Ann. Inst. Statist. Math.
35 (1983), Part A, 415-423

ASYMPTOTIC MEAN EFFICIENCY OF A SELECTION OF REGRESSION VARIABLES

RITEI SHIBATA

(Received Aug. 10, 1982; revised Nov. 15, 1982)

Summary

To compare different procedures for selection of regression variables, a mean efficiency concept is introduced, which is an extension of the concept of efficiency previously introduced by the author (Shibata [13]). Without any stronger assumption, we can show that the FPE procedure or the AIC procedure or the C_p procedure are all shown to be asymptotically mean efficient, under the assumption that the number of regression variables be infinite or increase with the sample size.

1. Introduction

The present author showed (Shibata [13]) that when the loss function is the mean squared error of prediction or the squared error of estimated regression function, there is a lower bound in an asymptotic sense. Using that lower bound, an asymptotic efficiency of a selection procedures is defined as the limit of the ratio of the bound to the actual loss. It was also shown that the FPE procedure (Akaike [1]), the AIC procedure (Akaike, [2]) and the C_p procedure (Mallows [8]) are all asymptotically efficient.

However, the results of computer simulations (Shibata [12]) suggest that the above theory does not work so well in several cases. One of the reason is that the convergence is in the sense of probability and comparisons are not effective in small samples. Actually, in Shibata [12] comparisons were done by introducing a new concept of efficiency, namely the mean efficiency, by supposing that similar theorems will hold true even if the loss is replaced by its expectation in the definition of the efficiency.

This paper aims to give a rigorous proof of the above supposition

* This work was partly done during the time the author was staying in the Australian National University.

Key words: Selection of variables, regression analysis, multiple regression.

on the mean efficiency. Any stronger assumption is not needed for showing mean convergence.

We will use the same notations and assumptions as in Shibata [13]. Assume that the observational equation is

$$Y=\langle x, \beta\rangle+e \qquad x, \beta \in l_2 ,$$

where e is the error variable normally distributed with mean 0 and unknown variance $\sigma^2>0$, β is the vector of regression parameters, and l_2 is the Hilbert space of sequence of real numbers with the inner product $\langle\ ,\ \rangle$, and the norm $\|\cdot\|$. For convenience, we will call the above model the "true model" and the parameter β the "true parameter".

By $j=(j_1, j_2, \cdots, j_{k(j)})$, $(j_1<j_2<\cdots<j_{k(j)})$, we denote the model

$$Y=\langle x, \beta(j)\rangle+\varepsilon ,$$

where

$$\beta(j) \in V(j)=\{\beta(j);\ \beta(j)'=(0,\cdots, \beta_{j_1}, 0,\cdots,0, \beta_{j_2},\cdots, \beta_{j_{k(j)}}, 0,\cdots)\}$$

is the vector of regression parameters and $k(j)$ is the number of nonzero parameters, that is, the number of variables included in the model j.

Given n independent observations on Y at $x^{(1)}, x^{(2)},\cdots, x^{(n)}$, by fitting a model j we have the least squares estimates of nonzero parameters of β,

$$\hat{\beta}(j)'=(\hat{\beta}_{j_1},\cdots, \hat{\beta}_{j_{k(j)}})$$

which is a solution of

$$M_n(j)\hat{\beta}(j)=X(j)'y .$$

Here $y'=(y_1, y_2,\cdots, y_n)$ is the vector of observations,

$$X(j)=\{x_{\alpha j_i};\ 1\leqq\alpha\leqq n,\ 1\leqq i\leqq k(j)\}$$

is the $n\times k(j)$ design matrix generated by subvectors of

$$x^{(\alpha)'}=(x_{\alpha 1}, x_{\alpha 2},\cdots) , \qquad \alpha=1,\cdots, n ,$$

and $M_n(j)=X(j)'X(j)$ is the $k(j)\times k(j)$ information matrix. For convenience, the vector $\hat{\beta}(j)$ is sometimes considered as an infinite dimensional vector, putting undefined coordinates zeros.

In our problem, the choice of the loss function is crucial for discussing the goodness of a selection procedure. One of commonly used loss functions is the 0, 1 loss function, which takes the value 0 if the selection is correct, otherwise 1. However such loss function is not always appropriate. For instance, consider the case when the number of nonzero coordinates of β is finite but some of them are very close to zero. For such case, is it still meaningful to know correctly the true model? Furthermore, if β is truly infinite dimensional, then the correctness of the selection itself is not well defined. Therefore in this paper we adopt a loss function which well reflects the goodness of the final estimate $\hat{\beta}(\tilde{j})$ of β, where $\tilde{j}$ denotes a selection from a given family J_n of models.

One such reasonable loss function is the squared loss

$$L_n\{\beta, \hat{\beta}(\tilde{j})\} = \|X\beta - X\hat{\beta}(\tilde{j})\|^2 .$$

Here, X is the $n\times\infty$ design matrix generated by the vectors $x^{(1)},\cdots,$ $x^{(n)}$. We denote the risk for a non-random selection $\tilde{j}\equiv j$ by

$$R_n(j) = \mathrm{E}\, L_n\{\beta, \hat{\beta}(j)\} .$$

Consider the "risk minimum model" j^* which minimizes $R_n(j)$ in J_n. The model j^* can be thought of the best model in J_n, which balances the "bias" term and the "variance" term of

$$R_n(j) = \|X\beta - X\beta^*(j)\|^2 + k(j)\sigma^2 \tag{1.1}$$

where $\beta^*(j) = \mathrm{E}\,\hat{\beta}(j)$. We can not apply this model j^* in practice, since which depends on unknown parameters β and σ^2. But in Theorem 2.1 we will prove that $R_n(j^*)$ asymptotically gives a lower bound for the risk $R_n(\tilde{j})$ even when $\tilde{j}$ is a random selection which depends on the observations $y_1, y_2, \cdots, y_n$.

Let us define a mean efficiency by the ratio

$$\mathrm{eff}\,(\tilde{j}) = \frac{R_n(j^*)}{\mathrm{E}\, L_n\{\beta, \hat{\beta}(\tilde{j})\}} \tag{1.2}$$

and an asymptotic mean efficiency by the limit,

$$\mathrm{a.eff}\,(\tilde{j}) = \liminf_{n\to\infty} \mathrm{eff}\,(\tilde{j}) .$$

A selection $\tilde{j}$ with $\mathrm{a.eff}\,(\tilde{j})\equiv 1$, is called asymptotically mean efficient. Examples of such selection are given by the minimum FPE procedure, the minimum AIC procedure (Akaike [1], [2]), and the C_p procedure (Mallows [8]). On the other hand, the BIC procedure (Schwarz [9]) or the ϕ procedure (Hannan and Quinn [6]) can be shown not to be asymptotically mean efficient in our sense, even though they are consistent under the 0, 1 loss function.

2. Asymptotic mean efficiency for a large number of variables

The following theorem shows that an asymptotic lower bound for the risk $\mathrm{E}\,L_n\{\beta, \hat{\beta}(\tilde{j})\}$ is given by $R_n(j^*)$ when the true number of variables is infinite or increases with the sample size n. The condition (2.1) is easily satisfied in such cases. For example, the case of the selection of the number of variables, a sufficient condition is that β is infinite dimensional and both the largest number of variables in J_n and the minimum eigenvalue of the information matrix diverges to infinity as n tends to infinity. For more detailed discussion on the condition (2.1), see Shibata [13].

THEOREM 2.1. *Assume that for any model j in J_n, $k(j)<n$ and $M_n(j)$ is of full rank. If*

$$\sum_{j\in J_n} \delta^{R_n(j)} \tag{2.1}$$

converges to zero as n tends to infinity for any $0<\delta<1$, then for any selection procedure $\tilde{j}$ from J_n,

$$\liminf_{n\to\infty} \frac{\mathrm{E}\,L_n\{\beta, \hat{\beta}(\tilde{j})\}}{R_n(j^*)} \geqq 1 .$$

PROOF. From the definition of j^*, we have

$$\frac{\mathrm{E}\,L_n\{\beta, \hat{\beta}(\tilde{j})\}}{R_n(j^*)} \geqq \mathrm{E}\left[\frac{L_n\{\beta, \hat{\beta}(\tilde{j})\}}{R_n(\tilde{j})}\right] = 1 + \mathrm{E}\left[\frac{L_n\{\beta, \hat{\beta}(\tilde{j})\} - R_n(\tilde{j})}{R_n(\tilde{j})}\right] .$$

Put

$$\xi_j = \frac{L_n\{\beta, \hat{\beta}(j)\} - R_n(j)}{R_n(j)}$$

then it is sufficient to show $\mathrm{E}|\xi_{\tilde{j}}|$ converges to zero as $n\to\infty$. We first rewrite ξ_j as

$$\xi_j = \frac{\|X\hat{\beta}(j) - X\beta^*(j)\|^2 - k(j)\sigma^2}{R_n(j)} .$$

Since

$$\|X\hat{\beta}(j) - X\beta^*(j)\|^2$$

is distributed as $\sigma^2\chi^2_{k(j)}$, applying the Schwartz inequality and Lemma 2.1 of Shibata [13], [14] we have

$$\begin{aligned} \mathrm{E}\{|\xi_j| I_{(|\xi_j|>\delta)}\} &\leqq \Pr(|\xi_j|>\delta)^{1/2}\{\sigma^4\cdot 2k(j)/R_n(j)^2\}^{1/2} \\ &\leqq 2\sigma R_n(j)^{-1/2} \exp\{-R_n(j)\delta^2/(8\sigma^2)\} , \end{aligned} \tag{2.2}$$

where I_A is the indicator function of a set A. The sum of the right hand side of (2.2) over J_n converges to zero as $n\to\infty$, by the assumption. It is enough to note that

$$\mathrm{E}\,|\xi_{\tilde{j}}| \leqq \sum_{j\in J_n} \mathrm{E}\,\{|\xi_j| I_{(|\xi_j|>\delta)}\} + \delta \ .$$

The above theorem justifies the definition (1.2) of the asymptotic mean efficiency. The selection previously proposed by Shibata [13] as an asymptotically efficient selection was defined by the $\tilde{j}$ which minimizes

$$S_n(j) = n\hat{\sigma}^2(j) + 2k(j)\hat{\sigma}^2(j) \ . \tag{2.3}$$

Although asymptotically equivalent, for small samples the FPE procedure (Akaike [1]) behaves better, in which $\tilde{\sigma}^2(j) = n\hat{\sigma}^2(j)/\{n-k(j)\}$ is used in the place of the last $\hat{\sigma}^2(j)$ on the right hand side of (2.3). This is because $\hat{\sigma}^2(j)$ is a biased estimate of σ^2. Therefore, in this section we will only prove the asymptotic mean efficiency of the selection $\hat{j}$ which minimizes

$$\mathrm{FPE}\,(j) = n\hat{\sigma}^2(j) + 2k(j)\tilde{\sigma}^2(j) \ .$$

As is seen from the proof, the other procedures, like the AIC or the C_p, which are shown to be asymptotically efficient in Shibata [13] are all asymptotically mean efficient, too.

THEOREM 2.2. *Under the same assumptions of Theorem* 2.1, *if* $\max_{j\in J_n} k(j) = o(n)$, *then*

$$\text{a.eff}\,(\hat{j}) = \lim_{n\to\infty} \frac{R_n(j^*)}{\mathrm{E}\,L_n\{\beta, \hat{\beta}(\hat{j})\}} = 1 \ .$$

PROOF. The conculsion follows if the expectation and the variance of $R_n(\hat{j})/R_n(j^*)$ converge to 1 and 0, respectively. In fact, using the same technique as in the proof of Theorem 2.1 we can easily prove that

$$\lim_{n\to\infty} \mathrm{E}\left[\frac{L_n\{\beta, \hat{\beta}(\hat{j})\} - R_n(\hat{j})}{R_n(\hat{j})}\right]^2 = 0 \ .$$

Then the desired result follows. We will only prove the covergence of the expectation of $R_n(\hat{j})/R_n(j^*)$ to 1, as the proof for the convergence of the variance is very similar. Put

$$J^{(1)} = \left\{ j \in J_n ;\ \frac{R_n(j)}{R_n(j^*)} \leqq (1+\eta) \right\}$$

and $J^{(2)} = J_n - J^{(1)}$.

Then

$$\frac{\mathrm{E}\,R_n(\hat{j})}{R_n(j^*)} - 1 = \left[\sum_{j\in J^{(1)}} \left\{\frac{R_n(j)}{R_n(j^*)} \Pr(\hat{j}=j)\right\} - 1\right] + \sum_{j\in J^{(2)}} \left\{\frac{R_n(j)}{R_n(j^*)} \Pr(\hat{j}=j)\right\} \ .$$

The first term of the right hand side is bounded by η from the definition of $J^{(1)}$ and bounded away from $\{\Pr(\hat{j}\in J^{(1)})-1\}$, which converges to zero as $n\to\infty$ (see Shibata [11]). It is sufficient to show that the second term converges to zero for any $\eta>0$. We rewrite the FPE statistic as

$$\begin{aligned}(2.4)\qquad \mathrm{FPE}(j)=&R_n(j)+\{k(j)\sigma^2-\|X\hat{\beta}(j)-X\beta^*(j)\|^2\}\\&+[\{2k(j)n\}/\{n-k(j)\}][\hat{\sigma}^2(j)-\{n-k(j)\}\sigma^2/n]\\&+n\{s^2(j)-\sigma^2(j)\}+n\sigma^2\,,\end{aligned}$$

where $ns^2(j)=\|y-X\beta^*(j)\|^2$ and $\sigma^2(j)=\mathrm{E}\,s^2(j)$. For the second term of the right hand side of (2.4) we have, from Lemma 2.1 of Shibata [13], [14], and noting $R_n(j)\geqq k(j)\sigma^2$

$$\Pr\left\{\frac{|k(j)\sigma^2-\|X\hat{\beta}(j)-X\beta^*(j)\|^2|}{R_n(j)}>\delta\right\}\leqq 2\exp\{-\delta R_n(j)/(8\sigma^2)\}\ .$$

The tail probability of the third term of the right hand side of (2.4) can be evaluated in the same way. The fourth term can not be evaluated in the same way, but the differences can be done in the same manner (see Shibata [13]). As a results, the tail probability of

$$\varDelta=\frac{\mathrm{FPE}(j^*)-\mathrm{FPE}(j)}{R_n(j)}-\frac{R_n(j^*)-R_n(j)}{R_n(j)}$$

is evaluated as

$$\Pr(|\varDelta|>\delta)\leqq C\exp\{-\delta R_n(j)/(8\sigma^2)\}\ ,$$

for some constant $C>0$ and for large n. On the other hand, for j in $J^{(2)}$,

$$\begin{aligned}\Pr(\hat{j}=j)&\leqq\Pr\left\{\frac{\mathrm{FPE}(j^*)-\mathrm{FPE}(j)}{R_n(j)}\geqq 0\right\}\\&=\Pr\{\varDelta\geqq 1-R_n(j^*)/R_n(j)\}\\&\leqq\Pr[\varDelta\geqq\{\eta/(1+\eta)\}]\ .\end{aligned}$$

Combining these results, we have

$$\begin{aligned}(2.5)\qquad &\sum_{j\in J^{(2)}}\frac{R_n(j)}{R_n(j^*)}\Pr(\hat{j}=j)\\&\qquad\leqq\frac{C}{R_n(j^*)}\sum_{j\in J^{(2)}}R_n(j)\exp[-\eta R_n(j)/\{8\sigma^2(1+\eta)\}]\ .\end{aligned}$$

Here, $R_n(j^*)$ diverges to infinity since

$$\delta^{R_n(j^*)}\leqq\sum_{j\in J_n}\delta^{R_n(j)}\,,$$

and

$$\sum_{j \in J_n} R_n(j)\delta^{R_n(j)}$$

converges to zero as n tends to infinity, because of the assumption (2.1). Therefore (2.5) converges to zero and the proof is complete.

It should be noted that the generalized FPE procedure (FPE_α) considered by Bhansali and Downham [5] is asymptotically mean efficient if and only if $\alpha=2$. It can be proved similarly as in Shibata [11]. Here the generalized FPE statistic is defined as

$$\mathrm{FPE}_\alpha(j)=n\hat{\sigma}^2(j)+\alpha k(j)\tilde{\sigma}^2(j) .$$

Another interesting attempt to justify $\alpha=2$ is made by Leonard and Ord [7] in the context of preliminary testing.

To help understanding in the theorem, a part of results by computer simulations are given in Table 1. Samples are generated from the model

$$Y=-\log(1-x)+e$$

Table 1. Mean efficiency and the mean length of selected Fourier series; $f(x)=-\log(1-x)$, $n=400$, $K=163$, 100 simulations

σ	FPE_2	FPE_3	FPE_4	FPE_5	AIC	BIC	ϕ	S_n	C_p	k^*
0.01	0.90	0.91	0.89	0.88	0.88	0.76	0.89	0.54	0.92	0.99
	78.1	72.5	67.5	65.1	80.9	58.8	65.2	147.8	75.2	74.0
0.02	0.91	0.92	0.90	0.88	0.88	0.73	0.89	0.47	0.92	1.00
	63.4	58.8	54.9	52.8	66.2	45.9	53.5	140.8	61.1	60.0
0.03	0.88	0.91	0.88	0.85	0.87	0.71	0.87	0.45	0.91	0.99
	56.9	49.6	46.9	45.0	58.8	39.7	46.1	131.9	54.0	54.0
0.04	0.93	0.93	0.90	0.86	0.91	0.72	0.89	0.42	0.95	1.03
	51.4	45.1	41.5	39.8	52.7	35.2	40.8	130.0	49.1	48.0
0.05	0.90	0.90	0.87	0.82	0.88	0.70	0.87	0.41	0.90	0.99
	47.0	41.1	37.9	35.7	48.2	31.9	37.6	119.5	45.5	44.0
0.06	0.88	0.87	0.83	0.79	0.86	0.68	0.82	0.41	0.89	0.95
	43.2	37.8	34.6	32.9	44.0	29.3	34.4	109.0	41.7	42.0
0.07	0.86	0.88	0.84	0.80	0.85	0.66	0.84	0.40	0.88	0.99
	41.8	35.7	32.6	30.6	42.3	26.8	32.1	107.4	40.3	40.0
0.08	0.86	0.89	0.86	0.81	0.86	0.67	0.86	0.42	0.89	0.99
	39.8	33.7	31.4	29.0	39.9	25.4	31.3	96.8	37.1	37.0
0.09	0.89	0.91	0.85	0.80	0.89	0.68	0.85	0.33	0.92	1.03
	37.5	31.3	28.6	27.3	37.8	23.9	28.4	121.7	35.8	36.0
0.10	0.88	0.88	0.83	0.79	0.87	0.68	0.83	0.37	0.89	1.02
	35.0	30.5	27.0	25.7	35.1	22.8	27.0	102.8	33.6	34.0
0.20	0.86	0.84	0.78	0.72	0.86	0.62	0.80	0.29	0.87	1.00
	25.9	20.6	18.3	16.7	26.1	14.8	18.5	92.3	24.7	24.0
0.30	0.85	0.84	0.80	0.74	0.84	0.64	0.82	0.33	0.86	0.99
	21.1	17.2	14.8	13.2	21.3	11.6	15.1	67.0	20.8	20.0
0.40	0.80	0.82	0.79	0.73	0.80	0.66	0.80	0.33	0.81	0.97
	18.0	14.2	12.6	11.4	18.0	10.1	12.8	57.6	17.4	16.0

Table 1. Continued

σ	FPE_2	FPE_3	FPE_4	FPE_5	AIC	BIC	ϕ	S_n	C_p	k^*
0.50	0.85 14.7	0.84 11.9	0.79 10.4	0.73 9.5	0.85 14.8	0.65 8.7	0.80 10.7	0.33 49.1	0.86 14.4	1.03 14.0
0.60	0.83 14.0	0.84 10.5	0.77 8.9	0.72 8.1	0.83 14.0	0.63 7.2	0.79 9.4	0.26 56.2	0.84 13.6	1.03 13.0
0.70	0.83 11.6	0.80 9.4	0.75 8.2	0.71 7.4	0.83 11.6	0.66 6.8	0.77 8.5	0.30 43.7	0.83 11.5	0.98 12.0
0.80	0.79 11.8	0.84 8.8	0.78 7.7	0.72 6.9	0.79 11.8	0.67 6.3	0.79 7.9	0.34 37.0	0.80 11.5	1.02 10.0
0.90	0.79 10.6	0.78 7.8	0.75 6.7	0.69 6.0	0.79 10.6	0.63 5.4	0.76 6.9	0.21 54.7	0.79 10.6	0.99 10.0
1.00	0.82 9.4	0.82 7.3	0.76 6.1	0.72 5.6	0.82 9.4	0.65 5.1	0.78 6.3	0.26 40.3	0.82 9.2	1.03 9.0
2.00	0.72 6.3	0.75 4.8	0.72 4.0	0.66 3.5	0.72 6.3	0.62 3.1	0.73 4.3	0.20 33.9	0.72 6.4	1.00 6.0
3.00	0.66 4.9	0.78 3.4	0.77 2.9	0.75 2.7	0.66 4.9	0.72 2.4	0.78 3.0	0.17 29.9	0.64 5.0	1.03 4.0
4.00	0.67 4.2	0.80 2.9	0.79 2.5	0.78 2.3	0.67 4.2	0.74 2.1	0.80 2.7	0.18 24.2	0.68 4.1	1.00 4.0
5.00	0.60 3.9	0.71 2.4	0.76 2.2	0.73 2.0	0.60 3.9	0.63 1.8	0.77 2.3	0.15 24.0	0.60 3.9	1.04 3.0
6.00	0.51 3.8	0.63 2.5	0.63 2.1	0.58 1.8	0.51 3.8	0.56 1.6	0.66 2.3	0.11 29.4	0.51 3.8	0.98 3.0

for the points,

$$x_1=\delta/(n+1),\ x_2=2\delta/(n+1),\cdots,\ x_n=n\delta/(n+1)\ .$$

A model is selected from Fourier regression models;

$$Y=\sum_{l=0}^{k-1}\frac{\cos\{\pi l(x/\delta)\}}{l+1}\beta_{l+1}+\varepsilon\ ,\qquad k=1,2,\cdots,K\ .$$

Each model can be specified only by the length k of the Fourier series so that the notation k is used in the place of j. In this example, $\sigma=0.99$, $n=400$ and $K=163$. The true number of variables is infinite since the regression function in (2.6) has an infinite Fourier series expansion. The risk minimum length $k^*=k(j^*)$ changes from 3 to 74 as σ varies from 6.0 to 0.01. Efficiencies are obtained for the following selection procedures. A model k is selected so as to minimize the corresponding statistic;

$\mathrm{AIC}(k)=n\log\hat{\sigma}^2(k)+2k\ ,$ (Akaike [2])

$\mathrm{BIC}(k)=n\log\hat{\sigma}^2(k)+(\log n)k\ ,$ (Schwarz [9], Akaike [3])

$\phi(k)=n\log\hat{\sigma}^2(k)+2(\log\log n)k\ ,$ (Hannan and Quinn [6])

$S_n(k)=n\hat{\sigma}^2(k)+2k\hat{\sigma}^2(k)\ ,$ (Shibata [13])

$C_k=n\hat{\sigma}^2(k)+2k\tilde{\sigma}^2(K)\ ,$ (Mallows [8], C_p)

$\mathrm{FPE}_\alpha(k)=n\hat{\sigma}^2(k)+\alpha k\tilde{\sigma}^2(k)$.

(Akaike [1], Bhansali and Downham [5], Atkinson [4])

For $n=400$, $\phi(k)$ is approximately $\mathrm{FPE}_{3.58}(k)$ and $\mathrm{BIC}(k)$ is approximately $\mathrm{FPE}_{5.98}(k)$. Among these procedures, the FPE, the AIC, the S_n and the C_p procedures are only asymptotically mean efficient. The k^* denotes the selection which always takes the model k^*. The mean efficiency of the k^* is always 1, so that the estimated efficiencies indicate how accurate our simulations are.

High efficiency of the FPE or the AIC, especially of the C_p is in contrast with the low efficiency of the S_n procedure. This may suggest a goodness of the replacement $\tilde{\sigma}^2(k)$ of $\hat{\sigma}^2(k)$ in $S_n(k)$, furthermore of the replacement $\tilde{\sigma}^2(K)$. Although the FPE_2 and the FPE_3 are not so clearly discriminated, the efficiency of the FPE quickly goes down as α increase to 4.5 and 5.98. The above observation is valid only for the case $k^*\geqq 9$. Otherwise the efficiencies behave differently.

UNIVERSITY OF PITTSBURGH
TOKYO INSTITUTE OF TECHNOLOGY

REFERENCES

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

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

[3] Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure, *Ann. Inst. Statist. Math.*, A, **30**, 9-14.

[4] Atkinson, A. C. (1980). A note on the generalized information criterion for choice of a model, *Biometrika*, **67**, 413-418.

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

[6] Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of autoregression, *J. R. Statist. Soc.*, B, **41**, 190-195.

[7] Leonard, T. and Ord, K. (1976). An investigation of the *F*-test procedure as an estimation short-cut, *J. R. Statist. Soc.*, B, **38**, 95-98.

[8] Mallows, C. L. (1973). Some comments on C_p, *Technometrics*, **12**, 591-612.

[9] Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Statist.*, **6**, 461-464.

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

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

[12] Shibata, R. (1980b). Selection of the number of regression parameters in small sample cases, *Statistical Climatology, Developments in Atmospheric Science*, **13**, (eds. S. Ikeda et al.), Elsevier, 137-148.

[13] Shibata, R. (1981). An optimal selection of regression variables, *Biometrika*, **68**, 45-54.

[14] Shibata, R. (1982). Correction to "An optimal selection of regression variables", *Biometrika*, **69**.