



Uncertain Gompertz regression model with imprecise observations

Zeyu Hu¹ · Jinwu Gao²

Published online: 2 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Regression is widely applied in many fields. Regardless of the types of regression, we often assume that the observations are precise. However, in real-life circumstances, this assumption can only be met sometimes, which means the traditional regression methods can result in significant imprecise or biased predictions. Consequently, uncertain regression models might provide more accurate and meaningful results under these circumstances. In this article, we provide the residual analysis of uncertain Gompertz regression model, as well as the corresponding forecast value and confidence interval. Finally, we give a numerical example of uncertain Gompertz regression model.

Keywords Regression analysis · Uncertainty theory · Uncertain variable · Residual · Confidence interval

1 Introduction

Regression is an important method in inference and prediction problems. Since the invention of method of least square by Legendre (1805), regression analysis is widely used to investigate relationships between independent variables and dependent variable. Statisticians also use method of maximum likelihood popularized by Wilks (1938) together with the method of least square. With the development of regression analysis, t test (Student 1908) and F test (Fisher 1925) were investigated to test the hypothesis of regression. Proposed by Neyman and Pearson (1933), likelihood ratio test also shows promising results in most cases. Furthermore, starting with the linear regression (Galton 1885), mathematicians also invented numerous nonlinear regression models to fit different scenarios. Well-known examples of nonlinear regression models include logistic regression model, polynomial regression model, etc. This paper will analyze another

important nonlinear regression model: Gompertz regression model.

Gompertz regression model was invented by Benjamin Gompertz in 1825 to elaborate his law of human mortality (Gompertz 1825). Gompertz model has a “S-shape” feature that enables it to be widely applied in many biological systems to determine the growth of population, because it can accurately describe the process of a relatively low growth rate in the early stage and late stage as well as a rapid growth in the intermediate stage (Nguimkeu 2014). Although Gompertz model shares similar properties with logistic model, it is symmetric while Logistic model is not, so one model might outperform the other in some circumstances (Nguimkeu 2014). For instances, Laird used Gompertz model to describe the growth of tumor (Laird 1964); Zwietering and other biologists investigated the bacterial growth curve based on the Gompertz model (Zwietering et al. 1990). It is then important to further analyze the property of Gompertz model to better describe related biological feature.

Most of the time, statisticians consider observations as precise data and ignore the fact that data points might be acquired in non-random way, which might lead to inaccurate inference or predictions. For instance, in a large biological system, it is always impossible to precisely measure the whole population. Scientists usually use the capture–recapture method to estimate the whole population size, which might result in biased estimation. In order to handle scenarios like this, Tanaka et al. (1982) first proposed a fuzzy linear regression model in 1982 and was later modified and

Communicated by Y. Ni.

✉ Jinwu Gao
jgao@ruc.edu.cn
Zeyu Hu
huzeyu@ucla.edu

¹ Department of Mathematics, University of California, Los Angeles 90024, USA

² Uncertain Systems Lab, School of Mathematics, Renmin University of China, Beijing 100872, China

improved by Corral and Gil (1984). In addition to the method proposed by Tanaka et al., uncertainty theory is shown to provide more accurate and reliable results at some cases (Liu 2012). Uncertainty theory was proposed to investigate the relationship between independent variables and dependent variable with uncertain observed data.

In uncertainty theory, the unknown parameters in models can be estimated by the principle of least squares (Yao and Liu 2018). Yang and Liu (2017) proposed uncertain time series analysis and estimated the unknown parameters using the principle of least squares. Besides, Lio and Liu have already deduced the method to find unknown parameters, forecast value, and confidence interval in several uncertain regression models (Lio and Liu 2018). In this article, we follow their process to analyze the properties of uncertain Gompertz regression model. In Sect. 2, we will introduce some preliminary knowledge of uncertainty theory. In Sects. 3 and 4, we will provide the method to find unknown parameters in the Gompertz regression model and its residual analysis. In Sect. 5, the confidence interval of the uncertain Gompertz regression models is provided. In Sect. 6, we will provide a numerical example to illustrate the application of the uncertain Gompertz regression model. Finally, some conclusions are made in Sect. 7.

2 Preliminaries

According to Liu, many surveys showed that subjective uncertainty cannot be modeled by fuzziness and hence cannot be processed by possibility theory (Liu 2009). In order to deal with subjective uncertainty, Liu proposed uncertainty theory to better capture the properties of imprecise observations that rely on degrees of belief (Liu 2007). In this section, we will provide some concepts and theorems of uncertainty theory which will be useful in the following analysis of Gompertz regression model.

Definition 1 (Liu 2007) Let \mathcal{L} be a σ -algebra on a nonempty set Γ . Then (Γ, \mathcal{L}) is a measurable space, and each element Γ in \mathcal{L} is called an event. An uncertain measure is a set function $\mathcal{M} : \mathcal{L} \rightarrow [0, 1]$ if it satisfies the following axioms:

Axiom 1 (Normality Axiom) $\mathcal{M}\{\Gamma\} = 1$ for the universal set Γ .

Axiom 2 (Duality Axiom) $\mathcal{M}\{A\} + \mathcal{M}\{A^c\} = 1$ for any event A .

Axiom 3 (Subadditivity Axiom) For every countable sequence of events A_1, A_2, \dots , we have

$$\mathcal{M}\left\{\bigcup_{i=1}^{\infty} A_i\right\} \leq \sum_{i=1}^{\infty} \mathcal{M}\{A_i\}.$$

Let Γ be a nonempty set, let \mathcal{L} be a σ -algebra over Γ , and let \mathcal{M} be an uncertain measure. Then the triplet $(\Gamma, \mathcal{L}, \mathcal{M})$ is called an uncertainty space. Furthermore, the product uncertain measure \mathcal{M} satisfies the following axiom:

Axiom 4 (Product Axiom) (Liu 2009) Let $(\Gamma_k, \mathcal{L}_k, \mathcal{M}_k)$ be uncertainty spaces for $k = 1, 2, \dots$. The product uncertain measure \mathcal{M} is an uncertain measure satisfying

$$\mathcal{M}\left\{\prod_{k=1}^{\infty} A_k\right\} = \bigwedge_{k=1}^{\infty} \mathcal{M}_k\{A_k\}$$

where A_k are arbitrarily chosen events from \mathcal{L}_k for $k = 1, 2, \dots$, respectively.

Definition 2 (Liu 2007) An uncertain variable ξ is a measurable function from an uncertainty space $(\Gamma, \mathcal{L}, \mathcal{M})$ to the set of real numbers, i.e., for any Borel set B of real numbers. The set

$$\{\xi \in B\} = \{\gamma \in \Gamma \mid \xi(\gamma) \in B\}$$

is an event.

We define the uncertainty distribution Φ of an uncertain variable ξ as $\Phi(x) = \mathcal{M}\{\xi \leq x\}$ for any real number x . An uncertainty distribution $\Phi(x)$ is called regular if it is a continuous and strictly increasing function with respect to x where $0 < \Phi(x) < 1$, and

$$\lim_{x \rightarrow -\infty} \Phi(x) = 0, \quad \lim_{x \rightarrow \infty} \Phi(x) = 1.$$

Let ξ be an uncertain variable with regular uncertainty distribution $\Phi(x)$. Then the inverse function $\Phi^{-1}(\alpha)$ is called the inverse uncertainty distribution of ξ (Liu 2010).

Now we introduce some special uncertainty distributions.

An uncertain variable ξ is called linear, denoted by $\mathcal{L}(a, b)$, if it has an uncertainty distribution

$$\Phi(x) = \begin{cases} 0, & \text{if } x \leq a \\ (x - a)/(b - a), & \text{if } a < x \leq b \\ 1, & \text{if } x > b \end{cases}$$

where a and b are real numbers with $a < b$. The inverse uncertainty distribution of $\mathcal{L}(a, b)$ is

$$\Phi^{-1}(\alpha) = (1 - \alpha)a + \alpha b.$$

An uncertain variable ξ is called zigzag, denoted by $\mathcal{Z}(a, b, c)$, if it has an uncertainty distribution

$$\Phi(x) = \begin{cases} 0, & \text{if } x \leq a \\ (x - a)/[2(b - a)], & \text{if } a < x \leq b \\ (x + c - 2b)/[2(c - b)], & \text{if } b < x \leq c \\ 1, & \text{if } x > c \end{cases}$$

where a, b and c are real numbers with $a < b < c$. Then the inverse uncertainty distribution of $\mathcal{Z}(a, b, c)$ is

$$\Phi^{-1}(\alpha) = \begin{cases} (1 - 2\alpha)a + 2\alpha b, & \text{if } \alpha < 0.5 \\ (2 - 2\alpha)b + (2\alpha - 1)c, & \text{if } \alpha \geq 0.5. \end{cases}$$

An uncertain variable ξ is called normal, denoted by $\mathcal{N}(e, \sigma)$, if it has an uncertainty distribution

$$\Phi(x) = \left(1 + \exp\left(\frac{\pi(e - x)}{\sqrt{3}\sigma}\right) \right)^{-1}, x \in \mathfrak{R}$$

where e and σ are real numbers with $\sigma > 0$. Then the inverse uncertainty distribution of $\mathcal{N}(e, \sigma)$ is

$$\Phi^{-1}(\alpha) = e + \frac{\sigma\sqrt{3}}{\pi} \ln \frac{\alpha}{1 - \alpha}.$$

Definition 3 (Liu 2009) Uncertain variables $\xi_1, \xi_2, \dots, \xi_n$ are independent if

$$\mathcal{M} \left\{ \bigcap_{i=1}^n (\xi_i \in B_i) \right\} = \bigwedge_{i=1}^n \mathcal{M} \{ \xi_i \in B_i \}$$

for any Borel sets B_1, B_2, \dots, B_n of real numbers.

Let $\xi_1, \xi_2, \dots, \xi_n$ be independent uncertain variables with regular uncertainty distributions $\Phi_1, \Phi_2, \dots, \Phi_n$, respectively. If $f(x_1, x_2, \dots, x_n)$ is a strictly monotonous function, then the inverse uncertainty distribution of the uncertain variable $f(\xi_1, \xi_2, \dots, \xi_n)$ can be calculated by the following theorem (Liu 2010).

Theorem 1 (Liu 2010) *Let $\xi_1, \xi_2, \dots, \xi_n$ be independent uncertain variables with regular uncertainty distributions $\Phi_1, \Phi_2, \dots, \Phi_n$, respectively. If a function $f(\xi_1, \xi_2, \dots, \xi_n)$ is strictly increasing with respect to $\xi_1, \xi_2, \dots, \xi_m$ and strictly decreasing with respect to $\xi_{m+1}, \xi_{m+2}, \dots, \xi_n$, then the uncertain variable $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$ has an inverse uncertainty distribution*

$$\begin{aligned} \Psi^{-1}(\alpha) &= f(\Phi_1^{-1}(\alpha), \dots, \Phi_m^{-1}(\alpha), \Phi_{m+1}^{-1}(1 - \alpha), \dots, \Phi_n^{-1}(1 - \alpha)). \end{aligned}$$

Definition 4 (Liu (2007)) Let ξ be an uncertain variable. The expected value of ξ is defined as

$$E[\xi] = \int_0^{+\infty} \mathcal{M}\{\xi \geq x\} dx - \int_{-\infty}^0 \mathcal{M}\{\xi \leq x\} dx$$

provided that at least one of the two integrals is finite.

Definition 5 (Liu 2007) Let ξ be an uncertain variable with finite expected value e . The variance of ξ is

$$V[\xi] = E[(\xi - e)^2].$$

Theorem 2 (Liu 2010) *If ξ is an uncertain variable with regular uncertainty distribution Φ , then we have*

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha) d\alpha, \tag{1}$$

$$E[\xi^2] = \int_0^1 (\Phi^{-1}(\alpha))^2 d\alpha, \tag{2}$$

$$V[\xi] = \int_0^1 (\Phi^{-1}(\alpha) - e)^2 d\alpha. \tag{3}$$

3 Uncertain Gompertz regression model

Let (x_1, x_2, \dots, x_p) be a vector of independent variables, and let y be the dependent variable. If the relationship between (x_1, x_2, \dots, x_p) and y can be expressed by a function, f , then the model is generally expressed as

$$y = f(x_1, x_2, \dots, x_p | \beta) + \epsilon \tag{4}$$

where β is a vector of unknown parameters, and ϵ is a disturbance term. If we have a set of imprecisely observed data,

$$(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i), \quad i = 1, 2, \dots, n \tag{5}$$

where $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i$ are uncertain variables with uncertainty distributions $\Phi_{i1}, \Phi_{i2}, \dots, \Phi_{ip}, \Psi_i, i = 1, 2, \dots, n$, respectively. In order to perform prediction or inference, it is necessary to find the vector of unknown parameters, β . Since it is impossible to find the precise value of the parameters, we are interested in obtaining an estimation of β , denoted by β^* , based on the imprecisely observed data.

As proposed by Yao and Liu (2018), the least squares estimate of β in the regression model (4) can be obtained by solving the minimization problem below:

$$\min_{\beta} \sum_{i=1}^n E[(\tilde{y}_i - f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip} | \beta))^2]. \tag{6}$$

If we denote the optimal solution of the minimization problem (6) as β^* , then the fitted regression model is given as

$$y = f(x_1, x_2, \dots, x_p | \beta^*). \tag{7}$$

Definition 6 The Gompertz regression model is defined as:

$$y = \beta_0 \exp(-\beta_1 \exp(-\beta_2 x)) + \epsilon, \quad \beta_0 > 0, \beta_1 > 0, \beta_2 > 0 \tag{8}$$

where $\beta_0, \beta_1, \beta_2$ are parameters.

Gompertz regression model is widely used in the biological system to determine the growth of population of a certain

specie. It can accurately capture the characteristics of the population growth in nature.

Theorem 3 Let $(\tilde{x}_i, \tilde{y}_i), i = 1, 2, \dots, n$, be a set of imprecisely observed data, where \tilde{x}_i, \tilde{y}_i are independent uncertain variables with regular uncertainty distributions $\Phi_i, \Psi_i, i = 1, 2, \dots, n$, respectively. Then the least squares estimate of β_0, β_1 and β_2 in the Gompertz regression model is the optimal solution of the following minimization problem:

$$\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^n \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0 \exp(-\beta_1 \exp(-\beta_2 \Phi_i^{-1}(1 - \alpha))))^2 d\alpha. \tag{9}$$

Proof By Eq. (6), the least squares estimate of β_0, β_1 and β_2 in the Gompertz regression model can be obtained by solving the minimization problem

$$\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^n E \left[(\tilde{y}_i - \beta_0 \exp(-\beta_1 \exp(\beta_2 \tilde{x}_i)))^2 \right]. \tag{10}$$

Since the function

$$\tilde{y}_i - \beta_0 \exp(-\beta_1 \exp(\beta_2 \tilde{x}_i)) \tag{11}$$

is strictly increasing with respect to \tilde{y}_i and strictly decreasing with respect to \tilde{x}_i for each i . According to Theorem 1, the inverse uncertainty distribution of function (11) is

$$F_i^{-1}(\alpha) = \Psi_i^{-1}(\alpha) - \beta_0 \exp(-\beta_1 \exp(-\beta_2 \Phi_i^{-1}(1 - \alpha))).$$

Then according to Theorem 2, we have

$$E \left[(\tilde{y}_i - \beta_0 \exp(-\beta_1 \exp(\beta_2 \tilde{x}_i)))^2 \right] = \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0 \exp(-\beta_1 \exp(-\beta_2 \Phi_i^{-1}(1 - \alpha))))^2 d\alpha.$$

Thus the minimization problem (10) is equivalent to

$$\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^n \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0 \exp(-\beta_1 \exp(-\beta_2 \Phi_i^{-1}(1 - \alpha))))^2 d\alpha.$$

The theorem is verified.

4 Residual analysis

In the regression model (4), there is always a disturbance term, ϵ , because it is usually impossible for our model to perfectly fit each observation point. The disturbance term indicates the distance that dependent variable y may deviate

from the regression. Because of the nature of disturbance, each observation has a different value of ϵ , and hence we are only interested in finding an estimation of ϵ for the given set of imprecisely observed data

$$(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i), \quad i = 1, 2, \dots, n. \tag{12}$$

For each i , the difference between the observed value \tilde{y}_i and $f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip} | \beta^*)$ represents the distance of observation and our regression model and hence is the disturbance term ϵ . Thus, we propose a definition of ϵ as follows:

Definition 7 (Lio and Liu 2018) Let $(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip}, \tilde{y}_i), i = 1, 2, \dots, n$ be a set of imprecisely observed data, and let the fitted regression model be

$$y = f(x_{i1}, x_{i2}, \dots, x_{ip} | \beta^*). \tag{13}$$

Then for each $i (i = 1, 2, \dots, n)$, the term

$$\hat{\epsilon}_i = \tilde{y}_i - f(\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ip} | \beta^*) \tag{14}$$

is called the i th residual.

In uncertainty theory, because the data are imprecisely observed, the i th residual ϵ is also assumed to be an uncertain variable. Because each observation has a different disturbance term, we use the average of the expected values of residuals:

$$\hat{\epsilon} = \frac{1}{n} \sum_{i=1}^n E[\hat{\epsilon}_i] \tag{15}$$

to estimate the expected value of ϵ , and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n E[(\hat{\epsilon}_i - \hat{\epsilon})^2] \tag{16}$$

to estimate the variance, where $\hat{\epsilon}_i$ are the i th residuals, $i = 1, 2, \dots, n$, respectively.

Theorem 4 Let $(\tilde{x}_i, \tilde{y}_i), i = 1, 2, \dots, n$ be a set of imprecisely observed data, where \tilde{x}_i, \tilde{y}_i are independent uncertain variables with regular uncertainty distributions $\Phi_i, \Psi_i, i = 1, 2, \dots, n$, respectively, and the fitted Gompertz regression model is denoted as

$$y = \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* x)), \quad \beta_0^* > 0, \beta_1^* > 0, \beta_2^* > 0 \tag{17}$$

Then the estimated expected value of ϵ is

$$\hat{\epsilon} = \frac{1}{n} \sum_{i=1}^n \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \Phi_i^{-1}(1 - \alpha)))) d\alpha \tag{18}$$

and the estimated variance is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \Phi_i^{-1}(1 - \alpha))))^2 d\alpha. \tag{19}$$

Proof Since the function

$$\tilde{y}_i - \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \tilde{x}_i)), \quad \beta_0^* > 0, \beta_1^* > 0, \beta_2^* > 0 \quad (*)$$

is strictly increasing with respect to \tilde{y}_i and strictly decreasing with respect to \tilde{x}_i for each i . According to Theorem 1, the inverse uncertainty distribution of the function (*) is

$$F_i^{-1}(\alpha) = \Psi_i^{-1}(\alpha) - \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \Phi_i^{-1}(1 - \alpha))).$$

By the definition of estimate of residuals and variance of uncertain variables and Theorem 2, the theorem is proved trivially.

5 Forecast value and confidence interval

The main purpose of regression is to make prediction given a new observation of independent variables. Suppose \tilde{x}_p is a new observation of independent variables, and \tilde{x}_p is a uncertain dependent variable with regular uncertainty distribution Φ_p . Since the model has been constructed by using previous imprecisely observed data $(\tilde{x}_i, \tilde{y}_i), i = 1, 2, \dots, n$, we can then make predictions and derive the forecast value for the new dependent variable \tilde{y}_p .

For example, scientists want to investigate the population of wild lions in a new plain. From previous imprecisely observed data set, the scientists choose the area of the plain, as independent variables, and the population of wild lions as dependent variable. Then, an uncertain Gompertz regression model can be constructed by these observations. By obtaining the data from the new plain, we can then use the model to predict the population of wild lions in the new plain.

Mathematically, we can denote the fitted uncertain Gompertz regression model constructed by using previous observations as:

$$y = \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* x)), \quad \beta_0^* > 0, \beta_1^* > 0, \beta_2^* > 0 \tag{20}$$

and the residual ϵ has estimated expected value \hat{e} and variance $\hat{\sigma}^2$, and is independent of \tilde{x}_p . Then the forecast uncertain variable of y with respect to \tilde{x}_p is determined by

$$\hat{y} = \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \tilde{x})) + \epsilon. \tag{21}$$

The intuitive attempt would be finding a point estimate for the new y . The forecast value of y is defined as

$$\mu = \beta_0^* E[\exp(-\beta_1^* \exp(-\beta_2^* \tilde{x}))] + \hat{e}, \tag{22}$$

or we can use the integral to express μ :

$$\mu = \int_0^1 (\beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \Phi^{-1}(1 - \alpha)))) d\alpha + \hat{e}. \tag{23}$$

The forecast value of y represents the expected value of the forecast uncertain variable \hat{y} . Assume that the disturbance term ϵ has a normal uncertainty distribution $N(\hat{e}, \hat{\sigma})$ with inverse uncertainty distribution $\Phi^{-1}(\alpha)$, i.e.,

$$\Phi^{-1}(\alpha) = \hat{e} + \frac{\hat{\sigma}\sqrt{3}}{\pi} \ln \frac{\alpha}{1 - \alpha}. \tag{24}$$

We can then derive the inverse uncertainty distribution of \hat{y} :

$$\hat{\Psi}^{-1}(\alpha) = \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \Phi^{-1}(1 - \alpha))) + \Phi^{-1}(\alpha) \tag{25}$$

Trivially, the uncertainty distribution, $\hat{\Psi}$, of \hat{y} can be obtained by $\hat{\Psi}^{-1}$.

The forecast value, μ , is a point estimation of y . However, we are more interested to find a range of values that y might fall within. Hence, the confidence interval of uncertain Gompertz regression model is proposed. Although we might loss some precision or accuracy of prediction, but we are more confident about the correctness of our prediction. Let α denote the confidence level, which indicates the belief degree of corresponding confidence intervals that contain the true value of y . We can derive the confidence interval by finding the minimum value b such that

$$\hat{\Psi}(\mu + b) - \hat{\Psi}(\mu - b) \geq \alpha. \tag{26}$$

Since

$$\mathcal{M}\{\mu - b \leq \hat{y} \leq \mu + b\} \geq \hat{\Psi}(\mu + b) - \hat{\Psi}(\mu - b), \tag{27}$$

it follows that $\mathcal{M}\{\mu - b \leq \hat{y} \leq \mu + b\} \geq \alpha$. Thus the α confidence interval of y can be represented as $[\mu - b, \mu + b]$, which can be simplified as

$$\mu \pm b. \tag{28}$$

6 Numerical example

In this section, we will use an example to illustrate the application of uncertain Gompertz regression model to make

Table 1 Imprecisely observed data where $\mathcal{L}(a, b)$ represents linear uncertain variable

| i | 1 | 2 | 3 | 4 | 5 | 6 |
|---------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| \tilde{x}_i | $\mathcal{L}(4.8, 5.2)$ | $\mathcal{L}(5.7, 6.2)$ | $\mathcal{L}(6.6, 7.0)$ | $\mathcal{L}(7.7, 8.1)$ | $\mathcal{L}(8.9, 9.5)$ | $\mathcal{L}(9.5, 10.5)$ |
| \tilde{y}_i | $\mathcal{L}(4.3, 5.0)$ | $\mathcal{L}(4.9, 5.3)$ | $\mathcal{L}(9.9, 11.2)$ | $\mathcal{L}(14.5, 16.5)$ | $\mathcal{L}(19.0, 21.0)$ | $\mathcal{L}(28.5, 31.5)$ |
| i | 7 | 8 | 9 | 10 | 11 | 12 |
| \tilde{x}_i | $\mathcal{L}(11.0, 11.8)$ | $\mathcal{L}(11.5, 12.4)$ | $\mathcal{L}(14.6, 15.7)$ | $\mathcal{L}(17.5, 18.5)$ | $\mathcal{L}(20.0, 22.0)$ | $\mathcal{L}(23.0, 25.0)$ |
| \tilde{y}_i | $\mathcal{L}(34.0, 38.0)$ | $\mathcal{L}(40.0, 45.0)$ | $\mathcal{L}(62.0, 66.0)$ | $\mathcal{L}(74.0, 78.0)$ | $\mathcal{L}(79.0, 82.0)$ | $\mathcal{L}(83.5, 84.5)$ |
| i | 13 | 14 | 15 | 16 | 17 | 18 |
| \tilde{x}_i | $\mathcal{L}(25.5, 29.0)$ | $\mathcal{L}(29.0, 31.5)$ | $\mathcal{L}(31.0, 34.0)$ | $\mathcal{L}(35.0, 37.0)$ | $\mathcal{L}(39.0, 42.0)$ | $\mathcal{L}(41.0, 45.0)$ |
| \tilde{y}_i | $\mathcal{L}(83.8, 84.6)$ | $\mathcal{L}(83.8, 84.9)$ | $\mathcal{L}(84.0, 85.1)$ | $\mathcal{L}(84.8, 85.9)$ | $\mathcal{L}(84.9, 85.8)$ | $\mathcal{L}(85.5, 86.5)$ |

prediction for a new independent data with imprecise observation. Furthermore, we will provide a confidence interval to make stronger and more reliable prediction when compared to the point estimation.

The data are originally about spores germination, where the independent variable is the day from the beginning of experiment and the dependent variable is the germination rate. The data can only be imprecisely observed because of the condition of experiment, and hence an uncertain Gompertz regression model should be applied under this circumstance.

We use $(\tilde{x}_i, \tilde{y}_i), i = 1, 2, \dots, 18$ to denote the set of imprecisely observed data of spores germination, where \tilde{x}_i, \tilde{y}_i are independent uncertain variables with linear uncertainty distributions, Φ_i, Ψ_i , respectively. \tilde{x}_i represents the days and \tilde{y}_i represents germination rate. The specific data are provided in Table 1.

We may use uncertain Gompertz regression model to forecast the germination rate for any given day. The uncertain Gompertz regression model is given as

$$y = \beta_0 \exp(-\beta_1 \exp(-\beta_2 x)) + \epsilon. \tag{29}$$

In order to obtain the least squares estimate of β_0, β_1 and β_2 in the Gompertz regression model, we need to solve the minimization problem (6), i.e.,

$$\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^n E \left[(\tilde{y}_i - \beta_0 \exp(-\beta_1 \exp(\beta_2 \tilde{x}_i)))^2 \right], \tag{30}$$

or equivalently,

$$\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^n \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0 \exp(-\beta_1 \exp(-\beta_2 \Phi_i^{-1}(1 - \alpha))))^2 d\alpha. \tag{31}$$

From Theorem 3, we can obtain the least squares estimate

$$(\beta_0^*, \beta_1^*, \beta_2^*) = (86.0493, 11.9398, 0.2391). \tag{32}$$

The fitted Gompertz regression model is then

$$y = 86.0493 \exp(-11.9398 \exp(-0.2391x)). \tag{33}$$

From Eq. (18), i.e.,

$$\hat{\epsilon} = \frac{1}{18} \sum_{i=1}^{18} \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \Phi_i^{-1}(1 - \alpha)))) d\alpha, \tag{34}$$

we can obtain the estimated expected value of the disturbance term ϵ , and from Eq. (19), i.e.,

$$\hat{\sigma}^2 = \frac{1}{18} \sum_{i=1}^{18} \int_0^1 (\Psi_i^{-1}(\alpha) - \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* \Phi_i^{-1}(1 - \alpha))) - \hat{\epsilon})^2 d\alpha, \tag{35}$$

we can obtain the variance of the disturbance term ϵ . The estimated expected value and variance of ϵ are

$$\hat{\epsilon} = 0.0000, \quad \hat{\sigma}^2 = 3.5725, \tag{36}$$

respectively. Now suppose

$$\tilde{x}_p \sim \mathcal{L}(24.0, 26.0) \tag{37}$$

is a new imprecisely observed data of a day. Assume the new day \tilde{x}_p is independent of ϵ , we can then obtain the forecast uncertain variable of the dependent variable y

$$\hat{y} = 86.0493 \exp(-11.9398 \exp(-0.2391\tilde{x}_p)) + \epsilon, \tag{38}$$

and the forecast value of y is 83.4837 which can be obtained from Eq. (22), i.e.,

$$\mu = \beta_0^* E[\exp(-\beta_1^* \exp(-\beta_2^* \tilde{x}_p))] + \hat{e}. \quad (39)$$

For the confidence level $\alpha = 95\%$, if we suppose that the disturbance term ϵ is a normal uncertain variable, then

$$b = 5.7271 \quad (40)$$

is the minimum value for Eq. (26), i.e.,

$$\hat{\Psi}(\mu + b) - \hat{\Psi}(\mu - b) \geq 95\%. \quad (41)$$

$\hat{\Psi}$ is the uncertainty distribution of \hat{y} and can be calculated by

$$\hat{\Psi}^{-1}(\alpha) = \beta_0^* \exp(-\beta_1^* \exp(-\beta_2^* (24(1 - \alpha) + 26\alpha))) \quad (42)$$

where $\Phi^{-1}(\alpha)$ is the inverse uncertainty distribution of normal uncertain variable $\mathcal{N}(\hat{e}, \hat{\sigma})$. The 95% confidence interval of dependent variable y is then

$$83.4837 \pm 5.7271. \quad (43)$$

7 Conclusion

This article introduced the uncertain Gompertz regression model with the method to obtain estimate least square betas, model's forecast value and confidence interval, and residual analysis of the model. We also provided a numerical example of spores germination to explain the application of the uncertain Gompertz regression model in the real-life problem.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 61374082).

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

References

- Corral N, Gil MA (1984) The minimum inaccuracy fuzzy estimation: an extension of the maximum likelihood principle. *Stochastica* 8:63–81
- Fisher RA (1925) *Statistical methods for research workers*. Oliver and Boyd, Edinburgh
- Galton F (1885) Regression towards mediocrity in hereditary stature. *J Anthropol Inst* 15:246–263
- 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 B* 115:513585
- Laird AK (1964) Dynamics of tumor growth. *Br J Cancer* 18:490502
- Legendre AM (1805) *New methods for the determination of the orbits of comets*. Firmin Didot, Paris
- Lio W, Liu B (2018) Residual and confidence interval for uncertain regression model with imprecise observations. *J Intell Fuzzy Syst* 35:2573–83
- Liu B (2007) *Uncertainty theory*, 2nd edn. Springer, Berlin
- Liu B (2009) Some research problems in uncertainty theory. *J Uncertain Syst* 3:3–10
- Liu B (2010) *Uncertainty theory: a branch of mathematics for modeling human uncertainty*. Springer, Berlin
- Liu B (2012) Why is there a need for uncertainty theory. *J Uncertain Syst* 6:3–10
- Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc A Math Phys Eng Sci* 231:289–337
- Nguimkeu P (2014) A simple selection test between the Gompertz and logistic growth models. *Technol Forecast Soc Change* 88:98–105
- Student (1908) The probable error of a mean. *Biometrika* 6:1–25
- Tanaka H, Uejima S, Asai K (1982) Linear regression analysis with fuzzy model. *IEEE Trans Syst Man Cybern* 12:903–907
- Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat* 9:60–62
- Yang X, Liu B (2017) Uncertain time series analysis with imprecise observations. *Fuzzy Optim Decis Mak* (in press)
- Yao K, Liu B (2018) Uncertain regression analysis: an approach for imprecise observations. *Soft Comput* 22:557982
- Zwietering MH, Jongenburger I, Rombouts FM, VanT Rie TK (1990) Modeling of the bacterial growth curve. *Appl Environ Microbiol* 56:197581

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.