KYBERNETIKA — VOLUME 17 (1981), NUMBER 6

ESTIMATING THE DIMENSION OF A LINEAR MODEL

JIŘÍ ANDĚL, MANUEL GARRIDO PEREZ, ANTONIO INSUA NEGRAO

A method for consistent estimation of the order of a linear model is derived in the paper. The procedure is analogous to modern criteria which are used in time series analysis. Some results of a simulation of polynomial regression are presented.

1. INTRODUCTION

Consider a regression model

$$Y_i = \beta_0 + \beta_1 x_i + \ldots + \beta_p x_i^p + e_i\,, \quad i = 1, 2, \ldots, N\,,$$

where $\mathbf{e} = (e_1, \ldots, e_N)' \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, $x_1, \ldots, x_N$ are given numbers and $\beta_0, \ldots, \beta_p, \sigma^2$ are unknown parameters such that $\beta_p \neq 0$, $\sigma^2 > 0$. The problem is to estimate the number $p + 1$ of regression parameters $\beta_0, \ldots, \beta_p$, when the pairs $(Y_1, x_1), \ldots$ $\ldots, (Y_N, x_N)$ are given. Usually, only indirect methods for determining the number of parameters are used. Such procedures are based on a set of tests of significance concerning the estimates of $\beta_0, \ldots, \beta_p$. However, the application of a long series of tests is rather an art than an objective statistical method. The statisticians also considered the estimating of the order as a multiple decision problem (see Anderson [4], for example). But it seems that these results have not become popular.

Another idea was proposed by Mallows [6]. Consider a linear model with p unknown regression parameters. If p grows, the bias in determining mean value is reduced, whereas the variances of estimators of parameters are larger. Denote s_p^2 the unbiased estimator for σ^2 in the model with p parameters and $\hat{\sigma}^2$ a suitable estimator for σ^2. Mallows advises to take the model which minimizes

$$C_p = (N - p)\, s_p^2/\hat{\sigma}^2 + 2p - N\,.$$

Similar problems appear also in time series analysis. Let $X_1, \ldots, X_N$ be a stationary autoregressive process generated by

(1) $$X_t = a_1 X_{t-1} + \ldots + a_p X_{t-p} + e_t ,$$

where e_t are again independent $N(0, \sigma^2)$ variables. The modern procedures for determining p are based on following ideas.

Assume that $0 \leqq p \leqq K$, where K is a given number. Denote s_k^2 an estimator of σ^2 in model (1) when k parameters $a_1, \ldots, a_k$ are taken into account. Usually, s_k^2 is the maximum likelihood estimator of σ^2. For $N \to \infty$ one can expect that s_k^2 approaches to σ^2 if $k \geqq p$, whereas s_k^2 remains larger than σ^2 if $k < p$. Nevertheless, the random behaviour of s_k^2 does not allow to determine the beginning of the asymptotically constant part of the function s_k^2, $k = 0, 1, \ldots, K$. The same problems arise in the variate difference method (see Anderson [4]).

Introduce a function

$$g_N(k) = s_k^2(1 + q_{k,N}) , \quad k = 0, 1, \ldots, K ,$$

where $q_{k,N}$ penalizes the growing number k of parameters in the model. Assume that $q_{k,N} \to 0$ as $N \to \infty$ for every fixed $k = 0, 1, \ldots, K$ and that $q_{k,N}$ is an increasing function of k, when N is fixed. Then the inequality $g_N(k) > g_N(p)$ for $k < p$ will asymptotically hold and, for a properly chosen $q_{k,N}$, the values of $g_N(k)$ for $k > p$ will also be greater than $g_N(p)$. For this reason we can estimate p by such a value $k = \hat{p}$, which minimizes the function $g_N(k)$, $k = 0, 1, \ldots, K$. Many authors use $\ln g_N(k) = G_N(k)$ instead of $g_N(k)$. Then they have the function

$$G_N(k) = \ln s_k^2 + Q_{k,N} ,$$

where $Q_{k,N} = \ln (1 + q_{k,N})$. For example, Akaike's FPE criterion [1] as well as his AIC criterion [2] lead to

(2) $$G_N(k) = \ln s_k^2 + 2kN^{-1} .$$

Schwarz [8] and Rissanen [7] derived the function

(3) $$G_N(k) = \ln s_k^2 + kN^{-1} \ln N .$$

Hannan and Quinn [5] proposed

(4) $$G_N(k) = \ln s_k^2 + 2kcN^{-1} \ln \ln N ,$$

where $c > 1$ is a constant. It was proved that (2) does not give the consistent estimator of the order of model (1) (see Shibata [10]), while the procedures based on (3) and (4) are consistent.

The aim of our paper is to derive by elementary means a similar method for consistent estimation of the order of a regression model and to present some results from simulated data.

2. PRELIMINARIES

In this section we introduce some general assertions which will be needed in the main part of the paper.

Theorem 1. Let ξ be an n-dimensional random vector with $\mathsf{E}\xi = \mu$, $\operatorname{Var} \xi = \boldsymbol{V}$. Then we have for every $n \times n$ matrix $\boldsymbol{A}$

$$\mathsf{E}\xi'\boldsymbol{A}\xi = \operatorname{Tr} \boldsymbol{AV} + \mu'\boldsymbol{A}\mu .$$

If ξ has a normal distribution, then the formula

$$\operatorname{Var} \xi'\boldsymbol{A}\xi = 2 \operatorname{Tr} (\boldsymbol{AV})^2 + 4\mu'\boldsymbol{AVA}\mu$$

holds.

Proof. See Searle [9], pp. 55–57. □

Theorem 2. Let $x_1, \ldots, x_N$ be a sample from a distribution with finite moments $\mu'_1, \ldots, \mu'_{2h}$. Denote

$$\boldsymbol{X} = \begin{Vmatrix} 1, & x_1, & \ldots, & x_1^h \\ 1, & x_2, & \ldots, & x_2^h \\ \ldots & \ldots & \ldots & \ldots \\ 1, & x_N, & \ldots, & x_N^h \end{Vmatrix}, \qquad \boldsymbol{M} = \begin{Vmatrix} 1, & \mu'_1, & \ldots, & \mu'_h \\ \mu'_1, & \mu'_2, & \ldots, & \mu'_{h+1} \\ \ldots & \ldots & \ldots & \ldots \\ \mu'_h, & \mu'_{h+1}, & \ldots, & \mu'_{2h} \end{Vmatrix}.$$

Then

$$N^{-1}\boldsymbol{X}'\boldsymbol{X} \xrightarrow{P} \boldsymbol{M}$$

as $N \to \infty$.

Proof. The assertion is a consequence of the law of large numbers. □

It happens also very often that $x_1, \ldots, x_N$ are equidistant points from a fixed interval $\langle a, b\rangle$, $-\infty < a < b < \infty$, such that $x_1 = a$, $x_N = b$. If $N \to \infty$, then $N^{-1}\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{M}$ again holds. This time the elements of matrix $\boldsymbol{M}$ are

$$\mu'_j = (b - a)^{-1} \int_a^b x^j \, \mathrm{d}x ,$$

i.e. the moments of the rectangular distribution on $\langle a, b\rangle$.

Theorem 3. Write

$$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2), \qquad \boldsymbol{M} = \begin{Vmatrix} \boldsymbol{M}_{11}, & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21}, & \boldsymbol{M}_{22} \end{Vmatrix},$$

where $\boldsymbol{X}_1$ is a $N \times k$ block and $\boldsymbol{M}_{11}$ is a $k \times k$ block, $k \leqq h$. Let $\boldsymbol{M}$ be regular. Then

(5) $$N^{-1}[\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} \boldsymbol{X}_1'\boldsymbol{X}_2] \xrightarrow{P} \boldsymbol{M}_k ,$$

where $\boldsymbol{M}_k = \boldsymbol{M}_{22} - \boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}$ is a positive definite matrix.

Proof. Let U be a random variable with moments $\mu'_1, \ldots, \mu'_{2h}$. For any vector $\mathbf{c} = (c_0, \ldots, c_h)'$ we have

$$0 \leqq \mathsf{E}\Big(\sum_{j=0}^{h} c_j U^j\Big)^2 = \sum_{j=0}^{h} \sum_{k=0}^{h} c_j c_k \mu'_{j+k} = \mathbf{c}'\mathbf{M}\mathbf{c}\,.$$

Therefore, $\mathbf{M}$ is a positive semidefinite matrix. We assume $\mathbf{M}$ to be regular and so it is positive definite. Then $\mathbf{M}_{11}$ as well as $\mathbf{M}_k$ are also positive definite matrices (see Anděl [3], p. 65 for details). Relation (5) follows from the law of large numbers. □

Let us remark that $\mathbf{M}$ is regular if and only if random variables $1, U, U^2, \ldots, U^h$ are linearly independent a.s.

If $x_1, \ldots, x_N$ are equidistant points from $\langle a, b\rangle$, then an analogous assertion to Theorem 3 holds. The assumption that $\mathbf{M}$ is regular is fulfilled automatically.

3. LINEAR MODEL

Consider a linear model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}\,, \tag{6}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ is a vector of observations, $\mathbf{X}$ is a given $N \times p$ matrix and $\mathbf{e} = (e_1, \ldots, e_N)'$ is a vector of disturbances. Assume that the rank of the matrix $\mathbf{X}$ is $\mathrm{r}(\mathbf{X}) = p$ and that $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Then

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}) \tag{7}$$

and the least squares estimator $\mathbf{b}$ of β is $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}$. It is well known that

$$\mathbf{b} \sim N[\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\,.$$

The unbiased estimator for σ^2 is

$$s_p^2 = (N - p)^{-1}\,(\mathbf{Y} - \mathbf{X}\mathbf{b})'\,(\mathbf{Y} - \mathbf{X}\mathbf{b})$$

and we shall use the fact that

$$(N - p)\, s_p^2/\sigma^2 \sim \chi^2_{N-p}\,. \tag{8}$$

Write $\mathbf{X}$ in the form $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1$ is a $N \times k$ block, $k < p$. Denote $\beta = (\beta^{1\prime}, \beta^{2\prime})'$, where β^1 has k components. If we try to fit to $\mathbf{Y}$ a wrong model

$$\mathbf{Y} = \mathbf{X}_1\beta^1 + \mathbf{e}\,,$$

then our estimator for β^1 is

$$\mathbf{b}^1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\,\mathbf{X}_1'\mathbf{Y}$$

and our estimator for σ^2 takes form

$$s_k^2 = (N - k)^{-1}\,(\mathbf{Y} - \mathbf{X}_1\mathbf{b}^1)'\,(\mathbf{Y} - \mathbf{X}_1\mathbf{b}^1)\,.$$

From here we get $s_k^2 = (N - k)^{-1} \mathbf{Y'AY}$, where $\mathbf{A} = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'$. The matrix $\mathbf{A}$ is symmetric and idempotent. The same is true for $\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'$. From $\mathrm{r}(\mathbf{X}_1) = k$ it follows that $\mathrm{r}[\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'] = k$ and then $\mathrm{Tr}\,\mathbf{A} = \mathrm{r}(\mathbf{A}) = N - k$. Denote $\sigma_k^2 = \mathsf{E}s_k^2$. With respect to (7) we get from Theorem 1 that

$$(9) \qquad \sigma_k^2 = \sigma^2 + (N - k)^{-1} \beta'\mathbf{X}'[\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'] \mathbf{X}\beta =$$
$$= \sigma^2 + (N - k)^{-1} \beta^{2\prime}[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'\mathbf{X}_2] \beta^2 .$$

Analogously,

$$(10) \qquad \mathrm{Var}\, s_k^2 = 2(N - k)^{-1} \sigma^4 +$$
$$+ 4(N - k)^{-2} \sigma^2\beta^{2\prime}[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'\mathbf{X}_2] \beta^2 .$$

Now, consider an overfitted model with $k > p$ parameters

$$\mathbf{Y} = \mathbf{Z}\gamma + \mathbf{e} ,$$

where $\mathbf{Z} = (\mathbf{X}, \mathbf{X}_3)$ and $\gamma = (\beta', \lambda')'$. Obviously, $\mathbf{X}_3$ is a $N \times (k - p)$ matrix and λ has $k - p$ components. Let $\mathrm{r}(\mathbf{Z}) = k$. Then $\mathbf{g} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}$ is an unbiased estimator for γ and if we put

$$s_k^2 = (N - k)^{-1} (\mathbf{Y} - \mathbf{Zg})' (\mathbf{Y} - \mathbf{Zg}) ,$$

then s_k^2 is an unbiased estimator for σ^2. Again, we have

$$(11) \qquad (N - k)\, s_k^2/\sigma^2 \sim \chi^2_{N-k} ,$$

which is quite analogous to (8).

Theorem 4. Assume that there exist such positive definite matrices $\mathbf{M}_0, \mathbf{M}_1, \ldots$ $\ldots, \mathbf{M}_{p-1}$ that

$$N^{-1}[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'\mathbf{X}_2] \to \mathbf{M}_k$$

for $k = 0, 1, \ldots, p - 1$ as $N \to \infty$. Define a function

$$A_k = s_k^2(1 + q_{k,N}) , \quad k = 0, 1, \ldots, K ,$$

where $q_{k,N} = kw_N$ and $w_N \to 0$, $N^{1/2}w_N \to \infty$ for $N \to \infty$. Then

$$\mathsf{P}(A_k > A_p \text{ for } k = 0, 1, \ldots, p - 1, p + 1, \ldots, K) \to 1$$

if $N \to \infty$.

Proof. Denote

$$(12) \qquad \delta_k = \lim_{N \to \infty} (N - k)^{-1} \beta^{2\prime}[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'\mathbf{X}_2] \beta^2 ,$$

$k = 0, 1, \ldots, p - 1$. Because we assume that the order of our model is exactly p, we have $\beta^2 \neq 0$ and thus $\delta_0, \ldots, \delta_{p-1}$ exist and are positive. Formulas (9) and (10) imply

$$(13) \qquad \sigma_k^2 \to \sigma^2 + \delta_k , \quad \mathrm{Var}\, s_k^2 = O(N^{-1}) , \quad k = 0, 1, \ldots, p - 1 .$$

Formulas (8) and (11) give

(14) $\sigma_k^2 = \sigma^2 , \quad \operatorname{Var} s_k^2 = 2(N - k)^{-1} \sigma^4 , \quad k = p, p + 1, \ldots, K .$

Denote

$$\eta_k = (s_p^2 - \sigma^2)(1 + q_{p,N}) - (s_k^2 - \sigma_k^2)(1 + q_{k,N}) ,$$
$$\varepsilon_k = \sigma_k^2(1 + q_{k,N}) - \sigma^2(1 + q_{p,N}) .$$

Let $k \neq p$. Then

$$\mathsf{P}(A_k > A_p) = \mathsf{P}(\eta_k < \varepsilon_k) .$$

Obviously, $\mathsf{E}\eta_k = 0$. For $k < p$ we have $\varepsilon_k \to \delta_k$ for $N \to \infty$. Denote $\delta^* = \min(\delta_0, \ldots, \delta_{p-1})$. There exists such N_k that for $N \geqq N_k$ the inequality $\varepsilon_k > \delta^*/2 > 0$ holds. For $k > p$ we see that $\varepsilon_k = \sigma^2(q_{k,N} - q_{p,N}) > 0$. Consider $N \geqq N^* = \max(N_0, \ldots, N_{p-1})$. Using Tchebyschev inequality we obtain

$$\mathsf{P}(A_k > A_p) \geqq \mathsf{P}(|\eta_k| < \varepsilon_k) \geqq 1 - \varepsilon_k^{-2} \operatorname{Var} \eta_k .$$

Since for any two random variables ξ_1, ξ_2 with finite second moments we have

$$\operatorname{Var}(\xi_1 \pm \xi_2) \leqq 2 \operatorname{Var} \xi_1 + 2 \operatorname{Var} \xi_2 ,$$

we can write

(15) $\mathsf{P}(A_k > A_p) \geqq 1 - 2\varepsilon_k^{-2}[(1 + q_{p,N})^2 \operatorname{Var} s_p^2 + (1 + q_{k,N})^2 \operatorname{Var} s_k^2] .$

If $k < p$, then $\varepsilon_k > \delta^*/2$. From (13) and (14) we have $\operatorname{Var} s_k^2 = O(N^{-1})$, $\operatorname{Var} s_p^2 = O(N^{-1})$, and thus formula (15) implies $\mathsf{P}(A_k > A_p) \to 1$.

If $k > p$, then using (14) we get from (15)

$$\mathsf{P}(A_k > A_p) \geqq$$
$$\geqq 1 - 4(q_{k,N} - q_{p,N})^{-2} [(N - p)^{-1} (1 + q_{p,N})^2 + (N - k)^{-1} (1 + q_{k,N})^2] .$$

Inserting $q_{k,N} = k w_N$ we obtain $\mathsf{P}(A_k > A_p) \to 1$.

The assertion that $\mathsf{P}(A_k > A_p$ simultaneously for all $k \neq p) \to 1$ follows from Bonferroni inequality. □

Theorem 3 shows that the condition $\delta_k > 0$ is fulfilled under quite general assumptions. The existence of positive limits of (12) for $k = 0, 1, \ldots, K$ can be proved also for other situations when x_i are chosen in a systematic way. It can be seen from the proof of Theorem 4, that the assertion remains true even under weaker conditions, namely if $\operatorname{Var} s_k^2 = O(N^{-1})$ and if for the smallest eigenvalue λ_N of the matrix

$$N^{-1}[\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} \mathbf{X}_1'\mathbf{X}_2]$$

the relation

$$\liminf_{N \to \infty} \lambda_N > 0$$

holds for all $k = 0, 1, \ldots, p - 1$.

Theorem 4 shows that the variable $k = \hat{p}$ which minimizes the function A_k is a consistent estimator of the order of the given linear model. It remains to choose the function $q_{k,N}$. We can take

$$q_{k,N} = ck/N^{\alpha},$$

where $c > 0$ and $\alpha \in (0, \frac{1}{2})$ are constants. In a simulation study (see Section 4) quite satisfactory results were obtained for $c = 1$, $\alpha = 0{\cdot}25$.

Let us consider in detail a special case of model (6), the classical linear regression

$$Y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \ldots, N.$$

Denote

$$\bar{x} = N^{-1} \sum x_i, \quad s_x^2 = (N - 1)^{-1} \sum (x_i - \bar{x})^2.$$

Inserting into above formulas we obtain

$$\sigma_0^2 = \sigma^2 + (\beta_0 + \beta_1 \bar{x})^2 + (N - 1) N^{-1} \beta_1^2 s_x^2,$$

$$\sigma_1^2 = \sigma^2 + \beta_1^2 s_x^2,$$

$$\operatorname{Var} s_0^2 = 2N^{-1}\sigma^4 + 4N^{-1}\sigma^2[(\beta_0 + \beta_1\bar{x})^2 + N^{-1}(N - 1)\beta_1^2 s_x^2],$$

$$\operatorname{Var} s_1^2 = 4(N - 1)^{-1} \beta_1^2 \sigma^2 s_x^2 + 2(N - 1)^{-1} \sigma^4.$$

We have

$$\sigma_0^2 - \sigma_1^2 = (\beta_0 + \beta_1 \bar{x})^2 - N^{-1} \beta_1^2 s_x^2.$$

If $\beta_0 + \beta_1 \bar{x} = 0$, then $\sigma_0^2 < \sigma_1^2$. It demonstrates a little surprising fact that σ_k^2 may not be decreasing for $k = 0, 1, \ldots, p - 1$.

4. A SIMULATION STUDY

A realization of the model

(16) $$Y_i = 2 - x_i - 2x_i^2 + x_i^3 + e_i$$

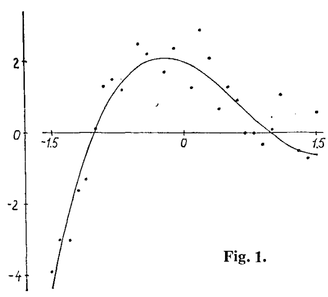

Fig. 1.

for $x_i = -1{\cdot}5\,(0{\cdot}1)\,1{\cdot}5$ with the corresponding theoretical regression function is shown in Fig. 1. The variables e_i are pseudorandom normal numbers with sample mean 0·032 and sample variance 0·408. The values of s_k^2 and $A_k = s_k^2(1 + kN^{-0.25})$ are given in Table 1 and in Figures 2 and 3. Fig. 3 clearly shows a minimum for $k = 4$ parameters. The estimated function is

$$y = 1{\cdot}915 - 1{\cdot}302x - 1{\cdot}866x^2 + 1{\cdot}224x^3$$

and the corresponding unbiased estimate for σ^2 is $s_4^2 = 0{\cdot}444$.

Table 1.

k	0	1	2	3	4	5	6	7	8
s_k^2	2·997	2·892	2·813	0·946	0·444	0·451	0·469	0·436	0·443
A_k	2·997	4·118	5·197	2·149	1·197	1·407	1·662	1·729	1·945

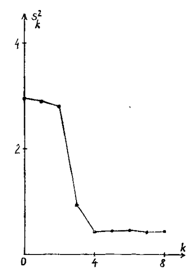

Fig. 2.

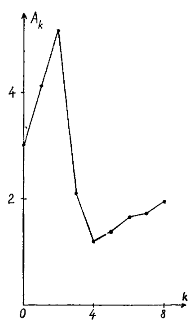

Fig. 3.

Other results concerning model (16) are collected in Tables 2–5. Each row corresponds to 100 simulations. N is the number of equidistant points from $\langle -1{\cdot}5, 1{\cdot}5 \rangle$. The first point is $-1{\cdot}5$, the last one is 1·5. We used functions $A_k = s_k^2(1 + ckN^{-\alpha})$ and $A_k = s_k^2(1 + ckN^{-\alpha} \ln N)$ with $c > 0$, $\alpha \in (0, 0{\cdot}5)$.

Tables 2a and 2b show that in the case of model (16) for $N = 31$ and $\sigma = 0{\cdot}65$ the results do not depend too much on α.

If we take $c = 1$, $\alpha = 0{\cdot}25$ and $N = 31$ or $N = 61$, then we can see from Tables 3a, 3b, 4a and 4b that for small σ both functions $A_k = s_k^2(1 + kN^{-0.25})$ and $A_k = s_k^2(1 + kN^{-0.25} \ln N)$ give similar results, whereas for large σ the former is substantially better than the latter.

Table 2a. $N = 31$, $\sigma = 0{\cdot}65$, $A_k = s_k^2(1 + kN^{-\alpha})$

α \ $\hat{p}$	1	2	3	4	5	6	7+
0·05	0	0	2	94	3	1	0
0·15	0	0	0	95	3	2	0
0·25	0	0	0	95	3	2	0
0·35	0	0	0	94	3	3	0
0·45	0	0	0	91	5	3	1

Table 2b. $N = 31$, $\sigma = 0{\cdot}65$, $A_k = s_k^2(1 + kN^{-\alpha} \ln N)$

α \ $\hat{p}$	1	2	3	4	5	6	7+
0·05	0	0	4	95	1	0	0
0·15	0	0	4	95	1	0	0
0·25	0	0	3	95	2	0	0
0·35	0	0	3	93	3	1	0
0·45	0	0	2	94	3	1	0

Table 3a. $N = 31$, $A_k = s_k^2(1 + kN^{-0.25})$

σ \ $\hat{p}$	1	2	3	4	5	6	7+
0·25	0	0	0	95	3	2	0
0·50	0	0	0	95	3	2	0
0·75	0	0	11	87	1	1	0
1·00	0	0	28	68	3	1	0
1·25	6	0	50	43	0	1	0
1·50	21	0	47	29	2	1	0

Table 3b. $N = 31$, $A_k = s_k^2(1 + kN^{-0.25} \ln N)$

σ \ $\hat{p}$	1	2	3	4	5	6	7+
0·25	0	0	0	98	2	0	0
0·50	0	0	0	98	2	0	0
0·75	0	0	17	82	0	1	0
1·00	8	0	40	50	2	0	0
1·25	41	0	36	23	0	0	0
1·50	72	0	19	9	0	0	0

Table 4a. $N = 61,\ A_k = s_k^2(1 + kN^{-20.5})$

σ \ $\hat{p}$	1	2	3	4	5	6	7+
0·25	0	0	0	100	0	0	0
0·50	0	0	0	100	0	0	0
0·75	0	0	1	99	0	0	0
1·00	0	0	18	82	0	0	0
1·25	0	0	52	48	0	0	0
1·50	9	0	65	26	0	0	0

Table 4b. $N = 61,\ A_k = s_k^2(1 + kN^{-0.25} \ln N)$

σ \ $\hat{p}$	1	2	3	4	5	6	7+
0·25	0	0	0	100	0	0	0
0·50	0	0	0	100	0	0	0
0·75	0	0	6	94	0	0	0
1·00	5	0	42	53	0	0	0
1·25	60	0	28	12	0	0	0
1·50	89	0	8	3	0	0	0

Table 5 summarizes some results with varying c. Again, the dependence on c does not seem to be very strong – only the value $c = 0{\cdot}5$ leads to a larger number of overfitted models.

Table 5. $N = 31,\ \sigma = 0{\cdot}65,\ A_k = s_k^2(1 + kcN^{-0.25})$

c \ $\hat{p}$	1	2	3	4	5	6	7+
0·5	0	0	0	91	5	3	1
1·0	0	0	0	95	3	2	0
1·5	0	0	1	95	3	1	0
2·0	0	0	2	94	3	1	0

The dependence of the estimates on the choice of a model was investigated for the following models:

I. $Y_i = 0{\cdot}2 + 0{\cdot}5x_i + 0{\cdot}2x_i^2 + e_i$;

II. $Y_i = 0{\cdot}2 + 0{\cdot}5x_i + 0{\cdot}5x_i^2 + e_i$;

III. $Y_i = 0{\cdot}2 + 0{\cdot}5x_i + 1{\cdot}0x_i^2 + e_i$;

IV. $Y_i = 0{\cdot}2 + 1{\cdot}0x_i + 0{\cdot}2x_i^2 + e_i$.

Here $e_i \sim N(0, \sigma^2)$. The corresponding regression functions for $1 \leqq x \leqq 6$ are given in Fig. 4. The points $x_i = 1{\cdot}0\ (0{\cdot}1)\ 5{\cdot}9$ were taken (i.e., $N = 50$). The results of

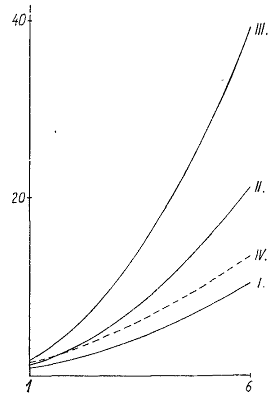

Fig. 4.

simulation are for $A_k = s_k^2(1 + kN^{-0.25} \ln N)$ in Tables 6a–6c. Each row corresponds to 30 simulations. Models I. and IV. led to the same table for $\sigma \leqq 2{\cdot}0$.

Table 6a.

Models I. and IV.

σ \ $\hat{p}$	1	2	3	4+
0·1	0	0	30	0
0·2	0	0	30	0
0·4	0	5	25	0
0·5	0	16	14	0
0·6	0	20	10	0
0·7	0	24	6	0
0·8	0	27	3	0
0·9	0	29	1	0
1·0	0	30	0	0
2·0	0	30	0	0

Table 6b.

Model II.

σ \ $\hat{p}$	1	2	3	4+
0·4	0	0	30	0
0·6	0	0	30	0
0·8	0	0	30	0
1·0	0	5	25	0
1·2	0	12	18	0
1·5	0	20	10	0
1·7	0	23	7	0
2·0	0	27	3	0
3·0	0	30	0	0
4·0	0	30	0	0

Table 6c.

Model III.

σ \ $\hat{p}$	1	2	3	4+
1·0	0	0	30	0
2·0	0	5	25	0
3·0	0	20	10	0
4·0	0	27	3	0
5·0	0	30	0	0

The table confirms our expectation, namely, that the order of the regression function can be better estimated even for greater σ, if the coefficient by x^2 is larger.

(Received May 3, 1981.)

REFERENCES

[1] H. Akaike: Fitting autoregressive models for prediction. Ann. Inst. Statist. Math. *21* (1969), 243–247.

[2] H. Akaike: Information theory and an extension of the maximum likelihood principle. Proc. 2nd Int. Symp. Information Theory, Supp. to Problems of Control and Information Theory (1972), 267–281.

[3] J. Anděl: Matematická statistika. SNTL, Praha 1978.

[4] T. W. Anderson: The Statistical Analysis of Time Series. Wiley, New York 1971.

[5] E. J. Hannan, B. G. Quinn: The determination of the order of an autoregression. J. Roy. Statist. Soc. *B 41* (1979), 190–195.

[6] C. L. Mallows: Some comments on C_p. Technometrics *15* (1973), 661–675.

[7] J. Rissanen: Modeling by shortest data description. Automatica *14* (1978), 465–471.

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

[9] S. R. Searle: Linear Models. Wiley, New York 1971.

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

Doc. RNDr. Jiří Anděl, CSc., Universita Karlova, fakulta matematicko-fyzikální (Charles University — Faculty of Mathematics and Physics), Sokolovská 83, 186 00 Praha 8. Czechoslovakia.

Professor Manuel Garrido Perez, Professor Antonio Insua Negrao, Polytechnical University, Faculty of Informatics, Carretera Valencia km. 7, Madrid 31. Spain.