Statistica Sinica **7**(1997), 407-423

ASYMPTOTIC EFFICIENCY OF THE ORDER SELECTION OF A NONGAUSSIAN AR PROCESS

Alexandros Karagrigoriou

University of Cyprus

Abstract: Motivated by Shibata's (1980) asymptotic efficiency results for the order selected for a zero mean Gaussian AR process this paper establishes the asymptotic efficiency of AIC-like model selection criteria for infinite order autoregressive processes with zero mean and unobservable errors that constitute a sequence of nongaussian random variables. Furthermore, from the spectral density point of view, the asympotic efficiency of AIC-like information criteria is established when the underlying process is an infinite order nonzero mean nongaussian AR process.

Key words and phrases: AR processes, asymptotic efficiency, Brillinger's mixing condition, model selection criteria, spectral density.

1. Introduction

The concept of asymptotic efficiency which was proposed by Shibata (1980), deals with the selection of a finite approximation to an infinite order process for which the average mean squared error of prediction is the smallest possible. Extensive work on this subject shows that AIC-like criteria (AIC: Akaike (1973), FPE: Akaike (1970), $S_n(k)$: Shibata (1980), CAT: Bhansali (1986b), Parzen (1974), AIC_C: Hurvich and Tsai (1989)) which have been proved to be inconsistent (see Shibata (1976), Hannan (1980), Bhansali (1986b)) possess this asymptotically optimal property (see Shibata (1980) and Bhansali (1986a) for zero mean processes, Hurvich and Tsai (1989) for regression and zero mean Gaussian AR processes, and Karagrigoriou (1995) for nonzero mean Gaussian processes). It should be noted here that the concept of asymptotic efficiency is closely associated not with an estimate of the order of the underlying process but rather with a finite approximation to the truly infinite order of the process. As a result, it is not a surprise that consistent criteria like BIC: Schwarz (1978) and ϕ: Hannan and Quinn (1979) fail to be asymptotically optimal.

Shibata (1981) also investigated the asymptotic efficiency of the spectral density estimate of a finite order AR model fitted to an infinite order Gaussian AR process. Taniguchi (1980) considered the nongaussian case and proved simultaneously the asymptotic efficiency of the order selected by AIC and that of a spectral density estimate by fitting an ARMA(p, q) model. The quasi-Gaussian maximum likelihood estimator of the vector of the unknown parameters was used and an asymptotic lower bound for the mean squared error of prediction was evaluated using Kolmogorov-Wiener theory. Finally, it was proved that this bound was attained if the order of the process was selected by AIC. Taniguchi's approach, although similar, is not based on the idea of the integrated relative squared error used by Shibata (1981).

The nongaussian case was also investigated by Härdle (1987), who established the efficient selection of regression variables when the unknown true error distribution is replaced by a likelihood function ρ which is believed to be close to the true one. Efficient selection was obtained using a special criterion which depends on the second moments of the true error distribution and is equivalent to the Mallows' (1973) C_p and AIC criteria.

This work is devoted to the asymptotic efficiency of the order of a zero mean nongaussian AR process as well as to the asymptotic efficiency of the spectral density estimate of a nonzero mean AR process. In Section 2, the definition of asymptotic efficiency, some preliminary results, and the assumptions required for asymptotic efficiency will be stated. In Section 3, the asymptotic efficiency of the order selected by the FPE, FPE_α: Bhansali and Downham (1977), AIC, $S_n(k)$ and AIC_C criteria will be established. Finally, in Section 4 the asymptotic efficiency of the spectral density estimate of a finite order AR model fitted to a nonzero mean nongaussian infinite order AR process will be shown.

2. Assumptions and Preliminary Results

Consider the linear zero mean infinite order AR process $\{X_t\}$ of the form:

$$X_{t+1} + a_1 X_t + a_2 X_{t-1} + \cdots = \varepsilon_{t+1}, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{2.1}$$

where $a_1, a_2, \ldots$ are real numbers and $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and variance σ^2. Let $r_{ij} = E(X_{t+i}X_{t+j})$ be the autocovariance and $R = (r_{ij}, 1 \le i, j < \infty)$ the infinite dimensional covariance matrix of $\{X_t\}$. Let $V = \{\mathbf{b} : \mathbf{b}' = (b_1, b_2, \ldots), \|\mathbf{b}\|_R < \infty\}$ be the vector space with norm

$$\|\mathbf{b}\|_R = \Big(\sum_{i,j=1}^{\infty} b_i b_j r_{ij}\Big)^{\frac{1}{2}}.$$

Also let $\|t\| = \|t\|_I$, for all t in V, where I is the infinite dimensional identity matrix and $\|R\|^2 = \sup_{t:\|t\|\le 1} \|Rt\|^2$.

Consider now the projection $\mathbf{a}(k) = (a_1(k), \ldots, a_k(k), 0, 0, \ldots)'$ of the parameter $\mathbf{a} = (a_1, a_2, \ldots)'$ with respect to the norm $\|\cdot\|_R$ on the k dimensional subspace $V(k) = \{\mathbf{c} : \mathbf{c}' = (c_1, \ldots, c_k, 0, 0, \ldots)\}$ and let $\gamma = \mathbf{a}(k) - \mathbf{a}$. The spectral density of the process $\{X_t\}$ is given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{|A(e^{\mathrm{i}\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi, \tag{2.2}$$

where $A(e^{\mathrm{i}\lambda}) = \sum_{j=0}^{\infty} a_j e^{\mathrm{i}j\lambda}$, $a_0 = 1$.

Given observations $x_1, \ldots, x_n$, if the kth order AR model, with $k \times k$ covariance matrix $R(k)$ and innovation variance σ_k^2,

$$X_{t+1} + a_1(k)X_t + a_2(k)X_{t-1} + \cdots + a_k(k)X_{t+1-k} = \varepsilon_{t+1,k}, \quad a_0(k) = 1 \tag{2.3}$$

is assumed, the least squares estimator $\hat{\mathbf{a}}(k) = (\hat{a}_1(k), \ldots, \hat{a}_k(k), 0, 0, \ldots)'$ of $\mathbf{a}(k)$ and the estimator

$$\hat{\sigma}_k^2 = \sum_{t=K_n}^{n-1} (X_{t+1} + \hat{a}_1(k)X_t + \cdots + \hat{a}_k(k)X_{t+1-k})^2/N$$

of σ_k^2 are obtained, where $N = n - K_n$ and K_n is a preassigned upper bound for the selected order. Let $\{Y_t\}$ be another realization of the process independent of, but with the same probabilistic structure as $\{X_t\}$. Then, an order $\hat{k}$ selected by any selection procedure is called asymptotically efficient if (Shibata (1980))

$$Q_n(\hat{k}) \Big/ L_n(k_n^\star) \xrightarrow{P} 1, \qquad \text{as} \qquad n \to \infty,$$

where $Q_n(k)$ is the penalty function defined by

$$Q_n(k) = E[(\hat{Y}_{t+1} - Y_{t+1})^2 | X_1, \ldots, X_n] - \sigma^2,$$

$L_n(k)$ the expectation of $Q_n(k)$, $\hat{Y}_{t+1}$ the 1-step ahead predictor of Y_{t+1} given by $\hat{Y}_{t+1} = -\sum_{i=1}^{k} \hat{a}_i(k) Y_{t+1-i}$, and $\{k_n^*\}_{n=1}^{\infty}$ a sequence of positive integers at each of which the minimum of $L_n(k)$ with respect to k $(1 \leq k \leq K_n)$ is attained. Finally, a spectral density estimate of $f(\lambda)$ is given by

$$\hat{f}_k(\lambda) = \frac{\hat{\sigma}_k^2}{2\pi} \cdot \frac{1}{|\hat{A}_k(e^{i\lambda})|^2}, \tag{2.4}$$

where $\hat{A}_k(e^{i\lambda}) = 1 + \hat{a}_1(k)e^{1i\lambda} + \cdots + \hat{a}_k(k)e^{ki\lambda}$. Then, the estimated covariance matrix is $\overline{R} = (\overline{r}_{l-m}, 1 \leq l, m < \infty)$ where $\overline{r}_l = \int_{-\pi}^{+\pi} e^{il\lambda} \hat{f}_k(\lambda) d\lambda$. For the goodness of fit of $\hat{f}_k$ we use the integrated relative squared error, namely

$$\widetilde{Q}_n(k) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \Big[\frac{\hat{f}_k(\lambda) - f(\lambda)}{f(\lambda)}\Big]^2 d\lambda. \tag{2.5}$$

From the point of view of the autoregressive spectral density, an order $\hat{k}$ is called asymptotically efficient if

$$\frac{\widetilde{Q}_n(\hat{k})}{L_n(k_n^*)} \xrightarrow{P} 2/\sigma^2, \qquad n \to \infty.$$

Assumptions

A0. $\{\varepsilon_t\}$ has finite moments up to the 16th order: $E\varepsilon_i = 0, \forall i$, $E(\varepsilon_i \varepsilon_j) = \delta_{i-j}\sigma^2$, where $\delta_k = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{otherwise} \end{cases}$ and

$$E|\varepsilon_1|^{16} < \infty. \tag{2.6}$$

A1. $\{X_t\}$ is a stationary not degenerate to a finite order autoregressive process.

A2. The associated power series $A(z) = \sum_{j=0}^{\infty} a_j z^j$ converges and is nonzero for $|z| \leq 1$.

A3. K_n is a sequence of positive integers such that $K_n \to \infty$ and $K_n/n^{\frac{1}{2}} \to 0$, as $n \to \infty$.

A4. $\sum_{j=0}^{\infty} j|a_j| < \infty$.

Note that (2.6) implies that $E|\prod_{j=1}^{\nu} \varepsilon_j| < \infty$, $\nu \leq 16$. Note also that assumptions **A1**$-$**A3** coincide with the assumptions required for the asymptotic efficiency in the nonzero mean Gaussian AR(∞) case (Karagrigoriou (1992) and (1995)). Assumptions **A2** and **A4** imply that $|A(e^{\mathrm{i}\lambda})|$ is bounded above and bounded away from zero. Therefore $A(e^{\mathrm{i}\lambda})$ and consequently $f(\lambda)$ are continuous for $-\pi \leq \lambda \leq \pi$. As a result, two constants Γ_1 and Γ_2 can be found such that $0 < \Gamma_1 < f(\lambda) < \Gamma_2$, $-\pi \leq \lambda \leq \pi$. If $\lambda_1 < \cdots < \lambda_p$ $(p \leq k)$ are the distinct eigenvalues of $R(k)$, then (see Berk 1974) $2\pi\Gamma_1 \leq \lambda_1 < \cdots < \lambda_p \leq 2\pi\Gamma_2$. Since the norm of $R(k)$ is dominated by the largest modulus of the eigenvalues of $R(k)$, we have $\|R(k)\| \leq 2\pi\Gamma_2$ and $\|R^{-1}(k)\| \leq (2\pi\Gamma_1)^{-1}$, where $R(k)^{-1}$ is the inverse of $R(k)$. Assumption **A4** implies that

$$\sum_{j=0}^{\infty} |a_j| < \infty \tag{2.7}$$

and therefore the expectation of $\{X_t\}$ is finite and in particular is equal to 0 since $E(\varepsilon_i) = 0$, for all i. Note that (2.7) is the condition imposed on the parameters of the process for the Gaussian case (Shibata (1980)). The cost to be paid for dropping the Gaussian assumption is the stronger condition **A4**.

Under assumptions **A0**$-$**A2** and condition (2.7), the penalty function $Q_n(k)$ in this case, can be easily derived from Lemma 2.1 of Karagrigoriou (1995):

$$Q_n(k) = \|\mathbf{a}(k) - \hat{\mathbf{a}}(k)\|_R^2 + \|\gamma\|_R^2.$$

As for the quantity asymptotically equivalent to the expectation of $Q_n(k)$, it is given by Shibata (1980) and Karagrigoriou (1992) $L_n(k) = k\sigma^2 N^{-1} + \|\gamma\|_R^2$. Since $\{X_t\}$ is a zero-mean AR(∞) process it has a moving average representation of the form:

$$X_{t+1} = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t+1-j}, \quad \beta_0 = 1. \tag{2.8}$$

The following lemma provides the relationship between the a_i's and the β_i's as well as their connection to the so called "Brillinger's mixing condition".

Lemma 2.1. *The following two statements are equivalent*

$$\text{(i)} \quad \sum_{j=0}^{\infty} j|a_j| < \infty, \qquad\qquad \text{(ii)} \quad \sum_{j=0}^{\infty} j|\beta_j| < \infty, \tag{2.9}$$

where the a_j's are defined in (2.1) *and the β_j's in* (2.8).

In addition, if assumptions **A0**–**A2** *are satisfied, then assumption* **A4** (*and consequently* (2.9 ii)) *implies Brillinger's mixing condition, namely*

$$\sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} \cdots \sum_{t_{k-1}=-\infty}^{\infty} |t_j cum_k| < \infty, \quad j = 1, \ldots, k-1, \quad k = 2, \ldots, 16, \tag{2.10}$$

where $cum_k = cum(X_{t_1+t}, X_{t_2+t}, \ldots, X_{t_{k-1}+t}, X_t)$ and $cum(X_1, \ldots, X_n)$ represents the joint cumulant of order n of $(X_1, \ldots, X_n)$ (see Rosenblatt (1983)).

Proof. For the proof of the first part refer to Hannan and Kavalieris (1986). To prove the second part of the Lemma we make use of the moving average representation of the process $\{X_t\}$ given in (2.8), so that

$$\sum_{t_1,\ldots,t_{k-1}} |t_j cum_k| = \sum_{l_1,\ldots,l_{k-1},l=0}^{\infty} \Big(\prod_{j=1}^{k-1} \beta_{l_j}\Big) \beta_l cum(\varepsilon_{t+t_1-l_1}, \ldots, \varepsilon_{t+t_{k-1}-l_{k-1}}, \varepsilon_{t-l}).$$

Due to independence, the cumulants are equal to zero unless all subscripts are equal. Furthermore, since by **A0** the moments up to the 16th order are finite the above quantity is bounded by $C\sum_{l=0}^{\infty} \beta_l \beta_{t_1+l}, \ldots, \beta_{t_{k-1}+l}$, for some constant C. Therefore,

$$\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty} |t_j cum_k| \le C \sum_{l=0}^{\infty} |\beta_l| \sum_{t_1} |\beta_{t_1+l}| \cdots \sum_{t_{k-1}} |\beta_{t_{k-1}+l}||t_j|,$$

where $j = 1, \ldots, k-1$ and $k = 2, \ldots, 16$. Since (2.9 ii) implies that $\sum_j |\beta_j| < \infty$, the mixing condition is verified.

Brillinger (1969), using the mixing condition given in (2.10), proved that

$$cum(m_{\eta_1\theta_1}(u_1), \ldots, m_{\eta_s\theta_s}(u_s)) = O(N^{-s+1}), \tag{2.11}$$

where $N = n - K_n$, $u_i = 0, \pm 1, \pm 2, \ldots$, $\eta_i, \theta_i = 1, \ldots, k$, $i = 1, \ldots, s$, $s = 1, 2, 3, \ldots$, $m_{\eta_i \theta_i}(u)$ is the η_ith$\times\theta_i$th entry of the matrix

$$N^{-1} \cdot \sum_{K_n \le t \le n-1} \mathbf{X}_{t+u} \mathbf{X}_t^{'}, u = 0, \pm 1 \pm 2, \ldots,$$

and $\mathbf{X}_t = (X_t, X_{t-1}, \ldots, X_{t+1-k})'$.

3. Asymptotic Efficiency of the Order Selection for Prediction

Lemma 3.1. *Under assumptions* **A0**$-$**A2** *and* **A4** *for the zero mean process* (2.1) *the following expressions hold:*

$$E\|N^{-1} \sum_{t=K_n}^{n-1} \mathbf{X}_t \varepsilon_{t+1}\|^2_{R^{-1}(k)} = N^{-1}\sigma^2 k,$$

$$E\|N^{-1} \sum_{t=K_n}^{n-1} \mathbf{X}_t \varepsilon_{t+1}\|^m_{R^{-1}(k)} = N^{-\frac{m}{2}} \sigma^m \prod_{j=0}^{\frac{m}{2}-1} (k+2j) + O(N^{-\frac{m}{2}-1}) k^{\frac{m}{2}},\ m = 4, 6, 8.$$

Proof.

$$E\|N^{-1} \sum_{t=K_n}^{n-1} \mathbf{X}_t \varepsilon_{t+1}\|^2_{R^{-1}(k)} = N^{-2} E\Big\{ \sum_{t_1=K_n}^{n-1} \mathbf{X}'_{t_1} \varepsilon_{t_1+1} R^{-1}(k) \sum_{t_2=K_n}^{n-1} \mathbf{X}_{t_2} \varepsilon_{t_2+1} \Big\}.$$

Note that due to the independence of the ε_t's the above expectation is equal to zero unless $t_1 = t_2$. Therefore, the above quantity is equal to

$$N^{-2} E\Big\{ \sum_{t=K_n}^{n-1} \mathbf{X}'_t \varepsilon_{t+1} R^{-1}(k) \mathbf{X}_t \varepsilon_{t+1} \Big\} = N^{-2} \sum_{t=K_n}^{n-1} E\{\mathbf{X}'_t R^{-1}(k) \mathbf{X}_t \varepsilon^2_{t+1}\}.$$

Since ε_t is independent of X_τ for every $\tau < t$ and $E\varepsilon_t^2 = \sigma^2$ for all t, the above expression becomes:

$$\begin{aligned}
\sigma^2 N^{-2} \sum_{t=K_n}^{n-1} E\{\mathbf{X}'_t R^{-1}(k) \mathbf{X}_t\} &= \sigma^2 N^{-2} \sum_{t=K_n}^{n-1} \mathrm{tr}(R^{-1}(k) E\{\mathbf{X}_t \mathbf{X}'_t\}) \\
&= \sigma^2 N^{-2} \sum_{t=K_n}^{n-1} \mathrm{tr}(R^{-1}(k) R(k)) = \sigma^2 N^{-2} N k,
\end{aligned}$$

where tr$A$ is the trace of the matrix A. Let r^{ij} be the ith$\times j$th entry of $R^{-1}(k)$. Since $\varepsilon_{t+1} = \sum_{j=0}^{\infty} a_j X_{t+1-j}$, the term $E\| \sum \mathbf{X_t} \varepsilon_{t+1}/N\|^4_{R^{-1}(k)}$ is equal to

$$\sum_{l_1,l_2,l_3,l_4=1}^{k} r^{l_1 l_2} r^{l_3 l_4} \sum_{j_1,j_2,j_3,j_4=0}^{\infty} a_{j_1} \cdots a_{j_4} \times E\Big\{ \sum_{t=K_n}^{n-1} \frac{X_{t+1-l_1} X_{t+1-j_1}}{N}$$

$$\times \sum_{t=K_n}^{n-1} \frac{X_{t+1-l_2}X_{t+1-j_2}}{N} \sum_{t=K_n}^{n-1} \frac{X_{t+1-l_3}X_{t+1-j_3}}{N} \sum_{t=K_n}^{n-1} \frac{X_{t+1-l_4}X_{t+1-j_4}}{N} \Big\}$$

$$= \sum_{l_1,\ldots,l_4=1}^{k} r^{l_1 l_2} r^{l_3 l_4} \sum_{j_1,\ldots,j_4=0}^{\infty} a_{j_1}\cdots a_{j_4} E\Big\{ \prod_{i=1}^{4} m_{1,l_i}(j_i) \Big\}$$

$$= \sum_{l_1,\ldots,l_4=1}^{k} r^{l_1 l_2} r^{l_3 l_4} \sum_{j_1,\ldots,j_4=0}^{\infty} a_{j_1}\cdots a_{j_4} \sum_{v} \Big\{ \prod_{h=1}^{p} cum(m_{1,l_i}(j_i)(i \epsilon v_h)) \Big\},$$

where the $\sum$ is over all partitions of $(v_1, \ldots, v_p)$, $p = 1,2,3,4$ of the integers $1,2,3,4$. By (2.11) the main terms are those involving the cumulants of order 2 i.e.:

$$\begin{aligned} &\sum r^{l_1 l_2} r^{l_3 l_4} \times \\ &\Big[E\frac{(\sum X_{t+1-l_1}\varepsilon_{t+1})(\sum X_{t+1-l_2}\varepsilon_{t+1})}{N^2} E\frac{(\sum X_{t+1-l_3}\varepsilon_{t+1})(\sum X_{t+1-l_4}\varepsilon_{t+1})}{N^2} \\ &\quad + E\frac{(\sum X_{t+1-l_1}\varepsilon_{t+1})(\sum X_{t+1-l_3}\varepsilon_{t+1})}{N^2} E\frac{(\sum X_{t+1-l_2}\varepsilon_{t+1})(\sum X_{t+1-l_4}\varepsilon_{t+1})}{N^2} \\ &\quad + E\frac{(\sum X_{t+1-l_1}\varepsilon_{t+1})(\sum X_{t+1-l_4}\varepsilon_{t+1})}{N^2} E\frac{(\sum X_{t+1-l_2}\varepsilon_{t+1})(\sum X_{t+1-l_3}\varepsilon_{t+1})}{N^2} \Big] \\ &+ O(N^{-3})k^2 \end{aligned}$$

which coincides with the expectation of $\| \sum \mathbf{X_t}\varepsilon_{t+1}/N \|^4_{R^{-1}(k)}$ for a zero mean Gaussian AR(∞) process $\{X_t\}$ (Shibata (1980)) and the proof is complete.

In the same way, we can easily verify the formula for m= 6 and 8.

Lemma 3.2. *Under assumptions* **A0**$-$**A4** *and for a divergent sequence* $k = k_n$*, the zero mean process* (2.1)*, as* $n \to \infty$*, satisfies*

(i) $$\|N^{-1} \sum_{t=K_n}^{n-1} \mathbf{X}_t(\varepsilon_{t+1,k} - \varepsilon_{t+1})\|^2 \xrightarrow{P} 0, \quad \textit{and}$$

(ii) $$\Big| \|N^{-1} \sum_{t=K_n}^{n-1} \mathbf{X}_t \varepsilon_{t+1,k}\|^2_{R^{-1}(k)} - \|N^{-1} \sum_{t=K_n}^{n-1} \mathbf{X}_t \varepsilon_{t+1}\|^2_{R^{-1}(k)} \Big| \xrightarrow{P} 0,$$

where $\varepsilon_{t+1,k}$ *is defined in* (2.3).

Proof. (i) Let $\gamma_i = a_i(k) - a_i$. Then, the expectation of the quantity under investigation is equal to

$$\sum_{l=1}^{k} \sum_{t_1,t_2=K_n}^{n-1} \sum_{m_1,m_2=0}^{\infty} N^{-2} \gamma_{m_1}\gamma_{m_2} E(X_{t_1+1-m_1}X_{t_1+1-l}X_{t_2+1-m_2}X_{t_2+1-l})$$

$$= \sum_{l=1}^{k} \sum_{m_1,m_2=0}^{\infty} \gamma_{m_1}\gamma_{m_2} E\Big\{\Big(\sum_{t=K_n}^{n-1} \frac{X_{t+1-m_1}X_{t+1-l}}{N}\Big)\Big(\sum_{t=K_n}^{n-1} \frac{X_{t+1-m_2}X_{t+1-l}}{N}\Big)\Big\}$$
$$= \sum_{l=1}^{k} \sum_{m_1,m_2=0}^{\infty} \gamma_{m_1}\gamma_{m_2} \Big\{cum\Big(\sum_{t=K_n}^{n-1} \frac{X_{t+1-m_1}X_{t+1-l}}{N}\Big)cum\Big(\sum_{t=K_n}^{n-1} \frac{X_{t+1-m_2}X_{t+1-l}}{N}\Big)$$
$$+cum\Big(\sum_{t=K_n}^{n-1} \frac{X_{t+1-m_1}X_{t+1-l}}{N}, \sum_{t=K_n}^{n-1} \frac{X_{t+1-m_2}X_{t+1-l}}{N}\Big)\Big\}$$
$$= \sum_{l=1}^{k} \Big(\sum_{m=0}^{\infty} \gamma_m N^{-1} E(\sum_{t=K_n}^{n-1} X_{t+1-m}X_{t+1-l})\Big)^2 + \sum_{l=1}^{k} \Big(\sum_{m=0}^{\infty} \gamma_m\Big)^2 O(N^{-1}),$$

where the cumulant of order 2 is $O(N^{-1})$ by (2.11). Then, the above expectation becomes:

$$\sum_{l=1}^{k} \Big(\sum_{m=0}^{\infty} \gamma_m \frac{1}{N} N r_{ml}\Big)^2 + \sum_{l=1}^{k} \Big(\sum_{m=0}^{\infty} \gamma_m\Big)^2 O(N^{-1})$$
$$\le \|\boldsymbol{\gamma}\|^2 \sum_{l=1}^{k} \Big(\sum_{m=0}^{\infty} \tilde{\gamma}_m r_{ml}\Big)^2 + \sum_{l=1}^{k} \Big(\sum_{m=0}^{\infty} \gamma_m\Big)^2 O(N^{-1}),$$

where $\tilde{\gamma}_m = \gamma_m/\|\boldsymbol{\gamma}\|$. Note that $\|\tilde{\gamma}\| = 1$ where $\tilde{\gamma} = (\tilde{\gamma}_1, \tilde{\gamma}_2, \ldots)'$. Using the definition of the norm of R and the fact that $\mathbf{a}(k)$ is the projection of $\mathbf{a}$ the above quantity is bounded by $\|\boldsymbol{\gamma}\|^2\|R\|^2 + \sum_j |a_j| k N^{-1} O(1)$ which converges to zero by **A3**, inequality (2.7), and the fact that $\|\boldsymbol{\gamma}\|$ converges to zero since k is a divergent sequence.

(ii) Let $\mathbf{C}_k = \sum \mathbf{X}_t \varepsilon_{t+1,k}/N$ and $\mathbf{C} = \sum \mathbf{X}_t \varepsilon_{t+1}/N$. Simple calculations show that

$$\Big|\|\mathbf{C}_k\|^2_{R^{-1}(k)} - \|\mathbf{C}\|^2_{R^{-1}(k)}\Big|$$
$$\le \Big(\Big|\|\mathbf{C}_k\|_{R^{-1}(k)} - \|\mathbf{C}\|_{R^{-1}(k)}\Big|\Big)^2 + 2\|\mathbf{C}\|_{R^{-1}(k)} \Big|\|\mathbf{C}_k\|_{R^{-1}(k)} - \|\mathbf{C}\|_{R^{-1}(k)}\Big|$$
$$\le \|\mathbf{C}_k - \mathbf{C}\|^2_{R^{-1}(k)} + 2\|\mathbf{C}\|_{R^{-1}(k)}\|\mathbf{C}_k - \mathbf{C}\|_{R^{-1}(k)}.$$

Note that $\|\mathbf{C}_k - \mathbf{C}\|_{R^{-1}(k)} \le \|\mathbf{C}_k - \mathbf{C}\| \cdot \|R^{-1}(k)\|$. The result follows from Lemma 3.1, the first part of this Lemma and the fact that the norm of the matrix $R^{-1}(k)$ is finite.

Lemma 3.3. *Under assumptions* **A0**–**A4** *the following statements hold as* $n \to \infty$:

(i) $\max_{1\le k\le K_n} \|\hat{R}(k) - R(k)\| \xrightarrow{P} 0,$ (ii) $\max_{1\le k\le K_n} \|\hat{R}^{-1}(k) - R^{-1}(k)\| \xrightarrow{P} 0,$

where $\hat{R}(k)$ is the $k \times k$ sample covariance matrix and $\hat{R}^{-1}(k)$ the inverse of $\hat{R}(k)$.

Proof. (i) It is easy to see that, for $\hat{r}_{ij} = \sum_{t=K_n}^{n-1} X_{t+i}X_{t+j}/N$,

$$\max_{1\le k\le K_n} \|\hat{R}(k) - R(k)\| \le \max_{1\le k\le K_n} \sum_{1\le i,j\le k} (\hat{r}_{ij} - r_{ij})^2 = \sum_{1\le i,j\le K_n} (\hat{r}_{ij} - r_{ij})^2.$$

Note that $\hat{r}_{ij}$ is an unbiased estimator for r_{ij} so that

$$E\Big(\sum_{i,j=1}^{K_n} (\hat{r}_{ij} - r_{ij})^2\Big) = \sum_{i,j=1}^{K_n} cum(\frac{\sum X_{t+i}X_{t+j}}{N}, \frac{\sum X_{t+i}X_{t+j}}{N}).$$

By assumption **A4** and (2.11) the above cumulant is of order $O(N^{-2+1})$ so that the expectation is of order $O(1)K_n^2/N$ which tends to 0 by **A3**. The proof of the second part of this Lemma is exactly the same as the one in Lemma 3.2 part (ii) (Karagrigoriou (1995)).

Note that the formulas for $E\|\sum \mathbf{X}_t\varepsilon_{t+1}/N\|^i_{R^{-1}(k)}$, $i = 2, 4, 6, 8$ are identical to the expressions that arise in the case where the errors have a Gaussian distribution (Shibata (1980), Karagrigoriou (1992) and (1995)). In addition, by Lemma 4.1 of Karagrigoriou (1995) both k_n^* and $\hat{k} \equiv \hat{k}_n$, the sequences that minimize $L_n(k)$ and the $S_n(k)$ criterion respectively, are divergent sequences. Consequently, the results presented by Shibata (1980) for a Gaussian process and, in general, for any process for which the moments of the error distribution coincide with those of the Normal distribution are also valid in this case where the Gaussian assumption has been dropped and the moments up to the 16^{th} order are only required to be finite. As a result, the following main Theorems hold:

Theorem 3.1. *Under assumptions* **A0**$-$**A4**, *for the processes* (2.1) *and* (2.3) *and for every divergent sequence $k = k_n$*

$$Q_n(k)\Big/L_n(k) \xrightarrow{P} 1, \qquad as \qquad n \to \infty.$$

Furthermore, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P[L_n(\hat{k})\Big/L_n(k_n^*) \le 1 + \varepsilon] = 1,$$

where $\hat{k}$ is the order that minimizes the criterion $S_n(k) = (N + 2k)\hat{\sigma}_k^2$ and k_n^* is the sequence that minimizes $L_n(k)$.

Remark 3.1. By Theorem 3.1 and the fact that $L_n(k_n^*) = \min_k L_n(k)$ we have

$$\lim_{n\to\infty} P\Big[Q_n(k)\Big/L_n(k_n^*) \ge 1 - \varepsilon\Big] = 1, \quad \text{for all } \varepsilon > 0,$$

which indicates that the mean squared error of prediction $Q_n(k)$ is asymptotically never below $L_n(k_n^*)$. The following theorem shows that this lower bound is attained when k is the order selected by the $S_n(k)$ criterion.

Theorem 3.2. *Under assumptions* **A0−A4** *and for the processes* (2.1) *and* (2.3)*, the order* $\hat{k}$ *selected by* $S_n(k)$ *is asymptotically efficient, i.e.,*

$$Q_n(\hat{k})\Big/L_n(k_n^*) \xrightarrow{P} 1, \quad as \quad n \to \infty,$$

where k_n^* *is the divergent sequence that minimizes* $L_n(k)$.

Theorem 3.3. *Let* $\delta_n(k)$ *be a real-valued random or nonrandom function of* k *and* k_n^* *be the divergent sequence that minimizes* $L_n(k)$*. Under assumptions* **A0−A4***, if the true underlying process is an* $AR(\infty)$ *process of the form* (2.1) *and an* $AR(k)$ *process of the form* (2.3) *is fitted to a set of observations and if*

$$\text{(i)} \quad \max_{1\le k\le K_n}\left|\frac{\delta_n(k)}{N}\right| \xrightarrow{P} 0 \qquad and \qquad \text{(ii)} \quad \max_{1\le k\le K_n}\left|\frac{\delta_n(k)-\delta_n(k_n^*)}{NL_n(k)}\right| \xrightarrow{P} 0,$$

then $\hat{k}^0$*, the order selected by the criterion* $S_n^0(k) = (N + \delta_n(k) + 2k)\hat{\sigma}_k^2$ *(Shibata* (1980)) *is asymptotically efficient.*

The asymptotic efficiency of the AIC-like criteria as well as the inefficiency of BIC and ϕ criteria is established in the following Corollary, which is the equivalent to Corollary 5.1 of Karagrigoriou (1995) for the zero mean nongaussian case and follows immediately from Theorem 3.3.

Corollary 3.1. *Under assumptions* **A0−A4***, if the true underlying process is zero mean* $AR(\infty)$ *of the form* (2.1) *and an* $AR(k)$ *process of the form* (2.3) *is fitted to a set of* n *observations then*

1. $AIC(k) = n\log\hat{\sigma}_k^2 + 2k$, $FPE(k) = \hat{\sigma}_k^2(1 + (2k)/n)$, $S_n'(k) = (n + 2k)\hat{\sigma}_k^2$, $AIC_C(k) = AIC(k)+2(k+1)(k+2)(n-k+2)^{-1}$ *and* $FPE_\alpha(k) = \hat{\sigma}_k^2(1+(\alpha k)/n)$, $\alpha \neq 2$ *are asymptotically efficient and*
2. $BIC(k) = n\log\hat{\sigma}_k^2 + k\log n$ *and* $\phi(k) = n\log\hat{\sigma}_k^2 + ck\log\log n$, $c > 2$, *are not asymptotically efficient criteria.*

Examples. The significance of the results presented above lies in the fact that the error distribution is allowed to be nongaussian as long as the moments up to the 16th order are finite. The simplest case of such a distribution that can be considered is a t−distribution with at least 17 degrees of freedom. Also, some assymetric distributions with finitely many moments fall into the same category.

The most interesting case is the one in which the error sequence is a mixture of Gaussian white noise processes. Such a mixture model is of the form $\varepsilon_t = Z_t/\sqrt{W/d}$ where Z_t is an i.i.d. Gaussian sequence and W is a chi-square variable with d degrees of freedom, $d > 16$. Note that in general W may be any nonnegative random variable independent of Z_t. Observe that, in the case of the chi-square distribution, the errors constitute a sequence of uncorrelated random variables following the t-distribution. Applications of the mixture models can be spotted in linear state space models like the seasonal adjustment models. In such models the estimation of the state is usually of great interest because it simplifies the problem of the analysis of the series. A procedure for the state estimation or equivalently, the fitting of the process, is provided by either the AIC criterion or its likeness.

Apart from these theoretical considerations which are well known, the practical applications and the implications and differences the nongaussian distributions entail are the key issues in the contributions of the present work.

4. Asymptotic Efficiency of the Order Selection for the Spectral Density Estimate

We turn now to the asymptotic efficiency from the point of view of the autoregressive spectral estimation. Consider a stationary process $\{X_t\}$ with expectation μ and covariance $r_l = E(X_t X_{t+l}) - \mu^2$ of the form

$$m + X_{t+1} + a_1 X_t + a_2 X_{t-1} + \cdots = \varepsilon_{t+1}, \qquad t = \ldots, -1, 0, 1, \ldots \tag{4.1}$$

where $a_1, a_2, \ldots$, and m are real numbers and $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables satisfying assumption **A0**. Note that $\{Z_t\} = \{X_t - \mu\}$ is a stationary zero mean process sharing the same covariance and autocovariance matrices with $\{X_t\}$.

Given observations $x_1, \ldots, x_n$, if the kth order AR model with $k \times k$ covariance matrix $R(k)$ and innovation variance σ_k^2

$$m(k) + X_{t+1} + a_1(k) X_t + a_2(k) X_{t-1} + \cdots + a_k(k) X_{t+1-k} = \varepsilon_{t+1,k} \tag{4.2}$$

is assumed, the L.S.E. $\hat{\mathbf{a}}(k)$ of $\mathbf{a}(k)$ and the estimators $\hat{m}(k)$ of $m(k)$ and

$$\hat{\sigma}_k^2 = \sum_{t=K_n}^{n-1} (\hat{m}(k) + X_{t+1} + \hat{a}_1(k) X_t + \cdots + \hat{a}_k(k) X_{t+1-k})^2 / N$$

of σ_k^2 are obtainable. Note that $\mathbf{b}'(k) = (m(k), a_1(k), \ldots, a_k(k), 0, \ldots)$ is the projection of $\mathbf{b}' = (m, a_1, a_2, \ldots)$ with respect to the norm $\| \cdot \|_R^*$ on the $k+1$ dimensional subspace $V(k+1) = \{\mathbf{c}(k) : \mathbf{c}'(k) = (c_0, c_1, \ldots, c_k, 0, \ldots)\}$, where

$$\|\mathbf{b}\|_R^* = \Big\{ \Big(b_0 + (\sum_{j=1}^{\infty} b_j)\mu \Big)^2 + \sum_{i,j=1}^{\infty} b_i b_j r_{ij} \Big\}^{\frac{1}{2}}.$$

Observe that $m(k) = -\mu \sum a_i(k)$ and the projections of $\mathbf{a}$ into $V(k+1)$ and $V(k)$ (see section 2) subspaces coincide so that (4.2) is equivalent to $Z_{t+1} + a_1(k)Z_t + \cdots + a_k(k)Z_{t+1-k} = \varepsilon_{t+1,k}$. In this case, the 1-step ahead predictor of Y_{t+1} is given by $\hat{Y}_{t+1} = -\hat{m}(k) - \sum_{i=1}^{k} \hat{a}_i(k) Y_{t+1-i}$. Furthermore, let $\hat{R}(k) = (\hat{r}_{ij}, 1 \le i,j < k)$ be the estimator of $R(k)$, where $\hat{r}_{ij} = \hat{r}_{ij}^z - N^{-2} \sum Z_{t+1-i} \sum Z_{t+1-j}$ and $\hat{r}_{ij}^z$ the sample covariance of Z_{t-i} and Z_{t-j}.

Under assumptions **A1**, **A2**, and **A4**, the penalty function $Q_n(k)$ is given by the formula (Lemma 2.1, Karagrigoriou (1995)):

$$Q_n(k) = \|\hat{\mathbf{a}}(k) - \mathbf{a}\|_R^2 + (\hat{m}(k) + \mu \sum_{i=0}^{k} \hat{a}_i(k))^2, \quad \hat{a}_0(k) = 1$$

and the quantity asymptotically equivalent to the expectation of $Q_n(k)$ is given by (Lemma 3.6, Karagrigoriou (1992)):

$$L_n(k) = kN^{-1}\sigma^2 + \|\gamma\|_R^2 + N^{-1}(\sigma m(k)/m)^2, \tag{4.3}$$

provided that assumptions **A0**$-$**A4** hold, where $\|\cdot\|_R$ is defined in section 2.

In addition to the assumptions **A0**$-$**A4**, we now impose the following assumption on the parameters of the process (4.1):

A5. $\sum_{k=1}^{\infty} \big(\sum_{j=k+1}^{\infty} |a_j| \big)^2 < \infty$.

The following result is stated without proof. The proof under the Gaussian assumption is given in Karagrigoriou (1995) but stands as is in the present set-up.

Lemma 4.1. *Under assumptions* **A0**$-$**A5** *and for the process* (4.2) *the following statements hold:*

$$\text{(i)} \quad \max_{1\le k\le K_n} \Big| \frac{(\sum_{t=K_n}^{n-1} \varepsilon_{t+1,k})^2}{N^2 L_n(k)} \Big| \xrightarrow{P} 0 \quad \textit{and} \quad \text{(ii)} \quad \max_{1\le k\le K_n} \Big(\sum_{t=K_n}^{n-1} \frac{\varepsilon_{t+1,k}}{N} \Big)^2 \xrightarrow{P} 0.$$

Furthermore, for any divergent sequence $k = k_n$ ($k_n \le K_n$), we have

$$\text{(iii)} \qquad Nk^{-1}\Big(\hat{m}(k) + \mu \sum_{i=0}^{k} \hat{a}_i(k)\Big)^2 \xrightarrow{P} 0.$$

Lemma 4.2. *Under assumptions* **A0**$-$**A5**, *for the process* (4.2) *and for any divergent sequence $k = k_n$ ($k_n \le K_n$), we have*

$$Nk^{-1}\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_R^2 \xrightarrow{P} \sigma^2$$

Proof. Let $\mathbf{Z}_t = (Z_t, Z_{t-1}, \ldots, Z_{t+1-k})'$,

$$\mathbf{B} = \sum \frac{\mathbf{Z}_t \varepsilon_{t+1}}{N}, \ \mathbf{B}_k = \sum \frac{\mathbf{Z}_t \varepsilon_{t+1,k}}{N}, \ \mathbf{W} = \sum \frac{\mathbf{Z}_t}{N}, \ \mathbf{E}_k = \sum \frac{\varepsilon_{t+1,k}}{N},$$

$c_1 = \hat{R}^{-1}(k)\mathbf{B}_k$, and $c_2 = \hat{R}^{-1}(k)\mathbf{W}\mathbf{E}_k$. Note that $\hat{\mathbf{a}}(k) - \mathbf{a}(k) = c_2 - c_1$. By Lemmas 3.2 and 3.3 the convergence of $P[|Nk^{-1}||c_1||_R^2 - \sigma^2| \geq \varepsilon]$ to zero is equivalent to the convergence of $E[N||\mathbf{B}||^2_{R^{-1}(k)} - k\sigma^2]^2/k^2\varepsilon^2$ to zero. Simple calculations show that the above expectation tends indeed to zero since, by Lemma 3.1, it is equal to $(2k\sigma^4 + O(N^{-3}))/\varepsilon^2 k^2$.

Furthermore, $||c_2||_R^2$ tends to zero by the boundedness of $||R||$ and $||\mathbf{W}||$, Lemma 4.1 (ii) and the fact that $||\mathbf{E}_k||^2$ is of order $O(N^{-1})$. The result is obtained by the squeezing theorem applied to the inequality:

$$\Big|\sum_{i=1}^{2} \|c_i\|_R^2 - 2\|c_1\|_R\|c_2\|_R\Big| \leq \|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_R^2 \leq \sum_{i=1}^{2} \|c_i\|_R^2 + 2\|c_1\|_R\|c_2\|_R.$$

Lemma 4.3. *Under assumptions* **A0**$-$**A4**,

(i) $\hat{\sigma}_k^2 - \sigma^2 \xrightarrow{P} 0$

(ii) $\sum_{j=0}^{k} |\hat{a}_j(k) - a_j(k)| \xrightarrow{P} 0, \quad for\ 1 \leq k \leq K_n$

(iii) *If* $\lim k_n = \infty$ *then* $\|\overline{R} - R\| \to 0$, *uniformly in* $k_n \leq k \leq K_n$.

Proof. (i) Let $s_k^2 = \sum \varepsilon_{t+1,k}^2/N$ and $\sigma_k^2 = Es_k^2$. Using the triangular inequality we have:

$$|\hat{\sigma}_k^2 - \sigma^2| \leq |\hat{\sigma}_k^2 - s_k^2| + |s_k^2 - \sigma_k^2| + |\sigma_k^2 - \sigma^2|. \tag{4.4}$$

The last term tends to zero since $\sigma_k^2 \to \sigma^2$ (Grenander and Szergö (1958)). It can easily be shown that the first term of the R.H.S. of (4.4) equals

$$\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|^2_{\hat{R}(k)} + \Big(\sum_{t=K_n}^{n-1} \frac{\varepsilon_{t+1,k}}{N}\Big)^2.$$

By adding and subtracting the term $k\sigma^2/N$, we have:

$$|\hat{\sigma}_k^2 - s_k^2| \leq \frac{k}{N}\Big|\frac{N}{k}\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|^2_{\hat{R}(k)} - \sigma^2\Big| + |k\sigma^2/N| + \Big(\sum_{t=K_n}^{n-1} \frac{\varepsilon_{t+1,k}}{N}\Big)^2.$$

Using Lemma 4.1 (ii), Lemma 4.2 and assumption **A3** the first term of the R.H.S of (4.4) converges to 0. For the second term of the R.H.S. of (4.4) using Chebychev's inequality we have that $P[|s_k^2 - \sigma_k^2| \geq \varepsilon] \leq \varepsilon^{-2}\mathrm{Var}\,(s_k^2)$. By the definition of s_k^2 the $\mathrm{Var}\,(s_k^2) = o_p(1)$ and thus the second term tends also to 0.

(ii) Using simple properties of $\|\cdot\|$ we have:

$$\begin{aligned}\sum_{j=0}^{k} |\hat{a}_j(k) - a_j(k)| &= \sum_{j=0}^{k} \sqrt{(\hat{a}_j(k) - a_j(k))^2} \\ &\leq [k\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|^2]^{1/2} \leq [k\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_R^2]^{1/2}\|R^{-1}\|^{1/2}.\end{aligned}$$

Using the same argument as in part (i) and the fact that the norm of R^{-1} is bounded we have the result.

(iii) The result follows from (Shibata (1981)) Lemma 2.3 and previous results.

Lemma 4.4. *Under assumptions* **A0**$-$**A4** *and for any divergent sequence* k*, the following statements hold as* $n \to \infty$*:*

$$\text{(i)}\quad \frac{\hat{\sigma}_k^2 - \sigma^2}{\sqrt{L_n(k)}} \xrightarrow{P} 0, \qquad \text{(ii)}\quad \frac{\|\hat{\mathbf{a}}(k) - \mathbf{a}\|_R^2}{L_n(k)} \xrightarrow{P} 1, \qquad \textit{and} \qquad \text{(iii)}\quad \frac{\tilde{m}^2}{L_n(k)} \xrightarrow{P} 0$$

where $\tilde{m} = \hat{m}(k) + \mu \sum \hat{a}_i(k)$.

Proof. (i) It can be easily seen that

$$\begin{aligned} \frac{(\sigma_k^2 - \sigma^2) + (\hat{\sigma}_k^2 - s_k^2)}{\sqrt{L_n(k)}} &= \frac{\|\hat{\mathbf{a}}(k) - \mathbf{a}\|_R^2}{\sqrt{L_n(k)}} + \frac{\left(\sum \varepsilon_{t+1,k}\right)^2}{N^2\sqrt{L_n(k)}} \\ &\le \sqrt{L_n(k)}\Big(\frac{Q_n(k)}{L_n(k)} + \max_k \frac{\left(\sum \varepsilon_{t+1,k}\right)^2}{N^2 L_n(k)}\Big) \end{aligned}$$

which converges to zero by Lemma 4.1 (i), Theorem 3.1, and the fact that $L_n(k)$ converges to zero. Furthermore,

$$P\Big[\max \frac{|s_k^2 - \sigma_k^2|}{\sqrt{L_n(k)}} \ge \varepsilon\Big] \le \frac{1}{\varepsilon^4}\sum_{k=1}^{K_n} E\Big[\frac{(s_k^2 - \sigma_k^2)^4}{L_n(k)^2}\Big] \le \frac{N^2}{\varepsilon^4\sigma^4}\sum_{k=1}^{K_n}\frac{1}{k^2}\text{Var}\Big(\frac{\sum \varepsilon_{t+1,k}^4}{N^4}\Big)$$

which tends to 0. Using the triangular inequality, the result is immediate.

(ii) Let $W = \|\hat{\mathbf{a}}(k) - \mathbf{a}\|_R^2 / L_n(k)$. Then, using (4.3), we have:

$$W - 1 = \frac{\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_R^2 - \|\gamma\|_R^2}{\frac{k\sigma^2}{N} + \|\gamma\|_R^2 + \frac{1}{N}(\frac{\sigma m(k)}{m})^2} - 1 \le \frac{\frac{N}{k}\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_R^2 - \sigma^2 - \frac{1}{k}(\frac{\sigma m(k)}{m})^2}{\sigma^2}$$

which converges to zero by Lemma 4.2 and the facts that k is a divergent sequence and $m(k)$ tends to m. The result follows from Lemma 4.3 (iii).

(iii) Using the formula for $L_n(k)$ we have $\tilde{m}^2/L_n(k) \le (N\tilde{m}^2)/(k\sigma^2)$ which tends to zero by Lemma 4.1 (iii).

Theorem 4.1. *Under assumptions* **A0**$-$**A4***, and for any divergent sequence* k *we have*

$$L_n^{-1}(k) \cdot |\widetilde{Q}_n(k) - \frac{2}{\hat{\sigma}_k^2} Q_n(k)| \xrightarrow{P} 0$$

If, in addition, assumption **A5** *is satisfied, then*

$$\frac{\widetilde{Q}_n(\hat{k})}{L_n(k_n^*)} \xrightarrow{P} 2/\sigma^2, \qquad n \to \infty$$

where $\hat{k}$ is the order selected by $S_n(k)$ and k_n^* the sequence that minimizes $L_n(k)$.

Proof. Consider the following known fact which holds for any spectral function $f(\cdot)$ and any type of estimators (see Shibata (1981) for details) :

$$\begin{aligned}&\frac{1}{L_n(k)}|\widetilde{Q}_n(k)-\frac{2}{\hat{\sigma}_k^2}\overline{Q}_n(k)|\\&\leq\frac{1}{L_n(k)}\Big\{\Big(\frac{\sigma^2-\hat{\sigma}_k^2}{\sigma^2}\Big)^2+\frac{2}{\hat{\sigma}_k^2}\tilde{m}^2+\frac{\|\hat{\mathbf{a}}(k)-\mathbf{a}\|_R^2}{\sigma^4}\\&\times\Big(|\mathbf{a}-\hat{\mathbf{a}}(k)|^2F^*+4|\mathbf{a}-\hat{\mathbf{a}}(k)|F^{*12}\hat{\sigma}_k+2|\sigma^2-\hat{\sigma}_k^2|(2+\frac{\sigma^2}{\hat{\sigma}_k^2})\Big)\Big\},\end{aligned}$$

where $F^*=\max_{-\pi\leq\lambda\leq\pi}\hat{f}_k(\lambda)$, $|\mathbf{a}-\hat{\mathbf{a}}(\cdot)|=\sum_{j=0}^k|\hat{a}_j(\cdot)-a_j(\cdot)|+\sum_{j=k+1}^\infty|a_j|$, and $\overline{Q}_n(k)=\|\hat{\mathbf{a}}(k)-\mathbf{a}\|_R^2+\tilde{m}^2$. By Lemmas 4.3 and 4.4 the above quantity tends to 0. By Lemma 4.3 (iii),

$$L_n^{-1}(k)\cdot|\widetilde{Q}_n(k)-\frac{2}{\hat{\sigma}_k^2}(\|\hat{\mathbf{a}}(k)-\mathbf{a}\|_R^2+\tilde{m}^2)|\xrightarrow{P}0$$

and the proof of the first assertion is established. Combining Theorem 3.1 and the fact that k_n^* minimizes $L_n(k)$ over $1\leq k\leq K_n$, we have

$$L_n(\hat{k})/L_n(k_n^*)\longrightarrow 1. \tag{4.5}$$

For the second part, we have

$$\frac{\widetilde{Q}_n(k)}{L_n(k_n^*)}=\frac{1}{L_n(k)}\Big\{\widetilde{Q}_n(k)-\frac{2}{\hat{\sigma}_k^2}Q_n(k)+\frac{2}{\hat{\sigma}_k^2}Q_n(k)\Big\}\cdot\frac{L_n(k)}{L_n(k_n^*)}.$$

Replacing k by the divergent sequence $\hat{k}$ (see Karagrigoriou (1995)) and using Theorem 3.2, the first part of this theorem, and (4.5) we have the asymptotic efficiency of the spectral density estimate.

Remark 4.1. Applying Theorem 3.3 to the zero mean nongaussian AR(∞) process $\{Z_t\}$, we conclude that Theorem 4.1 holds for all $AIC-$like criteria mentioned in Collorary 3.1, namely, AIC, FPE, FPE_α, AIC_C, and $S_n'(k)$.

5. Discussion

Under assumptions **A1**-**A4**, the asymptotic efficiency of $S_n(k)$ as well as that of $AIC-$like criteria is assured as long as the moments up to the 16th order exist, which is the only condition imposed on the true error distribution. Furthermore, even from the spectral density point of view, the asymptotic efficiency of AIC-like criteria is assured provided that the parameters of the process satisfy certain conditions. Note that the fact that this asymptotic result holds even for the nongaussian case is due to the use of a linear predictor in our criterion as well as to the use of the least squares estimators of the parameters of the fitted process. In conclusion, we can safely say that, provided that the moments up to the 16th order exist, the LSE's (i.e. the Gaussian MLE's) and consequently the AIC-like criteria derived under the Gaussian assumption can be used in the nongaussian case without affecting the asymptotic efficiency. The significance of Theorems 3.2, 3.3, and 4.1 lies in the fact that the Gaussian maximum likelihood estimators can be used, even if the true error distribution is not necessarily Gaussian, as long as the object of the analysis is to select a finite order process with optimal predictive performance.

Acknowledgement

I would like to express my deep appreciation to my advisor Professor Ching-Zong Wei for his guidance and support over the last several years and Professor Paul J. Smith for his helpful suggestions. I would also like to thank an associate editor and two anonymous referees for their constructive comments.

References

Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22**, 203-217.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (Edited by B. N. Petrov and F. Csáki), 267-281. Akadémia Kiadó, Budapest.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2**, 489-502.

Bhansali, R. J. (1986a). Asymptotically efficient selection of the order by the criterion autoregressive transfer function. *Ann. Statist.* **14**, 315-325.

Bhansali, R. J. (1986b). The criterion autoregressive transfer function of Parzen. *J. Time Ser.* **7**, 79-104.

Bhansali, R. J. and Downham, D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion. *Biometrika* **64**, 547-551.

Brillinger, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* **56**, 375-390.

Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and Their Applications.* Univ. California Press, Berkeley.

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.

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

Hannan, E. J. and Kavalieris, L. (1986). Regression, autoregression models. *J. Time Ser.* **7**, 27-49.

Härdle, W. (1987). An effective selection of regression variables when the error distribution is incorrectly specified. *Ann. Inst. Statist. Math.* **39**, 533-548.

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Karagrigoriou, A. (1992). Asymptotic efficiency of model selection procedures in time series. *Ph.D. Dissertation*, University of Maryland.

Karagrigoriou, A. (1995). Asymptotic efficiency of model selection criteria: the nonzero mean gaussian AR(∞) case. *Comm. Statist. Theory Methods* **24**, 911-930.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661-675.

Parzen, E. (1974). Some recent advances in time series modeling. *IEEE Trans. Automat. Control* **AC-19**, 723-730.

Rosenblatt, M. (1983). Cumulants and cumulant spectra. In *Handbook of Stat.* **V3** (Edited by D. R. Brillinger and P. R. Krishnaiah), 369-382.

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.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8**, 147-164.

Shibata, R. (1981). An optimal autoregressive spectral estimate. *Ann. Statist.* **9**, 300-306.

Taniguchi, M. (1980). On selection of the order of the spectral density model for a stationary process. *Ann. Inst. Statist. Math.* **32**, 401-419.

Department of Mathematics and Statistics, University of Cyprus, P. O. Box 537, CY-1678 Nicosia, Cyprus.

(Received October 1994; accepted December 1995)