Ridge Estimate Application to Growth Function

Ken-ichi Kamo¹*, Hirokazu Yanagihara²

Abstract: Growth functions are often used to describe longitudinal growth processes. In growth analysis, it is first necessary to estimate the parameters included in the growth function based on real life data. During the numerical estimation process, this is often accomplished using iterative algorithm. The initial value setting is the key to this iterative algorithm process. An improper setting of the initial value is risky, because it may lead to non-convergence. Another element of risk is the “convergence of an improper solution” without an error message. Under such a scenario, because the estimation process is correctly completed, it is difficult to identify the source of the error. To resolve this problem, we focus our attention to ridge estimation and how it applies to growth analysis. Ridge estimation method can be used to stabilize the estimation process by imposing the ℓ₂-norm condition as a constraint. In this paper we evaluated the performance of the ridge estimation technique using two cases of numerical experiments. Under several settings of the initial value, we were able to evaluate the behavior of the estimate, especially for dispersion. We concluded that the estimate from the ridge estimation process is more stable compared to the residual sum of square approach. Actually the mean squared error value in ridge estimate is smaller than the residual sum of square. Although ridge estimation method is not universal, this paper has shown us that it is possible to extend the range of setting initial values that converge to a proper estimate.

Keywords: growth function, numerical experiment, parameter estimate, ridge estimate

1. Introduction

Growth curves are found in a wide range of disciplines, such as biology, forestry and fishery research. Most living matter grows with successive lags, growth, and asymptotic phases; examples of quantities that follow such growth curves are tree growth and the extent of a population of fish. The role of the growth function is not only to describe the growth process, but to remove random fluctuations from the actual growth behavior and aggregate its characteristics into a function. The growth function is expected to reflect the underlying growth process as well as have high predictive ability. With the understanding that most growth curves are sigmoidal in shape, various growth functions have been proposed depending on the assumptions and constraints for longitudinal behavior. Zeide (1993) compared and investigated 12 types of growth functions. Each of these growth functions has parameters which characterize the property of the function. In the growth analysis, it is first necessary to determine the parameters based on real life growth data. Basic statistical approaches like least squared method or maximum likelihood method are employed in parameter estimation. In a numerical estimation process, the results are obtained through iterative algorithm (Nelder and Mead 1965). These iterative algorithms need a proper initial value-setting. If an improper initial value is set, there is a high risk that the iterative process would not converge. When this happens, the problem can easily be recognized because it will give rise to an error message, etc. However, there is the possibility that it may lead to an unfortunate situation where there may be “convergence of an improper solution” without an error message, even though the estimation process was correctly completed. On a small scale analysis, the error or defect can visually be identified through careful observations, however, with large scale analysis there is a high risk of overlooking the error, because the solution appears to have been converged. This scenario is explained in the example that follows:

Table 1 shows two types of virtual growth data for ages: 2, 3, ...,10. One of which represents a projected growth curve generated using the Bertalanffy growth function (Bertalanffy 1938), and the other, from the Gompertz growth function (Gompertz 1825). In this example, our focus is on the observations for the Bertalanffy growth function (see plot in Figure 1). We tried to fit the Bertalanffy growth function

\[ \text{Bert} \left( t : \theta = (\alpha, \beta) \right) = \alpha \left( 1 - e^{-\beta t} \right)^3 \]

Received Apr. 12, 2021; Accepted Apr. 29, 2021

¹Sapporo Medical University, Japan
²Hiroshima University, Japan
*Corresponding Author: kamo@sapmed.ac.jp
Kamo & Yanagihara

to the observation in Table 1, where \(t\) represents the independent variable related to the time element, and \(\mathbf{\theta} = (\alpha, \beta)'\) is the parameter which characterizes the property of the function. The parameter \(\alpha\) corresponds to the upper bound for the growth curve and \(\beta\) corresponds to the speed of the growth curve. Setting the initial value properly (for example \((\alpha, \beta) = (90, 0.3)\)) leads to a proper or true estimate of \((\hat{\alpha}, \hat{\beta}) = (102.258, 0.393)\) (dotted line “Proper estimate” in Figure 1). Here the hat-symbol denotes the estimate and underbar-symbol denotes the initial value. On the other hand, an improper setting of initial value (for example \((\alpha, \beta) = (20, 5)\)) leads to an improper estimate of \((\hat{\alpha}, \hat{\beta}) = (67.152, 12.657)\) (dotted line “Improper estimate” in Figure 1). It should be noted here that even in the latter estimate, the iterative process converges and no error is reported (We use “optim” function in R. When we use “nls” function, the error occurs). In other words, the fact that the estimates improperly converged is troublesome in that it is not revealed unless under a careful observation. This is an extreme and easy-to-understand example, but other initial values may lead to local convergence that is not so far from the true estimate. In this paper, we look at this problem through the lens of growth analysis. The above observations were made by using R, and the related script is shown in the Appendix.

Table 1. Virtual observation for Bertalanffy and Gompertz growth.

| Age | Bertalanffy | Gompertz |
|-----|-------------|----------|
| 2   | 14.7        | 13.9     |
| 3   | 34.7        | 33.4     |
| 4   | 48.2        | 48.2     |
| 5   | 69.7        | 71.4     |
| 6   | 76.2        | 79.0     |
| 7   | 80.2        | 83.4     |
| 8   | 89.8        | 92.8     |
| 9   | 94.4        | 96.9     |
| 10  | 96.4        | 98.5     |

Note) We set two situations, one is Bertalanffy \(y = 100(1 - e^{-0.4t})^3\), another is Gompertz \(y = 100e^{-5e^{-0.5t}}\). From these true settings, we create the virtual growth observation at age \(2, 3, \ldots, 10\) by adding random errors from \(N(0, 10)\) to the expected values.

In order to solve such a problem, we focus our attention on the ridge estimate proposed in Hoerl and Kennard(1970). In this paper, ridge estimate is a concept that attempts to stabilize the estimation process by imposing the \(\ell_2\)-norm condition as a constraint. Ridge estimate is often applied when the number of samples is small or when the explanatory variables are highly correlated with each other, thus making the estimate tends unstable. For example, in linear regression model \(y = X\beta + \epsilon\), the estimate of parameter \(\beta\) can be obtained from the expression: \(\hat{\beta} = (X'X)^{-1}X'y\). On the other hand, in ridge estimate, we adopt ridge parameter \(\lambda > 0\) and the estimate can be obtained from the expression: \(\hat{\beta}_\lambda = (X'X + \lambda I_p)^{-1}X'y\) (Hoerl and Kennard 1970). Here, \(y, \beta\) and \(\epsilon\) are the vectors of the response variable, parameter and error, and \(X\) and \(I_p\) are the matrices of the explanatory variable and \(p\)-dimensional unit matrix, respectively. Here, we apply this concept to the growth function. A detailed description of the mathematical formulation is shown in the section that follows.

2. Methods

2.1. Growth Model and Parameter Estimate

Let us describe the growth model as

\[ E[y(t_i)] = f(t_i | \theta) \quad (i = 1, 2, \ldots, n), \]

where \(y(t_i)\) is the growth amount at a given time \(t_i\), \(E[\cdot]\) represents the expected value, and \(n\) is the number of observations over time. On the right-hand side of equation [2], \(f(\cdot)\) is the growth function and \(\theta\) is a \(p\)-dimensional unknown parameter vector included in the growth function. Since
The solid line denotes the Gompertz function \( y = 100(1 - e^{-0.41})^3 \) and the dots denote the artificial observations. Two types of dotted lines are estimated curve, one is proper estimate \( (\hat{\alpha}, \hat{\beta}) = (102.258, 0.393) \) from the proper initial value \( (\alpha, \beta) = (90, 0.3) \), another is improper one \( (\check{\alpha}, \check{\beta}) = (67.152, 12.657) \) from the initial value \( (\alpha, \beta) = (20, 5) \).

The residual at time \( t_i \) is expressed as \( e_i(\theta) = y(t_i) - f(t_i \mid \theta) \), then parameter \( \theta \) can be estimated by minimizing residual sum of square (RSS)

\[
\text{RSS}(\theta) = \sum_{i=1}^{n} e_i(\theta)^2
\]

that is, \( \hat{\theta} = \arg\min_{\theta} \text{RSS}(\theta) \). Hereafter, we refer to such estimate as RSS-estimate. The value \( \hat{\theta} \) is known to have desirable asymptotic properties, consistency, asymptotic unbiasedness, and asymptotic normality, under some proper conditions.

2.2. Ridge Estimation

In ridge estimation, we consider the following PRSS (penalized residual sum of square)

\[
\text{PRSS}(\theta \mid \lambda) = \sum_{i=1}^{n} e_i(\theta)^2 + \lambda \|\theta\|^2
\]

instead of [3]. Here \( \lambda > 0 \) is ridge parameter and \( \|\cdot\| \) is \( \ell_2 \)-norm. Then, ridge estimation for the parameter in growth function is expressed as \( \hat{\theta}_{\lambda} = \arg\min_{\theta} \text{PRSS}(\theta \mid \lambda) \). Hereafter, we refer to such estimate as Ridge-estimate. Let \( g(t \mid \theta) \) and \( H(t \mid \theta) \) be the \( p \)-dimensional gradient vector and the \( p \times p \) Hessian matrix of \( f(t \mid \theta) \), respectively, which are defined as

\[
g(t \mid \theta) = \frac{\partial}{\partial \theta} f(t \mid \theta) \bigg|_{\theta = \theta} \quad \text{and} \quad H(t \mid \theta) = \frac{\partial^2}{\partial \theta \partial \theta^T} f(t \mid \theta) \bigg|_{\theta = \theta}
\]

and let \( I(\theta) \) and \( J(\theta) \) be \( p \times p \) matrices defined by
The equation \[7\] is derived as follows: Since \[8\] and \[9\] are satisfied:

\[ \frac{\partial}{\partial \theta} \left( \frac{1}{n} \sum_{i=1}^{n} e_i(\theta) H(t_i \mid \theta) \right) = \frac{1}{n} \sum_{i=1}^{n} e_i(\theta) \frac{\partial}{\partial \theta} H(t_i \mid \theta). \]

Note that Hessian matrices of \[3\] and \[4\] are \(n\{I(\theta) - J(\theta)\} + n\{I(\theta) - J(\theta)\} + \lambda I_p\). These Hessian matrices indicate that \[3\] does not guarantee convexity with respect to \(\theta\), while \[4\] does, when \(\lambda\) is greater than a certain value. We can also see from this result that the ridge estimation will be more stable than the RSS-estimation, because for convex functions there is only one local minimum. Unfortunately, it is widely known that \(\hat{\theta}_\lambda\) has a bias. However, under a fixed \(\lambda\), the difference between \(\theta\) and \(\hat{\theta}_\lambda\) converges to 0 with probability as \(n \to \infty\), because the following approximation holds:

\[ \hat{\theta}_\lambda \approx \hat{\theta} - \frac{\lambda}{n} \left\{ I(\hat{\theta}) - J(\hat{\theta}) + \frac{\lambda}{n} I_p \right\}^{-1} \hat{\theta}. \]

This indicates that \(\hat{\theta}_\lambda\) will have the same desired asymptotic properties as \(\hat{\theta}\) under a suitable \(\lambda\). The equation \[7\] is derived as follows: Since \(\hat{\theta}_\lambda\) is the minimizer of \[4\], the following equation is satisfied:

\[ 0_p = -\sum_{j=1}^{n} e_j(\hat{\theta}_\lambda) g(t_j \mid \hat{\theta}_\lambda) + \lambda \hat{\theta}_\lambda, \]

where \(0_p\) is the \(p\)-dimensional vector of zeros. By applying Taylor expansion around \(\hat{\theta}\) to the equation \[8\] and using the result \(0_p = -\sum_{j=1}^{n} e_j(\hat{\theta}) g(t_j \mid \hat{\theta})\) derived from the equation \[8\] with \(\lambda = 0\), we obtain equation \[7\].

Suppose \(g(t_1), \ldots, g(t_n)\) are independent random variables and \(g(t_i)\) is distributed according to a normal distribution with mean, \(f(t_i \mid \theta)\) and variance, \(\sigma^2\). In order to determine ridge parameter \(\lambda\), we minimize the following generalized cross-validation (GCV) criterion with the generalized degree of freedom (GDF) proposed by Ye (1998) as

\[ \text{GCV}(\lambda) = \left(1 - \frac{d_\lambda}{n}\right)^{-2} \sum_{i=1}^{n} \left( y(t_i) - f(t_i \mid \hat{\theta}_\lambda) \right)^2, \]

where \(d_\lambda\) is the generalized degree of freedom, which is defined as \(d_\lambda = \sum_{i=1}^{n} \frac{\partial f(t_i \mid \hat{\theta}_\lambda)}{\partial y_i} \). It is well known that the original GCV criterion was proposed by Craven and Wahba (1979). The GDF for the non-linear regression was derived by Kamo and Yoshimoto (2013). By using the GDF, Fukui, Yamamura and Yanagihara (2015) proposed the GCV criterion with GDF, for selecting growth functions. The GDF in the ridge estimation is obtained by the same derivation as in Kamo and Yoshimoto (2013). We will derive the specific from of \(d_\lambda\) in the following. Differentiating both sides of \[8\] with respect to \(y_i\) to yield

\[ 0_p = -g(t_i \mid \hat{\theta}_\lambda) + n I(\hat{\theta}_\lambda) \frac{\partial \hat{\theta}_\lambda}{\partial y_i} - n J(\hat{\theta}_\lambda) \frac{\partial \hat{\theta}_\lambda}{\partial y_i} + \lambda \frac{\partial \hat{\theta}_\lambda}{\partial y_i}. \]

The equation \[10\] implies that

\[ \frac{\partial \hat{\theta}_\lambda}{\partial y_i} = \frac{1}{n} \left\{ I(\hat{\theta}_\lambda) - J(\hat{\theta}_\lambda) + \frac{\lambda}{n} I_p \right\}^{-1} g(t_i \mid \hat{\theta}_\lambda). \]

Hence, we have

\[ \frac{\partial}{\partial y_i} f(t_i \mid \hat{\theta}_\lambda) = \frac{\partial \hat{\theta}_\lambda}{\partial y_i} g(t_i \mid \hat{\theta}_\lambda) = \frac{1}{n} g(t_i \mid \hat{\theta}_\lambda)^\prime \left\{ I(\hat{\theta}_\lambda) - J(\hat{\theta}_\lambda) + \frac{\lambda}{n} I_p \right\}^{-1} g(t_i \mid \hat{\theta}_\lambda). \]
Consequently, the GDF in the ridge estimation is given by

\[
d_{\lambda} = \sum_{i=1}^{n} \frac{\partial}{\partial y_i} f(t_i \mid \hat{\theta}_\lambda) = \frac{1}{n} \sum_{i=1}^{n} g(t_i \mid \hat{\theta}_\lambda)' \left\{ I(\hat{\theta}_\lambda) - J(\hat{\theta}_\lambda) + \frac{\lambda}{n} I_p \right\}^{-1} g(t_i \mid \hat{\theta}_\lambda)
\]

where \( \text{tr} \) is a trace operator.

3. Numerical Examinations

In this section, we present the results for applying the above procedures to artificially generated observations presented in Table 1. These observations were created by adding errors to the two growth functions; Bertalanffy and Gompertz. These errors were generated from \( N(0, 10) \). The Bertalanffy function is defined in equation [2] and the Gompertz function is defined as:

\[
\text{Gompertz: } Gomp(t : \theta = (\alpha, \beta, \gamma)) = \alpha \exp(-\beta e^{-\gamma t})
\]

We set up two scenarios based on the true parameter, \((\alpha, \beta) = (100, 0.4)\). In Bertalanffy, (Case-1) and \((\alpha, \beta, \gamma) = (100, 5, 0.5)\) and in Gompertz, (Case-2), i.e.,

\[\begin{align*}
\text{Case-1} & \quad \text{Bert}(t : 100, 0.4) = 100(1 - e^{-0.4t})^3, \\
\text{Case-2} & \quad \text{Gomp}(t : 100, 0.5) = 100\exp(-5e^{-0.5t}).
\end{align*}\]

From the above two settings, artificial observations along with ages 2, 3, ..., 10 were used to generate the results in Table 1. By using the information in Table 1, we compared RSS- and Ridge- estimate using numerical experiments.

In Case-1, under the proper initial value setting, we can obtain the proper estimate, which is shown as thicker dotted lines “Proper estimate” in Figure 1. On the other hand, improper initial value converges to improper estimate, which is shown as thinner dotted lines “Improper estimate” in Figure 1. When we apply ridge- estimation to this improper initial value, then the estimate is adjusted as close to the true parameters. On the other hand, in Case-2, under the proper initial value setting, we obtain the proper estimate, which is shown as thicker dotted lines “Proper estimate” in Figure 2. On the other hand, improper initial value converges to improper estimate, which is shown as thinner dotted lines “Improper estimate” in Figure 2. When we apply ridge- estimate to this improper initial value, then the estimate is adjusted as close to the true parameters. Table 2 summarize parameter estimates and estimated ridge parameter (\( \lambda \) in [4]).

| Function          | Parameter | True | Proper Initial Estimate | Proper RSS estimate | Improper Ridge estimate |
|-------------------|-----------|------|-------------------------|---------------------|-------------------------|
| Case-1 (Bertalanffy) | \( \alpha \) | 100  | 90 102.258              | 20 67.152           | 102.13                  |
|                   | \( \beta \) | 0.4  | 0.3 0.393               | 5 12.057            | 0.393                   |
|                   | \( \lambda \) | 0.0017 |
| Case-2 (Gompertz)  | \( \alpha \) | 100  | 90 100.871              | 80 68.604           | 100.656                 |
|                   | \( \beta \) | 5    | 4 5.435                 | 5 19.866            | 5.475                   |
|                   | \( \gamma \) | 0.5  | 0.3 0.519               | 5 9.053             | 0.522                   |
|                   | \( \lambda \) | 0.0023 |

Figures 3 and 4, show the behavior of estimate depending on the initial value setting. Figure 3 corresponds to Case-1 and Figure 4 to Case-2. Figure 3 is made up of 8 figures, which include 2 parameter estimate (a)-(d) for \( \alpha \) and (e)-(h) for \( \beta \), 2 pattern for initial value ((a), (c), (e), (g) for \( \alpha \) and (b), (d), (f), (h) for \( \beta \) and 2 pattern for the way for estimate ((a), (b), (e), (f) for
The solid line denotes the Bertalanffy function \( y = 100 \exp(-5e^{-0.5t}) \) and the dots denote the artificial observations. Two types of dotted lines are estimated curve, one is proper estimate \((\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (100.871, 5.435, 0.519)\) from the proper initial value \((\alpha, \beta, \gamma) = (90, 4, 0.3)\), another is improper one \((\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (68.604, 19.866, 9.053)\) from the initial value \((\alpha, \beta, \gamma) = (80, 5, 5)\).

RSS-estimate and (c), (d), (g), (h) for Ridge-estimate. Then Figure 3 includes \(2 \times 2 \times 2 = 8\) figures. The same is true for Case-2 and Figure 4. Figure 4 therefore includes \(2 \times 3 \times 3 = 18\) figures.

In Case-1, the range of initial value is set as \(\alpha \in [60, 140]\) and \(\beta \in [0.25, 1]\). We obtain the RSS and Ridge-estimate from the combination of \(\alpha\) and \(\beta\). Figure 3 shows that while there are a few outlier-type estimates in RSS-estimate, Ridge-estimate has no such estimate within this range. Regarding the outlier in RSS-estimate, it was observed that \(\hat{\alpha}\) tends to deviate downward, while \(\hat{\beta}\) tends to deviate upward. In Case-2, the range of initial value is set as \(\alpha \in [70, 130], \beta \in [2.2, 10]\) and \(\gamma \in [0.22, 1]\). No extreme outlier (as in Case 1) was observed, but for range variability differences were observed between RSS- and Ridge-estimate. It was observed that Ridge-estimate had less variability. For all three parameters \(\alpha, \beta\) and \(\gamma\), RSS-estimate generated a certain number of estimates that deviated from the appropriate estimates.
Figure 3. Estimated parameters for Bertalanffy.

Figures (a) and (b) are RSS-based $\hat{\alpha}$ (vertical axis) corresponding to the initial value $\alpha$ and $\beta$ (horizontal axis). Figures (c) and (d) are ridge-based $\hat{\alpha}$ corresponding to initial value $\alpha$ and $\beta$. Concerning the remaining figures (e)-(h), the arrangement is similar for $\beta$. The scales of the graph are consistent among same parameter for comparison.
Figure 4. Estimated parameters for Gompertz.

Figures (a), (b), (c) are RSS-based $\hat{a}$ (vertical axis) corresponding to the initial value $\alpha$, $\beta$ and $\gamma$ (horizontal axis). Figures (d),(e),(f) are ridge-based $\hat{a}$ corresponding to initial value $\alpha$, $\beta$ and $\gamma$. Concerning the remaining figures (g)-(l) are the same for $\beta$ and (m)-(r) are the same for $\gamma$. The scales of the graph are consistent among same parameter for comparison.
4. Conclusion

Let us consider the source of the error from an improper initial value setting of a nonlinear regression that leads to improper estimate, using the Case-1 scenario. The characteristics of Bertalanffy growth function is controlled by two parameters $\alpha$ and $\beta$. Figure 5 is a heatmap showing how RSS behaves depending on the parameters $\alpha$ and $\beta$. The horizontal axis is represented by $\alpha$ and the vertical axis is represented by $\beta$, and the RSS is expressed by shading and contour. The altitude of RSS in the heatmap is not a typical mortar-shaped concentric circle, centered on the true situation $(\alpha, \beta) = (100, 0.4)$. In the case of $\alpha$ as an example, there is a valley shaped bottom for RSS around $\alpha = 70$ regardless of the $\beta$ value. As observed before, the estimate from improper initial value $(\alpha, \beta) = (20, 5)$ converges to $\hat{\alpha} = 67.152$. This is because the iteration process cannot reach the best estimate after RSS decreases to a value around $\alpha = 70$ along the horizontal direction. Such apparent convergence to the local minimum of RSS depending on the initial values is the cause of the instability for the estimate.

Actually, when the numerical parameter estimation in a nonlinear function such as the growth function does not work well, the first countermeasure to be taken is to reselect the initial value based on previous experience. In such a scenario, a try and error process is repeated while considering the role or meaning of the parameters. This is a very effective way when the number of samples is not so large to pay attention to the characteristics of each sample. However, when there are many samples, to estimate everything at once (for example to apply the same estimation procedure to many samples, we use “for” in the loop statement), is risky such that the entire calculation process may stop as a result of just one error, or cannot identify improper convergence. Under such a situation, this paper applied ridge estimate approach to the growth function to stabilize the estimation procedure. Throughout the numerical examination introduced in Section 3, it was evident that the ridge-estimate produces more stable results than the usual residual-based estimate. In order to quantify the stability of ridge estimate, Table 3 shows the mean squared error (MSE) estimated by RSS- and Ridge-estimate in Cases-1 and 2 (Bertalanffy and Gompertz). Here, the MSE is defined by the expected value of the difference between observation and prediction as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} E \left[ (f(t_i) - f(t_i | \hat{\theta}))^2 \right].$$

This measures the average of the squares of the errors that is, the average squared difference between the estimated values and the actual value. The MSE is a risk function, corresponding to the expected value of the squared error loss. The MSE is a measure of the quality of an estimator. The closer
the value is to zero, the better. In both cases, the ridge-estimate results in a smaller MSE value, which suggests that the ridge-estimate is better than the RSS-estimate.

Table 3. Mean squared error.

|                | RSS  | Ridge |
|----------------|------|-------|
| Bertalanffy    | 0.977| 0.901 |
| Gompertz       | 1.190| 1.147 |

Note) Mean squared error $\frac{1}{n} \sum_{i=1}^{n} E [(f(t_i) - f(t_i | \hat{\theta}))^2]$ is obtained by RSS- and Ridge-estimate under the two situations (Bertalanffy and Gompertz).

From the aforementioned section we learnt that in the estimation of the growth function, there is a risk that improper initial value setting will result in improper result. In other words, the estimation of the growth function has instability that depends on the initial value. Until now the setting of the initial value has been left to trial and error, and it depends very much on the experience and intuition of the researcher. Finding an appropriate initial value is a difficult task and extremely difficult especially when considering a function with a large number of parameters. One way to solve this difficulty is outlined in this paper. In this paper, we propose a new application using the Ridge estimation approach. The Ridge estimation approach did not necessarily converge to a proper estimate for an initial value setting, but at least more stable convergence was observed than the RSS-estimate. Although the ridge-estimate has a bias, we know from equation [7] that the ridge-estimate is close to the RSS-estimate if an estimated $\lambda$ is not large. It can be seen from Figure 6 that estimated ridge parameters were very small, so we can say that the ridge-estimates are comparable to the RSS-estimates. Although ridge estimate is not universal, it seems to have made it possible to extend the range of initial value setting that converges to a proper estimate.

![Figure 6. Ridge parameter estimation.](image)

The vertical axis denotes the estimated value of Ridge parameter $\lambda$. 

10
Acknowledgement

This work was supported by Grant-in-Aid for Scientific Research (A)17H00806.

References

Bertalanffy, L.V. (1938) A quantitative theory of organic growth (inquiries on growth laws II), Hum. Biol. 10: 181-213.

Craven, P., Wahba, G. (1979) Smoothing noisy data with spline functions, Numer. Math. 31: 377-403.

Fukui, K., Yamanura, M., Yanagihara, H. (2015) Comparison with RSS-based model selection criteria for selecting growth functions, FORMATH 14: 27-39.

Gompertz, B. (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, Philos. Trans. R. Soc. B. 182: 513-85.

Hoerl, A.E., Kennard, R.W. (1970) Ridge regression: biased estimation for nonorthogonal problems, Technometrics 12: 55-67.

Kamo, K., Yoshimoto, A. (2013) Comparative analysis of growth functions based on Mallows’ Cp type criterion, FORMATH 12: 133-147.

Nelder, J.A., Mead, R. (1965) A simplex algorithm for function minimization, Comput. J. 7: 308-313.

Ye, J. (1998) On measuring and correcting the effects of data mining and model selection, J. Am. Stat. Assoc. 93: 120-131.

Zeide, B. (1993) Analysis of growth equations, For. Sci. 39: 594-616.
Appendix (R-script)

We used the statistical software “R” for the numerical examination. Since the growth functions are nonlinear functions, nonlinear regression approach is often used for growth function fitting. In “R”, the function “nls()” is prepared for nonlinear regression analysis. This function needs the initial value of the parameters for iterative estimation. The setting for initial value is so delicate that in some situations it is difficult to converge to a finite estimator. On the other hand, the optim function is more robust than nls function, then we show the script using optim function for future reference. Here we show only for Bertalanffy function due to space limitation. At first, we set the functions which give the growth amount for Bertalanffy function (Ber.func), and give the estimate of parameters in Bertalanffy function (Ber.est.func) as

Ber.func <- function(a,b,t){a*(1-exp(-b*t))^3}
Ber.est.func <- function(IC,t,y)
  {RSS.func <- function(b)
     {temp <- y-Ber.func(b[1], b[2], t); sum(temp ^2)}
     optim(IC, RSS.func)$par }

In the function Ber.func, arguments “a” and “b” denote the parameter $\alpha$ and $\beta$, and “t” denote the independent variable. In the function Ber.est.func, “IC”, “t” and “y” denotes initial value of parameter, observed time and growth amount. When we use the proper initial value $(\alpha, \beta) = (90, 0.3)$, then the parameters are estimated as $(\hat{\alpha}, \hat{\beta}) = (102.58, 0.393)$. On the other hand, when we use improper initial value, for example, $(\alpha, \beta) = (20, 5)$, then the estimate procedure converges to improper estimate as $(\hat{\alpha}, \hat{\beta}) = (67.152, 12.657)$. It is troublesome here that no warning or error message is displayed for conversing to an improper estimate.