Factors Affecting China Birth Rate: Taking Shanghai as an Example

Zhongxi Zhu

Department of Computing Mathematics, College of Science, City University of Hong Kong, Hong Kong, 999077, China

zhongxzhu2-c@my.cityu.edu.hk

Abstract. China's birth rate has fallen off a cliff in the past few years, and the total birth rate has fallen below 10 per thousand. On January 17, 2022, the Chinese National Bureau of Statistics released China's economic data for 2021. The annual birth population was 10.62 million, and the birth rate was 7.52‰. But in 2016, that figure was still 13.57‰. Previous studies have explored various influencing factors, but further research is needed on the degree of influence of multiple factors on the birth rate. Based on previous research, this study adopts mathematical-statistical methods such as principal component analysis, correlation analysis, stepwise multiple regression, and ridge regression to analyze more influencing factors to explore and test the effect of these factors on the birth rate and uses the data of the Shanghai Bureau of Statistics from 2002-2021 for 19 years to try and judge which factors significantly affect the birth rate under the joint action. The validity of the judgment conclusion is tested by various data from Beijing and Guangzhou in 2021. Results of the study found that the per capita GDP, unit housing price, per capita disposable income, and resident consumption index do have an impact on the birth rate. Still, the government's financial investment and expenditure correlate highly with the birth rate data. This may be because government revenue and expenditure are more closely related to other factors affecting the birth rate. Other factors were not very closely related to the birth rate.

Keywords: Birth rate, relevance, influencing factors, mathematical-statistical methods, Shanghai.

1. Introduction

Compared with the vast amount of data in the whole of China, this study chooses Shanghai as the city with the highest internationalization and the most developed information in China; its data has a specific index significance. Moreover, the data collection is more accurate and diverse, making it more convenient for subsequent mathematical and statistical analysis. Both the birth rate and the fertility rate are quantitative indicators that reflect the fertility level in this region. Therefore, the factors that affect both are considered in this study when seeking the influencing variables.

In the study, this paper collected data on Shanghai's birth rate since 1980 [1]. The reason why the last 20 years (2002-2021) of data were used for statistical analysis is that compared with the past, in 2002, it is more time-sensitive. The fertility rate showed a significant upward trend in 2016 and a substantial decline from 2016-2021. Moreover, the rest of the supporting data were relatively complete after 2000, so using the 20-year data from 2002 to 2021 to analyze the factors affecting the fertility rate is sufficient to show the trend in quantity [2].

In other words, this study needs to find suitable variables that affect the birth rate to explore whether they do affect the birth rate and the magnitude of their effect on the birth rate. Birth rates have rebounded in high-income cities in China over the past few years, such as major cities along the coast [3]. However, as the city with the highest per capita income in mainland China, Shanghai's birth rate has shown a downward trend. Per capita income can be seen as an essential factor. The education level of women is also related to the birth rate. Some studies have shown that women's education level (education time) greatly influences fertility, showing a significant negative correlation, which also requires setting variables [4]. Due to the difficulty of data collection, this study set this variable as the number of college students in school. The proportion of women admitted to college students in China has been relatively stable, reflecting the number of women in higher education to a certain extent. The more college students there are, the more women are in higher education.

In the study, it was observed that China's fertility policy is a very significant factor, that is, the regulation that a couple in China can only have one child. A comprehensive analysis method can explain the decline in China's birth rate after 2002 to a certain extent. Trends. However, this policy was canceled in 2015 and cannot fully explain the downward trend in China's birth rate after 2016. There is a specific link between banking and financial savings of Chinese families and having a second child [5, 6]. In turn, personal savings may also have a particular impact on the birth rate. Under the background of the current low birth rate, China tends to enter an aging society like Japan. In this case, the financial burden will increase, and the increase in pension subsidies will affect the birth subsidy and further affect the birth rate [7, 8]. Therefore, the government's fiscal revenue and fiscal expenditure may also have an impact on the birth rate.

Housing prices may also be an essential factor affecting the birth rate, and this manifestation can be expressed through intuitive mathematical simulations on a material basis. Female fertility also cannot be ignored, which may be related to air pollution levels or environmental factors such as temperature [9, 10]. The decline in human fertility is due to changes in the natural environment and not just sociological changes [11]. In addition, the most likely factor affecting the birth rate is income and expenditure, and there may be some inherent relationship between GDP per capita and the fertility rate [12]. Several factors actually affect each other. Income, housing prices, and consumer spending may all be related to the birth rate [13].

This paper combined with previous research to explore the factors that affect the birth rate with the following data from Shanghai on the websites of the National Bureau of Statistics of China and the Shanghai Municipal Bureau of Statistics for analysis: Shanghai population growth rate, population fertility rate, population mortality rate, Shanghai Resident population, Shanghai urban per capita disposable income, Shanghai rural per capita disposable income, Shanghai per capita income, Shanghai fiscal revenue, Shanghai fiscal expenditure, Shanghai unit price per square meter, Shanghai Gross Domestic Product (GDP), Shanghai per capita GDP, The number of college students in Shanghai, Shanghai Consumer Price Index. These factors have been analyzed in previous sociological studies on fertility and the birth rate. I will combine these factors with the birth rate data in Shanghai over the years and screen out the obvious factors. Finally, the most prominent factors affecting the birth rate were screened to fit the birth rate data of two other big cities in China-Beijing and Guangzhou, in 2021 to test and judge the effect.

2. Methodology

Its essence is to seek a comprehensive substitute object for related variables by correlating the original variables about the birth rate and ensuring the minimum loss of information in the transformation process.

Calculate the covariance matrix R from the standardized dataset:

$$R = (r_{ij})_{n\times n} = \begin{bmatrix} r_{11}\, r_{12} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1}\, r_{n2} & \cdots & r_{nn} \end{bmatrix} \tag{1}$$

Calculate the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$ of the matrix R and the corresponding eigenvectors $u_1,\ u_2,...,\ u_n$, where $u_j = (u_{1j},\ u_{2j},\ ...,\ u_{nj})$, u nj represents the nth component of the jth eigenvector; n new indicator variables consisting of the eigenvectors:

$$\begin{gathered} y_1 = u_{11}\overline{x_1} + u_{21}\overline{x_2} + \cdots + u_{n1}\overline{x_n} \\ y_2 = u_{12}\overline{x_1} + u_{22}\overline{x_2} + \cdots + u_{n1}\overline{x_n} \\ \cdots \\ y_n = u_{1n}\overline{x_1} + u_{2n}\overline{x_{12}} + \cdots + u_{nn}\overline{x_{nn}} \end{gathered} \tag{2}$$

In the formula, y_n is the nth principal component. Calculate the contribution rate b_j (j=1, 2, ..., n) of each main component y_j and the cumulative contribution rate αp of $y_1,\ y_2,...,\ y_n$ (p ≤ n).

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{n}\lambda_k}(j = 1,2,3,4,\dots,n)\ \ \alpha_p = \frac{\sum_{k=1}^{p}\lambda_j}{\sum_{k=1}^{n}\lambda_j}(p \le n) \tag{3}$$

The data related to the birth rate Pearson correlation coefficient is required to fit the normal distribution. The Spearman correlation coefficient is used when the data related to the birth rate is unable to provide the normal distribution.

Here we define the Pearson correlation coefficient between two variables as the ratio of the covariance to the standard deviation between the two variables:

$$\rho_{X,Y}\frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X} \tag{4}$$

Equation (4) defines the overall correlation coefficient, represented by the symbol ρ. The Pearson coefficient can be calculated by estimating the standard deviation and covariance of the sample. Here, we use the lowercase letter t to denote:

$$t = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{5}$$

Equal to:

$t = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma_X})\left(\frac{Y_i-\bar{Y}}{\sigma_Y}\right)$ is respected for X_i the sample standard deviation.

The Spearman correlation coefficient can be thought of as a rank variable between the Pearson correlation coefficients. A sample size of n samples, n raw data are converted into rank data, and the correlation coefficient is:

$$\rho = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{6}$$

In this research, in order to simplify the calculation process, the connection between the variables affecting the birth rate needs to be ignored, so ρ can is obtained through simple steps. Then ρ is:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{7}$$

Linear regression can be solved by the least squares method or gradient descent method, and a straight line for $y = bx + a$ can be calculated. Take the least squares method as an example. Assuming that there are x_1, x_2, ..., x_k, and the following linear relation can just be considered:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{8}$$

The uncorrelated random variables are all the same distribution random variable. To express the above equation in a matrix, let:

$$Y = X\beta + \varepsilon \tag{9}$$

Solution using least squares:

$$\hat{\beta} = (X^T X)^{-1}X^T Y\ \ u(b) = \frac{\sigma}{\sqrt{\sum(x-\bar{x})^2}} \tag{10}$$

In this research, the stepwise regression adopts the stepwise screening method: when a variable is introduced, first check whether the variable causes a significant change in the model (F test). When there is no substantial change due to the introduction of a variable added later, this variable is dropped to ensure that each time a new variable is introduced, only the variable with a relatively significant influence is included in the regression equation. Until neither critical explanatory variable are selected for the regression equation nor non-significant explanatory ones are eliminated from the regression equation. Finally, an optimal set of variables is obtained.

In general, the goal of solving the above regression problem using the least squares method is to minimize the following equation:

$$\hat{\beta} = argmin_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i)^2 \quad (11)$$

Ridge regression adds a penalty term to the minimization objective that affects the birth rate, where the parameter λ is a quantity to be determined. Ridge regression is a special kind of least squares regression with two norm penalties.

$$\hat{\beta}^{bridge} = argmin_{\beta} \{\sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2\} \quad (12)$$

For data fitting to judge which factors have a more significant impact on fertility. I chose to use SPSS as an analysis tool for data fitting in the software.

3. Results and Discussion

To use principal component analysis, first determine the number of principal components; here, we need to draw the scree plot, observe and analyze the scope. The scree plot:

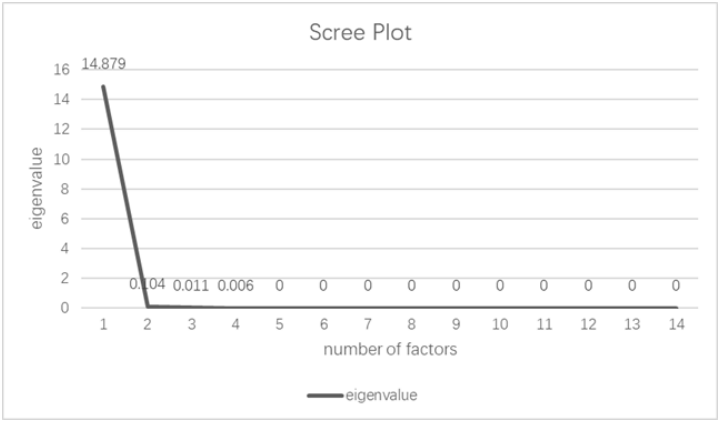


Figure 1. The scree plot of PCA

When the polyline suddenly becomes stable from steep, the number of factors corresponding to the steepness of the sound is the number of reference extraction factors. It can be seen from the figure that starting from the second principal component, the eigenvalue of the principal component begins to decrease slowly, and we can choose two principal components when the contribution rate of the cumulative explanation of the principal component reaches 90%.

Table 1. KMO Test

KMO and Bartlett test		
KMO		0.654
Bartlett test	Approx. Chi-Square	871.081
	df	105
	p value	0

The premise requirement that the principal component analysis is suitable is that the value of KMO is more significant than 0.6. It can be seen from the table above that the value of KMO this time is 0.716>0.6, which meets the initial conditions. To explore whether the data is suitable for principal component analysis, the Bartlett spherical test is also required. It can be seen from the above table that the p-value is 0<0.05, which is suitable for this analysis.

The formula of the model:

$$F1 = -0.067 \times Year + 0.067 \times growth\ rate + 0.067 \times birth\ rate + 0.067 \times resident\ population + 0.067 \times Death\ rate + 0.067 \times per\ income + 0.067 \times GDP + 0.067 \times per\ GDP + 0.067 \times per\ disposable\ income\ of\ rural + 0.067 \times per\ disposable\ income\ of\ urbon + 0.067 \times units\ house\ price + 0.067 \times College\ students + 0.067 \times General\ budget\ revenue + 0.067 \times General\ budget\ expenditure + 0.067 \times CPI \quad (13)$$

$$F2 = 0.305 \times Year + 2.404 \times growth\ rate - 0.305 \times birth\ rate - 0.305 \times resident\ population + 1.621 \times Death\ rate - 0.305 \times per\ income - 0.305 \times GDP - 0.305 \times per\ GDP - 0.305 \times per\ disposable\ income\ of\ rural - 0.305 \times per\ disposable\ income\ of\ urban - 0.305 \times units\ house\ price - 0.203 \times College\ students - 0.305 \times General\ budget\ revenue - 0.305 \times General\ expenditure\ budget - 0.368 \times CPI \quad (14)$$

From the above one can get:

$$F = (0.992/0.999) \times F1 + (0.007/0.999) \times F2 \quad (15)$$

3.1. Normality test and correlations

A normality test should be carried out when determining the correlation coefficient. Variables with normality are suitable for Pearson Correlation, and variables without normality are relevant for Spearman correlation.

Table 2. Normality test

Items	Mean	Std.	kurtosis	Kolmogorov-Smirnov test		Shapiro-Wilk test	
				Statistic D	p	Statistic W	p
Year	2011.5	5.916	-1.2	0.077	0.991	0.96	0.551
birth rate	6.722	1.343	-0.582	0.087	0.958	0.972	0.8
Death rate	5.749	0.842	10.254	0.246	0.003**	0.69	0.000**
resident population	2243.967	274.455	-0.893	0.222	0.011*	0.819	0.002**
per income	83025.405	45568.173	-0.804	0.104	0.853	0.947	0.353
growth rate	-1.142	1.113	0.011	0.154	0.243	0.944	0.28
per GDP	89515.789	38750.115	-0.984	0.092	0.943	0.947	0.345
GDP	21848.651	11647.916	-1.094	0.1	0.866	0.947	0.327
per disposable income of rural	41841.57	22258.872	-1.128	0.123	0.588	0.931	0.159
per disposable income of urban	18658.491	10237.822	-0.966	0.127	0.545	0.926	0.13
units house price	15565.352	9039.22	-0.714	0.156	0.262	0.929	0.166
General budget expenditure	4455.809	2696.505	-1.451	0.124	0.578	0.902	0.045*
College students	48.776	5.533	2.598	0.308	0.000**	0.782	0.000**
General budget revenue	3954.031	2407.704	-1.464	0.146	0.322	0.912	0.069
CPI	102.188	1.519	0.93	0.153	0.256	0.95	0.373

Year, birth rate, per income, growth rate, per GDP, GDP, per disposable income of rural, per disposable income of urban, units of house price, General budget revenue, CPI, a total of 11 items did not show significant ($p>0.05$), which means that Year, birth rate, per income, growth rate, per GDP, GDP, per disposable income of rural, per disposable income of urban, units house price, General budget revenue, and CPI have typical characteristics.

The normality here is not strictly required by definition; as long as the absolute value of kurtosis is less than ten and the total value of skewness is less than 3, it can be considered to be a normal distribution in principle by this research.

Summary: Death rate, resident population, General budget expenditure, and College students do not have normality characteristics. In addition, Year, birth rate, per income, growth rate, per GDP, GDP, per disposable income of rural, per disposable income of urban, units of house price, General budget revenue, and CPI have normal characteristics.

Table 3. Pearson Correlation

Pearson Correlation	Coefficient
growth rate	0.82
Year	0.28
per income	0.35
per disposable income of rural	0.16
units house price	0.31
CPI	0.52
per disposable income of urban	0.12
per GDP	0.39
GDP	0.19
general barget revenue	0.25
	birth rate

Table 4. Spearman Correlation

Spearman Correlation	Coefficient
Death rate	0.71
general budget expenditure	0.33
resident population	0.33
College students	0.19

From the above tables of the Pearson correlation coefficient and the Spearman correlation coefficient, the population growth rate, death rate, and CPI index have a strong correlation; the rest like Shanghai GDP, Shanghai urban per capita disposable income, the number of college students, and Shanghai rural per capita disposable income have weaker correlations. Similarly, it can be seen from the Spearman coefficient that there is a high correlation between the birth rate and the death rate.

3.2. Regression analysis

Take birth rate as the only dependent variable, perform multiple regression analysis on the remaining data as independent variables, and then remove the variables with weaker correlations and continue the research. After the results are obtained, the variables with moderate correlation will continue to be eliminated, and the operation observation data analysis will be carried out continuously.

Table 5. Parameter Estimates (n=19)

	Unstandardized Coefficients		Standardized Coefficients	t	p	VIF	F/R^2
	B	Std. Error	Beta				
Constant	8.813	5.379	-	1.638	0.162	-	
resident population	-0.001	0.003	-0.262	-0.38	0.72	766.114	
growth rate	0.836	0.096	0.742	8.721	0.000**	11.62	
per income	0	0	-1.093	-1.044	0.344	1758.262	
Death rate	0.302	0.167	0.203	1.809	0.13	20.256	
per GDP	0	0	-4.804	-1.214	0.279	25139.035	
GDP	0.001	0.001	6.955	1.192	0.287	54596.445	
per disposable income of urban	0	0	-2.009	-0.927	0.397	7535.127	F (13, 5) =123.053, p=0.000 R^2=0.997
per disposable income of rural	0	0	0.709	0.331	0.754	7357.994	
units house price	0	0	-0.028	-0.095	0.928	135.864	
General budget expenditure	-0.001	0.001	-1.155	-1.049	0.342	1946.177	
CPI	-0.007	0.029	-0.008	-0.245	0.816	1.9	
General budget revenue	0.001	0.001	1.934	1.756	0.139	1945.907	
College students	0.039	0.071	0.166	0.548	0.607	147.334	

It can be seen from the simulation that the growth rate will have a remarkable positive influence on the birth rate. Still, we know that because the birth rate itself is related to the data on the population

growth rate, that is, the birth rate minus the population death rate is equal to the population growth rate. Therefore, the influence of excessive population growth rate should be considered to be excluded.

After analyzing the variables related to the birth rate in the above table, the model formula is:

$$birth\ rate = 14.616 - 0.004 * resident\ population - 0.0001 * per\ income + 0.691 * Death\ rate - 0.0001 * per\ GDP + 0.0003 * GDP - 0.0001 * per\ disposable\ income\ of\ urban + 0.0001 * per\ Disposable\ income\ of\ rural + 0.0001 * units\ house\ price - 0.004 * General\ budget\ expenditure - 0.047 * CPI + 0.005 * General\ budget\ revenue + 0.110 * College\ students \quad (16)$$

The model R-square value is 0.949, which means resident population, per income, Death rate, per GDP, GDP, per disposable income of urban, per disposable income of rural, units house price, General budget expenditure, CPI, General budget revenue, College students can explain 94.9% of the transformation in the birth rate.

When the F-test was performed on the model, which passed the F-test (F=9.398, p=0.006<0.05), this means that the resident population, per income, Death rate, per GDP, GDP, per disposable income of urban, per disposable income of rural, units house price, General budget expenditure, CPI, General budget revenue, and College students have an effect on the birth rate, besides, the multicollinearity test of the model shows that there is a collinearity problem, which means the VIF value in the model is greater than 10, we still need to use ridge regression or stepwise regression to dislodge the disturbance.

General budget expenditure will have a remarkable negative impact on the birth rate, with the regression coefficient value of General budget expenditure being -0.004 (t=-2.946, p=0.026 < 0.05). General budget revenue has a significant positive impact on the birth rate, with the regression coefficient value of General budget revenue being 0.005 (t=3.756, p=0.009<0.01).

3.3. Stepwise regression

Stepwise regression requires retaining some variables that have been systematically analyzed with high correlation and deleting some variables that are less correlated in the regression.

From the analysis of the results of the F test, the P value is 0, the level is significant, and the null hypothesis that the regression coefficient is 0 is rejected. For variable collinearity performance, if the variables GDP, per disposable income of rural, General budget revenue, resident population, General budget expenditure, per disposable income of urban VIF are greater than 10, there is a collinear relationship, and the independent variables of collinearity are simply removed or Perform ridge regression.

The formula of the model is as follows:

$$Y(birth\ rate) = -1 + 0.00167 * GDP + 0.00167 * per\ disposable\ income\ of\ rural + 0.00167 * General\ budget\ revenue + 0.00167 * resident\ population + 0.00167 * General\ budget\ expenditure + 0.00167 * per\ disposable\ income\ of\ urban \quad (17)$$

3.4. Ridge regression

Ridge regression introduces k unit arrays. The introduction of unit arrays will lead to some loss of variable information. Still, at the same time, it can be exchanged for more reasonable estimates of the birth rate regression model.

Before the ridge regression analysis, it needs to confirm the K value in combination with the trace diagram.

The finding rule of the K value is the minimum K value when the standardized regression coefficient of each independent has more stability. The smaller the value of K, the smaller the deviation. The ridge trace plot was obtained by performing Ridge regression analysis on birth rate as the dependent variable. When the K value is 0.001, the standardized regression coefficient of the independent variable tends to be stable, so it is recommended to set the best K value. Take it as 0.001.

Table 6. ANOVA test

	Sum of Squares	df	Mean Square	F	p-value
Regression	26.913	12	2.243	4.637	0.036
Residual	2.902	6	0.484		
Total	29.815	18			

The ANOVA (F)test is used to determine whether the model is meaningful. It was found that the model passed the ANOVA test (F=4.637, p=0.036<0.05) p-value is less than 0.05, which means the model is meaningful. The Death rate, per income, GDP, resident population, per GDP, per disposable income of rural, per disposable income of At least one of urban, units of house price, College students, General budget expenditure, CPI, and General budget revenue will have an impact on the birth rate.

The model formula is:

$$birth\ rate = -6.948 + 0.152 * Death\ rate - 0.0001 * per\ income + 0.0001 * GDP + 0.005 * resident\ population + 0.0001 * per\ GDP + 0.0001 * per\ disposable\ income\ of\ rural - 0.0001 * per\ disposable\ income\ of\ urban - 0.000 * units\ house\ price + 0.020 * College\ students - 0.0001 * General\ budget\ expenditure + 0.020 * CPI + 0.002 * General\ budget\ revenue. \quad (18)$$

Table 7. ANOVA chart

	Sum of Squares	*df*	Mean Square	*F*	*p-value*
Regression	0	9	0	0	1
Residual	29.815	9	3.313		
Total	29.815	18			

It can be seen from the above table that the model did not pass the F test equals 0, which means that the death rate, units of house price, per GDP, per income, College students, General budget revenue, CPI, General budget expenditure, and resident population have no impact on birth rate, so the effect of the birth rate on dependent variables cannot be precisely analyzed.

3.5. Tested with the data of Beijing and Guangzhou

The following is the data about Beijing and Guangzhou in 2021. We need to bring these data into formulas (16), (17), and (18) to test whether the birth rate of the year can be better matched.

Table 8. Data of Guangzhou 2021 Part1

Year	growth rate (‰)	birth rate (‰)	resident population (10000 people)	Death rate (‰)	per income One person
2021	6.26	11.82	1881.06	5.46	144288

Table 9. Data of Guangzhou 2011-2021 Part2

GDP Guangzhou	per GDP One person	per disposable income of rural One person	per disposable income of urban One person
28231.97	150366	34533	74416

Table 10. Data of Guangzhou 2021 Part3

units house price RMB/m^2	College students 10000 persons	General budget revenue 100 million RMB	General budget expenditure 101 million RMB	CPI %
42959	141.26	1883.18	3020.72	101.1

Table 11. Data of Beijing 2021 Part 1

growth rate (‰)	birth rate (‰)	resident population (10000 people)	Death rate (‰)	per income One person	GDP Beijing	per GDP One person
0.96	6.35	2188.6	5.39	194651	40269.6	184000

Table 12. Data of Beijing 2021 Part 2

per disposable income of rural One person	per disposable income of urban One person	units house price RMB/m^2	College students 10000 persons
33303	81518	66299	59.6

Table 13. Data of Beijing 2021 Part 3

Year	General budget revenue 100 million RMB	General budget expenditure 101 million RMB	CPI %
2021	5932.3	7205.1	101.1

It can be found that after the formula is substituted, only the data Beijing in (16) and Guangzhou in (18) in terms of numerical values are close to the birth rate of that year, but the gap is still about 2 points. If it is expanded to the data of Beijing and Guangzhou from 2011-2021, it can be found that the effect of these formulas is not ideal, and the error is usually around 3‰.

4. Conclusion

Through mathematical and statistical research, it can be found that the above factors all impact the birth rate with different effects. The government's fiscal expenditure and income do affect the birth rate, similar to the factors mentioned in previous studies, such as personal GDP. The intrinsic connection of these factors interferes with the model's establishment. However, it can still be found through the data that there is a specific relationship between Shanghai's financial revenue and Shanghai's birth rate. Other negative correlations, such as the increase in house prices and changes in other variables, can be found to have statistical associations and influences. Still, they are not decisive factors for the birth rate. From a general rational point of view, there is no recognized relationship between death and birth rates. However, this study still judges that it shows a negative correlation in mathematics. It can be said that this is somewhat a rough result. Since the birth rate data's storage capacity is not large, it is impossible to expand further the database for calculation, which is also the shortcoming of this study. Questions about births are conducted on a household basis; it may be more accurate to use households as a unit in data collection. So, data such as per capita income may also be refined to household per capita income in future research. For this study, the above factors have a statistical and mathematical effect on the birth rate. But it is difficult to judge the impact of multiple factors. The data of Shanghai can only be used as a model, but the fitting effect for other large cities in China is not ideal. In addition to mathematical and statistical factors, sociological factors are emphasized in studying birth rates. The specific association of these factors with birth rate and the treatment details require follow-up research.

References

[1] Lucero-Prison, D. E., et al., "The challenges of the low birth rate in China," Public Health Challenges, 1(2): 8 (2022).

[2] Finlay, J. E., Mejía-Guevara, I., and Akachi, Y., "Inequality in total fertility rates and the proximate determinants of fertility in 21 sub-Saharan African countries," PloS one, 13(9): 2033 (2018).

[3] Cheng, H., et al., "Global trends in total fertility rate and its relation to national wealth, life expectancy and female education," BMC Public Health, 22(1): 1-13 (2022).

[4] Chen, N. and Xu, H., "Why has the birth rate relatively increased in China's wealthy cities," (2021).

[5] Koppes, D. R., "Influence of GDP per Capita on Birth Rates," Proceedings of the Jepson Undergraduate Conference on International Economics. 4(1): 1 (2022).

[6] Zhao, F., "The Impact of Socioeconomic Factors on Birth Rate and Aging Population in China," (2021).

[7] Brown, S., "Population Aging and Healthcare Costs in China and Japan," (2022).

[8] Gietel-Basten, S., Han, X. and Cheng, Y., "Assessing the impact of the "one-child policy" in China: A synthetic control approach, PLoS One, 14(11), e0220170. (2019).

[9] Clark, W. A.V., Yi, D. and Zhang, X., "Do house prices affect fertility behavior in China? An empirical examination," International Regional Science Review, 43(5): 423-449 (2020).

[10] Boland, M. R., "A model is investigating environmental factors that play a role in female fecundity or birth rate," Plos one, 13(11): e0207932. (2018).

[11] Skakkebæk, N. E., et al., "Environmental factors in declining human fertility," Nature Reviews Endocrinology, 18(3), 139-157 (2022).

[12] Baker, S. R., et al., "Fertility and Savings: The effect of China's two-child policy on household savings," National Bureau of Economic Research, (2022).

[13] Wu, J., "Prediction of Birth rate in China under three-child Policy based on neural network," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, 1652-1655 (2022).

[14] Su, L., et al., "Evolution Analysis of Large-Scale Spatiotemporal Configuration of Provincial Birth Rate in China," 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 217-222 (2022).