



## ARTICLE

## Research on China's Population Structure in the New Situation

Jingxuan Cui\* Mengshuai Yin Zerong Liu

North China University of Science and Technology, Tangshan, Hebei, 063210, China

## ARTICLE INFO

*Article history*

Received: 14 January 2020

Accepted: 17 January 2020

Published Online: 31 May 2020

*Keywords:*

Variance analysis

PDE model Differential equation model

Multiple regression model

Second child policy

## ABSTRACT

To analyze the impact of the “two-child policy” on the population size and structure, first of all, the birth rate, the ratio of men and women, and the ratio of urban and rural population are used as indicators. Before and after the dispersion, then establish a PDE model, and compare it with the population predicted by the gray forecast to analyze the mitigation of the ageing of the second child policy; continue to analyze the impact of changes in the population structure on the national economy, and select the male and female ratio and the labor population. The urban-rural population ratio is used as an index to establish a multiple regression equation for analysis, and a related regression equation is obtained. Finally, the future marriage problem is analyzed, considering only the difference in the number of men and women entering the marriageable period at the same time. The difference in the number of marriageable populations is analyzed through the difference in the number of men and women born at birth, focusing on a dynamic perspective.

### 1. introduction

Chinese population growth has undergone a historical change in just half a century after the founding of the People Republic of China. With the rapid development of our economy and the significant improvement in people living standards, China's rapid transition from a high fertility rate to a low fertility rate, the main contradiction of our population. Another is excessive growth, but imbalanced gender ratios and increasing population ageing. Against this background, the country has implemented a comprehensive second child policy.

The purpose of the policy is to increase the number of new students and adjust the population structure. This article establishes a model to determine the relevant impact indicators, to judge the short-term effects of the implementation of the new policy, and to the changes in China's population size and structure (gender, urban and rural,

aging, etc.) by 2035 or 2050. The impact of the national economy and the mitigation of the imbalance in the sex ratio of our population.

The comprehensive second child policy can adjust and improve the fertility policy, promote the increase of the population, improve the structure of the family, alleviate the problem of aging and the imbalance of gender ratio, and is conducive to the harmonious and stable development of society. In addition, the second child policy is conducive to maintaining a reasonable number and structure of the labor force, and is conducive to the balanced development of the population and the long-term development of the country.

### 2. Evaluation of Short-term Effects of the Second Child Policy

In order to analyze the short-term effects of the imple-

\*Corresponding Author:

Jingxuan Cui,

North China University of Science and Technology, No. 21 Bohai Avenue, Caofeidian District, Tangshan, Hebei, 063210, China;

Email: 1643360071@qq.com

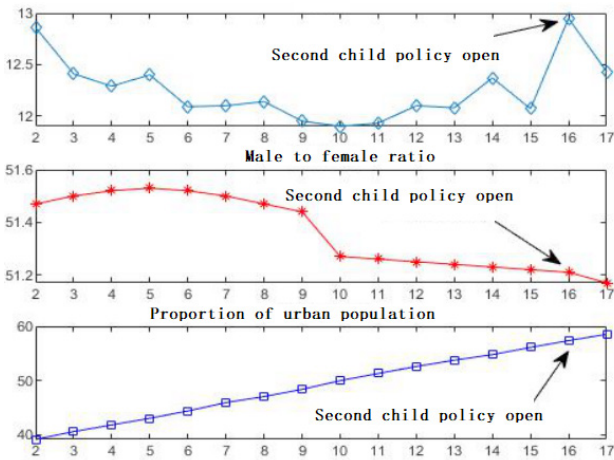
mentation of the second child policy, this model starts with the impact of the policy on the number of people and the population structure, and uses the birth rate, the ratio of men and women, and the ratio of urban and rural population as indicators. First, in order to visually and clearly observe the overall trend of changes in these three indicators before and after the introduction of the second-child policy, find out relevant data from the National Statistical Yearbook, plot their values and numerical change rates, and briefly analyze the second-child policy according to the trend of the graph Influence on each index, draw intuitive qualitative conclusions. Then, for the degree of dispersion of the specific observation data, the variance is calculated for quantitative analysis, and 2010-2014 and 2013-2017 are selected as two small samples to solve, and the quantitative analysis is obtained by comparison and analysis.

**2.1 Qualitative Analysis**

Considering the promulgation of the Family Planning Law in 2002 and the strengthening of national supervision, the data after 2002 were selected for analysis. The data on the birth rate, the proportion of male population and the proportion of urban population from 2002 to 2017 were obtained from national statistical maps, as follows: [1]

**Table 1.** Value of each indicator

Years	2002	2003	2004	2005	2006	2007	2008	2009
Birth rate(%)	12.86	12.41	12.29	12.40	12.09	12.10	12.14	11.95
male(%)	51.47	51.50	51.52	51.53	51.52	51.50	51.47	51.44
Proportion of urban population(%)	39.09	40.53	41.76	42.99	44.34	45.89	46.99	48.34
years	2010	2011	2012	2013	2014	2015	2016	2017
Birth rate(%)	11.90	11.93	12.10	12.08	12.37	12.07	12.95	12.43
male(%)	51.27	51.26	51.25	51.24	51.23	51.22	51.21	51.17
Proportion of urban population(%)	49.95	51.27	52.57	53.73	54.77	56.10	57.35	58.52



**Figure 1.** Overall trend of each indicator

According to Figure 1, from the overall trend of data changes, we can observe that the birth rate fluctuated between 2002 and 2015, and the overall trend was relatively stable. However, after the implementation of the second child policy in 2016, the birth rate increased significantly and reached a peak. The male to female ratio decreased at a constant rate from 2010 to 2016, but the rate of decline increased in 2017. As for the proportion of the urban population, the overall increase from 2002 to 2017 has been uniform.

**2.2 Evaluation of the Effect of the Second Child Policy Based on Analysis of Variance**

In order to quantitatively reflect the impact of the second-born child policy on the birth rate, the male-female ratio, and the urban-rural population ratio, it is mainly to analyze the changes in data before and after the policy is introduced, and to observe their degree of dispersion. Variance is introduced in this model, and five years is selected as a small sample for calculation. In order to highlight the impact of the second child policy on the indicators, one of the sample data for comparative analysis should only include the data before the introduction, and the other sample should include the data after the introduction. Therefore, the data of 2010-2014 and 2013-2017 are selected as two samples, and the variance is calculated.

**Table 2.** Indicator variance

Index	Birth rate(%)	Proportion of male population(%)	Proportion of urban population(%)
2010-2014	0.02052	0.0002	2.971984
2013-2017	0.10272	0.000584	3.322056
Relative change	4.00584	1.92	0.11779

Analysis: The variance of the birth rate from 2013 to 2017 is significantly different from that of 2010 to 2014, and the relative change is about 4 times, reflecting the large degree of dispersion in the birth rate data from 13 to 17 years. The birth rate of 15 to 17 years after the introduction of the policy is significant. The change indicates that the birth of the second child policy has a significant impact on the birth rate. The variance of male proportion in 13-17 years also changed to a certain extent compared to 10-14 years, the degree of change was 1.92 times, which also shows that the introduction of the second child policy also has a certain impact on the ratio of men and women. The difference in the proportion of the urban population between the two samples is small, indicating that the relative dispersion of the data is small, and within

the normal fluctuation range, it can be regarded as having almost no impact.

### 3. Long-term Effect Evaluation of the Second Child Policy

In order to predict the changes in the number and structure of China's population in 2035 or 2050, a PDE model was first established. Based on the data known in the statistical yearbook, a gray prediction model was used to predict the population and the number of women in all ages. Impact of population size. Then establish a differential equation to predict the number of the elderly population from 2035 to 2050, and analyze the future ageing degree. Finally, the impact of changes in the population structure on the national economy is analyzed, and multiple regression equations are established for analysis by selecting the ratio of men and women, the labor force, the ratio of urban and rural population, and the number of college graduates as indicators.

#### 3.1 PDE Forecast Population

Set fertility rate before second child policy  $A_{bt}$ , After the implementation of the second child policy, the fertility rate will change accordingly. If the change is expressed by an adjustment factor, the current fertility rate: <sup>[2]</sup>

$$A_{at} = A_{bt} \times I_t = A_{nbt1} \times 2 + \sum_{i=2}^{10} A_{nbt_i}$$

The total population in year t can be obtained from the population in year (t-1) + the number of newborns-the number of deaths in (t-1) years. The number of newborns is related to the fertility rate, assuming that 15-49 women are fertile The expression for the life cycle of this model is:

$$P_t = P_{t-1} + P_{bt} - P_{dt} = P_{t-1} - P_{dt} + \sum_{n=15}^{49} [P_{n(t-1)} \times R_{n(t-1)} A_{nt}]$$

Through the adjustment of related parameters, the population expression after the implementation of the second child policy can be obtained:

$$P_t = \sum_{n=15}^{49} [P_{n(t-1)} \times R_{n(t-1)} (A_{nbt1} \times 2 + A_{nbt_i})] + P_{t-1} - P_{dt}$$

Query the total population, mortality, and number of females of all ages through the statistical yearbook, and use the gray prediction data to formulate.: <sup>[4]</sup>

**Table 3.** Forecasted population

Years	2018	2019	2020	2021	2022	2023
Number of people (10,000)	140155	140939	141732	142535	143346	144168
Years	2024	2025	2026	2027	2028	2029
Number of people (10,000)	144998	145838	146687	147546	148415	149294
Years	2030	2031	2032	2033	2034	2035
Number of people (10,000)	150183	151029	151909	152794	153684	154579

According to the data in the table, the second child policy has significantly promoted the growth of the population.

#### 3.2 Prediction of the Elderly Population based on Differential Equations

To predict the degree of aging in the future, the number of aging populations should be predicted first. Since the second-child policy was implemented in 2015, the number of elderly people will not be affected by the second-child policy by 2035 or 2050. Therefore, it is not necessary to consider the second child policy when predicting the number of elderly people. Find out the number of people over the age of 65 from the statistical yearbook, treat it as the number of elderly people, and build a differential equation to predict.

##### Step1. The establishment of differential equations

Assume that the net growth rate of the elderly is a constant. Based on the population data from 2002 to 2015, and build a population index growth model, there is an equation

$$\begin{cases} \frac{dx}{dt} = rx \\ x(0) = x_0 \end{cases} \quad (1)$$

Solutions have to:

$$x(t) = x_0 e^{rt} \quad (2)$$

(2) indicates that the population will grow infinitely over time according to the exponential law, which is called an exponential growth model. But in the long run, the population cannot grow indefinitely, that is, the exponential model cannot describe or predict the evolution of the population over a long period of time. The population growth rate r decreases as the population x increases, and r is represented as a function r(x) of x and is a decreasing function. So (1) is written as:

$$\begin{cases} \frac{dx}{dt} = r(x)x \\ x(0) = x_0 \end{cases} \quad (3)$$

Let  $r(x)$  be a linear function of  $x$ ,

$$r(x) = r - ax (a > 0) \tag{4}$$

$r$  represents the growth rate when the population is small and is called the inherent growth rate. And when time tends to infinity and the population  $x$  reaches the maximum value  $x_m$ , the growth rate  $r(x_m) = 0$ , which can be solved in equation (4) to get  $a = r / x_m$ , so  $r(x) = r(1 - x / x_m)$ , substituting  $r(x)$  into equation

$$\begin{cases} \frac{dx}{dt} = rx(1 - \frac{x}{x_m}) \\ x(0) = x_0 \end{cases} \tag{5}$$

Among them,  $rx$  is the inherent growth trend of the population, and  $(1 - x / x_m)$  is the external retarding effect on population growth. Equation (5) is called a retarded growth model, and the solution to equation (5) is obtained by using the separation variable method as [5]:

$$x(t) = \frac{x_m}{1 + (\frac{x_m}{x_0} - 1)e^{-rt}} \tag{6}$$

当 when  $t \rightarrow \infty, e^{-rt} \rightarrow 0, 1 + (\frac{x_m}{x_0} - 1)e^{-rt} \rightarrow 1$ , 所以 because  $\lim_{t \rightarrow \infty} x(t) = x_m$ .

**Step 2. Model solving and prediction**

The number of aging population from 2002 to 2015 is summarized as follows:

**Table 4.** Number of elderly people

Years	2002	2003	2004	2005	2006	2007	2008
People	9377	9692	9857	10055	10419	10636	10956
Years	2009	2010	2011	2012	2013	2014	2015
People	11307	11894	12288	12714	13161	13755	14386

Estimate the parameters  $r$  and  $x_0$  in the model, take the logarithm of Equation (2), and get:

$$y = rt + s, y = \ln x, s = \ln x_0 \tag{7}$$

Based on the data from 2002 to 2015, fitting the formula (7) to the data, the solution obtained is  $r = 0.03264$ ,  $s = 9.093$ , and  $s = \ln x_0$ .

So there are prediction formulas:

$$x(t) = 8892.8245e^{0.03264t} \tag{8}$$

**Note:** Consider 2002 as  $t = 1$ , and  $t$  is incremented by one for every year

thereafter. For 2015,  $t = 14$ .

The number of elderly people in the country from 2035 to 2050 predicted by the above equation is as follows:

**Table 5.** Predicted number of elderly population

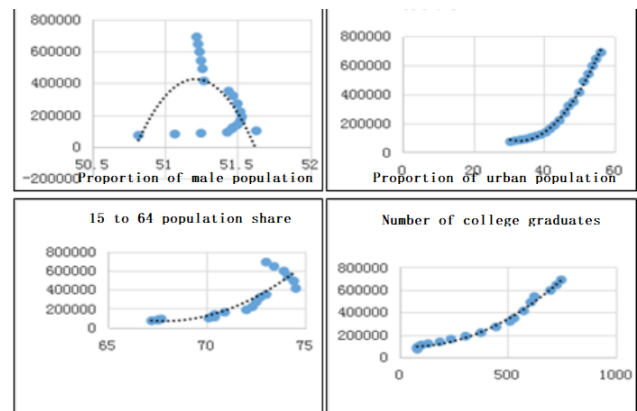
Years	2035	2036	2037	2038	2039	2040	2041	2042
Number of people (10,000)	26978	27873	28797	29753	30740	31760	32814	33902
Years	2043	2044	2045	2046	2047	2048	2049	2050
Number of people (10,000)	35027	36189	37390	38631	39912	41237	42605	44018

**4. National Economic Analysis Based on Multiple Regression**

Analyze the impact of changes in population size and structure on the national economy. The structure is roughly divided into gender structure, age structure, urban and rural structure, etc. Several factors affecting the development of the national economy are selected as independent variables: male and female ratio, labor force population, urban and rural population ratio, Number of college graduates. Due to multiple factors, a multiple regression equation was established for analysis. [6]

**Step 1. Remember the GDP**

Male population share is  $x_1$ , Urban population share is  $x_2$ , Number of working population is  $x_3$ , Number of college graduates is  $x_4$ . For rough analysis  $y$  and  $x_1, x_2, x_3, x_4$ . First use the data to make separately  $y, x_1, x_2, x_3, x_4$  Scatter plot, as follows:



**Figure 2.** Scatter plot

It can be found from the figure  $x_1$  increase,  $y$  There is a tendency to reduce downward bending, Select multiple fitting methods, When fitting a quadratic polynomial maximum, That is, the fitting is the best, so the quadratic polynomial fitting form is selected.

Similarly,  $x_2, x_3, x_4$  Also select the best fitting quadratic polynomial fit. The curve model in the figure:

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon \\ y = \beta_0 + \beta_3 x_2 + \beta_4 x_2^2 + \varepsilon \\ y = \beta_0 + \beta_5 x_3 + \beta_6 x_3^2 + \varepsilon \\ y = \beta_0 + \beta_7 x_4 + \beta_8 x_4^2 + \varepsilon \end{cases}$$

Based on the above analysis, the following regression model is established:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_3 + \beta_6 x_3^2 + \beta_7 x_4 + \beta_8 x_4^2 + \varepsilon$$

The MATLAB statistical toolbox can be used to obtain the regression coefficient prediction value and confidence interval (confidence level  $\alpha = 0.05$ ), test statistics  $R^2, F, p$ . The results are shown in the table:

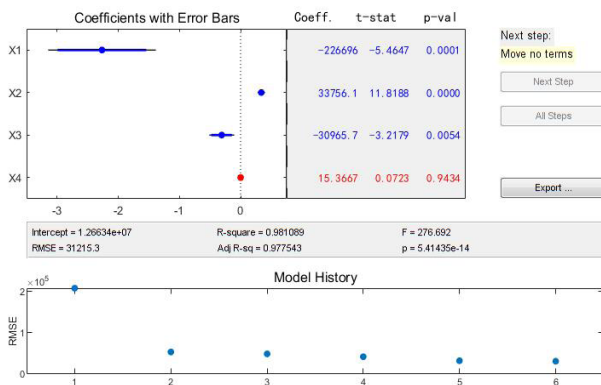
**Table 6.** Parameter values

parameter	Parameter estimates	Parameter confidence interval
$\beta_0$	-85933361.07	[-530644407,358777684]
$\beta_1$	3597195.167	[-13484200,20678590]
$\beta_2$	-34846.63792	[-201370,131677]
$\beta_3$	-61155.99952	[-110480,-11832]
$\beta_4$	939.0523233	[365,1513]
$\beta_5$	-168430.1336	[-470706,133846]
$\beta_6$	1221.789891	[-837,3317]
$\beta_7$	-195.632675	[-486,95]
$\beta_8$	0.375699668	[0,1]
<b>R2=0.9991 F=1456.3333. p=1.5801E-18</b>		

Analyzing the data in the graph can see the parameters  $\beta_0, \beta_1, \beta_2, \beta_5, \beta_6$ . The confidence intervals all contain zeros, so this model is not reasonable.

**Step 2. Use stepwise regression** [7]

A regression model was selected from the four independent variables that had a significant effect on the dependent variable y. Use the stepwise command in the MATLAB statistics toolbox to list the statistics:



**Figure 3.** Statistical table

It can be seen,  $\beta_5, \beta_6$ . The residual standard deviation (RMSE) is very small, close to zero. So that individual data does not affect the entire model, Abnormal indicators  $x_4$  Cull. Re-estimate the regression coefficient estimates and their confidence intervals for the model (confidence level  $\alpha = 0.05$ ). The data obtained are as follows:

**Table 7.** Improved model parameter values

parameter	Parameter estimates	Parameter confidence interval
$\beta_0$	-235482053	[-352936785,-118027320]
$\beta_1$	9374181	[4853357,13895005]
$\beta_2$	-90899	[-135048,-46751]
$\beta_3$	-95451	[-102377,-88526]
$\beta_4$	1356	[1284,1427]
$\beta_5$	-131628	[-216571,-46685]
$\beta_6$	978	[387,1569]
<b>R2=0.9988 F=1801.9085 p=3.31E-18</b>		

Analysis: According to the data in Table 7, 99.88% of the dependent variable y can be determined by this model ( $R^2=0.9988$ ), F far exceeds the critical value of F test, and p is much smaller than  $\alpha$ , so the improved model is available. The parameter values in Table 7 are compared with those in Table 6. The improved model F significantly increases and  $\alpha$  decreases significantly, indicating that the regression effect of this model is better.

The correlation equation between the national economy y and each indicator x can be expressed as:

$$y = -235482053 + 9374181x_1 - 90899x_2 - 95451x_3 + 1356x_4 - 131628x_5 + 978x_6^2$$

**Step 3. Model test**

The model does not consider that the data is a time series. Due to the regression analysis of the time series data, the model's random error term  $\varepsilon_t$  May be relevant, So it's D-W test. [8]

First according to  $e_t = y_t - \hat{y}_t$  false Calculating residuals, Then calculate the DW statistics based on the residuals as follows:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2}$$

According to the calculation, the DW value of this model is 1.8848. According to the sample size and the number of regression variables, check the DW distribution

table.  $\epsilon_t$  There is no autocorrelation, Therefore, the above regression equation is reasonable. The regression equation can reflect the impact of various indicators on the national economy.

### 5. The Number of Men and Women

In this model, all live-born babies born in a certain year are regarded as the same group of people, and it is regarded as a major factor in the study. Combined with the method of population change, it is inferred that the same group of people has changed over the past 20 years. What will happen to the sex ratio of the marriageable population?

If the marriageable population in 2020 is predicted, the total number of newborns in each year from 1986 to 2000 is calculated. In order to obtain the difference in the number of men and women in the marriageable population, we can calculate the number of male and female births and the number of births each year, and obtain the number of boys and girls born each year. Taken together, you can get the difference in the number of men and women in the marriageable population in a year.

According to the number of newborns from 1986 to 2017 and the proportion of male and female newborns in the corresponding year, the number of new males and females in the corresponding year can be calculated, and their number difference is obtained. The difference in the number of newborn men and women from 2003 to 2017 was calculated. Correspondingly, we can get the difference in the number of men and women in the marriageable population in 2020, 2021, ..., 2037.

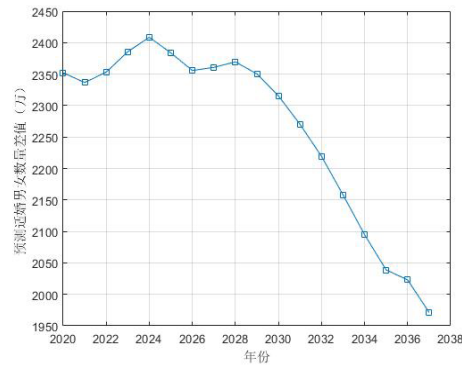
$$\sum_{i=t}^{t+14} Q_{ni} - Q_{vi}, t = 1, 2, \dots, 18$$

t is equal to 1, corresponding to 1986, corresponding t = 18 to 2003,  $Q_{ni}$  represents the number of new boys in the i-th year and  $Q_{vi}$  represents the number of new girls in the i-th year.

The following is the difference in the number of men and women in the marriageable population between 2020 and 2037.

**Table 8.** Gender difference in marriageable population

Years	2020	2022	2023	2025	2026	2028
Quantity difference	2352.020432	2353.159456	2385.520709	2383.694522	2355.715401	2369.388488
Years	2029	2031	2031	2034	2035	2037
Quantity difference	2350.293953	2270.145631	2219.259527	2094.805691	2038.870838	1971.152388



**Figure 4.** Curve of quantity difference

According to the chart, the year with the largest difference in the number of men and women in the marriageable population appears in 2024, with a difference of 2408.563753 million, and then there has been a downward trend. Therefore, we can preliminary judge that the “China’s marriageable population will appear in the future 30 million to 40 million bachelors”, with the change of people’s conception of fertility and the opening of policies, the ratio of male and female newborn babies gradually approaches 1:1, which is conducive to alleviating the population problem of imbalance between male and female.

### References

- [1] China Statistical Yearbook.
- [2] Jingyan Lv. The impact of the comprehensive second child policy on China’s population structure [A]. *Jingchu Academic* 2017, 7 (Fifteenth in total)[C]. Returning Intelligence(Wuhan)Strategic Investment Management Co., Ltd., 2017: 9.
- [3] Xingyu Mao. A Study of the Population Structure Before and After the Comprehensive Second Child Policy[J]. *China High-tech Zone*, 2018 (13): 278.
- [4] Zhizheng Zhang. Prediction model of population structure before and after comprehensive second child policy[J]. *China High-tech Zone*, 2018 (03).
- [5] Qiyuan Jiang, Jinxing Xie, Jun Ye. *Mathematical Model(Fourth Edition)*[M]. Beijing: Higher Education Press, 2011.
- [6] A rongGaowa. Estimation of passenger flow in A-level scenic spots in Inner Mongolia--5 single factor analysis of variance and multiple regression modeling[J]. *Arid Land Resources and Environment*, 2019 (12).
- [7] Zhengjiang Chen. A Comparative Study of Multiple Linear Regression and Stepwise Regression Analysis[J]. *Journal of Mudanjiang College of Education*, 2016 (05).
- [8] Shoukui Si, Zhaoliang Sun. *Mathematical Modeling Algorithms and Applications*[M]. Second Edition. National Defense Industry Press, 2015.04.