The Annals of Statistics
1986, Vol. 14, No. 1, 315–325

ASYMPTOTICALLY EFFICIENT SELECTION OF THE ORDER BY THE CRITERION AUTOREGRESSIVE TRANSFER FUNCTION

BY R. J. BHANSALI

University of Liverpool

The autoregressive orders selected by the criterion autoregressive transfer function (CAT) of Parzen (1974), a new version, CAT*, of CAT introduced by Parzen (1977), and the CAT_2 criterion of Bhansali (1985) are shown to be asymptotically efficient in the sense defined by Shibata (1980, 1981). A generalization of the penalty function considered by Shibata (1980) is introduced. The order selected by the CAT_α criterion of Bhansali (1985), with any fixed $\alpha > 1$, is asymptotically efficient with respect to this generalized penalty function.

1. Introduction. In an important paper, Shibata (1980) derived an asymptotic lower bound for the mean squared error of prediction of an infinite-order Gaussian autoregressive process when the order of the fitted autoregression is selected from data. He then used this bound to show that the order selection by minimizing the final prediction error criterion (FPE) of Akaike (1970) and the information criterion, AIC, of Akaike (1973) is asymptotically efficient in the sense that for either of these criteria the lower bound for the mean squared error of prediction is attained asymptotically. Shibata (1981) extended his (1980) results by obtaining an asymptotic lower bound for the integrated relative squared error of an autoregressive spectral estimate when the fitted order is determined from data and demonstrating that if the fitted order is selected by minimizing FPE, or AIC, then again the lower bound is attained asymptotically.

As is well-known, an alternative method for autoregressive order selection involves the use of the criterion autoregressive transfer function (CAT) of Parzen (1974). Indeed, this criterion was introduced by Parzen for implementing precisely the same "nonparametric" autoregressive model fitting approach to time series modelling as considered by Shibata (1980, 1981). In this approach, the behavior of an observed time series of length T, say, is modelled by an autoregressive process of order k. However, k is interpreted not as an estimate of the order of a finite autoregressive process, but as providing an optimal finite-order approximation to a truly infinite-order process; in theoretical arguments, k is treated as a function of T and assumed to approach infinity simultaneously with T.

Parzen (1977) has introduced a new version, CAT*, say, of CAT by slightly modifying the definition of the penalty function used for introducing CAT. Also, Bhansali (1985) has suggested an extension of the latter penalty function and, by examining the question of bias, he has introduced a new criterion, the CAT_α

Received May 1984; revised May 1985.
AMS 1980 *subject classifications*. Primary 62M20; secondary 62M15.
Key words and phrases. Order selection, efficiency, prediction, spectral estimation, autoregression, time series.

criterion, in which $\alpha > 1$ is an arbitrary constant. The choice $\alpha = 2$ corresponds to adopting the same penalty function as used for introducing CAT. However, the functional form of the corresponding CAT_2 criterion is not the same as that of CAT.

In this paper, we show that the optimality property derived by Shibata (1980, 1981) for FPE and AIC also holds for CAT, CAT* and CAT_2. Our result thus establishes an asymptotic equivalence between these criteria when the generating process is an infinite-order autoregression. Note that for a finite-order autoregressive process, the asymptotic equivalence of these criteria is established by Bhansali (1985), who shows that the asymptotic distributions of their selected orders are the same. Empirical support for these asymptotic results is provided by Parzen (1977), Beamish and Priestley (1981), and Bhansali (1985), who report that these criteria frequently select the same orders.

A related reference is Taniguchi (1980), who has suggested that the optimality property of AIC derived by Shibata for autoregressive model fitting may be extended to the nonparametric fitting of autoregressive–moving-average models; however, the proofs given there are not rigorous and appear to be incomplete in their present form.

2. Definition of optimality of a selected order. Consider a discrete-time second-order stationary process $\{x_t\}$ with mean 0, covariance function $R(u) = E(x_t x_{t+u})$ and satisfying the following assumption:

ASSUMPTION 1. The process x_t is an infinite-order autoregressive process

$$\sum_{u=0}^{\infty} a(u) x_{t-u} = \varepsilon_t, \qquad a(0) = 1,$$

where ε_t is a sequence of independent normal variates, each with mean zero and variance σ^2, the $\{a(u)\}$ are absolutely summable real coefficients, i.e., $\Sigma|a(u)| < \infty$, such that the polynomial

$$A(z) = \sum_{u=0}^{\infty} a(u) z^u$$

is nonzero for $|z| \leq 1$. Also, x_t does not degenerate to a finite-order autoregressive process.

Having observed $X_1, \ldots, X_T$, suppose that the order k is selected from the range $1 < k < K_T$, where K_T satisfies the following assumption:

ASSUMPTION 2. $\{K_T\}$ $(T = 1, 2, \ldots)$ is a sequence of positive integers such that $K_T \to \infty$, $K_T^2/T \to 0$ as $T \to \infty$.

In Section 3, we also require the following assumption.

ASSUMPTION 3.

$$\sum_{u=-\infty}^{\infty} |u||R(u)| < \infty.$$

The kth-order least-squares estimator $\hat{\mathbf{a}}(k) = [\hat{a}_k(1),\dots,\hat{a}_k(k)]'$ of the autoregressive coefficients is a solution of the equation

$$\hat{\mathbf{R}}(k)\hat{\mathbf{a}}(k) = -\hat{\mathbf{r}}(k), \tag{2.1}$$

where $\hat{\mathbf{R}}(k) = [D^{(T)}(u,v)]\ (u,v = 1,\dots,k)$, $\hat{\mathbf{r}}(k) = [D^{(T)}(0,1),\dots,D^{(T)}(0,k)]'$,

$$D^{(T)}(u,v) = N^{-1}\sum_{t=K_T+1}^{T} X_{t-u}X_{t-v},$$

and $N = T - K_T$. The corresponding theoretical parameter $\mathbf{a}(k) = [a_k(1),\dots,a_k(k)]'$ will also be needed and is defined by

$$\mathbf{R}(k)\mathbf{a}(k) = -\mathbf{r}(k),$$

where $\mathbf{R}(k) = [R(u-v)]\ (u,v = 1,\dots,k)$ and $\mathbf{r}(k) = [R(1),\dots,R(k)]'$. Note that the $a_k(j)$ are the coefficients of the kth-order linear least-squares predictor of x_t given $x_{t-1},\dots,x_{t-k}$ and

$$\sigma^2(k) = \sum_{j=0}^{k} a_k(j)R(j), \qquad a_k(0) = 1,$$

is the corresponding mean squared error of prediction. An estimate of $\sigma^2(k)$ is given by

$$\hat{\sigma}^2(k) = \sum_{j=0}^{k} \hat{a}_k(j)D^{(T)}(0,j), \qquad \hat{a}_k(0) = 1.$$

Put

$$e_{t,k} = \sum_{j=0}^{k} a_k(j)x_{t-j}.$$

We may write

$$\hat{\mathbf{a}}(k) - \mathbf{a}(k) = -\hat{\mathbf{R}}(k)^{-1}\left\{\sum_{t=K_T+1}^{T} \mathbf{X}_t(k)e_{t,k}/N\right\}, \tag{2.2}$$

where $\mathbf{X}_t(k) = [x_{t-1},\dots,x_{t-k}]'$.

For an arbitrary infinite-dimensional vector $\boldsymbol{\delta} = [\delta_1,\delta_2,\dots]'$, let

$$\|\boldsymbol{\delta}\|_{\mathbf{R}} = \left\{\sum_{u=1}^{\infty}\sum_{v=1}^{\infty}\delta_u\delta_v R(u-v)\right\}^{1/2}$$

denote its norm with respect to the matrix $\mathbf{R} = [R(u-v)]\ (u,v = 1,2,\dots)$. Also, let $\mathbf{a} = [a(1),a(2),\dots]'$ denote an infinite-dimensional vector. In the sequel, it will often be convenient to think of $\mathbf{a}(k)$ and $\hat{\mathbf{a}}(k)$ also as infinite-dimensional vectors with $\hat{a}_k(j) = a_k(j) = 0\ (j > k)$.

Parzen (1974) suggested selecting k by minimising the criterion

$$\text{CAT}(k) = 1 - \{\hat{\sigma}_\infty^2/\tilde{\sigma}^2(k)\} + (k/T), \tag{2.3}$$

where

$$\tilde{\sigma}^2(k) = T(T-k)^{-1}\hat{\sigma}^2(k),$$

$$\hat{\sigma}_\infty^2 = 2\pi \exp\left[\left\{n^{-1}\sum_{j=1}^{n}\log I^{(T)}(\omega_j)\right\} + \gamma\right],$$

$$I^{(T)}(\lambda) = (2\pi T)^{-1}|\sum_{t=1}^{T} X_t \exp(-it\lambda)|^2,$$

$\gamma = 0.57721$, $\omega_j = 2\pi j/T$, and n is the largest integer not greater than $(T-1)/2$, i.e., $n = [(T-1)/2]$.

As discussed in Section 1, Parzen (1977) later suggested that k may also be selected by minimising the criterion

$$\text{CAT}^*(k) = T^{-1}\sum_{j=1}^{k}\tilde{\sigma}^{-2}(j) - \tilde{\sigma}^{-2}(k). \tag{2.4}$$

The CAT_α criterion considered by Bhansali (1985) is of the form

$$\text{CAT}_\alpha(k) = 1 - \{\hat{\sigma}_\infty^2/\hat{\sigma}^2(k)\} + \alpha(k/T), \tag{2.5}$$

where $\alpha > 1$ is an arbitrary constant, and k is selected by minimising this criterion.

We note that the CAT* criterion has the advantage that its definition does not depend upon $\hat{\sigma}_\infty^2$, which is the case with the CAT and CAT_α criteria. Also, AIC has been defined without explicitly requiring the evaluation of $\hat{\sigma}_\infty^2$.

When Assumptions 1 and 3 hold, $\hat{\sigma}_\infty^2$ converges in probability to σ^2 as $T \to \infty$; see Bhansali (1985). As shown by Hannan and Nicholls (1977), Assumption 3 is strictly not necessary for this result to hold and may be replaced by an assumption requiring that the spectral density function of x_t satisfy a Lipschitz condition. However, Assumption 3 is made for ease of exposition and it is used only for ensuring that $\hat{\sigma}_\infty^2$ is consistent for σ^2. Also, this assumption has not been made for establishing the asymptotic efficiency of the CAT* criterion.

Shibata (1980) argues that if the objective of fitting the autoregressive model is prediction then the goodness of the fitted model may be evaluated by mean squared error prediction as defined by the following penalty function:

$$\begin{aligned} Q_T(k) &= \|\hat{\mathbf{a}}(k) - \mathbf{a}\|_{\mathbf{R}}^2 \\ &= \|\mathbf{a}(k) - \mathbf{a}\|_{\mathbf{R}}^2 + \|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_{\mathbf{R}}^2. \end{aligned} \tag{2.6}$$

Shibata shows that if $\{k_T\}$ is a sequence of integers such that $1 \le k_T \le K_T$ and $k_T \to \infty$ as $T \to \infty$ then

$$\operatorname*{plim}_{T\to\infty}\{Q_T(k_T)/L_T(k_T)\} = 1,$$

where

$$L_T(k) = \sigma^2(k) - \sigma^2 + k\sigma^2/N.$$

Now, let $\{k_T^*\}$ $(T = 1, 2, \dots)$ be a sequence of positive integers at each of which the minimum of $L_T(k)$ with respect to k is attained, i.e.,

$$L_T(k_T^*) = \min_{1 \le k \le K_T} L_T(k) \qquad (T = 1, 2, \dots).$$

Then $k_T^* \to \infty$ as $T \to \infty$.

A remarkable result established by Shibata is that for any random variable $\tilde{k}$, possibly depending on $X_1, \dots, X_T$, and for any $\delta > 0$,

$$\lim_{T \to \infty} \operatorname{pr}\{Q_T(\tilde{k})/L_T(k_T^*) \ge 1 - \delta\} = 1.$$

Therefore, a selected order $\tilde{k}$, is defined to be asymptotically efficient if

$$\operatorname*{plim}_{T \to \infty} \{Q_T(\tilde{k})/L_T(k_T^*)\} = 1. \tag{2.7}$$

This definition is also adopted in this paper.

As is well-known, another motivation for fitting an autoregressive model is the estimation of the spectral density function,

$$f(\lambda) = \sigma^2(2\pi)^{-1}|\sum_{u=0}^{\infty} a(u)\exp(-iu\lambda)|^{-2},$$

of x_t. Let

$$\hat{f}_k(\lambda) = \hat{\sigma}^2(k)(2\pi)^{-1}|\sum_{u=0}^{k} \hat{a}_k(u)\exp(-iu\lambda)|^{-2}$$

denote the autoregressive spectral estimate corresponding to the fitted kth-order model. Shibata (1981) suggests adopting the integrated relative squared error,

$$J_T(k) = (2\pi)^{-1}\int_{-\pi}^{\pi} \{(\hat{f}_k(\lambda) - f(\lambda))/f(\lambda)\}^2 \, d\lambda,$$

as a penalty function for $\hat{f}_k(\lambda)$ and determines a lower bound, $2L_T(k_T^*)/\sigma^2$, for $J_T(k)$. Therefore, an order selection $\tilde{k}$ is defined as asymptotically efficient from the point of view of autoregressive spectral estimation if

$$\operatorname*{plim}_{T \to \infty} \{J_T(k)/L_T(k_T^*)\} = 2/\sigma^2.$$

This second definition of asymptotic efficiency of a selected order is, however, related to that introduced at (2.7). Suppose that a selected order, $\tilde{k}$, is a random variable such that

$$\operatorname*{plim}_{T \to \infty} \{L_T(\tilde{k})/L_T(k_T^*)\} = 1. \tag{2.8}$$

It follows from Shibata's results that $\tilde{k}$ is asymptotically efficient simultaneously from the point of view of prediction and spectral estimation. In particular, (2.8) holds for the order selected by the criterion

$$S_T(k) = (N + 2k)\hat{\sigma}^2(k) \tag{2.9}$$

considered by Shibata (1980, 1981), and, also, for the orders selected by FPE and AIC. We show in Section 3 that it also holds for the orders selected by CAT, CAT*, and CAT_2.

We note that although a relationship between AIC(k) and $\text{CAT}_2(k)$ has been established by Bhansali (1985), this relationship may not be employed for deducing our theorems as a direct consequence of Theorem 4.2 of Shibata (1980). Also, the remark made in lines 7 and 8 of Shibata (1980, page 162) concerning CAT does not apply to the criterion (2.3), but to the CAT_α criterion with $\alpha = 1$.

3. Asymptotic efficiency of CAT, CAT*, and CAT_2. For two arbitrary random variables X and Y, we write $X \le Y$ if $\text{pr}(Y - X \ge 0) = 1$.

We need the following lemma, which is an extension of Lemma 4.1 of Shibata (1980).

LEMMA 3.1. *Suppose that Assumptions* 1 *and* 2 *hold. Then*

$$\operatorname*{plim}_{T\to\infty} K_T \max_{1\le k\le K_T} \{|\hat{\sigma}^2(k) - \sigma^2|/(NL_T(k))\} = 0.$$

PROOF. We have

$$\begin{aligned} K_T \sum_{k=1}^{K_T} E&\left\{\left\| N^{-1} \sum_{t=K_T+1}^{T} \mathbf{X}_t(k)e_{t,k}\right\|^2 \Big/ (NL_T(k))\right\} \\ &\le K_T \sum_{k=1}^{K_T} k^{-1} \sum_{l=1}^{k} \sum_{j=0}^{k} \sum_{s=0}^{k} |a_k(j)||a_k(s)| \\ &\quad \times \left|E\left[\{D^{(T)}(l,j) - R(l-j)\}\{D^{(T)}(l,s) - R(l-s)\}\right]\right| \\ &\le MK_T^2/T, \end{aligned}$$

where M denotes a bounded constant, and converges to 0 as $T \to \infty$. Hence, by (2.2) and Lemma 3.3 of Shibata (1980),

$$\operatorname*{plim}_{T\to\infty} K_T \max_{1\le k\le K_T} \left\{\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|^2_{\hat{\mathbf{R}}(k)} \Big/ (NL_T(k))\right\} = 0.$$

The lemma therefore follows by an adaptation of the proof of Lemma 4.1 of Shibata (1980). □

Let $\hat{k}_\alpha$ denote the order selected by minimizing the CAT_α criterion (2.5). The asymptotic efficiency of $\hat{k}_2$ is established below.

THEOREM 3.1. *Suppose that Assumptions* 1, 2, *and* 3 *hold. Then*

$$\operatorname*{plim}_{T\to\infty} \{Q_T(\hat{k}_2)/L_T(k_T^*)\} = 1$$

and

$$\operatorname*{plim}_{T\to\infty} \{J_T(\hat{k}_2)/L_T(k_T^*)\} = 2/\sigma^2.$$

PROOF. We have, for all $T \geq 1$,

$$\left[\left(\hat{\sigma}_\infty^2/\hat{\sigma}^2(k_T^*)\right)\left(1 - \{\hat{\sigma}^2(k_T^*)/\hat{\sigma}^2(\hat{k}_2)\}\right) + 2T^{-1}(\hat{k}_2 - k_T^*)\right] \leq 0, \tag{3.1}$$

with probability 1, because $\mathrm{CAT}_2(\hat{k}_2) \leq \mathrm{CAT}_2(\hat{k}_T^*)$. Now as $T \to \infty$, $\hat{\sigma}_\infty^2/\hat{\sigma}^2(k_T^*)$ converges in probability to 1. Also as $T \to \infty$, $\hat{\sigma}^2(\hat{k}_2)$ converges in probability to a bounded positive constant, because $D^{(T)}(0,0) > \hat{\sigma}^2(\hat{k}_2) \geq \hat{\sigma}^2(K_T)$, where $D^{(T)}(0,0)$ converges in probability to $R(0)$ and $\hat{\sigma}^2(K_T)$ to σ^2. Therefore, (3.1) implies that

$$\lim_{T\to\infty} \mathrm{pr}\left\{S_T(\hat{k}_2) - S_T(k_T^*) + 2k_T^*\left(\hat{\sigma}^2(k_T^*) - \hat{\sigma}^2(\hat{k}_2)\right) \leq 0\right\} = 1.$$

The theorem may now be established from Lemma 3.1 and an argument similar to that used by Shibata (1980) for proving his Theorem 4.1 by demonstrating that

$$\operatorname*{plim}_{T\to\infty} \{L_T(\hat{k}_2)/L_T(k_T^*)\} = 1. \qquad \square$$

Let $\hat{k}_C$ denote the order selected by minimizing CAT. The asymptotic efficiency of $\hat{k}_C$ is established below.

THEOREM 3.2. *Suppose that Assumptions* 1, 2, *and* 3 *hold. Then*

$$\operatorname*{plim}_{T\to\infty} \{Q_T(\hat{k}_C)/L_T(k_T^*)\} = 1$$

and

$$\operatorname*{plim}_{T\to\infty} \{J_T(\hat{k}_C)/L_T(k_T^*)\} = 2/\sigma^2.$$

PROOF. On arguing as in the proof of Theorem 3.1, we have, because $\mathrm{CAT}(\hat{k}_C) \leq \mathrm{CAT}(k_T^*)$,

$$\lim_{T\to\infty} \mathrm{pr}\Big\{\Big[\left(S_T(\hat{k}_C) - S_T(k_T^*)\right) + 2k_T^*\left(\hat{\sigma}^2(k_T^*) - \hat{\sigma}^2(\hat{k}_C)\right) + \hat{k}_C\left(\hat{\sigma}^2(k_T^*) - \hat{\sigma}^2(\hat{k}_C)\right)\Big] \leq 0\Big\} = 1.$$

The theorem may therefore be established by demonstrating that the last result implies

$$\operatorname*{plim}_{T\to\infty} \{L_T(\hat{k}_C)/L_T(k_T^*)\} = 1. \qquad \square$$

Let $\tilde{k}_C$ denote the order selected by minimizing the CAT* criterion (2.4). We have

$$\mathrm{CAT}^*(k) = T^{-2} \sum_{j=1}^{k} (T-j)\hat{\sigma}^{-2}(j) - T^{-1}(T-k)\hat{\sigma}^{-2}(k).$$

We need the following lemmas.

LEMMA 3.2. *Suppose that Assumptions* 1 *and* 2 *hold, and let*

$$U_{1T} = |\hat{\sigma}^2(\tilde{k}_C)\hat{\sigma}^2(k_T^*) \sum_{j=1}^{\tilde{k}_C} \hat{\sigma}^{-2}(j) - \tilde{k}_C\hat{\sigma}^2(\tilde{k}_C)|, \tag{3.2}$$

$$U_{2T} = |\hat{\sigma}^2(\tilde{k}_C)\hat{\sigma}^2(k_T^*) \sum_{j=1}^{k_T^*} \hat{\sigma}^{-2}(j) - k_T^*\hat{\sigma}^2(k_T^*)|. \tag{3.3}$$

Then

$$\operatorname*{plim}_{T\to\infty} \{U_{jT}/NL_T(\tilde{k}_C)\} = 0 \qquad (j = 1,2). \tag{3.4}$$

PROOF. We need only consider (3.4) for $j = 1$, since the proof for $j = 2$ is similar. We have

$$U_{1T}/(NL_T(\tilde{k}_C)) \le F_1 + F_2 + F_3,$$

where

$$F_1 = \{NL_T(\tilde{k}_C)\}^{-1}\tilde{k}_C|\hat{\sigma}^2(k_T^*) - \sigma^2|,$$

$$F_2 = \{NL_T(\tilde{k}_C)\}^{-1} \sum_{j=1}^{\tilde{k}_C} |\sigma^2(j) - \sigma^2|,$$

$$F_3 = \{NL_T(\tilde{k}_C)\}^{-1} \sum_{j=1}^{\tilde{k}_C} |\hat{\sigma}^2(j) - \sigma^2(j)|.$$

Now

$$\begin{aligned} F_1 &\le \tilde{k}_C|\hat{\sigma}^2(k_T^*) - \sigma^2|/(NL_T(k_T^*)) \\ &\le K_T \max_{1\le k\le k_T} \{|\hat{\sigma}^2(k) - \sigma^2|/(NL_T(k))\} \end{aligned}$$

and converges to 0 in probability as $T \to \infty$ by Lemma 3.1. Also

$$F_3 \le \max_{1\le k\le k_T} |\hat{\sigma}^2(j) - \sigma^2(j)|$$

and converges to 0 in probability as $T \to \infty$ by Lemma 2.1 of Shibata (1981). Finally, consider F_2. Since $k_T^* \to \infty$ as $T \to \infty$ and $\{NL_T(\tilde{k}_C)\}^{-1} \le$

$\{NL_T(k_T^*)\}^{-1} \le (k_T^*)^{-1}$, we have, for all $T > T_0$, say,

$$F_2 \le (k_T^*)^{-1} \sum_{j=1}^{k_T^*} |\sigma^2(j) - \sigma^2|$$

and converges to 0 as $T \to \infty$ because $\sigma^2(j) \to \sigma^2$ as $j \to \infty$. □

LEMMA 3.3. *Suppose that Assumptions* 1 *and* 2 *hold, and put*

$$(3.5) \quad U_{3T} = (\tilde{k}_C - k_T^*)\{\hat{\sigma}^2(\tilde{k}_C) + \hat{\sigma}^2(k_T^*)\} - 2\tilde{k}_C\hat{\sigma}^2(\tilde{k}_C) + 2k_T^*\hat{\sigma}^2(k_T^*).$$

Then

$$\operatorname*{plim}_{T\to\infty} \{|U_{3T}|/(NL_T(\tilde{k}_C))\} = 0.$$

PROOF. We have

$$|U_{3T}|/(NL_T(\tilde{k}_C)) \le 4K_T \max_{1 \le k \le k_T} \{|\hat{\sigma}^2(k) - \sigma^2|/NL_T(k)\}$$

and converges to 0 in probability as $T \to \infty$ by Lemma 3.1. □

LEMMA 3.4. *Suppose that Assumptions* 1 *and* 2 *hold and let*

$$(3.6) \qquad U_{4T} = T^{-1}\hat{\sigma}^2(\tilde{k}_C)\hat{\sigma}^2(k_T^*) \sum_{j=1}^{\tilde{k}_C} j\hat{\sigma}^{-2}(j)$$

and

$$(3.7) \qquad U_{5T} = T^{-1}\hat{\sigma}^2(\tilde{k}_C)\hat{\sigma}^2(k_T^*) \sum_{j=1}^{k_T^*} j\hat{\sigma}^{-2}(j).$$

Then

$$\operatorname*{plim}_{T\to\infty} \{|U_{jT}|/(NL_T(\tilde{k}_C))\} = 0 \qquad (j = 4,5).$$

PROOF. The lemma follows directly from Lemma 3.2, by noting that $K_T/T \to 0$ as $T \to \infty$ and $1 \le \tilde{k}_C$, $k_T^* \le K_T$. □

THEOREM 3.3. *Suppose that Assumptions* 1 *and* 2 *hold. Then*

$$\operatorname*{plim}_{T\to\infty} \{Q_T(\tilde{k}_C)/L_T(k_T^*)\} = 1$$

and

$$\operatorname*{plim}_{T\to\infty} \{J_T(\tilde{k}_C)/L_T(k_T^*)\} = 2/\sigma^2.$$

PROOF. On arguing as in the proof of Theorem 4.1, we have, because $\mathrm{CAT}^*(\tilde{K}_C) \le \mathrm{CAT}^*(k_T^*)$,

$$\lim_{T\to\infty} \mathrm{pr}\{S_T(k_C) - S_T(k_T^*) + U_T \le 0\} = 1,$$

where

$$U_T = U_{1T} - U_{2T} + U_{3T} + U_{4T} - U_{5T}$$

and U_{jT} ($j = 1, \dots, 5$) are as in Lemmas 3.2–3.4. The theorem may therefore be established from these lemmas by demonstrating that the last result implies

$$\operatorname*{plim}_{T \to \infty} \{L_T(\tilde{k}_C)/L_T(k_T^*)\} = 1. \qquad \square$$

4. Discussion. As in Shibata (1980), let $\hat{k}^{(\alpha)}$ be the order selected by minimizing the criterion

$$(4.1) \qquad S_T^{(\alpha)}(k) = (N + \alpha k)\hat{\sigma}^2(k) \qquad (1 \le k \le K_T),$$

where $\alpha > 0$ is on arbitrary constant. It follows from Theorem 3.1 that the order selected by minimising CAT is asymptotically equivalent to $\hat{k}^{(2)}$ rather than to $\hat{k}^{(1)}$ as suggested by Shibata (1980).

For any fixed α, $\hat{k}^{(\alpha)}$ and $\tilde{k}_\alpha$—the order selected by minimizing the CAT_α criterion—are asymptotically equivalent. Hence, by repeating the arguments of Shibata, it follows that if $\alpha \neq 2$ and, as $k \to \infty$, $\{\sigma^2(k) - \sigma^2\} \to 0$ geometrically then $\tilde{k}_\alpha$ is not asymptotically efficient in the sense defined in Section 2. However, $\tilde{k}_\alpha$ is still asymptotically efficient if, as $k \to \infty$, $\sigma^2(k) - \sigma^2$, goes down to 0 exponentially, which is the case if $f(\lambda)$ coincides in $-\pi \le \lambda \le \pi$ almost everywhere with a function that is analytic for real λ and has no real zeroes; see Grenander and Szegö (1958). In particular, if x_t is a Gaussian autoregressive moving average process of order (p, q), with $q > 0$, to ensure that it does not degenerate to a finite autoregression, then $\sigma^2(k) \to \sigma^2$ exponentially as $k \to \infty$ and Assumptions A1 and A3 hold. Thus, for this important class of processes, $\tilde{k}_\alpha$ is still asymptotically efficient, in the sense defined in Section 2, for any $\alpha > 1$.

The order of decrease of $\sigma^2(k)$ to σ^2 is of course usually unknown. However, whether x_t is an autoregressive process of infinite, or finite, order is also usually unknown. If the order is finite then a choice of $\alpha = 2$ is not necessarily optimal because it leads to an inconsistent estimator of the order, and, also, because, with a finite T, no one choice of α is always optimal for all processes and all values of T; see Bhansali (1985).

We finally note that if instead of (2.6), a generalized penalty function

$$Q_T^{(\alpha')}(k) = \|\mathbf{a}(k) - \mathbf{a}\|_{\mathbf{R}}^2 + \alpha'\|\hat{\mathbf{a}}(k) - \mathbf{a}(k)\|_{\mathbf{R}}^2$$

is defined, where $\alpha' > 0$ is an arbitrary constant, then the arguments given in Section 3 show that for $\alpha = \alpha' + 1$ the order selected by minimising the CAT_α criterion is asymptotically efficient with respect to this generalized penalty function, and so are the orders selected by minimising the FPE_α and AIC_α criteria of Bhansali and Downham (1977) and Akaike (1979). A motivation for considering this generalized penalty function is that by varying α' the two terms on the right of (2.6) may be given unequal weights; see Bhansali (1985) for a discussion of the reasons for considering this possibility. The question of how best to choose α is considered by several authors; see, for example, Atkinson (1980), Smith and Spiegelhalter (1980), Akaike (1979), Shibata (1983), and Bhansali (1979).

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.* (B. N. Petrov and F. Csaki, eds.) 267–281, Akademia Kiado, Budapest.

AKAIKE, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika* **66** 237–242.

ATKINSON, A. C. (1980). A note on the generalized information criteria for choice of a model. *Biometrika* **67** 413–418.

BEAMISH, N. and PRIESTLEY, M. B. (1981). A study of autoregressive and window spectral estimation. *Appl. Statist.* **30** 41–58.

BHANSALI, R. J. (1979). A mixed spectrum analysis of the Lynx data. *J. Roy. Statist. Soc. A* **142** 199–209.

BHANSALI, R. J. (1985). The criterion autoregressive transfer function of Parzen. To appear in *J. Time Series Analysis.*

BHANSALI, R. J. and DOWNHAM, D. Y. (1977). Some properties of the order of an autoregressive model selected by a generalisation of Akaike's FPE criterion. *Biometrika* **64** 547–551.

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

HANNAN, E. J. and NICHOLLS, D. F. (1977). The estimation of the prediction error variance. *J. Amer. Statist. Assoc.* **72** 834–840.

PARZEN, E. (1974). Some recent advances in time series modelling. *IEEE Trans. Automat. Control* **19** 723–730.

PARZEN, E. (1977). Multiple time series: Determining the order of approximating autoregressive schemes. In *Multivariate Analysis*, **IV** (P. R. Krishnaiah, ed.) 283–295. North-Holland, Amsterdam.

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.

SHIBATA, R. (1983). A theoretical view of the use of AIC. In *Time Series Analysis: Theory and Practice* 4 (O. D. Anderson, ed.) North-Holland, Amsterdam.

SMITH, A. F. M. and SPIEGELHALTER, D. J. (1980). Bayes factors and choice criteria for linear models. *J. Roy. Statist. Soc. B* **42** 213–220.

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

DEPARTMENT OF STATISTICS
AND COMPUTATIONAL MATHEMATICS
UNIVERSITY OF LIVERPOOL
POB 147
LIVERPOOL L69 3BX
ENGLAND