TR-No. 14-07, Hiroshima Statistical Research Group, 1–12

Comparison with RSS-Based Model Selection Criteria for Selecting Growth Functions

Keisuke Fukui[2*], Mariko Yamamura[1] and Hirokazu Yanagihara[2]

[1]Department of Mathematics Education, Graduate School of Education, Hiroshima University
1-1-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8524, Japan
[2]Department of Mathematics, Graduate School of Science, Hiroshima University
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8626, Japan

(Last Modified: August 15, 2014)

Abstract

A growth curve model used for analyzing quantity of growth is characterized by a mathematical function with respect to a time, which is called a growth function. Since results of analysis from a growth curve model strongly depend on which growth functions are used for the analysis, a selection of growth functions is important. A choice of growth function based on the minimization of a model selection criterion is one of the major selection methods. In this paper, we compare the performances of growth-function-selection methods using these criteria (e.g., Mallows' C_p criterion) through Monte Carlo simulations. As a result, we recommend the use of the selection method using the BIC-type model selection criterion for a selection of growth functions.

Key words: Growth-function-selection, growth curve model, model selection criterion, residual sum of squares

*Corresponding author
E-mail address: d126313@hiroshima-u.ac.jp (Keisuke Fukui)

1. Introduction

A growth curve model used for analyzing quantity of growth is specified by a mathematical function, which is called the growth function. Since there are many growth functions used for analysis, a growth-function-selection (GF-selection) is important because results of analysis from a growth curve model are different if used growth functions are different. A growth function with high prediction performance is regarded as a better growth function. Hence, in the GF-selection, the best model should be chosen to improve the accuracy of a prediction.

A choice of growth function based on the minimization of a model selection criterion (MSC) is one of the major selection methods. A MSC consists of two terms; one is the goodness-of-fit term, the other is the penalty term imposing the complexity of a model. In particular MSC whose goodness-of-fit term is a residual sum of squares (RSS) is called a RSS-based MSC in this paper.

A RSS-based MSC is often used for selecting the best model in many fields. Since several RSS-based MSC can be regarded as an estimator of the risk function assessing the standardized mean square error (MSE) of prediction, we can expect that the accuracy of a growth prediction will be improved in the sense of making MSE small by minimizing a RSS-based MSC. However, there are many RSS-based MSCs, e.g., Mallows' C_p criterion (Mallows, 1973). If MSC used for GF-selection is different, a chosen growth function will be changed. Hence, the purpose of this study is to compare the performances of GF-selection methods using RSS-based MSCs through Monte Carlo simulations.

This paper is organized in the following ways: In Section 2, we introduce the growth curve model and used growth functions. In section 3, we describe the RSS-based MSCs for GF-selection. In Section 4, we compare GF-selection methods using the RSS-based MSCs by conducting numerical experiments and give discussions.

2. Growth Curve Model

2.1. True and Candidate Models

Let $y(t_i)$ be a growth amount at the time t_i $(i = 1, \ldots, n)$, where n is the sample size. Suppose that $y(t_i)$ is generated from the following true model:

$$y(t_i) = \mu_*(t_i) + \varepsilon_*(t_i),$$

where $\mu_*(t_i)$ is the true expected value of $y(t_i)$, and $\varepsilon_*(t_1), \ldots, \varepsilon_*(t_n)$ are mutually independent true error variables from the same distribution with the mean 0 and the variance σ_*^2. Since $\mu_*(t)$ expresses the average variation of the true growth amount, $\mu_*(t)$ is denoted by the growth function. However, nobody knows the true model. Hence, the following candidate model is assumed to $y(t_i)$:

$$y(t_i) = \mu(t_i) + \varepsilon(t_i),$$

where $\mu(t_i)$ is the expected value of $y(t_i)$ under the candidate model, and $\varepsilon(t_1), \ldots, \varepsilon(t_n)$ are mutually independent error variables from the same distribution with the mean 0 and the variance σ^2. Here, we call $\mu(t_i)$ the candidate growth function. In practice, we have to prepare a specific function with respect to t, whose shape is determined by unknown parameters, as the candidate growth function.

Let $\mu(t; \boldsymbol{\theta}_\mu)$ denote the candidate growth function, where $\boldsymbol{\theta}_\mu$ is a $q(\mu)$-dimensional vectors. It should be kept in mind that $q(\mu)$ denotes the number of unknown parameters of a candidate growth function μ. In order to use the growth curve model, it is necessary to estimate $\boldsymbol{\theta}_\mu$ from a growth data. In this paper, we estimate $\boldsymbol{\theta}_\mu$ by the least square (LS) estimation. Let the residual sum of squares be denoted by

$$\mathrm{RSS}(\boldsymbol{\theta}_\mu; \mu) = \sum_{i=1}^{n} \left\{ y(t_i) - \mu(t_i; \boldsymbol{\theta}_\mu) \right\}^2 .$$

Then, the LS estimator of $\boldsymbol{\theta}$ is derived by minimizing $\mathrm{RSS}(\boldsymbol{\theta}; \mu)$ as

$$\hat{\boldsymbol{\theta}}_\mu = \arg\min_{\boldsymbol{\theta}_\mu} \mathrm{RSS}(\boldsymbol{\theta}_\mu; \mu).$$

By using $\hat{\boldsymbol{\theta}}_\mu$, a growth curve can be estimated by $\mu(t; \hat{\boldsymbol{\theta}}_\mu)$.

2.2. Selection of Growth Functions

There are many available growth functions proposed in the literatures. In this paper, we consider the following twelve candidate growth functions, which were described in Ziede (1993).

(1) Bertalanffy: $\mu_1(t; \boldsymbol{\theta}) = \alpha(1 - e^{-\beta t})^3$ $(\boldsymbol{\theta} = (\alpha, \beta)')$.

(2) Chapman-Richards: $\mu_2(t; \boldsymbol{\theta}) = \alpha(1 - e^{-\beta t})^\gamma$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(3) Gompertz: $\mu_3(t; \boldsymbol{\theta}) = \alpha \exp(-\beta e^{-\gamma t})$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(4) Hossfeld-4: $\mu_4(t; \boldsymbol{\theta}) = \alpha(1 + \beta t^{-\gamma})^{-1}$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(5) Korf: $\mu_5(t; \boldsymbol{\theta}) = \alpha \exp(-\beta t^{-\gamma})$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(6) Levakovic-3: $\mu_6(t; \boldsymbol{\theta}) = \alpha(1 + \beta t^{-2})^{-\gamma}$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(7) Logistic: $\mu_7(t; \boldsymbol{\theta}) = \alpha(1 + \beta e^{-\gamma t})^{-1}$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(8) Monomolecular: $\mu_8(t; \boldsymbol{\theta}) = \alpha(1 - \beta e^{-\gamma t})$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(9) Weibull: $\mu_9(t; \boldsymbol{\theta}) = \alpha(1 - e^{-\beta t^\gamma})$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma)')$.

(10) Levakovic-1: $\mu_{10}(t; \boldsymbol{\theta}) = \alpha(1 + \beta t^{-\gamma})^{-\delta}$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta)')$.

(11) Sloboda : $\mu_{11}(t; \boldsymbol{\theta}) = \alpha \exp(-\beta e^{-\gamma t^\delta})$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta)')$.

(12) Yoshida-1 : $\mu_{12}(t; \boldsymbol{\theta}) = \alpha(1 + \beta t^{-\gamma})^{-1} + \delta$ $(\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta)')$.

In the above list, t denotes a time and all parameters are restricted to positive values. The candidate growth functions have been listed in order by the number of unknown parameters, i.e., the function μ_1 includes two, the functions μ_2 to μ_9 include three and the functions μ_{10} to μ_{12} include four parameters.

Although an estimate of a growth curve can be obtained by the LS estimation, it is important that which growth function is the most suitable to an obtained growth data. In this paper, we select the best growth function by the RSS-based MSC minimization method. Let $\mathrm{MSC}_{\mathrm{RSS}}(\mu)$ denote a general form of a RSS-based MSC. Then the best growth function is determined as

$$\hat{\mu} = \arg \min_{\mu \in \{\mu_1, \ldots, \mu_{12}\}} \mathrm{MSC}_{\mathrm{RSS}}(\mu).$$

2.3. Underspecified and Overspecified Models

From the equation of each growth function, it is seen that several growth functions have inclusion relations (e.g., Bertalanffy with $\gamma = 3$ corresponds perfectly Chapman-Richards). In model selection, these relationships sometimes play key roles, since several MSCs are derived under the assumption that a candidate model includes the true model. We define following two specific candidate models,

- An overspecified model: a growth function of a candidate model includes that of the true model, i.e., the true growth function can be expressed as the special case of the growth function of the overspecified model. In general, the true model is the overspecified model. However, in this paper, we rule out the true model from the definition of an overspecified model.
- An underspecified model: the model is not the overspecified model and the true model.

In practice, there is no overspecified model in most cases. An overspecified model does not exist except the following four cases:

(i) When the true growth function is Bertalanffy, the candidate model whose growth function is Chapman-Richards is the overspecified model.

(ii) When the true growth function is Gompertz, the candidate model whose growth function is Sloboda is the overspecified model.

(iii) When the true growth function is Hossfeld-4, the candidate model whose growth function is Levakovic-1 is the overspecified model.

(iv) When the true growth function is Levakovic-3, the candidate model whose growth function is Levakovic-1 is the overspecified model.

3. RSS-Based Model Selection Criteria

In this section, we describe an explicit form of used RSS-based MSC for GF-selection.

When the penalty for the complexity of a model is imposed additively, an estimator of σ^2 is required to use a RSS-based MSC. In the general regression model, an estimator of σ^2 in the full model is usually used. However, it is difficult to construct the full model in the growth curve model because there is no candidate model which includes all candidate models. Hence, we use the following estimator of σ^2 derived from a local linear fitting, which was proposed by Gasser, Sroka and Jennen-Steinmetz (1986),

$$\hat{\sigma}_{\mathrm{L}}^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} \frac{(a_i y_{i-1} + b_i y_{i+1} - y_i)^2}{a_i^2 + b_i^2 - 1},$$

where coefficients a_i and b_i are given by

$$a_i = \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}}, \quad b_i = \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}.$$

The $\hat{\sigma}_{\mathrm{L}}^2$ has a desirable property as an estimator of σ^2, e.g., $\hat{\sigma}_{\mathrm{L}}^2$ converges to σ^2 as $n \to \infty$ in probability if $\mu_*(t)$ is twice continuously differentiable, $\limsup_{n\to\infty} \max_{i=2,\ldots,n-1} |t_i - t_{i-1}| < \infty$ and $E[\varepsilon_*(t_i)^4] < \infty$.

3.1. Mallows' C_p Criterion

By using $2q(\mu)$ as the penalty term, Mallows' C_p criterion is defined as

$$C_p(\mu) = \frac{\mathrm{RSS}(\hat{\boldsymbol{\theta}}_\mu; \mu)}{\hat{\sigma}^2_{\mathrm{L}}} + 2q(\mu). \tag{1}$$

The $2q(\mu)$ was derived as the bias of $\mathrm{RSS}(\hat{\boldsymbol{\theta}}_\mu; \mu)/\hat{\sigma}^2_{\mathrm{L}}$ to the risk function assessing the standardized MSE of prediction under the assumption that the candidate model considered is not an underspecified model. Hence, there is a possibility that the C_p may not evaluate correctly the complexity of an underspecified model.

3.2. Modified C_p Criterion

The weakness of the C_p may be overcome by using the generalized degree of freedom (GDF) proposed by Ye (1998) instead of $q(\mu)$. The GDF of the growth curve model was calculated by Kamo and Yoshimoto (2013) as

$$df(\mu) = q(\mu) + \mathrm{tr}\left\{\left(\boldsymbol{I}_\mu(\hat{\boldsymbol{\theta}}_\mu) - \boldsymbol{J}_\mu(\hat{\boldsymbol{\theta}}_\mu)\right)^{-1} \boldsymbol{I}_\mu(\hat{\boldsymbol{\theta}}_\mu)\right\},$$

where $\boldsymbol{I}_\mu(\hat{\boldsymbol{\theta}}_\mu)$ and $\boldsymbol{J}_\mu(\hat{\boldsymbol{\theta}}_\mu)$ are matrices given by

$$\boldsymbol{I}_\mu(\hat{\boldsymbol{\theta}}_\mu) = \frac{1}{n}\sum_{i=1}^n \frac{\partial \mu(t_i; \boldsymbol{\theta}_\mu)}{\partial \boldsymbol{\theta}_\mu} \frac{\partial \mu(t_i; \boldsymbol{\theta}_\mu)}{\partial \boldsymbol{\theta}'_\mu}\bigg|_{\boldsymbol{\theta}_\mu = \hat{\boldsymbol{\theta}}_\mu},$$

$$\boldsymbol{J}_\mu(\hat{\boldsymbol{\theta}}_\mu) = \frac{1}{n}\sum_{i=1}^n \left\{y(t_i) - \mu(t_i; \boldsymbol{\theta}_\mu)\right\} \frac{\partial^2 \mu(t_i; \boldsymbol{\theta}_\mu)}{\partial \boldsymbol{\theta}_\mu \partial \boldsymbol{\theta}'_\mu}\bigg|_{\boldsymbol{\theta}_\mu = \hat{\boldsymbol{\theta}}_\mu}.$$

In this paper, "$\boldsymbol{a}'$" denotes a transpose of a vector $\boldsymbol{a}$. Kamo and Yoshimoto (2013) proposed the following modified C_p (MC_p) expressed by replacing $q(\mu)$ with $df(\mu)$ in (1) as

$$MC_p(\mu) = \frac{\mathrm{RSS}(\hat{\boldsymbol{\theta}}_\mu; \mu)}{\hat{\sigma}^2_{\mathrm{L}}} + 2df(\mu).$$

The terminology "modified" means that the bias of $\mathrm{RSS}(\hat{\boldsymbol{\theta}}_\mu; \mu)/\hat{\sigma}^2_{\mathrm{L}}$ to the risk function is corrected even under an underspecified model. A modified C_p criterion was originally proposed by Fujikoshi and Satoh (1997) in the multivariate linear regression model. Since the MC_p was derived under the assumption that the candidate model may be an underspecified model, the MC_p may evaluate correctly the complexity of an underspecified model. If the candidate model considered is an overspecified model, then $df(\mu)$ converges to $q(\mu)$ as $n \to \infty$ in probability.

3.3. BIC-Type C_p Criterion

The Bayesian information criterion (BIC) proposed by Schwarz (1978) is one of famous MSCs. In the BIC, the penalty term is "(the number of parameters)$\times \log n$". By using $q(\mu)\log n$ instead of $2q(\mu)$ in (1), the BIC-type C_p (BC_p) can be proposed as

$$BC_p(\mu) = \frac{\mathrm{RSS}(\hat{\boldsymbol{\theta}}_\mu; \mu)}{\hat{\sigma}^2_{\mathrm{L}}} + q(\mu) \log n.$$

Recall our purpose of GF-selection is to choose the growth function so that the growth-prediction of the selected model will be improve. However, a consistency property, that a selection probability of the true model by MSC goes to 1 asymptotically, is also important property of the model selection. Since BIC has a consistency property, we can expect that BC_p also has a consistency property.

3.4. Generalized Cross-Validation Criterion

A generalized cross-validation (GCV) criterion proposed by Craven and Wahba (1979) is one of RSS-based MSCs. In GCV, the penalty to the complexity of a model is imposed not additively but multiplicatively. The GCV based the GDF was proposed by Ye (1998). The GCV for GF-selection is defined by

$$\mathrm{GCV}(\mu) = \frac{\mathrm{RSS}(\hat{\boldsymbol{\theta}}_\mu; \mu)}{\{1 - df(\mu)/n\}^2}. \tag{2}$$

If $\hat{\sigma}^2_{\mathrm{L}}$ does not work well, there are possibilities that C_p, MC_p and BC_p become instable. However, even if $\hat{\sigma}^2_{\mathrm{L}}$ does not work well, the GCV does not become instable because GCV in (2) is defined without an estimator of σ^2.

4. Numerical Study

4.1. Setting

In this section, we compare the performance each criteria by conducting numerical experiments under several sample sizes, variances and true growth functions. At first, we prepared the following twelve true growth functions:

Case 1: $\mu_*(t)$ is Bertalanffy as $\mu_*(t) = 100(1 - e^{-0.5t})^3$.

Case 2: $\mu_*(t)$ is Chapman-Richards as $\mu_*(t) = 100(1 - e^{-0.4t})^{3.8}$.

Case 3: $\mu_*(t)$ is Gompertz as $\mu_*(t) = 90 \exp(-0.4e^{-0.1t})$.

Case 4: $\mu_*(t)$ is Hossfeld-4 as $\mu_*(t) = 100(1 + 5t^{-1.5})^{-1}$.

Case 5: $\mu_*(t)$ is Korf as $\mu_*(t) = 100 \exp(-3t^{-1})$.

Case 6: $\mu_*(t)$ is Levakovic-3 as $\mu_*(t) = 100(1 + 5t^{-2})^{-1.5}$.

Case 7: $\mu_*(t)$ is Logistic as $\mu_*(t) = 100(1 + 5e^{-0.4t})^{-1}$.

Case 8: $\mu_*(t)$ is Monomolecular as $\mu_*(t) = 100(1 - 1.35e^{-0.25t})$.

Case 9: $\mu_*(t)$ is Weibull as $\mu_*(t) = 100(1 - e^{-0.6t^{0.7}})$.

Case 10: $\mu_*(t)$ is Levakovic-1 as $\mu_*(t) = 100(1 + 3t^{-2.3})^{-2}$.

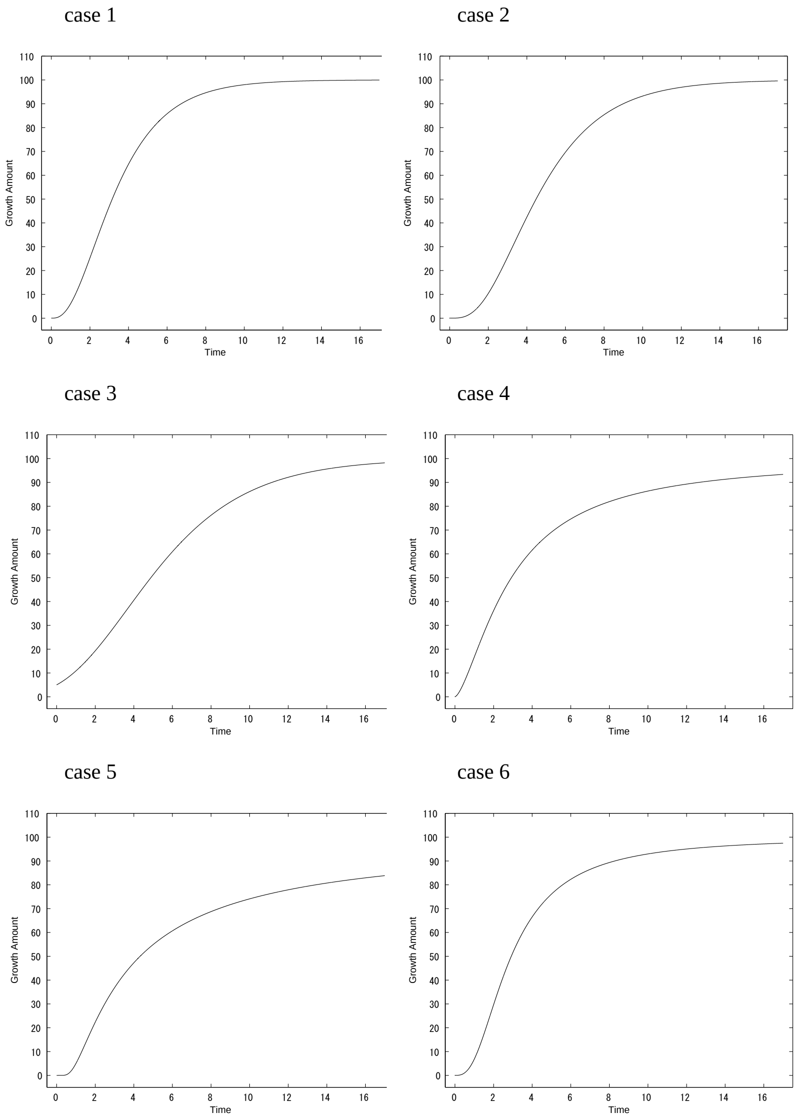

Figure 1. The shapes of the true growth curves (case 1 to case 6)

Case 11: $\mu_*(t)$ is Sloboda as $\mu_*(t) = 100\exp(-4e^{-0.5t^{0.8}})$.

Case 12: $\mu_*(t)$ is Yoshida-1 as $\mu_*(t) = 80(1 + 5t^{-1.4t})^{-1} + 20$.

We used $t_i = 2 + 18i/(n-1)\,(i = 1, \ldots, n)$ as time series with $n = 30, 50, 100, 300$ and 500, we generated error variables of the true model from $N(0, \sigma_*^2)$ with $\sigma_*^2 = 1$ and 2. The shapes of true growth curves are shown in figures 1 and 2. In this paper, we assessed performances of GF-selection methods by the following two properties that was derived from $1,000$ repetitions.

- The prediction error (PE) of the best growth function chosen by minimizing MSC.

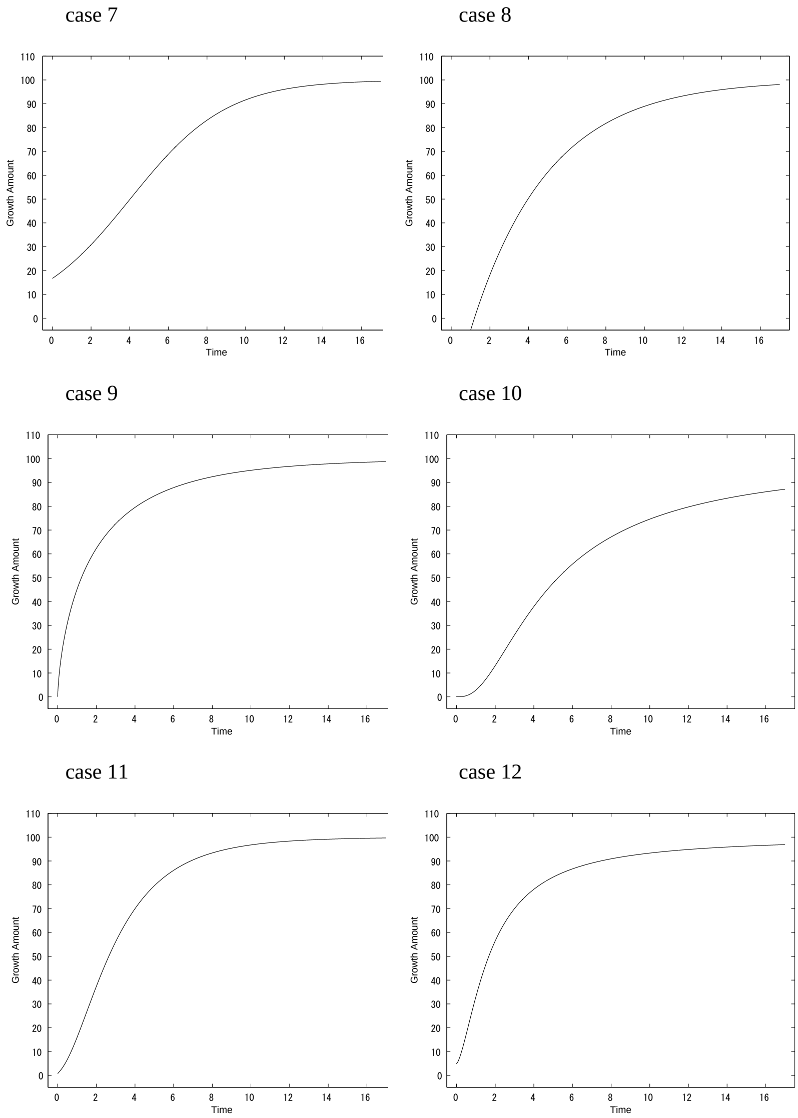

Figure 2. The shapes of the true growth curves (case 7 to case 12)

- The selection probability (SP) of the true growth function chosen by minimizing the criterion.

Here, the PE is defined by

$$\mathrm{PE} = \frac{1}{n} \sum_{j=n+1}^{n+3n/10} \left\{ \mu_*(t_j) - \hat{\mu}(t_j; \hat{\theta}_{\hat{\mu}}) \right\}^2 ,$$

where $t_j = 2 + 18j/(n-1)$. Note that PE is more important property, since the aim of our study is to select the growth function so that the growth prediction of the selection model will be improved.

Table 1. The prediction error under each case when $\sigma_*^2 = 1$

case	n	C_p	MC_p	BC_p	GCV	case	n	C_p	MC_p	BC_p	GCV
1*	30	1.13	1.14	**1.11**	1.14	7	30	1.32	1.34	**1.26**	1.33
	50	1.09	1.09	**1.06**	1.09		50	1.21	1.21	**1.15**	1.21
	100	1.04	1.04	**1.02**	1.04		100	1.10	1.10	**1.06**	1.10
	300	**1.01**	**1.01**	**1.01**	**1.01**		300	**1.02**	**1.02**	**1.02**	**1.02**
	500	1.01	1.01	**1.00**	1.01		500	**1.01**	**1.01**	**1.01**	1.02
2	30	**1.42**	1.43	**1.42**	1.43	8	30	1.53	1.56	**1.52**	1.55
	50	**1.23**	**1.23**	**1.23**	**1.23**		50	1.34	1.34	**1.30**	1.33
	100	1.09	1.09	**1.08**	1.09		100	1.15	1.15	**1.12**	1.15
	300	**1.02**	**1.02**	**1.02**	**1.02**		300	1.04	1.04	**1.03**	1.04
	500	**1.01**	1.03	**1.01**	1.02		500	1.02	1.03	**1.01**	1.03
3*	30	1.53	1.53	**1.41**	1.54	9	30	**1.40**	1.45	**1.40**	1.45
	50	1.33	1.33	**1.22**	1.34		50	**1.29**	1.31	**1.29**	1.31
	100	1.18	1.18	**1.11**	1.17		100	**1.15**	1.17	**1.15**	1.16
	300	1.06	1.06	**1.03**	1.05		300	**1.05**	**1.05**	**1.05**	**1.05**
	500	1.02	1.02	**1.01**	1.03		500	1.03	1.03	**1.02**	1.03
4*	30	1.49	1.49	1.49	**1.48**	10	30	**1.22**	1.25	1.23	1.25
	50	1.29	**1.27**	1.29	1.28		50	**1.15**	1.17	1.16	1.17
	100	1.13	1.13	1.13	**1.12**		100	**1.07**	1.08	1.09	1.08
	300	1.03	1.03	**1.02**	1.03		300	**1.02**	**1.02**	1.03	**1.02**
	500	1.02	1.02	**1.01**	1.02		500	1.14	1.14	1.16	**1.13**
5	30	**1.36**	**1.36**	**1.36**	**1.36**	11	30	**1.94**	2.01	**1.94**	2.03
	50	**1.21**	**1.21**	1.22	**1.21**		50	**1.68**	1.71	1.70	1.71
	100	**1.09**	**1.09**	**1.09**	1.10		100	**1.42**	1.45	1.52	1.45
	300	1.03	1.03	**1.02**	1.03		300	**1.26**	1.30	1.35	1.30
	500	**1.01**	1.02	**1.01**	1.02		500	**1.04**	**1.04**	1.05	1.05
6*	30	**1.31**	**1.31**	**1.31**	**1.31**	12	30	1.60	**1.58**	1.60	**1.58**
	50	1.17	**1.16**	**1.16**	**1.16**		50	**1.43**	**1.43**	1.44	**1.43**
	100	**1.06**	**1.06**	**1.06**	**1.06**		100	1.27	**1.26**	1.30	**1.26**
	300	**1.02**	**1.02**	**1.02**	**1.02**		300	**1.12**	**1.12**	1.24	**1.12**
	500	**1.01**	**1.01**	**1.01**	**1.01**		500	**1.02**	**1.02**	**1.02**	**1.02**

4.2. Results

Tables 4.2 and 4.2 show PEs of the best growth functions when $\sigma_*^2 = 1$ and 2, respectively, and tables 4.2 and 4.2 show SPs of the true growth functions when $\sigma_*^2 = 1$ and 2, respectively. The number in the column named "case" shows which growth function used as the true growth function. For example, the number 1 indicates that simulation data are generated from the true growth function of the case 1, i.e., Bertalanffy. Furthermore, * denotes the case that there is the overspecified model. In the tables, bold fonts indicate the smallest PEs of the best growth functions, and the highest SPs of the true growth functions.

From tables, we obtained the following results:

- When the number of parameters of the true growth function was not large, i.e., cases 1 to 9, BC_p was the high-performance MSC in most cases. In particular, when the sample size was not

Table 2. The prediction error under each case when $\sigma_*^2 = 2$

case	n	C_p	MC_p	BC_p	GCV	case	n	C_p	MC_p	BC_p	GCV
1*	30	2.53	2.57	**2.43**	2.57	7	30	3.20	3.24	**3.09**	3.25
	50	2.39	2.40	**2.27**	2.40		50	2.85	2.87	**2.67**	2.87
	100	2.16	2.17	**2.11**	2.17		100	2.48	2.49	**2.31**	2.49
	300	2.06	2.06	**2.03**	2.06		300	2.13	2.13	**2.08**	2.13
	500	2.03	2.06	**2.02**	2.06		500	2.06	2.06	**2.04**	2.07
2	30	**3.47**	3.55	3.51	3.57	8	30	4.12	4.26	**4.03**	4.30
	50	**2.91**	2.95	2.95	2.95		50	**3.49**	3.61	**3.49**	3.59
	100	**2.50**	2.52	**2.50**	2.52		100	2.81	2.83	**2.76**	2.84
	300	2.14	2.14	**2.12**	2.14		300	2.22	2.22	**2.16**	2.21
	500	2.06	2.12	**2.04**	2.11		500	2.11	2.12	**2.08**	2.12
3*	30	3.83	3.92	**3.79**	3.95	9	30	**3.18**	3.22	**3.18**	3.24
	50	3.00	3.10	**2.88**	3.09		50	**2.80**	2.84	**2.80**	2.84
	100	2.52	2.53	**2.34**	2.54		100	**2.43**	2.48	**2.43**	2.48
	300	2.23	2.23	**2.12**	2.23		300	**2.18**	2.21	**2.18**	2.21
	500	2.11	2.14	**2.06**	2.15		500	**2.12**	2.14	2.13	2.14
4*	30	3.53	**3.49**	3.55	3.50	10	30	**2.76**	2.78	2.92	2.79
	50	2.99	**2.96**	2.98	**2.96**		50	**2.46**	2.47	2.51	2.47
	100	2.56	**2.55**	**2.55**	**2.55**		100	**2.25**	2.27	2.26	2.27
	300	**2.18**	**2.18**	**2.18**	**2.18**		300	**2.09**	2.11	2.11	2.11
	500	**2.10**	**2.10**	2.11	**2.10**		500	2.38	2.38	2.72	**2.36**
5	30	3.58	**3.52**	3.59	3.53	11	30	**4.36**	4.56	4.44	4.60
	50	2.95	**2.92**	2.95	**2.92**		50	**3.57**	3.69	3.62	3.71
	100	**2.51**	**2.51**	**2.51**	**2.51**		100	**2.95**	3.03	3.02	3.03
	300	**2.13**	**2.13**	**2.13**	**2.13**		300	**2.53**	2.56	2.57	2.56
	500	2.08	2.08	**2.07**	2.08		500	**2.10**	2.13	**2.10**	2.12
6*	30	3.12	3.05	3.14	**3.04**	12	30	3.59	**3.51**	3.59	3.53
	50	2.69	2.68	2.70	**2.67**		50	3.06	**3.04**	3.07	**3.04**
	100	2.34	2.34	**2.33**	2.34		100	2.66	**2.65**	2.67	2.66
	300	**2.10**	**2.10**	2.11	**2.10**		300	2.33	**2.31**	2.34	**2.31**
	500	**2.05**	**2.05**	**2.05**	**2.05**		500	**2.08**	**2.08**	2.09	**2.08**

small, SPs of the true growth function by BC_p was always highest among all the MSCs. The differences of SPs were large in the case that the overspecified model exists, i.e., cases 1, 3, 4 and 6. This is because BC_p has a consistency and C_p, MC_p and GCV do not have a consistency, i.e., the SPs of BC_p converge to 1 asymptotically although those of C_p, MC_p and GCV do not in the cases 1, 3, 4 and 6.

- When the number of parameters of the true growth function was large, i.e., cases 10 to 12, BC_p was not the high-performance MSC. This is because the penalty term of BC_p is too large in the cases 10 to 12. In general, BC_p tends to choose the model having the smaller number of known parameters than the true model. Reversely, C_p, MC_p and GCV tend to choose the model having the larger number of known parameters than the true model. In the cases 10 to 12, there were no models having the larger number of known parameters than the true model. Hence, the SPs of C_p, MC_p and GCV tended to be higher than those of BC_p. Although PEs of the best models

Table 3. The selection probability under each case when $\sigma_*^2 = 1$

case	n	C_p	MC_p	BC_p	GCV	case	n	C_p	MC_p	BC_p	GCV
1*	30	71.9	71.8	**85.5**	71.1	7	30	80.8	79.7	**89.0**	79.4
	50	72.7	72.1	**90.9**	73.3		50	85.5	85.0	**93.8**	85.2
	100	74.8	74.7	**93.8**	74.9		100	89.6	89.3	**97.6**	89.5
	300	81.0	80.7	**97.9**	80.6		300	93.1	93.0	**99.5**	92.7
	500	82.8	83.0	**98.9**	64.7		500	89.3	89.2	**99.3**	77.6
2	30	74.1	74.1	**78.8**	74.2	8	30	76.1	75.3	**77.9**	75.7
	50	79.2	79.3	**85.3**	80.3		50	80.8	80.3	**83.6**	80.7
	100	88.7	89.3	**95.1**	89.2		100	88.2	88.2	**90.1**	88.2
	300	93.8	94.0	**98.8**	93.8		300	97.1	96.8	**97.9**	96.8
	500	97.2	96.5	**98.9**	95.1		500	98.8	98.3	**99.3**	98.2
3*	30	63.1	63.2	**74.2**	63.8	9	30	25.9	18.0	**26.0**	17.9
	50	67.0	67.0	**80.9**	66.8		50	28.2	21.3	**28.5**	21.1
	100	73.8	73.6	**89.6**	73.6		100	38.2	32.0	**38.8**	32.0
	300	77.0	77.0	**96.1**	77.5		300	52.8	50.2	**58.6**	49.9
	500	88.6	88.3	**98.5**	81.2		500	60.8	60.3	**67.0**	55.9
4*	30	57.3	55.2	**57.7**	55.5	10	30	2.3	4.0	0.2	**5.6**
	50	70.3	67.3	**70.9**	67.7		50	**12.7**	12.1	1.6	11.7
	100	80.0	75.7	**83.6**	76.0		100	**38.6**	36.2	7.2	36.4
	300	81.2	75.6	**98.6**	75.5		300	**77.5**	77.3	55.7	77.1
	500	87.2	77.2	**98.9**	68.4		500	49.3	50.1	46.7	**50.4**
5	30	85.3	55.7	**87.5**	56.1	11	30	**1.2**	**1.2**	0.5	**1.2**
	50	87.9	56.9	**90.5**	57.2		50	4.6	**4.8**	1.0	4.5
	100	89.1	55.4	**95.7**	54.9		100	12.9	13.2	2.9	**13.4**
	300	87.4	52.8	**98.4**	52.4		300	24.9	25.3	15.2	**25.6**
	500	95.8	79.4	**99.6**	77.1		500	61.1	61.6	38.0	**64.4**
6*	30	54.8	54.1	**55.4**	54.5	12	30	1.5	**4.9**	0.0	3.3
	50	63.7	63.1	**65.6**	63.0		50	3.6	**6.5**	0.1	5.3
	100	71.4	70.3	**77.4**	70.8		100	12.8	**17.5**	0.3	17.1
	300	83.5	81.9	**90.4**	82.1		300	**53.8**	52.3	11.8	52.0
	500	88.3	88.1	**95.2**	82.3		500	8.7	10.8	2.8	**16.7**

chosen by C_p, MC_p and GCV tended to be smaller than those chosen by BC_p, the differences were not so large.

From the simulation results, the use of the selection method using BC_p for a selection of growth functions.

References

Casser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression model. *Biometrika*, **73**, 625–633.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31**, 377–403.

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

Kamo, K. and Yoshimoto, A. (2013). Comparative analysis of GFs based on Mallows' C_p type criterion. *FORMATH*, **12**, 133–147.

Table 4. The selection probability under each case when $\sigma_*^2 = 2$

case	n	C_p	MC_p	BC_p	GCV	case	n	C_p	MC_p	BC_p	GCV
1*	30	62.9	61.6	**79.1**	62.1	7	30	65.8	64.2	**75.0**	64.7
	50	65.2	64.0	**85.5**	65.1		50	72.4	71.1	**83.0**	70.5
	100	69.1	68.1	**91.1**	68.2		100	76.2	75.3	**91.6**	76.4
	300	75.7	75.1	**95.8**	75.3		300	87.6	87.0	**98.0**	87.3
	500	79.2	78.6	**98.4**	73.2		500	86.5	86.4	**98.8**	82.3
2	30	**46.2**	45.5	44.9	**46.2**	8	30	48.5	46.0	**48.7**	45.7
	50	56.7	56.1	**56.9**	56.2		50	54.3	52.9	**54.6**	53.0
	100	69.9	69.5	**73.6**	69.9		100	66.0	64.8	**67.1**	65.3
	300	84.6	84.6	**93.8**	84.9		300	83.3	83.3	**86.6**	83.4
	500	90.6	88.3	**98.5**	86.8		500	89.7	89.1	**93.5**	88.4
3*	30	50.9	50.7	**57.8**	51.1	9	30	**10.9**	7.6	**10.9**	7.6
	50	54.2	53.7	**66.1**	53.5		50	14.6	8.2	**14.8**	8.0
	100	61.2	61.2	**76.4**	61.3		100	19.9	12.9	**20.0**	12.9
	300	72.2	72.3	**90.0**	72.1		300	35.2	28.8	**35.5**	28.8
	500	84.4	83.0	**95.2**	79.1		500	42.5	39.9	**42.8**	39.5
4*	30	29.8	27.7	**30.4**	27.9	10	30	1.0	**7.7**	0.5	7.4
	50	37.5	35.0	**38.1**	35.0		50	1.8	7.3	0.6	**7.4**
	100	47.5	45.1	**48.4**	45.1		100	5.2	**10.1**	0.9	10.0
	300	74.2	71.0	**75.0**	71.5		300	**22.5**	20.0	1.1	19.6
	500	83.6	72.7	**86.3**	70.0		500	35.0	37.0	13.0	**38.1**
5	30	69.6	40.3	**71.1**	40.2	11	30	0.1	0.2	0.1	**0.3**
	50	71.7	45.6	**73.2**	45.6		50	0.0	**0.1**	0.0	**0.1**
	100	78.2	47.9	**80.9**	47.9		100	0.1	**0.3**	0.0	**0.3**
	300	86.6	54.1	**93.7**	54.3		300	**5.8**	**5.8**	0.3	5.6
	500	91.5	75.1	**97.7**	74.5		500	17.9	18.3	0.6	**24.1**
6*	30	28.3	28.4	28.4	**28.5**	12	30	0.5	**2.3**	0.0	2.1
	50	37.5	37.3	**37.6**	**37.6**		50	0.7	**3.1**	0.0	3.0
	100	47.0	45.9	**47.1**	45.7		100	1.1	4.3	0.0	4.0
	300	65.9	64.3	**68.7**	64.4		300	5.8	**10.3**	0.0	9.4
	500	74.1	73.3	**78.7**	70.8		500	2.6	6.7	0.3	**7.7**

Mallows, C. L. (1973). Some Comments on C_p. *Technometrics*, **15**, 611–675.
Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.
Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *J. Amer. Statist. Assoc.*, **93**, 120–131.
Zeide, B. (1993). Analysis of growth equations. *Forest Sci.* **39**, 594–616.