JOURNAL OF MULTIVARIATE ANALYSIS 27, 40–52 (1988)

On Determination of the Order of an Autoregressive Model

Z. D. BAI, K. SUBRAMANYAM, AND L. C. ZHAO*

University of Pittsburgh

Communicated by the Editors

To determine the order of an autoregressive model, a new method based on information theoretic criterion is proposed. This method is shown to be strongly consistent and the convergence rate of the probability of wrong determination is established.

1. INTRODUCTION

Consider an autoregressive (AR) model of order p ($p \geqslant 1$, unknown) generated by a purely random process $e(n)$ given by

$$\sum_{j=0}^{p} \alpha(j)\, X(n-j) = e(n), \qquad \alpha(0) = 1. \tag{1.1}$$

Assume that $\{e(n)\}$ is a sequence of i.i.d. random variables with $Ee(1) = 0$, $Ee^2(1) = \sigma^2$ and $0 < \mathrm{Var}(e^2(1)) < \infty$. Suppose the coefficients in the model $\alpha(0), \alpha(1), ..., \alpha(p)$ satisfy

$$g(z) = \sum_{j=0}^{p} \alpha(j) z^j \neq 0 \qquad \text{for} \quad |z| \leqslant 1. \tag{1.2}$$

In time series analysis, AR models play an important role. An interesting problem in the analysis of AR models is the determination of the order p of the model. There is a considerable amount of research work done on this topic. To name a few, the reader is referred to Akaike [1], Hannan [3], Hannan and Quinn [4], and Shibata [6].

Received January 5, 1988; accepted April 28, 1988.

AMS 1980 subject classifications: Primary 62M10; Secondary 60G10.

Key words and phrases: autoregressive model, convergence rate, strong consistency, time series.

* Present address: Dept. of Mathematics, University of Science & Technology of China, Hefei, Anhui, Peoples Republic of China.

0047-259X/88 $3.00

Let $X(1)$, $X(2)$, ..., $X(N)$ denote a random sample drawn from an AR model of order p. Assume that the order p is known a priori to be $p \leqslant K < \infty$. Using Yule–Walker equations and a recursive computing procedure, Hannan and Quinn [4] obtained an estimate $\hat{\sigma}_p^2$ of σ^2. To estimate p, the following criterion based on $\hat{\sigma}_p^2$ is proposed,

$$\psi(p) = \log \hat{\sigma}_p^2 + 2p\, CN^{-1} \log \log N, \tag{1.3}$$

where $C > 1$ is a constant. An estimate $\hat{p}$ of p is chosen as that one which minimises $\psi(p)$. Under weaker conditions than mentioned above, strong consistency of $\hat{p}$ is obtained.

In this article a new criterion to estimate the order of the AR model is proposed. Strong consistency as well as the convergence rate of the estimate $\hat{p}$ is established.

The paper is organized as follows. In Section 2, a new method to determine the order AR model is described. In Section 3, convergence rates of $P\{\hat{p} \neq p\}$ is derived. Some general remarks, including the strong consistency of $\hat{p}$, are made in Section 4.

2. Determination of the Order p

Let $X(1)$, $X(2)$, ..., $X(N)$ be a random sample from an AR series. Define

$$L_p(\boldsymbol{\alpha}_p) = \sum_{n=p+1}^{N} \left(X(n) + \sum_{i=1}^{p} \alpha(i)\, X(n-i) \right)^2, \tag{2.1}$$

where $\boldsymbol{\alpha}_p = (\alpha(1), ..., \alpha(p))'$. The true order p of the model and the true regression coefficients $\alpha(1), ..., \alpha(p)$ will be denoted as p_0, $\alpha_0(1), ..., \alpha_0(p_0)$, respectively.

For each $p \leqslant K$ choose $\hat{\boldsymbol{\alpha}}_p = (\hat{\alpha}(1), ..., \hat{\alpha}(p))'$ such that

$$L_p(\hat{\boldsymbol{\alpha}}_p) = \min_{\boldsymbol{\alpha}_p} L_p(\hat{\boldsymbol{\sigma}}_p^2) \triangleq N\hat{\sigma}_p^2. \tag{2.2}$$

Since L_p is a quadratic form of $\boldsymbol{\alpha}_p$, it is easy to compute $\hat{\boldsymbol{\alpha}}_p$ and $L_p(\hat{\alpha}_p)$. Define

$$\phi(p) = N \log \left[\frac{1}{N} L_p(\hat{\boldsymbol{\alpha}}_p) \right] + pC_N, \tag{2.3}$$

where constants C_N will be chosen suitably. Then any $\hat{p}$ minimizing

$$\phi(\hat{p}) = \min_{p \leqslant K} \phi(p) \tag{2.4}$$

will be taken as the estimate of the order p of the AR series.

Remark 2.1. In fact, $(1/N)\, L_p(\hat{\boldsymbol{\alpha}}_p)$ is an estimate of σ^2, which is slightly different from that used by Hannan and Quinn [4]. When N is not very large, $(1/(N-p))\, L_p(\hat{\alpha}_p)$ is a better estimate of σ^2 as compared to $(1/N)\, L_p(\hat{\alpha}_p)$. Since we are interested in the large sample properties, there is no harm in using $(1/N)\, L_p(\hat{\alpha}_p)$ as an estimate of σ^2.

Define

$$\begin{aligned} \hat{q}_p(i, j) &= \frac{1}{N} \sum_{n=p+1}^{N} X_{n-i} X_{n-j}, \quad i, j = 0, 1, 2, ..., p. \\ \hat{Q}_p &= (\hat{q}_p(i, j))_{i, j = 1, 2, ..., p} \\ \hat{\beta}_p &= (\hat{q}_p(0, 1), ..., \hat{q}_p(0, p))'. \end{aligned} \tag{2.5}$$

By differentiating $L_p(\hat{\boldsymbol{\alpha}}_p)$, we get

$$\hat{Q}_p \hat{\boldsymbol{\alpha}}_p = -\hat{\boldsymbol{\beta}}_p$$

or, equivalently,

$$\hat{\boldsymbol{\alpha}}_p = -\hat{Q}_p^{-1} \hat{\boldsymbol{\beta}}_p \tag{2.6}$$

provided $\hat{Q}_p$ is nonsingular. In the proof of our main result, it is shown that with probability one, for large N, $\hat{Q}_p$ is nonsingular. Hence we can use (2.6).

Using the above notation, the main theorems are stated below. Proofs are given in the next section.

THEOREM 2.1. *Suppose*

$$E \exp\{te(1)^2\} < \infty \qquad \textit{for some} \quad t > 0, \tag{2.7}$$

and choose C_N *such that*

$$C_N/N \to 0, \qquad C_N \to \infty. \tag{2.8}$$

Then

$$P(\hat{p} \neq p_0) \leqslant C_1 \exp\{-C_2 C_N\}, \tag{2.9}$$

where C_1, C_2 *are two positive constants independent of* N.

THEOREM 2.2. *Suppose* (2.8) *holds and*

$$E\, |e(1)|^{2t} < \infty, \qquad \textit{for some} \quad t \geqslant 2. \tag{2.10}$$

Then

$$P(\hat{p} \neq p_0) \leqslant C_1/(N^{t/2-1} C_N^{t/2}) + C_2 e^{-C_3 C_N}, \tag{2.11}$$

where C_1, C_2, C_3 *are positive constants independent of* N.

3. Proof of the Theorems

LEMMA 3.1. *Let $y_1, ..., y_n$ be independent random variables with $Ey_i = 0$ and $E|y_i|^t < \infty$, $i = 1, ..., n$, for some $t \geq 2$. Denote*

$$S_n = \sum_{i=1}^{n} y_i, \quad B_n^2 = \sum_{i=1}^{n} \mathrm{Var}(y_i), \quad A_{t,n} = \sum_{i=1}^{n} E|y_i|^t.$$

Then for any $a > 0$,

$$P\{S_n \geq a\} \leq C_t^{(1)} A_{t,n} a^{-t} + \exp\{-C_t^{(2)} a^2/B_n^2\},$$

where

$$C_t^{(1)} = (1 + 2/t)^t \qquad \text{and} \qquad C_t^{(2)} = 2(t+2)^{-2} e^{-t}.$$

Proof. Refer to Corollary 4 of Fuk and Nagaev [2].

Let $\boldsymbol{\alpha}_{p_0} = (\alpha_0(1), ..., \alpha_0(p_0))'$ and σ^2 be the true parameters of the model. Let

$$\begin{aligned} &\gamma(i-j) = E(X(n-i)\, X(n-j)), \\ &\Gamma_p = ((r(i-j)))_{i,j=1,...,p}, \boldsymbol{\gamma}_p = (\gamma(1), ..., \gamma(p)),\ p \leq K. \end{aligned} \tag{3.1}$$

Suppose $p \geq p_0$, then from

$$\sum_{i=0}^{p_0} \alpha_0(i)\, X(n-i) = e(n),$$

it follows that

$$\sum_{i=0}^{p_0} \alpha_0(i)\, \gamma(i-j) = \delta_{0,j}\sigma^2, \qquad j = 0, 1, 2, ..., p, \tag{3.2}$$

where $\delta_{i,j}$ is Kronecker's delta. Thus, if we take $\boldsymbol{\alpha}_p^* = (\alpha_0(1), ..., \alpha_0(p_0), 0, ..., 0)'$, then $\boldsymbol{\alpha}_p = \boldsymbol{\alpha}_p^*$ is a unique solution of the equation

$$\Gamma_p \boldsymbol{\alpha}_p = -\boldsymbol{\gamma}_p. \tag{3.3}$$

It is well known that, under the conditions (1.1) and (1.2), for $0 \leq p \leq K$,

$$\lim_{N \to \infty} \hat{Q}_p = \Gamma_p \text{ a.s.}, \qquad \lim_{N \to \infty} \hat{\boldsymbol{\beta}}_p = \boldsymbol{\gamma}_p \text{ a.s.} \tag{3.4}$$

and

$$\lim_{N \to \infty} \hat{\boldsymbol{\alpha}}_p \overset{\text{a.s.}}{=} -\Gamma_p^{-1} \boldsymbol{\gamma}_p \triangleq \boldsymbol{\alpha}_p^* = (\alpha^*(1), ..., \alpha^*(p))'. \tag{3.5}$$

Note that for $p_0 \leqslant p \leqslant K$,

$$\boldsymbol{\alpha}_p^* = (\alpha_0(1), ..., \alpha_0(p_0), 0, ..., 0)', \tag{3.6}$$

and that

$$\lim_{N \to \infty} \hat{\sigma}_p^2 \overset{\text{a.s.}}{=} \gamma(0) - \boldsymbol{\alpha}_p^{*\prime} \Gamma_p \boldsymbol{\alpha}_p^* = \begin{cases} \sigma^{*2} > \sigma^2, & \text{if } p < p_0, \\ \gamma(0) - \boldsymbol{\alpha}_{p_0}' \Gamma_{p_0} \boldsymbol{\alpha}_{p_0} = \sigma^2, & \text{if } p \geqslant p_0. \end{cases} \tag{3.7}$$

It is easily seen that,

$$\begin{aligned} \hat{\sigma}_p^2 &= \min_{\boldsymbol{\alpha}_p} \frac{1}{N} \sum_{n=p+1}^{N} \left(X(n) + \sum_{i=1}^{p} \alpha(i) X(n-i) \right)^2 \\ &\geqslant \min_{\boldsymbol{\alpha}_{p+1}: \alpha(p+1)=0} \frac{1}{N} \sum_{n=p+2}^{N} \left(X(n) + \sum_{i=1}^{p+1} \alpha(i) X(n-i) \right)^2 \\ &\geqslant \min_{\boldsymbol{\alpha}_{p+1}} \frac{1}{N} \sum_{n=p+2}^{N} \left(X(n) + \sum_{i=1}^{p+1} \alpha(i) X(n-i) \right)^2 = \hat{\sigma}_{p+1}^2. \end{aligned} \tag{3.8}$$

First we establish the following proposition which will be used to prove our main theorems.

PROPOSITION 3.1. *Under conditions* (1.1), (1.2), *and* (2.8), *there exists a constant* $\varepsilon > 0$ *such that for large* N,

$$P\{\hat{p} \neq p_0\} \leqslant P_1 + P_2 + P_3 + P_4,$$

where

$$\begin{aligned} P_1 &= \sum_{i,j=0}^{K} P\{|\hat{q}_K(i,j) - \gamma(i,j)| > \varepsilon \sqrt{C_N/N}\} \\ P_2 &= \sum_{i=1}^{K} P\left\{ \left| \frac{1}{N} \sum_{n=K+1}^{N} e(n) X(n-i) \right| > \varepsilon \sqrt{C_N/N} \right\} \\ P_3 &= 2KP\{|e(0)| > \varepsilon \sqrt{C_N}\} \end{aligned} \tag{3.9}$$

and

$$P_4 = 2KP\{|X(0)| > \varepsilon \sqrt{C_N}\}.$$

Proof. Denote

$$\begin{aligned} A_1(\varepsilon) &= \{|\hat{q}_K(i,j) - \gamma(i-j)| \leqslant \varepsilon \sqrt{C_N/N} \quad \text{for all} \quad i, j \leqslant K\} \\ A_2(\varepsilon) &= \left\{ \left| \frac{1}{N} \sum_{n=K+1}^{N} e(n) X(n-i) \right| \leqslant \varepsilon \sqrt{C_N/N} \quad \text{for all} \quad 1 \leqslant i \leqslant K \right\} \\ A_3(\varepsilon) &= \{|e(n)| \leqslant \varepsilon \sqrt{C_N} \quad \text{for all} \quad n \leqslant 2K\} \\ A_4(\varepsilon) &= \{|X(n)| \leqslant \varepsilon \sqrt{C_N} \quad \text{for all} \quad n \leqslant 2K\}. \end{aligned}$$

For $p < p_0$, since $\hat{\sigma}_p^2$, as a function of $\hat{q}_K(i, j)$'s and $X(n)X(n-l)$, is continuously differentiable, we have

$$\begin{aligned}
\hat{\sigma}_p^2/\hat{\sigma}_{p_0}^2 &\geqslant \hat{\sigma}_{p_0-1}^2/\hat{\sigma}_{p_0}^2 \\
&= \frac{1}{N}\sum_{i=p_0}^{N}\left(X(n)+\sum_{i=1}^{p_0-1}\hat{\alpha}_{p_0-1}(i)\,X(n-i)\right)^2\Big/\hat{\sigma}_{p_0}^2 \\
&= \{\hat{q}_{p_0-1}(0,0)-\hat{\boldsymbol{\alpha}}'_{p_0-1}\hat{Q}_{p_0-1}\hat{\boldsymbol{\alpha}}_{p_0-1}\}/\hat{\sigma}_{p_0}^2 \\
&\geqslant \{\gamma(0)-\boldsymbol{\alpha}^{*\prime}_{p_0-1}\Gamma_{p_0-1}\boldsymbol{\alpha}^{*}_{p_0-1}\}/\sigma^2 \\
&\quad -C\left\{\sum_{i,j=0}^{K}|\hat{q}_K(i,j)-\gamma(i,j)|+\frac{1}{N}\sum_{n=1}^{2K}X^2(n)\right\}. \qquad (3.10)
\end{aligned}$$

Hereafter, C denotes a constant independent of N, but may take a different value at each appearance even in the same expression.

From (3.7), noting (3.10), there exists $\varepsilon > 0$ such that if $A_1(\varepsilon)\cap A_4(\varepsilon)$ holds then for any $p < p_0$ and large N,

$$\begin{aligned}
\log(\hat{\sigma}_p^2/\hat{\sigma}_{p_0}^2) &\geqslant \log(\hat{\sigma}_{p_0-1}^2/\sigma_{p_0}^2) \\
&\geqslant \log(\sigma^{*2}/\sigma^2)-C\varepsilon\sqrt{C_N/N} > (p_0-p)\,C_N/N. \qquad (3.11)
\end{aligned}$$

Now assume that $p_0 < p \leqslant K$. Put $\Delta\hat{\alpha}_p(i)=\hat{\alpha}_p(i)-\alpha^*(i)$, $\Delta\boldsymbol{\alpha}_p=\hat{\boldsymbol{\alpha}}_p-\boldsymbol{\alpha}_p^*$. By (2.2) and (3.8),

$$\begin{aligned}
0 &\geqslant \hat{\sigma}_p^2-\hat{\sigma}_{p_0}^2 \geqslant \hat{\sigma}_K^2-\hat{\sigma}_{p_0}^2 \\
&\geqslant \frac{1}{N}\sum_{n=K+1}^{N}\left(X(n)+\sum_{i=1}^{K}\hat{\alpha}_K(i)\,X(n-i)\right)^2 \\
&\quad -\frac{1}{N}\sum_{n=p_0+1}^{N}\left(X(n)+\sum_{i=1}^{p_0}\alpha_0(i)\,X(n-i)\right)^2 \\
&= \frac{1}{N}\sum_{n=K+1}^{N}\left(e(n)+\sum_{i=1}^{K}\Delta\hat{\alpha}_K(i)\,X(n-i)\right)^2 \\
&\quad -\frac{1}{N}\sum_{n=p_0+1}^{N}e(n)^2 \\
&\geqslant -\frac{1}{N}\sum_{n=1}^{K}e(n)^2-\hat{\psi}'\hat{Q}_K^{-1}\hat{\psi}, \qquad (3.12)
\end{aligned}$$

where $\hat{Q}_K$ is defined in (2.5) and $\hat{\psi} = (\hat{\psi}_1, ..., \hat{\psi}_K)'$, $\hat{\psi}_j = (1/N)\sum_{n=K+1}^{N} e(n)\,X(n-j)$, $j = 1, 2, ..., K$.

From this, one can see that, there exists $\varepsilon > 0$ such that for large N, if $A_1(\varepsilon)\cap A_2(\varepsilon)\cap A_3(\varepsilon)$ holds then for any $p_0 < p \leqslant K$,

$$\frac{\hat{\sigma}_{p_0}^2-\hat{\sigma}_p^2}{\hat{\sigma}_p^2} < \frac{C_N}{2N} \qquad (3.13)$$

which in turn implies that for any $p_0 < p \leqslant K$,

$$\log(\hat{\sigma}_p^2/\hat{\sigma}_{p_0}^2) > -C_N/N \geqslant -(p-p_0)\, C_N/N. \tag{3.14}$$

From (3.11) and (3.14), Proposition 3.1 follows.

Proof of Theorem 2.1. Hereafter, C is a positive constant independent of N which can be assigned as large as you wish, but may take a different value at each appearance. To prove Theorem 2.1, it is enough to show that

$$P_\eta < C \exp\{-CC_N\}, \quad \eta = 1, 2, 3, 4, \tag{3.15}$$

where P_η's are defined in (3.9). It is easy to see that (3.15) is true for $\eta = 3, 4$ using (2.7). By (2.4),

$$X(n) = \sum_{j=0}^{\infty} a_j e(n-j), \quad |a_j| \leqslant M\rho^j, \quad j = 0, 1, 2, ..., \tag{3.16}$$

where $\rho \in (0, 1)$ and $M > 0$ are constants. In order to prove (3.15) for $\eta = 1, 2$, it is enough to show that for any $\varepsilon > 0$,

$$P\left\{\left|\frac{1}{N}\sum_{n=1}^{N} X(n)\, X(n-l) - \gamma(l)\right| > \varepsilon\sqrt{C_N/N}\right\} \leqslant C\exp\{-CC_N\}, \quad l = 0, 1, 2, ..., K, \tag{3.17}$$

and

$$P\left\{\left|\frac{1}{N}\sum_{n=1}^{N} e(n)\, X(n-l)\right| > \varepsilon\sqrt{C_N/N}\right\} \leqslant C\exp\{-CC_N\}, \quad l = 1, 2, ..., K. \tag{3.18}$$

By (3.16), $\gamma(l) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{l+j}$, and

$$\frac{1}{N}\sum_{n=1}^{N} X(n)\, X(n-l) = \sum_{i=l}^{\infty} a_i a_{i-l} \frac{1}{N}\sum_{n=1}^{N} e(n-i)^2 + \sum_{(i,j)\, i \neq l+j} a_i a_j \times \frac{1}{N}\sum_{n=1}^{N} e(n-i)\, e(n-j-l). \tag{3.19}$$

Fix $l \leqslant K$. Take $\rho_1 \in (\rho, 1)$ and set

$$B(\varepsilon_1) = \left\{ \left| \frac{1}{N} \sum_{n=1}^{N} e(n-i)^2 - \sigma^2 \right| < (\rho_1/\rho)^{2i-l} \varepsilon_1 \sqrt{C_N/N} \right.$$
$$\left. \text{for } i = l, l+1, ... \right\}$$

$$D(\varepsilon_1) = \left\{ \left| \frac{1}{N} \sum_{n=1}^{N} e(n-i)\, e(n-j-l) \right| \leqslant (\rho_1/\rho)^{i+j} \varepsilon_1 \sqrt{C_N/N} \right.$$
$$\left. \text{for any } i \neq l+j,\ i, j = 0, 1, 2, ... \right\}.$$

Take $\varepsilon_1 < \varepsilon M^{-2}(1-\rho_1)^2$. If $B(\varepsilon_1) \cap D(\varepsilon_1)$ occurs, using (3.19) we get

$$\left| \frac{1}{N} \sum_{n=1}^{N} X(n)\, X(n-l) - \gamma(l) \right|$$
$$\leqslant M^2 \sum_{i=l}^{\infty} \rho^{2i-l} \varepsilon_1 \sqrt{C_N/N}\, (\rho_1/\rho)^{2i-l}$$
$$+ M^2 \sum_{(i,j): i \neq l+j} \rho^{i+j} \varepsilon_1 \sqrt{C_N/N}\, (\rho_1/\rho)^{i+j}$$
$$\leqslant M^2 \left(\sum_{i=0}^{\infty} \rho_1^i \right)^2 \varepsilon_1 \sqrt{C_N/N} < \varepsilon \sqrt{C_N/N}.$$

Thus, taking $\lambda = \rho_1/\rho (>1)$ and $\varepsilon_2 = (\rho_1/\rho)^{-l} \varepsilon_1$ we get

$$P\left\{ \left| \frac{1}{N} \sum_{n=1}^{N} X(n)\, X(n-l) - \gamma(l) \right| > \varepsilon \sqrt{C_N/N} \right\}$$
$$\leqslant \sum_{i=0}^{\infty} P\left\{ \left| \frac{1}{N} \sum_{n=1}^{N} e(n)^2 - \sigma^2 \right| \geqslant \lambda^{2i} \varepsilon_2 \sqrt{C_N/N} \right\}$$
$$+ \sum_{i \neq j} P\left\{ \left| \frac{1}{N} \sum_{n=1}^{N} e(n)\, e(n-j+i) \right| \geqslant \lambda^{i+j} \varepsilon_2 \sqrt{C_N/N} \right\}. \quad (3.20)$$

Setting $f(\tau) = E \exp\{\tau(e(i)^2 - \sigma^2)\}$, $\tau \in (0, t)$, we have $f(\tau) = 1 + f'(0)\tau + \frac{1}{2} f''(\tau_1)\tau^2$, where $\tau_1 \in (0, \tau)$. Hence

$$f(\tau) \leqslant 1 + C\tau^2 \leqslant \exp\{C\tau^2\}. \quad (3.21)$$

Thus,

$$P\left\{\frac{1}{N}\sum_{n=1}^{N}(e(n)^2-\sigma^2)\geqslant\lambda^{2i}\varepsilon_2\sqrt{C_N/N}\right\}$$
$$\leqslant\exp(-\tau\lambda^{2i}\varepsilon_2\sqrt{NC_N})\,f(\tau)^N$$
$$\leqslant\exp\{-\tau\lambda^{2i}\varepsilon_2\sqrt{NC_N}+C\tau^2N\}. \tag{3.22}$$

Taking $\tau=\delta\sqrt{C_N/N}\,\lambda^{2i}$, where $\delta>0$ is small, one can see that

$$P\left\{\frac{1}{N}\sum_{n=1}^{N}(e(n)^2-\sigma^2)\geqslant\lambda^{2i}\varepsilon_2\sqrt{C_N/N}\right\}$$
$$\leqslant\exp\{-C\lambda^{4i}C_N\}\leqslant C\lambda^{-2i}\exp(-CC_N). \tag{3.23}$$

In the same way,

$$P\left\{\frac{1}{N}\sum_{n=1}^{N}(e(n)^2-\sigma^2)\leqslant-\lambda^{2i}\varepsilon_2\sqrt{C_N/N}\right\}$$
$$\leqslant C\lambda^{-2i}\exp(-CC_N). \tag{3.24}$$

In a similar fashion it follows that if $\tau\in(0,t)$

$$E\exp(\tau e(0)\,e(i-j))\leqslant\exp(C\tau^2). \tag{3.25}$$

For $i>j$, by (3.25),

$$P\left\{\sum_{n=1}^{N}e(n)\,e(n+i-j)\geqslant\lambda^{i+j}\varepsilon_2\sqrt{NC_N}\right\}$$
$$\leqslant\sum_{m=0}^{i-j}P\left\{\sum_{n\leqslant N,n\equiv m(\operatorname{mod}(i-j+1))}e(n)\,e(n+i-j)\right.$$
$$\left.\geqslant\frac{\varepsilon_2}{i-j+1}\lambda^{i+j}\sqrt{NC_N}\right\}$$
$$\leqslant(i-j+1)\exp\left\{-\tau\frac{\varepsilon_2}{i-j+1}\lambda^{i+j}\sqrt{NC_N}\right\}$$
$$\times\exp\left(C\frac{N}{i-j+1}\tau^2\right). \tag{3.26}$$

Taking $\tau=\delta\sqrt{C_N/N}\,\lambda^{i+j}$, where $\delta>0$ is small enough to get

$$P\left\{\sum_{n=1}^{N}e(n)\,e(n+i-j)\geqslant\lambda^{i+j}\varepsilon_2\sqrt{NC_N}\right\}$$
$$\leqslant C(i+j)\exp\left\{-C\frac{1}{i-j+1}\lambda^{i+j}C_N\right\}$$
$$\leqslant C\exp\{-CC_N\}\lambda^{-i-j}. \tag{3.27}$$

Similarly,

$$P\left\{\sum_{n=1}^{N} e(n)\,e(n+i-j) \leqslant -\lambda^{i+j}\varepsilon_2\sqrt{NC_N}\right\}$$
$$\leqslant C\exp\{-CC_N\}\lambda^{-i-j}. \tag{3.28}$$

Note that (3.27), (3.28) hold for $i<j$. Thus, by (3.20), (3.23), (3.24), (3.27), and (3.28),

$$P\left\{\left|\frac{1}{N}\sum_{n=1}^{N} X(n)\,X(n-l)-\gamma(l)\right| \geqslant \varepsilon\sqrt{C_N/N}\right\}$$
$$\leqslant 2C\sum_{i=0}^{\infty}\lambda^{-2i}\exp(-CC_N)+2C\exp(-CC_N)\sum_{i,j=0}^{\infty}\lambda^{-i-j}$$
$$\leqslant C\exp(-CC_N), \tag{3.29}$$

which is (3.17). The proof of (3.18) is similar. That completes the proof of Theorem 2.1.

Proof of Theorem 2.2. The line of proof is similar to that of Theorem 2.1. Here Lemma 3.1 is used. For example, in order to prove

$$P\left\{\left|\frac{1}{N}\sum_{n=1}^{N} X(n)^2-\gamma(0)\right| > \varepsilon\sqrt{C_N/N}\right\}$$
$$\leqslant CN^{-t/2+1}(C_N)^{-t/2}+C\exp(-CC_N), \tag{3.30}$$

we use $\gamma(0)=\sigma^2\sum_{j=0}^{\infty}a_j^2$ and

$$\frac{1}{N}\sum_{n=1}^{N} X(n)^2=\sum_{j=0}^{\infty}a_j^2\frac{1}{N}\sum_{n=1}^{N} e(n-j)^2$$
$$+\sum_{i\neq j}a_i a_j\frac{1}{N}\sum_{n=1}^{N} e(n-i)\,e(n-j). \tag{3.31}$$

Take $\rho_1\in(\rho, 1)$ and set

$$B(\varepsilon_1)=\left\{\left|\frac{1}{N}\sum_{n=1}^{N} e(n-j)^2-\sigma^2\right| < (\rho_1/\rho)^{2j}\right.$$
$$\left.\times\,\varepsilon_1\sqrt{C_N/N}\text{ for } j=0, 1, 2, \ldots\right\},$$

$$D(\varepsilon_1)=\left\{\left|\frac{1}{N}\sum_{n=1}^{N} e(n-i)\,e(n-j)\right| \leqslant (\rho_1/\rho)^{i+j}\right.$$
$$\left.\times\,\varepsilon_1\sqrt{C_n/N}\text{ for any } i\neq j,\ i, j=0, 1, 2, \ldots\right\}.$$

As before, by taking $\varepsilon_1 < \varepsilon M^{-2}(1-\rho_1)^2$, we get, when $B(\varepsilon_1) \cap D(\varepsilon_1)$ occurs

$$\left| \frac{1}{N} \sum_{n=1}^{N} X(n)^2 - \gamma(0) \right| < \varepsilon \sqrt{C_N/N}.$$

Thus, with $\lambda = \rho_1/\rho$ (>1), we have

$$\begin{aligned} P&\left\{ \left| \frac{1}{N} \sum_{n=1}^{N} X(n)^2 - \gamma(0) \right| > \varepsilon \sqrt{C_N/N} \right\} \\ &\leqslant \sum_{i=0}^{\infty} P\left\{ \left| \frac{1}{N} \sum_{n=1}^{N} e(n)^2 - \sigma^2 \right| \geqslant \lambda^{2i} \varepsilon_1 \sqrt{C_N/N} \right\} \\ &\quad + \sum_{i,j=0,\, i \neq j}^{\infty} P\left\{ \left| \frac{1}{N} \sum_{n=1}^{N} e(n)\, e(n-j+i) \right| \geqslant \lambda^{i+j} \varepsilon_1 \sqrt{C_N/N} \right\}. \end{aligned}$$

By Lemma 3.1,

$$\begin{aligned} \sum_{j=0}^{\infty} & P\left\{ \left| \frac{1}{N} \sum_{n=1}^{N} (e(n)^2 - \sigma^2) \right| \geqslant \lambda^{2j} \varepsilon_1 \sqrt{C_N/N} \right\} \\ &\leqslant (1+2/t)^t \sum_{j=0}^{\infty} NE\, |e(1)^2 - \sigma^2|^t\, \varepsilon_1^{-t} \lambda^{-2jt} (NC_N)^{-t/2} \\ &\quad + \sum_{j=0}^{\infty} \exp\{-2(t+2)^{-2}\, e^{-t} \varepsilon^2 \lambda^{4j} NC_N / (N \operatorname{Var} e(1)^2)\} \\ &\leqslant C \sum_{j=0}^{\infty} \lambda^{-2j} N^{-t/2+1} C_N^{-t/2} + C \sum_{j=0}^{\infty} \lambda^{-j} \exp(-CC_N) \\ &\leqslant CN^{-t/2+1} C_N^{-t/2} + C \exp(-CC_N). \end{aligned} \tag{3.33}$$

For the last term of the right-hand side of (3.32), we can obtain the same bound. The proof of the rest is similar to that Theorem 2.1. This completes the proof of the theorem.

4. Some Remarks

From Theorem 2.1 and Theorem 2.2, it is easily seen that, under the restriction $C_n = o(N)$, the larger the magnitude of C_N, the better the detection is in the large sample cases. By the same way, if (1.1), (1.2), and (2.8) hold then the detection is weakly consistent.

Now we point out that, if (1.1), (1.2) hold and

$$\lim_{N \to \infty} C_N/N = 0 \qquad \text{and} \qquad \lim_{N \to \infty} C_N/\log\log N = \infty, \tag{4.1}$$

then $\hat{p}$ determined by (2.4) is a strongly consistent estimate of p_0. In fact, if $p < p_0$, then by (2.8), (3.7), and $\lim_{N \to \infty} C_N/N = 0$,

$$\lim_{N \to \infty} [\phi(p) - \phi(p_0)]/N \geqslant \log(\sigma^{*2}/\sigma^2) > 0.$$

It follows that, with probability one for N large,

$$\phi(p_0) < \phi(p), \qquad \text{for} \quad p < p_0. \tag{4.2}$$

Now we assume $p_0 < p \leqslant K$. Under the conditions (1.1) and (1.2), by the law of the iterated logarithm,

$$|\hat{q}_K(i, j) - \gamma(i-j)| = O\left(\sqrt{\frac{\log\log N}{N}}\right) \quad \text{a.s.,}$$

$$\left|\frac{1}{N} \sum_{n=K+1}^{N} e(n)\, X(n-i)\right| = O\left(\sqrt{\frac{\log\log N}{N}}\right) \quad \text{a.s.}$$

for $i, j = 0, 1, ..., K$. Thus, by (3.12),

$$0 \geqslant \hat{\sigma}_p^2 - \hat{\sigma}_{p_0}^2 = O\left(\frac{\log\log N}{N}\right) \quad \text{a.s.} \tag{4.3}$$

By (3.7), (3.12), and $\lim_{N \to \infty} C_N/\log\log N = \infty$, with probability one for large N,

$$\begin{aligned} \phi(p) - \phi(p_0) &\geqslant N \log \hat{\sigma}_K^2/\hat{\sigma}_{p_0}^2 + (p - p_0) C_N \\ &= N \log\{1 + (\hat{\sigma}_K^2 - \hat{\sigma}_{p_0}^2)/\hat{\sigma}_{p_0}^2\} + (p - p_0) C_N \\ &= O(\log\log N) + (p - p_0) C_N > 0, \quad p_0 < p \leqslant K. \end{aligned} \tag{4.4}$$

From (4.2) and (4.4), it follows that with probability one for N large,

$$\hat{p} = p_0. \tag{4.5}$$

This shows strong consistency of $\hat{p}$.

Note that for strong consistency of $\hat{p}$, the last condition of (4.1) can be weakened as

$$C_N \geqslant 2C \log\log N \qquad \text{with} \quad C > 1. \tag{4.6}$$

But this needs more accurate calculations.

REFERENCES

[1] AKAIKE, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21** 243–247.

[2] FUK, D. K. H., AND NAGAEV, S. V. (1971). Probability inequalities for sums of independent random variables. *Theory Probab. Appl.* **16**, No. 4, 643–660.

[3] HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

[4] HANNAN, E. J., AND QUINN, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B* **41** 190–195.

[5] HANNAN, E. J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8** 1071–1081.

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