Non-equigap Stage Sequence GM(1,1, α) Model and its Application

Liao Lijun^{1,a}, Gao Mingyun^{2,b}

¹Department of Statistics, Wuhan University of Technology, Wuhan, 430070, China

²Department of Statistics, Wuhan University of Technology, Wuhan, 430070, China

^aemail: 1213126637@qq.com, ^bemail: 1198915787@qq.com

Keywords: Non-equigap; grey model; GM(1,1)

Abstract. To predict the number of newborns, we establish an non-equigap stage sequence GM (1,1) model. Above all, we construct a non-equigap cumulative operator and calculate the optimal multiplier. Then we use the genetic algorithms to solve the model. On this basis, we improve it and obtain the GM(1,1, α) model. We make a contrast of the predicted number of newborns in two models above and the data in the traditional GM(1,1) model.

Introduction

The population problem is a world issue of universal concern. The population situation is an important national condition of a country or region. To realize the coordinated development of population, economy, society, resources and environment, it is of great significance to control the population. In order to achieve the target of population control and ensure the national sustainable development strategy is implemented smoothly, the Chinese government must make its efforts to stabilize the current low birthrate that a couple gives to 1.8 children averagely. At the same time, the government should take measures to improve the quality of the population. It shows that it is necessary to predict and analyze the national population, not only to know the status of our population growth, also to provide basis decision for the government's population control.

GM(1,1) model is a single-variable first-order linear model; it is a basic model in the grey prediction model. It's simpler than multi-variable multi-stage prediction models or other forecasting methods; it also has the characteristics of higher forecast accuracy and no strict requirements of the sample size and the probability distribution. Therefore, it can predict efficiently, adapt well and can be widely applied to the population, disease prediction and provide advanced services for government decision-making departments. As for certain issues, such as the prediction containing unequal intervals or stages, we can create non-equigap stage sequence GM(1,1) model to improve the prediction accuracy.

However, the grey prediction method has its own limitations, that it mainly reflects the regularity of the data but cannot fully reflect the impact which various irregular social factors make on the prediction index. Therefore, the decision cannot totally depend on its forecast results.

Methodology

Supposing that $x^{(0)}$ is a non-equigap stage sequence

$$x^{(0)} = (x^{(0)}(t_1), \dots, x^{(0)}(t_k), x^{(0)}(t_{k+1}), \dots, x^{(0)}(t_n)), \Delta t_1 = 1, \Delta t_i = t_i - t_{i-1} \neq \text{const}, i = 2, 3, \dots, n, \quad (1)$$

In which it starts to present as double-stage type before and after from the $k+1$ th point. Then we build a cumulative matrix $x^{(1)}$

$$x^{(1)} = \left(g\Delta t_1 x^{(0)}(t_1), \dots, \sum_{i=1}^k g\Delta t_i x^{(0)}(t_i) + \sum_{i=k+1}^{k+1} h\Delta t_i x^{(0)}(t_i), \dots, \sum_{i=1}^k g\Delta t_i x^{(0)}(t_i) + \sum_{i=k+1}^n h\Delta t_i x^{(0)}(t_i) \right) \quad (2)$$

Commonly used $z^{(1)}$ is the mean generated sequence of $x^{(1)}$. That is $z_a^{(1)}(k) = \alpha x^{(1)}(k) + (1-\alpha)x^{(1)}(k+1)$, $z^{(1)} = (z^{(1)}(t_1), z^{(1)}(t_2), \dots, z^{(1)}(t_n))$, to explore the most general case, we establish the weighted generated sequence of $Z_a^{(1)}$.

We establish whitening differential equation for the generate sequence: $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$

$$\text{Discrete models } \frac{\Delta x^{(1)}(t_i)}{\Delta t_i} + az^{(1)}(t_i) = b$$

$$\text{When } i = 2, 3, \dots, k, \frac{\Delta x^{(1)}(t_i)}{\Delta t_i} = \frac{x^{(1)}(t_i) - x^{(1)}(t_{i-1})}{\Delta t_i} = \frac{g\Delta t_i x^{(0)}(t_i)}{\Delta t_i} = gx^{(0)}(t_i)$$

$$\text{When } i = k+1, k+2, \dots, n, \frac{\Delta x^{(1)}(t_i)}{\Delta t_i} = \frac{x^{(1)}(t_i) - x^{(1)}(t_{i-1})}{\Delta t_i} = \frac{h\Delta t_i x^{(0)}(t_i)}{\Delta t_i} = hx^{(0)}(t_i)$$

Therefore the discrete model can be written as

$$\begin{cases} gx^{(0)}(t_i) + az^{(1)}(t_i) = b & i = 2, 3, \dots, k \\ hx^{(0)}(t_i) + az^{(1)}(t_i) = b & i = k+1, k+2, \dots, n \end{cases} \quad (3)$$

Substituting $z^{(1)}(k) = ax^{(1)}(k) + (1-\alpha)x^{(1)}(k+1)$ into the formula (3), we get

$$gx^{(0)}(t_j) + a[ax^{(1)}(j) + (1-\alpha)x^{(1)}(j+1)] = b, \text{ where } j = 2, 3, \dots, k-1,$$

$$gx^{(0)}(t_j) + a \sum_{i=1}^{j+1} g\Delta t_i x^{(0)}(t_i) - aag\Delta t_{j+1}(t_{j+1}) = b$$

$$\text{Where } j = k+1, k+2, \dots, n, hx^{(0)}(t_j) + a \sum_{i=1}^k g\Delta t_i x^{(0)}(t_i) + a \sum_{i=k+1}^j h\Delta t_i x^{(0)}(t_i) + (1-\alpha)h\Delta t_{j+1}x^{(0)}(t_{j+1}) = b$$

Proposition For non-equipag stage sequence GM (1,1) model, the following conclusion is established

Supposing that the first-level parameter pack of the non-equipag stage sequence GM (1,1) model is (a, b) . Supposing: $\mathbf{P}=(a,b)^T, \mathbf{Y}=(x^{(0)}(t_2), x^{(0)}(t_3), \dots, x^{(0)}(t_n))^T$, The least squares estimation parameters sequence of the staged GM (1,1) model $x^{(0)}(t_k) + az^{(1)}(t_k) = b$ is to meet: $\mathbf{P}=(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y}$.

$$\text{Where } \mathbf{B} = \begin{pmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \dots & -z^{(1)}(n) \\ 1 & 1 & \dots & 1 \end{pmatrix}^T, \quad \mathbf{Y} = (x^{(0)}(t_2) \ x^{(0)}(t_3) \ \dots \ x^{(0)}(t_n))^T.$$

$$\text{And } \mathbf{B} = \begin{bmatrix} -\alpha & \alpha-1 & 0 & \dots & 0 \\ 0 & -\alpha & \alpha-1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -\alpha & \alpha-1 \end{bmatrix} \times \begin{bmatrix} g\Delta t_1 & \dots & 0 & 0 & \dots & 0 \\ g\Delta t_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g\Delta t_1 & \dots & g\Delta t_k & 0 & \dots & 0 \\ g\Delta t_1 & \dots & g\Delta t_k & h\Delta t_{k+1} & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g\Delta t_1 & \dots & g\Delta t_k & h\Delta t & \dots & h\Delta t_n \end{bmatrix} \times \begin{bmatrix} x_1^{(0)}(1) & -1/g\Delta t_1 \\ x_1^{(0)}(2) & 0 \\ \vdots & \vdots \\ x_1^{(0)}(n) & 0 \end{bmatrix},$$

From the fact above that accumulating the non-equal interval stage sequence, it generates to the matrix decomposition. We can found that the matrix can exactly reflect three features as accumulated generating, equally spaced and stage. The final matrix is consisted by the product of two corresponding matrix (an n-order lower triangular matrix and two n-order diagonal matrix).

The whitening responsive of the non-equipag interval stage sequence grey GM (1,1) model is $\hat{x}^{(1)}(t_i) = \left[x^{(1)}(t_1) - \frac{b}{a} \right] e^{-a(t_i - t_1)} + \frac{b}{a}$, $i = 1, 2, \dots, n$. The reduced form is $\hat{\mathbf{x}}^{(0)} = \mathbf{A}^{-1} \hat{\mathbf{x}}^{(1)}$.

Case Study

A large number of newborns is born each year. Although the number in China is substantial reducing in recent years. Now, we study the diversification characteristics of the number of newborns in China from 1988 to 1997. We take the newborn's data of the recent 10 years as an example to make model. The data is in Table 1.

Table.1. The newborns population in China from 1988 to 1997 (millions)

years	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Newborn population	24.58	25.14	26.21	20.08	18.75	17.91	16.47	16.93	15.22	14.45

Modeling of the stage non-equigap GAGM (1,1). It can be seen from Figure 1, the newborns population curve is a two-stage type which jumps apparently at the third point. Therefore, we take $k=3$ in Generalized accumulation generated matrix and then create a stage-type non-equigap GAGM (1, 1) model as follow. The construction of the optimal multiplier calculation and non-equigap accumulate operator.

Assuming that $\mathbf{P}=(a,b)^T=f(g,h)$, the average relative error can be seen as a function of g and h . We optimize the model by minimizing the average relative error to determine the values of g and h , and work out the values of a and b using the least squares method. The optimization model is:

$$\min E(g, h) = \frac{1}{n} \left[\sum_{i=1}^n \left| \frac{\hat{x}^{(0)}(t_i) - x^{(0)}(t_i)}{x^{(0)}(t_i)} \right| \right]$$

Using intelligent algorithm such as genetic algorithms, taboo search algorithm, particle swarm optimization and etc. to solve the model, we can get parameters g and h

Using the genetic algorithm search, we can obtain $g=1.0714, h=1.2493$, (In fact, due to the predictive value of grey GM (1,1) model has nothing to do with the initial value, values of g, h are only related to the ratio of the two, therefore, both have an infinite number of values). Consequently, we can construct stages operator (diagonal matrix) as follows:

$$S = \text{diag}(1.0714, 1.0714, 1.0714, 1.2493,$$

$$1.2493, 1.2493, 1.2493, 1.2493, 1.2493, 1.2493)$$

We also notice that the time interval between the data is non-equigap. So we construct generalized accumulation operator. According to the model above, it could be found that $a=0.0539$ and $b=30.1196$.

Table 2 Model prediction of non-equigap stage GM (1,1) and traditional GM(1,1)

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Original value	24.58	25.14	26.21	20.08	18.75	17.91	16.47	16.93	15.22	14.45
Tradition model	24.58	25.3	23.44	21.71	20.11	18.63	17.26	15.99	14.81	13.72
Non-equigap	24.58	26.08	24.71	20.08	19.03	18.03	17.08	16.19	15.34	14.53
Average accuracy	Tradition model 95.12%					Non-equigap 97.90%				

Based on the genetic algorithms, we calculate the optimal parameters α whose value is 0.59433. The optimal fitness value is 0.47528.

Fig.1. The optimal fitness evolution curve

Fig.2. Comparison of the predicted value and the original value in two models

It can be seen from two tables and graphs above that it can improve the accuracy apparently when using the non-equigap stage sequence GM (1,1) model. It shows that the model has great reference value when we encounter such non-equigap stage type in the future.

Reference

- [1] Dai Wenzhan, Li Junfeng. non spacing GM (1,1) modeling system [J]. Engineering theory and practice, 2005, (9).
- [2] Song Zhongmin, Tong Xiao Jun, Xiao Xinping center approximation-grey GM (1,1) model [J]. Theory and practice of systems engineering, 2001, (5): 110-113.
- [3] Liu Sifeng, Dang Yaoguo, Fang Zhigeng. Grey system theory and its application [M] Beijing: Science Press, 2004.
- [4] Jiang Weidong, Li Xibing, Zhao Guoyan. Research on the time-data-separation of the forecast of unequal interval sequence[J].Systems Engineering.Theory & Practice, 2003, 23(1) : 68-71.
- [5] Wang Zhongxian, Wu Chundu, Shi Xuerong. A grey model for non-equidistant sequence[J] . Mathematics in Practice and Theory,2003, 33(10) : 16-20.
- [6] Mu Yong.White Exponential law coincidence GM (1, 1) model [J]. Mathematics in Practice and Theory, 2002, (1): 15-19.
- [7] Ji Peirong, Huang Weisong, Hu Xiangyong.Grey prediction model characteristics [J]. Theory and practice of systems engineering, 2001, (9): 105-108.
- [8] Deng Julong .Grey system prediction and decision-making. Wuhan: Huazhong University of Science and Technology Press, 1990,133-142.