Universal Library of Multidisciplinary Research Article

ISSN: 3064-9986 | Volume 1, Issue 2

Open Access | PP: 01-11

DOI: https://doi.org/10.70315/uloap.ulmdi.2024.0102001

A Novel Numerical Modeling Chinese Population Evolution

Liangxin Li[1], Zaishu Cheng[2]

[1]Business School, Chongqing College of Humanities, Science & Technology, Hechuan District, Chongqing 401524, China.
[2]Shenzhen TCL Industrial Research Institute, TCL International E City, Shenzhen 518000, Guangdong province, China.

Abstract

In this paper, we build a very accurate mortality model with a constant time decay function and the optimal goal of minimizing the total number of deaths, then solve it by a numerical approach. China's vast territory, large population, complex flow of people, and the concept of returning to the hometown greatly affect the accuracy of survey population mortality. Furthermore, the instability of the solution of the Lee-Carter mortality model makes it difficult to predict the future population mortality in China. Based on the population data published in China from 1990 to 2000 with corrections, the mortality results from 2001 to 2019 are predicted by using the new model. Compared with the population and deaths in the National Statistical Bulletin, the average death rate error is less than 0.64%, the maximum annual error is less than 6.11%. According to this result, the total population and mortality results from 2001 to 2020 are predicted. Most of the prediction error rates of annual total population is 1 ‰, and the maximum error rate is less than 3.4 ‰. The accuracy of the model for Chinese population prediction is around 10 times higher than any other current models.

Keywords: *Mortality, Lee-Carter, Chinese Population Forecast, Numerical Optimization Method, Time Series*

INTRODUCTION

Mortality modeling and prediction are the basis for population prediction, risk management of life insurance companies, and social pension. The most used and popular model is the Lee-Carter model[1-2], but this model has many imperfect aspects. For a long time, many researchers have been modifying and expanding the Lee-Carter model to form a series of lee-carter models. However, the central idea of these models has not changed which needs long term historic and accurate mortality data. China's historic conditions limits these two requirements, as the history of mortality records is not very long and the mortality rate data of surveys and censuses is not accurate enough.

The assessment of the quality of death data is the first step in the analysis of population deaths. The data mainly comes from the 1% spot check every ten years, every 5 years, and the 1‰ spot check in other years since 1990. The most accurate data is the census. Even so, the fourth, fifth, and sixth census data in 1990, 2000, and 2010 all had various degrees of underreporting of deaths in [3-4]. According to the results of the seventh census, the national population separated from households was 492.76 million. The separated population in the municipal area was 116.94 million, and the floating population was 37.582 million. Among them, the inter-provincial floating population was 124.84 million, accounting for 35%, 8.35%, 26.8%, and 8.91% of the country's total population. More than a quarter of the floating population brings difficulties to death statistics, and the problem of underreporting will exist for a long time in the future. In order to calculate the degree of underreporting of death data, it is necessary to find a data benchmark for comparison. This article uses the data from the National Statistical Bulletin, including total population, death population, 65+ total population. These data meet: 1) raw data for consecutive years, and 2) the smoothness of first-order differential data and the second-order differential data. If these data are expressed as an analytical function with the year as the independent variable, it is multi-order differentiable. The connection curve of the multi-order differentials of the data is smooth. In the bulletin, there is a strict logical relationship between the total population, the number of births and the number of deaths, $P_{t+1}=P_t+B_{t+1}-D_{t+1}$. This article calculates the total number of deaths according to the mortality rate obtained from the census and random checks considering the percentage of the age group in the total population. Furthermore, the total number of deaths is compared with the statistical bulletin.

It is well known that Lee-Carter and its optimization models in [1-6] are widely used to predict mortality, but the accuracy of forecasting mortality for China is affected by short historical data, inaccurate data, and poor logarithmic correlations in most age groups found in this paper. In fact, from 1994 to 2017, China did not hold strong mortality correlations for any age group. The survey age group mortality data from 2005 to 2017 were clearly inconsistent with the Chinese Statistical Bulletin population death data. The earlier papers 'calculated death toll by using Lee-Carter model in China is far from the data by more than 20% error away from the statistical bulletin data which is quite large for a serious research in [7].

The Lee-carter model is ln (m (x, t)) = a (x) + b (x) * k (t) + e (x, t). By taking the logarithm of the mortality rate, the optimal solution is found for the parameters a and b. The same approach is used for the Lee-carter extension models. Although this method is statistically feasible, when it is applied to the mortality rate of a specific country or region, the following problems will be encountered when it is applied to the mortality rate of a specific country or region: 1 . The ill-posedness of the calculation process when solving the model parameters. First, because the mortality rate of a certain age in a certain year may be close to zero, then the rate of change of the mortality of this age in the random check is significant. After the death rate is taken to the logarithm, the parameter solution is severely affected. Second, the death rate of 91-100 is the highest in the elderly group, and the proportion of the population is small, but it has a great influence on the value of the parameter solution. Together, the age group with the fewest deaths and the age group with the largest rate of deaths and smallest proportion of the population have the greatest impact to the value of the parameter solution. Therefore, taking the logarithm of the mortality rate distorts the importance of the mortality rate of each age group. 2.After solving m (x, t)) = exp(a (x) + b (x) * k (t) + e (x, t)) for a and b, the variables a and b do not have the meaning of the dimension. Also, a_i and a_j belong to the parameters in the mortality index of different age groups, and the operation a_i+a_j do not have any mathematical meaning. 3. The data samples in the Lee-carter model are usually using consecutive 5 years as an age group, which avoids the instability of the data due to the small number of age-specific samples in the death rate spot check, but it is difficult to predict the population for the next year, because people in the previous year's age group (usually 5 years a group) are broken up in the next year and need to be regrouped. 4. In the Lee-carter model, the mortality rate of any age group is completely linearly correlated after taking the logarithm, which is proved to be invalid in practice.

A considerable proportion of scholars use the lee-carter model to predict the mortality rate in China. Li Zhisheng et al. in [7] used the Lee-Carter model to fit and predict the mortality rate of the Chinese population. The article used the best fitting model and Bootstrap method performs interval estimation. Wang Xiaojun et al. (2008) in [6] summarized the progress of mortality prediction models. Lu Fangxian et al. (2005) in [8] used the Lee-Carter model to predict the mortality rate of the Chinese population based on China's sex-disaggregated mortality data from 1986 to 2002. He used the data from 1986 to 2002, but among them, the data from 1987 to 1988, 1991 to 1993 and 2000 are missing, and the missing data is not processed in the use. To make Lee-Carter model more suitable for the current population mortality prediction in China, Han Meng et al. (2010) in [9] improved the Lee-Carter model, using a double stochastic process to model the time term in the Lee-Carter model, and taking into account the impact of insufficient sample size on the prediction results. They believed that the weighted least squares method has better fitting effect and prediction effect. Zhu Wei et al. (2009) in [10] used the method of Brouhans et al. (2002) to fit and forecast China's urban population mortality data from 1989 to 2006 in view of the lack of some annual mortality data in China. However, due to the flaws in the theoretical design of the lee-carter model and the low quality of mortality data in China, when the calculated total number of deaths based on the prediction results of these models is compared with the data from the National Bureau of Statistics, the error is large. Actually, the error exceeds 20%.

In this paper, by constructing a single age group mortality model, using numerical optimization methods with a main mortality function decaying over time and applying time series analysis, the 1990 and 2000 national census data and the 1990-2000 National Statistical Bulletin data in [20-22] are used to predict mortality during 2001-2019. The total error for death rate is 0.45%, and the annual average error is 2.36%.

THE MODEL

Symbol Meaning

1). Subscript

t year, such as 1990, 1991, for calculation convenience, set t=0,1,...

x age, 0, 1, 2, ..., 100

q Age group, 0-4, 5-9,...,94-99,100+

g Gender group, 1=male, 0=female

o 65+ years old age group

2). Numerical value

P is total population reported by the National Bureau of Statistics,

P_t represents the total population of year t,

$P_{t,x}$ represents the total population of year t, age x,

$P_{t,x,g=1}$ represents year is t, age is x, male total population, other numerical subscripts are the same

D, the total number of deaths reported by the National Bureau of Statistics

B, the number of births reported by the National Bureau of Statistics

M, the mortality rate of the random check of the population yearbook

P^c, the number of the population of a random checks in the population yearbook

D^c, the total number of deaths in a random check of the population yearbook

D^k, the total number of deaths calculated by the death rate in random check of the population yearbook

P^0, the total population calculated or predicted in this paper

D^0, the total number of deaths calculated or predicted in this paper

B^0, the number of births quoted or predicted in this article

M^0, the mortality rate of calculation or prediction in this paper.

M^i, the mortality rate calculated or predicted by literature i

F, total fertility rate

r, sex ratio at birth

3). function

h(x,y)=(y-x)/x*100%, which is the error rate with %

Methods

1) the model

In this paper, $M^0_{x,t} = m(x, t)$. Defining $m(x, t)$ as the central mortality for year t, age x, where $x= 0,1,2,..., 99,0 < m(x, t) < 1$, suppose

$$\lim_{t\to\infty} m(x,t) = 0$$

$m(x, t)$ is a decreasing function with respect to t, and the degree of decrease is approximately proportional to $m(x, t)$. Let

$$\frac{\partial m(x,t)}{\partial t} = l(x,t) * m(x,t)$$

the solution of this equation can be written as following

$m(x, t) = e^{g(x, t)}$.

Let $g(x, t) = ln(m(x, t)) = c_x + d_x t + \varepsilon(x, t)$,

same as the Lee-Carter model.

$m(x, t) = exp(c_x)\,[exp(d_x)]^{t + \varepsilon(x, t) / dx}$

Let

$a_x = exp(c_x)$, $b_x = exp(d_x)$

be expressed by the time series value $T(x, t)$,

$m(x, t) = a_x * b_x^{t + T(x, t)}$

Then

$T(x, t) = ln(m(x, t) / a_x) / lnb_x - t$

treat $a_x * b_x^t$ as the main function, and $T(x, t)$ as the residual time series value, its significance is that for any main function, ($a_x > 0$, $b_x > 0$, $b_x \neq 1$), it is always possible to make the death rate equal to the given death rate through the translation of time. Regardless of the agreement between the theoretical calculation value and the actual value in the past, the results can always be modified by $T(x, t)$,

$T(x, t) = T1(x, t) + e(x, t)$

$T1(x, t)$ can be adjusted or predicted time series, $e(x, t)$ is the residual time series value. By adjusting the predicted time series value $T1(x, t)$ to adapt to the change of future mortality, assuming that $T1(x, t)$ is approximately linearly expressed as, when $t > t_n$,

$T1(x, t) = c1_x + d1_x(t - t_n)$,

$$m(x,t) = a_x * b_x^{t+T(x,t)} = a_x * b_x^{t+c1_x+d1_x(t-tn)+e(x,t)}$$

$$= \left(a_x * b_x^{c1_x - tn*d1_x}\right) * \left(b_x^{1+d1_x}\right)^{t+\frac{e(x,t)}{1+d1_x}} = a'_x * (b'_x)^{t+e'(x,t)}$$

This is equivalent to readjusting a_x and b_x, which means we can always get as close as possible to the real $m(x, t)$.

In addition to representing $T(x, t)$ with mortality, it can also be expressed as group deaths D^0_t and the population distribution number $P_{x,t}$ of the group which can be obtained by observing.

$$D_t^0 = \sum_x P_{x-1,t-1} a_x b_x^{t+T(x,t)}$$

$$\frac{\partial D_t^0}{\partial t} = \sum_x P_{x-1,t-1} a_x b_x^{t+T(x,t)} ln b_x \approx \sum_x P_{x-1,t-1} a_x b_x^t ln b_x$$

$$D_t^0 \approx \sum_x P_{x-1,t-1} a_x b_x^t + \sum_x T(x,t) P_{x-1,t-1} a_x b_x^t ln b_x$$

Simplifying models and calculations,

Suppose T(x,t) =T(0,t) ,then

$$T(0,t) = \frac{D_t^0 - \sum_x P_{x-1,t-1} a_x b_x^t}{\sum_x P_{x-1,t-1} a_x b_x^t ln b_x} \approx \frac{D_t^0 - \sum_x P^0_{x-1,t-1} a_x b_x^t}{\sum_x P^0_{x-1,t-1} a_x b_x^t ln b_x} \quad (2.1)$$

$T(0, t)$ and the number of deaths can be calculated from the death rate through theses equations.

In special case, let $a_x = 1$, $b_x = e$, $T(x, t) = ln(m(x, t)) - t$, this degenerates into a deformation of the Lee-Carter model.

Consider a completely different scenario, by adjusting the parameters a_x and b_x, the effort is to make $T(x, t)$ series values close to zero. The solution process is as follows:

Take the two years with accurate single-age mortality as the initial time ($t = 0$) and end time ($t = t_n$), $t_n > 9$,(t_n *also written astn*), let

$$m(x,t) = a_x * b_x^t \quad (2.2)$$

For $t = 0, t_n$, then

$$a_x = m(x,0) \quad (2.3)$$

$$m(x,tn) = a_x * b_x^{tn}$$

then

$$b_x = \left(\frac{m(x,tn)}{m(x,0)}\right)^{\frac{1}{tn}} \quad (2.4)$$

With distinguishing between genders, then

$$a_{x,g} = m(x,0,g) \quad (2.3')$$

$$b_{x,g} = \left(\frac{m(x,tn,g)}{m(x,0,g)}\right)^{\frac{1}{tn}} \quad (2.4')$$

Treat the entire population as a group, adjust the multiples α and β of a_x and b_x, and note

$$D_t^0 = \sum_{x=1}^{100} P_{x-1,t-1} * m(x,t) \approx \sum_{x=1}^{100} P^0_{x-1,t-1} * \alpha * a_x (\beta b_x)^t \quad (2.5)$$

With distinguishing between genders, then

$$D_t^0 \approx \sum_{g=0}^{1} \sum_{x=1}^{100} P^0_{x-1,t-1,g} * \alpha * a_{x,g} (\beta b_{x,g})^t \quad (2.5')$$

Equation (2.1) can be written as

$$T(0,t) \approx \frac{D_t^0 - \sum_{g=0}^{1} \sum_x P_{x-1,t-1,g}^0 a_{x,g} b_{x,g}^t}{\sum_{g=0}^{1} \sum_x P_{x-1,t-1,g}^0 a_{x,g} b_{x,g}^t ln b_{x,g}} \quad (2.1')$$

find α and β such that

$$err1 = abs(\sum_{t=1}^{tn}(D_t - D_t^0)/D_t) \quad (2.6)$$

and

$$err2 = \sum_{t=1}^{tn} \frac{(D_t - D_t^0)^2}{(D_t)^2} \quad (2.7)$$

takes the minimum value. Because the analytical solution is not easy to obtain, the numerical solution (initial value α = 1, β = 1) is used as approximate solutions, which can be solved by numiericla methods such as the fastest gradient method, multi-point numerical comparison method, and annealing method.

2) The numerical method

The calculation method used in this paper is to traverse the l and n in α=1+nΔa, β=1+lΔb, where l,n are integers, the step size Δa is set to 0.01, the step size Δb is set to 0.001, -21<n<21, -21< l<21, a total of 1681 sets of err1 and err2 data are generated. When looking for the minimum value of err2 in err1<0.02, l1 and n1 are obtained, and the numerical solution is α=1+n1*0.01, β=1+l1*0.001, and α is obtained with β.

The test results show that the optimized parameters make the annual death rate error rate from 2001 to 2019 less than 6.1%, and the total error rate is less than 0.36% as shown in Table 1. The error rate fluctuates up and down around the 0 axis, and the amplitude tends to increase, but the total error rate and the annual error rate are very small. Although the error rate is expanding after 2013, the number of those rates is small, the error rates are not large, and there is no need to adjust the time series. If it needs an adjustment in the future, for example: from the errors for the years of 2013-2019, by linear regression, we get *T1(x, t)* = 0.7707 * (*t*-2012)-1.0314, then recalculate the total population mortality rate, and find the annual death population error rate based on the calculated mortality rate as shown in Table 1.

Table 1. Comparison of the prediction results of the total number of deaths with the data in the statistical bulletin (unit: ten thousands)

year	D_t	D_t^0	D_t^0-D_t	h(D_t^0- D_t)	*T(0,t)*	*T1(0,t)*	h^1 (D_t^0- D_t)
1991	776	770	-6	-0.77	-0.462	0	-0.77
1992	778	769	-9	-1.16	-0.692	0	-1.16
1993	787	773	-14	-1.78	-1.085	0	-1.78
1994	778	789	11	1.41	0.859	0	1.41
1995	796	790	-6	-0.75	-0.469	0	-0.75
1996	803	800	-3	-0.37	-0.234	0	-0.37
1997	805	811	6	0.75	0.469	0	0.75
1998	811	815	4	0.49	0.313	0	0.49
1999	810	817	7	0.86	0.543	0	0.86
2000	817	826	9	1.10	0.698	0	1.10
2001	818	827	9	1.10	0.698	0	1.10
2002	821	835	14	1.71	1.085	0	1.71
2003	825	842	17	2.06	1.308	0	2.06
2004	832	852	20	2.40	1.527	0	2.40
2005	849	864	15	1.77	1.145	0	1.77
2006	892	879	-13	-1.46	-0.985	0	-1.46
2007	913	886	-27	-2.96	-2.030	0	-2.96
2008	935	903	-32	-3.42	-2.388	0	-3.42
2009	943	911	-32	-3.39	-2.370	0	-3.39
2010	953	922	-31	-3.25	-2.279	0	-3.25
2011	960	941	-19	-1.98	-1.387	0	-1.98
2012	966	950	-16	-1.66	-1.159	0	-1.66
2013	972	965	-7	-0.72	-0.504	-0.261	-0.41
2014	977	981	4	0.41	0.284	0.509	-0.31
2015	975	998	23	2.36	1.620	1.279	0.51
2016	977	1011	34	3.48	2.361	2.049	0.61
2017	986	1030	44	4.46	3.034	2.819	0.61
2018	993	1044	51	5.14	3.469	3.589	0.40
2019	998	1059	61	6.11	4.094	4.359	0.50

In Figure 4a, b, we construct the logarithm mortality with time curve through the historical data and forecasting data. We find that they are a set of straight line with negative slope which means the falling rate of mortality with time is constant. Figure 1a is demonstrating the model and data results for the years from 1994 to 2017 which we have data to compare the truth with the model results. Figure 1b, we show the predictions for the years from 2018 to 2098 which we do not have any data to make the comparison. The slopes are the keys to determine how fast the mortality will be halved. It seems the age groups for ages 15-19 and 60-64 are getting faster in lowering mortality than other age groups. It is a very interesting result for our research. Due to the extreme accuracy of our methods, we conclude this result as a law in the following discussions.

DATA ANALYSIS

Correlation Analysis of Mortality in Each Age Group

The lee-carter model is expressed as

$ln\ (m\ (x, t)) = a\ (x) + b\ (x) * k\ (t) + e\ (x, t).$

For any $i, j, i \neq j$, $ln\ (m\ (x_i, t)) = a\ (x_i) + b\ (x_i) * k\ (t) + e\ (x_i, t)$,

Get

$ln\ (m\ (x_j, t)) = a\ (x_j)\text{-}b\ (x_j) / b\ (x_i) * (a\ (x_i) + e\ (x_i, t)) + e\ (x_j, t) + b\ (x_j) / b\ (x_i) * ln\ (m\ (x_i, t))$

It is found that $ln\ (m\ (x_i, t))$ can be linearly expressed by $ln\ (m\ (x_j, t))$, and $ln\ (m\ (x_i, t))$ is completely linearly related to $ln\ (m\ (x_j, t))$.

The correlation coefficients of different age groups' mortality is verified. The original mortality data is derived from the age group mortality data, death population and exposed population data in the 1994-2017 China Demographic Yearbook in [20-22], including 0 to 4 years, 5 to 9 years ... ... 85-89 years old with total 18 age groups, the mortality correlation coefficients between each age group is calculated. The calculation indicates that the 0-4 age group is closely related to the 20-49 age group and the 60-74 age group. The reason may be due to the fact that young children often live together with their parents, grandparents, and therefore mortality is much related. The 5-9,10-14,15-19 age group mortality rate is hardly related to other age groups, because they are in the growth and learning period. They are in primary school, junior high school, high school or technical school or employment, with different characteristics. The age groups of 20-49 years old as the main force of social employment are in the strong correlation among these age groups, 35-49 age groups is strongly correlated with 65-79 age groups, but 50-54 age group is special, and the correlation with other groups is weak. The group of 55-84 years old with retirement age are related to each age group. The 85-89 age group is relatively independent and usually their children have retired.

The correlation table indicates that the mortality rate of all age groups in the Chinese population does not meet the strong correlation requirements in the lee-carter model, and these maybe the reason the lee-carter model accuracy for Chinese mortality rate is not high.

Analysis of the Contribution Ratio of Deaths in Each Age Group

In the Lee-carter model, the high volatility of the death rate of the young age group over time and the large death rate of the elderly group enlarge their effects to the solution parameters proportionally. In fact, the number of deaths in these two groups is small. Taking the recent more accurate mortality rate in the 2014 spot check as an example, calculations of the proportion of deaths in each age group based on the spot check mortality rate and the proportion of the population are shown in Table 1.

Table 1. Contribution ratio of deaths by age group in 2014

age	year	Proportion of the population	Random check death rate	Death rate* Proportion	Percentage of death
0~4	2014	5.69	1.23	7.00	1.13
5~9	2014	5.61	0.24	1.35	0.22
10~14	2014	5.18	0.28	1.45	0.23
15~19	2014	5.76	0.40	2.30	0.37
20~24	2014	8.07	0.40	3.23	0.52
25~29	2014	8.79	0.35	3.08	0.50
30~34	2014	7.34	0.57	4.18	0.68
35~39	2014	7.27	1.04	7.56	1.22
40~44	2014	9.07	1.50	13.61	2.20
45~49	2014	8.83	2.28	20.13	3.26
50~54	2014	6.93	4.17	28.90	4.67
55~59	2014	5.91	6.87	40.60	6.57
60~64	2014	5.48	10.45	57.27	9.26
65~69	2014	3.71	17.62	65.37	10.57
70~74	2014	2.59	27.90	72.26	11.69

75~79	2014	1.9	49.08	93.25	15.08
80~84	2014	1.18	77.77	91.77	14.84
85~89	2014	0.5	129.89	64.95	10.50
90~94	2014	0.19	211.12	40.11	6.49

Table 1 shows that the age groups 5-9 and 10-14 have the lowest mortality rate, which is prone to large fluctuations in the number of deaths, but the proportion of deaths is only about 0.22%, while the mortality rate of the 90-94 age group is 30 times that of the 60 to 64 groups. In the numerical solution of the lee-carter model, the weight is 3 times (ln30) of the group of 60-64, but the 90-94 group death toll accounts only for 70% of the group 60 to 64. Therefore, when the lee-carter model is solved for predicting the number of deaths, the elderly group will be distorted because the proportion of the population of the elderly group in demography is ignored.

Differences in the Statistical Bulletin

The total population, death population, birth population, 65+ population in the national annual statistical bulletin reflect the full picture of our country's population. The accuracy of basic data has a wide impact, is authoritative, and is the most reliable source of data. This article takes the continuity analysis of the first-order difference and the second-order difference of the data, and lists first-order difference and second-order difference of the total population, death population, birth population, 65+ population data from 1991 to 2019 as follows

Table 2. the first-order and second-order difference of population data in the Statistical Bulletin

year	P	B	P_o	D	ΔD	Δ(ΔD)	ΔP	Δ(ΔP)	ΔB	Δ(ΔB)	ΔP_o	$\Delta(\Delta P_o)$
1991	115823	2279		776								
1992	117171	2137		778	2		1348		−142			
1993	118517	2144		787	9	7	1346	−2	7	149		
1994	119850	2121		778	−9	−18	1333	−13	−23	−30		
1995	121121	2074		796	18	27	1271	−62	−47	−24		
1996	122389	2078	7833	803	7	−11	1268	−3	4	51		
1997	123626	2048		805	2	−5	1237	−31	−30	−34		
1998	124761	1951	8375	811	6	4	1135	−102	−97	−67		
1999	125786	1909	8688	810	−1	−7	1025	−110	−42	55		
2000	126743	1778		817	7	8	957	−68	−131	−89		
2001	127627	1702	9062	818	1	−6	884	−73	−76	55		
2002	128453	1647	9377	821	3	2	826	−58	−55	21	315	
2003	129227	1599	9692	825	4	1	774	−52	−48	7	315	0
2004	130000	1593	9857	832	7	3	773	−1	−6	42	165	−150
2005	130756	1617	10055	849	17	10	756	−17	24	30	198	33
2006	131448	1584	10419	892	43	26	692	−64	−33	−57	364	166
2007	132129	1594	10636	913	21	−22	681	−11	10	43	217	−147
2008	132802	1608	10956	935	22	1	673	−8	14	4	320	103
2009	133450	1615	11309	943	8	−14	648	−25	7	−7	353	33
2010	134091	1596	11892	953	10	2	641	−7	−19	−26	583	230
2011	134735	1604	12288	960	7	−3	644	3	8	27	396	−187
2012	135404	1635	12714	966	6	−1	669	25	31	23	426	30
2013	136072	1640	13161	972	6	0	668	−1	5	−26	447	21
2014	136782	1687	13755	977	5	−1	710	42	47	42	594	147
2015	137462	1655	14386	975	−2	−7	680	−30	−32	−79	631	37
2016	138271	1786	15003	977	2	4	809	129	131	163	617	−14
2017	139008	1723	15831	986	9	7	737	−72	−63	−194	828	211
2018	139538	1523	16658	993	7	−2	530	−207	−200	−137	827	−1
2019	140005	1465		998	5	−2	467	−63	−58	142		

Table 2 shows that the ratio of the first-order difference and the second-order difference of the number of deaths to the total number of deaths shows a high degree of stability. The first-order difference of the total population gradually decreases from high to low, which expresses the gradual decrease of the total population increment in China, reflecting the effectiveness of the population reduction policy in the early stage of family planning. The second-order differential data of the total population shows the magnitude of the population increment change, which is almost all negative. On January 1, 2016, the two-child policy was fully implemented, and the birth population was with an increase of 1.31 million for the first time within 20 years. The ratio of the second-order difference to the first-order difference of the total population is almost very small for all cases showing the reliability of the total population data. The first-order difference in the birth population was significantly negative in 2000 and 2018, and the second-order difference in 2017 and 2018 was very negative which means the full implementation of the two-child policy has a short duration. The first-order difference of 65+ population has been calculated since 2003. Except for a few years, it

has exceeded 3 million. After 2014, it exceeded 6 million. In 2017, the growth rate shifted to more than 8 million. The second-order difference data is large and sometimes small, showing the instability of the 65+ population growth rate, reflecting that the quality of the 65+ population data is lower than the quality of the total population and the number of deaths data.

Error Analysis of the Number of Deaths Calculated from the Mortality of Random Surveys and Censuses

Every year, the National Bureau of Statistics conducts random checks on the mortality rate according to age and gender group. The calculated total number of deaths in the country according to the results of the random check D_t^k is different from the data D_t^c in the statistical bulletin. The error rate is different in different years. Write D^k_t as

$$D_t^k = D_t^c * \frac{P_t}{P_t^c}$$

The 65+ number of annual deaths of the population is written as

$$D_{t,o}^k = D_{t,o}^c * \frac{P_{t,o}}{P_{t,o}^c}$$

The 65+ raw death rate of the population is written as

$$M_{t,o}^k = 1000 * \frac{D_{t,o}^c}{P_{t,o}} * \frac{P_{t,o}}{P_{t,o}^c}$$

The results of 1994-2017 as in Table 3.

Table 3. Error analysis of the number of deaths calculated from the mortality of random surveys and censuses

year	P	D	P_o	P^c*1000	D^c*1000	D^k	h(D^k,D)	D_o^k	D_o^k/D^k	$M_{t,o}^k$
1994	119850	778		749048.5	4840	774	-0.46	423	0.546	
1995	121121	796		12352691	79622	781	-1.92	423	0.542	
1996	122389	803	7833	1241700	8166	805	0.24	384	0.477	49.02
1997	123626	805		1234566	7842	785	-2.45	435	0.554	
1998	124761	811	8375	1235373	7736	781	-3.67	444	0.568	53.01
1999	125786	810	8688	1204472	7420	775	-4.33	446	0.576	51.34
2000	126743	817		7313081	43293	750	-8.16	487	0.649	
2001	127627	818	9062	1213758	7135	750	-8.28	443	0.590	48.89
2002	128453	821	9377	1252230	7787	799	-2.71	474	0.593	50.55
2003	129227	825	9692	1253923	7581	781	-5.30	462	0.591	47.67
2004	130000	832	9857	1246413	7355	767	-7.80	442	0.576	44.84
2005	130756	849	10055	16950030	101739	785	-7.56	434	0.553	43.16
2006	131448	892	10419	1190388	6385	705	-20.96	356	0.505	34.17
2007	132129	913	10636	1186609	6696	746	-18.34	387	0.519	36.39
2008	132802	935	10956	1176458	6759	763	-18.40	397	0.520	36.24
2009	133450	943	11309	1162782	6302	723	-23.30	367	0.507	32.45
2010	134091	953	11892	1330613984	7421990	748	-21.52	452	0.604	38.01
2011	134735	960	12288	1143063	6670	786	-18.10	434	0.552	35.32
2012	135404	966	12714	1122077	6606	797	-17.48	458	0.575	36.02
2013	136072	972	13161	1116709	6545	798	-17.95	467	0.586	35.48
2014	136782	977	13755	1122265	6791	828	-15.28	457	0.552	33.22
2015	137462	975	14386	21281109	102913	665	-31.82	358	0.539	24.89
2016	138271	977	15003	1155237	6179	740	-24.30	388	0.525	25.86
2017	139008	986	15831	1141889	5807	707	-28.30	361	0.511	22.80

Table 3 shows that the total number of deaths underreporting rate h(D_k,D)calculated based on the random check data and the statistical bulletin increased gradually with the increase of the year, and showed obvious time interval characteristics. The underreporting rate during 1994-2005 is less than 8.2%, but since 2006, it suddenly increased to 20%, and then hovered around 20% to 2014. After 2015, the underreporting rate rose sharply to 31%, and has maintained a high underreporting rate ever since. The 65+ total death toll ratio D_o^k/D^k calculated based on the spot check data was significantly different in 1996, 2000, and 2010 from around years. The 65+ death rate calculated based on the spot check $M_{t,o}^k$ have been drastic changes in the years between 2005-2006, 2009-2010,and 2014-2015. The results showed that 2005 was a watershed. If the statistical bulletin is used as the standard, the random check mortality data after 2005 must be revised significantly before it can be used.

Results Test of Application of Lee-Carter Model to Predict Mortality

Mortality prediction plays an important role in population prediction, social security and insurance business. Many papers use the lee-carter model to predict the future population mortality rate in China. This article tests two

early papers' mortality data results with later data with labels as literature 1 [6] and literature 2 [7].The total number of deaths calculated based on the predicted data of the literatures is written as:

$$D_t^i = \frac{P_t}{P_t^c} * \sum_q P_{t,q}^c * M_{t,q}^i$$

The under-reporting rates of D^1_{20}=783,(2010), D^1_{25}=827(2015) and D_{20}=953,D_{25}=975 were -17.8 and -15.2% respectively. Compared with the death data based on random inspections, the missed reports were greatly corrected. The literature 1 uses very accurate random and census mortality data from 1981, 1986, 1989, 1995, 2000, and 2005. Nevertheless, the total number of deaths calculated using the mortality rate predicted by the Lee-carter model is with more than 15% of the under-reporting rate compared with the bulletin results. Literature 2 predicted the mortality rate from 2008 to 2015, and the results are shown in Table 4.

Table 4. Literature 2 total number of deaths calculated by mortality rate from 2008 to 2015

year	P	D	D^k	D^2	$h(D^2,D)$	$h(D^2,D^k)$
2008	132802	935	763	733	-21.58	-3.90
2009	133450	943	723	722	-23.40	-0.12
2010	134091	953	748	667	-30.00	-10.80
2011	134735	960	786	657	-31.55	-16.42
2012	135404	966	797	651	-32.61	-18.33
2013	136072	972	798	650	-33.16	-18.54
2014	136782	977	828	648	-33.66	-21.69
2015	137462	975	665	638	-34.52	-3.96

The total number of deaths calculated according to the Lee-carter model forecast is obviously much smaller. Compared with the statistical bulletin data, almost all of them exceed 30% less and compared with the total number of deaths calculated by random checks, the number of underreporting is aggravated with the number of underreporting exceeding 10%. Two results show that, based on random inspection of mortality data, the results of using the lee-carter model to predict mortality are not very ideal

Test of 65+ Death Toll Result

Among the deaths, 65+ accounted for the main cause, and many papers pointed out that 60+ underreported seriously in [1-2]. Using three indicators 65+ crude death rate, the ratio of death number to D in the bulletin, and the error between the number of spot check deaths D^k and D^0_o calculated in this paper for analysis, the results are shown in Table 5.

Table 5. 65+ death toll result

year	D	P_o	D^k	D_o^0	D_o^k	D_o^k/D	$M_{t,o}^k$	D_o^0/D	D_o^0/P_o*1000	$h(D_o^k,D_o^0)$
1994	778		774	424	423	0.544		0.545		-0.24
1995	796		781	432	423	0.531		0.543		-2.08
1996	803	7833	805	442	384	0.478	49.02	0.550	56.43	-13.12
1997	805		785	451	435	0.540		0.560		-3.55
1998	811	8375	781	461	444	0.547	53.01	0.568	55.04	-3.69
1999	810	8688	775	472	446	0.551	51.34	0.583	54.33	-5.51
2000	817		750	483	487	0.596		0.591		0.83
2001	818	9062	750	494	443	0.542	48.89	0.604	54.51	-10.32
2002	821	9377	799	506	474	0.577	50.55	0.616	53.96	-6.32
2003	825	9692	781	517	462	0.560	47.67	0.627	53.34	-10.64
2004	832	9857	767	529	442	0.531	44.84	0.636	53.67	-16.45
2005	849	10055	785	540	434	0.511	43.16	0.636	53.70	-19.63
2006	892	10419	705	551	356	0.399	34.17	0.618	52.88	-35.39
2007	913	10636	746	562	387	0.424	36.39	0.616	52.84	-31.14
2008	935	10956	763	574	397	0.425	36.24	0.614	52.39	-30.84
2009	943	11309	723	584	367	0.389	32.45	0.619	51.64	-37.16
2010	953	11892	748	596	452	0.474	38.01	0.625	50.12	-24.16
2011	960	12288	786	608	434	0.452	35.32	0.633	49.48	-28.62
2012	966	12714	797	620	458	0.474	36.02	0.642	48.77	-26.13

2013	972	13161	798	632	467	0.480	35.48	0.650	48.02	-26.11
2014	977	13755	828	646	457	0.468	33.22	0.661	46.96	-29.26
2015	975	14386	665	660	358	0.367	24.89	0.677	45.88	-45.76
2016	977	15003	740	677	388	0.397	25.86	0.693	45.12	-42.69
2017	986	15831	707	693	361	0.366	22.80	0.703	43.77	-47.91

Table 5 shows that there is a significant difference between the D^k calculated based on the spot check and the D^0_o calculated in this paper. The difference is not obvious before 1996 and 2000, since then it become significant. In terms of the crude mortality rate, the spot check results fluctuate significantly. It has been 53‰ in 1998 but by 2017, it dropped to 22.8‰, which is unbelievable. The result of this article is that it has slowly dropped from 55‰ in 1998 to 43.77‰ in 2017. In terms of the proportion of deaths, the result of random check calculation was higher than 50% before 2005, but after 2005, there was a sudden big change to around 40%. However, the result of this article is a gentle increase from 54.5% in 1994 to 70.3% in 2017. The error between D^k and D^0_o was 23% on average during 2001-2013, but during 2014-2017, it reached 45%.

POPULATION FORECAST

China's population accounts for nearly one-fifth of the world. It is already overpopulated for China. The accurate population forecast is very important for government policy making in the world. From 1980s, China's efforts are to promote birth control to reduce the total fertility rate in order to feed a large population with China's resources, especially food. Therefore, China's future population is affected by factors such as economy, science and technology, and government birth control policy. It is not easy to use statistical methods to accurately predict Chinese future population in [11-14]. The U.S. Census Bureau website in[32] fits historical data based on existing Chinese population historical data, with the largest population error exceeding 10 million and the maximum cumulative death error exceeding 4 million.

The population forecast is determined by four factors include basic data, fertility, mortality and international net migration. The 1990 census data was later tested to be the most accurate in [23-27] in the census. This data was corrected as the basic data. The birth population in the statistical bulletin was used as the fertility data to deduce the total fertility rate. According to the data from 1997 to 2014, the number of Chinese students studying abroad minus returning population was 1.7 million. Therefore in this article, the net outflow of population is calculated at 100,000 per year since 1997. By assuming that after 2000, the total fertility rate will remain 1.6approximately adjusted according to 2016 family planning policy, we forecast the future population based on data before 2000, 2001-2019 as a test, according to multiple documents in [28-30].

$$P^0_{x+1,t+1,g} = P^0_{x,t,g} * \left(1 - 0.5 * M^0_{x,t+1,g} - 0.5 * M^0_{x+1,t+1,g}\right) + MGR^0_{x+1,t+1,g}$$

$$P^0_{0,t,g} = \gamma \sum_{x=15}^{49} P^0_{x,t,0} * ATFR_{x,t})$$

When g = 0, γ = 1 / 2.18, when g = 1, γ = 1.18 / 2.18, *ATFR (x, t)* is the fertility rate from the 2000 fertility rate data. $MGR^0_{x,t,g}$ as international net migration.

Figure 1 shows that the error rate between the total population predicted in this article and the total population reported by the Bureau of Statistics is less than 1‰ during the 17 years from 2001 to 2017. In the years of 2018-2020, due to the family planning policy, the number of births is more than expected. As a result, the predicted number of deaths in 2020 and the difference between the census data have led to an increase in the number of births predicted in the years of 2018-2019. The predicted number of deaths in 2020 is higher than the number in the census. The predicted total population has been "corrected". Due to the over optimistic estimation of the impact of the family planning policy on the birth population, the birth population is probably over 2 million more than the actual population since 2019. The peak of the predicted total population is 1.44 billion in 2027, which may be over 20 million more than the true value in the future. During the 30 years from 1991 to 2020, the median age, the average life expectancy of men, and the average life expectancy of women have changed between 24.5-37.3, 68.7-73.6, and 72-78.1, respectively. From 2020 to 2030, their future change ranges are predicted between 37.3- 41.5, 73.6-75, 78.1-79.5,

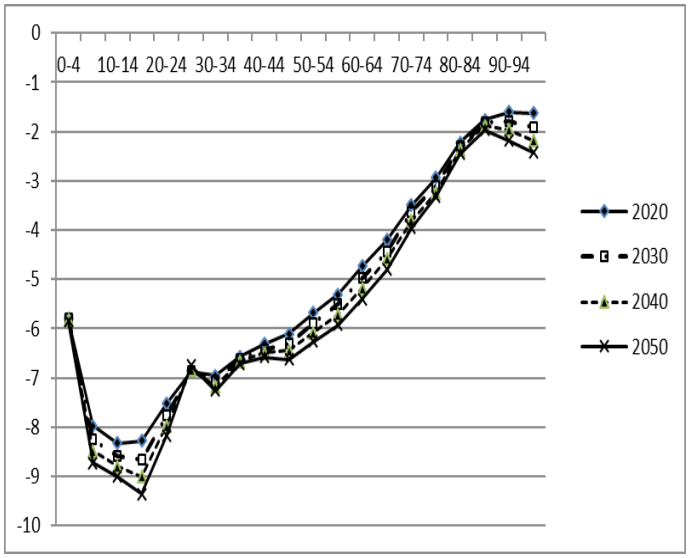

Figure 1. Future logarithm mortality of age group by 5.

In Figure 2, the total population appears around 2025, with more than 1.41 billion, and the number of births has been declining. It has declined rapidly between 2015-2030 and 2040-2055, and the number of deaths has continued to rise to the highest peak around 2055. The median age is not optimistic, going from 26 years old in 1995 to 49 years old in 2045.

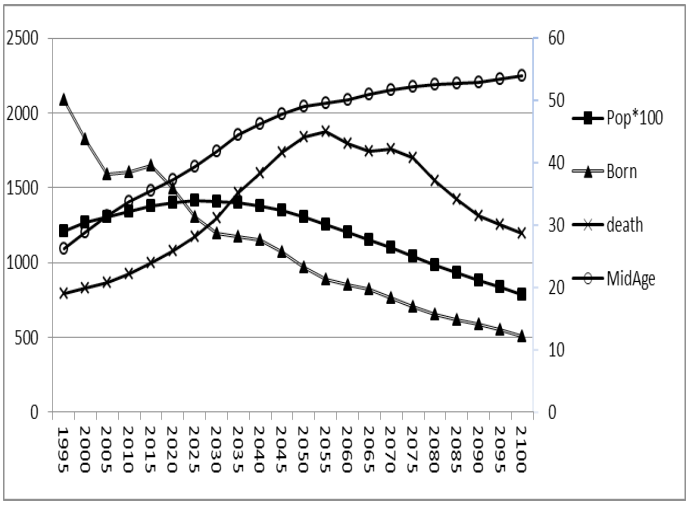

Figure 2. The total population forecast with time (unit: ten thousands; years old)

Figure 3 shows that after the dependency ratio fell rapidly from 0.26 in 2000 to 2010 to 0.19, it slowly falls to 0.13 by the year 2035, and then keeps stable. After the dependency ratio increased slightly from 0.06 to 0.10 from 1995 to 2015, it will increase to 0.32 in the year 2060 almost at an annual rate of 0.02. This shows that China is under increasing pressure from the elderly.

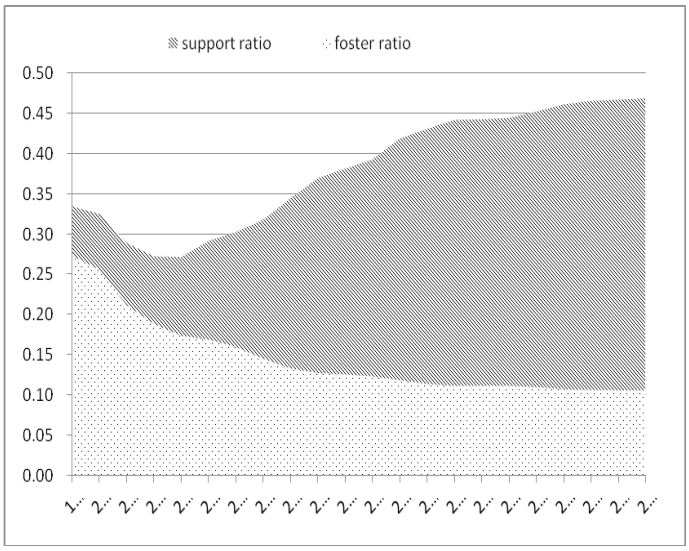

Figure 3. the foster ratio and support ratio

Although there are many obstacles in the future forecast of the Chinese population affected by fertility and mortality errors, it is possible to increase the accuracy by establishing a data model and using time series analysis. Test results show that the uncertainty can be greatly reduced.

CONCLUDING REMARKS

Since our results show that the simple single age model with a lineal decrease of logarithm mortality rate provide a remarkable enhancement of the accuracy for forecasting the future population which is about 10 times more precise than the Lee-Carter model results, we propose the following law as an inference of our dramatic progress of the modeling as the key conclusion of this

$$\ln(m) = a + bt \Rightarrow \frac{\partial \ln(m)}{\partial t} = b$$

$$\Rightarrow \frac{\nabla m / \nabla t}{m} = b$$

i.e., the decrease speed of percentage change of the mortality is a constant due to b negative which is our following law.

Law: The world overall decrease rate of the mortality is in a constant speed. Every 52-65 years, the total mortality rate is halved from Chinese results.

Although the age structure has little influence on the mortality halving rate, it is all in the range of 52.4 to 65.8, with a typical average 60.

It is similar to the Moore's law for the CPU technology, so we make a brave suggestion that it is the medical technology that makes people's life longer, not anything else. For all technology, the acceleration rate is constant just like the gravity.

As the conclusion, our model is better than all present models in accuracy by a huge advancement and our law for the mortality is the direct results of our model. The higher accuracy will provide a key important value to the government policy of the The world and generate fundamental change for the humankind future.

REFERENCES

1. Lee, RD and Carter, LR. Modeling and Forecasting US Mortality[J]. Journal of the American Statistical Association , 1992, 419: 659 -671
2. Li, N. , Lee, R. , and Tuljapurkar, S. Using the Lee-Carter method to forecast mortality for populations with limited data.[J]. International Statistical Review, 2004 (72):19-36.
3. Robert D. Retherford, Minja Kim Choe, Chen Jiajian, Li Xiru, Cui Hongyan. Fertility in China: How Much Has It Really Declined?[J]. Population Research, 2004, (28): 3-15
4. Wang Jinying. Trends in Life Expectancies and Mortality Patterns in China since 1990[J]. Population Research,2013,37(4): 3-18
5. Zhang Wenjuan,Wei Meng. The Evaluation of the Mortality and Life Expectancy of Chinese Population[J]. Population Journal, 2013,(38): 18-28
6. Hong Li, Johnny S.H. Li. Optimizing the Lee-Carter Approach in the presence of structural changes in the time- and age-patterns of mortality improvements[J]. Demography, 2017, (54): 1073-1095
7. Wang Xiaojun, Ren Wendong. Application of Lee-Carter Method in Forecasting the Mortality of Chinese Population with limited data[J]. Statistical Research, 2012,(29): 87-94
8. Li Zhisheng, Liu Hengjia. Estimation and Application of the Lee-Carter Model: Based on Demographic Data

of China[J]. Chinese Journal of Population Science, 2010,(3): 46-56

9. Lu Fangxian, Yin Sha. Application of Lee-Carter Method in Predicting China's Population Mortality[J]. Journal of Insurance Vocational College, 2005, (6): 9-11
10. Han Meng, Wang Xiaojun. Application and Improvement of Lee-Carter Model in the Prediction of China's Urban Population Mortality [J]. Insurance Studies,2010. 10: 3-9.
11. Zhu Wei, Chen Bingzheng. Prediction of the mortality rate of China's urban population[J]. Mathematical Statistics and Management. 2009 4:736-744.
12. Zhu Xingzao, Pang Feiyu. Application of autoregressive logistic discrete model in Chinese population forecasting [J]. Statistics and Decision, 2009,289: 157-159
13. JIANG Ruo-fan, JIANG Yu-mei, LI Fei-ya. Study and Application of Population Forecast Model Based on Grey system and PSO-BP Neural Network[J]. Nothwest Population, 2011,32: 23-26
14. Jiang Yuanying. Population Forecast Based on Age-shift Algorithm[J]. Statistics and Decision, 2012,361: 82-84
15. Chen Wei. China's Future Population Development Trend: 2005-2050 [J]. Population Research, 2006, 30:16-25
16. Tu Xiongling, Xu Haiyun. ARIMA and exponential smoothing for comparative research in China's population forecast. Statistics and Decision, 2009,(292): 21-23
17. Furong Li. Application of Improved Dynamic GM (1,1) Model in Population Forecasting[J]. Statistics and Decision, 2013, (391): 72-74
18. Yu Li, Yang Lintao. Chinese Population Development Forecast Based on Bilinear Model[J]. Statistics and Decision, 2014, (418): 90-92
19. Xi Wei, Yu Xueting. Forecast and Analysis of Population Age Structure in China[J]. Statistics and Decision, 2013,(423): 112-116
20. Duan Kefeng. Chinese Population Forecast Model Based on a Compound Model[J]. Statistics and Decision, 2012, (368): 30-32
21. National Bureau of Statistics. Department of Population and Employment Statistics China Population2004[M]. Beijing: China Statistics Press ,2005.
22. Department of Population and Employment Statistics and National Bureau of Statistics[M]. China Demographic Yearbook 1995-2012. Beijing: China Statistics Press, 1995-2012
23. Department of Population and Employment Statistics and National Bureau of Statistics[M]. China Population and Employment Statistics Yearbook 2013-2017. Beijing: China Statistics Press, 2013-2017
24. Cui Hongyan, Xu Lan, Li Rui. Estimation of the Accuracy of the 2010 Census Data[J]. Population Research, 2013,(37): 10-21
25. Yu Xuejun. Estimation of Total Amount and Structure in the Fifth National Census Data[J]. Population Research, 2002,(26): 9-15
26. Zhang Weimin, Cui Hongyan. Estimation of the accuracy of China's 2000 census[J]. Population Research, 2003,(27): 25-35
27. Cui Hongyan. On the Total Population of China[J]. Population Research, 2000,(24): 1-4
28. Cheng Xi. Evaluation of the Data Quality of Employed Population in the 4th Population Census[J]. Statistical Research, 1993,(10): 41-44
29. Xia Leping. Chinese Fertility Trends 1979 ~ 2000: A Comparative Analysis of Birth Numbers and School Data[J]. Population Research, 2005,(129): 2-15
30. Zhang Lipin, Wang Guangzhou. A Research on the Second Childbirth Expectation and the Birth Plan for the Fertility Age Population of Chinese[J]. Population and Economics, 2015,(213): 21-27
31. http://www.census.gov/population/international/data/idb/informationGateway.php
32. http://www.stats.gov.cn/tjsj/zxfb/201602/t20160229_1323991.html
33. http://www.stats.gov.cn/tjsj/zxfb/201701/t20170120_1455942.html
34. http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm
35. http://www.stats.gov.cn/tjsj/pcsj/rkpc/5rp/index.htm
36. http://www.stats.gov.cn/tjsj/ndsj/2013/indexch.htm
37. http://www.who.int/countries/chn/zh/
38. http://www.eol.cn/html/lx/2014baogao/content.html

Citation: Liangxin Li, Zaishu Cheng, "A Novel Numerical Modeling Chinese Population Evolution", Universal Library of Multidisciplinary, 2024; 1(2): 01-11. DOI: https://doi.org/10.70315/uloap.ulmdi.2024.0102001.