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

Liao Lijun\textsuperscript{1,a}, Gao Mingyun\textsuperscript{2,b}

\textsuperscript{1}Department of Statistics, Wuhan University of Technology, Wuhan, 430070, China
\textsuperscript{2}Department of Statistics, Wuhan University of Technology, Wuhan, 430070, China

\textsuperscript{a}email: 1213126637@qq.com, \textsuperscript{b}email: 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

\[
\begin{align*}
  x^{(0)} = \left( x^{(0)}(t_1), \ldots, x^{(0)}(t_k), x^{(0)}(t_{k+1}), \ldots, x^{(0)}(t_n) \right), \\
  \Delta t_i = 1, \Delta t_{i+1} = t_i - t_{i+1} = \text{const.}, i = 2, 3, \ldots, n, \ldots
\end{align*}
\]  

(1)

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

\[
\begin{align*}
  x^{(1)} = \left[ g\Delta t_i x^{(0)}(t_i), \ldots, \sum_{i=1}^{k} g\Delta t_i x^{(0)}(t_i) + \sum_{i=k+1}^{k+1} h\Delta t x^{(0)}(t_i), \ldots, \sum_{i=k+1}^{k+1} g\Delta t x^{(0)}(t_i) + \sum_{i=k+1}^{k+1} h\Delta t x^{(0)}(t_i) \right]
\end{align*}
\]  

(2)

Commonly used \(z(1)\) is the mean generated sequence of \(x(1)\). That is \(z^{(0)}(k) = ax^{(0)}(k) + (1-a)x^{(0)}(k+1), z^{(1)} = \left( z^{(0)}(t_1), z^{(1)}(t_2), \ldots, z^{(1)}(t_n) \right)\), to explore the most general case, we establish the weighted generated sequence of \(Z^{(1)}\).
We establish whitening differential equation for the generate sequence: 
\[
\frac{dx^{(i)}(t)}{dt} + ax^{(i)}(t) = b
\]

Discrete models 
\[
\frac{\Delta x^{(i)}(t)}{\Delta t} + az^{(i)}(t) = b
\]

When \( i = 2,3,\cdots,k \), 
\[
\frac{\Delta x^{(i)}(t)}{\Delta t} = x^{(i)}(t) - x^{(i)}(t_{i-1}) = g\Delta t x^{(0)}(t_i) = g x^{(0)}(t_i)
\]

When \( i = k+1,k+2,\cdots,n \), 
\[
\frac{\Delta x^{(i)}(t)}{\Delta t} = x^{(i)}(t) - x^{(i)}(t_{i-1}) = h\Delta t x^{(0)}(t_i) = h x^{(0)}(t_i)
\]

Therefore the discrete model can be written as 
\[
\begin{cases}
x^{(i)}(t_i) + az^{(i)}(t_i) = b & i = 2,3,\cdots,k \\
hx^{(i)}(t_i) + az^{(i)}(t_i) = b & i = k+1,k+2,\cdots,n
\end{cases}
\]

Substituting \( x^{(i)}(k) = ax^{(i)}(k) + (1-a)x^{(i)}(k+1) \) into the formula (3), we get 
\[
gx^{(i)}(t_j) + a(\alpha x^{(i)}(j)+(1-a)x^{(i)}(j+1)) = b, \text{ where } j = 2,3,\cdots,k-1,
\]

\[
gx^{(i)}(t_j) + a \sum_{j=1}^{j-1} g\Delta t x^{(0)}(t_i) - aao\Delta t_{j+1}(t_{j+1}) = b
\]

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

Supposing that the first-level parameter pack of the non-equigap stage sequence GM (1,1) model is \((a, b)\). Supposing: 
\[
P = (a, b)^T, \quad Y = (x^{(0)}(t_2), x^{(0)}(t_3), \cdots, x^{(0)}(t_n))^T
\]

The least squares estimation parameters sequence of the staged GM (1,1) model \(x^{(0)}(t_k)+az^{(1)}(t_k)=b x^{(0)}\) is to meet: 
\[
P = (B^T B)^{-1} B^T Y
\]

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)

|-----------|------|------|------|------|------|------|------|------|------|------|

Modeling of the stage non-equipag 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-equipag GAGM (1, 1) model as follow. The construction of the optimal multiplier calculation and non-equipag accumulate operator.

Assuming that $P=(a,b)^T=f(g,h)$, 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:

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

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:

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

We also notice that the time interval between the data is non-equipag. 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-equipag stage GM (1,1) and traditional GM(1,1)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-equipag</td>
<td>24.58</td>
<td>26.08</td>
<td>24.71</td>
<td>20.08</td>
<td>19.03</td>
<td>18.03</td>
<td>17.08</td>
<td>16.19</td>
<td>15.34</td>
<td>14.53</td>
</tr>
</tbody>
</table>

Average accuracy  | Tradition model | 95.12% | | Non-equipag | 97.90% |

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

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


