Statistical analysis and prediction based on recursive calculation under short and long term models

Zhang Liu
School of Computer and Information Engineering, Jiangxi Agricultural University, Nanchang 330045, China

Corresponding author: Zhang Liu, liuzhang1006@163.com

Abstract. One of the most critical factors for economic growth is population issue, which is also a key factor restricting the development of a country. Currently, the population aging process has accelerated, the number of birth sex ratios has continued to rise, and the rural population has been urbanized. Based on the recursive calculation method, the relevant population short-term growth prediction model and long-term growth prediction model are established in this paper. According to the census data, the future possible number is predicted. Practice has proved that our prediction results are basically in line with China's development.

1. Introduction
For society to develop, economic activities are the main body, and the changing trend of population growth has a great impact on social and economic development. Therefore, the prediction method based on data analysis for economic and social development has a very important role in practice. The accurate prediction of a country’s zero or regional population is an important basis for formulating corresponding macroeconomic policies [1]. Based on the available data and the use of mathematical models, analysis and prediction of Chinese population also have important practical applications.

In recent years, some new characteristics have emerged in China’s population development, the acceleration of the aging process, the continuing increase in the sex ratio of the birth population, and the urbanization of the rural population have all affected the growth of our population. The "National Population Development Strategy Research Report" released also made further analysis. There have been many studies on China's population issues, and many scholars have done a lot of work. For example, Rozell [2] compared the UN and IIASA population projections, and found that the UN’s probabilistic population projections should be used with caution as they tend to underestimate the uncertainty in long-range population forecasts. The scenario projections were better suited to long-range climate change policy analysis. The climate change integrated assessment models that performs better than current alternatives. Zhang et al. [3] used the gray prediction model of residuals to predict the state of China's population aging from 2012 to 2020, providing a scientific theoretical basis for the Chinese government to formulate correct relevant policies. Based on the historical bulletin data and census data of Gansu Province, Jia et al. [4] used the linear regression model, block growth model, gray forecasting model and China Population Forecasting Software (CPPS) to forecast and analyze development trends. Through the establishment of a population prediction equation, Han and Chen [5] conducted a comprehensive forecast analysis of the total population and structure of the
registered population of Guangzhou, and deeply studied the impact of population migration on the population growth and population structure of Guangzhou. Yang [6] analyzed the current situation of Shanghai’s population aging and conducted empirical analysis based on the GM (1,1) model. Tang et al. [7] selected data from 2003 to 2015 to build a static gray model, and then built an equal-dimensional replenishment dynamic model and a static gray model, and predicted the population from 2018 to 2028. Zhang [8] used the data from the 5th National Census and the 6th National Census of Jiangxi Province to predict the population development trend of Jiangxi Province from 2011 to 2060, and to take into account the size of the elderly population, the coefficient of the elderly population, the ratio of old to young, the ratio of elderly to elderly Coefficients, social support coefficients and other aspects of the Jiangxi Province population aging degree in the next 50 years. Shi Xianjin [9] combed 16 cities' previous urban master plans to predict their future population size, and found that nearly 80% of the samples underestimated the future population size. Therefore, by constructing a labor transfer model and using a spatial econometric model to analyze the economic drivers of urbanization in 31 provinces and cities in China from 2000 to 2017. In this paper, we set up a mathematical model of China’s population growth based on the actual situation of China and the above characteristics of population growth, and thus make predictions on the short-term and long-term trends of China’s growth, and analyze and discuss the relevant results.

2. Analysis and prediction of population growth based on recursive calculation

2.1. Analysis of short-term population prediction

The influencing factors of short-term population growth mainly include population-based birth rate, mortality rate, male-female ratio, population age composition, and population migration. According to the specific data of each age, these data can be regarded as continuous. However, if the continuous form is used, it involves the problem of solving partial differential equations, it is difficult to obtain the prediction results, and at the same time, the amount of data is also large. Therefore, we directly establish a discrete form of population development model to predict the short-term population. Since the birth rate and death rate of the population are different, the age structure is also an important factor affecting population growth. Therefore, this model is mainly to grasp the development trend of China's short-term population and make a short-term population forecast.

2.1.1. Population prediction equation. This model uses a recursive calculation process, which is, inferring future data at a certain time based on known data at a certain time. When estimating population problems, you can use the age data of the population for the current year to calculate the population data for the corresponding year with a increase of one year. The specific process is as follows:

(1) The number of people who are i years old in year x to i + 1 years in year x + 1 is given by

\[ a_{i+1}(x+1) = \beta_i(x) \alpha_i(x) \]  

Where \( i = 0,1, ..., m-1; \ k = 0,1, .... \). When considering the impact of women of childbearing age and fertility rate on the number of births, the number of births in year x is

\[ a_{00}(x) = \sum_{i=0}^{m-1} \gamma_i(x) \alpha_i(x) \]  

Due to the survival rate and the number of births impact, the total number of babies in year x is given by \( a_0(x) = \beta_{00}(x) a_{00}(x) \).

(2) The number of people in k + 1 year who is 1 year old is given by

\[ a_k(x+1) = \beta_k(x) \alpha_{00}(x) \]  

(3) Let the female fertility rate be \( \gamma_i = \beta(x) h_i(x) \), \( h_i(x) \) is the fertility model, and

\[ \chi(x) = \sum_{i=1}^{m-1} \gamma_i(x) \]  

which means the average fertility rate of all women of childbearing age in the kth year Therefore, we can rewrite (2) as
\[ a_i(x+1) = \beta_0(x) \beta_{i0}(x) \sum_{i=0}^{k} \gamma_i(x) \lambda_i(x) a_i(x) = \beta(x) \sum_{i=0}^{k} \gamma_i(x) a_i(x), \]  

Where \( \gamma_i(x) = \beta_0(x) \beta_{i0}(x) h_i(x) \lambda_i(x) \).

(4) By introducing the vector \( A(x) = [a_1(x), a_2(x), \ldots, a_m(x)]^T \), the discrete form of the population prediction equation can be obtained as

\[ A(x+1) = M(x) \cdot A(x) + \chi(x) N(x) \cdot A(x) \]

where

\[
M(x) = \begin{bmatrix}
0 & 0 & \cdots & 0 & 0 \\
\beta_1(x) & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \cdots & 0 & 0 \\
0 & \cdots & \cdots & \beta_{m-1}(x) & 0
\end{bmatrix},
\]

\[
N(x) = \begin{bmatrix}
0 & \cdots & \gamma_1'(x) & \cdots & \gamma_m'(x) & 0 & \cdots & 0 \\
0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0
\end{bmatrix}.
\]

Note that this population prediction equation is a first-order difference equation. In practice, if the initial population \( A(0) \) of each age is known, the matrix \( M(k) \) and \( N(k) \) are independent of time (that are constants), the population prediction equation becomes

\[ A(x+1) = M \cdot A(x) + \lambda(x) N \cdot A(x). \]

Let be the weight-bearing factor for females of age \( i \), that is, the birth pattern. In a stable environment, it can be approximated as having nothing to do with \( x \), i.e., shows at what ages the fertility rate is high and at which ages the fertility rate is low. From the demographic data, we can know the current actual fertility pattern. Generally, the probability density function of the birth pattern adopts the distribution density function as

\[ h_i(x) = \frac{(i - i_t)^{x-1} e^{-i \cdot x}}{\theta^x \Gamma(x)}, \quad i > i_t, \]

Where \( i_t = i + n - 2 \), it represents the age at the highest value of the fertility rate, that is, the legal age of marriage, and \( n \) is an index reflecting the reproductive age of women of childbearing age.

### 2.1.2. Solving the model

According to the appeal model, MATLAB is used to solve the problem, and the results of the total population prediction under different total fertility rates within 30 years can be obtained. Here we can predict the number of people in the next 30 years based on the population data at the end of 2019, and randomly give prediction results every 3 or 5 years, as shown in Table 1. It can be seen from Table 1 that the different values of the total fertility rate have a great influence on the future population.

**Table 1.** Predicted population values with different total fertility rates.

| Population Fertility rate | 1.3  | 1.8  | 2.1  |
|---------------------------|------|------|------|
| 2025                      | 13.829 | 13.982 | 14.121 |
| 2030                      | 13.458 | 13.876 | 14.136 |
| 2035                      | 13.149 | 13.672 | 14.156 |
| 2040                      | 12.710 | 13.571 | 14.362 |
| 2045                      | 12.244 | 13.490 | 14.448 |
| 2050                      | 11.712 | 13.314 | 14.512 |
As shown in Table 1, when the fertility rate is 1.8, the population growth is relatively fast, and this value is relatively close to the current situation in China. When the fertility rate is 2.1, the population has reached a stable state. However, due to the relatively large population base in China, when the fertility rate is 2.1, the total population will remain high in the short term. The population problem should become a serious social problem, which is not conducive to national development.

2.2. Analysis of long-term population prediction

To make a long-term prediction of population growth, factors such as fertility pattern, fertility rate, urbanization of rural population, sex ratio of the population, and aging must be considered. These factors affect the long-term number of the population, but also have a future impact on the development of the population. Different constraints and influences.

As far as the long-term population development prediction model is concerned, it is necessary to analyze the impact of factors such as the fertility rate, sex ratio, and the proportion of women of childbearing age on the population based on relevant data. According to the different data analysis of cities and towns, the future population of each population is predicted. Fertility rate is mainly affected by factors such as population quality, urban-rural ratio, and government policies. There is unknown information, ambiguous mechanisms, and insufficient data between these factors. Therefore, the GM (1, 1) model in the grey system theory can be used to predict the development trend of fertility.

Taking into account the difference in the peak fertility of women of childbearing age in urban and rural areas, it has a greater impact on the fertility pattern. According to the difference of women of childbearing age when they reach the peak, the parameters of the fertility pattern are changed to indicate the impact on the future population growth.

The urbanization of the rural population is only considered in terms of the number of people. In fact, the impact of the rural population on the total urban population after moving into the urban area. Considering the total number of people in society, the migration of rural population and the migration of urban population are balanced. With reference to relevant data, the ratio of urban population to total population is increasing at a rate of 1% per year. In other words, the ratio of the urban population to the total population increases at a rate of 1% per year. In other words, the total annual urban population is constantly changing, and it will have a certain impact on the long-term population forecast. We can consider predicting and analyzing in two steps.

The continuous increase in the sex ratio of the birth population will lead to a decline in the female ratio of the birth population. In the long run, the reduction in the proportion of female population will lead to a decrease in the number of people of childbearing age. Therefore, the total population will be affected by adjusting the total fertility rate. In addition, some social factors, such as total fertility rate and social aging, can be used to analyze the impact of population growth, and a harmonious model of population growth can be considered. When the urbanization of the rural population, the continuously rising gender ratio of the birth population, and the acceleration of the aging process have been identified, the urban population, the town population, and the rural population are now analyzed and predicted.

2.2.1. Discussion of fertility patterns \( h_i(x) \). Consider the different situations of population development in cities, towns and townships:

(1) Assuming that the fertility pattern \( h_i(x) \) follows the \( \chi^2 \) distribution and takes \( \theta = 2, \alpha = n/2 \), so we have

\[
h_i(x) = \frac{(i-i_1)^{n-1} e^{-\frac{i-i_1}{2}}}{2^{n/2} \Gamma(n/2)}, i > i_1.
\]
2.2.2. Development trend of fertility rate $X_i(x)$. Fertility rate $X_i(x)$ is mainly affected by factors such as population quality, urban-rural ratio, and government policies. There is unknown information, ambiguous mechanisms, and insufficient data between these factors. Therefore, the GM (1, 1) model in the grey system theory is used here to predict the development trend of the fertility rate.

Adding once to the known data reference sequence, we can get $A^{(1)}(x) = \sum_{m=1}^{k} A_m(m)$ (k = 1, 2, 3), then we get a set of generated sequence $A^{(1)} = \{A^{(1)}(1), \ldots, A^{(1)}(3)\}$, the corresponding albino differential equation $\frac{dA^{(1)}}{dt} + aA^{(1)} = u$, denote that

$$\hat{a} = (a, u)^T, \ y_1 = [A_0(2), A_0(3)]^T, \ N = \begin{bmatrix} -\frac{1}{2} (A_i^{(1)}(2) + A_i^{(1)}(1)) & \cdots & 1 \\ \frac{1}{2} (A_i^{(1)}(3) + A_i^{(1)}(2)) & \cdots & 1 \end{bmatrix}.$$  

According to the least square method, we have $\hat{a} = (a, u)^T = [N^T N]^{-1} N^T y_1$, then applying SPSS to solve can get the deterministic growth trend of the sequence by the formula

$$A^{(1)}(x) = \left[ A_0(1) - \frac{u}{a} \right] e^{-a(x-1)} + \frac{u}{a}.$$  

2.2.3. Influence of factors such as birth pattern, fertility rate and urbanization. The discretized population development equations were used to establish models for the populations of cities, towns and townships. According to the above results, the population development equations are given by

$$A_c(x + 1) = MA_c(x) + X_c(x)N_c(x)A_c(x),$$  

$$A_x(x + 1) = MA_x(x) + X_x(x)N_x(x)A_x(x),$$  

$$A_h(x + 1) = MA_h(x) + X_h(x)N_h(x)A_h(x),$$
where $M$ represents the mortality rate. According to the data of the population yearbook of the national bureau of statistics of China, there has been little change in the population mortality rate in the past 20 years, so $M$ is taken as a constant.

The total number of women of childbearing age in cities, towns, and townships $N(x)$ can be predicted by the model of 2.2.2, and their respective fertility patterns can be determined by formula (8), where the fertility index $n$ is 13, 12, 10. Then, the $x$ The total population in $+1$ year is given by

$$A(x + 1) = A_c(x + 1) + A_z(x + 1) + A_x(x + 1).$$

For the predicted data, using the above steps to adjust, we can get the actual population in year $x + 1$.

3. Solving the model

According to the data of the sixth population census in China, using MATLAB to program and solve the two stages of population prediction, the corresponding results can be obtained as shown in Table 3.

Table 3. China's population forecast in the next 50 years.

| Year | Urban | Town | Countryside | Total |
|------|-------|------|-------------|-------|
| 2031 | 4.145 | 3.288| 6.525       | 13.958|
| 2036 | 4.224 | 3.293| 6.492       | 14.009|
| 2041 | 4.374 | 3.329| 6.328       | 14.031|
| 2046 | 4.413 | 3.378| 6.136       | 13.927|
| 2051 | 4.507 | 3.461| 5.923       | 13.891|
| 2056 | 4.593 | 3.479| 5.815       | 13.887|
| 2061 | 4.572 | 3.509| 5.767       | 13.848|
| 2066 | 4.558 | 3.511| 5.304       | 13.373|
| 2071 | 4.515 | 3.514| 5.127       | 13.156|
| 2076 | 4.482 | 3.501| 5.119       | 13.102|

It can be seen from this forecast table that the population in 2031 is 13.958 billion, which is roughly the same as the current population of China. This result is consistent with the future status of China's population; after 20 years, the total population of China will gradually decline from the peak. The data shows that reasonable measures must be taken to curb the significant decline in the population. On the other hand, it also shows that our discrete population prediction model is more feasible, and can more accurately predict the short-term population.

When taking variable values of total fertility rate and mortality rate, we make a new prediction of the future population, and we find that the total population of the next ten years is basically consistent with the predicted value of the above model. However, the data in the next ten years is quite different from the prediction results of the model. The results are shown in Table 4.

Table 4. Comparison of the predicted value of $x$ and the fixed value.

| Year | Predicted value | Fixed value |
|------|-----------------|-------------|
| 2031 | 13.071          | 13.016      |
| 2036 | 13.404          | 13.296      |
| 2041 | 13.726          | 13.547      |
| 2046 | 13.916          | 13.518      |
| 2051 | 14.328          | 13.419      |
| 2056 | 14.243          | 13.303      |
| 2061 | 14.855          | 13.108      |
| 2066 | 14.051          | 12.816      |
For short-term population forecasting, the values of total fertility and mortality can be considered unchanged, so short-term forecasting within ten years is feasible. However, the prediction of the total population in 20 to 50 years will not be reliable [10]. Also, the longer the predicted years, the greater the data error.

4. Conclusions
In this paper, we give the statistical analysis and prediction based on recursive calculation under short and long term models, it can be seen that the total fertility rate will have a significant impact on the future population. If the total fertility rate is too small, it will cause problems such as aging society. However, if the total fertility rate is large, because China’s population base is very large, the total population will grow rapidly, making the total population unable to be effectively controlled. In addition, the short-term and long-term prediction models used in this article can not only predict the number of the future population, but can also be applied to other fields, such as the field of biology, the field of chemistry, etc. The quantity in a certain period in the future can be used as a reference method for predicting the number of rare populations in China in the future, so as to provide better solutions for taking effective measures to protect.

Acknowledgement
This work is supported by Jiangxi Science and Technology Plan Fund Project (GJJ20180201).

References
[1] Jiang H., China Population Forecast Analysis, Science and Technology Management Research, 11: 142-145, 2005.
[2] Rozell, D. Using population projections in climate change analysis. Climatic Change 142, 521–529, 2017.
[3] Zhang R.Y., Ma Y.Q., Hao S.S. Application of Residual Error Gray Forecast Model on Prediction of Aging Development Trend in China% Application of Residual Error Gray Forecast Model in Prediction of Aging Development Trend in China. Mathematics in Practice and Theory, 043 (16): 162-166, 2013.
[4] Jia H.W., Xie Z.J., Gao Y.G. Gansu Province Population Forecast and Development Trend Analysis. Northwest Population, 181 (03): 122-130, 2018.
[5] Han Z.Z., Chen X.H. Guangzhou Population Forecast and Analysis from 2006 to 2020. Southern Population, 21 (3): 58-64, 2006.
[6] Yang M.R. Prediction of Shanghai population aging trend based on GM (1,1) model. Journal of Economic Research, 000 (018): 48-49,134, 2019.
[7] Tang X.F., Cui Y., Zhang S.L. Population prediction analysis based on the equal-dimension complementary gray model, Journal of Luoyang Normal University, 39 (2): 5-8, 2020.
[8] Zhang X.X. Jiangxi population forecast and development trend analysis. Jiangxi Social Sciences, 10: 233-237, 2012.
[9] Shi X.J. Why can't the urban master plan accurately predict the future population size? —— A spatial econometric analysis of the economic drivers of urbanization in China. Beijing Social Sciences (11), 2019.
[10] Howell P. E., Hossack B. R., Muths E., et al. A statistical forecasting approach to metapopulation viability analysis. Ecological Applications, 30, 2020.