The Annals of Statistics
1984, Vol. 12, No. 4, 1425–1433

ORDER SELECTION IN NONSTATIONARY AUTOREGRESSIVE MODELS

BY RUEY S. TSAY

Carnegie-Mellon University

In Hannan (1980), some limiting properties of the order selection criteria, AIC, BIC, and $\phi(p, q)$ for modeling stationary time series were derived. In this paper, we generalize these properties to the case in which the underlying process follows a nonstationary autoregressive model. We show that BIC and $\phi(p, 0)$ are weakly consistent. For the AIC, we prove that the asymptotic distribution given by Shibata (1976) for the stationary autoregressive models continues to hold.

1. Introduction. Consider a nonstationary or stationary autoregressive model of order p,

$$\Phi(B)X_t = a_t \tag{1.1}$$

with $\Phi(B) = U(B)\phi(B)$ where $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_p B^p$, $U(B) = 1 - U_1 B - \cdots - U_d B^d$, and $\phi(B) = 1 - \phi_1 B - \cdots - \Phi_{p-d} B^{p-d}$ are polynomials in B, B is the backshift operator such that $BX_t = X_{t-1}$, and $\{a_t\}$ is a process of independent and identically distributed continuous random variables with mean zero, variance σ^2 and finite fourth moment. We shall require that all the zeros of $U(B)$ are on and those of $\phi(B)$ are outside the unit circle. In mode (1.1), if $U(B) = 1$, the process X_t is stationary; otherwise, it is nonstationary. For the nonstationary case, we further assume that the process X_t starts at a finite time point t_0 with fixed initial values. Note that when $U(B) = (1 - B)^d$, model (1.1) reduces to the well-known autoregressive integrated moving average, ARIMA($p - d$, d, 0), model of Box and Jenkins (1976).

In the literature, two main approaches have been proposed to estimate the true order p of model (1.1). The first one is the so-called Box-Jenkins approach which employs the sample autocorrelation and partial autocorrelation functions as its identification statistics. Example references of this approach are Box and Jenkins (1976), Gray, Kelley and McIntire (1978), and Beguin, Gourieroux and Monfort (1980). Following this school, Tsay and Tiao (1984) recently proposed an extended sample autocorrelation function method to handle the nonstationary models. The second approach is the information criterion approach. Akaike (1969) is the first to advocate this information criterion procedure. Now there are several order selection criteria available. In particular, those in Akaike (1974, 1977), Rissanen (1978), Schwarz (1978), Hannan and Quinn (1979), Hannan (1980), and Hannan and Rissanen (1982) are of special interest. This approach has advantages of being objective and automatic. However, all the criteria

Received July 1983; revised June 1984.
AMS 1980 *subject classifications.* Primary 62M10; secondary 60F15.
Key words and phrases. AIC, autoregressive model, BIC, model identification, regression.

mentioned above were derived mainly for the stationary processes and their performance in modeling nonstationary series is relatively unknown. Ozaki (1975) perhaps is the first one to apply empirically Akaike's information criterion (AIC) to select possible orders for nonstationary series. The purpose of this paper, therefore, is to study the limiting properties of some order selection criteria when the underlying process follows a nonstationary autoregressive model.

The criteria considered are:

$$\mathrm{AIC}(k) = \log \hat{\sigma}_k^2 + 2k/n, \quad \mathrm{BIC}(k) = \log \hat{\sigma}_k^2 + k \log n/n$$

$$\phi(k, 0) = \log \hat{\sigma}_k^2 + ck \operatorname{loglog} n/n, \quad c > 2,$$

where n is the number of observations and $\hat{\sigma}_k^2$ is an estimate of σ^2 under the assumption that X_t is an AR(k) process. These are the three criteria considered in Hannan (1980). The AIC is the well-known Akaike's information criterion, and $\phi(k, 0)$ is proposed by Hannan and Quinn (1979). The BIC, on the other hand, is introduced independently by Schwarz (1978), Akaike (1977), and Rissanen (1978) using different derivations.

In this paper, we make use of some stochastic order properties of nonstationary AR models shown in Tiao and Tsay (1983). These properties are useful in handling nonstationary time series and will be discussed first. Then, we prove that the asymptotic distribution of AIC given in Shibata (1976) for the stationary AR models continues to hold in the nonstationary situation. For BIC and $\phi(k, 0)$ we show that they are weakly consistent in selecting the true order p of model (1.1). For simplicity, we assume that $p \le P$ with P being a known positive integer.

2. The main result. In this paper, $\hat{\sigma}_k^2$, an estimate of the variance σ^2 in (1.1), is computed in the following manner. For $k = 0$, $\hat{\sigma}_0^2$ is the sample variance of X_t, i.e.,

$$\hat{\sigma}_0^2 = \textstyle\sum_{t=1}^n (X_t - \bar{X})^2/(n - 1)$$

where $\bar{X} = n^{-1} \sum_1^n X_t$ is the sample mean of X_t. For a given positive integer k $(1 \le k \le P)$, $\hat{\sigma}_k^2$ is the least squares residual variance of the usual AR(k) regression, i.e.,

$$\hat{\sigma}_k^2 = \textstyle\sum_{t=k+1}^n \{X_t - \hat{\Phi}_{1(k)}X_{t-1} - \cdots - \hat{\Phi}_{k(k)}X_{t-k}\}^2/(n - k - 1) \tag{2.1}$$

where $\hat{\Phi}_{i(k)}$'s are the estimates of $\Phi_{i(k)}$'s obtained by minimizing the sum of squares of $e_{k,t}$ in the autoregression

$$X_t = \Phi_{1(k)}X_{t-1} + \cdots + \Phi_{k(k)}X_{t-k} + e_{k,t}, \quad t = k + 1, \cdots, n. \tag{2.2}$$

The existence of the above least squares estimates in the nonstationary case has been shown in Lemma 2.3 of Tiao and Tsay (1983) provided that n is sufficiently large.

For such estimates of σ^2, Shibata (1976) has derived the asymptotic distribution of AIC(k) under the assumptions that a_t is Gaussian and X_t is stationary.

The limiting probability is

$$\lim_{n\to\infty}\Pr\{\hat{p} = k\} = \pi(k - p, P - k), \quad \text{for} \quad 0 \le k \le P, \tag{2.3}$$

where $\pi(k - p, P - k) = 0$ if $k < p$ and it only depends on $k - p$ and $P - k$ if $p \le k \le P$. This result has been extended to a more general setting in Hannan (1980) where the Gaussian assumption is removed. For BIC and $\phi(k, 0)$, Hannan and Quinn (1979) have shown that they are strongly consistent in modeling stationary AR models.

Now, for the nonstationary AR models, we have the following result.

THEOREM 1. *Suppose that X_t is a nonstationary AR(p) process and satisfies the conditions of Section* 1, *and that σ^2 is estimated by the least squares method mentioned above. Then,* (i) *for AIC(k), the limiting probability* (2.3) *continues to hold, i.e.,*

$$\lim_{n\to\infty}\Pr\{\hat{p} = k\} = \pi(k - p, P - k), \quad \textit{for} \quad 0 \le k \le P.$$

(ii) *BIC(k) and $\phi(k, 0)$ are weakly consistent.*

To prove this theorem, we use some properties of the nonstationary ARMA models derived by Tiao and Tsay (1983). For a nonstationary AR process X_t in (1.1), let d be the true order of $U(B)$ and m the highest multiplicity of the roots of $U(B)$.

LEMMA 1. *Suppose that X_t is a nonstationary AR process and satisfies the conditions of Section* 1. *Then,*

(i) $\sum_{t=1}^{n} X_t^2 = O_p(n^{2m})$,
(ii) $(\sum_{t=1}^{n} X_t^2)^{-1} = O_p(n^{-2m})$,
(iii) $X_t^2 = O_p(t^{2m-1})$.

PROOF. The results (i) and (ii) are, respectively, Lemmas 2.4 and 2.6 of Tiao and Tsay (1983) while (iii) can be easily proved by using the same techniques as those of Lemma 2.4 mentioned above. □

Using this lemma, we consider the AR(k) regression of X_t in (2.2) in the following two cases.

CASE 1. $0 \le k \le d$.

In what follows, unless otherwise stated, the summation notation $\sum$ is summing over t from $d + 1$ to n.

LEMMA 2. *Suppose that X_t is a nonstationary AR process and satisfies the conditions of Section* 1. *Then,*

(i) $\sum \hat{e}_{k+1,t}^2 \le \sum \hat{e}_{k,t}^2$,

(ii) $\hat{\sigma}_k^2 = \sum \hat{e}_{k,t}^2/(n-d-1) + o_p(1)$,

where $\hat{e}_{j,t}$ ($j = k$ *or* $k+1$) *is the least squares residual of autoregression of order* j *in the form of* (2.2), *and* $\hat{\sigma}_k^2$ *is defined in* (2.1).

PROOF. Part (i) is a well-known result in the least squares theory. For Part (ii), we notice that the two quantities only differ in finitely many terms in the beginning of the summation and Lemma 1 shows that the effect of such difference is of order $O_p(n^{-1})$. □

Based on this lemma, it will be seen later that properties of the least squares estimates of the following two autoregressions are of importance:

$$X_t = \Phi_{1(d-1)}X_{t-1} + \cdots + \Phi_{d-1(d-1)}X_{t-d+1} + e_{d-1,t}, \quad t = d+1, \cdots, n \tag{2.4}$$

$$X_t = \Phi_{1(d)}X_{t-1} + \cdots + \Phi_{d(d)}X_{t-d} + e_{d,t}, \quad t = d+1, \cdots, n. \tag{2.5}$$

Given the nonstationary polynomial $U(B)$ of (1.1), we may factor it as

$$U(B) = \prod_{v=1}^{m} U_v(B) \tag{2.6}$$

where $U_v(B) = 1 - U_{1(v)}B - \cdots - U_{d_v(v)}B^{d_v}$ are polynomials in B of degrees d_v such that (a) $\sum_{v=1}^{m} d_v = d$, (b) $U_v(B)$ is a factor of $U_{v+1}(B)$, and (c) the multiplicity of any root of $U_v(B)$ is 1. Also, let us define

$$X_{1,t} = X_t,\ X_{j,t} = U_{j-1}(B)X_{j-1,t}, \quad \text{for} \quad j = 2, \cdots, m+1. \tag{2.7}$$

The factorization (2.6) and the transformation (2.7) have been found useful in deriving the stochastic orders of certain statistics of the nonstationary ARMA models, see Tiao and Tsay (1983, page 862) for details. Note that for $1 \le v \le m$, $m - v + 1$ is the highest multiplicity of the nonstationary roots, roots with modulus 1, of the series $X_{v,t}$ and $X_{m+1,t} = U(B)X_t$ is a stationary AR$(p-d)$ process.

From (2.7), the least squares autoregressions (2.4) and (2.5) can, respectively, be linearly transformed into

$$X_t = \sum_{v=1}^{m-1} \sum_{i=1}^{d_v} \eta_{i(v)}X_{v,t-i} + \sum_{i=1}^{d_m-1} \eta_{i(m)}X_{m,t-i} + e_{d-1,t} \tag{2.8}$$

$$X_t = \sum_{v=1}^{m-1} \sum_{i=1}^{d_v} \gamma_{i(v)}X_{v,t-i} + \sum_{i=1}^{d_m} \gamma_{i(m)}X_{m,t-i} + e_{d,t}. \tag{2.9}$$

Let $\mathbf{Y}_{d,t-1} = (X_{1,t-1}, \cdots, X_{1,t-d_1}, X_{2,t-1}, \cdots, X_{2,t-d_2}, \cdots, X_{m,t-1}, \cdots, X_{m,t-d_m})'$ *be the vector of regressors of* (2.9) *and* $\mathbf{\Gamma}_d$ be the corresponding vector of coefficients, i.e., $\mathbf{\Gamma}_d$ is defined analogously to $\mathbf{Y}_{d,t-1}$ with $\gamma_{i(v)}$ in the place of $X_{v,t-i}$. Also, let $\mathbf{Y}_{d-1,t-1}$ be the vector of regressors of (2.8) obtained by deleting the last element of $\mathbf{Y}_{d,t-1}$ and $\boldsymbol{\eta}_{d-1}$ the corresponding vector of coefficients. Then, the least squares normal equations of (2.8) and (2.9) can be written, respectively, as

$$\mathbf{A}_{11}\hat{\boldsymbol{\eta}}_{d-1} = \mathbf{B}_{11} \tag{2.10}$$

$$\mathbf{A}_{11}\hat{\mathbf{\Gamma}}_{d-1}^{(d)} + \mathbf{A}_{12}\hat{\gamma}_{d_m(m)} = \mathbf{B}_{11} \tag{2.11a}$$

$$\mathbf{A}_{21}\hat{\mathbf{\Gamma}}_{d-1}^{(d)} + \sum X_{m,t-d_m}^2 \hat{\gamma}_{d_m(m)} = \sum X_{m,t-d_m}X_t, \tag{2.11b}$$

where

$\mathbf{A}_{11} = \sum \mathbf{Y}_{d-1,t-1}\mathbf{Y}'_{d-1,t-1}$, $\mathbf{B}_{11} = \sum Y_{d-1,t-1}X_t$, $\mathbf{A}_{12} = \mathbf{A}'_{21} = \sum \mathbf{Y}_{d-1,t-1}X_{m,t-d_m}$, $\hat{\boldsymbol{\Gamma}}^{(d)}_{d-1}$ denotes the first $d-1$ elements of $\hat{\boldsymbol{\Gamma}}_d$, and the notation $\hat{\omega}$ of the parameter (or vector) ω denotes its least squares estimate.

From (2.10), (2.11) and the ordinary least squares theory, we have that

$$\hat{\gamma}_{d_m(m)}\left[\sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12}\right] = \sum X_{m,t-d_m}X_t - \mathbf{A}_{21}\hat{\boldsymbol{\eta}}_{d-1}, \tag{2.12}$$

$$\hat{\boldsymbol{\Gamma}}^{(d)}_{d-1} = \hat{\boldsymbol{\eta}}_{d-1} - (\mathbf{A}_{11})^{-1}\mathbf{A}_{12}\hat{\gamma}_{d_m(m)}, \tag{2.13}$$

$$\sum \hat{e}^2_{d-1,t} = \sum X^2_t - \mathbf{B}'_{11}\hat{\boldsymbol{\eta}}_{d-1} = \sum X^2_t - \mathbf{B}'_{11}(\mathbf{A}_{11})^{-1}\mathbf{B}_{11}, \tag{2.14}$$

$$\sum \hat{e}^2_{d,t} = \sum \hat{e}^2_{d-1,t} - \hat{\gamma}^2_{d_m(m)}\left[\sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12}\right]. \tag{2.15}$$

In the above, (2.12) is obtained by premultiplying $\mathbf{A}_{21}(\mathbf{A}_{11})^{-1}$ to (2.11a) and then subtracting the result from (2.11b), (2.13) follows from (2.10) and (2.11a), (2.14) can be easily obtained by using the property $\sum \hat{e}^2_{d-1,t} = \sum \hat{e}_{d-1,t}X_t$, and (2.15) follows from (2.12)–(2.14).

LEMMA 3. *Suppose that X_t is a nonstationary autoregressive process and satisfies the conditions of Section* 1. *Then, the least squares residual variances of* (2.4) *and* (2.5) *are related by*

$$\sum \hat{e}^2_{d,t}\Big/\sum \hat{e}^2_{d-1,t} = 1 - \hat{\gamma}^2_{d_m(m)} + o_p(1).$$

Clearly, this lemma is a generalization of equation (D.5) of Ramsey (1974, page 1297), see also Shibata (1976, page 119). Now, to prove the above lemma we only need to show, from (2.15), that

$$\left(\sum \hat{e}^2_{d-1,t}\right)^{-1}\left[\sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12}\right] = 1 + o_p(1). \tag{2.16}$$

To do this, we first establish the following two lemmas.

LEMMA 4. *Let $W_t = U_{d_m(m)}X_{m,t-d_m} + X_{m+1,t}$. Then, the sum of squares of the least squares residuals of the autoregression* (2.8) *can be written as*

$$\sum \hat{e}^2_{d-1,t} = \sum W^2_t - \mathbf{R}_{21}(\mathbf{A}_{11})^{-1}\mathbf{R}_{12},$$

where $\mathbf{A}_{11}$ is defined in (2.10), *and $\mathbf{R}'_{21} = \mathbf{R}_{12} = \sum \mathbf{Y}_{d-1,t-1}W_t$ with $\mathbf{Y}_{d-1,t-1}$ also defined in* (2.10).

PROOF. From the transformation (2.7), we have that

$$\begin{aligned} X_t &= \textstyle\sum_{v=1}^{m}\sum_{i=1}^{d_v} U_{i(v)}X_{v,t-i} + X_{m+1,t}, \\ &= \textstyle\sum_{v=1}^{m-1}\sum_{i=1}^{d_v} U_{i(v)}X_{v,t-i} + \sum_{i=1}^{d_m-1} U_{i(m)}X_{m,t-i} + W_t. \end{aligned} \tag{2.17}$$

Substituting (2.17) into $\mathbf{B}_{11} = \sum \mathbf{Y}_{d-1,t-1}X_t$, one can easily obtain that

$$(\mathbf{A}_{11})^{-1}\mathbf{B}_{11} = \mathbf{U}_{d-1} + (\mathbf{A}_{11})^{-1}\mathbf{R}_{12}, \tag{2.18}$$

where $\mathbf{U}_{d-1}$ is defined analogously to $\boldsymbol{\eta}_{d-1}$ of (2.10) with the true coefficient $U_{i(v)}$

in the place of $\eta_{i(v)}$. In (2.18), we use the property that for a nonsingular $(d-1) \times (d-1)$ matrix $\mathbf{H}$, $\mathbf{H}^{-1}\mathbf{H}_i = \zeta_i$ where $\mathbf{H}_i$ is the ith column of $\mathbf{H}$ and ζ_i is the ith unit vector in the $(d-1)$-dimensional Euclidean space.

By (2.14) and using (2.18) and (2.17) twice, we complete the proof. □

LEMMA 5. *Suppose that X_t is a nonstationary AR process and satisfies the conditions of Section* 1. *Also, h is a fixed integer. Then,*

(i) $\sum X_{v,t}X_{m+1,t+h} = O_p(n^{m-v+1})$, for $1 \le v \le m$,
(ii) $\sum X_{v,t}X_{j,t+h} = O_p(n^{2m-v-j+2})$ for $1 \le v, j \le m$,
(iii) $[\sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12}]^{-1} = O_p(n^{-2})$.

PROOF. Parts (i) and (ii) are Lemma 2.5 of Tiao and Tsay (1983). For Part (iii), we note that the stochastic orders of the inverse of the $\mathbf{X}'\mathbf{X}$-matrix, $\mathbf{A} = \sum \mathbf{Y}_{d,t-1}\mathbf{Y}'_{d,t-1}$, of the autoregression (2.9) can readily be obtained from the results of Tiao and Tsay (1983), see for example (3.8), (3.9), and Lemmas 2.7 and 3.2 there. In particular, the stochastic order of the (d, d)-th element of $\mathbf{A}^{-1}$ is $O_p(n^{-2})$. On the other hand, from the inverse identity of a symmetric matrix

$$\begin{bmatrix} \mathbf{M} & \mathbf{N} \\ \mathbf{N}' & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{M}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{bmatrix}$$

where $\mathbf{E} = \mathbf{D} - \mathbf{N}'\mathbf{M}^{-1}\mathbf{N}$, $\mathbf{F} = \mathbf{M}^{-1}\mathbf{N}$, one can see that the (d, d)-th element of $\mathbf{A}^{-1}$ is $[\sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12}]^{-1}$. This completes the proof of Part (iii). □

PROOF OF LEMMA 3. Since $U_{d_m(m)}$ is either 1 or -1, we have, from Lemma 4, that

$$\begin{aligned} \sum \hat{e}^2_{d-1,t} = {} & \sum X^2_{m,t-d_m} + 2U_{d_m(m)} \sum X_{m,t-d_m}X_{m+1,t} + \sum X^2_{m+1,t} \\ & - (U_{d_m(m)}\mathbf{A}_{21} + \mathbf{A}^*_{21})(\mathbf{A}_{11})^{-1}(U_{d_m(m)}\mathbf{A}_{12} + \mathbf{A}^*_{12}) \end{aligned} \tag{2.19}$$

where $(\mathbf{A}^*_{21})' = \mathbf{A}^*_{12} = \sum \mathbf{Y}_{d-1,t-1}X_{m+1,t}$. Next, we note that the stochastic orders of the elements of $\mathbf{A}_{11}^{-1}$ can readily be obtained from (a) the order of the determinant of $\mathbf{A}^{-1}$ (Lemmas 2.7 and 3.2 of Tiao and Tsay, 1983), (b) orders of the elements of $\mathbf{A}$ (Lemma 5(ii) and Lemma 1(i) of the present paper), (c) the result $0 < \sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12} < \sum X^2_{m,t-d_m} = O_p(n^2)$ (the inequalities hold because $\mathbf{A}^{-1}$ and $\mathbf{A}_{11}^{-1}$ are positive definite which in turn follow from the positive definiteness of $\mathbf{A}$, Lemma 2.3 of Tiao and Tsay; 1983), and (d) the relationship of determinants

$$\det[\mathbf{A}] = \det[\mathbf{A}_{11}][\sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12}],$$

where $\det[\mathbf{A}]$ denotes the determinant of the matrix $\mathbf{A}$. More specifically, (a), (c) and (d) provide the stochastic order of $\det^{-1}[\mathbf{A}_{11}]$ while (b) can be used to obtain the orders of the cofactors of the elements $\mathbf{A}_{11}$. Using these stochastic orders and Lemma 5(i), one can easily show that

$$(\mathbf{A}_{21} + \mathbf{A}^*_{21})(\mathbf{A}_{11})^{-1}(\mathbf{A}_{12} + \mathbf{A}^*_{12}) = \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12} + O_p(n). \tag{2.20}$$

Moreover, since $X_{m+1,t}$ is stationary, we have $\sum X^2_{m+1,t} = O_p(n)$ and, from Lemma

5(i), $\sum X_{m,t-d_m} X_{m+1,t} = O_p(n)$. By these two order properties and (2.20), (2.19) can be simplified as

$$\sum \hat{e}^2_{d-1,t} = \sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12} + O_p(n). \tag{2.21}$$

From (2.21) and Lemma 5(iii), we have that

$$[\sum X^2_{m,t-d_m} - \mathbf{A}_{21}(\mathbf{A}_{11})^{-1}\mathbf{A}_{12}]^{-1} \sum \hat{e}^2_{d-1,t} = 1 + O_p(n^{-1}) = 1 + o_p(1). \tag{2.22}$$

Lemma 3 then follows directly from (2.22) and (2.16). □

CASE 2. $d \le k \le P$.

We now consider the AR(k) regression of X_t for $P \ge k \ge d$. It is shown in Section 4 of Tiao and Tsay (1983) that the least squares estimates of the AR(k) regression in (2.2) when $k \ge d$ can asymptotically be obtained from those of (2.5) and the following autoregression:

$$Z_t = \beta_{1(k-d)} Z_{t-1} + \cdots + \beta_{k-d(k-d)} Z_{t-k+d} + h_t, \quad t = k+1, \cdots, n \tag{2.23}$$

where $Z_t = U(B)X_t$. More precisely, the least squares estimates involved are related by

$$(1 - \textstyle\sum_{i=1}^{k} \hat{\Phi}_{i(k)} B^i) = (1 - \sum_{i=1}^{d} \hat{\Phi}_{i(d)} B^i)(1 - \sum_{h=1}^{k-d} \hat{\beta}_{h(k-d)} B^h) + o_p(1). \tag{2.24}$$

A detailed proof of (2.24) can be found in Lemma 4.1 and equation (4.19) of Tiao and Tsay (1983). Also, for the autoregression (2.5) or (2.9), Theorem 3.1 of the above paper shows that

$$\hat{\gamma}^2_{d_m(m)} = 1 + o_p(1). \tag{2.25}$$

Here, again, we make use of the fact that $U_{d_m(m)}$ is either 1 or -1.

Next, let $\hat{\xi}^2_{k-d}$ be the least squares residual variance of the autoregression (2.23) if $k > d$, and be the sample variance of the stationary process Z_t when $k = d$.

LEMMA 6. *Suppose that X_t is a nonstationary AR process and satisfies the conditions in Section* 1. *Then, for any integer $k \ge d$, the least squares estimates and residual variance of AR(k) regression in the form of* (2.2) *satisfy*

(i) $\hat{\Phi}_{k(k)} = \hat{\beta}_{k-d(k-d)} + O_p(n^{-1}) = \hat{\beta}_{k-d(k-d)} + o_p(1)$,
(ii) $\hat{\sigma}^2_k = \hat{\xi}^2_{k-d} + O_p(n^{-.5})$,

where $\hat{\beta}_{0(0)} = 1$ *or* -1.

PROOF. The result (i) follows directly from (2.24). (ii) can be proved by the linear transformation (2.7) and (2.24), see Lemma 6.1 of Tsay and Tiao (1984) for details.

PROOF OF THEOREM 1. Let us start with Part (i). First, consider $\hat{p} = k$ with $P \ge k \ge d$. In this case, Lemma 6(ii) and Lemma 2(ii) show that

$$\hat{\sigma}^2_{k+1}/\hat{\sigma}^2_k = \hat{\xi}^2_{k+1-d}/\hat{\xi}^2_{k-d} + O_p(n^{-.5}).$$

Using this result and Theorem 1 of Shibata (1976), we have

$$\lim_{n\to\infty}\Pr\{\hat{p} = k\} = \pi(k-p, P-k), \quad \text{if} \quad d \le k \le P. \tag{2.26}$$

Next, consider $\hat{p} = d - 1$. From Lemma 3 and Lemma 2(ii),

$$\hat{\sigma}_d^2/\hat{\sigma}_{d-1}^2 = 1 - \hat{\gamma}^2_{d_m(m)} + o_p(1).$$

It is then clear, by (2.25), that

$$n \log(\hat{\sigma}_{d-1}^2/\hat{\sigma}_d^2) \to_P \infty, \tag{2.27}$$

which in turn implies that

$$\lim_{n\to\infty}\Pr\{n \log(\hat{\sigma}_{d-1}^2/\hat{\sigma}_d^2) \le 2\} = 0.$$

In (2.27), $\to_P$ denotes convergence in probability. Therefore, for the AIC criterion

$$\lim_{n\to\infty}\Pr\{\hat{p} = d-1\} = 0.$$

Finally, for $\hat{p} = k$ with $o \le k < d - 1$, we have, by Lemma 2, that

$$\hat{\sigma}_k^2 \ge \hat{\sigma}_{d-1}^2 + o_p(1).$$

Hence, by (2.27),

$$\lim_{n\to\infty}\Pr\{\hat{p} = k\} = 0 \quad \text{for} \quad 0 \le k < d-1.$$

This completes the proof of Theorem 1(i). On the other hand, Theorem 1(ii) follows immediately from Lemma 6(i) and the following result of Hannan and Quinn (1979).

LEMMA 7. *Suppose that Z_t is a stationary $AR(p)$ process and satisfies the conditions of Section* 1. *Then, there is an integer N,* $\Pr(N < \infty) = 1$, *such that for $n > N$,* $\log\{1 - \hat{\beta}^2_{v(v)}\} + 2c \log\log n/n > 0$, *almost surely, $p < v \le P$. Of course, here $\hat{\beta}_{v(v)}$ is the least squares estimate of $\beta_{v(v)}$ in the autoregression* (2.23) *with $v = k - d$.*

3. Concluding remarks. In this paper, we generalized the order selection criteria AIC, BIC and $\phi(k, 0)$ to the nonstationary autoregressive models. One obvious advantage of such generalizations is that they eliminate the need to determine the order of differencing in practical time series modeling. We remark here that Theorem 1(i) has also been investigated by Yajima (1982) for the special case where $U(B) = (1 - B)^d$ and a_t is Gaussian.

Acknowledgment. The author sincerely wishes to thank the Associate Editor for many constructive suggestions and his colleague Mark Schervish for helpful comments. These suggestions and comments have improved the presentation of this paper.

REFERENCES

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

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **AC-19** 716–723.

AKAIKE, H. (1977). On entropy maximization principle. In *Applications of Statistics* 27–41. Ed. P. R. Krishnaiah. North-Holland, Amsterdam.

BEGUIN, J. M., GOURIEROUX, C. and MONFORT, A. (1980). Identification of a mixed autoregressive moving average process: the corner method. *Time Series* 423–435. Ed. O. D. Anderson. North-Holland, Amsterdam.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

GRAY, H. L., KELLEY, G. D. and MCINTIRE, D. D. (1978). A new approach to ARMA modeling. *Commun. Statist.* **B7** 1–77.

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

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

HANNAN, E. J. and RISSANNEN, J. (1982, 1983). Recursive estimation of mixed autoregressive moving average order. *Biometrika* **69** 81–94, **70** 303.

OZAKI, T. (1975). On ARIMA model identification. Proceeding of the 8-th Hawaii International Conference on System Sciences, Western Periodical Company, 221–223.

RAMSEY, F. L. (1974). Characterization of the partial autocorrelation function. *Ann. Statist.* **2** 1296–1301.

RISSANEN, J. (1978). Modeling by shortest data description. *Automatica* **14** 465–471.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

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

TIAO, G. C. and TSAY, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Statist.* **11** 856–871.

TSAY, R. S. and TIAO, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *J. Amer. Statist. Assoc.* **79** 84–96.

YAJIMA, Y. (1982). Estimation of the number of differencing of an ARIMA process by an AR model fitting. Research Report, Department of Information Sciences, Tokyo Institute of Technology.

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
SCHENLEY PARK
PITTSBURGH, PENNSYLVANIA 15213