Research on Factors Affecting China's Birth Rate and The Relation Between Economic Development and The Birth Rate

Xinyu Yang*

Beijing Royal School, Beijing, China

* Corresponding Author Email: yangxinyu@whsu.edu.cn

Abstract. Many scholars have previously studied the birth rate and its development, but there are still gaps in the research on which factors play a dominant role in the birth rate. Therefore, this paper will provide a clear study of this issue, using China as an example to help major countries solve the problem of low birth rates and economic development. The population issues caused by low birth rate were researched and the main factors for birth rate were found from research. The decisive factors that determine the birth rate in China were analyzed. The link between birth rate and economic development is examined. This study figures out that the two main factors that influence the birth rate are the divorce rate and the female employment rate. There is a relatively strong correlation between the birth rate and economic development. The formula and the principle of Vector Autoregressive (VAR), Bayesian Information Criterion (BIC) and Times Series Graphical lasso (TSGlasso), were researched and used.

Keywords: Birth rate, China, decisive factors, VAR mode, TSGlasso, BIC, Economic development.

1. Introduction

The low birth rates have become a more widespread and serious problem. Many countries are experiencing low birth rates, economic and social deterioration problems, like Singapore, Japan and South Korea [1]. China is also a country that has a low birth rate problem. Therefore, the Chinese government is aware of this problem of negative population growth and liberalized the two-child policy in 2016 and wants to raise the birth rate of the country.

However, the implementation of the two-child policy has not led to a significant increase in the population and the good impact of the two-child policy on China's birth rate is only short-lived [2]. Also, many people believe that more government subsidies for large families will be effective in alleviating the low birth rate, but this is not the case.

The low birth rate might lead to a vast number of economic problems, such as a decline in employment, innovation and also a higher pressure on local government [3]. In addition, there are also several social issues such as the pressure of living and the inability to raise children [4]. Thus, this article will analyze the factors that lead to the change of birth rate.

A large number of factors can contribute to low birth rates, such as religious beliefs, female employment, gender ratio and other relative factors that might all have contributed to the decline in population, demonstrated by Bailey et al. [5]. This paper will first introduce the population issues brought on by a low birth rate in section 2, which shows the emergency of the event to the world. Then, this paper will choose China as a good example for investigation and analysis of the impact of the low birth rate on China. Secondly, the paper will visualize and study the data using the VAR model, BIC and TSGlasso methods in section 3. In section 4, the paper concludes and writes up the final results of the study. Finally, the reference is placed in the end section.

2. Literature Review

2.1. Population Issues

Population issues are various problems that affect the survival and development of a population, such as low birth rates [5]. These issues have always been closely related to the economic development, social stability and ecological environment of a region, and bring great concern to all countries in the world.

2.1.1 The impact of the low birth rate in the world

The economic problems that might be caused by the low birth rates in both the short-run and the long-run. In the short run, a low birth rate is not only bad for society in some respects [3]. It also brings some benefits, such as the possibility of higher per capita wage payments, making the country more welcoming to immigrants and reducing pressure on the environment. But in the long run, a low birth rate may lead to less innovation in the economy as a whole. Not only that, but the government would also be under more pressure to pay for overspending on pensions and health services.

Many famous people and scholars throughout history have studied issues related to population and birth rates. Mr. Musk has appeared in the news to warn the world of the dangers posed by the collapse of global birth rates [7]. One of the greatest risks to civilization is a low and rapidly declining birth rate [7]. While Mr. Musk may be the best-known person to raise the alarm about depopulation, more and more politicians and economists are beginning to realize the reality that a world without newborns poses huge challenges and risks [7].

Why is a declining population due to a declining birth rate a bad thing? What Mr. Musk acknowledges is that, even from a purely practical point of view, the population is a major asset. That is, people are not just resource-consuming creatures. People also use their creativity and imagination to create things that are better for the world and make the world a better place. Also, economic development and the low birth rate may be mutually beneficial. For example, the increased pressure exerted on families by the economy as a result of the low birth rate may have led to an awakening of women's sense of self-independence. They are seeking independence and a high quality of life. They begin to be open to marriage and give up the option of childcare. This is likely to lead to a low birth rate once again. This also leads to the fact that a low birth rate is likely to bring negative effects to most countries, as well as China.

2.1.2 The data on low birth rate in China

China is one of the most populous of the many countries. Predicting of Birth rates are declining globally (2021), illustrates that it could lose between 600 million and 700 million people in population by 2100, which means that China's population in 2100 will likely be half the size of its population in 2021 [7]. Not only that but also the Chinese population will be 141.6 million at the end of 202 [4]. It is 480,000 more than at the end of 2020, which means the Chinese population is 1.4 million less than in 2020. The birth rate is 7.52 per 1,000 population.

2.2. Several Factors Influence the Birth Rate in China

In this section, the paper will analyze several factors that influence the birth rate. This will help the paper to complete the identification of the most influential factors in China's birth rate and the identification of solutions and policies that will ultimately address the problem of low birth rates.

2.2.1 Economic factors and birth rate in China

With the rapid development of China's economy,women are entering the workforce, which leads to an increase in the female employment rate. As more and more women are working, they may not want to get married and have children. Thus, the employment rate of Chinese females may also be a factor in the birth rate of the population [9].

Also, there might be a strong link between the Chinese birth rate and housing prices. This is because there might be a strong correlation between the pressure on the lives of the population and their desire to raise children.

2.2.2 Social factors and birth rate in China

Firstly, gender distribution may play a decisive role in determining the birth rate in China. Many data show that in China, the ratio of males to females at fertile age is usually greater than the ratio of males. Many scholars have studied the gender ratio and have analyzed the relationship between the gender ratio and the birth rate. Marriage may also be an important relevant factor in addressing birth rates. The relationship between birth and divorce rates has been investigated by many academics before [9]. Domestic violence and family discord might have the potential to stop people from getting married. At the same time, these family problems can also have a deep impact on the children in the family. Children may also no longer be attracted to marriage.

2.3. The Methods and Mathematical Models Used by Previous Scholars

A vast number of models have been used frequently in the study of statistical demographic problems. They are logistic models, deep learning models, regression models and time series models [6]. The paper will analyze and compare the pros and cons of these models, which have been used more often in the study of birth rates and apply them to solve the problem of the low birth rate in China.

2.4. Autoregressive Integrated Moving Average（ARIMA）

Autoregressive integrated moving average (ARIMA) is one of the most common types of time series model which is the process of recording data over time. ARIMA can be used to predict the future course of events [10]. For ARIMA models, there must be part of the necessary process to define the suitability of the data. However, ARIMA is relatively rare to use in multivariate data [11]. As a result, the paper will use the Vector Autoregressive (VAR) model for the investigation of the low birth rate in China.

3. Data Collection and Analysis

In this section, the paper will investigate the relationship between birth rate and other socio-economic factors. The information on the correlated variables is from [5]. The variables can be divided into two parts. One part is social factors which conclude the gender ratio and divorce ratio. Another part is economic factors which contained female employment rate and housing prices. The variables so far were discussed in section 2.2. Data on the variables are mainly collected from China Statistical Yearbook data. The paper will use the PSC, BIC and TGlasso which are based on VAR to analyze the data and get the most important factors which bring the relatively largest impact on the birth rate.

3.1. Data Analysis

3.1.1 Birth rate

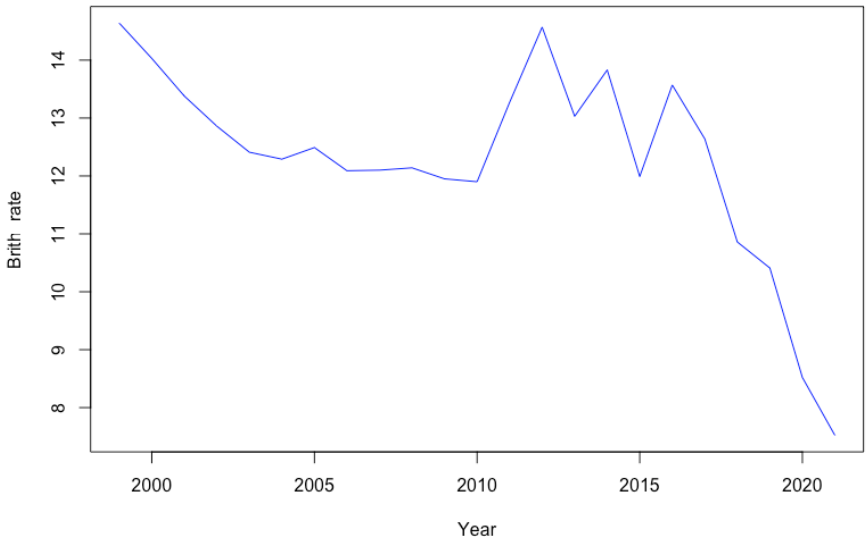

Fig. 1 Birth rate data of China from 1999 to 2021

As the figure 1 above indicates, there is an overall downward trend in the birth rate. The birth rate has a flat downward trend interval from 2000 to 2010. Also, as the Chinese government introduced family planning in around 1980, it can be seen from the graph that the family planning policy has had the effect of reducing the birth rate. This can be seen in the period from approximately 2010 to 2017 and the birth rate has been very volatile. This is because this period was followed by the opening of the two-child policy in 2016. However, as shown in the figure 1, the two-child policy appears to have been effective only in the short term, with a peak in the birth rate between 2016 and 2017 and a steady decline there. Therefore, these periods will be important objects to observe when analyzing the data.

3.1.2 Gender ratio

Fig. 2 Gender ratio of China from 1999 to 2021 (unit: female=100)

The birth rate may be linked to marriage. Take out the years from the 2010s to 2017s for observation. First of all, as shown in figure 2 the gender ratio shows a relatively flat gradient over this period, which may indicate that the gender ratio has not changed much over this period. However, as shown in figure 1, the birth rate has not been stable over the period from 2010 to 2017. Therefore, the gender ratio may not have had a decisive effect on the birth rate over this period.

3.1.3 Female employment rate

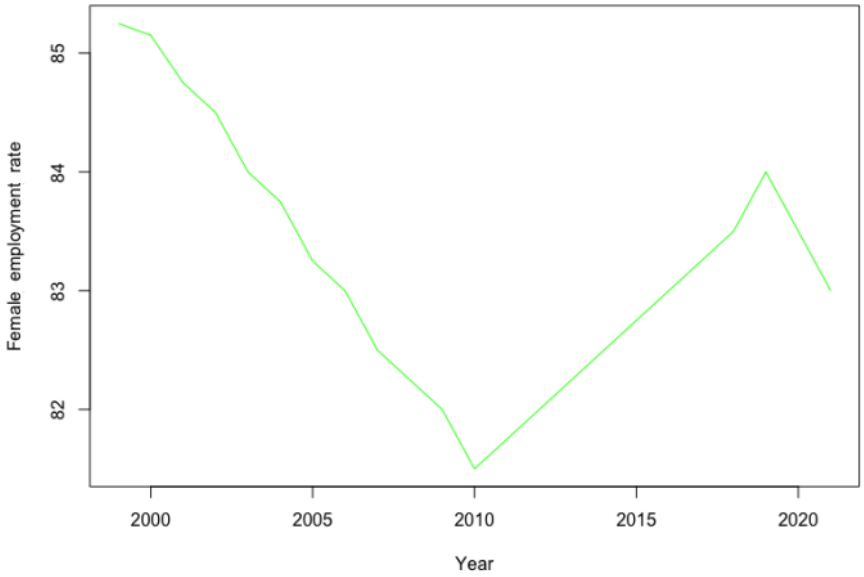

Fig. 3 Female employment rate of China from 1999 to 2021(unit: male=100)

As figure3 above illustrates, the employment rate for women has increased gradually over the period from 2010 to 2017. However, the period from 2010 to 2017 in figure 1 is choppier. Therefore, there may not be a visible relationship between female employment and birth rate interval from 2010 to 2017. In the graph, there is a significant increase in the female employment rate after about 2010 due to the implementation of the Women's Development Programmer policy in China around 2011 [12]. There might be a negative relationship between the female employment rate and the birth rate interval [2000,2010].

3.1.4 Divorce rate

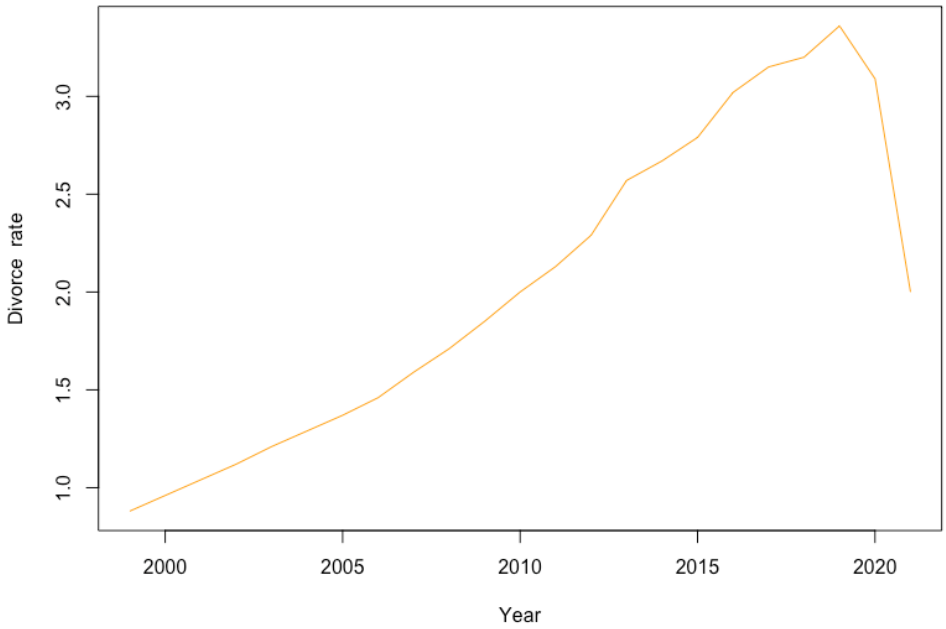

Fig 4 Divorce rate of China from 1999 to 2021(sample=100, unit: ‰)

Over the period 2000 to 2010, the divorce rate and the birth rate may be inversely proportional. This is because, when the divorce rate rises, the birth rate falls. This might make logical sense, because when the divorce rate rises, people may become disillusioned with marriage and therefore not want to have children. However, between 2010 and 2017, there was no clear relationship between the birth rate and divorce rate.

3.1.5 Housing price

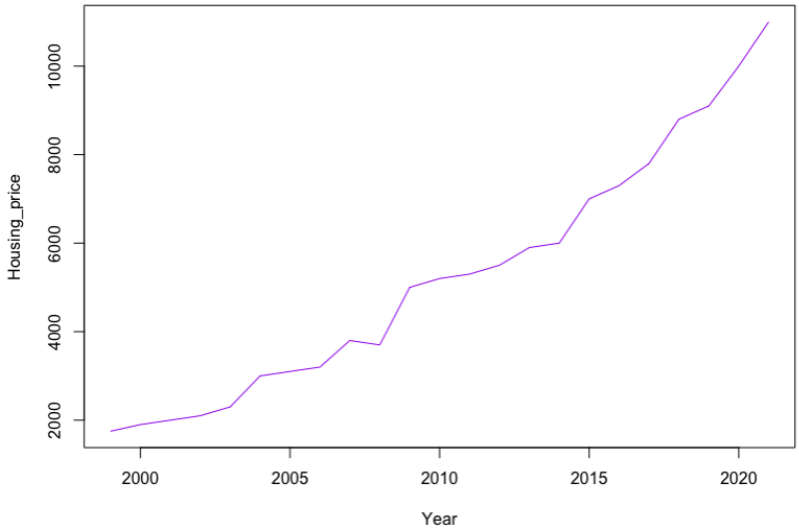

Fig. 5 Housing price of China from 1999 to 2021(unit: rmb/m^2)

From figure 5, house prices have been rising in a flat trend. Between 2000 and 2010, the housing price and the birth rate have a negative relationship by linking figure 5 and figure 1. Logically speaking, a negative relationship is plausible. This is because when housing prices rise, the pressure on people's lives may be gradually increasing. This may lead to people feeling that they cannot afford to raise a child. As a result, the birth rate might fall if the housing price increases.

3.2. Analytical Methods

The paper will use the TSGlasso based on the VAR model. Different stages for investigation and calculation are shown below.

Stage 1: Set an initial algebraic model with the AR coefficient and intercept coefficient (Section 3.3 for details)

Stage 2: Apply the Bayesian information criterion (BIC) by using R to obtain the value of p which is the number of years for data of variables that the investigation should take consideration (Section 3.4 for details)

Stage 3: TSGlasso is then used to rank variables that are not relatively correlated or have small AR coefficients by using R. (Section 3.5 for details)

Stage 4: The final model will be obtained and then come up with the best solution for solving the problem of the low birth rate. (Section 3.6 and 4 for details)

3.3. Vector Autoregressive (VAR)

Vector Autoregressive models (VAR) are a powerful tool for analyzing multivariate time series data and it is primarily used as a time series model with multiple variables [13]. In this section, the paper uses high-dimensional VAR. This is because the investigation has more than one variable in the model. In high-dimensional VAR, a sequence of t-statistics is created and the BIC is used to select m of these to eliminate spurious values in the non-zero AR coefficients and reduce them to obtain the final model [13]. The investigation needs to set the initial model first as shown below (linear as p=1):

$$B_t = a_1 + A_{1,1}B_{t-1} + A_{1,2}D_{t-1} + A_{1,3}F_{t-1} + A_{1,4}G_{t-1} + A_{1,5}H_{t-1} + u_{t,1} \quad (1)$$

$$D_t = a_2 + A_{2,1}B_{t-1} + A_{2,2}D_{t-1} + A_{2,3}F_{t-1} + A_{2,4}G_{t-1} + A_{2,5}H_{t-1} + u_{t,1} \quad (2)$$

$$F_t = a_3 + A_{3,1}B_{t-1} + A_{3,2}D_{t-1} + A_{3,3}F_{t-1} + A_{3,4}G_{t-1} + A_{3,5}H_{t-1} + u_{t,3} \quad (3)$$

$$G_t = a_4 + A_{4,1}B_{t-1} + A_{4,2}D_{t-1} + A_{4,3}F_{t-1} + A_{4,4}G_{t-1} + A_{4,5}H_{t-1} + u_{t,4} \quad (4)$$

$$H_t = a_5 + A_{5,1}B_{t-1} + A_{5,2}D_{t-1} + A_{5,3}F_{t-1} + A_{5,4}G_{t-1} + A_{5,5}H_{t-1} + u_{t,5} \quad (5)$$

where B denotes birth rate, D denotes divorce rate, F denotes female employment rate, G denotes gender ratio and H denotes the prices of housing. B_t, D_t, F_t, G_t, H_t which are the related factors in a particular year. $a_n = a_1, a_2, a_3, a_4, \dots, a_p$ which are the intercept coefficients in the functions. $A = A_{1,1}, A_{1,2}, A_{1,3}, A_{1,4}, \dots, A_{n,p}$ which are called AR coefficient matrices, they are the coefficient matrices of variables. A higher AR value indicates a stronger association. $u_{t,n} = u_{t,1}, u_{t,2}, u_{t,3}, u_{t,4}, \dots, u_{t,n}$ are the noise term with the expected mean of zero. Traditional VAR models cannot select the relatively important variables that affect the birth rate, thus the paper use TSGlasso to model different time series based on the VAR model and Bayesian Information Criterion (BIC) to do the model selection.

3.4. Bayesian Information Criterion (BIC)

Bayesian information criterion (BIC) might help the paper to find out the value of p that the investigation should use. p is the number of years that the investigation needs to consider for the variables for several years. For example, when $p = 1$ the model only needs to consider the data corresponding to the relevant factors for the previous year. Model is shown as follows:

$$B_t = a_1 + A_{1,1}B_{t-1} + A_{1,2}G_{t-1} + A_{1,3}E_{t-1} + A_{1,3}D_{t-1} + A_{1,4}H_{t-1} + u_{t,1} \quad (6)$$

However, when $p = 2$ the model needs to consider all the data corresponding to the relevant factors for the previous two years, and the model becomes like what is shown below:

$$B_t = a_1 + A_{1,1}B_{t-1} + A_{1,2}B_{t-2} + A_{1,3}G_{t-1} + A_{1,4}G_{t-2} + A_{1,5}E_{t-1} + A_{1,6}E_{t-2} \quad (7)$$
$$+ A_{1,7}D_{t-1} + A_{1,8}D_{t-2} + A_{1,9}H_{t-1} + A_{1,10}H_{t-2} + u_{t,1}$$

All BIC calculating needs to do is bring the five values of $p = 1, p = 2, p = 3, p = 4$ and $p = 5$ into the following equation for BIC to obtain the minimum BIC value and finally arrive at the ideal p value by using R. The BIC formula is shown here [14]:

$$\mathrm{BIC} = \frac{\mathrm{RSS} + \log(n) d\hat{\sigma}^2}{n} \quad (8)$$

Where d denotes the number of predictors and n denotes the total observations, $\hat{\sigma}^2$ is the estimation of the variance of the error associated with each response measurement in a regression model. RSS denotes the residual sum of squares of the regression model. By following the formula of BIC above, the investigation receives the value of BIC shown below in the table 1:

Table 1. BIC value of the different values of p

p	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
BIC	72.86218	80.89422	72.06764	32.19645	-Inf

From the table of BIC values above, $p = 4$ might be the best choice for the model because it has the smallest BIC value. The smaller the BIC value, the better the model fits the data. The reason for not going for the value of p after $p = 4$ is that p could go on infinitely. Therefore, when $p = 4$, the investigation receives the new model shown below:

$$B_t = a_1 + A_{1,1}B_{t-1} + A_{1,2}B_{t-2} + A_{1,3}B_{t-3} + A_{1,4}B_{t-4} + A_{1,5}G_{t-1} + A_{1,6}G_{t-2} + A_{1,7}G_{t-3} + A_{1,8}G_{t-4} + A_{1,9}F_{t-1} + A_{1,10}F_{t-2} + A_{1,11}F_{t-3} + A_{1,12}F_{t-4} + A_{1,13}D_{t-1} + A_{1,14}D_{t-2} + A_{1,15}D_{t-3} + A_{1,16}D_{t-4} + A_{1,17}H_{t-1} + A_{1,18}H_{t-2} + A_{1,19}H_{t-3} + A_{1,20}H_{t-4} + u_{t,1} \quad (9)$$

3.5. Time Series Graphical Lasso (TSGlasso)

In this section, the paper will first receive a mean squared error (MSE) diagram for the estimation of the model, and then the investigation will use the method for calculating TSGlasso to get the final AR coefficients of the model. The advantage of using TSGlasso is that scholars can find out the most important factor that influences the birth rate as TSGlasso will help the model to shrink the

insignificance AR coefficient into zero, which might then help the model come up with some of the solutions for low birth rate problem.

3.5.1 The calculation of MSE

If the model assumes that y is the value of B_t assumes x are $B_{t-1}, B_{t-2}, B_{t-3}, B_{t-4}, G_{t-1}, G_{t-2}, G_{t-3}, G_{t-4}, F_{t-1}, F_{t-2}, F_{t-3}, F_{t-4}, D_{t-1}, D_{t-2}, D_{t-3}, D_{t-4}, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-4}$. Then, the investigation gets the MSE figure as shown in figure 6:

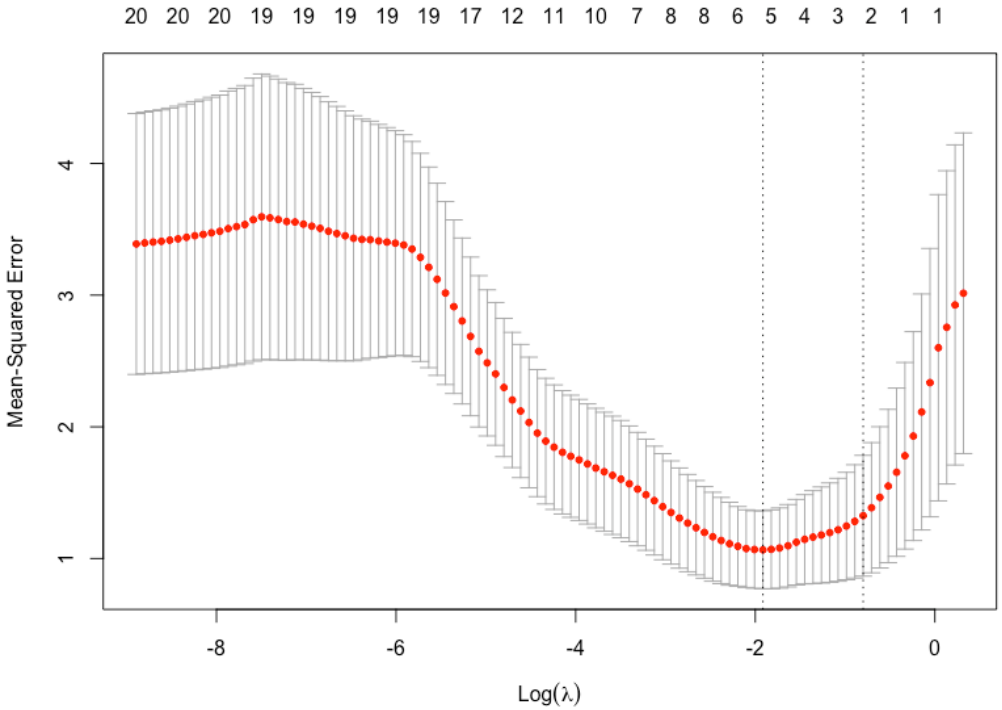

Fig. 6 The MSE value of the final model.

3.5.2 Methods of TSGlasso

Lasso means the least absolute shrinkage and selection operator. The Time Series Graphical lasso mainly helps the investigation to find factors that are weakly related to the birth rate. Even when the AR factor is not equal to zero, the investigation can still use TSGlasso to shrink the factor to zero. The general method for calculating TSGlasso is shown below [15]:

$$\text{RSS} = \Sigma(y_i - \hat{y}_i)^2 \tag{10}$$

$$\text{RSS} + \lambda\Sigma|\beta_j| \tag{11}$$

The equation above indicates the shrinkage penalty method. In the lasso method. Where RSS means the squared residuals, Σ means the sum of the total number of $(y_i - \hat{y}_i)^2$, y_i denotes the actual response value for i^{th} observation and $\hat{y}_i$ denotes the predicted response value based on the multiple linear regression model. The investigation has a λ for producing the lowest possible mean squared error (MSE) as shown in figure 6.

Setting the alpha equal to 1 and lambda equal to the best lambda, the calculation then receives all the data of the AR coefficients of the final model as shown in the tables below: (where "." means the AR coefficient is small enough and shrink to zero and data are recorded for 5 significance level)

Table 2. AR coefficient calculated by TSGlasso

	(intercept)	B_{t-1}	B_{t-2}	B_{t-3}	B_{t-4}
AR	27.401	0.56223	0.36012	.	.
		D_{t-1}	D_{t-2}	D_{t-3}	D_{t-4}
AR		.	.	.	-0.33320
		G_{t-1}	G_{t-2}	G_{t-3}	G_{t-4}
AR		0.0032359	.	.	.
		F_{t-1}	F_{t-2}	F_{t-3}	F_{t-4}
AR		-0.23580	-0.066858	.	.
		H_{t-1}	H_{t-2}	H_{t-3}	H_{t-4}
AR	.	.	.		-0.00010329

3.6. Final model

$$B_t = 27.401 + 0.56223B_{t-1} + 0.36012B_{t-2} + 0.0032359\,G_{t-1} - 0.23580F_{t-1} - 0.066858F_{t-2} - 0.33320D_{t-4} - 0.00010329H_{t-4} \quad (12)$$

where B_t denotes the birth rate in the current year, B_{t-1} denotes the last year birth rate and B_{t-2} denotes the birth rate of the year before last year. G_{t-1} denotes the gender ratio in the last year, F_{t-1} denotes the female employment rate of last year and F_{t-2} denotes the female employment rate of the year before last year. D_{t-4} is the divorce rate four years ago and H_{t-4} is the housing price four years ago. The best lambda for the final model is 0.1223379.

4. Conclusion

As can be seen from the final model, the calculation has used TSGlasso to shrink many of the AR coefficients into zero. As a result, there are fewer variables for the final model. By looking at the AR coefficients in the model, the variables that are more related to the birth rate can be found. The main aim is to look at the largest AR coefficient in the model other than the two variables B_{t-1} and B_{t-2}. Thus, the third largest absolute value of the AR coefficient of the model is -0.23580 for F_{t-1}. It illustrates a negative value that indicates a negative relationship between F_{t-1} and B_t. This might show that when the female employment rate rises in the current year, the birth rate is likely to fall in the next year. ExcludingB_{t-1}, B_{t-2} and F_{t-1}, the fourth largest absolute value of AR coefficient is -0.33320 of D_{t-4}, which means that D_{t-4} and the current year's birth rate is in a negative relationship, so that when the divorce rate in previous year increases, the birth rate of current year decrease.

The government can help improve the birth rate by protecting pregnant women and women with children and reducing discrimination against them by companies. The Government could also organize classes for young teenagers to develop proper concepts of marriage to help reduce the divorce rate and thus improve the birth rate.

References

[1] Khan, H. Top 10 Lowest Birth Rate Countries in the World. World Blaze.2018.

[2] Admin. The second child is ineffective & another major adjustment to fertility policy & 700 million people may benefit. Aiziw. 2022.

[3] Garza, A. The Economic Impact of Low Birth Rates. Human Defense Initiative. 2019.

[4] Wang, P. P. The total population has been growing and the level of urbanization has been steadily increasing. National Bureau of Statistics.2022.

[5] Bailey, L., Bliese, C., Davis, J., Grams, K., Huth, B., Mcdonald, K., … & Short, L. The Factors Affect the Total Fertility Rate & TFR. Population Education.2019.

[6] Tian, Z. R., Yu, J., Zhang, S. M. Application of mathematical methods to population growth models in the new context. Open access. 2020.

[7] Boque, J. A Warning About Declining Birth Rates. Human life.2021.

[8] Birthrates are declining globally & here is why it matters. Birthrates are declining globally & here is why it matters. World economic forum. 2021.

[9] Marriage and divorce rates. Marriage and divorce rates. National Bureau of Statistics.2021.

[10] China Women's Development Programme (2011-2020) Statistical Monitoring Report.. China Women's Development Programme (2011-2020) Statistical Monitoring Report. National Bureau of Statistics.2020.

[11] Wang, C. Y. & Lee, S. J. Regional Population Forecast and Analysis Based on Machine Learning Strategy. Entropy. 2021.

[12] China Women's Development Programme 2011-2020 Final Statistical Monitoring Report. China Women's Development Programme 2011-2020 Final Statistical Monitoring Report. Central People's Government of the People's Republic of China. 2021.

[13] Vector Autoregressive models VAR(p) models. (n.b.). Vector Autoregressive models VAR(p) models. PennState. Section 11.2.

[14] Zach. How to Calculate BIC in R. Statology. 2021.

[15] Zach. Lasso Regression in R (Step-by-Step). Statology.2020.