

Change trend of birth rate of our country population and influence factor analysis

Ran Yu ^a, Dan Li ^b, Hongyun Gao ^c and Shuang Chen ^{d, *}

College of Information Engineering, Dalian University, Dalian China

^a ranyu@dlu.edu.cn, ^b lidandlm@163.com, ^c gaohongyun@dlu.edu.cn,

^d chenshuang0707@163.com

* Corresponding author

Abstract. On the ground of time series analyses, this paper excavates the trend in preliminary and carries out some short-term forecasts about future birthrate variation in wish to shed light on its solutions. This paper attempts to excavate the influence that several dominant factors exert on birthrates via regression analyses. The cluster analyses are utilized to filter out representative provinces as samples for subsequent research in order to tackle the problem that the correlations between dependant variable and each independent variable is not significant. Wielding SPSS as an analytical tool, we construct the regression model about how price level in combination with education level influences the birthrates.

Keywords: Birth Rates; Time Series Analyses; Regression Analyses; Cluster Analyses.

1. Introduction

The birth rate plays an important role in a country's economic development. In recent years, due to the rapid development of China's economy, the prices of commodity and house have risen rapidly. More and more young people are no longer willing to have children because of the pressure from housing, children's education and other social aspects, which accelerates the population aging of our country and has a certain impact on economic growth. With the help of the theories and methods of statistics and economics, combined with relevant laws, regulations and policies, we can deeply explore the changing trend and influencing factors of China's population birth rate through data, which can further understand the main reasons why people are unwilling to have children, and moreover provide the clues and guidance for their solutions.

A large number of excellent documents focusing on this topic have emerged before. Some scholars try to use time series analysis and other methods to fit the changing trend of China's birth rate and make a reasonable prediction. At the same time, it also analyzes the most favorable level for economic development when the population birth rate is stable. For example, through data analysis in 2009, Dong Yongquan predicted that China would achieve zero population growth around 2032 and begin to experience negative growth[1]. For another example, in 2017, Tang Jieru and others used eviews software to establish time series models such as AR model, MA model and ARIMA model to predict the changing trend of population structure in a certain area in the next 20 years without implementing the two-child policy, and concluded in their academic papers that the fitted model of birth rate and sex ratio is $Tt = 5.9580 + 2.8273Lt$; The optimal birth rate is about 15%[2]. Other scholars have established the optimal regression model through various methods to find the factors that have the greatest influence on the birth rate, and put forward relevant schemes to control these factors so as to keep the birth rate in a reasonable range. For example, in the academic paper published by Wang Dehuai and others in 2018, it is pointed out that the level of economic development, the level of social security development and the level of education have a negative impact on the birth rate; The living standard of residents has a positive impact on the birth rate[3]. For another example, Qi Tianjiao said in his English thesis that the birth rate in a province is positively correlated with the illiteracy rate and negatively correlated with the quality of basic education; However, the proportion of women with

higher education has no significant relationship with the fertility rate[4]. The above research results have important reference significance and inspiration for this study.

This paper aims to explore the changing trend of China's birth rate through time series analysis and make a short-term prediction, and then establish a multiple regression model of birth rate about price level and education level through regression analysis.

2. Research method

2.1. Time series analysis

This paper adopts an autoregressive moving average model, namely ARMA model, which is widely used in important fields such as econometrics and engineering forecasting [5]. ARMA model is an important method to study time series, which is based on the autoregressive model (AR model) and the moving average model (MA model), and is often used in the research of long-term tracking data in market research. The general expression of the ARMA (p,q) model is:

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Where, $\phi_1, \dots, \phi_p$ respectively represent each autoregressive coefficient, $\theta_1, \dots, \theta_q$ respectively represent the moving average coefficients, $\varepsilon_t \sim N(0, \sigma^2)$ for $\forall s < t, E(X_s \cdot \varepsilon_t) = 0$.

2.2. Cluster analysis

Cluster analysis, also known as point group analysis, is one of the key technologies for building personalized information service systems [6]. Cluster analysis is mainly based on the difference or similarity of features to achieve the purpose of grouping sample data. Its primary use is knowledge discovery rather than prediction. The main criteria for evaluating the quality of the results of cluster analysis are as follows: the similarity degree of the data within the group is as large as possible, and the mutual degree of the data between the groups is as large as possible, which indicates that the cluster analysis has a better effect. There are several commonly used clustering analysis algorithms: K-means clustering algorithm, matrix clustering algorithm, class hierarchy algorithm. Cluster analysis is widely used in text mining, data analysis, pattern recognition and other Internet fields [7]. In this paper, the system clustering method is adopted, and 23 "representative" regions are selected according to the pedigree output results.

2.3. Regression analysis

Unitary linear regression model: assuming that there are two related variables x and y , then the established model, used by $y = \alpha + \beta x + \varepsilon$, $E(\varepsilon) = 0$ and $D(\varepsilon) = \sigma^2$, is called the unitary linear regression model [8]. Where, α and β are the parameters in the model, called regression coefficients, and ε is the random disturbance term. For any set of observations (x_i, y_i) , $i = 1, 2, \dots, n$, $y_i = \alpha + \beta x_i + \varepsilon_i$ is satisfied, where each ε_i is independent of each other and subject to a normal distribution, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, \dots, n$.

In addition, this paper also adopts the curve regression model, which is the generalization form of linear regression model.

Multiple linear regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon$, Where ε is a random perturbation, independent of each other and subject to the normal distribution $N(0, \sigma^2)$. The equation is expressed in matrix form $Y = X\beta + \varepsilon$, where

$$Y = [y_1, y_2, \dots, y_n]^T, \beta = [\beta_0, \beta_1, \dots, \beta_m]^T, \varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m]^T, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}$$

3. Analysis of birth rate change trend in China

3.1. Data selection

The main source of data in this paper is secondary data, which is faster and more convenient to collect, and the timeliness also meets the requirements of this study. The main way of collection is online collection, because the scope of online collection is wider, the amount of information is larger, so we can obtain the data we need; In addition, the access channels are basically official channels, which ensures the reliability, authenticity and objectivity of data [9]. The Data of China's birth rate over the years come from World Bank Open Data.

According to the preliminary observation and analysis of the data, the birth rate showed a small U-shaped increase in the years after 1982, and then showed a decreasing trend year by year since 1987. And 1982 is just the initial implementation of our country's reform and opening policy, the stage of economic development began to take off. It is also from 1982 that our country listed family planning as a basic national policy and wrote it into the constitution. Since the small section of data of U-shaped growth is obviously inconsistent with the overall law in the later period, which may be the manifestation of the initial implementation of the family planning policy, this time series analysis abandons that section of data and selects the data from 1987 to 2019 as samples for analysis. Some sample data are shown in Table 1.

Table 1. China's annual birth rate from 1987 to 2019 (part)

Year	Birth Rate (%)	Year	Birth Rate (%)	Year	Birth Rate (%)
1987	23.33	1994	17.7	2013	12.08
1988	22.37	1995	17.12	2014	12.37
1989	21.58	1996	16.98	2015	12.07
1990	21.06	1997	16.57	2016	12.95
1991	19.68	1998	15.64	2017	12.43
1992	18.27	1999	14.64	2018	10.94
1993	18.09			2019	10.5

3.2. Example analysis

3.2.1. Forecasting models

As can be seen from the time series chart, China's birth rate has been showing a downward trend since 1987, but the deceleration is slowing down, and has remained relatively stable since 2004. Based on the analysis of the actual situation, in the early stage, the main factor leading to the decline in the birth rate may be the influence of the family planning policy, while in the later stage, it may be due to the great changes in people's ideas, as well as the pressure from prices, housing prices and other social aspects, and no longer willing to have children. Since 2015, the birth rate has experienced a small increase and then dropped rapidly, possibly due to the influence of the two-child policy.

According to the ADF test, since $p = 0.0000 < 0.05$, the null hypothesis that the time series has a unit root is rejected at the significance level of $\alpha = 0.05$, so the time series is stationary.

According to the autocorrelation and partial autocorrelogram, it is reasonable to determine the maximum lag order according to the autocorrelation and partial autocorrelogram, and take 1 for Pmax and 6 for Qmax. The information criterion minimization principle is used to determine the best ARMA model, Pmax=1, Qmax=6, so there are six combinations as follows: ARMA(1,0); ARMA(1,1); ARMA(1,2); ARMA(1,3); ARMA(1,4); ARMA(1,5). According to the information criterion, the model with the smallest values of AIC, SC, HQ and other indicators should be selected. After comprehensive comparison of the test effects of the above models, it is concluded that the application effect of the ARMA(1,1) model is the most ideal.

The model estimation expression is $x_t = 16.74427 + 0.989815x_{t-1} + \varepsilon_t - 0.660060\varepsilon_{t-1}$.

According to the residual test, because the P value of the first 28 q statistics is > 0.05 , we cannot reject the null hypothesis that there is no correlation between the residual series at the significance level of $\alpha = 0.05$, so the final test result can be considered as: There is no correlation in the residual series, the white noise test passes, and the model is significantly effective. Therefore, the model we constructed is relatively good.

According to the final fitting effect diagram of the model, the fit between the fitted value and the real value is relatively ideal. From the residual plot, it can be seen that the residual of this model also fluctuates randomly around 0, so the extraction of residual information is also sufficient.

3.2.2. Short-term projections of the birth rate

According to the ARMA model based on the sample data from 1987 to 2019, the short-term prediction of China's birth rate from 2020 to 2022 is 10.5‰ (2020), 10.6‰ (2021) and 10.6‰ (2022). However, the actual values of China's birth rate in 2020 and 2021 are 8.5 per thousand and 7.5 per thousand, respectively, and the estimated residual values are large and significantly underestimate the decline in the birth rate. Therefore, we can speculate based on this that without the introduction of the three-child policy, in the next few years, the birth rate in our country may continue to decline and lead to negative population growth, which means that the domestic market may shrink. At present, the anti-globalization trend begins to emerge, the import and export trade is blocked, and it is urgent to expand domestic demand to stimulate the economy, which is obviously very worrying. Therefore, it is not difficult to understand that the government has introduced a two-child policy and a three-child policy in order to stimulate population growth and stabilize the economy. Therefore, in the coming years after the introduction of the three-child policy, which may temporarily stimulate population growth, the declining trend of the birth rate may slow down somewhat. However, whether the population will continue to grow in the future needs more policy support, such as stabilizing housing prices and prices, increasing the employment rate, improving the social security system, improving social welfare and other practical measures to fundamentally solve the problem that people are not willing to have children.

4. Analysis of influencing factors of birth rate

4.1. Monadic regressive analysis

According to the linear regression and quadratic regression of the birth rate with respect to the price level, the p value of the linear regression model is 0.157, which is greater than α , and the significance test is not passed. However, the p value of the quadratic regression model is basically less than α , so it can be considered that the significance test of the quadratic regression passes, and the model form is $Br = 0.430PL^2 - 0.347PL - 0.411$.

According to the linear regression and the quadratic regression of the birth rate on the education level, the p value of the linear regression model is 0.053, which is greater than but very close to 0.05. Moreover, the fitting effect of the curve model is not ideal. Therefore, this paper still considers the establishment of a unary linear model, whose model form is $Br = 0.408EL$. This shows that the relationship between the birth rate and the level of education is not only close to linear, but also

positive, which is not consistent with the previous research and our traditional concept, the specific reasons need to be further research.

The linear regression of the birth rate on the proportion of young people shows that the p value of Pearson's correlation test is $0.165 > 0.05$, which fails. Therefore, at the significance level of $\alpha = 0.05$, it can be assumed that there is no significant linear relationship between the birth rate and the proportion of young people. In addition, from the scatter plot, the distribution of the scatter plot is disorderly, indicating that the correlation between the two variables is weak.

4.2. Multiple regression analysis

Multiple regression analysis was conducted with price level and education level as independent variables and birth rate as dependent variable. The regression equation obtained is:

$$Br = 0.379PL^2 - 0.285PL + 0.220EL - 0.362.$$

In this paper, the rank correlation coefficient method (Spearman test) is used to diagnose heteroscedasticity. According to the rank correlation coefficient test, it can be seen from the test results that the p values of the rank correlation coefficient test between the two independent variables of price level and education level and the residual error are 0.703 and 0.914 respectively, which are significantly greater than $0.05 = \alpha$. Therefore, the size of the residual error does not change with the change of the size of the independent variable. There is no heteroscedasticity, and the model effect is better.

It can be seen from the collinearity test that VIF is significantly less than 10.

It can be seen from the scatter plot that most of the points are near the line $y = x$, indicating that most of the fitted values are close to the real values, and the fitting effect of the model is good. The adjusted goodness of fit is 0.395, which also indicates that the goodness of fit is within the acceptable range.

5. Conclusion

This paper uses Eviews software as an analysis tool, combined with the theoretical knowledge of time series analysis, constructs the ARMA model, analyzes the changing trend of China's birth rate from 1987 to 2019, and makes a short-term forecast. The obtained ARMA model is:

$$x_t = 16.74427 + 0.989815x_{t-1} + \varepsilon_t - 0.660060\varepsilon_{t-1}.$$

It is concluded that since the family planning policy was fully implemented and the reform and open policy was carried out, the change trend of the population birth rate in our country has shown a decreasing trend on the whole, but the deceleration is slowing down and has been relatively stable since 2004. However, there are various internal reasons for the downward trend: in the early stage, the main reason for the sharp decline in the birth rate may be the influence of the family planning policy, which indicates that the family planning policy in our country has been implemented and the problem of over-rapid population growth has been successfully controlled; In the later period, especially after the liberalization of the two-child policy and the three-child policy, the main reasons for the decline in the birth rate may be various economic and social pressures, as well as fundamental changes in people's ideas and ideas, and people's willingness to have children constantly decreases. Moreover, despite the good fitting effect of the ARMA model, the final forecast results still underestimate the decline in China's birth rate in 2020 and 2021, indicating that although the country has begun to encourage fertility, people's subjective willingness to have children is still too low, and more effective measures should be taken to stimulate fertility. Otherwise, Birth rates are likely to fall further.

At the same time, this paper attempts to conduct regression analysis based on SPSS to explore the main factors affecting the birth rate. The main factors considered are: the price level (measured by the CPI index), the education level (measured by the average public finance budget for education of

ordinary high school students), and the proportion of the young population (measured by the proportion of the population aged 15-59). During the correlation analysis, it is found that the correlation is not significant, so this paper excludes 8 abnormal regions through cluster analysis, and only retains the remaining 23 representative regions as samples for subsequent analysis. Finally, this paper finds that the main factors affecting the birth rate in China are price level and education level, and the established multiple regression model is

$$Br = 0.379PL^2 - 0.285PL + 0.220EL - 0.362 .$$

Based on the regression model, this paper takes Hunan Province as an example and concludes that the consumer price index should be controlled at around 101.92 in order to stabilize the birth rate at a level conducive to the current economic growth under the assumption that the current education level remains unchanged.

Acknowledgment

The research were partially supported by the National Natural Science Foundation of China No.11701061, the High-level Talents Innovation Support Program of Dalian No.2021RQ062, the Basic Research Project of the Education Department of Liaoning Province No. LJKQZ2021170, and Interdisciplinary project of Dalian University (DLUXK-2023-ZD-001).

References

- [1] Dong, Y. Forecast of population growth in China [J]. Journal of Tangshan Normal University,2009,31(2):46-48.
- [2] Tang, J., Zhu, X. & Zhou, L. Prediction and research of population structure based on time series [J]. Journal of Jiamusi University (Natural Science Edition),2017,35(4):696-699.
- [3] Wang, D. & Li, X. Spatial measurement of birth rate and its influencing factors in China [J]. Guizhou Science,2017,35(6):66-71.
- [4] TIANJIAO QI. An Empirical Analysis of the Effect of Education on Birth Rate Based on Chinese Provincial Panel Data[C]. //Proceedings of the 2nd International Workshop on Education Reform and Social Sciences (ERSS 2019). 2019.
- [5] Fang, R. Modeling and application based on ARMA time series theory [J]. Science and Technology Information,2012(19):197-198.
- [6] PAN Wei. Key technology of personalized information service: cluster analysis [J]. Modern Intelligence,2007,27(10):212-214.
- [7] Cao, K., Xu, T., Liu, Y. et al. Review of cluster analysis [J]. Chinese Journal of Intelligent Health,2016,2(10):50-53.
- [8] Duan Nannan. Research on Problems based on Regression Analysis and its Implementation in MATLAB [J]. Journal of Luoyang Normal University,2014(5):33-35.
- [9] Jia, J., He, X. & Jin, Y. Statistics[M].Beijing, China Renmin University Press,2018.