American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN: 2320-0936
Volume-14, Issue-10, pp-175-180
www.ajer.org
Research Paper Open Access

Least Squares Estimation of Uncertain Growth Model

Han Wang
College of Mathematics and Statistics, Shanxi Datong University, Datong 037009, China.
Corresponding Author: Han Wang.

ABSTRACT: *Uncertain regression analysis holds a significant position in the field of statistics, enabling in-depth analysis of the intrinsic relationships between variables. In the study of uncertain growth models, how to estimate the unknown parameters and the uncertain disturbance term in the models is a key issue. This paper constructs a statistical invariant by combining the uncertain distributions of observed values and disturbance terms, which is a normal uncertain variable. The least squares principle is applied to estimate the parameters of the statistical invariant to determine the unknown parameters and the uncertain disturbance terms in the uncertain growth model. This improves the traditional least squares method, which only takes into account the reduction the deviation between predicted and observed values, without fully considering the relationship between observed values and disturbance terms under the uncertain framework. In addition, a numerical algorithm was designed to calculate the corresponding estimators, and an uncertain hypothesis test is proposed to determine whether the estimated uncertain growth model is reasonable. Meanwhile, point prediction and interval prediction are also conducted. Finally, an empirical study on the death toll of COVID-19 in China as an example is given to illustrate the effectiveness of the proposed method in solving practical problems.*
KEYWORDS *uncertain growth model, least squares principle, uncertain hypothesis testing, COVID-19*

Date of Submission: 13-10-2025 Date of acceptance: 27-10-2025

I. INTRODUCTION

Regression analysis is one of the core methods in the field of statistics. Although stochastic regression analysis has been developed for a long time, it has always been considered within the framework of probability theory. Its premise is that the estimated probability distribution function needs to be sufficiently close to the true frequency. However, the real world is complex and changeable, and this assumption is difficult to meet in many situations, especially when dealing with unpredictable events, where the limitations of traditional methods become increasingly evident.

To this end, Liu first proposed the uncertainty theory in 2007 and further improve and perfect it in 2009 [1][2]. uncertain regression analysis is a mathematical method that utilizes uncertainty theory to model the inherent relationships between variables. To achieve this, Yao and Liu initiated research on uncertain regression analysis in 2018, with its core innovation lying in redefining the disturbance term in the regression model from a random variable to an uncertain variable, thus adapting to the uncertainty modeling needs within a non-probabilistic framework[3].Since then, uncertain regression models have been continuously developed and refined. For instance, Ye and Liu designed a multivariate uncertain regression model to handle complex situations involving multivariate response variables[4]. Jiang and Ye innovatively constructed an uncertain panel regression model, which spans the two dimensions of time and space and thoroughly reveals the complex interactive relationships hidden in panel data [5]. Fang and Hong propose the uncertain revised regression model [6].

In addition, it is necessary to estimate the unknown parameters and disturbance terms in the uncertain regression model. To estimate uncertainty disturbance terms, Liu proposed the least squares estimation [7]. Lio and Liu initially tried the moment estimation method [8]. Subsequently, Lio and Liu further introduced the uncertain maximum likelihood estimation [9]. To test the goodness of fit estimated in uncertain regression models, Ye and Liu introduced uncertain hypothesis testing [10]. Additionally, uncertainty regression analysis has been applied to describe epidemic transmission. Liu utilized an uncertainty growth model to describe the cumulative number of COVID-19 infections in China [11].

Currently, scholars have conducted research on the least squares estimation of unknown parameters and disturbance terms for uncertain regression models, but they have not simultaneously considered the estimation of both unknown parameters and disturbance terms, nor have they fully considered the relationship between observed values and disturbance terms. Therefore, based on the uncertain distribution of observed data and disturbance terms, this paper constructs corresponding statistical invariant and applies the least squares principle to the parameter estimation process of statistical invariant, in order to compensate for the deficiencies of existing methods.

II. UNCERTAIN GROWTH MODEL

When exploring growth patterns in complex environments, constructing a logistic growth model can capture the trend of gradual acceleration in early growth, acceleration in the middle stage, and deceleration in the late stage. The uncertain logistic growth model is given by:

$$y=\frac{\beta_0}{1+\beta_1\exp(-\beta_2 t)}+\varepsilon \tag{1}$$

where β_0, β_1, and β_2 represent unknown parameters, and ε denotes the uncertain disturbance term. Assuming ε follows a normal uncertain distribution $N(0, \sigma^2)$, i.e.,

$$y-\frac{\beta_0}{1+\beta_1\exp(-\beta_2 t)} \sim N(0,\sigma). \tag{2}$$

Based on the normality assumption of the uncertain disturbance term, equation (2) can be rewritten as:

$$\frac{y-\dfrac{\beta_0}{1+\beta_1\exp(-\beta_2 t)}}{\sigma} \sim N(0,1) \tag{3}$$

Equation (3) fully considers the correlation between unknown parameters β_0, β_1, and β_2 and the uncertain disturbance term ε. Assuming we have a series of observed values (t, y_t), $t=1,2,\cdots,n$, let

$$h_t(\beta,\sigma)=\frac{y_t-\dfrac{\beta_0}{1+\beta_1\exp(-\beta_2 t)}}{\sigma}, \tag{4}$$

where $h_t(\beta,\sigma)$ are real functions of the parameter vectors $\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2)$ and σ, while $h_1(\beta,\sigma)$, $h_2(\beta,\sigma)$, $\cdots$, $h_n(\beta,\sigma)$ can be regarded as the n samples of the standard normal uncertain distribution. That is, $h_t(\beta,\sigma) \sim N(0,1)$, $t=1,\cdots,n$. where the standard normal uncertain distribution $N(0,1)$ is a statistical invariant that we want to construct.

2.1 Parameter Estimation

Based on the principle of least squares, this study investigates the estimation of unknown parameters and disturbance terms in an uncertain logistic growth model. Liu estimates the unknown parameter vector θ by minimizing the sum of squared deviations between the empirical distribution of observations and the population distribution [7], as given in

$$\min_{\theta}\sum_{i=1}^{n}(\Phi_\theta(x_i)-F(x_i))^2, \tag{5}$$

Where

$$F(x)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(x_i\le x)$$

is the empirical distribution function of observations x_i and the indicative function denoted as $\mathrm{I}(x_i\le x)=\begin{cases}1, & x_i\le x\\ 0, & x_i>x\end{cases}$.

Because the expression of the uncertainty distribution of $N(0,1)$ is

$$\Phi(x)=\left(1+\exp\left(\frac{-\pi x}{\sqrt{3}}\right)\right)^{-1}$$

and the empirical distribution function of the sample $h_t(\beta,\sigma)$ is $F(x)=\frac{1}{n}\sum_{t=1}^{n}\mathrm{I}(h_t(\beta,\sigma)\le x)$, so equation (5) leads to the optimization problem

$$\min_{\beta,\sigma>0}\sum_{t=1}^{n}\left(\left(1+\exp\left(\frac{-\pi h_t(\beta,\sigma)}{\sqrt{3}}\right)\right)^{-1}-\frac{1}{n}\sum_{s=1}^{n}\mathrm{I}\left(h_s(\beta,\sigma)\le h_t(\beta,\sigma)\right)\right)^2. \quad (6)$$

Because there is a nonlinear relationship in the objective function, it is difficult to directly find the optimal solution by using the traditional method. In order to solve this problem, numerical algorithm 1 is designed to effectively process and approximate the solution in equation (6).

Algorithm 1 Numerical algorithms for least squares estimation

Step 0: Input observational data (t, y_t), $t=1,\cdots,n$.

Step 1: Determine the feasible regions Θ of unknown parameters vector $\mathbf{\beta}=(\beta_0,\beta_1,\beta_2)$.

Step 2: For each $\mathbf{\beta}\in\Theta$ and $\sigma>0$, compute $h_1(\beta,\sigma), h_2(\beta,\sigma),\cdots,h_n(\beta,\sigma)$ by

$$h_t(\beta,\sigma)=\frac{y_t-\dfrac{\beta_0}{1+\beta_1\exp(-\beta_2 t)}}{\sigma},\quad t=1,\cdots,n.$$

Step 3: Set $t=t+1$ and $E(\beta_0,\beta_1,\beta_2,\sigma)=0$.

Step 4: Set $t=t+1$ and

$$E(\beta_0,\beta_1,\beta_2,\sigma)=E(\beta_0,\beta_1,\beta_2,\sigma)+\left(\left(1+\exp\left(\frac{-\pi h_t(\beta_0,\beta_1,\beta_2,\sigma)}{\sqrt{3}}\right)\right)^{-1}-\frac{1}{n}\sum_{s=1}^{n}\mathrm{I}\left(h_s(\beta_0,\beta_1,\beta_2,\sigma)\le h_t(\beta_0,\beta_1,\beta_2,\sigma)\right)\right)^2.$$

Step 5: If $t\le n$, then go to Step 4.

Step 6: Find $\hat\beta_0,\hat\beta_1,\hat\beta_2$ and $\hat\sigma$ such that $E(\beta_0,\beta_1,\beta_2,\sigma)$ reaches its minimum value.

Step 7: Output $\hat\beta_0,\hat\beta_1,\hat\beta_2$ and $\hat\sigma$.

Thus, the estimated uncertain logistic growth model is obtained as

$$y=\frac{\hat\beta_0}{1+\hat\beta_1\exp(-\hat\beta_2 t)}+N(0,\hat\sigma). \quad (7)$$

2.2 Uncertain Hypothesis Testing

To test the rationality of model (7), we can test whether a set of residual sequences follows a normal uncertainty distribution $N(0,\hat\sigma)$. Let

$$\varepsilon_t=y_t-\frac{\hat\beta_0}{1+\hat\beta_1\exp(-\hat\beta_2 t)}, t=1,\cdots,n,$$

That is, assuming

$$H_0:\sigma=\hat\sigma \quad vs \quad H_1:\sigma\ne\hat\sigma \quad (8)$$

Given a significance level α, we use the rejection region constructed by Ye [10] to test hypothesis (8), and obtain the corresponding rejection region as

$$W=\left\{(z_1,\cdots,z_n):\text{至少}\lfloor n\times\alpha+1\rfloor\text{个}t,\ 1\le t\le n,\text{使得}z_t<\Phi^{-1}\left(\frac{\alpha}{2}\right)\text{或}\ z_t>\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right\}, \quad (9)$$

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

Through this test, if the observed value ε_t falls into the rejection region W, it indicates that the model (7) can not fit the observed values well. Conversely, If it does not fall into the rejection region W, it indicates that the model (7) can fit the observed values well, which means that the model (7) is reasonable.

2.3 Point Prediction and Interval Prediction

Based on model (7), we can derive the point prediction value and prediction interval for the response variable y. According to the research results of Yao [3], the predicted uncertain variable y is defined as

$$\hat{y} = \frac{\hat{\beta}_0}{1+\hat{\beta}_1 \exp(-\hat{\beta}_2 t)} + N(0,\hat{\sigma}), \qquad (10)$$

which follows a normal uncertain distribution, and its expectation is

$$\frac{\hat{\beta}_0}{1+\hat{\beta}_1 \exp(-\hat{\beta}_2 t)}.$$

Thus, we can infer that the uncertainty distribution of the forecast uncertain variable $\hat{y}$ is

$$\hat{\Phi}(x) = \frac{1}{1+\exp(\frac{\pi}{\sqrt{3}\hat{\sigma}}(\hat{\beta}_0/(1+\hat{\beta}_1\exp(-\hat{\beta}_2 t))-x))},$$

and inverse uncertainty distribution is

$$\hat{\Phi}^{-1}(\alpha) = \frac{\hat{\beta}_0}{1+\hat{\beta}_1 \exp(-\hat{\beta}_2 t)} + \frac{\sqrt{3}\hat{\sigma}}{\pi}\ln\frac{\alpha}{1-\alpha} \qquad (11)$$

Then, we can choose the expected value of $\hat{y}$ as its the point prediction value $\hat{\mu}$, i.e.,

$$\hat{\mu} = \int_0^1 \Phi^{-1}(\alpha)d\alpha = \frac{\hat{\beta}_0}{1+\hat{\beta}_1 \exp(-\hat{\beta}_2 t)}. \qquad (12)$$

Given a confidence level α (e.g., 95%), since

$$M\left\{\hat{\Phi}^{-1}\left(\frac{1-\alpha}{2}\right) \le \hat{y} \le \hat{\Phi}^{-1}\left(\frac{1+\alpha}{2}\right)\right\} \ge M\left\{\hat{y} \le \hat{\Phi}^{-1}\left(\frac{1+\alpha}{2}\right)\right\} - M\left\{\hat{y} \le \hat{\Phi}^{-1}\left(\frac{1-\alpha}{2}\right)\right\}$$
$$\le \hat{\Phi}\left(\hat{\Phi}^{-1}\left(\frac{1+\alpha}{2}\right)\right) - \hat{\Phi}\left(\hat{\Phi}^{-1}\left(\frac{1-\alpha}{2}\right)\right) = \alpha$$

The above formula means that

$$\left[\hat{\Phi}^{-1}\left(\frac{1-\alpha}{2}\right), \hat{\Phi}^{-1}\left(\frac{1+\alpha}{2}\right)\right]$$

is a reasonable fluctuation range of the value of $\hat{y}$ at the α confidence level. Therefore, we can choose the above interval, i.e.,

$$\left[\hat{\mu} - \frac{\sqrt{3}\hat{\sigma}}{\pi}\ln\frac{1+\alpha}{1-\alpha}, \hat{\mu} + \frac{\sqrt{3}\hat{\sigma}}{\pi}\ln\frac{1+\alpha}{1-\alpha}\right]$$

as the forecast interval (α confidence interval) of the response variable.

III. Empirical research

COVID-19 has become an epidemic and a public health emergency. As of April 15,2020, China had reported 82,341 confirmed cases and 3,342 deaths. Therefore, analyzing the evolution of the cumulative death toll of COVID-19 is of great significance. The following section uses an uncertain Logistic growth model to model the death toll of COVID-19 in China. The data collected from the National Health Commission's website are shown in Table 1.

Table. 1. The death toll of COVID-19 in China from January 20 to April 15,2020

6	9	17	25	41	56	80	106
32	170	213	259	304	361	425	490
563	636	722	811	908	1016	1113	1367
1380	1523	1665	1770	1868	2004	2118	2236
2345	2442	2592	2663	2715	2744	2788	2835
2870	2912	2943	2981	3012	3042	3070	3097
3119	3136	3158	3169	3176	3189	3199	3213

3226	3237	3245	3248	3255	3261	3270	3277
3281	3287	3292	3295	3300	3304	3305	3312
3318	3322	3326	3329	3331	3331	3333	3335
3336	3339	3339	3341	3341	3342	3342	

According to the observational data of the death toll of COVID-19 in China from January 20 to April 15,2020, the 87 samples of the standard normal uncertain distribution obtained from Equation (4) are $h_1(\beta,\sigma)$, $\cdots, h_{87}(\beta,\sigma)$. Using equation (6), we can obtain the following optimization problem:

$$\min_{\beta,\sigma>0}\sum_{t=1}^{87}\left(\left(1+\exp\left(\frac{-\pi h_t(\beta,\sigma)}{\sqrt{3}}\right)\right)^{-1}-\frac{1}{n}\sum_{s=1}^{87}\mathrm{I}\left(h_s(\beta,\sigma)\le h_t(\beta,\sigma)\right)\right)^2 \qquad (13)$$

By solving the above optimization problem, we can obtain the least squares estimations of β_0,β_1,β_2 and σ, which are

$\hat{\beta}_0=3302.558,\ \hat{\beta}_1=69.666,\ \hat{\beta}_2=0.157,\ \hat{\sigma}=48.943$

Therefore, the fitted Logistic growth model is given by

$$y=\frac{3302.558}{1+69.666\exp(-0.157t)}, \qquad (14)$$

and the uncertain Logistic growth model is expressed as

$$y=\frac{3302.558}{1+69.666\exp(-0.157t)}+N(0,48.943) \qquad (15)$$

The fitting results of the Logistic growth model (14) and the observational data in Table 1 are shown in Figure 1. which shows a good fit between the fitted Logistic growth model and the observational data.

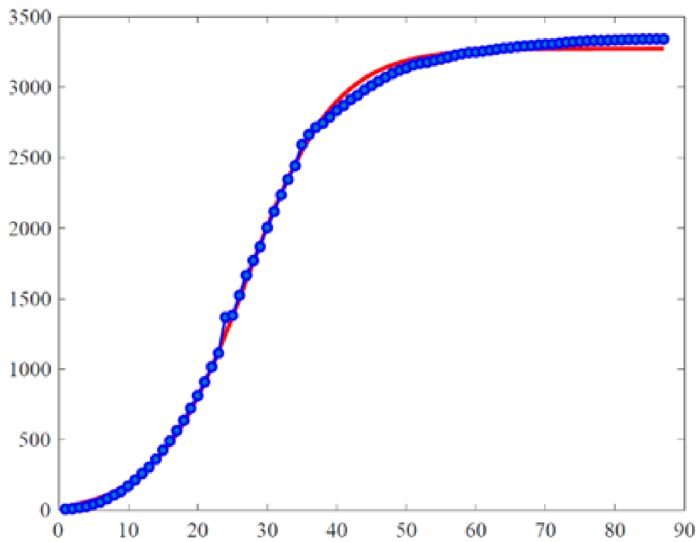

Fig. 1. Fitted Logistic growth model and observational data

To validate the rationality of model (15), by using

$$\varepsilon_t=y_t-\frac{3302.558}{1+69.666\exp(-0.157t)}$$

for $t=1,2,\cdots,87$, we can generate 87 residuals $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{87}$, as shown in Figure 2.

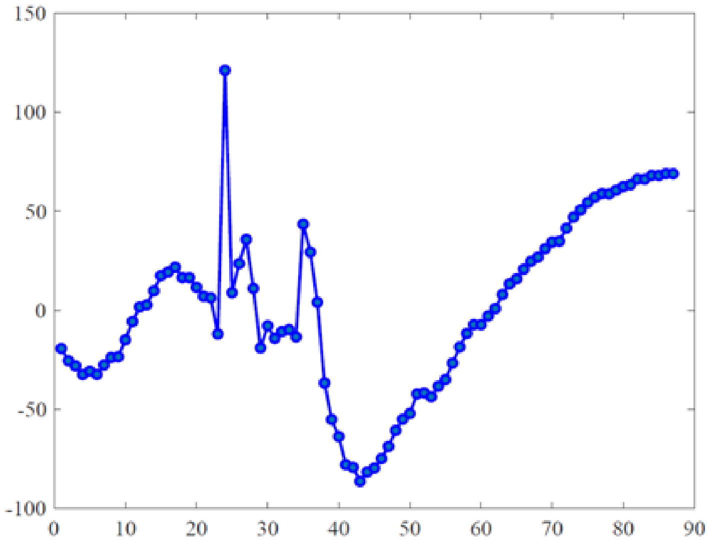

Fig. 2. Residual plot of the estimated uncertain Logistic growth model (15)

we need to test whether the above 87 residuals, $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{87}$ is set of samples of the population. Suppose the significance level is set to $\alpha = 0.05$. Then, it follows from $\alpha \times 87 = 4.35$ and Equation (9) that the test is

$$W = \{(z_1, z_2, \cdots, z_{87}): \text{至少4个} t, 1 \le t \le 87, \text{使得} z_t < -98.856 \text{或} z_t > 98.856)$$

As shown in Figure 2, it is evident that only $\varepsilon_{24} \notin [-98.856, 98.856]$.

Consequently, we have $(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{87}) \notin W$ and the estimated uncertain Logistic growth model (15) is a good fit for observational data.

Based on model (15), the predicted value of the death toll from COVID-19 in China on April 16, 2020 is obtained, that is

$$\hat{y}_{88} = \frac{3302.558}{1 + 69.666\exp(-0.157 \times 88)} + N(0, 48.943) \sim N(3302, 48.943).$$

Therefore, the predicted death toll from COVID-19 in China on April 16, 2020 is 3,302, and the 95% prediction interval is $3302 \pm \frac{48.943\sqrt{3}}{\pi} \ln \frac{1+0.95}{1-0.95}$, that is $3302 \pm 99 = [3203, 3401]$.

IV. CONCLUSION

To address the limitations of traditional least squares methods in studying uncertain growth models, this paper is the construction of a statistical invariant for uncertain growth models based on the observed data and uncertainty distribution of disturbance terms, and applying the least squares method to the parameter estimation of this statistical invariant. A dedicated numerical algorithm is developed to compute the corresponding estimators. Subsequently, this paper applies the hypothesis test for uncertainty to validate the applicability of the estimated uncertainty logistic growth model, while its effectiveness is demonstrated by successfully predicting China's COVID-19 death toll.

Funding: This work was supported by Shanxi Datong University Project (No. 2022Q15).

REFERENCES

[1]. Liu, B.. Uncertainty Theory. 2nd ed., Springer: Berlin, Germany (2007).
[2]. Liu, B.. Some research problems in uncertainty theory. J. Uncertain System., 3, 3–10(2009).
[3]. Yao K., Liu B.. Uncertain regression analysis: An approach for imprecise observations. Soft Computing, 22(17), 5579-5582 (2018).
[4]. Ye T., Liu Y.. Multivariate uncertain regression model with imprecise observations. Journal of Ambient Intelligence and Humanized Computing, 11, 4941-4950(2020).
[5]. Jiang B., Ye T.. Uncertain panel regression analysis with application to the impact of urbanization on electricity intensity. J. Ambient. Intell. Humaniz. Comput. , 14, 13017–13029(2023).
[6]. Fang L., Hong Y..Uncertain revised regression analysis with responses of logarithmic, square root and reciprocal transformations. Soft Computing, 24(4), 2655-2670(2020).
[7]. Liu Y., Liu B.. Estimation of uncertainty distribution function by the principle of least squares. Communication Statistics-Theory and Methods, 53(21), 7624-7641(2024).
[8]. Lio W., Liu B.. Residual and confidence interval for uncertain regression model with imprecise observations. J. Intell. Fuzzy Syst.,, 35, 2573–2583(2018).

[9]. Lio W., Liu B.. Uncertain maximum likelihood estimation with application to uncertain regression analysis. Soft Comput., 24, 9351–9360(2020).

[10]. Ye T., Liu B.. Uncertain hypothesis test with application to uncertain regression analysis. Fuzzy Optimization and Decision Making, 22, 157-174(2022).

[11]. Liu Z.. Uncertain growth model for the cumulative number of COVID-19 infections in China. Fuzzy Optimization and Decision Making. 20, 229-242(2021).