Ann. Inst. Statist. Math.
37 (1985), Part A, 389-408

ESTIMATION OF THE DEGREE OF DIFFERENCING OF AN ARIMA PROCESS

YOSHIHIRO YAJIMA

(Received May 4, 1984; revised Mar. 6, 1985)

Summary

A method of estimating the degree of differencing of an ARIMA process is proposed. This is based on fitting an AR model to the original and to each differenced series and calculating the residual sum of squares. As an application, we suggest an identification method of an ARI (p, d) process combining our method of estimating the degree of differencing with Akaike's Information Criterion.

1. Introduction

A stochastic process w_t is called an ARIMA (p, d, q) process if it satisfies

$$\phi(B)\nabla^d w_t = \theta(B)e_t ,$$

where B is the backward shift operator such that $Bw_t = w_{t-1}$ and

$$\phi(B)=1-\phi_1-\cdots-\phi_p B^p , \qquad \theta(B)=1-\theta_1-\cdots-\theta_q B^q ,$$

∇ is the difference operator, $\nabla w_t=(1-B)w_t$ with $\nabla^d w_t=\nabla(\nabla^{d-1}w_t)$, and where e_t is a sequence of independent normally distributed random variables with means zero and variances σ_e^2. We assume that $\phi(B)$ and $\theta(B)$ have all the roots outside the unit circle and have no common roots.

The time series modeling strategy developed by Box and Jenkins [7] consists of model identification, model estimation, and diagnostic checks. In their procedure model identification, that is the determination of the order (p, d, q), is an important problem. The main purpose of this paper is to propose a method of estimating the degree of differencing d.

For this problem, there seems to have been only a small amount

Key words: ARIMA processes, AR model, differencing, order selection.

of literature published. Box and Jenkins suggested to estimate d by visual examination of the sample autocorrelations and their method depends on subjective judgement. Anderson ([3], [4], Ch. 12) discussed this problem by adopting $\hat{\delta}$ as an estimator of the degree of differencing, when the sample variance of δth differenced series attains its minimum at $\hat{\delta}=\delta$. These methods are useful in some cases, however there seems to have been hardly anything known about the statistical properties of these methods because of the nonstationarity of ARIMA processes. On the other hand, though Anderson's method determines the degree of differencing automatically, it misleads to overdifferencing in some cases as is remarked in Anderson [3], [4]. Overdifferencing of the ARIMA process implies a transformation of the series to a non-invertible ARMA process, which causes some difficulties in estimation and prediction.

Thus we shall propose a new method to avoid overdifferencing. Hereafter we assume that the possible degrees of differencing are bounded by some known number D. Let $\{w_1, w_2, \cdots, w_T\}$ be observations from an ARIMA(p, d, q) process. Let $\hat{\sigma}_k^2(\delta)$ be the residual sum of squares obtained by fitting an AR(k) model to $\nabla^\delta w_t$. That is, it satisfies

$$\hat{\sigma}_k^2(\delta)=\min_{\beta_i} \sum_{t=k+\delta+1}^{T} \left(\nabla^\delta w_t-\sum_{i=1}^{k} \beta_i \nabla^\delta w_{t-i}\right)^2 \Big/ (T-k-\delta) .$$

Next let $\sigma_k^2(\delta)$ be the mean square error of prediction of $\nabla^\delta w_{k+1}$ given $\nabla^\delta w_k, \cdots, \nabla^\delta w_1$. Then it can be expected that as $k\to\infty$, $\sigma_k^2(\delta)$ converges to σ_e^2 most rapidly at $\delta=d$. $\hat{\sigma}_k^2(\delta)$ can be viewed as an estimator of $\sigma_k^2(\delta)$. Thus we shall construct an estimator of d based on $\hat{\sigma}_k^2(\delta)$.

Let $\hat{\delta}$ be the value of δ which minimizes

$$\hat{\sigma}_k^2(\delta)-f(\delta)\hat{\sigma}_l^2(0)/g(k) ,$$

$0\leqq\delta\leqq D$, where $f(\delta)$ is an arbitrary strictly increasing function of δ and $g(k)$ an arbitrary function satisfying

$$\lim_{k\to\infty} k/g(k)=0 , \quad \text{and} \quad g(k)>0 .$$

In this paper, we shall show that $\hat{\delta}$ is a consistent estimator for d under some conditions on the parameters if k and l are sufficiently large and fixed. $f(\delta)$ is introduced to avoid underdifferencing since the relation,

$$p-\lim_{T\to\infty} \hat{\sigma}_k^2(\delta)=p-\lim_{T\to\infty} \hat{\sigma}_k^2(d)=\sigma_e^2 ,$$

can occur for finite k if $q=0$ and $\delta<d$, as is shown later. However if $f(\delta)$ is too large, our method may mislead to overdifferencing. Hence $\hat{\sigma}_l^2(0)/g(k)$ is added since

$$p-\lim_{T\to\infty}\hat{\sigma}_k^2(\delta)-\sigma_e^2=\sigma_e^2 O(1/k)\,, \qquad \delta>d\,,$$

which will be shown later, and $\hat{\sigma}_l^2(0)$ can be an estimator of σ_e^2 if l is sufficiently large.

In Section 2, the asymptotic behavior of our estimator is discussed. This is based on the asymptotic distribution of the sample covariance matrix of an ARIMA process given in the Appendix, which is a generalization of the results obtained by Hasza and Fuller [10].

After d is estimated by $\hat{\delta}$, we have to determine p and q. One method is to apply the order determination procedures proposed by Akaike [1], [2] or Hannan [9] to $\nabla^{\hat{\delta}}w_t$, since $\nabla^{\hat{\delta}}w_t$ could be assumed as ARMA (p, q) if the estimator $\hat{\delta}$ is consistent.

In Section 3, we propose another method by combining $\hat{\delta}$ with AIC (Akaike's Information Criterion) when w_t is an ARI (p, d) process. We regard w_t as a nonstationary AR process with the true order $k_0=p+d$, and let $\hat{k}$ be the order selected by applying AIC to the original process, w_t. Then we estimate p by $\hat{p}=\hat{k}-\hat{\delta}$. It is shown that the asymptotic distribution of the selected order of a stationary AR process given by Shibata [12] still holds for $\hat{k}$. Hence $\hat{p}$ is a reasonable estimator of p if $\hat{\delta}$ is consistent.

Finally some numerical results are given in Section 4 to illustrate the performance of the proposed estimation procedure.

Our method intends to determine the degree of differencing d before considering p and q. Unlike us, Ozaki [11] applies AIC effectively to estimate (p, d, q) simultaneously.

2. Asymptotic behavior

Let $\{e_t\}_{t=-\infty}^{\infty}$ be a sequence of independent normally distributed random variables with means zero and variance σ_e^2. A process w_t, defined for $t\geqq -d+1$, is said to be ARIMA if $\nabla^d w_t$, $t\geqq 1$, satisfies

$$\phi(B)\nabla^d w_t=\theta(B)e_t\,,$$

given the initial random variables w_t with $\mathrm{E}\,w_t^2<\infty$, $t=-d+1,\cdots,1,0$. This indicates that the process $x_t=\nabla^d w_t$ is a stationary ARMA (p, q) process.

We shall show the asymptotic properties of $\hat{\delta}$ given in the preceding section as an estimator of the true degree of differencing d. For this purpose, we have to investigate $\hat{\sigma}_k^2(\delta)$. Let $\mathcal{D}(x, 1, l)$ be the linear space spanned by $\{x_t;\, 1\leqq t\leqq l\}$ and $P_{\mathcal{D}(x,1,l)}$ the projection into $\mathcal{D}(x, 1, l)$. Let W_T be the $d\times d$ matrix whose (i, j) element is

$$\sum_{t=1}^{T-k}\nabla^{d-i}w_{t+d-i}\nabla^{d-j}w_{t+d-j}\,,$$

and $D_T=\operatorname{diag}(T, T^2, \cdots, T^d)$ be the diagonal matrix with diagonal elements, $T, T^2, \cdots, T^d$. Further let W be the $d\times d$ random matrix to which $D_T^{-1}W_TD_T^{-1}$ converges in distribution as $T\to\infty$. W will be given explicitly in Appendix Finally, put

$$\sigma_l^2=\mathrm{E}\{x_{l+1}-P_{\mathcal{D}(x,1,l)}x_{l+1}\}^2 .$$

Then we have the following theorem about the asymptotic behavior of $\hat{\sigma}_k^2(\delta)$ when $\delta\leqq d$.

THEOREM 2.1. *Assume that* $\Pr\{\det W\neq 0\}=1$.
(i) *If* $\delta\leqq d$, $k\geqq d-\delta$, *then* $p-\lim_{T\to\infty}\hat{\sigma}_k^2(\delta)=\sigma_{k-(d-\delta)}^2$.
(ii) *If* $\delta<d$, *and* $k<d-\delta$, *then* $\hat{\sigma}_k^2(\delta)/T^{2(d-\delta-k)-1}$ *has an asymptotic distribution* $F_k(x)$ *with* $F_k(0)=0$ *as* $T\to\infty$.

PROOF. (i) First we put $y_t=\nabla^\delta w_t$. Then y_t is an ARIMA (p, d^*, q) process with $d^*=d-\delta$. Let

$$D_{d^*,k}=(\boldsymbol{c}_{d^*}, L\boldsymbol{c}_{d^*}, \cdots, L^{k-d^*}\boldsymbol{c}_{d^*}, L^{k-d^*+1}\boldsymbol{c}_{d^*-1}, \cdots, L^{k-1}\boldsymbol{c}_1\, L^k\boldsymbol{c}_0)' ,$$

where L is the $(k+1)\times(k+1)$ matrix whose (i, j)th element is 1 for $i-j=1$ and is 0 for $i-j\neq 1$ and $\boldsymbol{c}_s$ is the $(k+1)$ component vector such that

$$\boldsymbol{c}_s=(c_{s,0}, c_{s,1}, \cdots, c_{s,s}, 0, 0, \cdots, 0)' ,$$

and $c_{s,i}$, $0\leqq i\leqq s$, $0\leqq s\leqq d$, are defined by $(1-B)^s=\sum_{i=0}^{s} c_{s,i}B^i$. Since $\{y_t, y_{t-1}, \cdots, y_{t-k}\}'$ is transformed to

$$\{x_t, \cdots, x_{t-k+d^*}, \nabla^{d^*-1}y_{t-k+d^*-1}, \cdots, \nabla y_{t-k+1}, y_{t-k}\}' ,$$

by $D_{d^*,k}$, we have

$$\begin{aligned}\hat{\sigma}_k^2(\delta)&=\min_{\beta_i}\sum_{t=k+\delta+1}^{T}\left\{y_t-\sum_{i=1}^{k}\beta_i y_{t-i}\right\}^2\Big/\{T-(k+\delta)\}\\ &=\min_{\alpha_i}\sum_{t=k+\delta+1}^{T}\left\{x_t-\sum_{i=1}^{k-d^*}\alpha_i x_{t-i}-\sum_{i=k-d^*+1}^{k}\alpha_i\nabla^{k-i}y_{t-i}\right\}^2\Big/\{T-(k+\delta)\} .\end{aligned}$$

Now we put $\tilde{\boldsymbol{x}}_T=(x_{k+\delta+1}, x_{k+\delta+2}, \cdots, x_T)'$ and $Z_T=[X_T, Y_T]$ where X_T is the $\{T-(k+\delta)\}\times(k-d^*)$ matrix and Y_T is the $\{T-(k+\delta)\}\times d^*$ matrix such that the (i, j)th element of X_T and Y_T are $x_{k+\delta+i-j}$ and $\nabla^{d^*-j}y_{\delta+d^*+i-j}$ respectively. Then

$$\begin{aligned}(1)\qquad \hat{\sigma}_k^2(\delta)&=\{\tilde{\boldsymbol{x}}_T'\tilde{\boldsymbol{x}}_T-\tilde{\boldsymbol{x}}_T'Z_T(Z_T'Z_T)^{-1}Z_T'\tilde{\boldsymbol{x}}_T\}/\{T-(k+\delta)\}\\ &=\{\tilde{\boldsymbol{x}}_T'\tilde{\boldsymbol{x}}_T-\tilde{\boldsymbol{x}}_T'X_T(X_T'X_T)^{-1}X_T'\tilde{\boldsymbol{x}}_T\}/\{T-(k+\delta)\}\\ &\quad-\{\tilde{\boldsymbol{x}}_T'X_TP_{1T}P_{2T}P_{1T}'X_T'\tilde{\boldsymbol{x}}_T-2\tilde{\boldsymbol{x}}_T'Y_TP_{2T}P_{1T}'X_T'\tilde{\boldsymbol{x}}_T\\ &\quad+\tilde{\boldsymbol{x}}_T'Y_TP_{2T}Y_T'\tilde{\boldsymbol{x}}_T\}/\{T-(k+\delta)\} ,\end{aligned}$$

where

$$P_{1T}=(X_T'X_T)^{-1}X_T'Y_T \ ,$$

and

$$P_{2T}=\{Y_T'Y_T-Y_T'X_T(X_T'X_T)^{-1}X_T'Y_T\}^{-1} \ .$$

In (1), it is well known that the first term of the right hand side converges to $\sigma^2_{k-d^*}$ in probability as $T\to\infty$. Thus we have only to show that the second term converges to zero in probability as $T\to\infty$. Since this result can be proved for each portion of the second term, we illustrate the result for the portion, $\tilde{\boldsymbol{x}}_T'Y_TP_{2T}Y_T'\tilde{\boldsymbol{x}}_T/\{T-(k+\delta)\}$. This can be rewritten as

$$\tilde{\boldsymbol{x}}_T'Y_TD_T^{-1}(D_T^{-1}P_{2T}^{-1}D_T^{-1})^{-1}D_T^{-1}Y_T'\tilde{\boldsymbol{x}}_T/\{T-(k+\delta)\} \ .$$

It is shown by evaluating higher order moments that

$$\tilde{\boldsymbol{x}}_T'Y_TD_T^{-1}=O_p(1) \ .$$

Since $p-\lim_{T\to\infty} X_T'X_T/T=R$, where R is the nonsingular matrix with (i, j)th element equal to $\mathrm{E}\, x_ix_j$, it is similarly shown that in $D_T^{-1}P_{2T}^{-1}D_T^{-1}$,

$$D_T^{-1}Y_T'X_T(X_T'X_T)^{-1}X_T'Y_TD_T^{-1}=O_p(1/T) \ .$$

In $D_T^{-1}P_{2T}^{-1}D_T^{-1}$, the limiting distribution of $D_T^{-1}Y_T'Y_TD_T^{-1}$ is that of the upper-left $d^*\times d^*$ matrix of W in Proposition A.1. Then the result is immediately obtained by assumption and Corollary 1 of Billingsley ([6], p. 31).

(ii) In a similar way, we find

$$\begin{aligned}\hat{\sigma}_k^2(\delta)&=\min_{\alpha_i}\sum_{t=k+\delta+1}^{T}\left\{\nabla^k y_t-\sum_{i=1}^{k}\alpha_i\nabla^{k-i}y_{t-i}\right\}^2\Big/\{T-(k+\delta)\}\\&=\{\tilde{\boldsymbol{y}}_T'\tilde{\boldsymbol{y}}_T-\tilde{\boldsymbol{y}}_T'Y_T(Y_T'Y_T)^{-1}Y_T'\tilde{\boldsymbol{y}}_T\}/\{T-(k+\delta)\} \ ,\end{aligned}$$

where

$$\tilde{\boldsymbol{y}}_T=\{\nabla^k y_{k+\delta+1}, \nabla^k y_{k+\delta+2}, \nabla^k y_T\}' \ ,$$

and Y_T is the $\{T-(k+\delta)\}\times k$ matrix whose (i, j)th element is $\nabla^{k-j}\cdot y_{k+\delta+i-j}$. Then it follows from Proposition A.1 that the limiting distribution of $\hat{\sigma}_k^2(\delta)/T^{2(d-\delta-k)-1}$ is the distribution of

$$w_{d^*-k,d^*-k}-\boldsymbol{w}_{d^*,k}'\tilde{W}_{d^*-k,k}^{-1}\boldsymbol{w}_{d^*,k} \ ,$$

where

$$\boldsymbol{w}_{d^*-k}=(w_{d^*-k+1,d^*-k}, w_{d^*-k+2,d^*-k}, \cdots, w_{d^*,d^*-k})' \ ,$$

and w_{ij} is the (i, j)th element of W, and $\tilde{W}_{s,l}$ is the $l\times l$ matrix whose

(i, j)th element is $w_{s+i,s+j}$. It follows from the assumption that $\tilde{W}_{d^*-1-k,k+1}$ is positive definite with probability one. Then noting that

$$\det \tilde{W}_{d^*-1-k,k+1}=(w_{d^*-k,d^*-k}-\boldsymbol{w}'_{d^*,k}\tilde{W}^{-1}_{d^*-k,k}\boldsymbol{w}_{d^*,k}) \det \tilde{W}_{d^*-k,k} \ ,$$

we have the result.

Remark 2.1. If d is given explicitly we can check whether or not the assumption $\Pr\{\det W\neq 0\}=1$ in Theorem 2.1 holds by the same manner as that for Theorem 3.2 of Hasza and Fuller [10]. For example we can show for $d\leqq 5$ that the assumption holds.

Next we consider the case $\delta\geqq d$. Let us put

$$\sigma^2_{k,l}=\mathrm{E}\,\{\nabla^l x_{k+1}-P\mathscr{D}(\Delta x, 1, k)\nabla^l x_{k+1}\}^2 \ .$$

THEOREM 2.2. (i) *If* $\delta=d$, *then* $p-\lim_{T\to\infty}\hat{\sigma}^2_k(\delta)=\sigma^2_k=\sigma^2_e\{1+O(\max_i|h_i|^{2k})\}$ *where* $\{h_i\}$ *are the roots of* $\theta(z^{-1})=0$.
(ii) *If* $\delta>d$, *then* $p-\lim_{T\to\infty}\hat{\sigma}^2_k(\delta)=\sigma^2_{k,\delta-d}=\sigma^2_e\{1+(\delta-d)^2c^*/k+o(1/k)\}$ *where* c^* *is a positive constant depending on* $(\phi_1,\cdots,\phi_p,\theta_1,\cdots,\theta_q)$.

PROOF. Except for the assertion that $c^*>0$, all of the results have already been proved by Grenander and Rosenblatt [8] and Yajima [13]. Now we shall show that $c^*>0$. Let us denote

$$\sum_{i=0}^{\infty}\pi_i z^i=\theta(z)^{-1}\phi(z) \ .$$

And let

$$\eta_\nu(z)=\sum_{i=0}^{\nu}a_{i,\nu}z^{\nu-i}/\sigma_\nu \ , \qquad \nu\geqq 0 \ ,$$

be the orthonormal polynomials of order ν obtained by the Gram-Schmidt procedure from $1, z, z^2,\cdots$, whose inner product is defined by

$$(g, h)=\{\sigma^2_e/(2\pi)\}\int_{-\pi}^{\pi} g(e^{i\lambda})\bar{h}(e^{i\lambda})|\theta(e^{i\lambda})/\phi(e^{i\lambda})|^2 d\lambda \ .$$

Then it follows from Proposition 1 of Yajima [13] that

$$(2) \qquad c^*=\left\{\sigma_e\Big/\left(\sum_{i=0}^{\infty}\pi_i\right)\right\}^2\lim_{k\to\infty}s^2_{12}(k+\delta-d)/s_{11}(k+\delta-d) \ ,$$

where

$$s_{11}(l)=\sum_{\nu=0}^{l}\eta^2_\nu(0)=\sigma_l^{-2} \ ,$$

which converges to σ_e^{-2} as $l\to\infty$, and

$$s_{12}(l)=\sum_{\nu=0}^{l}\eta_\nu(0)\eta_\nu(1) .$$

Therefore c^* is independent of $(\delta-d)$, and we can put $\delta-d=1$. First we show $c^*>0$ for $q=0$. Since $a_{\nu,\nu}=0$ for $\nu>p$,

$$\lim_{k\to\infty} s_{12}^2(k)=s_{12}^2(p)=\left\{\sum_{\nu=0}^{p} a_{\nu,\nu}\left(\sum_{i=0}^{\nu} a_{i,\nu}\right)\Big/\sigma_\nu^2\right\}^2 .$$

The relation (2) assures that it suffices to show $s_{12}(p)\neq 0$. Let $s_{12}(p)=0$. Since

$$\sigma_{k,1}^2=s_{22}(k+1)/\{s_{11}(k+1)s_{22}(k+1)-s_{12}^2(k+1)\} ,$$

with $s_{22}(l)=\sum_{\nu=0}^{l}\eta_\nu^2(1)$ (Theorem 1 of Grenander and Rosenblatt [8]), we have

$$\sigma_{p-1,1}^2=s_{11}^{-1}(p)=\sigma_p^2=\sigma_e^2 .$$

This means that ∇x_t is an AR $(p-1)$ process. But this contradicts the fact that ∇x_t is an ARMA $(p, 1)$ process. Hence $s_{12}(p)\neq 0$. Next we discuss the case $q\neq 0$. Let us introduce the random variables v_t such that $\phi(B)v_t=e_t$. Since ∇x_t is expressed as $\nabla x_t=\theta(B)\nabla v_t$, we get

$$\begin{aligned}\mathrm{E}\{\nabla x_{k+1}-P_{\mathcal{D}(\nabla x,1,k)}\nabla x_{k+1}\}^2 &\geqq \mathrm{E}\{\nabla v_{k+1}-P_{\mathcal{D}(\nabla v,-q+1,k)}\nabla v_{k+1}\}^2\\ &=\sigma_e^2\{1+c_v/k+o(1/k)\} ,\end{aligned}$$

in which $c_v>0$ since v_t is an AR (p) process. Thus the proof is complete.

Remark 2.2. Grenander and Rosenblatt [8] proved in Theorem 5 that $c^*>0$ if $p=q=0$ and $\delta-d=1$. Theorem 2.2 implies that $c^*>0$ still holds for any p, q, and $\delta-d>0$.

Now we can clarify the asymptotic properties of our method. Since σ_k^2 is a nonincreasing function of k, Theorem 2.1 asserts that

$$\lim_{T\to\infty}\Pr\{\hat\delta=\delta\}=0 ,$$

for any k if $\delta<d$. On the other hand, Theorem 2.2 implies that σ_k^2 converges to σ_e^2 exponentially as $k\to\infty$ and $\sigma_{k,\delta-d}^2$ dominates $f(\delta)\sigma_{l-d}^2/g(k)$ for sufficiently large k. Hence

$$\lim_{T\to\infty}\Pr\{\hat\sigma_k^2(\delta)-f(\delta)\hat\sigma_l^2(0)/g(k)>\hat\sigma_k^2(d)-f(d)\hat\sigma_l^2(0)/g(k)\}=0 .$$

holds for sufficiently large k. However the rates of convergence of σ_k^2 and $\sigma_{k,n}^2$ depend on unknown parameters and are not uniform. Hence the number k satisfying the above relation also depends on unknown parameters. Accordingly in order to use this method in practice, we have to assume the following conditions:

(i) p and q are bounded by known quantities P and Q respectively.
(ii) Let $\{g_i\}$ and $\{h_i\}$ be the roots of $\phi(z^{-1})=0$ and $\theta(z^{-1})=0$ respectively. Then

$$|g_i|\leqq 1-\varepsilon\,, \qquad 1\leqq i\leqq p\,,$$
$$|h_i|\leqq 1-\varepsilon\,, \qquad 1\leqq i\leqq q\,,$$

where ε is arbitrarily given small quantity.

Similar conditions are imposed when we estimate the order of a stationary process (Hannan [9]). Let us define A ($\subset R^{P+Q}$) by

$$A=\{(\phi_1,\cdots,\phi_P,\theta_1,\cdots,\theta_Q)\,|\,|g_i|\leqq 1-\varepsilon,\ 1\leqq i\leqq P,\ |h_i|\leqq 1-\varepsilon,\ 1\leqq i\leqq Q\}\ .$$

Then we have the following result for $\delta>d$ and $l\geqq d$.

PROPOSITION 2.1. *There exist a positive number K such that*

$$\sigma^2_{k,\delta-d}-f(\delta)\sigma^2_{l-d}/g(k)>\sigma^2_k-f(d)\sigma^2_{l-d}/g(k)\ ,$$

holds for any $k\geqq K$ and any $(\phi_1\cdots,\phi_P,\theta_1,\cdots,\theta_Q)\in A$.

PROOF. First we consider the case of $\theta_1=\cdots=\theta_Q=0$. Put

$$\tilde{A}=\{(\phi_1,\cdots,\phi_P)\,|\,|g_i|\leqq 1-\varepsilon,\ 1\leqq i\leqq P\}\ (\subset R^P)\ .$$

c^* is a continuous function of $(\phi_1,\cdots,\phi_P)$ and its minimum in $\tilde{A}$ is bounded away from zero since $\tilde{A}$ is a compact set. Now we put

$$\varepsilon_{k,\phi,\delta-d}=k[\sigma^2_{k,\delta-d}-\sigma^2_e\{1+c^*(\delta-d)^2/k\}]\ .$$

By using the method in Theorem 1 of Grenander and Rosenblatt [8], we can show that $\varepsilon_{k,\phi,\delta-d}$ converges to zero uniformly in $(\phi_1,\cdots,\phi_P)\in\tilde{A}$ and $\delta\leqq D$. Let $\tilde{c}$ be the minimum of c^* in $\tilde{A}$. Then since $\sigma^2_k=\sigma^2_e$ for $k\geqq P$, we have

$$\begin{aligned}\{\sigma^2_{k,\delta-d}-f(\delta)\sigma^2_{l-d}/g(k)\}-\{\sigma^2_k-f(d)\sigma^2_{l-d}/g(k)\}\\ \geqq\sigma^2_e\{\tilde{c}(\delta-d)^2/k+\varepsilon_{k,\phi,\delta-d}/k\}-\sigma^2_{l-d}\{f(\delta)-f(d)\}/g(k)\ .\end{aligned}$$

Since $\lim_{k\to\infty}k/g(k)=0$, the right hand side is positive for sufficiently large k. Thus we have the result. Next consider the general case.

Let $\{v_t\}$ be a sequence of random variables such that $\phi(B)v_t=e_t$. Then as in Theorem 2.2, we have

$$\sigma^2_{k,\delta-d}\geqq\sigma^2_e\{1+c_v(\delta-d)^2/(k+Q)+\varepsilon_{k+Q,\phi,\delta-d}/(k+Q)\}\ .$$

We have just proved that the minimum of c_v in $\tilde{A}$ is bounded away from zero and

$$\lim_{k\to\infty} \varepsilon_{k,\phi,\delta-d}=0 ,$$

uniformly in $(\phi_1,\cdots,\phi_P)\in\tilde{A}$ and $\delta\leqq D$. On the other hand,

$$\sigma_k^2=\sigma_\varepsilon^2\{1+O((1-\varepsilon)^{2k})\} ,$$

uniformly in $(\phi_1,\cdots,\phi_P,\theta_1,\cdots,\theta_Q)\in A$. Then the result follows immediately.

Hence $\hat{\delta}$ is a consistent estimator of d under the preceding conditions if k and l are sufficiently large and fixed. However the performance of $\hat{\delta}$ depends on $f(\delta)$, $g(k)$, k, and l heavily. For practical use, we shall recommend the following policy to determine these amounts.
(i) In order to choose $f(\delta)$ and $g(k)$ more easily, we only investigate $f(\delta)$ and $g(k)$ of the form $f(\delta)=c\delta$ and $g(k)=k^\beta$ with $c>0$ and $\beta>1$. This restriction is reasonable since $\sigma_\varepsilon^2(\delta-d)^2c^*/k$ is of order δ^2/k.
(ii) We can put $\sigma_\varepsilon^2=1$ without loss of generality since

$$\sigma_{k-(\delta-d)}^2-\sigma_{l-d}^2 f(\delta)/g(k) ,$$

and

$$\sigma_{k,\delta-d}^2-\sigma_{l-d}^2 f(\delta)/g(k) ,$$

are proportional to σ_ε^2. When the difference between $\sigma_{k-j}^2-c(d-j)\sigma_{l-d}^2/k^\beta$ and $\sigma_k^2-cd\sigma_{l-d}^2/k^\beta$, and that between $\sigma_{k,n}^2-c(d+n)\sigma_{l-d}^2/k^\beta$ and $\sigma_k^2-cd\sigma_{l-d}^2/k^\beta$ are large simultaneously, $\hat{\delta}$ has good performance. On the other hand,

$$\lim_{k\to\infty}(\sigma_{k,n}^2-\sigma_{k-j}^2)=0 .$$

Hence first we choose the smallest value of k' as k, which satisfy

$$\sigma_{k'}^2<\sigma_{k',l}^2 ,$$

for any $(\phi_1,\cdots,\phi_P,\theta_1,\cdots,\theta_Q)\in A$. After k is fixed, we determine c, β, and l so that both

$$\sigma_{k-j}^2-(\sigma_k^2-cj\sigma_{l-D}^2/k^\beta) ,$$

and

$$(\sigma_{k,n}^2-cn\sigma_{l-D}^2/k^\beta)-\sigma_k^2 ,$$

are as large as possible. Since $\sigma_{l-D}^2\geqq\sigma_{l-d}^2$,

$$(\sigma_{k,n}^2-cn\sigma_{l-D}^2/k^\beta)-\sigma_k^2>0 ,$$

assures

$$(\sigma_{k,n}^2-cn\sigma_{l-d}^2/k^\beta)-\sigma_k^2>0 .$$

The choice of β and l is not so important since we can adjust $c\sigma_{l-D}^2/k^\beta$ by using c.

3. Identification of an ARI (p, d) process

Throughout this section, let w_t be an ARI (p, d) process. We propose a method of determining p and d by using $\hat{\delta}$ and AIC (Akaike [1], [2]). We adopt $\hat{\delta}$ as an estimator of d. Next we regard w_t as an AR process with the true order $k_0=p+d$, and, applying the AIC method, select the order for which

$$\mathrm{AIC}\,(k)=T\log\hat{\sigma}_e^2(k)+2k\ , \qquad (k=0, 1, \cdots, K)\ ,$$

attains its minimum as a function of k, where K is a preassigned order and

$$\hat{\sigma}_e^2(k)=\min_{\beta_i(k)}\sum_{t=K+1}^{T}\{w_t-\beta_1(k)w_{t-1}-\cdots-\beta_k(k)w_{t-k}\}^2/T\ .$$

Let $\mathrm{AIC}\,(\hat{k})=\min_k \mathrm{AIC}\,(k)$. Then we adopt $\hat{p}=\hat{k}-\hat{\delta}$ as an estimator of p. It should be noted here that $\log\hat{\sigma}_e^2(k)$ is an approximate likelihood. Thus, strictly speaking, the employed AIC is an approximate one. We have to investigate $\hat{k}$ to show the asymptotic properties of $\hat{p}$. Then it is shown that the asymptotic distribution of $\hat{k}$ for a stationary AR process given by Shibata [12] still holds in this case.

PROPOSITION 3.1. *Let the assumption of Theorem* 2.1 *be satisfied. Let* w_t *be an* ARI (p, d) *process and put*

$$p_n=\sum_n^{*}\left\{\prod_{i=1}^{n}(\alpha_i/i)^{r_i}/r_i!\right\}\ , \qquad \textit{and} \qquad q_n=\sum_n^{*}\left[\prod_{i=1}^{n}\{(1-\alpha_i)/i\}^{r_i}/r_i!\right]\ ,$$

where $\alpha_i=p_r\{\chi_i^2>2i\}$ *and* $\sum_n^{*}$ *extends over all n-tuples* $(r_1,\cdots,r_n)$ *of non-negative integers such that* $\sum_{i=1}^{n}r_i\times i=n$. *Then*

$$\lim_{T\to\infty}p_r\{\hat{k}=k\}=\begin{cases} p_{k-(p+d)}q_{K-k} & (p+d\leq k\leq K)\ ,\\ 0 & \textit{otherwise}\ .\end{cases}$$

PROOF. By the same argument as that for $\hat{\sigma}_k^2(\delta)$ in Theorem 2.1, we can show that

$$\hat{\sigma}_e^2(k)=O_p(T^{2(d-k)-1})\ , \qquad k<d\ ,$$

and

$$p-\lim_{T\to\infty}\hat{\sigma}_e^2(k)=\sigma_{k-d}^2\ , \qquad k\geq d\ .$$

Thus the assertion is easily shown if $k<p+d$. Next consider the case, $k\geqq p+d$. Define $\{\beta_i;\ 1\leqq i\leqq p+d\}$ by

$$1-\sum_{i=1}^{p+d}\beta_i z^i=(1-z)^d\phi(z)\ ,$$

and $\beta(k)$ by

$$\beta(k)=(\beta_1,\beta_2,\cdots,\beta_k)'\ ,$$

with $\beta_i=0$ for $i>p+d$. Let

$$\hat\beta(k)=\{\hat\beta_1(k),\hat\beta_2(k),\cdots,\hat\beta_k(k)\}'\ ,$$

and $\hat\beta_1(k)=\beta_1(k),\cdots,\hat\beta_k(k)=\beta_k(k)$ be the values that minimize

$$\sum_{t=K+1}^{T}\left\{w_t-\sum_{i=1}^{k}\beta_i(k)w_{t-i}\right\}^2\Big/T\ ,$$

with the minimum $\hat\sigma_e^2(k)$. Now let us put $x_t=\nabla^d w_t$ and

$$\Phi_0(k)=(\phi_1,\phi_2,\cdots,\phi_k)'\ ,$$

with $\phi_i=0$ for $i>p$ and let

$$\tilde\Phi(k)=\{\tilde\phi_1(k),\tilde\phi_2(k),\cdots,\tilde\phi_k(k)\}'\ ,$$

where $\tilde\phi_1(k)=\phi_1(k),\cdots,\tilde\phi_k(k)=\phi_k(k)$ are the values that minimize

$$\sum_{t=K+1}^{T}\left\{x_t-\sum_{i=1}^{k}\phi_i(k)x_{t-i}\right\}^2\Big/T\ ,$$

with the minimum $\tilde\sigma_e^2(k)$.

We shall show

$$\{\hat\sigma_e^2(k+1)/\hat\sigma_e^2(k)\}-\{\tilde\sigma_e^2(k+1-d)/\tilde\sigma_e^2(k-d)\}=o_p(1/T)\ ,\qquad k\geqq p+d\ .$$

Since x_t is a stationary AR(p) process, the assertion follows from Theorem 1 of Shibata [12]. Direct calculation shows that

$$\hat\sigma_e^2(k+1)/\hat\sigma_e^2(k)=1-\hat\beta_{k+1}^2(k+1)\{\hat s_e^2(k)/\hat\sigma_e^2(k)\}\ ,$$

$$\tilde\sigma_e^2(k+1-d)/\tilde\sigma_e^2(k-d)=1-\tilde\phi_{k+1-d}^2(k+1-d)\{\tilde s_e^2(k-d)/\tilde\sigma_e^2(k-d)\}\ ,$$

where

$$\hat s_e^2(l)=\min_{b_i(l)}\sum_{t=K}^{T-1}\left\{w_{t-l}-\sum_{i=1}^{l}b_i(l)w_{t-l+i}\right\}^2\Big/T\ ,$$

and

$$\tilde s_e^2(l)=\min_{b_i(l)}\sum_{t=K}^{T-1}\left\{x_{t-l}-\sum_{i=1}^{l}b_i(l)x_{t-l+i}\right\}^2\Big/T\ .$$

By the same argument as that for $\hat\sigma_e^2(l)$ and $\tilde\sigma_e^2(l)$, we can show that

$$p-\lim_{T\to\infty} \hat{s}_e^2(l)=\sigma_{l-d}^2 ,$$

and

$$p-\lim_{T\to\infty} \tilde{s}_e^2(l)=\sigma_l^2 .$$

Thus it suffices to show that

$$\hat{\beta}_{k+1}^2(k+1)=\tilde{\phi}_{k+1-d}^2(k+1-d)+o_p(1/T) .$$

Since

$$p-\lim_{T\to\infty} \hat{s}_e^2(k)/\hat{\sigma}_e^2(k)=p-\lim_{T\to\infty} \tilde{s}_e^2(k-d)/\tilde{\sigma}_e^2(k-d)=1 ,$$

noting the fact,

$$\tilde{\phi}_{k+1-d}^2(k+1-d)=O_p(1/T) ,$$

for $k\geqq p+d$, (Anderson [5], Sec. 5.6.3), we have

$$\begin{aligned}
&|\hat{\sigma}_e^2(k+1)/\hat{\sigma}_e^2(k)-\tilde{\sigma}_e^2(k+1-d)/\tilde{\sigma}_e^2(k-d)|\\
&\quad\leqq|\hat{\beta}_{k+1}^2(k+1)-\tilde{\phi}_{k+1-d}^2(k+1-d)|\{\hat{s}_e^2(k)/\hat{\sigma}_e^2(k)\}\\
&\qquad+\tilde{\phi}_{k+1-d}^2(k+1-d)|\hat{s}_e^2(k)/\hat{\sigma}_e^2(k)-\tilde{s}_e^2(k-d)/\tilde{\sigma}_e^2(k-d)|\\
&\quad=o_p(1/T) .
\end{aligned}$$

Let us put $Z_T=[X_T, Y_T]$ where X_T is the $(T-k)\times(k+1-d)$ matrix and Y_T is the $(T-k)\times d$ matrix whose (i, j) element are x_{K+i-j} and $\nabla^{d-j} w_{K-k+d-1+i-j}$ respectively. Define e_T by

$$e_T=(e_{K+1}, e_{K+2}, \cdots, e_T)' .$$

Then we have

$$\begin{aligned}
(3)\qquad &\hat{\beta}(k+1)-\beta(k+1)\\
&=D_{d,k}'\begin{pmatrix}(X_T'X_T)^{-1}X_T'e_T+P_{1T}P_{2T}P_{1T}'X_T'e_T-P_{1T}P_{2T}Y_T'e_T\\ -P_{2T}P_{1T}'X_T'e_T+P_{2T}Y_T'e_T\end{pmatrix}
\end{aligned}$$

$$(4)\qquad \tilde{\phi}(k+1-d)-\Phi_0(k+1-d)=(X_T'X_T)^{-1}X_T'e_T ,$$

where $D_{d,k}$ and P_{iT}, $i=1, 2$, are defined in the same manner as those in Theorem 2.1. All the elements except for $(X_T'X_T)^{-1}X_T'e_T$ in (3) can be shown to be at most of order $1/T$ in probability by the same argument as those in Theorem 2.1. Since $c_{d,d}=(-1)^d$, (3) and (4) imply

$$\hat{\beta}_{k+1}^2(k+1)=\tilde{\phi}_{k+1-d}^2(k+1-d)+o_p(1/T) .$$

Thus we have the result.

It follows from Proposition 3.1 that

$$\lim_{T\to\infty} \Pr\{\hat{\delta}=d,\ \hat{p}=p\}=p_0q_{K-(p+d)} ,$$

since $\hat{\delta}$ is a consistent estimator of d. Hence $\hat{p}$ is a reasonable estimator of p.

Remark 3.1. (3) also holds for $k=p+d-1$. Accordingly, (3) implies that the asymptotic distribution of $T^{1/2}\{\hat{\beta}(k)-\beta(k)\}$, $k \geqq p+d$, is the degenerate k-dimentional normal distribution with zero mean vector and covariance matrix,

$$\sigma_e^2 \tilde{D}'_{d,k-1} R_{k-d}^{-1} \tilde{D}_{d,k-1} ,$$

of rank $k-d$, where $\tilde{D}_{d,k-1}$ is the $(k-d)\times k$ matrix being composed of the first $(k-d)$ rows of $D_{d,k-1}$ and R_{k-d} is the $(k-d)\times(k-d)$ matrix whose (i, j)th element is $\mathrm{E}\, x_i x_j$.

4. Computational experiments

For estimating the degree of differencing, all of the computational results are obtained by putting $p, q \leqq 1$, $d=2$, and $D=5$. Hence the original process, w_t, is expressed as $(1-\phi_1 B)\nabla^2 w_t=(1-\theta_1 B)e_t$. We choose the values of ϕ_1 and θ_1 to be close to 1, which makes the estimation of d more difficult. And we set $f(\delta)=0.5\delta$, $g(k)=k^2$, $k=5$, and $l=12$. These values are chosen according to the method described at the end of the Section 2. We set $\varepsilon=0.1$, $P=Q=1$ and, next, evaluate σ_k^2 and $\sigma_{k,i}^2$ exactly for every pair of ϕ_1 and θ_1 of the form $\phi_1=-0.9+0.3j$, $\theta_1=-0.9+0.3m$, $0 \leqq j, m \leqq 6$. Then $k=5$ is the smallest value of k which satisfies

$$\sigma_k^2 < \sigma_{k,i}^2 ,$$

for every pair of ϕ_1 and θ_1, and, $1 \leqq i \leqq 5$. $g(k)=k^2$ and $l=12$ are arbitrarily chosen since we can adjust $c\sigma_{l-D}^2/k^\beta$ by using c. Finally we set $f(\delta)=c\delta=0.5\delta$ so that

$$\sigma_{5-j}^2-(\sigma_5^2-cj\sigma_{12-D}^2/5^2) ,$$

and

$$(\sigma_{5,i}^2-ci\sigma_{12-D}^2/5^2)-\sigma_5^2 ,$$

are as large as possible simultaneously for every pair of ϕ_1 and θ_1.

All of the simulations reported below were performed by using pseudo normal random numbers.

We see from Table 1 that this method works effectively except the case when $\phi_1=0.9$ and $\theta_1=-0.9$ as the sample size increases. When $\phi_1=0.9$ and $\theta_1=-0.9$, $|h_1|$ is close to 1 while ϕ_1 is close to 1 and, hence, c^* is very small. Consequently our method does not work well in this case.

On the other hand, if $\rho=\mathrm{Cor}\,(x_t, x_{t+1})$ is near to or greater than 1/2, the degree of differencing which minimizes sample variance often misleads to overdifferencing, since $\mathrm{Var}\,(\nabla x_t)=(2-2\rho)\,\mathrm{Var}\,(x_t)$ is equal to or less than $\mathrm{Var}\,(x_t)$ if $\rho\geqq 1/2$. This is supported by the computational results for $k=0$ in Table 1.

Table 1. The frequency of the degree of differencing in 100 realizations of an ARIMA (1, 2, 1) process

ϕ	θ	ρ	T	k	$f(\delta)\ \diagdown\ \delta$	0	1	2	3	4	5
0.90	0.00	0.90	200	0	0.0δ	0	0	0	100	0	0
				5	0.5δ	2	9	76	13	0	0
			500	0	0.0δ	0	0	0	100	0	0
				5	0.5δ	0	0	98	2	0	0
0.00	0.90	−0.50	200	0	0.0δ	0	43	57	0	0	0
				5	0.5δ	25	22	53	0	0	0
			500	0	0.0δ	0	17	83	0	0	0
				5	0.5δ	8	8	84	0	0	0
0.90	0.40	0.82	200	0	0.0δ	0	0	0	100	0	0
				5	0.5δ	0	10	76	14	0	0
			500	0	0.0δ	0	0	0	100	0	0
				5	0.5δ	0	0	97	3	0	0
0.40	0.90	−0.33	200	0	0.0δ	0	2	98	0	0	0
				5	0.5δ	23	29	48	0	0	0
			500	0	0.0δ	0	0	100	0	0	0
				5	0.5δ	5	8	87	0	0	0
0.90	−0.90	0.95	200	0	0.0δ	0	0	0	86	14	0
				5	0.5δ	0	1	58	41	0	0
			500	0	0.0δ	0	0	0	95	5	0
				5	0.5δ	0	0	57	43	0	0

Now let us see the results of simulations on the identification of an ARI (p, d) process. We considered three ARI (1, 2) processes and set $K=12$ in order to refer to the results of Shibata [12]. It can be seen from Table 2 that although the processes are nonstationary and the number of realizations is small, the "Total" values are close to the asymptotic values in Table 1 of Shibata [12]. And the values for $\hat{k}=k$, and, $\hat{\delta}=d$, that is, $\hat{p}=k-d$ and $\hat{\delta}=d$, are almost proportional to the asymptotic values when $T=200$, and are closer to them when $T=500$ since $\hat{\delta}$ is a consistent estimator of d.

Table 2. The frequency of the identification in 100 realizations of an ARI (1, 2) process

ϕ	T	$\hat{\delta}\ \diagdown\ \hat{k}$	0	1	2	3	4	5	6	7	8	9	10	11	12
0.9	200	0	0	0	0	0	2	0	0	0	0	0	0	0	0
		1	0	0	0	6	3	0	0	0	0	0	0	0	0
		2	0	0	0	52	13	3	2	1	1	2	2	0	0
		3	0	0	0	4	3	1	1	1	2	1	0	0	0
		4	0	0	0	0	0	0	0	0	0	0	0	0	0
		5	0	0	0	0	0	0	0	0	0	0	0	0	0
		Total	0	0	0	62	21	4	3	2	3	3	2	0	0

Table 2. (Continued)

ϕ	T	$\hat{\delta}$ \ $\hat{k}$	0	1	2	3	4	5	6	7	8	9	10	11	12
	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		1	0	0	0	0	0	0	0	0	0	0	0	0	0
		2	0	0	0	64	17	5	4	2	2	0	2	1	1
		3	0	0	0	0	1	0	0	0	0	0	0	1	0
		4	0	0	0	0	0	0	0	0	0	0	0	0	0
		5	0	0	0	0	0	0	0	0	0	0	0	0	0
		Total	0	0	0	64	18	5	4	2	2	0	2	2	1
0.8	200	0	0	0	0	1	0	0	0	0	0	0	0	0	0
		1	0	0	0	4	2	0	0	0	0	1	0	0	0
		2	0	0	0	56	15	6	2	2	1	3	2	0	0
		3	0	0	0	1	2	0	0	0	2	0	0	0	0
		4	0	0	0	0	0	0	0	0	0	0	0	0	0
		5	0	0	0	0	0	0	0	0	0	0	0	0	0
		Total	0	0	0	62	19	6	2	2	3	4	2	0	0
	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		1	0	0	0	0	0	0	0	0	0	0	0	0	0
		2	0	0	0	61	18	7	2	3	1	2	2	3	1
		3	0	0	0	0	0	0	0	0	0	0	0	0	0
		4	0	0	0	0	0	0	0	0	0	0	0	0	0
		5	0	0	0	0	0	0	0	0	0	0	0	0	0
		Total	0	0	0	61	18	7	2	3	1	2	2	3	1
0.3	200	0	0	0	0	2	0	0	0	0	0	0	0	0	0
		1	0	0	0	1	0	2	0	0	0	1	0	0	1
		2	0	0	0	62	13	7	2	2	3	2	0	1	0
		3	0	0	0	0	0	0	1	0	0	0	0	0	0
		4	0	0	0	0	0	0	0	0	0	0	0	0	0
		5	0	0	0	0	0	0	0	0	0	0	0	0	0
		Total	0	0	0	65	13	9	3	2	3	3	0	1	1
	500	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		1	0	0	0	1	0	0	0	0	0	0	0	0	0
		2	0	0	0	74	13	4	1	4	0	1	1	1	0
		3	0	0	0	0	0	0	0	0	0	0	0	0	0
		4	0	0	0	0	0	0	0	0	0	0	0	0	0
		5	0	0	0	0	0	0	0	0	0	0	0	0	0
		Total	0	0	0	75	13	4	1	4	0	1	1	1	0
Asymptotic			0	0	0	72	11	6	4	2	2	1	1	1	1

The total of the asymptotic values is 101. This is due to the rounding errors.

Appendix: The sample covariance matrix

In this appendix, we shall derive the asymptotic distribution of $D_T^{-1}W_T D_T^{-1}$, a sample covariance matrix of an ARIMA process. This is a generalization of the results given by Hasza and Fuller [10]. This has its own interest and is necessary to investigate whether or not the assumption imposed on the main results in Section 2 holds. Define w_t' by

$$w_t'=\begin{cases} \sum_{l=1}^{t}(l+d-2)!/\{(l-1)!(d-1)!\}x_{t+1-l}\,, & t\geqq 1\,, \\ 0\,, & t\leqq 0\,. \end{cases}$$

And let M be the $d\times d$ matrix whose (i, j) element is $(-i+1)^{j-1}$. Then w_t is expressed as

(A.1) $$w_t = w'_t + \sum_{j=0}^{d-1} v_j t^j , \qquad t \geqq -d+1 ,$$

where $\{v_0, \cdots, v_{d-1}\}$ satisfies

$$\{w_0, \cdots, w_{-d+1}\} = \{v_0, \cdots, v_{d-1}\} M' .$$

Then we have that $\mathrm{E}\, v_i^2 < \infty$, $0 \leqq i \leqq d-1$. Now we shall prepare a lemma which is used frequently later.

LEMMA A.1. (i) *If* $0 < i, j \leqq d$,

$$p-\lim_{T\to\infty} \left(\sum_{t=1}^{T-k} \nabla^{d-i} w_{t+d-i} \nabla^{d-j} w_{t+d-j} - \sum_{t=1}^{T-k} \nabla^{d-i} w'_{t+d-i} \nabla^{d-j} w'_{t+d-j} \right) \Big/ T^{i+j} = 0 .$$

(ii) *If* $0 < i, j \leqq d$,

$$p-\lim_{T\to\infty} \left(\sum_{t=1}^{T-k} \nabla^{d-i} w'_{t+d-i+h} \nabla^{d-j} w'_{t+d-j+l} - \sum_{t=1}^{T-k} \nabla^{d-i} w'_{t+d-i} \nabla^{d-j} w'_{t+d-j} \right) \Big/ T^{i+j} = 0 .$$

for any finite integers h *and* l.

The proof is easily shown by using (A.1) and, hence, is omitted.

Now we shall introduce some notations according to Hasza and Fuller [10]. B_T is the $(T-1)\times(T-1)$ lower triangular matrix with $B_{ij}=1$ for $i \geqq j$ and A_T is defined by $A_T = B'_T B_T$. J_T is the $(T-1)\times 1$ vector defined by

$$J_T = (1, 1, \cdots, 1)' .$$

Further let $\lambda_T = (\lambda_{1T}, \cdots, \lambda_{T-1,T})'$ denote the vector of the eigenvalues of A_T and $\tau_{iT} = (\tau_{iT,1}, \cdots, \tau_{iT,T-1})'$ be the eigenvector associated with λ_{iT}. Then it can be shown that

$$\lambda_{iT} = \sec^2 \{(2T-1)^{-1}(T-i)\pi\}/4 ,$$

and

$$\tau_{iT,t} = 2(2T-1)^{-1/2} \cos \{(2T-1)^{-1}(2i-1)\pi(t-1/2)\} ,$$

(Hasza and Fuller [10], p. 1111). Then we have the following lemma.

LEMMA A.2. *Let* $\gamma_i = 2\{(2i-1)\pi\}^{-1}(-1)^{i+1}$ *for* $i \geqq 1$ *and*

$$\nu_k = 2^{2(k+1)}(2^{2(k+1)}-1)\tilde{B}_{2(k+1)}/(2k+2)! ,$$

where $\tilde{B}_{2(k+1)}$ *is Bernoulli's number. Then*

(i) $\lim_{T\to\infty} \sum_{i=1}^{T-1} \lambda_{iT}^k / T^{2k} = \sum_{i=1}^{\infty} \gamma_i^{2k} = \nu_{k-1}/2$,

(ii) $\lim_{T\to\infty} J'_T A_T^k J_T / T^{2k+1} = 2 \sum_{i=1}^{\infty} \gamma_i^{2(k+1)} = \nu_k$.

PROOF. We see that $J_T' A_T^k J_T / T^{2k+1} = \sum_{i=1}^{T-1} \lambda_{iT}^k \left(\sum_{t=1}^{T-1} \tau_{iT,t} \right)^2 \Big/ T^{2k+1}$. Since

$$\lambda_{iT}^k / T^{2k} \leqq C_k / (2i-1)^{2k} ,$$

and

$$\lambda_{iT}^k \left(\sum_{t=1}^{T-1} \tau_{iT,t} \right)^2 \Big/ T^{2k+1} \leqq C_k' / (2i-1)^{2k+2} ,$$

where C_k and C_k' are constants, we have the assertion by the bounded convergence theorem.

Now we give the asymptotic distribution of $D_T^{-1} W_T D_T^{-1}$ as $T \to \infty$.

PROPOSITION A.1. *Let $\{V_i\}_{i=1}^{\infty}$ be a sequence of independent normal $(0, \sigma_e^2)$ random variables. And let $S_1(m)$, $S_2(m)$, and $S_3(m)$ be defined by*

$$S_1(m) = \sum_{l=1}^{\infty} \gamma_l^{2m} V_l^2 , \qquad S_2(m) = \sum_{l=1}^{\infty} \gamma_l^{2m+1} V_l , \qquad S_3(m) = \sum_{l=1}^{\infty} \mu_{l,m} V_l ,$$

respectively, where

$$\mu_{l,m} = \begin{cases} \sum_{i=0}^{n-1} (-1)^{n-1-i} \gamma_l^{2(n-i)} / (2i)! , & m = 2n , \\ \sum_{i=0}^{n-1} (-1)^{n-1-i} \gamma_l^{2(n-i)} / (2i+1)! + (-1)^n \gamma_l^{2n+1} , & m = 2n+1 , \end{cases}$$

And let W be the $d \times d$ random matrix whose (i, j) component W_{ij} $(= W_{ji})$ is defined by

$$\begin{aligned} W_{ij} = {} & c_0 (-1)^{m+n} \Bigg[S_1(m+n) + 2 \left\{ \sum_{l=1}^{m} (-1)^l S_2(m+n-l) S_3(2l) \right. \\ & \left. + \sum_{l=1}^{n} (-1)^l S_2(m+n-l) S_3(2l) \right\} \\ & + 2 \sum_{l=1}^{m} \sum_{a=1}^{n} (-1)^{l+a} \nu_{m+n-l-a} S_3(2l) S_3(2a) \Bigg] , \\ & \qquad (i = 2m,\ j = 2n,\ m, n \geqq 1) , \\ = {} & c_0 (-1)^{m+n} \Bigg[S_1(m+n+1) + 2 \left\{ \sum_{l=1}^{m} (-1)^l S_2(m+n+1-l) S_3(2l) \right. \\ & \left. + \sum_{l=1}^{n} (-1)^l S_2(m+n+1-l) S_3(2l) \right\} \\ & + 2 \sum_{l=1}^{m} \sum_{a=1}^{n} (-1)^{l+a} \nu_{m+n+1-l-a} S_3(2l) S_3(2a) \Bigg] , \\ & \qquad (i = 2m+1,\ j = 2n+1,\ m, n \geqq 0) , \\ = {} & c_0 \left\{ \sum_{l=0}^{m-1} (-1)^l 2 S_3(i+l+1) S_3(j-l) + (-1)^m S_3^2(j-m) \right\} , \\ & \qquad (j = i+2m+1,\ m \geqq 0) , \end{aligned}$$

where $c_0=(1-\theta_1-\cdots-\theta_q)^2/(1-\phi_1-\cdots-\phi_p)^2$. *Then we have*

$$D_T^{-1}W_TD_T^{-1}\xrightarrow[L]{} W$$

as $T\to\infty$ *where* L *means convergence in distribution.*

PROOF. The proof for the most part is essentially the same as the one given by Hasza and Fuller [10]. Thus we only show the outline. Define

$$u_t=\begin{cases}\sum_{l=1}^{t}(l+d-2)!/\{(l-1)!(d-1)!\}e_{t+1-l}\,, & t\geqq 1\,,\\ 0\,, & t\leqq 0\,,\end{cases}$$

$$e_T=(e_1, e_2,\cdots, e_{T-1})'\,,$$

$$\tau_T=(\tau_{1T}, \tau_{2T},\cdots, \tau_{T-1,T})'\,,$$

$$V_T=(V_{1T}, V_{2T},\cdots, V_{T-1,T})'=\tau_Te_T\,,$$

$$\Lambda_T=\mathrm{diag}\,(\lambda_{1T}, \lambda_{2T},\cdots, \lambda_{T-1,T})\,.$$

The random variables discussed later can be written by using $e_T'A_T^le_T$, $e_T'A_T^rJ_T$ and $\nabla^{d-a}u_{T-b}$. After these terms are expressed in terms of V_T and $\tau_T'\Lambda_T\tau_T$ $(=A_T)$, it can be shown in a similar way to the analogous part of Hasza and Fuller [10] that

$$(e_T'A_T^le_T/T^{2l}, e_T'A_T^rJ_T/T^{2r+1/2}, \nabla^{d-a}u_{T-b}/T^{a-1/2})\xrightarrow[L]{}\{S_1(l), 2^{1/2}S_2(r), 2^{1/2}S_3(a)\}\,, \tag{A.2}$$

as $T\to\infty$ for any finite integers l, r, b, and a with $1\leqq a\leqq d$.

Consider the case $i=2m$, $j=2n$. Putting

$$\nabla^sU_{Ti}=\{\nabla^su_{1-i}, \nabla^su_{2-i},\cdots, \nabla^su_{(T-1)-i}\}'\,,$$

we can show that the asymptotic distribution of

$$\sum_{t=1}^{T-k}\nabla^{d-i}w_{t+d-i}\nabla^{d-j}w_{t+d-j}/T^{i+j}\,,$$

is equal to that of

$$c_0\nabla^{d-2m}U_{Tm}\nabla^{d-2n}U_{Tn}/T^{i+j}\,,$$

by the repeated use of Lemma A.1. Here $\nabla^{d-2m}U_{Tm}$ is expressed as

$$(-1)^m\nabla^{d-2m}U_{Tm}=A_T^me_T+\sum_{l=1}^{m}(-1)^lA_T^{m-l}J_T\nabla^{d-2l}u_{T-l}\,. \tag{A.3}$$

Intuitively, A_T is the matrix representation of the "operator" $\tilde{A}_T$ defined by

$$\tilde{A}_T \nabla^s U_{Ti} = J_T(I-B)^{-2}\nabla^s u_{T-1-i} - (I-B)^{-2} B \nabla^s U_{Ti}$$
$$= J_T \nabla^{s-2} u_{T-1-i} - \nabla^{s-2} U_{T,i+1} \ .$$

By operating A_T on e_T $(=\nabla^d U_{T,0})$ m times, $\tilde{A}_T$ takes it into (A.3). Then the assertion follows from (A.2) and (A.3). Lemma A.2 is used to evaluate $\sum_{i=1}^{T-1} \lambda_{i,T}^l / T^{2l}$ and $J_T' A_T^l J_T / T^{2l+1}$. The result for $i=2m+1$ and $j=2n+1$ is shown in a similar way by noting $\nabla^{d-2m-1} U_{Tm} = B_T(\nabla^{d-2m} U_{Tm})$. Finally consider the case $j=i+2m+1$. This case is proved in a different way. We have

$$\nabla^{d-l} u_T \nabla^{d-n} u_T = \sum_{t=1}^{T} (\nabla^{d-l} u_t \nabla^{d-n} u_t - \nabla^{d-l} u_{t-1} \nabla^{d-n} u_{t-1})$$
$$= \sum_{t=1}^{T} (\nabla^{d-l+1} u_t \nabla^{d-n} u_t + \nabla^{d-l} u_t \nabla^{d-n+1} u_t) + o_p(T^{l+n-1}) \ .$$

Consequently

$$\sum_{t=1}^{T} \nabla^{d-j} u_t \nabla^{d-i} u_t = \sum_{l=0}^{m-1} (-1)^l \nabla^{d-j+l} u_T \nabla^{d-j+2m-l} u_T$$
$$+ (-1)^m (\nabla^{d-j+m} u_T)^2 / 2 + o_p(T^{i+j}) \ .$$

Then the assertion follows from (A.2).

Remark A.1. If we put $i=j=2$, then $W_{22}/c_0 = \{S_1(2) - 4S_2(1)S_3(2) + 2S_3^2(2)\} = \left\{\sum_{l=1}^{\infty} \gamma_l^4 V_l^2 - 4\left(\sum_{l=1}^{\infty} \gamma_l^3 V_l\right)\left(\sum_{l=1}^{\infty} \gamma_l^2 V_l\right) + 2\left(\sum_{l=1}^{\infty} \gamma_l^2 V_l\right)^2\right\}$ which is equal to the limiting distribution of $\sum_{t=1}^{n} Y_{t-1}^2 / n^4$ in Lemma 3.1. (vi) of Hasza and Fuller [10].

Acknowledgement

The author is grateful to Professor M. Huzii for his valuable advice. He is also grateful to the referee and an editor for their helpful comments which led to improvements in this paper.

TOKYO INSTITUTE OF TECHNOLOGY

REFERENCES

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, in *2nd International Symposium on Information Theory*, (eds. B. N. Petrov and F. Csáki), Budapest, Akadémia Kiado, 267-281.

[2] Akaike, H. (1974). A new look at the statistical model identification, *I.E.E.E. Trans. Auto. Control*, AC-**19**, 716-723.

[3] Anderson, O. D. (1975). A note on differencing autoregressive moving average (p, q) processes, *Biometrika*, **62**, 521-523.

[4] Anderson, O. D. (1976). *Time Series Analysis and Forecasting, The Box-Jenkins Approach*, Butterworths, London.

[5] Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley, New York.

[6] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[7] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control* (Revised ed.), Holden Day, San Francisco.

[8] Grenander, U. and Rosenblatt, M. (1954). An extension of a theorem of G. Szegő and its application to the study of stochastic processes, *Trans. Amer. Math. Soc.*, **76**, 112-126.

[9] Hannan, E. J. (1980). The estimation of the order of an ARMA process, *Ann. Statist.*, 8, 1071-1081.

[10] Hasza, D. P. and Fuller, W. A. (1979). Estimation for autoregressive processes with unit roots, *Ann. Statist.*, **7**, 1106-1120.

[11] Ozaki, T. (1977). On the order determination of ARIMA models, *Appl. Statist.*, **26**, 290-301.

[12] Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika*, **63**, 117-126.

[13] Yajima, Y. (1980). On an effect of overdifferencing on the prediction of ARIMA processes, in *Recent Developments in Statistical Inference and Data Analysis*, (ed. K. Matusita), North-Holland, Amsterdam, 335-345.