DOI: 10.18686/ahe.v6i15.5142

The Improved Leslie Model for Population Forecasting

Jinjing Ma*, Yongkang Peng*, Lianyu Wu

Jingdezhen Ceramic University, Jingdezhen 333403, Jiangxi China

Abstract: Based on China's population data from 1953 to 2020, the Leslie model combines the fertility rate of women of childbearing age by region and age, the sex ratio of the birth population, the mortality rate, the migration rate between urban and rural areas by age, the curve fitting migration function, and the application of ARIMA to predict mortality rates to construct a discrete population dynamics system in order to predict China's future population development trajectory. The improved Leslie, Leslie, BP and Malthus models were compared in terms of error rates. The improved Leslie model was more stable than the rest of the models and had an average error rate of 0.09%, with good model generalization ability. The results show that the improved Leslie model predicts that the total population will slowly increase under the national regulation policy, and will reach a peak by around 2045 and then decline.

Keywords: Leslie; Population Prediction; ARIMA; Mortality Rate; Mortality Rate

Fund project: This work was partly supported by the key scientific research project of the Educational Commission of Jiangxi Province, China (Grant NO. GJJ201302, GJJ201340), and partly supported by the Science and Technology Program of Jingdezhen City, China (Grant NO.20202GYZD013-01).

1. Intrlduction

China is a country with a large population and the population issue has always been one of the key factors limiting the country's development. Therefore, in order to grasp data on the changes in the existing population, gender ratio and urban and rural population in different regions, the country has conducted a national population census every ten years since the founding of New China. However, in recent years, China's population has gradually shown an ageing population, an imbalance in the sex ratio between men and women and an urbanisation of the rural population. To solve this problem, the existing population prediction models mainly include curve fitting method, BP neural network algorithm, Malthus, Leslie model and so on. BP neural networks are a local search optimization method, and even though they can solve complex non-linear problems, they can lead to network training failure because they can trap the algorithm in local extremes. The Malthus model [1] is an exponential growth model that fits the characteristics of population growth, but ignores external conditions such as limited resources and is only suitable for projections in the short term when there is no population turnaround.

Leslie model are based on age structure, Chen Shujun et al [2], Wang Yifang [3] and Ni Xuanming et al [4] revised the Leslie model by using regression analysis to correct for fertility and mortality, normal distribution to correct for fertility, using constant fertility, and mortality predicted by the Lee-Carter model to improve the Leslie model. All of the above models are only partially corrected for the influence factors of the Leslie model, and do not consider combining the main dynamic factors with large deviations for accurate analysis and prediction. To solve this problem, we combined factors such as fertility rate of women of reproductive age by region and age, sex ratio of births, mortality rate and migration rate by age, curve-fitted migration function and used ARIMA to predict mortality rate [5], thus improving the Leslie model. The results showed that the improved Leslie model was more stable than the rest of the models, with an average error rate of 0.09% and a maximum error rate of 4.26% higher than that proposed by Hou Ruihuan et al

2 .The improved Leslie model

2.1 Model building

We assume that fertility rates are considered only for women aged 15-49, and zero for those not in this range; that only population change from rural to urban migration is considered; and that the effect of factors such as dinks and unexpected disasters on population change is not considered.

Step1: Divide the predicted time period, i.e., let time $t = 1,2,3,...$, and at the same time, divide the age into m segments by considering the elderly aged 90 or above as a whole. Let the total number of people in the ith age group at time t be $n_i(t), i = 1,2,3,......,m$. From this, the total population vector $n(t) = [n_1(t), n_2(t), ...n_m(t)]^T$.

Step2: Let the fertility rate at age i be b_i and the mortality rate be $d(i,t)$, then the survival rate at age i is $s(i,t) = 1 - d(i,t)$. Therefore, based on the correlation between b_i, $s(i,t)$ and $n_i(t)$, the following relationship can be derived.

$$\begin{cases} n_i(t+1) = \sum_{i=1}^{m} b_i n_i(t) \\ n_{i+1}(t+1) = s(i,t) n_i(t) \end{cases}, i = 1,2, ... m-1 \quad (1)$$

It can be reduced to equation (2):

$$n(t+1) = Ln(t) \quad (2)$$

where L is the Leslie matrix:

$$L = \begin{bmatrix} b_1 & b_2 & \dots & b_{m-1} & b_m \\ s(1,1) & 0 & \dots & 0 & 0 \\ 0 & s(2,2) & & 0 & \\ & & 0 & & \\ 0 & 0 & 0 & s(m-1,t) & 0 \end{bmatrix} \quad (3)$$

Therefore, when L and $n(0)$ are known, equation (4) is available for any $t = 1,2,3,\dots$.

$$n(t) = L^t n(0) \quad (4)$$

2.2 Parameter setting and solution

Step1: Based on the number of female population per age group, population mortality rate, female urban, town and rural population, fertility rate and female ratio in urban, town and rural areas in 2020 published by the National Bureau of Statistics. The number of babies born to women of childbearing age in each age group in urban, town and rural areas is calculated separately with the following formula:

$$\begin{cases} H_C(i) = b_c(i) * (n_c(0) + p_c M(i,0)) \\ H_Z(i) = b_z(i) * (n_z(0) + p_z M(i,0)) \\ H_X(i) = b_x(i) * (n_x(0) - M(i,0)) \end{cases} \quad (5)$$

Where $b_c(i)$, $b_z(i)$ and $b_x(i)$ denote the fertility rate of women of reproductive age by age in urban and rural areas respectively, $n_c(0)$, $n_z(0)$ and $n_x(0)$ denote the total population in urban and rural areas in 2020 respectively, p_c and p_z denote the proportion of women in urban areas respectively, and $M(i,t)$ denotes the migration function.

According to the data related to towns, towns and villages from the National Bureau of Statistics, it is found that there is a part of the population that migrates. Therefore, the migrating population $M(i,t)$ was set when calculating the number of babies born. We fitted the migration data at each age to derive a fitted migration function based on a confidence interval of 95%.

$$M(i,t) = \begin{cases} 1500(|i-30|+30) + \left(\sin(i) + \frac{\ln(t)}{10}\right), & 0 < i \le 60 \\ 1200\sin(i+t), & i = 0; 60 \le i \le 90 \end{cases} \quad (6)$$

Step2: Calculate the birth rate of female infants in urban, town and rural areas separately with the following formula.

$$\begin{cases} V_C = \frac{c_C}{100+c_C} \\ V_Z = \frac{c_Z}{100+c_Z} \\ V_X = \frac{c_X}{100+c_X} \end{cases} \quad (7)$$

Where c_C, c_Z and c_X denote the proportion of men and women in urban, town and rural areas respectively.

Step3: Calculate the female fertility rate for each age group with the following formula.

$$b_i = \frac{H_C(i)*V_C + H_Z(i)*V_Z + H_X(i)*V_X}{n_i(0)}, \quad i = 15,16,\dots\dots,49 \quad (8)$$

Step4: Calculate the survival rate of females at each age group. The mortality data by age population by urban and rural areas from 1953-2020 were selected and the mortality rate was predicted using ARIMA [5] $d(i,t)$, and $arima(i,t)$ was introduced, then $d(i,t) = arima(i,t)$. The mortality rate is normalised by the formula $D(t) = \frac{\sum_{i=0}^{m} d(i,t)}{1000} > 1$ for the female population at all ages, therefore, the mortality rate needs to be normalised by the following formula.

$$d1(i,t) = d(i,t)/D(t) \quad (9)$$

Step5: Construct the Leslie matrix based on equation (3).

Step6: Calculate the total population for the year in which the projection is required according to equation (4).

3. Experimental results and analysis

According to the current stage of China's development, the population is divided by age as follows: 0-14 years old as young, 15-64 years old as young-adult, 65 years old and above as old, respectively, using Y_i, T_i, O_i, where Y denotes the young population, T denotes the young-adult population, O denotes the old population, i denotes the year, noting i=1,2,..,N (i=1 for 2010).

We used the constructed age-specific population projection model to calculate the distribution of the number of young, young-adult and elderly population in 2010-2020 with the following equations.

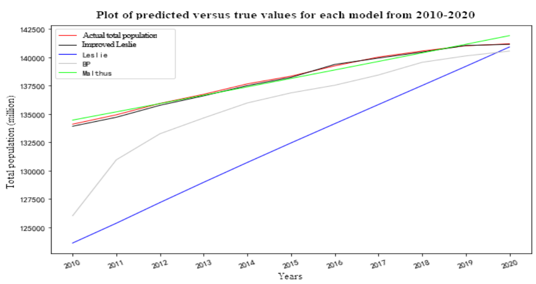
$$\begin{cases} Y_i = \sum_{j=1}^{14} Z_j \\ T_i = \sum_{j=15}^{64} Z_j, \ i=1,2,\ldots\ldots,N \\ O_i = \sum_{j=65}^{M} Z_j \end{cases} \quad (10) \qquad \begin{cases} \varepsilon_{Yi} = (Y_i - Y)/Y \\ \varepsilon_{Ti} = (T_i - T)/T \\ \varepsilon_{Oi} = (O_i - O)/Y \end{cases} \quad (11)$$

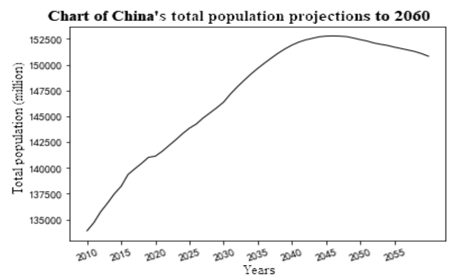
In order to better reflect the relationship between the true and predicted values, the error rate ε_i between the two is calculated by equation (11).

Where Y, T and O denote the true values of the young, young-adult and elderly populations respectively; ε_{Yi}, ε_{Ti} and ε_{Oi} denote the error rates of these three population categories respectively.

In order to further verify the accuracy of the model predictions, we compare and analyse the improved Leslie, Leslie, BP and Malthus predicted data for 2010-2020 with the real values, and by fitting the data, the predicted values are compared with the real value results as shown in Figure 1.

Figure 1: It can be seen that in the Malthusian model, the population is subject to infinite growth, but, in context, it is unlikely that the population of any area will grow indefinitely. Therefore, the exponential growth model is only suitable for short-term forecasting. In the BP neural network, the model tends to fall into local minima and overtraining tends to cause overfitting phenomena, making the accuracy rate drop. In the Leslie model, the impact of dynamic factors such as urbanisation and mortality on population development





is not taken into account, making the model less accurate. The improved Leslie model is more stable and has a lower prediction error rate than the rest of the models, with an average error rate of only 0.09%, and the model has strong generalisation ability and high accuracy.

As the sex ratio between male and female births, the degree of urbanisation and the fertility rate all have an impact on the size of the population, and based on the data comparison, the improved Leslie model is found to be more suitable for predicting the population size in any period. Therefore, the improved Leslie has better results in predicting the total population of China. The specific prediction results are shown in Figure 2.

Figure 2 China's population projections to 2060: It shows that the total population will reach 1.463 billion in 2030, compared to 1.411 billion in 2020, with a slow increase in the population under the national regulation policy.

4. Conclusion

We base our projections of China's total population mainly on population-related data available on the website of the National Bureau of Statistics. We based our projections of the total population on the Leslie model. Through comparative analysis with Leslie, BP and Malthus, we decided to use the Leslie matrix model with sub-regional mortality, fertility, infant sex ratio and migration rate for the projection of China's total population, and the projections show that China's population reaches 1.463 billion by 2030 and around 2045 reaching a peak.

References:

[1] WANG Ying, WANG Yanxia, ZHANG Huining, DU Yongqiang. Analysis of population aging prediction and influencing factors based on Malthus model[A]. China Society of Statistical Education, 2015:20.

[2] CHEN Shu-jun, WANG Yi-jiang, LIU Chun-mei. Population Structure and Development Trend of Heilongjiang Province under Different Fertility Policy[J].Journal of Harbin University of Commerce(Social Science Edition),2017(03):28-37.

[3] WANG Yifang, JIANG Yong, LIN Yuyang. The Allocation of Preschool Educational Resources under the Policy of Universal Two-children in China——Based on Leslie Model[J].Journal of Educational Science of Hunan Normal University,2018,17(03):59-66.

[4] NI Xuan-ming, SHEN Xin-ru, HUANG Song, ZHANG Jun-yu.The Convergence Trend of the Age Structure and the Aging Population in China[J].Journal of Applied Statistics and Management,2020,39(02):191-205.

[5] Xiao Hongmin, Ma Haifei, Kang Yanling.Comparison of Two Mortality Prediction Methods[J].Statistics & Decision,2020,36(23):5-8.