Ann. Inst. Statist. Math.
32 (1980), Part A, 401-419

# ON SELECTION OF THE ORDER OF THE SPECTRAL DENSITY MODEL FOR A STATIONARY PROCESS

MASANOBU TANIGUCHI

(Received Nov. 13, 1979; revised Sept. 3, 1980)

## Summary

Let $\{X(t)\}$ be a stationary process with mean zero and spectral density $g(x)$. We shall use a $k$th order parametric spectral model $f_{\tau(k)}(x)$ for this process. Without Gaussianity we can obtain an estimate of $\tau(k)$, say $\hat{\tau}(k)$, by maximizing the quasi-Gaussian likelihood of this model. We can then construct the best linear predictor of $X(t)$, which is computed on the basis of the estimated spectral density $f_{\hat{\tau}(k)}(x)$. An asymptotic lower bound of the mean square error of the estimated predictor is obtained. The bound is attained if $k$ is selected by Akaike's information criterion.

## 1. Introduction

There has been much discussion on the fitting of parametric models to a time series. Most of papers are based on the assumption that the data come from an autoregressive or autoregressive moving average process of a known finite order. But it is rather natural assuming the data come from a linear process with infinitely many unknown parameters. From this point of view, Shibata [8] investigated a finite order autoregressive model fitting to a Gaussian linear process with infinitely many parameters.

In this paper we shall extend the results of Shibata [8] to the case when the process is not necessarily Gaussian and the model is not necessarily autoregressive. We evaluate the goodness of the spectral estimate $f_{\hat{\tau}(k)}(x)$ by the mean squared error of prediction which is obtained by Kolmogorov-Wiener theory. In general the predictor has an infinite series expansion in time domain, so that it could not be applied in practice. However the purpose of this paper is not the construction of a realizable predictor but the order selection and estimation of the spectral density.

In Section 4, we shall give an asymptotic lower bound and define a concept of asymptotically efficient order selection by extending the definition of Shibata [8]. In Section 5, if $k$ is selected by Akaike's information criterion [2], the mean squared error of the estimated predictor attains the lower bound.

## 2. Preliminaries

Let $\{X(t);\ t=\cdots,-1,0,1,\cdots\}$ be a linear process

$$X(t)=\sum_{j=0}^{\infty} a(j;\theta)e(t-j)\ , \qquad a(0;\theta)=1\ , \tag{2.1}$$

satisfying

$$\sum_{j=0}^{\infty} j^{\beta}|a(j;\theta)|<\infty\ , \qquad \text{for some } \beta>1\ , \tag{2.2}$$

where $a(j;\theta)$'s are known functions of an infinite dimensional parameter $\theta=(\theta_1,\theta_2,\cdots)'$, and $e(j)$'s are independently and identically distributed with finite cumulants $\kappa_s$, $s=1,\cdots,16$. The spectral density of $\{X(t)\}$ is then written as

$$g(x;\theta)=\frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty} a(j;\theta)\exp(ijx)\right|^2 , \tag{2.3}$$

where $\sigma^2=\mathrm{E}[e^2(j)]$. For notational convenience, sometimes we write simply $g(x)$ in the place of $g(x;\theta)$.

We need the following assumptions (A.1)–(A.3):

(A.1) $g(x;\theta)$ is three times differentiable with respect to each coordinates of $\theta\in\Theta$, and the third order derivative is continuous function of $(x,\theta)\in[-\pi,\pi]\times\Theta$, where $\Theta$ is a compact set in $R^{\infty}$.

(A.2) The associated power series

$$A(z)=1+a(1;\theta)z+a(2;\theta)z^2+\cdots$$

in not zero for $|z|\leqq 1$.

(A.3) The true parameter $\theta$ belongs to the interior of $\Theta$, denoted by Int$(\Theta)$, and $\theta$ has infinitely many nonzero elements.

We commence to set down the following proposition. As for the definition of the cumulant $\{\cdot\}$, see Brillinger [4].

PROPOSITION 1. For the process $\{X(t)\}$ which satisfies (2.2), we have

$$\sum_{t_1,\cdots,t_p=-\infty}^{\infty} |t_l||\mathrm{cum}(t_1,\cdots,t_p)|<\infty\ , \qquad \text{for } l,p=1,\cdots,15\ , \tag{2.4}$$

where $\text{cum}\,(t_1,\cdots,t_p)=\text{cumulant}\,\{X(t_1+t),\cdots,X(t_p+t),X(t)\}$.

PROOF. For $p=1,\cdots,15$,

$$\begin{aligned}(2.5)\quad &\text{cum}\,\{X(t_1+t),\cdots,X(t_p+t),X(t)\}\\ &=\text{cum}\left\{\sum_{j_1=0}^{\infty}a(j_1;\theta)e(t+t_1-j_1),\cdots,\sum_{j_p=0}^{\infty}a(j_p;\theta)e(t+t_p-j_p),\right.\\ &\qquad\left.\sum_{j=0}^{\infty}a(j;\theta)e(t-j)\right\}.\end{aligned}$$

Noting (2.2), and that the joint cumulant is linear with respect to each variables (Brillinger [4]), we see that (2.5) is equal to

$$(2.6)\qquad \sum_{j_1,\cdots,j_p,j=0}^{\infty}a(j_1;\theta)\cdots a(j_p;\theta)a(j;\theta)\,\text{cum}\,\{e(t+t_1-j_1),\cdots,e(t-j)\}\,.$$

From the independence of $e(t)$'s, (2.6) can be written as

$$\sum_{j=0}^{\infty}a(t_1+j;\theta)\cdots a(t_p+j;\theta)a(j;\theta)\kappa_{p+1}\,.$$

Thus (2.4) is bounded by

$$(2.7)\qquad \sum_{j=0}^{\infty}\sum_{t_1,\cdots,t_p\geq -j}^{\infty}|t_l||a(t_1+j;\theta)\cdots a(j;\theta)||\kappa_{p+1}|\,,\qquad l=1,\cdots,p\,.$$

The result follows from (2.2).

*Remark* 1. The condition (2.2) is stronger than Brillinger [3] type mixing condition (2.4). Of course we can get this proposition under a milder condition than (2.2).

## 3. Model and estimation

Suppose that a stretch, $X(t)$ $(t=0,\cdots,n-1)$ of the time series $X(t)$ is given. Let $f_{\tau(k)}(x)$ be a spectral density with $(k+1)$-dimensional unknown parameter vector $\tau(k)=(\sigma^2(k),\theta(k)')'$, $\sigma^2(k)>0$, $\theta(k)=(\theta_1(k),\cdots,\theta_k(k))'$, by which $f_{\tau(k)}(x)$ is parameterized as

$$f_{\tau(k)}(x)=\frac{\sigma^2(k)}{2\pi}h_{\theta(k)}(x)=\frac{\sigma^2(k)}{2\pi}|\dot{h}_{\theta(k)}(e^{ix})|^2\,,$$

where $\dot{h}_{\theta(k)}(0)=1$.

A parameter vector $\underline{\tau}(k)=(\underline{\sigma}^2(k),\underline{\theta}(k)')'$ gives the best approximation $f_{\underline{\tau}(k)}(x)$ to $g(x)$ in the following sense.

$$(3.1)\qquad \int_{-\pi}^{\pi}\left\{\log f_{\underline{\tau}(k)}(x)+\frac{g(x)}{f_{\underline{\tau}(k)}(x)}\right\}dx$$

$$=\min_{\tau(k)\in\Theta_{k+1}}\int_{-\pi}^{\pi}\left\{\log f_{\tau(k)}(x)+\frac{g(x)}{f_{\tau(k)}(x)}\right\}dx\ ,$$

where $\Theta_{k+1}$ is a compact set of $R^{k+1}$. We can estimate $\underline{\tau}(k)$ by $\hat{\tau}(k)=(\hat{\sigma}^2(k), \hat{\theta}(k)')'$ which is a solution of

$$\min_{\tau(k)\in\Theta_{k+1}}\int_{-\pi}^{\pi}\left\{\log f_{\tau(k)}(x)+\frac{I_n(x)}{f_{\tau(k)}(x)}\right\}dx \tag{3.2}$$

$$=\int_{-\pi}^{\pi}\left\{\log f_{\hat{\tau}(k)}(x)+\frac{I_n(x)}{f_{\hat{\tau}(k)}(x)}\right\}dx\ ,$$

where $I_n(x)=(2\pi n)^{-1}\left|\sum_{t=0}^{n-1}X(t)\exp(-itx)\right|^2$ is the periodogram of $X(t)$. The estimator $\hat{\tau}(k)$ is called the quasi-Gaussian maximum likelihood estimator of $\underline{\tau}(k)$ since Gaussianity of $\{X(t)\}$ is not assumed (see Walker [10], Dunsmuir and Hannan [5]).

In our analysis, the number $k$ of parameters is not always fixed. Therefore, in addition to (A.1)–(A.3) we put the following assumptions (A.4)–(A.10):

(A.4) The number of the parameters $k$ is in $1\leqq k\leqq K_n$, where $K_n\to\infty$ and $K_n/\sqrt{n}\to 0$ as $n\to\infty$.

(A.5) The spectral density $f_{\tau(k)}(x)$ is three times differentiable with respect to $\tau(k)\in\Theta_{k+1}$. The third order derivative is continuous function of $(x, \tau(k))\in[-\pi, \pi]\times\Theta_{k+1}$, and, as a function of $x\in[-\pi, \pi]$, the first and second derivatives satisfy the Lipschitz condition of order 1.

(A.6) $|\dot{h}_{\theta(k)}(z)|$ is bounded and bounded away from zero for $|z|\leqq 1$.

(A.7) For any $1\leqq k\leqq K_n$, the $k\times k$ matrix

$$H_k=\int_{-\pi}^{\pi}\frac{\partial^2 h_{\theta(k)}(x)^{-1}}{\partial\theta(k)\partial\theta(k)'}\bigg|_{\underline{\theta}(k)}g(x)dx\ ,$$

is non-singular. If a symmetric $k\times k$ matrix $A=\{a_{rj}\}$ satisfies that $\sum_{j=1}^{k}|a_{rj}|$ is bounded uniformly in $r$ as $k\to\infty$ we denote $A\in l(k\times k)$.

(A.8) For any $1\leqq k\leqq K_n$, $H_k$, $H_k^{-1/2}$ and $\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)'}\bigg|_{\underline{\theta}(k)}g^2(x)dx$ belong to $l(k\times k)$.

(A.9) The sum $\sum_{j,m=1}^{k}|W(r, j, m)|$ is bounded uniformly in $r$ as $k\to\infty$ where

$$W(r, j, m)=\int_{-\pi}^{\pi}\frac{\partial^3 h_{\theta(k)}(x)^{-1}}{\partial\theta_r(k)\partial\theta_j(k)\partial\theta_m(k)}\bigg|_{\underline{\theta}(k)}g(x)dx\ .$$

(A.10) For the $\beta$ in (2.2),

$$|f_{\tau(k)}(x)-g(x)|=O(k^{-\beta}), \qquad \text{for all } x\in[-\pi,\pi].$$

*Example* 1. Suppose that the true spectral density $g(x;\theta)$ is parameterized such that

$$g(x;\theta)=\frac{\sigma^2}{2\pi}\frac{\left|\sum_{j=0}^{\infty}\mu_{1,j}\exp(ijx)\right|^2}{\left|\sum_{j=0}^{\infty}\mu_{2,j}\exp(ijx)\right|^2},$$

$\mu_{1,0}=\mu_{2,0}=1$, $\theta=(\sigma^2,\mu_{1,1},\mu_{2,1},\mu_{1,2},\mu_{2,2},\cdots)'$, satisfying $\sum_{j=0}^{\infty}j^{\beta}|\mu_{1,j}|<\infty$, $\sum_{j=0}^{\infty}j^{\beta}\cdot|\mu_{2,j}|<\infty$. Also we assume that $\sum_{j=0}^{\infty}\mu_{1,j}z^j$ and $\sum_{j=0}^{\infty}\mu_{2,j}z^j$ are not zero for $|z|\leqq 1$. Of course noting Theorem 3.8.3 of Brillinger ([4], p. 78) we can express the above $g(x;\theta)$ in the form (2.1) satisfying (2.2). We choose an autoregressive moving average spectral model

$$f_{\tau(k)}(x)=\frac{\sigma^2(k)}{2\pi}\frac{\left|\sum_{j=0}^{p}\theta_{1,j}\exp(ijx)\right|^2}{\left|\sum_{j=0}^{p}\theta_{2,j}\exp(ijx)\right|^2},$$

$\theta_{1,0}=\theta_{2,0}=1$, $\tau(k)=(\sigma^2(k),\theta_{1,1},\cdots,\theta_{1,p},\theta_{2,1},\cdots,\theta_{2,q})'$, $p+q=k$, where $\sum_{j=0}^{p}\theta_{1,j}z^j$ and $\sum_{j=0}^{q}\theta_{2,j}z^j$ are not zero for $|z|\leqq 1$. Further assume that $\sum_{j=0}^{p}j^{\beta}|\theta_{1,j}|$ and $\sum_{j=0}^{q}j^{\beta}|\theta_{2,j}|$ are bounded as $p=p(k)$ and $q=q(k)$ tend to infinity such that $p(k)/q(k)\to 1$. Then it is not so hard to show $f_{\tau(k)}(x)$ satisfies (A.4)–(A.10).

Now we shall present the following lemma without the proof because it is easy.

LEMMA 1. (i) *If* $A, B\in l(k\times k)$, *then* $AB\in l(k\times k)$. (ii) *Let* $C=\{C_{rj}\}$ *be a* $k\times k$ *matrix in which each* $C_{rj}$ *is at most of an order* $\mathcal{O}$. *If* $L\in l(k\times k)$, *then any element in* $CL$ *is at most of the order* $\mathcal{O}$.

## 4. Asymptotic properties of the estimated mean square error

We can obtain a predictor $\hat{X}(t)$ by fitting the spectral density $f_{\hat{\tau}(k)}(x)$, and the mean square error is

$$\begin{aligned}&\mathrm{E}\,|X(t)-\hat{X}(t)|^2\\&=\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{g(x)}{f_{\hat{\tau}(k)}(x)}dx\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\hat{\tau}(k)}(x)}{g(x)}dx\right\}-1\right]\sigma^2+\sigma^2\end{aligned}$$

(see Grenander and Rosenblatt [6], p. 261). Although $\hat{X}(t)$ can not be directly applied in practice, as it has an infinite series expansion in time domain, we can measure the goodness of the estimated spectral model $f_{\hat{\tau}(k)}(x)$ by

$$(4.1)\quad D(f_{\hat{\tau}(k)}, g)=\frac{1}{2\pi}\left\{\int_{-\pi}^{\pi}\frac{g(x)}{f_{\hat{\tau}(k)}(x)}dx\right\}\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\hat{\tau}(k)}(x)}{g(x)}dx\right\}-1$$
$$=\frac{1}{2\pi}\left\{\int_{-\pi}^{\pi}\frac{g(x)}{h_{\hat{\theta}(k)}(x)}dx\right\}\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{h_{\hat{\theta}(k)}(x)}{g(x)}dx\right\}-1\,.$$

PROPOSITION 2. Assume (A.1)-(A.9). Then

$$(4.2)\quad D(f_{\hat{\tau}(k)}, g)=D(h_{\underline{\theta}(k)}, g)+\frac{1}{4\pi}(\hat{\theta}(k)-\underline{\theta}(k))'H_k(\hat{\theta}(k)-\underline{\theta}(k))$$
$$\times\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{h_{\theta(k)}(x)}{g(x)}dx\right\}+\text{lower order terms}\,,$$

where "lower order terms" means stochastically lower order terms as $n\to\infty$ compared with the second term in the right-hand side of (4.2) uniformly in $k$.

PROOF. By the definition of $\underline{\theta}(k)$, (A.3) and (A.6), noting that $\theta(k)$ is independent of $\sigma^2(k)$, we have

$$\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\bigg|_{\underline{\theta}(k)}g(x)dx=0\,,$$

$$\frac{\partial}{\partial\theta(k)}\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{h_{\theta(k)}(x)}{g(x)}dx\right\}\bigg|_{\underline{\theta}(k)}=0\,.$$

Therefore by Taylor expansion of the second equation of (4.1) with respect to $\theta(k)$ and noting (A.5) and (A.9) we have this proposition.

From Proposition 2 we can measure the goodness of $f_{\hat{\tau}(k)}(x)$, neglecting the lower order terms, by

$$(4.3)\quad M(f_{\hat{\tau}(k)}, g)=D(h_{\underline{\theta}(k)}, g)+\frac{1}{4\pi}(\hat{\theta}(k)-\underline{\theta}(k))'H_k(\hat{\theta}(k)-\underline{\theta}(k))$$
$$\times\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{h_{\theta(k)}(x)}{g(x)}dx\right\}$$

instead of $D(f_{\hat{\tau}(k)}, g)$. The first term on the right-hand side of (4.3) represents the bias between the spectral density $f_{\underline{\tau}(k)}(x)$ and the true spectral density $g(x)$. The second term represents the variance of estimation. To investigate the asymptotic behavior of $\hat{\theta}(k)$ we shall prepare the following lemma, which is essentially due to Grenander and Rosenblatt [6].

LEMMA 2. *Let $\psi(x)$ be a $k\times 1$ vector of functions on $[-\pi, \pi]$ each of which satisfies the Lipschitz condition of order 1, and $\psi(x)=\psi(-x)$. Then, under the assumptions* (A.1) *and* (A.2), *we have*

$$\begin{aligned} n\,\mathrm{E}&\left\{\int_{-\pi}^{\pi}\psi(x)(I_n(x)-g(x))dx\right\}\left\{\int_{-\pi}^{\pi}\psi(x)(I_n(x)-g(x))dx\right\}' \\ &=4\pi\int_{-\pi}^{\pi}\psi(x)\psi(x)'g(x)^2dx+\frac{\kappa_4}{\sigma^4}\int_{-\pi}^{\pi}\psi(x)g(x)dx \\ &\quad\times\int_{-\pi}^{\pi}\psi(x)'g(x)dx+O((\log n)^2/n)\ , \end{aligned}$$

*where $\psi(x)'$ stands for the transpose of the vector $\psi(x)$, and $O((\log n)^2/n)$ is a $k\times k$ matrix in which each element is $(\log n)^2/n$ in order.*

Now we introduce the following notations for simplicity;

$$A(x;\theta(k))=\frac{\partial^2 h_{\theta(k)}(x)^{-1}}{\partial\theta(k)\partial\theta(k)'}\ ,\qquad B(x;\theta(k))=\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)'}\ .$$

Using Lemma 2 we have the following theorem.

THEOREM 1. *Assume* (A.1)-(A.9). *Then*

$$\hat\theta(k)-\underline{\theta}(k)=-H_k^{-1}\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\bigg|_{\underline{\theta}(k)}(I_n(x)-g(x))dx+O_p(n^{-1})\ ,$$

*where $O_p(n^{-1})$ means a $k$ dimensional vector whose elements are at most of order $n^{-1}$ in probability.*

PROOF. By the definition of $\hat\theta(k)$, (A.3) and (A.6), noting that $\theta(k)$ is independent of $\sigma^2(k)$, we have

$$\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta(k)}h_{\theta(k)}(x)^{-1}\bigg|_{\hat\theta(k)}I_n(x)dx=\mathbf{0}\ ,\qquad \text{for sufficiently large } n\ . \tag{4.4}$$

Then (4.4) can be written as

$$\int_{-\pi}^{\pi}\left[\frac{\partial}{\partial\theta(k)}h_{\theta(k)}(x)^{-1}\bigg|_{\underline{\theta}(k)}+A(x;\tilde\theta(k))(\hat\theta(k)-\underline{\theta}(k))\right]I_n(x)dx=\mathbf{0}\ , \tag{4.5}$$

where $\tilde\theta(k)=\underline{\theta}(k)+\lambda(\hat\theta(k)-\underline{\theta}(k))$ and $\lambda$ is a bounded $k\times k$ matrix. Also by the definition of $\underline{\theta}(k)$ we have

$$\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta(k)}h_{\theta(k)}(x)^{-1}\bigg|_{\underline{\theta}(k)}g(x)dx=\mathbf{0}\ . \tag{4.6}$$

Therefore (4.5) and (4.6) yields

$$\hat\theta(k)-\underline{\theta}(k)=-\left[\int_{-\pi}^{\pi}A(x;\tilde\theta(k))I_n(x)dx\right]^{-1}\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\bigg|_{\underline{\theta}(k)}(I_n(x)-g(x))dx\ .$$

Putting $\psi(x)=(\partial h_{\theta(k)}(x)^{-1}/\partial\theta(k))|_{\underline{\theta}(k)}$ in Lemma 2, and noting (A.8), we have

$$\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\bigg|_{\underline{\theta}(k)}I_n(x)dx=\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\bigg|_{\underline{\theta}(k)}g(x)dx+O_p(1/\sqrt{n})\ .$$

Then we have $\hat{\theta}(k)=\underline{\theta}(k)+O_p(1/\sqrt{n})$. Therefore $\tilde{\theta}(k)=\underline{\theta}(k)+O_p(1/\sqrt{n})$. Since the derivative of $A(x;\theta(k))$ satisfies (A.9), we have

$$\hat{\theta}(k)-\underline{\theta}(k)=-\left[\int_{-\pi}^{\pi}A(x;\underline{\theta}(k))g(x)dx\right]^{-1}\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\bigg|_{\underline{\theta}(k)}(I_n(x)-g(x))dx$$
$$+O_p(1/\sqrt{n})V_k\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\bigg|_{\underline{\theta}(k)}(I_n(x)-g(x))dx\ ,$$

where $V_k\in l(k\times k)$, and $O_p(1/\sqrt{n})$ means a scalar quantity which is at most of order in probability $1/\sqrt{n}$. This completes the proof.

Using Lemma 2 and Theorem 1 and noting $\int_{-\pi}^{\pi}\frac{\partial h_{\theta(k)}(x)^{-1}}{\partial\theta(k)}\Big|_{\underline{\theta}(k)}g(x)dx=0$, we have the following proposition.

PROPOSITION 3. Assume (A.1)-(A.9). Then

$$n\,\mathrm{E}\,[\hat{\theta}(k)-\underline{\theta}(k)][\hat{\theta}(k)-\underline{\theta}(k)]'$$
$$=H_k^{-1}\left[4\pi\int_{-\pi}^{\pi}B(x;\underline{\theta}(k))g(x)^2dx+O((\log n)^2/n)\right]H_k^{-1}+O(1/\sqrt{n})\ .$$

*Remark* 2. Note that the covariance matrix of $\hat{\theta}(k)-\underline{\theta}(k)$ is independent of the cumulant $\kappa_4$.

PROPOSITION 4. Assume (A.1)-(A.10). Then

$$\int_{-\pi}^{\pi}B(x;\underline{\theta}(k))g(x)^2dx=\frac{\underline{\sigma}^2(k)}{2\pi}H_k+O(k^{-\beta})\ .$$

PROOF. We can rewrite

$$H_k=\int_{-\pi}^{\pi}A(x;\underline{\theta}(k))f_{\underline{\tau}(k)}(x)dx+\int_{-\pi}^{\pi}A(x;\underline{\theta}(k))(g(x)-f_{\underline{\tau}(k)}(x))dx\ .$$

Noting (A.10) we have

$$H_k=\int_{-\pi}^{\pi}A(x;\underline{\theta}(k))f_{\underline{\tau}(k)}(x)dx+O(k^{-\beta})\ .$$

Similarly we have

$$\int_{-\pi}^{\pi}B(x;\underline{\theta}(k))g(x)^2dx$$
$$=\int_{-\pi}^{\pi}B(x;\underline{\theta}(k))f_{\underline{\tau}(k)}(x)^2dx+\int_{-\pi}^{\pi}B(x;\underline{\theta}(k))(g(x)^2-f_{\underline{\tau}(k)}(x)^2)dx$$

$$=\int_{-\pi}^{\pi} B(x;\underline{\theta}(k))f_{\underline{\theta}(k)}(x)^2dx+O(k^{-\beta}) .$$

Notice that

$$A(x;\theta(k))h_{\theta(k)}(x)=B(x;\theta(k))h_{\theta(k)}(x)^2-\frac{\partial^2 \log h_{\theta(k)}(x)}{\partial\theta(k)\partial\theta(k)'} .$$

Since $\theta(k)$ is independent of the innovation we have

$$\int_{-\pi}^{\pi} \frac{\partial^2 \log h_{\theta(k)}(x)}{\partial\theta(k)\partial\theta(k)'}dx=0 .$$

Therefore we have

$$\int_{-\pi}^{\pi} A(x;\theta(k))h_{\theta(k)}(x)dx=\int_{-\pi}^{\pi} B(x;\theta(k))h_{\theta(k)}(x)^2dx . \tag{4.7}$$

Multiplying (4.7) by $\{\underline{\sigma}^2(k)/(2\pi)\}^2$, we have this proposition.

THEOREM 2. *Assume* (A.1)–(A.10). *Then*

$$\begin{aligned} &\mathrm{E}\,[(n/4\pi)H_k^{1/2}(\hat{\theta}(k)-\underline{\theta}(k))(\hat{\theta}(k)-\underline{\theta}(k))'H_k^{1/2}] \\ &\quad=\left(\frac{\underline{\sigma}^2(k)}{2\pi}\right)I_k+O(k^{-\beta})+O(1/\sqrt{n}) , \end{aligned} \tag{4.8}$$

*where $I_k$ is the $k\times k$ unit matrix.*

PROOF. By Proposition 3, (4.8) is equal to

$$\begin{aligned} &H_k^{-1/2}\left[\int_{-\pi}^{\pi} B(x;\underline{\theta}(k))g(x)^2dx+O((\log n)^2/n)\right]H_k^{-1/2} \\ &\qquad+H_k^{1/2}O(1/\sqrt{n})H_k^{1/2} . \end{aligned} \tag{4.9}$$

Remembering Proposition 4 and Lemma 1 together with (A.8) we have this theorem.

Theorem 2 implies the following corollary.

COROLLARY 1.

$$\mathrm{E}\left[\frac{n}{4\pi k}(\hat{\theta}(k)-\underline{\theta}(k))'H_k(\hat{\theta}(k)-\underline{\theta}(k))\right]=\frac{\underline{\sigma}^2(k)}{2\pi}+O(k^{-\beta})+O(1/\sqrt{n}) .$$

Now let $\psi(x)=(\psi_1(x),\cdots,\psi_k(x))'$ be a vector of continuous functions such that $\psi(x)=\psi(-x)$. Putting $F_j=\int_{-\pi}^{\pi}\psi_j(x)(I_n(x)-g(x))dx$, and using arguments similar to that of Theorem 4.1 in Brillinger [3] under the same condition (2.4) as in Proposition 1, we have the following lemma.

LEMMA 3. *Under the condition* (2.4),

$$\text{cumulant}\,\{F_{j_1},\cdots,F_{j_p}\}=O(n^{-p+1}) ,\qquad p=1,\cdots,8 ,$$

*where* $\{j_1, \cdots, j_p\} \subset \{1, \cdots, k\}$.

For the notational convenience we put $Y(k) = (Y_1, \cdots, Y_k)' = (\sqrt{n}/(\sqrt{2}\,\underline{\sigma}(k)))H_k^{1/2}(\hat{\theta}(k) - \underline{\theta}(k))$.

LEMMA 4. *Assume* (A.1)-(A.10). *Then*

$$\mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k)-\underline{\theta}(k))' H_k(\hat{\theta}(k)-\underline{\theta}(k)) - k\right\}^4$$
$$= 48k + 12k^2 + k^4\{O(k^{-\beta}) + O(1/\sqrt{n})\}\ .$$

PROOF. Consider

$$\mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k)-\underline{\theta}(k))' H_k(\hat{\theta}(k)-\underline{\theta}(k))\right\}^4 \tag{4.10}$$
$$= \mathrm{E}\left(\sum_{j=1}^{k} Y_j^2\right)^4$$
$$= \sum_{j_1,\cdots,j_8=1}^{k} \delta(j_1, j_2)\cdots\delta(j_7, j_8)\,\mathrm{E}\,(Y_{j_1}\cdots Y_{j_8})\ ,$$

where

$$\delta(j_1, j_2) = \begin{cases} 1\ , & j_1 = j_2 \\ 0\ , & j_1 \neq j_2\ . \end{cases}$$

It is known that

$$\mathrm{E}\,(Y_{j_1}\cdots Y_{j_8}) = \sum \mathrm{cum}\,(Y_{l_1};\ l_1 \in \upsilon_1)\cdots \mathrm{cum}\,(Y_{l_p};\ l_p \in \upsilon_p)\ , \tag{4.11}$$

where the summation extends over all partitions $(\upsilon_1, \cdots, \upsilon_p)$, $p = 1, \cdots, 8$, of $(j_1, \cdots, j_8)$. Here, $\mathrm{cum}\,(Y_{l_1};\ l_1 \in \upsilon_1)$ denotes the joint cumulant of $\{Y_{l_1}\}_{l_1 \in \upsilon_1}$. From Lemma 3, we can find the main order terms on the right-hand side of (4.11). They are the terms of the following type,

$$\mathrm{cum}\,(Y_{j_{P(1)}}, Y_{j_{P(2)}})\,\mathrm{cum}\,(Y_{j_{P(3)}}, Y_{j_{P(4)}})\,\mathrm{cum}\,(Y_{j_{P(5)}}, Y_{j_{P(6)}}) \tag{4.12}$$
$$\times\,\mathrm{cum}\,(Y_{j_{P(7)}}, Y_{j_{P(8)}})\ , \qquad \text{say}\ ,$$
$$\{= \mathrm{cum}\,(P(1), P(2))\,\mathrm{cum}\,(P(3), P(4))\,\mathrm{cum}\,(P(5), P(6))$$
$$\times\,\mathrm{cum}\,(P(7), P(8))\}\ ,$$

which contribute mainly in the sum (4.10). Here $P(\cdot)$ denotes the partition of $(1, \cdots, 8)$. Therefore the main order terms in (4.10) can be written as

$$\sum_{j_1}\cdots\sum_{j_8} \delta(j_1, j_2)\cdots\delta(j_7, j_8) \sum_{P} \mathrm{cum}\,(P(1), P(2)) \cdots \mathrm{cum}\,(P(7), P(8))\ . \tag{4.13}$$

By Theorem 2 we have

$$\text{cum}\,(P(r),\, P(j)) = \delta(r,\, j) + O(k^{-\beta}) + O(1/\sqrt{n}) \;. \tag{4.14}$$

Thus if we consider (4.14) and all partitions in the second sum in (4.13), we can show that (4.13) is equal to

$$k^4 + 12k^3 + 44k^2 + 48k + k^4\{O(k^{-\beta}) + O(1/\sqrt{n})\} \;. \tag{4.15}$$

In the same way we have

$$\begin{aligned} &\mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k) - \underline{\theta}(k))' H_k(\hat{\theta}(k) - \underline{\theta}(k))\right\}^3 \\ &\quad = k^3 + 6k^2 + 8k + k^3\{O(k^{-\beta}) + O(1/\sqrt{n})\} \;. \end{aligned} \tag{4.16}$$

$$\begin{aligned} &\mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k) - \underline{\theta}(k))' H_k(\hat{\theta}(k) - \underline{\theta}(k))\right\}^2 \\ &\quad = k^2 + 2k + k^2\{O(k^{-\beta}) + O(1/\sqrt{n})\} \;. \end{aligned} \tag{4.17}$$

$$\begin{aligned} &\mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k) - \underline{\theta}(k))' H_k(\hat{\theta}(k) - \underline{\theta}(k))\right\} \\ &\quad = k + k\{O(k^{-\beta}) + O(1/\sqrt{n})\} \;. \end{aligned} \tag{4.18}$$

Combining (4.15)–(4.18) we have this lemma.

The result of Lemma 4 means that if $k_n \to \infty$ as $n \to \infty$ and $k_n \leqq K_n$, then

$$p-\lim_{n\to\infty} \frac{n}{2\underline{\sigma}^2(k)k}(\hat{\theta}(k) - \underline{\theta}(k))' H_k(\hat{\theta}(k) - \underline{\theta}(k)) = 1 \;, \qquad \text{for } k_n \leqq k \leqq K_n \,.$$

Therefore, $M(f_{\hat{\tau}(k)},\, g)$ behaves like to

$$R(n,\, k) = \frac{k}{n} \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\frac{f_{\tau(k)}(x)}{g(x)}\,dx\right\} + D(h_{\underline{\theta}(k)},\, g) \;,$$

that is,

PROPOSITION 5. Assume (A.1)–(A.10). For $k_n \leqq k \leqq K_n$, such that $k_n \to \infty$,

$$p-\lim_{n\to\infty} M(f_{\hat{\tau}(k)},\, g)/R(n,\, k) = 1 \;.$$

DEFINITION 1. A sequence $\{k_n^*\}$ is defined by

$$R(n,\, k_n^*) = \min_{1 \leqq k \leqq K_n} R(n,\, k) \;.$$

Of course we can see $k_n^* \to \infty$ as $n \to \infty$.

THEOREM 3. *Assume* (A.1)–(A.10). *Then*

$$p-\lim_{n\to\infty}\{\max_{1\le k\le K_n}|(M(f_{\hat{\tau}(k)}, g)/R(n, k))-1|\}=0 .$$

PROOF. We shall evaluate

$$(4.19)\quad \sum_{1\le k\le K_n} \mathrm{E}\{(M(f_{\hat{\tau}(k)}, g)/R(n, k))-1\}^4$$
$$=\sum_{1\le k\le K_n} \mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k)-\underline{\theta}(k))'H_k(\hat{\theta}(k)-\underline{\theta}(k))-k\right\}^4$$
$$\times\left[\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}dx\right\}\right]^4\Big/\{nR(n, k)\}^4 .$$

By the definition of $k_n^*$ we can see

$$(4.20)\quad (\log k_n^*)/\{nD(f_{\underline{\tau}(k)}, g)\}=o(1) \qquad \text{for all } 1\leqq k\leqq \log k_n^* .$$

Using Lemma 4 and (A.10) we have

$$(4.21)\quad \sum_{\log k_n^*\le k\le K_n} \mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k)-\underline{\theta}(k))'H_k(\hat{\theta}(k)-\underline{\theta}(k))-k\right\}^4$$
$$\times\left[\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}dx\right\}\right]^4\Big/\{nR(n, k)\}^4$$
$$\leqq \sum_{\log k_n^*\le k\le K_n}\{12k^{-2}+48k^{-3}+O(k^{-\beta})+O(1/\sqrt{n})\}\{1+O(k^{-\beta})\} .$$

Since $\beta>1$, $K_n=o(\sqrt{n})$ and $\log k_n^*\to\infty$, (4.21) tends to zero as $n\to\infty$. Also we have

$$(4.22)\quad \sum_{1\le k<\log k_n^*} \mathrm{E}\left\{\frac{n}{2\underline{\sigma}^2(k)}(\hat{\theta}(k)-\underline{\theta}(k))'H_k(\hat{\theta}(k)-\underline{\theta}(k))-k\right\}^4$$
$$\times\left[\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}dx\right\}\right]^4\Big/\{nR(n, k)\}^4$$
$$\leqq \sum_{1\le k<\log k_n^*}\{12k^2+48k+O(k^{-\beta+4})+O(k^4/\sqrt{n})\}$$
$$\times\left[\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}dx\right\}\right]^4\Big/\{nD(f_{\underline{\tau}(k)}, g)\}^4 .$$

By (4.20) and (A.10), (4.22) is dominated by

$$(4.23)\quad \sum_{1\le k<\log k_n^*}\{12k^2+48k+O(k^{-\beta+4})+O(k^4/\sqrt{n})\}\{1+O(k^{-\beta})\}/(\log k_n^*)^4 ,$$

which tends to zero as $n\to\infty$. Therefore (4.19) tends to zero as $n\to\infty$. This completes the proof.

Using Theorem 3 the following corollary is easily shown.

COROLLARY 2. *For any integer valued random variable $\tilde{k}$, and for any $\varepsilon>0$,*

$$\lim_{n\to\infty} \mathrm{P}\,(M(f_{\tau(\tilde{k})}, g)/R(n, k_n^*) \geqq 1-\varepsilon)=1 .$$

Define an asymptotically efficient order selection extending the concept of Shibata [8] in the case of autoregressive model.

DEFINITION 2. We call an order selection $\tilde{k}$ asymptotically efficient if

$$p-\lim_{n\to\infty} M(f_{\hat{\tau}(\tilde{k})}, g)/R(n, k_n^*)=1 .$$

## 5. Asymptotically efficient order selection

Akaike [2] proposed a criterion AIC (Akaike's information criterion)

$$\text{AIC}\,(k)=-2\log\,(\text{maximum likelihood})+2k , \tag{5.1}$$

where $k$ is the number of independently adjusted parameters with in the model. A selection $\hat{k}$ is defined as the $\hat{k}$ which minimizes AIC $(k)$. Of course this criterion has been proposed for a Gaussian ARMA model, but it can be applied for our generalized situation.

Remembering (3.2),

$$\frac{\partial}{\partial\sigma^2(k)}\int_{-\pi}^{\pi}\left\{\log\,(\sigma^2(k)h_{\hat{\theta}(k)}(x)/2\pi)+\frac{I_n(x)}{(\sigma^2(k)/2\pi)h_{\hat{\theta}(k)}(x)}\right|_{\sigma^2(k)=\hat{\sigma}^2(k)}dx=0 .$$

Thus we have

$$\hat{\sigma}^2(k)=\int_{-\pi}^{\pi}\frac{I_n(x)}{h_{\hat{\theta}(k)}(x)}dx . \tag{5.2}$$

Also we define the following value

$$\underline{\sigma}^2(k)=\int_{-\pi}^{\pi}\frac{g(x)}{h_{\underline{\theta}(k)}(x)}dx . \tag{5.3}$$

Then the AIC for $f_{\tau(k)}(x)$ can be written as

$$\text{AIC}\,(k+1)=n\log\hat{\sigma}^2(k)+2k . \tag{5.4}$$

Minimizing AIC $(k+1)$ is equivalent to minimizing

$$n\exp\left\{\frac{1}{n}\,\text{AIC}\,(k+1)\right\}=n\hat{\sigma}^2(k)\exp\frac{2k}{n} . \tag{5.5}$$

By (5.3) we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} g(x)/f_{\underline{\tau}(k)}(x)dx=1 , \tag{5.6}$$

and

$$\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}dx\right\}=\underline{\sigma}^2(k)/\sigma^2 .$$

Thus we have the following proposition.

PROPOSITION 6. Assume (A.1)-(A.3), (A.5) and (A.6). Then

$$D(f_{\underline{\tau}(k)}, g)=\{\underline{\sigma}^2(k)-\sigma^2\}/\sigma^2 .$$

Now we define

$$s^2(k)=\int_{-\pi}^{\pi}\frac{I_n(x)}{h_{\underline{\theta}(k)}(x)}dx .$$

*Remark* 3. We can see

$$\begin{aligned}(5.7)\qquad & n\exp(2k/n)\hat{\sigma}^2(k)/\sigma^2\\ &=nR(n, k)+n\{\exp(2k/n)-1\}\{(\hat{\sigma}^2(k)-\sigma^2)/\sigma^2\}\\ &\quad+\{k-n(s^2(k)-\hat{\sigma}^2(k))/\sigma^2\}+n(s^2(k)-\underline{\sigma}^2(k))/\sigma^2\\ &\quad+n\left[\exp\left(\frac{2k}{n}\right)-1-k\left\{\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}dx\right)+1\right\}\Big/n\right]\\ &\quad+n .\end{aligned}$$

Since

$$\lim_{n\to\infty}\max_{1\leq k\leq K_n} n\left[\exp\left(\frac{2k}{n}\right)-1-k\left\{\exp\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}dx\right)+1\right\}\Big/n\right]\Big/\{nR(n, k)\}=0 ,$$

we shall show that the second, third and fourth terms on the right-hand side of (5.7) are negligible with respect to $nR(n, k)$.

LEMMA 5. *Assume* (A.1)-(A.10). *Then*

$$(5.8)\qquad p-\lim_{n\to\infty}\max_{1\leq k\leq K_n}\{k-n(s^2(k)-\hat{\sigma}^2(k))/\sigma^2\}/\{nR(n, k)\}=0 .$$

PROOF. By Proposition 6 and (A.10) we have

$$(5.9)\qquad \underline{\sigma}^2(k)-\sigma^2=O(k^{-\beta}) .$$

Notice that

$$(5.10)\qquad \int_{-\pi}^{\pi}\phi(x)I_n(x)dx=\int_{-\pi}^{\pi}\phi(x)g(x)dx+O_p(1/\sqrt{n}) ,$$

$$(5.11)\qquad \hat{\theta}(k)=\underline{\theta}(k)+O_p(1/\sqrt{n}) .$$

If we apply the Taylor theorem we have

$$(5.12)\quad \{s^2(k)-\hat{\sigma}^2(k)\}/\sigma^2 = \frac{1}{2\sigma^2}(\hat{\theta}(k)-\underline{\theta}(k))' \int_{-\pi}^{\pi} \frac{\partial^2 h_{\theta(k)}(x)^{-1}}{\partial\theta(k)\partial\theta(k)'}\bigg|_{\hat{\theta}(k)} I_n(x)dx(\hat{\theta}(k)-\underline{\theta}(k)) + \text{lower order terms}.$$

Noting (5.9)–(5.12) and using an argument essentially the same as that of Theorem 3 we have (5.8).

This lemma means that the third term on the right-hand side of (5.7) is negligible with respect to $nR(n, k)$.

LEMMA 6. *Assume* (A.1)–(A.10). *Then*

$$(5.13)\quad p-\lim_{n\to\infty}\max_{1\leq k\leq K_n} n\{\exp(2k/n)-1\}\{\hat{\sigma}^2(k)-\sigma^2\}/\{nR(n, k)\}=0.$$

PROOF. Notice that $n\{\exp(2k/n)-1\}=O(k)$, and that

$$|\hat{\sigma}^2(k)-\sigma^2|\leqq|\hat{\sigma}^2(k)-s^2(k)|+|s^2(k)-\underline{\sigma}^2(k)|+|\underline{\sigma}^2(k)-\sigma^2|.$$

By Lemma 5 we have

$$p-\lim_{n\to\infty}\max_{1\leq k\leq K_n} k(\hat{\sigma}^2(k)-s^2(k))/\{nR(n, k)\}=0.$$

Also by Proposition 6

$$\lim_{n\to\infty}\max_{1\leq k\leq K_n} k|\underline{\sigma}^2(k)-\sigma^2|/\{nR(n, k)\} \leqq \lim_{n\to\infty}\max_{1\leq k\leq K_n} k|\underline{\sigma}^2(k)-\sigma^2|/\{nD(f_{\underline{\tau}(k)}, g)\}=0.$$

While we have

$$(5.14)\quad \mathrm{E}\,\{k(s^2(k)-\underline{\sigma}^2(k))\}^2=\mathrm{E}\left[k\int_{-\pi}^{\pi}\{(I_n(x)-g(x))/h_{\underline{\theta}(k)}(x)\}dx\right]^2.$$

By Lemma 2, we see that the order of (5.14) is $O(k^2/n)$. Thus $\sum_{1\leq k\leq K_n} \mathrm{E}\,\{k(s^2(k)-\underline{\sigma}^2(k))\}^2/\{nR(n, k)\}^2=\sum_{1\leq k\leq K_n} O(k^2/n)/\{nR(n, k)\}^2\leqq\sum_{1\leq k\leq K_n} O(k^2/n)/k^2 \to 0$. This completes the proof.

LEMMA 7. *Assume* (A.1)–(A.6) *and* (A.10). *Then*

$$D(f_{\underline{\tau}(k)}, g)=\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{f_{\underline{\tau}(k)}(x)-g(x)}{f_{\underline{\tau}(k)}(x)}\right\}^2dx+\text{lower order terms}.$$

PROOF. Remembering (A.10) we can write

$$(5.15)\quad \log\frac{f_{\underline{\tau}(k)}(x)}{g(x)}=-\log\left[1+\frac{g(x)}{f_{\underline{\tau}(k)}(x)}-1\right]$$

$$=-\frac{g(x)}{f_{\underline{\theta}(k)}(x)}+1+\frac{1}{2}\left[\frac{g(x)}{f_{\underline{\theta}(k)}(x)}-1\right]^2+\text{lower order terms}\,.$$

Noting $\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{g(x)}{f_{\underline{\theta}(k)}(x)}dx=1$, and that

$$\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f_{\underline{\theta}(k)}(x)}{g(x)}dx\right\}$$
$$=\exp\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\frac{g(x)}{f_{\underline{\theta}(k)}(x)}-1\right]^2dx+\text{lower order terms}\right\}$$
$$=1+\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\frac{f_{\underline{\theta}(k)}(x)-g(x)}{f_{\underline{\theta}(k)}(x)}\right]^2dx+\text{lower order terms}\,,$$

we have this lemma.

Now we shall show that the fourth term in (5.7) is negligible with respect to $nR(n, k)$. Since the behavior of $\hat{k}$ is determined only by the differences of (5.7) it is sufficient to show that $n\{s^2(k)-\underline{\sigma}^2(k)-(s^2(k_n^*)-\underline{\sigma}^2(k_n^*))\}$, $k=1,\cdots, K_n$, are uniformly negligible. To do so we shall prepare the following lemma.

LEMMA 8. *Assume* (A.1)–(A.10). *Then*

$$p-\lim_{n\to\infty}\max_{1\leq k\leq K_n} n\left|(s^2(k)-\underline{\sigma}^2(k))\left(\frac{1}{\sigma^2}-\frac{1}{\underline{\sigma}^2(k)}\right)\right|\Big/\{nR(n, k)\}=0\,.$$

PROOF. We evaluate the following

$$\sum_{k=1}^{K_n}\mathrm{E}\,\{n(s^2(k)-\underline{\sigma}^2(k))(\underline{\sigma}^2(k)-\sigma^2)\}^2/\sigma^2\underline{\sigma}^2(k)nR(n, k)\}^2 \tag{5.16}$$
$$=\sum_{k=1}^{K_n}\mathrm{E}\left\{n\int_{-\pi}^{\pi}h_{\underline{\theta}(k)}(x)^{-1}(I_n(x)-g(x))dx\right\}^2(\underline{\sigma}^2(k)-\sigma^2)^2$$
$$/\{\sigma^2\underline{\sigma}^2(k)nR(n, k)\}^2\,.$$

By Lemma 2 and Proposition 6, (5.16) can be written as

$$\sum_{k=1}^{K_n}O(n)D(f_{\underline{\theta}(k)}, g)^2/\{nR(n, k)\}^2\,. \tag{5.17}$$

Since $R(n, k)^2\geqq D(f_{\underline{\theta}(k)}, g)^2$, (5.17) is dominated by

$$\sum_{k=1}^{K_n}O(n^{-1})=O(K_n/n)\to 0\,,\qquad \text{as } n\to\infty\,.$$

This completes the proof.

LEMMA 9. *Assume* (A.1)–(A.10). *Then*

$$p-\lim_{n\to\infty}\max_{1\leq k\leq K_n} n|s^2(k)-\underline{\sigma}^2(k)-(s^2(k_n^*)-\underline{\sigma}^2(k_n^*))|/\{nR(n, k)\}=0\,.$$

PROOF. By Lemma 8 if we show that

$$n\left\{\frac{s^2(k)-\sigma^2(k)}{\sigma^2(k)}-\frac{s^2(k_n^*)-\sigma^2(k_n^*)}{\sigma^2(k_n^*)}\right\}$$
$$=\frac{n}{2\pi}\int_{-\pi}^{\pi}\{f_{\varepsilon(k)}(x)^{-1}-f_{\varepsilon(k_n^*)}(x)^{-1}\}(I_n(x)-g(x))dx$$

is negligible with respect to $nR(n, k)$, then the proof is completed. Now we shall evaluate the following

$$\mathrm{E}\left[\int_{-\pi}^{\pi}\{f_{\varepsilon(k)}(x)^{-1}-f_{\varepsilon(k_n^*)}(x)^{-1}\}(I_n(x)-g(x))dx\right]^4. \tag{5.18}$$

Considering the identity (4.11) and Lemma 3, the main contributive terms in (5.18) are typically written as

$$\left[\mathrm{cum}\left\{\int_{-\pi}^{\pi}(f_{\varepsilon(k)}(x)^{-1}-f_{\varepsilon(k_n^*)}(x)^{-1})(I_n(x)-g(x))dx,\right.\right.$$
$$\left.\left.\int_{-\pi}^{\pi}(f_{\varepsilon(k)}(x)^{-1}-f_{\varepsilon(k_n^*)}(x)^{-1})(I_n(x)-g(x))dx\right\}\right]^2$$
$$=\left\{4\pi\int_{-\pi}^{\pi}(f_{\varepsilon(k)}(x)^{-1}-f_{\varepsilon(k_n^*)}(x)^{-1})^2g(x)^2dx\right\}^2\Big/n^2$$
$$+\text{lower order terms}. \tag{5.19}$$

Noting the formula $(a+b)^2\leqq 2(a^2+b^2)$, the main order term in (5.19) is not greater than

$$\left\{8\pi\int_{-\pi}^{\pi}(f_{\varepsilon(k)}(x)^{-1}-g(x)^{-1})^2g(x)^2dx\right.$$
$$\left.+8\pi\int_{-\pi}^{\pi}(f_{\varepsilon(k_n^*)}(x)^{-1}-g(x)^{-1})^2g(x)^2dx\right\}^2\Big/n^2. \tag{5.20}$$

Similarly (5.20) is not greater than

$$\frac{128\pi^2}{n^2}\left[\left\{\int_{-\pi}^{\pi}(f_{\varepsilon(k)}(x)^{-1}-g(x)^{-1})^2g(x)^2dx\right\}^2\right.$$
$$\left.+\left\{\int_{-\pi}^{\pi}(f_{\varepsilon(k_n^*)}(x)^{-1}-g(x)^{-1})^2g(x)^2dx\right\}^2\right]. \tag{5.21}$$

Therefore the proof is complete if we can show that

$$\sum_{1\leqq k\leqq K_n}n^2\left\{\int_{-\pi}^{\pi}(f_{\varepsilon(k)}(x)^{-1}-g(x)^{-1})^2g(x)^2dx\right\}^2\Big/\{nR(n,k)\}^4 \tag{5.22}$$

goes to zero. In fact (5.22) is bounded by

$$\sum_{1\leqq k<\log k_n^*}\left\{\int_{-\pi}^{\pi}(f_{\varepsilon(k)}(x)-g(x))^2/f_{\varepsilon(k)}(x)^2dx\right\}^2\Big/\{n^2D(f_{\varepsilon(k)},g)^4\} \tag{5.23}$$

$$+\sum_{\log k_n^*\le k\le K_n}\left\{\int_{-\pi}^{\pi}(f_{\hat{\tau}(k)}(x)-g(x))^2/f_{\hat{\tau}(k)}(x)^2dx\right\}^2\Big/\{6k^2D(f_{\hat{\tau}(k)},g)^2\}\,.$$

Noting Lemma 7 and $(\log k_n^*)^2<n^2D(f_{\hat{\tau}(k)},g)^2$ for all $1\le k<\log k_n^*$, the first term in (5.23) is bounded by

$$\sum_{1\le k<\log k_n^*}O(1)/(\log k_n^*)^2\to 0 \qquad \text{as } n\to\infty\,. \tag{5.24}$$

Noting Lemma 7 the second term in (5.23) is bounded by

$$\sum_{\log k_n^*\le k\le K_n}O(1)/6k^2\to 0 \qquad \text{as } n\to\infty\,.$$

This completes the proof.

Using these lemmas we have the following theorem.

THEOREM 4. *Assume* (A.1)–(A.10). *Then*

$$p-\lim_{n\to\infty}M(f_{\hat{\tau}(\hat{k})},g)/R(n,k_n^*)=1\,.$$

*That is, the order selection $\hat{k}$ is asymptotically efficient in the sense of Definition* 2.

PROOF. Remembering Lemmas 5–9 and noting that

$$\{n\exp(2\hat{k}/n)\}\hat{\sigma}^2(\hat{k})\le\{n\exp(2k_n^*/n)\}\hat{\sigma}^2(k_n^*)\,,$$

we have for any $\varepsilon>0$,

$$\lim_{n\to\infty}\mathrm{P}\,(R(n,\hat{k})/R(n,k_n^*)\le 1+\varepsilon)=1\,.$$

Of course by the definition of $k_n^*$ we have $R(n,\hat{k})/R(n,k_n^*)\ge 1$. Thus $p-\lim_{n\to\infty}R(n,\hat{k})/R(n,k_n^*)=1$. By Theorem 3 we have this theorem.

## 6. Concluding remarks

Needless to say all results in the previous sections can be applied for the fitting of ARMA spectral density model $f_{\tau(k)}(x)$ described in Example 1. Theorem 4 means that the order selection $\hat{k}$, minimizing Akaike's information criterion (5.4) constructed by Gaussian likelihood, is asymptotically efficient although Gaussianity of $\{X(t)\}$ is not assumed.

Consider an order selection $\hat{k}^{(\alpha)}$ which attains the minimum of

$$\{n\exp(\alpha k/n)\}\hat{\sigma}^2(k)\,, \qquad \text{for some } \alpha>0, \text{ with respect to } k\,. \tag{6.1}$$

Then using an argument essentially the same as that of Shibata [8] we can show that $\hat{k}^{(\alpha)}$ is asymptotically efficient if and only if $\alpha=2$ under the assumptions (A.1)–(A.10).


## Acknowledgement

The author wishes to thank the referee, Professor Y. Fujikoshi and Mr. R. Shibata for their many profound comments.


HIROSHIMA UNIVERSITY

## REFERENCES


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

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

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

[4] Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, New York.

[5] Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models, *Adv. Appl. Prob.*, **8**, 339-364.

[6] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, Wiley, New York.

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

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

[9] Taniguchi, M. (1979). On estimation of parameters of Gaussian stationary processes, *J. Appl. Prob.*, **16**, 575-591.

[10] Walker, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series, *J. Aust. Math. Soc.*, **4**, 363-384.