

Data Mining Research on Beijing Population Growth

Xinfang Song

General Ability Teaching Department, Beijing Information Technology College, Beijing, 100015

songxf@bitc.edu.cn

Keywords: Beijing population growth; Correlation analysis of SPSS; Logistic model; Granger causality test

Abstract. In this paper, a significant increase in population in Beijing the status quo, analysis of Beijing Population significant increase in the status quo, to identify relevant factors by reading literature, summed up the result of the rapid population growth in Beijing 30 factors, and to all the factors grouped into six Primary categories: ① population-related classes, ② economic level class, ③ consumption level class, ④ employment category, ⑤ Education, ⑥ income class. The use of official statistics, SPSS software through this six categories of basic data correlation detection and analysis of the factors that influence the focus, and then focus by Eviews software calibration factors affecting Granger causality, ultimately derived six kinds caused a significant increase in population in Beijing strong correlation factor people use in recent years, Beijing building data using logistic models to predict population growth in Beijing.

Introduction

Since reforming and opening up of China, in the overall national population development strategy under the guidance of the population of Beijing, it is fully developed. In recent years, along with urbanization, Beijing population growth is too fast and urban population density is too high that's given enormous pressure for Beijing's resource supply. The size and structure of Beijing population development have more restrictive relationship with the economic development. Therefore, controlling the excessive growth of population, improving population quality and achieving rational distribution of population are related to better promote the capital of social and economic development. From the official data to better promote the social and economic development of the capital. The main problems for Beijing are to study the rapid population growth and to strengthen population control. From the CPPCC option, the size of the population carrying capacity of the environment cause many "urban disease" root. The key to solving the population problem and breaking the "urban disease" ideas in one continuous line. In this paper it analyzes the current situation of population increase and uses SPSS software to analyze the data of seven categories of related factors which lead to population increase and to obtain the key influencing factors for Beijing. In the article the following hypothesis is established. In the stable social environment, there are no events such as war, virus and other large-scale harming the population, and there is no force majeure such as natural disasters. In the premise of social stability, fertility and mortality are relatively stable.

The Correlation Test on the Data Preconditioning of the Basic Factors of Population Growth in Beijing

With the accelerating integration of Beijing and Tianjin, the population of Beijing has experienced a rapid expansion. The main reasons for the population growth in Beijing are as follows: the development of Beijing's urban function and the growth of Beijing's entry into the project [1]; the high level of economic development, the amount of employment Increase; urban and rural development of regional imbalances, extensive economic growth; survival needs and self-realization. The adjustment of population in Beijing should be based on the combination of market and administrative methods, with structural adjustment as the guidance, moderate control of urban population growth, optimization of urban population structure, reasonable adjustment of

urban population spatial layout. The above researches laid the foundation for further study of population growth in Beijing, but did not excavate the factors that affect the population growth in Beijing comprehensively, deeply and systematically. On the basis of the characteristics of population growth in Beijing it will excavate, sort out and classify the factors that affect population growth in the paper. It's mainly excavated six basic categories and 30 basic factors that may affect the population growth in Beijing.

The data used in this paper are derived from the Beijing Statistical Yearbook. The most representative data segment of Beijing's population spike from 2001 to 2011 was selected in the yearbook. Because of the redundant data, significance (2-tailed) which indicated that there was significant correlation between the factor and population at the end of the year in Beijing was less than 0.05 before the statistical analysis. The data were analyzed by the software SPSS. In the process of correlation analysis, the dependent variable is the population at the end of the year, and the remaining variables are explanatory variables.

(1) Population-related categories: In this paper, the first factor class that will affect the population growth in Beijing is defined as Proportion of Migrant Population, Average Life Expectancy, Sex Ratio, Proportion of Urban and Rural Population, Birth Rate, Natural Growth Rate, Infant Mortality, Mortality, Household Population and other nine related factors. According to the correlation test, it is found that there are four significant correlations which are the proportion of foreign population, household population, birth rate, natural growth rate among the nine categories of population-related factors. Therefore we screen these four factors as the representative factors of population-related categories.

(2) Economic categories: In this paper, the second factor class that will affect the population explosion in Beijing is defined as the level of economics GDP, GDP per capita, Gross Regional Product Index, Gross Primary Product Index, Secondary Industry Gross Product Index, Gross Industrial Product Index, Gross Industrial Product Index, Building Index, Beijing Gross Domestic Product Index Industrial structure, composition of the primary industry, the composition of the secondary industry, the composition of the tertiary industry, the industrial structure, the construction industry and other 15 categories of factors. According to the correlation test, it is found that only the gross regional product (GDP) (Unit: 100 million yuan), per capita GDP, the composition of the primary industry and the composition of the tertiary industry are significant in the nine categories of population related categories. Therefore, we screen these four factors as the representative factors of population-related categories.

(3) Consumption level categories: In this paper, the third factor that affects the population growth in Beijing is defined as the level of consumption. The three main factors are: the consumption level of residents, the consumer price index and the Engel coefficient of urban households. Through the correlation test, there are the level of consumption, the consumer price index correlation significantly, and take them as the main consideration factors.

(4) Employment categories: the fourth factor that affects the population growth in Beijing is defined as the employment category. The main factors are the employees at the end of the year and the employment rate. Through the correlation test, it is found that only the employees at the end of the year have significant correlation and take it as Employment of the main consideration.

(5) Education categories: the fifth factor that affects the population growth in Beijing is defined as Education category, mainly constitute the number of ordinary colleges and universities, the number of college enrollment; Through the correlation test it is found only the number of significant correlation between colleges and universities have significant correlation and take it as the main consideration.

(6) Income categories: the sixth factor that affects the population growth in Beijing is defined as the income category which mainly consists of residents' income, per capita household income and per capita disposable income. The correlation test shows that the per capita household income and per capita disposable income are significant, so the per capita household income and per capita disposable income are the representative factors of the income class.

According to the results of the correlation analysis, the above six categories 30 factors were further excavated: elimination of the dependent variable is only the "end of the population," the description of the explanatory variables. Some indicators are related to Beijing's population at the end of the year, but these indicators are only descriptive for Beijing's population at the end of the year and are not explanatory. Based on the correlation test, 14 factors which have strong correlation with Beijing population are extracted from the original 30 factors. The results show that these factors are closely related to the population growth in Beijing.

Table 1 Passenger volume

| Categories | Factors |
|-------------------------------|---|
| Population-related categories | The population of Beijing at the end of the year, |
| | Household registration population |
| | infant mortality |
| | natural growth rate |
| Economic categories | the gross regional product (GDP) (Unit: 100 million yuan) |
| | per capita GDP |
| | the composition of the primary industry |
| | the composition of the tertiary industry |
| Consumption level categories | the consumer price index and |
| | the Engel coefficient of urban households |
| Employment categories | the employees at the end of the year |
| Education categories | the number of ordinary colleges and universities |
| Income categories: | the per capita household income |
| | per capita disposable income |

Granger Causality Test on the Factors of Population Growth in Beijing

Firstly we use the Eviews software for stationary verification, and then carry out regression analysis of variables co-integration test, so that we can do further Granger causality test.

The Granger causality test is conducted to determine the causal relationship between the 14 factors that contributed to the rapid population growth in Beijing and the population at the end of Beijing.

By using the Eviews for the stability check and the regression analysis of the covariance between the variables, so that you can further Granger causality test.

In order to determine whether there is a causal relationship between the 14 factors that have led to a sharp increase in population and the number of people in Beijing at the end of the year, the Granger causality test is carried out for the above 14 major influencing factors.

The Granger causality test was created by Clive WJ Granger in 2003 Nobel laureate in economics which analyze the Granger causality between economic variables. He gave the Granger causality a definition which is the variance of the optimal least squares prediction that relies on the use of all the information at some point in the past. The basic requirement for the Granger causality test is a smooth time series that is preceded by a Granger causality test 14 factors affect the stability test when the test results of the exact value of less than or equal to 0.05, that the variable is stable. It selected the Eviews for smoothness test and the default ADF method in this article. In order to

carry on the further analysis, the first and second order differences of these 30 variables are tested by ADF and these variables are found to be second order single order sequences. After the smoothness test, these factors and the number of people in Beijing at the end of the year to return in Fig. 1. Through the regression analysis of the amendment R2 we can be seen that 14 influencing factors on the number of people in Beijing at the end of the fitting effect is very good. Through the F test can be seen that the whole equation is significant. Through the t-test of these 14 variables, we can see that the following variables have a significant effect on the population of Beijing at the end of the year. These variables include: population at the end of Beijing, household population, infant mortality rate , The natural growth rate, the regional GDP (unit: billion), per capita GDP, the composition of the primary industry, the tertiary industry, the level of consumption, the consumer price index, the number of employees at the end of the year, Per capita disposable income.

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.999648 | Mean dependent var | 1262.245 |
| Adjusted R-squared | 0.999561 | S.D. dependent var | 311.4215 |
| S.E.of regression | 7.367432 | Akaike info criterion | 6.921222 |
| Sum squared resid | 381.4539 | Schwarz criterion | 8.167434 |
| Log Likelihood | -89.87345 | F-statistic | 2278.615 |
| Durbin-Watson stat | 2.993614 | Prob(F-statistic) | 0.000000 |

Figure 1. Regression results

| | | | |
|--|-----------|-----------|--------|
| Augmented Dickey-Fuller test statistic | | -9.017432 | 0.0000 |
| Test critical values: | 1% level | -3.914721 | |
| | 5% level | -3.011521 | |
| | 10% level | -2.514483 | |

Figure 2. Cointegration test results

After the regression analysis, we need to test whether the variables have a cointegration relationship. In this paper, the sequence c is tested by ADF. The test results are shown in Fig 2. It can be seen from the test results, sequence c has a smooth, can show that there is a cointegration between the variables. Therefore, we can further make the Granger causality test after all the above tests, you can further Granger causality test. According to the results of the Granger causality test, the confidence interval is 25%. Then the following eight indicators are obtained: the population at the end of Beijing, the natural growth rate, the regional GDP (unit: billion), the per capita GDP, the tertiary industry, residents' consumption level, the consumer price index, practitioners of the end of the year.

Therefore, we first through the ADF test and co-integration test, dig out 14 variables on the impact of the number of people in Beijing at the end of the more significant indicators, namely: the number of population in Beijing at the end of the population, household population, infant mortality, natural growth rate, GDP, per capita GDP, per capita disposable income, per capita disposable income. The Secondly, through the Granger causality test, excavated eight indicators which are caused by the Beijing population at the end of the year the Granger reasons. The eight indicators are: the number of people in Beijing at the end of the year, the natural growth rate, the regional GDP (unit: billion), per capita GDP, the tertiary industry, the level of consumption, the consumer price index, the end of the practitioners. By the results of the Granger causality test we can see what's digged out the eight indicators are reflected in the economic level and consumption level which will be collectively referred to as "economic indicators." This part of the economic indicators directly reflects the Beijing residents income and consumption levels. These

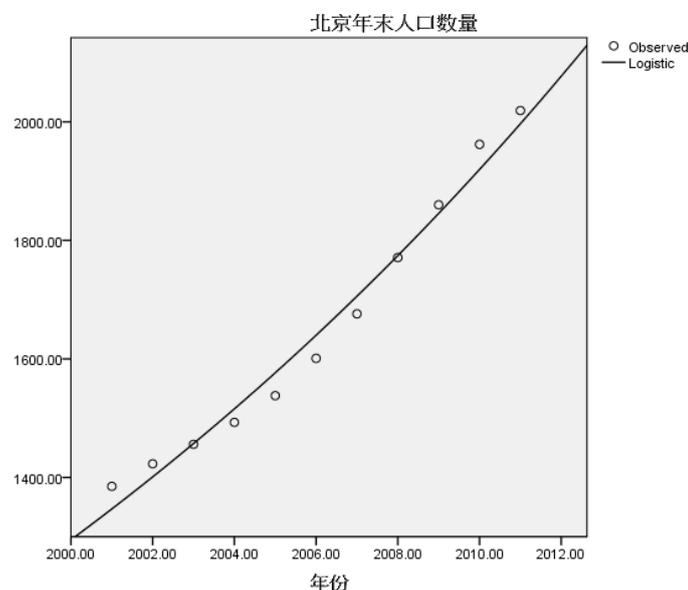
reasons are precisely to attract foreign population to Beijing the most important reason. Therefore, they are the main factors affecting the population growth in Beijing. However, it is undeniable that, in addition to economic factors, the natural growth of the natural population in Beijing and the reasons for education, such as Beijing's population growth has a very great impact. Therefore, by virtue of the statistical significance of the Granger causality test 8 Which is not enough to comprehensively include the main factors that affect the population growth in Beijing. On this basis, combined with the results of correlation analysis and co-integration test, the population growth of the natural population and the cause of education two factors into the further induction, and excavate the basic factors affecting the population growth : Beijing population at the end of the year, household population, infant mortality, natural growth rate, regional GDP (unit: billion), per capita GDP, The composition of the primary industry, the composition of the tertiary industry, the consumption level of residents, the consumer price index, the number of employees at the end of the year, the number of ordinary colleges and universities, the per capita household income, per capita disposable income.

The Statistical Analysis of the Population Data of Beijing Leads to the Regression Equation

In order to study the problem of population growth, the number of people in Beijing at the end of the year was chosen, and the logistic model was used to predict the population of Beijing in the next few years.

| year | Population (million) |
|------|----------------------|
| 2001 | 1385.00 |
| 2002 | 1423.00 |
| 2003 | 1456.00 |
| 2004 | 1493.00 |
| 2005 | 1538.00 |
| 2006 | 1601.00 |
| 2007 | 1676.00 |
| 2008 | 1771.00 |
| 2009 | 1860.00 |
| 2010 | 1962.00 |
| 2011 | 2019.00 |

The regression image is:



Model Summary and Parameter Estimates

Dependent Variable: Population at the end of Beijing

| Equation | Model Summary | | | | | Parameter Estimates | |
|----------|---------------|---------|-----|-----|------|---------------------|------|
| | R Square | F | df1 | df2 | Sig. | Constant | b1 |
| Logistic | .980 | 449.135 | 1 | 9 | .000 | 1.229E31 | .961 |

The independent variable is the year.

Sig value <0.05, so that the regression equation is significant.

So the regression equation is $y = 1.299e31 + 0.961x$.

Through the above images we can see that the population of Beijing is more and more like rising and reaching the high point, but we know that the population expansion is relatively fast, the contradiction between population and resources and environment is becoming more and more prominent, and the population is increasing, leading to overwhelming urban infrastructure. City public security brings problems!

On the impact of population growth on the environment, in 1970, Metos proposed that the population surge will lead to the following three crises simultaneously: 1) excessive land use, and thus cannot be used, food production decline; 2) natural resources due to the world's population; 3) serious environmental pollution, damage, so that the food to speed up production, a large number of human deaths, population growth to stop. In proposing the assertion, Metos did not take into account the ability of man to control his own development and the creativity of mankind, but it was thus clear that ecological balance and population growth were of great importance.

References

- [1] Liu Laifu, Huang Haiyang. Analysis on Basic Boundary Value Problem of functions with Two Displacement_ Liu laifu [J].Journal of Beijing Normal University(Natural Science), 987(2): -8.
- [2] Chen Qiang. Analysis of Population System Models and Population Status _ Chen Qiang [J].Nanjing University of Science and Technology,2004(5):15-20
- [3] Yu Taofang. China's Urban Growth: 2000-2010[J].Journal of Urban and Regional Planning, 012(2):70-87.
- [4] Li Huazhong. Application of Prediction Logistic Model in Population [J].Journal of Jiangsu Institute of Petrochemical Technology, 1998(2):13-15.
- [5] Wang Zhifu, Guan Jie, Su Zaixing. China's Population Growth Predictions Based on Logistics Model [J].Journal of Bohai University (Natural Science Edition), 010(4):326-330.
- [6] Huang Rongqing, Duan Cheng Rong, Lu Jiehua, and so on. Beijing Population Control [J]. opulation and Economics,2011(3):26-38.
- [7] Wang Zhouxi, Hu Bin, Wang Hongping. Study on Nonlinear Dynamics of Population Forecast Model_ Wang Zhouxi [J].The Journal of Quantitative and Technical Economics, 002(8): 3-56.
- [8] Liu Bo. Characteristics, Problems and Countermeasures of Beijing's Population Control [J].Ecological Economy, 2013(5):184-186.
- [9] Feng Shouping. Construction and Application of Prediction Model of Population Development in China [J].Statistics and Decision, 2010(15):24-27.
- [10] Zhu Yanwei, Zhang Yongli. Forecasting Model and Its Improvement of China's Population Growth [J].Statistics and Decision, 2010(16):20-21.

- [11] Long Chengxing, Zhang Bo. A Study of the Population Forecasting Model for Kunming [J]. *Journal of Yunnan University of Nationalities (Natural Sciences Edition)*, 011(4): 58-261.
- [12] Shen Wei, Wu Xin. Study on Influence Factors of Population Growth in Beijing Based on Data Digging [J]. *China Population Resources and Environment*, 2013(0):479-482.
- [13] Hu Xisheng, Fan Hailan, Song Ping, and so on. Application of Modified Logistic Model in Prediction of Urban Population_ Hu XiSheng [J]. *Journal of Beihua University (Natural Science)*, 008(4):370-373.
- [14] Chu Wenjie. Beijing Trend and Countermeasure Research on Changes in Population Size_ Chu Wenjie [J]. *China Population Resources and Environment*, 2009(5):431-436.