Bridging the gap between B-spline and polynomial regression model

Kenichi Satoh , Hirokazu Yanagihara and **Megu Ohtaki**
Hiroshima University, JAPAN

SUMMARY. In this paper, we propose a method to select the better of two types of models: a polynomial with low degree and a B-spline model, using the common information criterion. The methodology can be directly applied to semi-parametric multiple regression analysis.

Keywords: B-spline, information criterion, nonparametric regression, semi-parametric regression, simple regression,

†*Address for correspondence*: Department of Environmetrics and Biometrics, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan.
E-mail: ksatoh@hiroshima-u.ac.jp

1. Introduction

B-splines are often used in nonparametric regression as a flexible smoother (see de Boor (1978), Eilers and Marx (1996)). They can provide various styles of curves, from approximate linear curves to complicated nonlinear curves, but it is impossible for B-spline to express a genuine linear curve or a constant line, which plays a role as the null trend curve in nonparametric regression models. These curves have the advantage of easy handling: estimation, confidence bound, prediction, *etc.*, and putting interpretations on actual situations. Therefore we sometimes face the problem of choosing between a simple regression model and a B-spline model. In this paper, we propose a method to select the best model among those models. Models and information criterion are discussed in Section 2. Some illustrative examples shown in Section 3 demonstrate the good performance of the procedure. The methodology can be directly applied to semi-parametric multiple regression analysis, for example, through the ACE-algorithm of Breiman and Friedman (1985).

2. Models and Criterion

The B-spline model is essentially a multiple linear regression model with a special design matrix. For given n pairs of observation points: $(x_1, y_1), \cdots (x_n, y_n)$, let $\mathbf{y} = (y_1, \cdots, y_n)'$ be a $n \times 1$ response vector and $\mathbf{x} = (x_1, \cdots, x_n)'$ be an $n \times 1$ explanatory vector. The design matrix consists of m B-spline basis functions expressed as $X = \{\mathbf{b}_1(\mathbf{x}) \cdots \mathbf{b}_m(\mathbf{x})\}$ with $\mathbf{b}_j(\mathbf{x}) = \{B_j(x_1) \cdots B_j(x_n)\}'$. Then, the model is written as

$$\mathbf{y} = X\mathbf{a} + \mathbf{e}, \qquad \mathbf{e} \sim N_n(0, \sigma^2 I_n),$$

where $\mathbf{a}$ is an unknown coefficient vector, $\mathbf{e}$ is an error vector, and each element of $\mathbf{e}$ is independently distributed as normal distribution with mean zero and unknown variance σ^2. B_j is the jth basis function, which is defined on an equidistant mesh as

$$B_j(x) = B_0\{h, a + h(j-2), x\}, \qquad j = 1, \cdots, m,$$

where $h = (b-a)/(m-3)$ for $m \geq 4$ with $a = \min_i(x_i)$ and $b = \max_i(x_i)$. B_0 is a symmetric function on x_0 defined by

$$B_0(h, x_0, x) = \begin{cases} \dfrac{1}{6h}\left\{\left(2 - \dfrac{|x-x_0|}{h}\right)^3 - 4\left(1 - \dfrac{|x-x_0|}{h}\right)^3\right\} & (|x-x_0| \leq h) \\ \dfrac{1}{6h}\left(2 - \dfrac{|x-x_0|}{h}\right)^3 & (h < |x-x_0| \leq 2h) \\ 0 & (otherwise) \end{cases}.$$

Then, the predictor of the response with a penalty term is given by

$$\hat{\mathbf{y}}_{(\lambda)} = S_{(\lambda)}\mathbf{y}, \qquad S_{(\lambda)} = X(X'X + n\hat{\sigma^2}\lambda D_2'D_2)^{-1}X',$$

where $\hat{\sigma^2} = ||\mathbf{y} - \hat{\mathbf{y}}_{(\lambda)}||/n$, λ is a smoothing parameter and D_2 is a $(m-2) \times m$ second difference matrix expressed as

$$D_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

Note that the when $\lambda = 0$ the fitted predictor is the same as that of maximum likelihood estimator.

Since the minimum number of B-spline components is commonly four and the B-spline can't itself express a linear curve, we suggest using the *extended B-spline model* where the B-spline design is replaced by that of a polynomial regression model when the number of components is less than four, as follows way:

$$X = (\mathbf{x}^0 \cdots \mathbf{x}^{m-1}), \qquad m = 1, 2 \text{ or } 3,$$

where $\mathbf{x}^j = (x_1^j, \cdots, x_n^j)'$. Thus, the quadratic model ($m = 3$) is a bridge between the simple regression models ($m = 1, 2$) and the B-spline models ($m \geq 4$). Note that the predictor of the response under the quadratic model has the penalty term as given in Table 1, because the curve is twice continuously differentiable.

- Insert Table 1 Here -

In order to select the best model among the extended B-spline models, we suggest using an improved Akaike information criterion (AIC, Akaike, 1973) $CAIC$, due to Sugiura (1978). The properties of $CAIC$ ware investigated by Bedrick and Tsai (1994) and modified by Fujikoshi and Satoh (1997). Hurvich, Simonoff and Tsai (1998) proposed using it in nonparametric regression models and it was compared to other criteria by Imoto and Konishi (1999). The criterion is written in the following simple form:

$$CAIC = n \log \hat{\sigma}^2 + \frac{n(\mathrm{tr}S + n)}{n - \mathrm{tr}S - 2}.$$

Note that $\mathrm{tr}S$ is m, the number of basis functions, and it is the same criterion as that of linear regression models when the smoothing parameter is unused or estimated to be zero. The trace term might be considered as the dimension of the projection space spaned by $S_{(\lambda)}$. The optimization of the smoothing parameter is obtained by minimizing the criterion for each model. Thus, we can choose the model that minimizes the criterion among the extended B-spline models: low order polynomials and B-spline models.

3. Examples

In this section we use Monte Carlo simulation to investigate the behavior of the criterion. We simulations examine the performance of the criterion as it relates to the true regression function and confirm the role of the quadratic model as a bridge.

In our simulation study, sample sizes $n = 30$ and 300 were examined. Note that consideration of the larger sample size is equivalent to that of a smaller error variance. First, x_i, $(i = 1, \cdots, n)$ were generated from uniform distribution on (-10,10) and the following five types of the true regression models were considered,

1) constant line: $y_i = 0 + e_i$,

2) linear curve: $y_i = x_i + 4e_i$,

3) quadratic curve: $y_i = x_i - x_i^2 + 40e_i$,

4) cubic curve: $y_i = x_i - x_i^2 + x_i^3 + 400e_i$

5) sine curve: $y_i = x_i \sin(x_i) + 4e_i$,

where e_i is independently distributed as $N(0, 1)$ and independent of x_i. We obtained n pairs of observation points from each true regression model and Sselected the best model by $CAIC$ among the extended B-spline models with $m = 1, \cdots, 15$. The typical shape of the criterion values and the fitted curve under the best model for $n = 30$ are shown in Figures 1-5.

- Insert Fig. 1. Here -
- Insert Fig. 2. Here -
- Insert Fig. 3. Here -
- Insert Fig. 4. Here -
- Insert Fig. 5. Here -

The optimazation of λ using $CAIC$ was performed on $\lambda = 0, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}$ and 10^{2}. The best models selected by $CAIC$ seem to give suitable fits to the observation points. The value of $CAIC$ for the quadratic model might take on a reasonable value so that there is a continuum in our extended B-spline models because of the smoothing term. Tables 2 and 3 show the frequencies of the best model selected by $CAIC$ among the extended B-spline models. We examined 1000 repetitions for each of the above five true regression models.

- Insert Table 2. Here -
- Insert Table 3. Here -

In the small sample case, the criterion tended to choose simpler models even if the true model was more complicated, especially the cubic and sine curves. The results depended on the fact that $CAIC$ or AIC was based on a predictive density function. In the large sample case, we can easily see that the criterion did not select the models that are simpler than the true model. The frequencies for a true constant line is remarkable, since the total frequencies for $m = 1$ and 2 attain nearly 80% for the case of small sample and 70% even for the case of large sample. Therefore it is possible for the criterion to choose a linear curve from the extended B-spline models even in situations where we unknowingly do not have to use B-spline models, i.e., nonparametric regression. As a result, we can take advantage of these simple nonparametric curves to analyze real data.

References

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.Csáki), 267-281, Akademia Kiado, Budapest. (Reproduced in *Breakthroughs in Statistics*, Vol. 1 (eds. S. Kotz and N. L. Johnson), Springer New York, (1992).)

Bedrick, E. J. and Tsai, C. L. (1994) Model selection for multivariate regression in small samples, *Biometrics*, **50**, 226-231.

Breiman, L. and Friedman, J. H. (1985) Estimating optimal transformations for multiple regression and correlation (with comment), *J. Am. Statist. Assoc*, **80**, 580-599.

de Boor, C. (1978) A Practical Guide to Splines, Springer, Berlin.

Eilers, P. and Marx, B. (1996) Flexible smoothing with B-splines and penalties (with discussion), *Statistical Science*, **11**, 89-121.

Fujikoshi, Y. and Satoh, K. (1997) Modified AIC and C_p criterion in multivariate linear regression, *Biometrika*, **84**, 707-716.

Imoto S. and Konishi, S. (1999) Nonlinear regression models using B-spline and information criteria, *Proceedings of the Institute of Statistical Mathematics*, **47**, 359-373.

Hurvich, C. M., Simonoff, J. S. and Tsai, C. L. (1998) Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion, *J. R. Statist. Soc.* B, **60** , 271-293.

Sugiura, N. (1978) Further analysis of the data by Akaike's information criterion and the finite corrections, *Commun. Statist. -Theory Meth.*, **7**, 13-26.

Table 1. The role of the quadratic model as a bridge between simple regression models and B-spline models.

Model	#Basis functions	Smoothing parameter
Constant, Linear	m=1,2	$\lambda = 0$
Quadratic	m=3	$\lambda \geq 0$
B-spline	$m \geq 4$	$\lambda \geq 0$

Table 2. The frequencies of various models selected by $CAIC$ for each true regression model when $n = 30$. The best model is expressed by its number of basis functions (m).

Best model: m	1	2	3	4	5	6-10	11-15
Constant	693	91	115	18	17	55	11
Linear	0	647	13	119	75	99	47
Quadratic	14	5	763	51	48	93	26
Cubic	6	342	9	265	94	237	47
Sine	164	28	94	13	18	518	165

Table 3. The frequencies of various models selected by $CAIC$ for each true regression model when $n = 300$. The best model is expressed by its number of basis functions.

Best model: m	1	2	3	4	5	6-10	11-15
Constant	619	98	134	23	21	69	36
Linear	0	533	12	147	83	134	91
Quadratic	0	0	687	51	88	126	48
Cubic	0	0	0	571	108	263	58
Sine	0	0	0	0	0	2	998

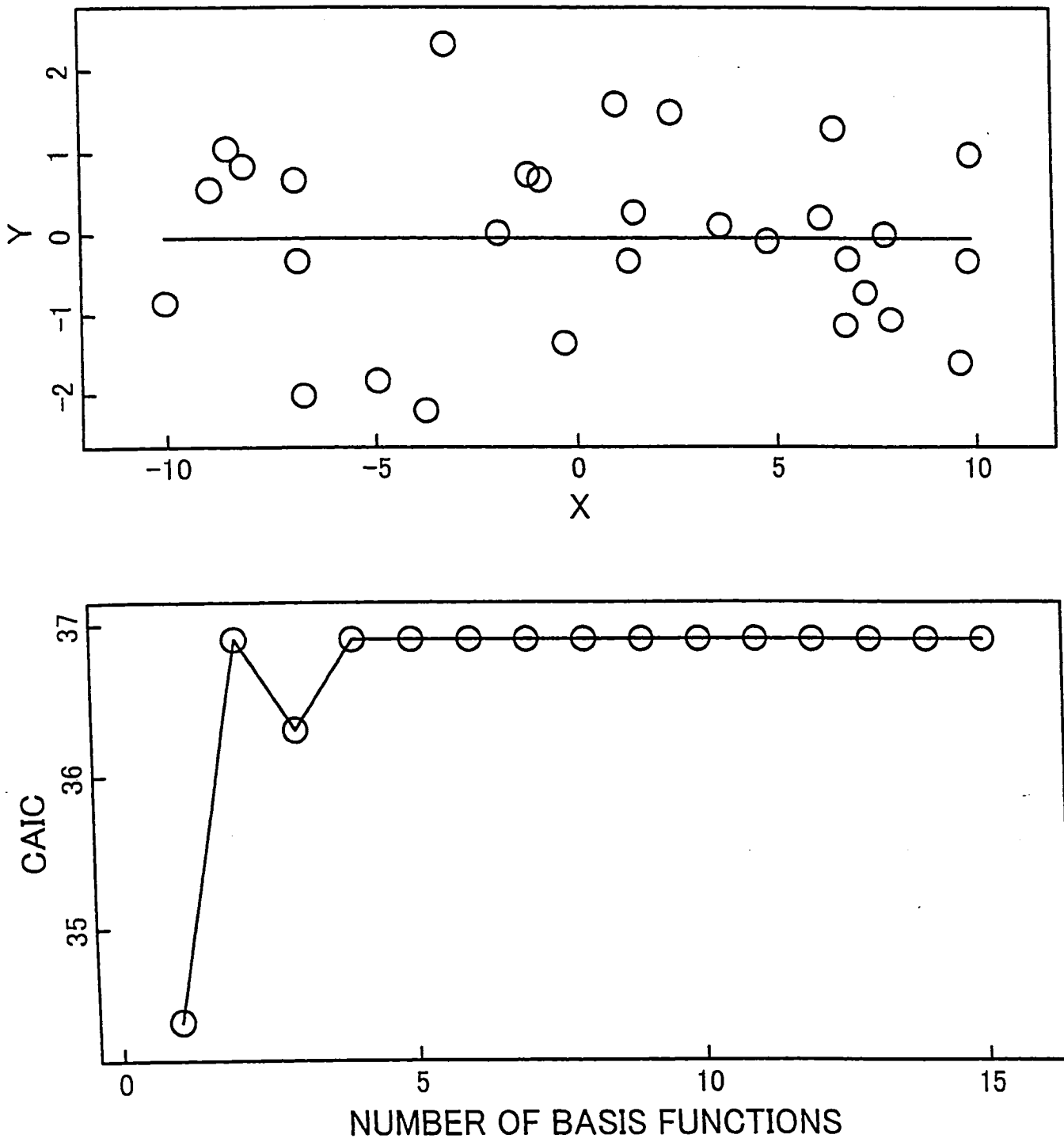

Fig. 1. The fitted spline under the best model and the profile of $CAIC$ for extended models; the true regression model: constant line, the best model: $m = 1$, the smoothing parameter: λ is unused;

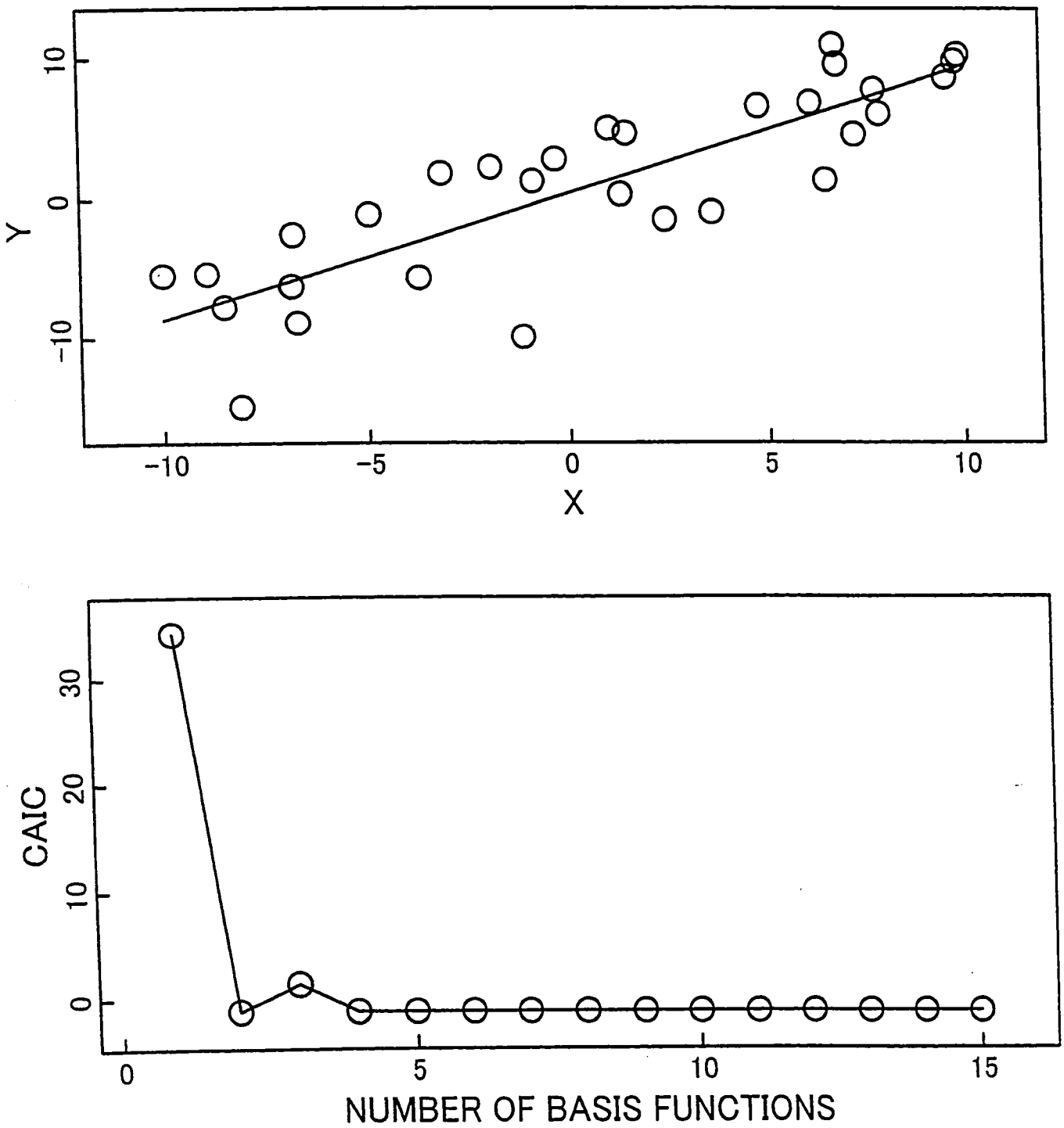

Fig. 2. The fitted spline under the best model and the profile of $CAIC$ for extended models; the true regression model: linear curve, the best model: $m = 2$, the smoothing parameter: λ is unused;

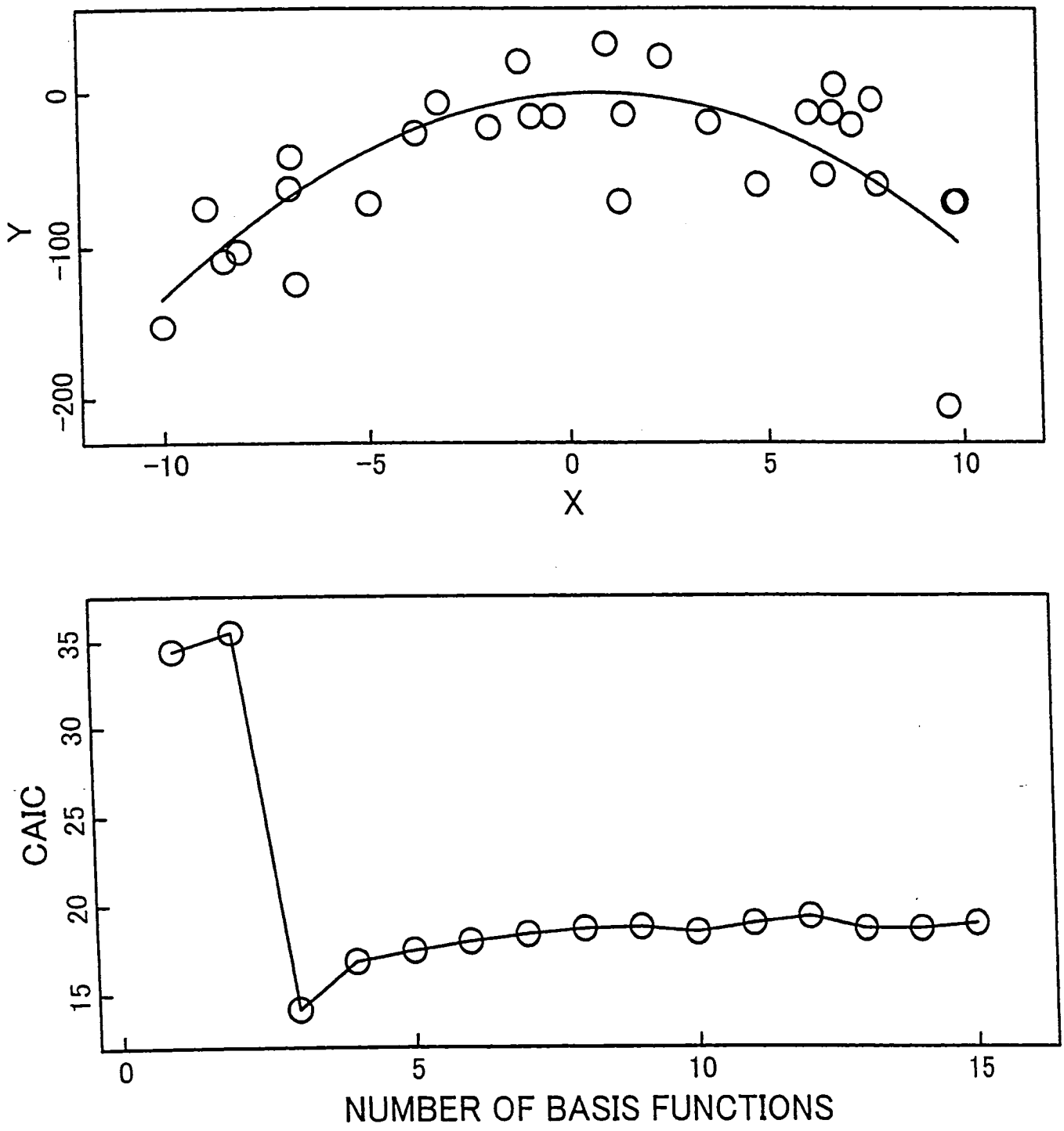

Fig. 3. The fitted spline under the best model and the profile of $CAIC$ for extended models; the true regression model: quadratic curve, the best model: $m = 3$, the optimal smoothing parameter: $\lambda = 100$;

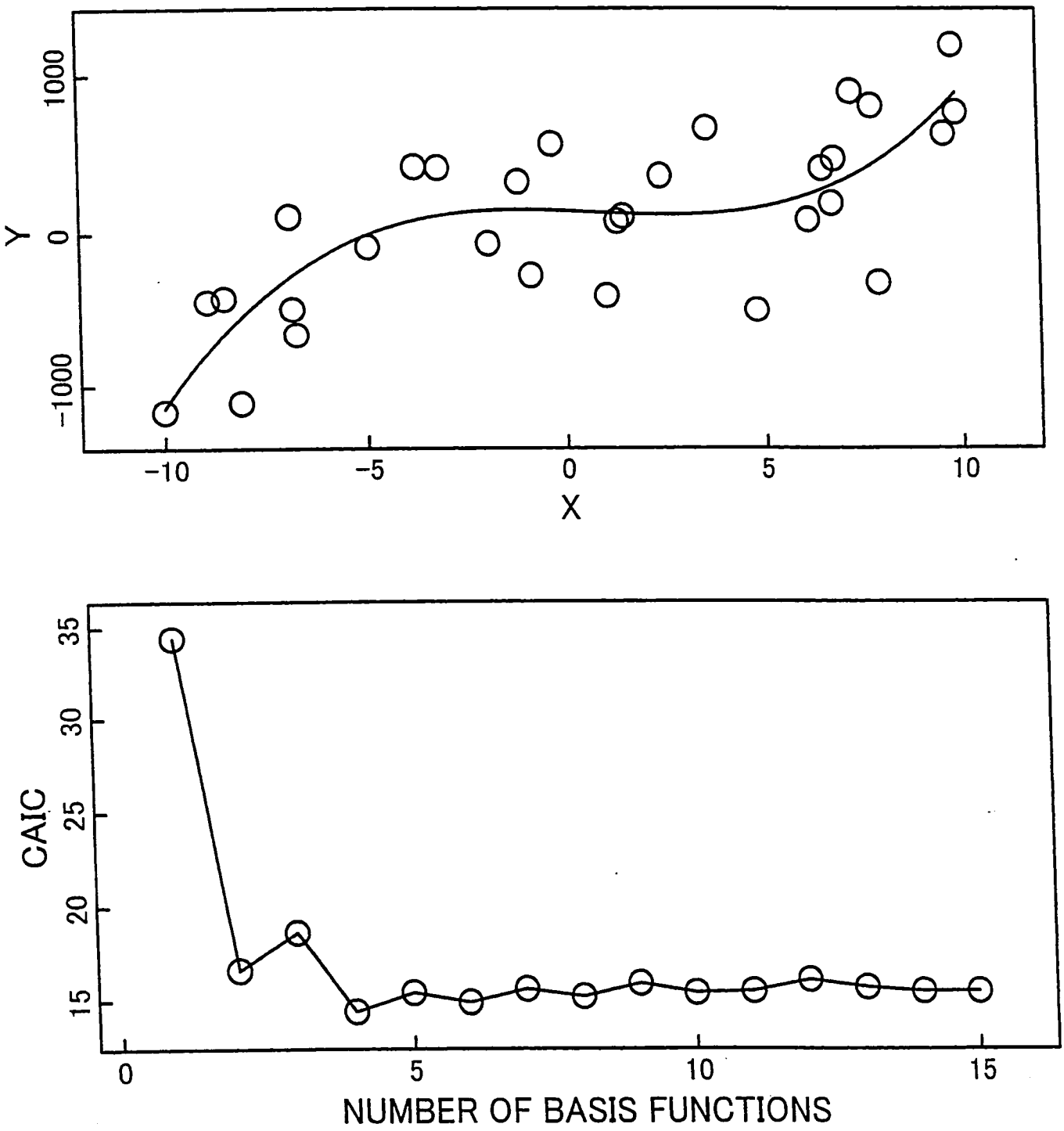

Fig. 4. The fitted spline under the best model and the profile of $CAIC$ for extended models; the true regression model: cubic curve, the best model: $m = 4$, the optimal smoothing parameter: $\lambda = 0$;

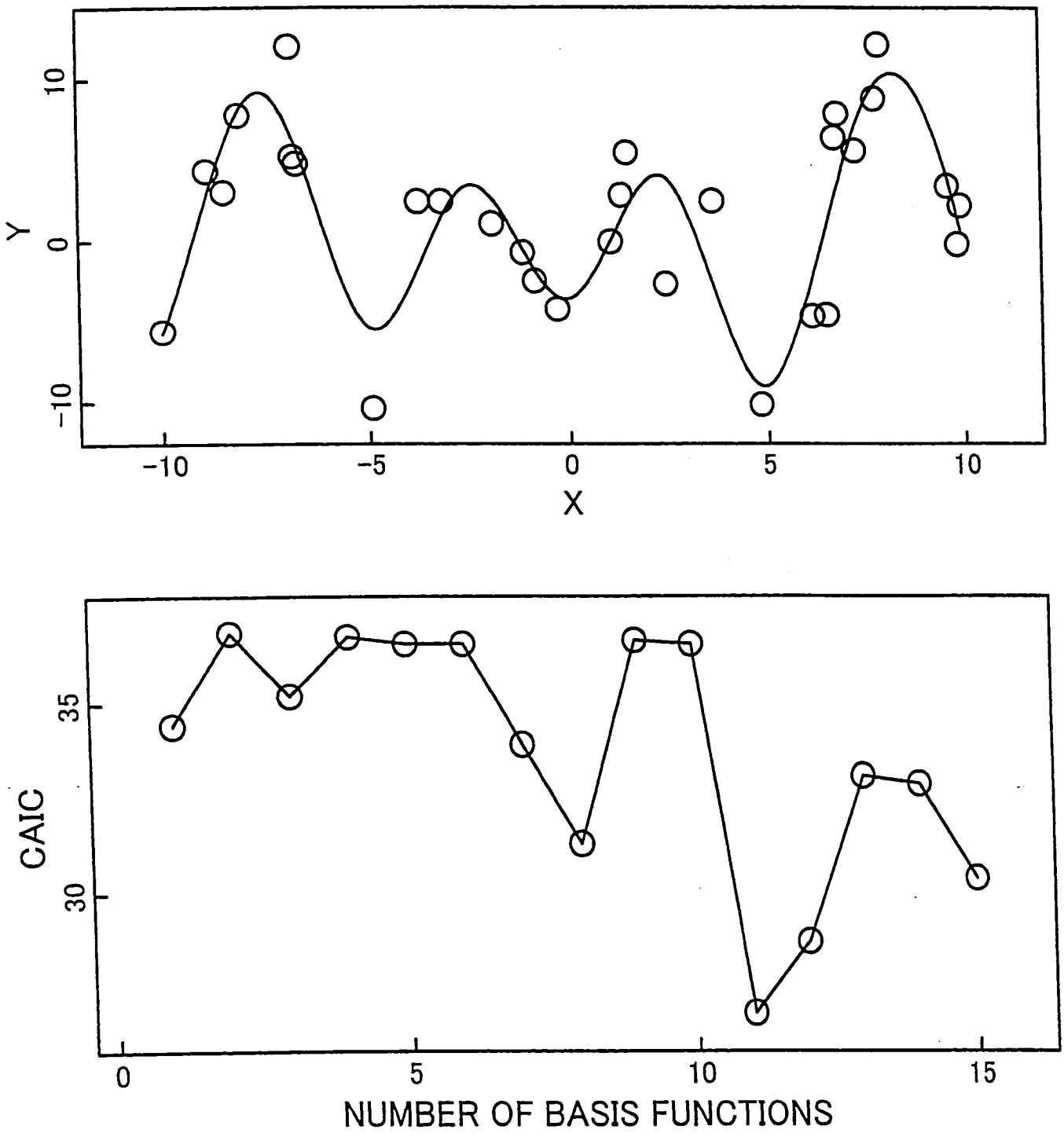

Fig. 5. The fitted spline under the best model and the profile of $CAIC$ for extended models; the true regression model: complicated curve, the best model: $m = 11$, the optimal smoothing parameter: $\lambda = 100$;