Comparison of Grey Prediction with Age Shift Equation Used in Population Forecast

Yonghong Zou\textsuperscript{1,a} and Jiatao Huang\textsuperscript{2,b}

\textsuperscript{1}School of Statistics, Chengdu University of Information Technology, Chengdu 610103, China; \textsuperscript{2}MBA, Magdeburg University, Magdeburg, Germany; \textsuperscript{a}zyonghong@cuit.edu.cn, \textsuperscript{b}770169197@qq.com

Keywords: Grey prediction; Age shift equation; Population forecast; Comparison

Abstract. Population is big national power. It helps the government making macro-economic decision if the population dynamic change trend is hold. It is necessary to predict population. Age shift equation is one of the population prediction methods. Also, grey model is used by some scholars to predict population. These two methods are used to forecast the future ten-year population in Chengdu, Sichuan province, China based on the same data. The comparison is made from the prediction results to the prediction method. The conclusion was found that the result made by grey prediction is little higher than age shift equation’s. The influence of the fertility policy adjustment and mechanical change of population can be mixed together in grey prediction. But this method cannot analysis deeply on the changes in women of childbearing age and the fertility rate change of policy factors. The influence on fertility rate from the fertility policy adjustment can be mixed together in the forecast made by age shift equation. But some data is acquired difficultly.

Introduction

There are many methods to forecast population, but there is less do research of the comparison on method application. The methods of Grey Prediction and Age Shift Equation are used to predict the future ten years in Chengdu based on the 1% population sampling survey data in 2015 of Chengdu. Then the comparison is made from the prediction results to the prediction method.

Grey Prediction

Grey Prediction GM (1, 1) Model and its Principle. Grey model is prediction method that long-time describe things development rule using a small amount of incomplete information base on establishing grey differential prediction equation. It is characteristic of less information. Not only can be chaotic discrete original sequence is converted into an orderly sequence, but also the prediction precision is high. It can maintain the characteristics of the original system and better reflect the actual situation of the system.

The urban resident population quantity is affected by many factors, and these factors are not completely certain, so there are many difficulties to collect data. The factors of grey model can meet the require of population forecast.

The steps to establish GM (1, 1) model are the follow:
1. To assume that the original data sequence \( X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)) \)

   First-order accumulative generated on the original data sequence \( X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)) \)

   Among them, \( x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \quad k = 1, 2, \ldots, n \)

2. construct \( Z^{(1)} \) order
\( x^{(1)}(k) = \frac{1}{2} [x^{(1)}(k) + x^{(1)}(k-1)] \), have to
\[ Z^{(0)} = (z^{(0)}(1), z^{(0)}(2), \cdots, z^{(0)}(n)) \]
(3) An albino equation
\[ \frac{dx^{(1)}}{dt} + ax^{(1)} = b \]
(4) Strives for the parameters \( a \) and \( b \)
\[ B = \begin{bmatrix} -\frac{1}{2}z^{(1)}(2) & 1 \\ -\frac{1}{2}z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}z^{(1)}(n) & 1 \end{bmatrix} \]
If \( \hat{a} = [a, b]^T \) is a parameters of the sequence, and solve it by the least square method \( a = [a, b]^T = (B^T B)^{-1} B^T Y_n \)
(5) Dispersed the albinism differential equation, differential variable difference, then we got the GM(1, 1) differential equation
\[ x^{(0)}(k) + az^{(1)}(k) = b \]
It is called the basic form of GM (1, 1).
(6) To solve the albinism differential equation
Get the solution of the albinism differential equation
\[ x^{(1)}(t) = (x^{(1)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a} \]
The time response equation of \( GM(1,1) \) predict model \( x^{(0)}(k) + az^{(1)}(k) = b \) is:
\[ \hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \]
The reduced value is
\[ \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \]
\( \hat{a} \) is the development coefficient, \( \hat{a} \in [-2, 2] \), reflects the development tendency of \( \hat{x}^{(1)}(k) \) and \( \hat{x}^{(0)} \). \( b \) is grey action.

**Grey Prediction Test**

We also use the after-test residue checking in Grey prediction test. The propose is to test the reliability of the results. The steps are:
(1) To caculate the mean value of the original sequence and mean-square deviation are:
\[ \bar{x} = \frac{1}{n} \sum_{k=1}^{n} x^{(0)}(k), \quad S_1 = \sqrt{\frac{\sum_{k=1}^{n} (x^{(0)}(k) - \bar{x}^{(0)}(t))^2}{n-1}} \]
(2) To caculate the mean of residual errors and mean-square deviation are:
\[ \Delta^{(0)}(k) = \frac{1}{n} \sum_{k=1}^{n} \Delta^{(0)}(k), \quad S_2 = \sqrt{\frac{\sum_{k=1}^{n} (\Delta^{(0)}(k) - \bar{\Delta}^{(0)}(t))^2}{n-1}} \]
To calculate the posterior error ratio

\[ C = \frac{S_2}{S_1} \]

(4) \( p = P(\varepsilon(k) - \bar{\varepsilon} < 0.6745S_1) \) is called the small error probability.

(5) To determine the model level, the indicators are in Table 1:

<table>
<thead>
<tr>
<th>Level</th>
<th>The average relative error ( E )</th>
<th>Variance ratio ( C )</th>
<th>Small error probability ( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>&lt;0.01</td>
<td>&lt;0.35</td>
<td>&gt;0.95</td>
</tr>
<tr>
<td>II</td>
<td>&lt;0.05</td>
<td>&lt;0.5</td>
<td>&lt;0.80</td>
</tr>
<tr>
<td>III</td>
<td>&lt;0.10</td>
<td>&lt;0.65</td>
<td>&lt;0.70</td>
</tr>
<tr>
<td>IV</td>
<td>&gt;0.20</td>
<td>&gt;0.80</td>
<td>&lt;0.60</td>
</tr>
</tbody>
</table>

Instructions: The value of \( C \) is better as small as possible, that is to say \( S_1 \) is smaller enough than \( S_0 \), indicates the discrete degree of the original data is big, the discrete degree of the predicat error is small, so the prediction accuracy is high. \( P \) is better as bigger as possible, that is to say the probability of small error is large, indicates the fitting precision is higher. If the residual errors and the posterior error all passed the test the model can be used to predicate data. Or the residual errors of the model must be corrected.

** Forecast Using Grey Model **

We forecast the population development tendency of Chengdu in the future 10 years below using the statistical data of Chengdu from 2006 to 2015 (include Jianyang City).

<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>Number (10 thousand)</td>
<td>1613.93</td>
<td>1641.51</td>
<td>1669.77</td>
<td>1698.71</td>
<td>1728.35</td>
<td>1758.69</td>
<td>1789.75</td>
<td>1821.53</td>
<td>1854.06</td>
<td>1887.33</td>
</tr>
</tbody>
</table>

The sum of squared residuals = 7259.112
The average relative error = 1.738749 %
Relative accuracy = 98.26125 %
Posterior differential ratio test:
\[ C = 0.2111388 \]
\( C < 0.35 \), The prediction accuracy level of GM(1,1) is good.

From the results predicted the posterior error all passed the test. So we can say the result is accurate.

** Age Move Calculate Equation Method **

** Age Move Calculate Equation Method and its Principle.** Age move calculate equation method is a method that predict the population according to certain survival recursive year by year based on for each age group of actual population. The principle is the population is a function of time. In particular, the age of people is said using time. A person's age is increased by one year old. So, the age of the population is in constant changing group with time passing. The number of population is changing with time passing under a certain level of mortality. According to this principle, we can calculate the next near/group people’s quantity according to the certain level of mortality in some year/group.
Age Move Calculate Equation

The expression of age move calculate equation is

\[ P_{x+1(t+1)} = P_{x(t)} \cdot s_x \]

Here: \( P_{x+1(t+1)} \) is the population of the predicted year \((x+1)\). \( P_{x(t)} \) is the real population of the predicted base year \((x)\). \( s_x \) is the survival rate of \( x \) year. \( s_x = 1 - m_x \), \( m_x \) is the mortality of \( x \) year. \( P_{\omega-1(t+1)} \) is the estimate population of the highest age group in estimated annual. \( P_{\omega-1(t+1)} = P_{\omega-2(t)} \cdot s_{\omega-2} \).

To sum can get prediction total population. The model is:

\[ P_{c(t+n)} = \sum_{x=0}^{\omega-1} P_{x(t+n)} \]

Predict Condition Assumptions

Mortality rates set. Assume that all ages mortality rates remain unchanged 1% population sampling survey in 2015 in Chengdu.

The fertility rate set. Assume that all ages fertility rates remain unchanged 1% population sampling survey in 2015 in Chengdu.

Prediction Based on Age Move Calculate Equation

The steps are:

Set the childbearing age women’s fertility in Chengdu. The childbearing age women is the women that they are age during 15-49. Fertility rates by 1% population sampling survey of chengdu in 2015 fertility rates by age group.

Forecast the number of women of childbearing age. The numbers of women of childbearing age of Chengdu in 2016~2026 based on Age move calculate equation are: 454.35, 447.05, 435.31, 425.42, 413.21, 401.14, 391.6, 382.14, 372.33, 361.89 ten-thousand.

Forecast the number of 0-year-group in 2016~2026 according to the childbearing age women’s fertility in Chengdu and the number of the childbearing age women in the future 10 years are:454.35, 447.05, 435.31, 425.42, 413.21, 401.14, 391.6, 382.14, 372.33, 361.89 ten-thousand.

The number of the future 10-year population in Chengdu is below in table 3:

<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>Population (ten-thousand)</td>
<td>1576.26</td>
<td>1584.02</td>
<td>1590.52</td>
<td>1596.93</td>
<td>1602.86</td>
<td>1608.11</td>
<td>1612.71</td>
<td>1616.5</td>
<td>1619.28</td>
<td>1620.96</td>
</tr>
</tbody>
</table>

These predictal number is made based on the population number in 2015. Only consider the natural population growth factor, and did not reflect mechanical population growth factors (net immigration). The average annual growth rate of Chengdu permanent residents from 2006 to 2016 the was 16.2 ‰. So, we use 16.2 ‰ as the adjustment coefficient to reflect the influence of the mechanical population growth factors. The results below in Table 4:

<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>Population (ten-thousand)</td>
<td>1601.80</td>
<td>1627.75</td>
<td>1654.12</td>
<td>1680.92</td>
<td>1708.15</td>
<td>1735.82</td>
<td>1763.94</td>
<td>1792.52</td>
<td>1821.56</td>
<td>1851.07</td>
</tr>
</tbody>
</table>
Macros. Do not use any macros for the figures and tables. (We will not be able to convert such papers into our system)

The Comparison of Two Methods

First of all, the comparison results. From the results of table 1 and table 2, we can see the grey prediction results overall move slightly higher than the age calculation equation method. Investigate its reason mainly is that age to move to calculate equation method is based on the structure of resident population of Chengdu in 2015 as a benchmark, not considering the mechanical changes of population factors (migration). For the national center of Chengdu city, the urbanization level and the level of population migration is in very high level. But due to the defect of statistical yearbook data, the factor cannot blend in age move calculate equation method, only by other way to adjust. Results of age move calculate equation is slightly lower than the results of gray forecast.

Second, compared the merits of the method. The advantage of age move calculates equation: will melt into the influence of the fertility policy adjustment to population reproduction because the fertility rate in age of childbearing women are grasped very accurately. The shortage is: The data is insufficient. Generally, there is year-group data only in census data or population sampling survey data, and there is no year-group data of the changes in the mechanical population (immigration and emigration). We can only adjust it by other method. The advantage of grey prediction: Can be unified the influence into fertility policy adjustment factor and mechanical changes factor of population, but it will not focus on investigation analysis of the effects on the change of childbearing women and the fertility policy adjustment.

Acknowledgements

The study result is by the project of 1% population sampling survey of Chengdu funding.

References


