

Analysis of a modified Gompertz model based on nonlinear least squares method.

Abstract

This paper proposes a modified Gompertz model with a constant free term based on the classical Gompertz model. Different from the three-sum method for determining the parameters of classical Gompertz model, this paper constructs an optimization problem with the help of nonlinear least squares method. Moreover, employing the Levenberg-Marquardt method and the MATLAB software, the numerical solution of the optimization problem is found. A numerical example is provided in this paper. Finally, this paper uses this model to forecast the consumption level of Chinese rural residents, and the results illustrate the modified Gompertz model provides accurate prediction.

Keywords

Gompertz model, optimization problem, nonlinear least squares method, Levenberg-Marquardt method

1. Introduction

The Gompertz curve was first proposed by Benjamin Gompertz who is a British mathematician and insurer when studying population mortality^[1]. The Gompertz curve was first used by insurers to calculate the cost of life insurance. The curve is characterized by a slower increase, then a faster growth, and finally the growth tends to be slower. It can be applied in some new products that have just come out, the growth of industrial production, the life cycle of products, the population growth in a certain period of time, and so on. Hongmei Zhao^[2] used the Gompertz curve model to predict the development trend of China's automobile ownership in the next few decades by combination with per capita GDP and vehicle ownership. Xu Wang^[3] shows that the supply of fire extinguishing agent in the fire meets the Gompertz curve based on the relationship between the successful fire fighting and the dose of fire extinguishing agent in the oil tank area. Ye et al.^[4] used three curve models including the von Bertalanffy, the Gompertz and the Logistic to fit the growth curves of 60 pigeons which are randomly selected from 0 to 30 days old. The calculation results show that the Gompertz curve has a high degree of fit than other models. Wang et al.^[5] took the beef cattle industry as an example to examine the development stage and spatial differentiation characteristics of the beef industry from the national and provincial scales by using the Gompertz model, the Gini coefficient and the spatial econometric model.

The common method for parameter estimation of Gompertz model is the three-sum method, which can be seen in the literature^[6]. Jin Kaizheng^[7] initially used the polynomial and differential theory, and the least squares method to propose a better fitting method for parameter estimation of Gompertz curve. Zhu Yuren^[8] applied the GNL method of the implicit function curve, the coordinates of the Gompertz curve at the inflection point, and the relationship between the derivative and the parameter to perform the least square fitting. The author presented the system parameters, and provided some numerical examples. Yin et al.^[9] proposed a nonlinear regression least squares method for parameter estimation of improved Gompertz model based on Gauss-Newton least squares method and multiple linear regression least

squares method, and then verified by experimental data.

Based on the Gauss-Newton algorithm, the matrix is required to be a full rank in the iterative process of the numerical model. In this paper, the Levenberg-Marquardt method is used to discuss the modified Gompertz model with constant free terms. Under the nonlinear equations of system parameters are satisfied, the numerical solution of the model parameters is given.

2. Classic Gompertz model

This section presents a classic Gompertz model proposed by Benjamin Gompertz, a British mathematician and statistician, and the three-sum method to determine model parameters.

The classical Gompertz curve equation is

$$Y_t = ka^{b^t} \quad (1)$$

where k , a , b are unknown arguments, $k > 0$, $0 < a \neq 1$, $0 < b \neq 1$.

To solve the parameters in equation (1), take the logarithm of the two sides of the equation and get:

$$\ln Y_t = \ln k + b^t \ln a \quad (2)$$

Let the three local sum of the model be S_1 , S_2 , S_3 , that is

$$\begin{cases} S_1 = \sum_{t=1}^m \ln Y_t \\ S_2 = \sum_{t=m+1}^{2m} \ln Y_t \\ S_3 = \sum_{t=2m+1}^{3m} \ln Y_t \end{cases} \quad (3)$$

As a result, the equations can be obtained as follows

$$\begin{cases} b = \left(\frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ a = e^{\frac{(S_2 - S_1) \frac{b-1}{(b^m - 1)^2}}{m}} \\ k = \frac{1}{m} \left(S_1 - \frac{b^m - 1}{b - 1} \right) \ln a \end{cases} \quad (4)$$

Once given the original sequence, three local and values can be obtained, and then through (4) the expression of the model parameters can be obtained. Further, substituting obtained values into equation (1) to predict.

In the process of solving the model parameters, the three-sum method is very cleverly obtained, but such a method can only handle expressions of special structures which is not applicable to the more general model three-sum method. To this end, this paper combines the multivariate function to find the

extremum, and the nonlinear least squares theory and other theories to solve the general model parameters.

3. Modified Gompertz model

3.1. A modified Gompertz model based on non-linear least squares method

Firstly, the constant free term is introduced into the classical Gompertz model, and the modified Gompertz curve model is obtained as follows

$$Y_t = ka^{b^t} + c \quad (5)$$

where k , a , b are unknown parameters, $k > 0, 0 < a \neq 1, 0 < b \neq 1, c \in R$.

In fact, if the free term $c = 0$, it reduces to the classical Gompertz curve model. However, due to the introduction of the free term, the three-sum method is no longer applicable. For this reason, this paper gives a general method to solve the new model. Let \hat{Y}_t be the calculated value under the parameters k , a , b , c are determined, and Y_t is the given value, so the square sum of error between the two values is provided as follows

$$H(k, a, b, c) = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - ka^{b^t} - c)^2 \quad (6)$$

where n represents the number of raw data. From the Eq. (6), the problem of solving k , a , b , c is transformed into an optimization problem given below.

$$\min_{k, a, b, c} H(k, a, b, c) = \min_{k, a, b, c} \sum_{t=1}^n (Y_t - ka^{b^t} - c)^2 \quad (7)$$

According to the theory of extreme value of multivariate function, the derivation of H about parameters k , a , b , c , respectively, are obtained

$$\frac{\partial H}{\partial k} = \sum_{t=1}^n (2a^{b^t} c + 2^{2b^t} k - 2a^{b^t} Y_t) \quad (8)$$

$$\frac{\partial H}{\partial a} = \sum_{t=1}^n (2a^{-1+b^t} \cdot b^t c k + 2a^{-1+2b^t} \cdot b^t k^2 - 2a^{-1+b^t} \cdot b^t k Y_t) \quad (9)$$

$$\frac{\partial H}{\partial b} = \sum_{t=1}^n (2a^{b^t} b^{-1-t} c t k \ln a + 2a^{2b^t} b^{-1-t} t k^2 \ln a - 2a^{b^t} b^{-1-t} t k Y_t \ln a) \quad (10)$$

$$\frac{\partial H}{\partial c} = \sum_{t=1}^n (2c + 2a^{b^t} k - 2Y_t) \quad (11)$$

$$\text{Setting } \frac{\partial H}{\partial k} = f_1(k, a, b, c), \frac{\partial H}{\partial a} = f_2(k, a, b, c), \frac{\partial H}{\partial b} = f_3(k, a, b, c), \frac{\partial H}{\partial c} = f_4(k, a, b, c).$$

Then the formulae (8)-(11) are converted to the following equations

$$\begin{cases} f_1(k, a, b, c) = 0 \\ f_2(k, a, b, c) = 0 \\ f_3(k, a, b, c) = 0 \\ f_4(k, a, b, c) = 0 \end{cases} \quad (12)$$

Equation (12) is a nonlinear equation, it is generally impossible to obtain k, a, b, c analytical. The commonly method is used to solve the problem is the Gaussian Newton algorithm, but this algorithm in the iterative process requires the matrix is full rank, and this condition limits its application. In order to overcome this difficulty, we use the Levenberg-Marquardt algorithm to solve the formula (12). Thence the equation (12) is transformed into an equivalent form

$$\min_{x \in R^4} f(x) = \frac{1}{2} \|f(x)\|^2 = \frac{1}{2} \sum_{i=1}^4 f_i^2(x), x = (k, a, b, c) \quad (13)$$

Its search direction

$$d_k = -\left(J_k^T \cdot J_k + u_k I\right)^{-1} J_k^T f_k \quad (14)$$

which $J_k = f'(x) = [f_1'(x), f_2'(x), f_3'(x), f_4'(x)]$

The gradient of the formula (13) is given by

$$g(x) \square \nabla f(x) = \nabla \left(\frac{1}{2} \|f(x)\|^2 \right) = J^T(x) f(x) = \sum_{i=1}^4 f_i(x) \nabla f_i(x) \quad (15)$$

3.2. Model error testing criteria

In order to test the prediction accuracy of the modified Gompertz model, the absolute percent error (APE) and mean absolute percent error (MAPE) are used in this paper. They are defined as follows

$$APE(k) = \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%, \quad k = 2, 3, \dots, n \quad (16)$$

$$MAPE = \frac{1}{m-l+1} \sum_{k=l}^m \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%, \quad m \leq n \quad (17)$$

3.3. Modelling steps

In order to demonstrate more clearly the specific solution process for the modified Gompertz model, the following steps are given.

Step 1

Given $\rho, \sigma \in (0, 1)$ and $u_0 > 0, x_0 \in R^4$, and let $k = 0$;

Step 2

If $g(x_k) = 0$, stop the calculation, otherwise, turn to step 3;

Step 3

Solve equations $(J_k^T J_k + u_k I) d_k = -J_k^T f_k$, get d_k ;

Step 4

Let m_k is the least nonnegative integers m satisfying the following inequalities

$$f(x_k + \rho^m d_k) \leq f_k + \sigma \rho^m g_k^T d_k$$

Step 5

Let $\alpha_k = \rho^{m_k}$;

Set to $x_{[k]} = x_k + \alpha_k d_k, k = k + 1$, go to step 2.

4. Numerical example

In this paper, the classical Gompertz model and the modified Gompertz model are used to verify the accuracy. Assuming original sequences is $Y_t = (15.4490, 15.9636, 16.5700, 17.2892, 18.1478, 19.1809, 20.4346, 21.9709, 23.8737, 26.2593, 29.2908, 33.2022, 38.3354, 45.2015, 54.5838)$. The first 12 data are used to construct the models, and the left 3 data are checked. All the calculations are done by the Matlab software.

Classic Gompertz model

$$Y_t = 1.4168 \times 1.2515^{1.1443^t}$$

Modified Gompertz model

$$Y_t = 7.2574 \times 1.5041^{1.1077^t} + 3.8096$$

The calculated numerical results are shown in table 1 and Figure 1.

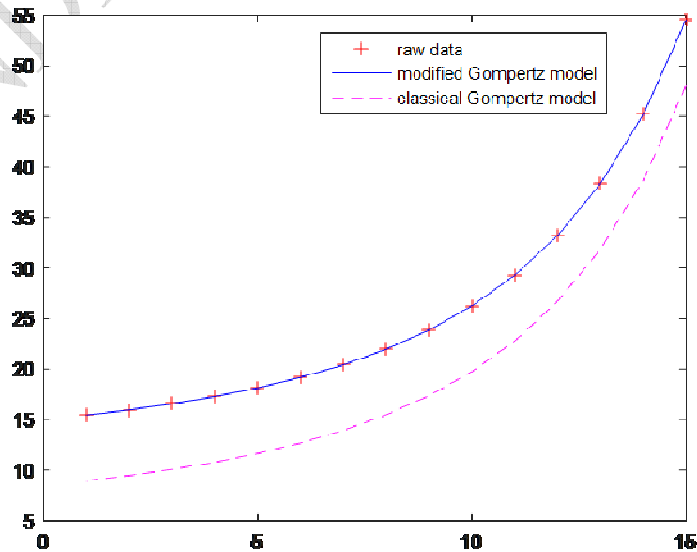


Figure 1. Results by classic Gompertz model and modified Gompertz model

Table 1. Numerical results obtained by classical Gompertz model and modified Gompertz model

ordina	data	Classical Gompertz	Modified Gompertz
1	15.4490	1.8314	15.2155
2	15.9636	1.9005	15.7844
3	16.5700	1.9828	16.4477
4	17.2892	2.0814	17.2254
5	18.1478	2.2002	18.1428
6	19.1809	2.3445	19.2324
7	20.4346	2.5212	20.5362
8	21.9709	2.7399	22.1095
9	23.8737	3.0135	24.0254

The results show that the modified Gompertz has a higher accuracy than the classical Gompertz, and our generalization and calculation process is feasible.

5. Forecasting the consumption level of Rural residents in China

5.1. Data Sources

In order to verify the practicability of the above methods and models, this paper selects the consumption level of Chinese rural residents as an application. The raw data come from the China Statistical Yearbook 2017^[10] which given in Table 2.

Table 2. Raw data of the consumption level of rural residents in China from 1998 to 2014

year	consumption level of Rural residents in China/yuan	year	consumption level of Rural residents in China/yuan	year	consumption level of Rural residents in China/yuan
1998	1778	2004	2521	2010	4941
1999	1793	2005	2784	2011	6187
2000	1917	2006	3066	2012	6964
2001	2032	2007	3538	2013	7773
2002	2157	2008	4065	2014	8711
2003	2292	2009	4402		

5.2. Prediction results

With the help of the Matlab software, we choose the first 13 data from 1998 to 2010 to develop the models, and the left data from 2011 to 2014 to test. The computational results of the classical Gompertz model and the modified Gompertz model are tabulated in Table 3 and figure 2. From the results, two models have been successfully caught the trend of the consumption level of Chinese rural residents, and they can reflect the trend of consumption level of Chinese rural residents from 2011 to 2014. However, the predicted value of the modified Gompertz model is closer to the real value than the classical Gompertz model.

Table 3. Prediction values of consumption levels of rural residents in China from 2011 to 2014

year	actual value/yuan	Classical Gompertz	Modified Gompertz
2011	6187	5438.8945	5711.9054
2012	6964	6316.2939	6589.3048
2013	7773	7394.8612	7667.8721
2014	8711	8731.8057	9004.8166

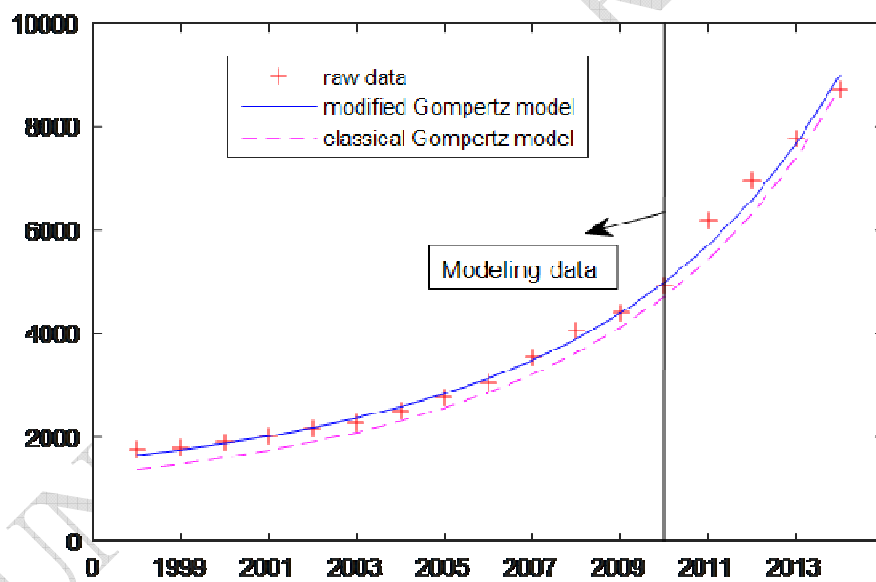


Figure 2. Simulation and prediction values of consumption levels of rural residents in China

5.3. Precision comparison

In order to compare the prediction accuracy between the classical Gompertz model and the modified Gompertz model, the APE and MAPE of the two models are calculated by the Eqs. (16) and (17). The results shown in Table 3a. In addition, Figure 3 shows a comparison of the relative errors between the two models.

Table 3a. Relative error of Chinese rural residents' consumption level in 2011-2014 years

year	Classical Gompertz	Modified Gompertz
2011	12.091	7.6789
2012	9.3007	5.3804
2013	4.8647	1.3524
2014	0.2388	3.3729
MAEP	9.8537	2.7821

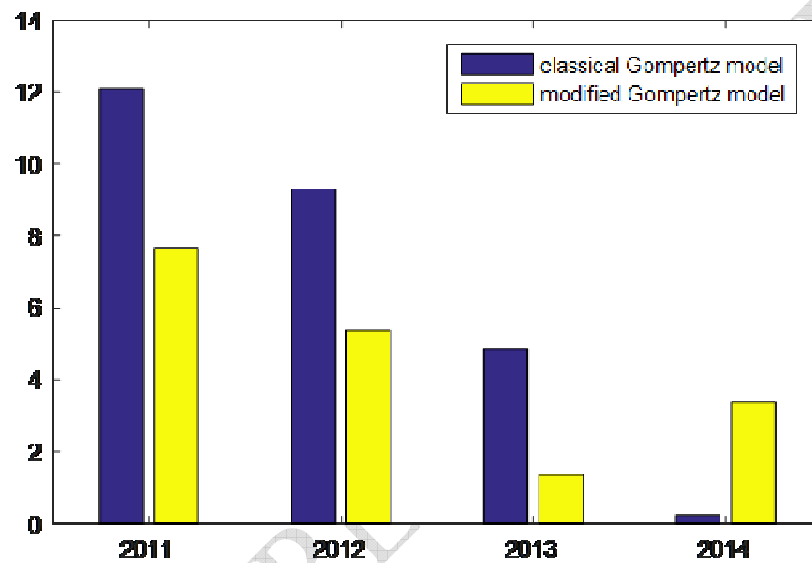


Figure 3. Relative error of consumption level of rural residents in China in 2011--2014 years

From tables 3 and Figure 3, we can see that the MAPA of the classical Gompertz model is 9.853737%, while the MAAP of the modified Gompertz model is 2.7821%. Therefore, results show that the modified Gompertz model is more accurate than the classical Gompertz model to forecast the consumption level of Chinses rural residents'.

6. Conclusion

In this paper, a modified Gompertz model is proposed based on the classic Gompertz model, the multi-function's extremum, the method of nonlinear least squares, and the Levenberg-Marquardt algorithm. We given a general method to determine system parameters. In future, the specific application of the new model is also our research topics.

Reference

- [1] Gompertz B. 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[J]. Philosophical Transactions of The Royal Society of London, 1825, 115.
- [2] HongmeiZhao. The Medium and Long Term Forecast of China's Vehicle Stock per 1000 Person Based on the Gompertz Model[J].Journal of Industrial Technological

Economics 2012, 31(07):7-23.

- [3] Xu Wang. Prediction on the supply quantity of fire extinguishing agent for petroleum storage tank based on the Gompertz curve[J]. Journal of Safety Science And Technology, 2013, 9(05):21-24.
- [4] Chao Ye, Xichao Lv, Sibao Ji, Wei Luo, Ang Li. 0-30-day-old pigeon growth of observation and analysis of growth curve fitting[J]. Poultry and Poultry Disease Control, 2014 (09) : 6-10.
- [5] Guogang Wang, Mingli Wang, Chun Yang. China beef industry development stage of identifying empty differentiation characteristics in a timely manner[J]. Economic Geography, 2014, 34 (10): 131-136+170.
- [6] Yanshan Song, Jianxi Wang, Yanhuang Xi, Siliang Guo. Applied Statistics -- to Excel for analysis tool (2nd Edition). Tsinghua University Press, 2018.
- [7] Kaizheng Jin. Study on the Gompertz curve fitting Method[J]. The economy of agricultural technology, 1991 (4) : 61-62.
- [8] Minren Zhu. Fitting of the Gompertz and Logistic models[J]. Mathematics in Practice and Theory, 2002, 32(5):705-709.
- [9] Yin Yi, Jiehui Xu, Jinshun Duanmu. Modified Gompertz based on Nonlinear regression least squares estimation of model parameters[J]. Practice and cognition of mathematics, 2002, 32 (5) : 705-709.
- [10] People's Republic of China National Bureau of Statistical Yearbook[M]. Beijing: China Statistics Press, 2017