Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 2
May 2008

A Prediction Model of China Population Growth

Hao Zhang, Chao Wang & Xiumin Zhang
College of Information and Communication Engineering
Harbin Engineering University
Harbin 150001, China
E-mail: boshaos66@163.com

Abstract

In this article, we develop a prediction model of China population growth, and notice that the modified index curve is a sort of curve that possesses the growth limit K in the statistics, which is the same with the problem of China population growth. Considering influences of urbanization, population aging and sex proportion, we adopt and improve the modified index model, and add some coefficients to adjust the mathematical equation according to the data. The computation results show that in a short period, the population quantity will increase slowly and approach a fixed value, and in a long term, with the influence of population aging, some factors will put up periodic fluctuations.

Keywords: Improved modified index model, Data fitness, Sex proportion, Urbanization, Population aging

1. Introduction

The population growth prediction mainly includes the prediction of total population and birthrate, death rate, sex proportion. In China, many special problems still exist in the prediction of population growth, such as national population policy, education degree, economic environment and human ideas. All population predictions are implemented based on certain data, and available data generally include sex proportion, death rate and birthrate. But these data are always incomplete and random, so in this article, we select and deal with these data and introduce some authorized data, and overcome the difficulties of incomplete information and accuracy.

The population growth is influenced by many aspects, in this article, we mainly consider following aspects.

(1) The influence of sex proportion. The difference of sex proportion is larger and the negative influence to the total population is larger.

(2) The influence of village population urbanization. The urbanization of village population cannot but induces the change of village and urban population structure and influences the development tendency of total population.

(3) The influence of birthrate to the total population. The national policies have large influences to the birthrate, and influence the change tendency of the total population.

(4) In the long-term population growth prediction, some influencing factors put up tendency fluctuation, which makes the population growth bring the element of tendency fluctuation, and some factors present periodic fluctuation, which makes the population growth bring the element of periodic fluctuation, and some factors present occasional fluctuation, which make the population growth bring the element of random fluctuation.

In this article, we establish the model based on above factors which are presented in the equation through the mathematical expression, and approximated and weakened to a certain extent in the middle and short term model.

2. Symbol explanations

t: The year needed to be predicted (In this article, the year of 1984 represents t=1 which is the initial value.).

e_{it} : The proportion of the influence factor i in the t'th year (i=1 represents the city, i=2 represents the town, and i=3 represents the village.).

p_{it} : The sex proportion in the t'th year.

k_i : The association degree of city population, town population and village population with the total population.

3. Establishment of the Model

3.1 The model of middle and short term

According to the actuality of China and the development of population model, we develop the equation based on the modified index, and the equation is as follows:

$$y_t = K + ab^t - \sum_{i=1}^{3} e_{it} k_i (p_{it} - 1)^2 ,$$

and to explain problems, we evolve the above equation and we can obtain this equation.

$$y_t = K + ab^t - e_1 k_1 (p_{1t} - 1)^2 - e_2 k_2 (p_{2t} - 1)^2 - e_3 k_3 (p_{3t} - 1)^2$$

First, we briefly introduce the modified index model, and the general form of the modified index curve is

$$\hat{Y}_t = K + a b^t .$$

Where, K, a and b are unknown constants, K>0, a≠0, 0<b≠1.

Supposed that every part has m periods, and the sums of various observation value are respectively S_1, S_2 and S_3, i.e.

$$s_1 = \sum_{t=1}^{m} Y_t ,\quad s_2 = \sum_{t=m+1}^{2m} Y_t ,\quad s_3 = \sum_{t=2m+1}^{3m} Y_t .$$

So, we can obtain

$$S_1 = mK + ab + ab^2 + \cdots + ab^m = mK + ab(1 + b + b^2 + \cdots + b^{m-1})$$

$$S_2 = mK + ab^{m+1} + ab^{m+2} + \cdots + ab^{2m} = mK + ab^{m+1}(1 + b + b^2 + \cdots + b^{m-1})$$

$$S_3 = mK + ab^{2m+1} + ab^{2m+2} + \cdots + ab^{3m} = mK + ab^{2m+1}(1 + b + b^2 + \cdots + b^{m-1}) .$$

And we can obtain solutions.

$$b = \left(\frac{s_3 - s_2}{s_2 - s_1} \right)^{\frac{1}{m}}$$

$$a = (s_2 - s_1) \frac{b - 1}{b(b^m - 1)^2}$$

$$K = \frac{1}{m} \left[s_1 - \frac{ab(b^m - 1)}{b - 1} \right]$$

We respectively consider the city population, town population and village population, and $e_{1t} k_1 (p_{1t} - 1)^2 , e_{2t} k_2 (p_{2t} - 1)^2 , e_{3t} k_3 (p_{3t} - 1)^2$ represent the sex population of city, town and village in turn and the influence of village population urbanization to the distribution of the population.

In the following text, we will concretely explain the function of every item and their influences to the development of the population.

3.1.1 Explanations of e_{it} and its computation method

e_{it} represents the weight of the influence of the population in the structure i occupying the total influences of city, town and village to the development of the population, i.e. we can confirm the influencing degree of city, town and village in every year to the development tendency of total population according to the available data. For example, in a certain year, e_{1t}, e_{2t} and e_{3t} are respectively 0.4, 0.3 and 0.4.

e_{it} = the sum of female and male in i in the t'th year/ the sum of female and male in the t'th year.

The urbanization of village population is also embodied in e_i because in the urbanization process of village population, changes of various weight e_1, e_2 and e_3 will certainly occur. For example, above various weights are respectively changed to 0.4, 0.4 and 0.2, which indicates that when village population enters into the city and town, the proportion of city and town population will increase, i.e. the influence of city and town to the development tendency of the total population increase.

So e_i represents the influence of city, town and village to the total population development tendency of China, and fully embodies the influence of village population urbanization to the development tendency of the total population in China.

Because of the middle and short term model, the urbanization of village population can not achieve saturation, so we approximate e_i as the linear function.

According to the data from 2001 to 2005 in China population sampling data, we fit the data to the function which takes t as the variable, and the fitted equations are:

$e_1 = 0.1278 + 0.0066t$

$e_2 = -0.0734 + 0.011t$

$e_3 = -0.0177t + 0.94857$.

3.1.2 Explanations of p_{it} and its computation method

p_{it} represents the sex proportion in city, town and village. According to the bearing age and other information embodied in the date, we can confirm values of p_{it} from 2001 to 2005, and fit the development tendency in the middle and short term, and we also fit these changes to the linear function. But up to 2001, the sex proportion has gone to be stable, so we can take it as the constant which is the value fitted on the curve in 2010. We use the tool of Mathematica to fit the data, and the fitted equations are as follows:

$p_{1t}=0.800885+0.0104723t$, (t<23)

$p_{2t} = 1.13934 - 0.0071676t$, (t<23)

$p_{3t} = 1.38723 - 0.0174645t$, (t<23)

When t>23, p_{it} are respectively $p_{1t}(23)$, $p_{2t}(23)$ and $p_{3t}(23)$.

3.1.3 Explanations of $(p_{it}-1)^2$

Because the sex will influence the total population, so we introduce the influencing factor $(p_{it}-1)^2$ which represents the difference among sex proportion in every year, and the sex proportion difference are larger, the negative influence to the total population is more obvious and the sex proportion is more closed to 1, so the influence of the sex proportion to the total population is smaller. We select $(p_{it}-1)^2$ to describe the deviation degree of sex proportion of city, town and village with 1. Table 1 shows values of $(p_{it}-1)^2$ with time from 2001 to 2005.

3.1.4 Explanations of k_i

We use the parameter k_i to denote the association degree of city, town and village with the total population, which is relative with human education degree, concept, living level and other factors. k_i is computed through the statistical data from 2001 to 2005 of China Statistics Bureau and the equation. The rapid change occurred in the data in 2003 because of the large sized SARS and the reason that Chinese thought the Yang year went against the birth. So we select data of 2002, 2003 and 2005 to compute k_i.

The computation method is as follows:

$$128453 = 128513.7977 - 0.262 \times 0.00034203k_1 - 0.126 \times 0.0001206k_2 - 0.613 \times 0.00305k_3$$

$$129988 = 130237.1933 - 0.258 \times 0.001132724k_1 - 0.154 \times 0.000266555k_2 - 0.588 \times 0.001306106k_3$$

$$130756 = 131044.088 - 0.005592994 \times 0.277k_1 - 0.171 \times 0.000448304k_2 - 0.000303747 \times 0.552k_3 \cdot$$

From above equation group, we can obtain

k_1=429433

$k_2=3.10905\times10^0$

k_3=-14585.6.

3.1.5 Confirmations of parameter k, a and b

Because we suppose the China population policy in middle and short term is stable, so its influence to the population growth tendency is stable, for example, the family planning policy and other factors are embodied in the coefficient a and b. The method to compute parameters is as follows:

$$b = \left(\frac{\sum_3 y_t - \sum_2 y_t}{\sum_2 y_t - \sum_1 y_t} \right)^{\frac{1}{n}}$$

$$a = \left(\sum_2 y_t - \sum_1 y_t\right) \frac{b-1}{b(b^n-1)^2}$$

$$K = \frac{1}{n}\left[\sum_1 y_t - ab\frac{(b^n-1)^2}{b-1}\right].$$

So we can obtain

$$b = 7\sqrt{\frac{892585-838497}{838497-765078}} = 0.95729$$

$$a = (838497-765078)\times\frac{0.95729-1}{0.95729\times\left[(0.95729)^2-1\right]^2} = -47246.9$$

$$K = \frac{1}{7}\times\left[765079-(-47264.9)\times 0.95729\times\frac{(0.95729)^7-1}{0.95729-1}\right] = 149129.6.$$

From the equation

$$y_t = \mathrm{K} + ab^t - e_1k_1(p_{1t}-1)^2 - e_2k_2(p_{2t}-1)^2 - e_3k_3(p_{3t}-1)^2,$$

we compute and figure the Figure 1 and Figure 2, and the Figure 1 denotes the tendency of population growth when $t<23$, and the Figure 2 denotes the tendency of population growth when $t>=23$.

3.2 The model of long term

The population model of long term is got based on the modification and perfection to the model of middle and short term which is as follows:

$$y_t = \mathrm{K} + ab^t - \sum_{i=1}^{3} e_i k_i (p_{it}-1)^2.$$

Where, e_i goes to the balance in the long-term process, and according the status of developed country and the development tendency of China, we thought e_i can be approximated as a constant, we take that e_1 is 0.4, e_2 is 0.2 and e_3 is 0.4.

P_{it} is the sex proportion, and we find that it shakes in a certain value through large of data, and we take it as 1.03 in the long-term prediction.

In the long-term prediction, the influence of replacement rate to the population is more obvious. So we add $e^{-0.09t}$ to embody its influence in the equation. We confirm that the value of c is 200000 according to the influence of the aging population.

Therefore, we put forward the long-term model as follows.

$$y_t = \mathrm{K} + ab^t - \sum_{i=1}^{3} e_i k_i (p_{it}-1)^2 - ce^{-0.09t}$$

In the long-term prediction, some influencing factors put up tendency fluctuation, which makes the population growth bring the element of tendency fluctuation, and some factors present periodic fluctuation, which makes the population growth bring the element of periodic fluctuation, and some factors present occasional fluctuation, which make the population growth bring the element of random fluctuation.

4. The solution of the Model

The computation results are seen in Table 3.

The prediction data are seen in Table 4.

We compute the standard deviation S and the mean absolute percent error (MAPE) utilizing the data in above tables and review the prediction effect of the model.

$$S_y = \sqrt{\frac{1}{n-1}\sum_{t=1}^{21} e_t^2} = 0.168083737$$

$$\mathrm{MAPE} = \frac{1}{21}\sum_{t=8}^{22}\frac{|e^t|}{y_t} = 0.000001356$$

References

China National Bureau of Statistics. (2006). *China Statistical Yearbook of 2006.* Beijing: China Statistics Press.

Jiang, Qiyuan. (1993). *Mathematical Model.* Beijing: China Higher Education Press.

Li, Xiaofeng. (2006). The Model of Population Prediction by Modified Index Curve. *Journal of Jingmen Technical College.* No. 6. p. 85-88.

Wang, Moran. (2003). *Matlab and Science Computation.* Beijing: Electric Industry Press.

Xu, Annong. (2004). *Mathematical Experiment.* Beijing: Publishing House of Electrics Industry.

Yang, Qifan. (2005). *Mathematical Modeling.* Beijing: Machine Industry Press.

Table 1. Values of $(p_{it}-1)^2$ with time from 2001 to 2005

$(p_{it}-1)^2$ / t	$(p_{it}-1)^2$	$(p_{it}-1)^2$	$(p_{it}-1)^2$
2001	0.000000106	0.000001021	0.003639929
2002	0.00034203	0.000120592	0.003054233
2003	0.000000111	0.000029677	0.00306567
2004	0.001132474	0.000266555	0.001306106
2005	0.005592994	0.000448304	0.000303747

Table 2. Data from China Statistical Yearbook

year	time	total population	growth rate	first order difference proportion of population	sum of three stages
year	t	y_t	y_t/y_{t-1}	$\Delta y_t/y_{t-1}$	$\sum y_t$
1984	1	104357			765078
1985	2	105851	0.01431624		
1986	3	107507	0.01564463	1.108433735	
1987	4	109300	0.01667798	1.082729469	
1988	5	111026	0.0157914	0.96263246	
1989	6	112704	0.01511358	0.972190035	
1990	7	114333	0.01445379	0.97079857	
1991	8	115823	0.01303211	0.914671578	
1992	9	117171	0.01163845	0.904697987	
1993	10	118517	0.01148748	0.99851632	
1994	11	119850	0.01124733	0.990341753	838497
1995	12	121121	0.01060492	0.953488372	
1996	13	122389	0.01046887	0.997639654	
1997	14	123626	0.01010712	0.97555205	
1998	15	124761	0.00918092	0.917542441	
1999	16	125786	0.00821571	0.9030837	
2000	17	126743	0.00760816	0.933658537	
2001	18	127627	0.00697474	0.923719958	892585
2002	19	128453	0.00647198	0.93438914	
2003	20	129227	0.00602555	0.937046005	
2004	21	129988	0.00588886	0.983204134	

Table 3. Computation results

Year	Actual data	Prediction data	Relative errors (%)
1991	115823	116120	0.25
1992	117171	117306	0.11
1993	118517	118507	0.008
1994	119850	119714	0.11
1995	121121	120919	0.16
1996	122389	122113	0.22
1997	123626	123287	0.27
1998	124761	124431	0.26
1999	125786	125538	0.19
2000	126743	126598	0.11
2001	127627	127600	0.021
2002	128453	128537	0.065
2003	129227	129398	0.13
2004	129988	130175	0.14
2005	130756	130856	0.07

Table 4. Prediction data

Year	Prediction data
2006	132004
2007	132736
2008	133436
2009	134107
2010	134748
2011	135362
2012	135949
2013	136511
2014	137048
2015	137562

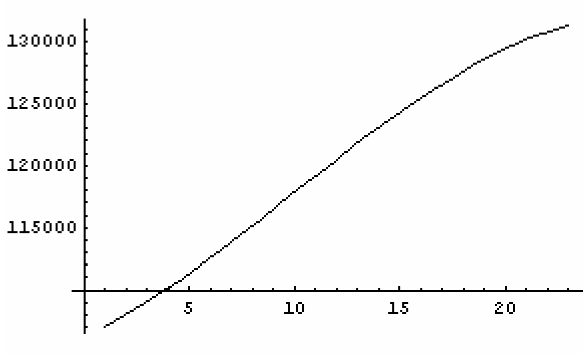

Figure 1. The Tendency of Population Growth When t<23

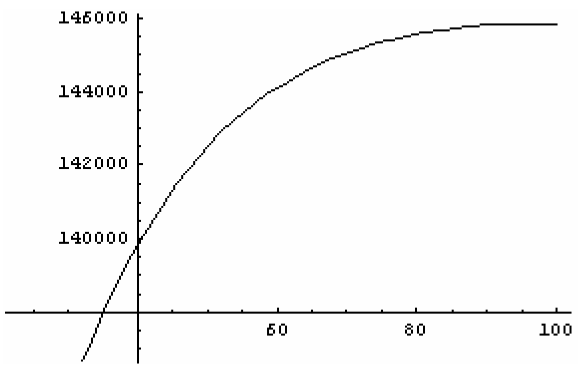

Figure 2. The Tendency of Population Growth When t>=23