s}}(k, i) = (\hat{\sigma}(i-k), \cdots, \hat{\sigma}(i-2), \hat{\sigma}(i-1))^t$. Let $\mathbf{z} = \hat{\Sigma}(k, i)^{-1}\hat{\mathbf{s}}(k, i)$. Then, the following two equations are equivalent;

$$\hat{\Sigma}(k, i-1)\hat{\phi}(k, i-1) = \hat{\sigma}(k, i-1),$$
$$\hat{\Sigma}(k, i)\mathbf{z} = \hat{\mathbf{s}}(k, i).$$

Thus, the j-th element of $\mathbf{z}$ is

$$z_j = -\hat{\phi}^{(i-1)}_{k,j-1}/\hat{\phi}^{(i-1)}_{k,k}, \quad j = 1, 2, \cdots, k.$$

It completes the derivation. □

Since the process $\{y_t\}$ is not purely deterministic, the estimate $\hat{\Sigma}(k, i)$ is nonsingular with probability 1 for any fixed T. Thus, one can define $\hat{\phi}(k, i)$ for $k = 1, 2, \cdots$ and $i = 0, 1, \cdots$, and then, the algorithm is practically applicable.

ACKNOWLEDGEMENT

The authors thank the anonymous referees and the associate editor for their helpful suggestions and comments.

REFERENCES

(1) Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

(2) Choi, B. S. (1992). *ARMA Model Identification*, Springer-Verlag, New York.

(3) Faddeeva, V. N. (1959). *Computational Methods of Linear Algebra*, Dover, New York.

(4) Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables, *Duke Mathematical Journal*, **15**, 773-780.

(5) Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations, *Econometrica*, **11**, 173-220.