Influencing Factors Analysis of Birth Rate Spatial Distribution in China's Provinces

Liyun Su[1(✉)], Haiqin Yu[1], Fenglan Li[2], and Yun Jiang[2]

[1] School of Science, Chongqing University of Technology, Chongqing, China
suliyun@cqut.edu.cn

[2] Chongqing University of Technology, Chongqing, China
{lifenglan,jiangyun}@cqut.edu.cn

Abstract. The long-term development of China's economy is inseparable from the labor force, and the birth rate is an important factor to promote China's economic development. This paper uses spatial statistics to analyze the potential social causes of influencing the spatial distribution of birth rates in Chinese provinces. The influence factors of economic development, social security and cultural life constructed by factor analysis and the spatial interaction effect of birth rate in provinces were incorporated into the spatial statistical model, and the factor space regression model was established and compared. The results show that: the spatial distribution of birth rate in China's provinces is mainly affected by economic development, social security and cultural life; in terms of economic development, especially the per capita disposable income of residents, has a significant role in hindering the rise of birth rate; in terms of social security, it promotes the rise of birth rate; in terms of cultural life, especially the aging rate, has a significant impact on birth rate, the higher the aging rate is, the lower the birth rate is; for the spatial analysis of birth rate in Chinese provinces, the factor spatial lag model is more suitable.

Keywords: birth rate · spatial regression · factor analysis · factor spatial lag model

1 Introduction

In recent years, although China's total population is growing slowly, the birth rate is also changing. Since the early 1990s, it has dropped below the replacement level and has continued to decline in recent years, which is significantly related to China's unique family planning policy and other factors. In order to optimize the population structure, the state has adopted a certain fertility policy to control the population stability. Although it has played a certain role, it has not changed the current situation of China's low fertility rate, and the population problem has become diversified and complicated. According to the summary of research literature, China's birth rate is mainly affected by economic development [1], living standards [2], culture education [3] and social security [4]. We can use these four aspects to study the impact of social comprehensive indicators on the birth rate.

Z. Zhan et al. (Eds.): ICBDIE 2022, AHCS 5, pp. 1184–1195, 2023.
https://doi.org/10.2991/978-94-6463-034-3_122

This paper uses factor analysis to explore influence birth rate of the correlation between each variable, build the impact factor of birth rate, and using spatial weight matrix said provinces adjacent relations, realize the provinces spatial interaction effect of birth rate, Based on the traditional regression model, the spatial weight matrix of each province was incorporated into the model, and the spatial lag model, spatial error model and spatial Durbin model were introduced to analyze our country provinces the influence factors of the birth rate, and comparing with the ordinary least squares regression model analysis.

2 Method and Data Description

TO solve the spatial dependence in linear regression analysis, the spatial effect is integrated into the standard linear regression model. Anselin [5] proposed different spatial regression models according to different spatial interaction effects. If the interaction effects of endogenous, exogenous and error terms are considered at the same time, the expression is obtained:

$$\mathrm{Y} = \rho WY + \alpha 1_N + X\beta + WX\theta + \mu, \mu = \lambda W\mu + \varepsilon, \varepsilon \sim N\left(0, \sigma^2 I_N\right) \tag{1}$$

The Humanities and Social Sciences Research Program Foundation of the Ministry of Education in China(19XJA840002) and Humanities and Social Sciences Research Project of Chongqing Municipal Commission of Education of China in 2020 (20SKGH174).

In Eq. (1), y is the vector of order N × 1, which is composed of the observation value y corresponding to the explained variable of each spatial observation unit y_i in the sample; X is the N × K order exogenous explanatory variable matrix, where x_{ij} (i = 1, 2,…, N; j = 1, 2,…K) is the (i, j) element of the matrix X; W is an N × N order spatial weight matrix; 1_N is a column vector with N × 1 order element 1, which corresponds to the parameter α of constant term to be estimated; Spatial regression coefficients of ρ and λ; Error term in μ space regression; The parameter β mainly reflects the influence of independent variable X on dependent variable Y. θ is the spatial autocorrelation coefficient of exogenous variables of order K × 1; ε is a random perturbation term.

2.1 Spatial Durbin Model (SDM)

$$Y = \rho WY + a1_N + X\beta + WX\theta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_N) \tag{2}$$

In Eq. (1), let λ = 0 to obtain the spatial Durbin model, which takes into account the spatial correlation between the explained variables and the explanatory variables.

2.2 Spatial Autocorrelation Model (SAC)

$$\mathrm{Y} = \rho WY + \alpha 1_N + X\beta + \mu, \mu = \lambda W\mu + \varepsilon, \varepsilon \sim N\left(0, \sigma^2 I_N\right) \tag{3}$$

In Eq. (1), let θ = 0 to obtain the spatial autocorrelation model, which includes the endogenous interaction effect and error term effect, that is, the model includes both spatial lag and spatial error correlation structure.

2.3 Spatial Lag Model (SLM)

$$\mathrm{Y} = \rho WY + \alpha 1_N + X\beta + \varepsilon, \varepsilon \sim N\left(0, \sigma^2 I_N\right) \tag{4}$$

In Eq. (2), let θ = 0 to get the spatial lag model. This model only considers the spatial correlation between the explained variables, and mainly studies whether the variables have spillover effect in a certain area. The ρ is a spatial regression coefficient, which reflects the relationship between the spatial units. WY is the spatial lag dependent variable of the spatial weight matrix W, and it is an endogenous variable.

2.4 Spatial Error Model (SEM)

$$Y = a1_N + X\beta + \mu, \mu = \lambda W\mu + \varepsilon, \varepsilon \sim N\left(0, \sigma^2 I_N\right) \tag{5}$$

In Eq. (2), let $\theta = —\rho\beta$ to obtain the space error model, taking into account the correlation between error terms, mainly studying the situation where the interaction between regions is different due to their relative positions. Parameter λ is the correlation coefficient of spatial error, which measures the impact of the error impact of dependent variable in the adjacent area on the observed value of the region.

2.5 Ordinary Least Square Model (OLS)

$$Y = a1_N + X\beta + \varepsilon, \varepsilon \sim N\left(0, \sigma^2 I_N\right) \tag{6}$$

In Eq. (4), let $\lambda = 0$ to get the ordinary least square regression.

The research objects of this paper are 31 provinces of the People's Republic of China except Hong Kong, Macao and Taiwan. Use sectional data for spatial statistical analysis, the adopted birth rate is the ratio of the number of births in a certain area in a certain period (usually a year) to the average number of people (or the number of people in the Interim). Considering the authenticity and validity of the relevant data, all the variable data involved in this paper are from the statistical data of various provinces in 2017 published on the official website of the National Bureau of Statistics [6].

Based on the comprehensive literature [7, 8] and considering the extensiveness and rationality of variables, paper considers the influencing factors of the spatial distribution of birth rate from three aspects of economic development, social security and cultural life, and selects 11 indicators to illustrate the influence on the spatial distribution of birth rate (Shown in Table 1).

Table 1. Variable Interpretation

Variables	Unit
Birth rate(y)	‰
Per capita disposable income(x_1)	yuan
Average selling price of housing(x_2)	yuan/square meter
GDP per capita (x_3)	yuan/person
Urbanization rate(x_4)[a]	%
GDP of tertiary industry(x_5)	%
Resident population(x_6)	ten thousand person
Number of students in ordinary primary schools(x_7)	ten thousand person
financial education expenditure(x_8)	hundred million
Residential sales area(x_9)	hm^2
Aging rate(x_{10})[b]	%
Divorce rate(x_{11})[c]	‰

a. The proportion of the urban population (the total permanent population living within the urban area) in the permanent population at the end of the year, usually expressed in %.
b. The ratio of the population aged 65 and above to the total population, usually expressed in %.
c. The proportion of the number of divorces in the annual average population of a region in that year, which is expressed by %.

3 Construct Spatial Statistical Model

Firstly, the correlation between 11 explanatory variables x_j (j = 1,2, …,11) and the correlation between birth rate y and explanatory variables x_j (j = 1,2, …,11) are analyzed. Because of the large number of explanatory variables and the strong correlation between explanatory variables, it is reasonable to establish a linear regression model between the birth rate and multiple explanatory variables.

3.1 Factor Analysis

KMO test and Bartlett test were used to judge the applicability of factor analysis (Shown in Table 2).

It can be seen from Table 2 that the p value of Bartlett's spherical test is approximately equal to 0, obviously rejecting the original hypothesis that the variables are independent of each other, and it can be seen that there is an obvious correlation between the variables. Moreover, the statistic of KMO test is 0.713, which indicates that the correlation between variables is strong, and the original variables are suitable for factor analysis. The results of factor analysis of variables are shown in Table 3.

The cumulative explanation rate is 86.8%, indicating that there is less information missing. The public factors proposed are also relatively strong in explanatory power for each variable, which illustrates the applicability of factor analysis to a certain extent.

Table 2. Variable Interpretation

KMO Statistic		0.713
Bartlett Sphericity Test	The approximate chi-square	453.763
	Df /P value	55 /4.2e-64

Table 3. Factor Analysis Results of Variables

Variables	**Factor loading**		
	Factor 1	***Factor 2***	***Factor 3***
Per capita disposable income(x_1)	0.992	0.065	0.096
Average selling price of housing(x_2)	0.927	−0.072	−0.019
GDP per capita (x_3)	0.927	0.082	0.086
Urbanization rate(x_4)	0.878	0.042	0.227
GDP of tertiary industry(x_5)	0.812	−0.217	0.007
Resident population(x_6)	−0.070	0.987	0.118
Number of students in ordinary Primary schools(x_7)	−0.169	0.957	−0.071
Financial education expenditure (x_8)	0.198	0.952	−0.027
Residential sales area(x_9)	−0.041	0.940	0.115
Aging rate(x_{10})	0.334	0.260	0.904
Divorce rate(x_{11})	−0.020	−0.072	0.699
New Named Factor	**Economic development (F_1)**	**Social security(F_2)**	**Cultural life (F_3)**
Eigenvalues	4.325	3.818	1.406
Factor explanation ratio (%)	39.3	34.7	12.8
Cumulative interpretation ratio(%)	39.3	74.0	86.8
Per capita disposable income(x_1)	0.992	0.065	0.096

According to Table 3, we can see the correlation between factor F(F_1,F_2,F_3) and each explanatory variable x_j (j = 1,2, …,11), so three factor score functions are:

$$\begin{aligned} F_1 &= 0.988x_1 + 0.017x_2 + 0.016x_3 + 0.010x_4 + 0.007x_5 - 0.040x_6 \\ &\quad - 0.006x_7 + 0.010x_8 - 0.002x_9 - 0.100x_{10} - 0.020x_{11} \\ F_2 &= 0.090x_1 - 0.003x_2 + 0.003x_3 - 0.001x_4 - 0.005x_5 + 0.748x_6 \\ &\quad + 0.118x_7 + 0.115x_8 + 0.054x_9 - 0.096x_{10} - 0.004x_{11} \\ F_3 &= -0.382x_1 - 0.009x_2 - 0.007x_3 + 0.001x_4 - 0.002x_5 - 0.171x_6 \\ &\quad - 0.046x_7 - 0.049x_8 - 0.013x_9 + 1.152x_{10} + 0.009x_{11} \end{aligned} \tag{7}$$

3.2 Establishment of Spatial Statistical Model

The variance inflation factor (VIF) is used to study the multicollinearity among the three factors of economic development (F_1), social security (F_2) and life culture (F_3). The variance inflation factor of the three variables is very low, and the mean value is close to 1. It is much less than 10, indicating that the multicollinearity between variables is very small and can be used to construct a regression model.

The spatial interaction between provinces and cities is realized through the spatial weight matrix, which is incorporated into the regression model, and the factor analysis is combined with the spatial regression model to construct the factor spatial regression model.

3.2.1 Factor Space Durbin Model (FSDM)

$$\begin{gathered} Y = \rho WY + \beta_0 1_N + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \theta_1 WF_1 + \\ \theta_2 WF_2 + \theta_3 WF_3 + \varepsilon, \qquad \varepsilon \sim \mathrm{N}\left(0, \sigma^2 \mathrm{I_N}\right) \end{gathered} \tag{8}$$

3.2.2 Factor Space Autocorrelation Model (FSAC)

$$\begin{gathered} Y = \rho WY + \beta_0 1_N + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \mu \\ \mu = \lambda W\mu + \varepsilon, \varepsilon \sim \mathrm{N}(0, \sigma^2 \mathrm{I_N}) \end{gathered} \tag{9}$$

3.2.3 Factor Space Lag Model (FSLM)

$$Y = \rho WY + \beta_0 1_N + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \varepsilon, \varepsilon \sim \mathrm{N}\left(0, \sigma^2 \mathrm{I_N}\right) \tag{10}$$

3.2.4 Factor Space Error Model (FSEM)

$$\begin{gathered} Y = \beta_0 1_N + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \mu \\ \mu = \lambda W\mu + \varepsilon, \varepsilon \sim \mathrm{N}(0, \sigma^2 \mathrm{I_N}) \end{gathered} \tag{11}$$

3.2.5 Ordinary Least Square Method (OLS)

$$Y = \beta_0 1_N + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \varepsilon, \varepsilon \sim \mathrm{N}\left(0, \sigma^2 \mathrm{I_N}\right) \quad (12)$$

where F_1, F_2 and F_3 are respectively

$$\begin{cases} F_1 = (F_{1\mathrm{i}}) = (\alpha_1 x_{\mathrm{i}1} + \alpha_2 x_{\mathrm{i}2} + \cdots + \alpha_{11} x_{\mathrm{i}11}) \\ F_2 = (F_{2\mathrm{i}}) = (\alpha_1 x_{\mathrm{i}1} + \alpha_2 x_{\mathrm{i}2} + \cdots + \alpha_{11} x_{\mathrm{i}11}) \\ F_3 = (F_{3\mathrm{i}}) = (\varphi_1 x_{\mathrm{i}1} + \varphi_2 x_{\mathrm{i}2} + \cdots + \varphi_{11} x_{\mathrm{i}11}) \\ \mathrm{i} = 1, 2, \cdots \mathrm{N} \end{cases} \quad (13)$$

In these models, Y represents the birth rate, which is an N × 1 order vector, composed of y_i (i = 1, 2, …, N); F_1 represents the factor score function in economic development, which is an N × 1 order vector In the same way, F_2 represents the factor scoring function of social security, F_3 represents the factor scoring function of cultural life, W is the spatial weight matrix constructed according to the Queen proximity criterion, ρ、θ and λ are the spatial regression coefficients, x_1, x_2, …,x_{11} are selected explanatory variables.

4 Empirical Analysis

4.1 Ordinary Least Squares Regression

The birth rate and factors were analyzed by OLS. The estimated results are shown in Table 4.

It can be seen from Table 4 that the adjusted R^2 after OLS regression is 0.4425, which is not a good fitting result, but the real factors affecting the birth rate in real life can't be simply inferred from the R^2. The constant term, economic development (F_1) and cultural life (F_3) all passed the significance test with a significance level of 0.05, while the social security (F_2) failed the significance test. The p value of J-B test of OLS model is 0.8101, which doesn't reject the original hypothesis that the error term is normal; the p value of B-P test is 0.4058, which does not reject the original hypothesis of homogeneity of variance. At the same time, the p value of Moran's I statistic is 0.04158, which indicates that the error of OLS model has obvious spatial correlation, which further proves that the birth rate in China has obvious spatial influence on geographical location. At the same time, it indicates the independence of each element in the cross-section data, and the results will be biased if OLS model is used directly.

For LM Test, it can be found that the LMlag of spatial lag model is more significant than that the LMerr of spatial error model, and RLMlag is more significant than RLMerr, which indicates that the birth rate of China's provinces is directly affected by neighboring provinces, rather than from some unrecognized spatial aggregation features. Therefore, it can be inferred that spatial lag model is more suitable for explaining the spatial dependence of birth rate than spatial error model.

4.2 Factor Space Regression Analysis

In order to further compare and verify that the factor spatial regression model is more suitable for explaining the spatial distribution of birth rates, the factor spatial regression model is used to analyze the birth rate. The estimated results are shown in Table 5.

Table 4. Ols Estimation Results.

	Regression coefficient	Standard deviation	T statistics	P values
Constant	12.1210	0.3988	30.396	2e-16
F_1	−1.1338	0.4064	−2.790	0.00954
F_2	0.6668	0.4066	1.640	0.11262
F_3	−1.6463	0.4076	−4.039	0.00040
R^2	0.4983	*RMSE*	11.536	
Adjusted R^2	0.4425		AIC	143.143
F value	8.938		BIC	150.313
J-B test	3		0.4212	0.8101
B-P test	3		2.9091	0.4058
Moran's I	0.1697		1.7326	0.04158
LMerr	1		1.7830	0.1818
LMlag	1		7.3544	0.0067
RLMerr	1		2.3661	0.1240
RLMlag	1		7.9375	0.0048

According to Anselin(1998) proposed classical LM test and Anselin and Smirnov(1996) proposed robust LM test, it is considered that the spatial lag model is more suitable than the spatial error model. In order to choose the best model to fit the data, a spatial Durbin model is established and Wald test and LR test are used to verify whether the spatial Durbin model can be reduced to a spatial lag model. Wald test (6.808, $p = 0.0783$) and LR test (7.0296, $p = 0.0709$) both show that the null hypothesis is not rejected. Similarly, the original hypothesis of spatial Durbin model simplified to spatial error model is H_0: $\theta + \rho\beta = 0$. At this time, the original hypothesis is rejected (Wald test (34.667, $P = 1.432e\text{-}07$) and LR test (12.197, $P = 0.0067$)), indicating acceptance of spatial lag model, rejection of spatial error model and spatial Durbin model. Considering SAC model, the statistical value of LR test for SAC model and SLM model is 2.188, and its degree of freedom is 1, and the statistical value of LR test for sac model and SEM model is 7.356, and its degree of freedom is 1. However, in both cases, the critical value is 3.84 at the 5% level, which further indicates that SLM model is better than sac model.

Combined with other common test criteria, SDM and SAC models are obviously more advantageous than SLM and SEM, because both models are nested SLM and SEM models. However, the coefficient of variable WY in SAC model is significant, but the coefficient of variable Wu is not significant, and the coefficient of WX in SDM model is not significant, so SLM model is a better choice. The results show that the spatial lag model considering the spatial lag term can better explain the spatial interaction effect of the population birth rate in China's provinces. The expression of FSLM is obtained by.

Table 5. Comparison Between Factor Space Regression Model

	Factor Spatial Lag Model (FSLM)			**Factor Spatial Error Model (FSEM)**			**Factor Space Durbin Model (FSDM)**			**Factor Space Autocorrelation Model(FSAC)**		
	Regression coefficient	***Z statistics***	***P values***	***Regression coefficient***	***Z statistics***	***P values***	***Regression coefficient***	***Z statistics***	***P values***	***Regression coefficient***	***Z statistics***	***P values***
Constant	4.007	2.479	0.013	12.066	11.860	0.000	5.057	2.629	0.008	2.633	1.636	0.102
F_1	-0.875	−2.774	0.006	−1.176	−2.853	0.004	−0.651	−1.626	0.104	−0.798	−3.17	0.002
F_2	0.320	1.070	0.285	−0.135	−0.424	0.671	0.243	0.790	0.429	0.567	2.310	0.002
F_3	-1.048	−3.297	0.001	−1.099	−2.742	0.006	−0.963	−2.609	0.009	−0.955	−3.514	0.021
WF_1	4.007	2.479	0.013	12.066	11.860	0.000	−1.125	−1.379	0.167			
WF_2	-0.875	−2.774	0.006	−1.176	−2.853	0.004	1.265	2.276	0.023			
WF_3	0.320	1.070	0.285	−0.135	−0.424	0.671	-0.198	−0.281	0.779			
ρ	0.665	5.116	0.000				0.548	3.396	0.001	0.773	5.861	0.000
λ				0.684	4.964	0.000				−0.500	−1.378	0.168
Test	Degree of freedom	Statistics	P values	Degree of freedom	Statistics	P values	Degree of freedom	Statistics	P values	Degree of freedom	Statistics	P values
R^2		0.689			0.636			0.740			0.741	
LogL		−61.143			−63.727			-57.629			−60.048	
LR	1	10.856	0.001	1	5.689	0.017	1	6.786	0.009	1	13.045	0.001
AIC		134.29			139.45			133.26			134.1	
RMSE		9.086			9.829			8.307			8.284	

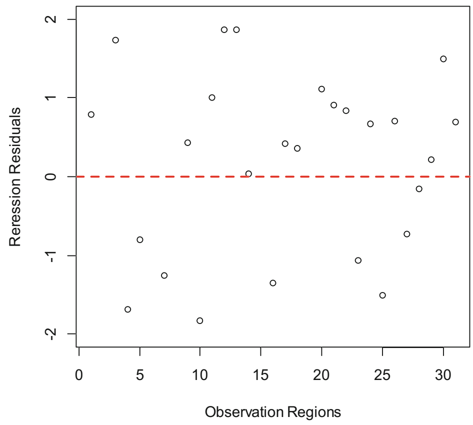

Fig. 1. Residual plot.

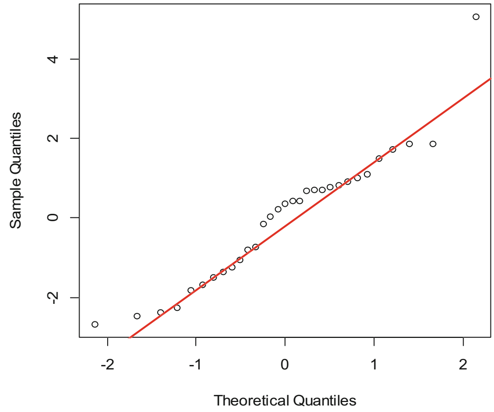

Fig. 2. Residual plot.

Equation (7), Eq. (10) and Eq. (13):

$$
\begin{aligned}
Y = 0.665WY + 4.007 \times 1_N - 0.4353x_1 - 0.0150x + 0.0029x_3 \\
-0.0129x_4 - 0.0200x_5 + 0.4535x_6 + 0.0912x_7 \\
+ 0.0794x_8 + 0.0326x_9 - 1.4269x_{10} - 0.0204x_{11}
\end{aligned}
\tag{14}
$$

4.3 Normality Test and Randomness Test of Residuals

In order to verify the rationality of Eq. (14) of FSLM, the residual test was carried out. Through the residual graph (shown in Fig. 1), it is found that the residuals fluctuate around 0 without obvious trend, and the residuals have randomness; The points of QQ chart (shown in Fig. 2) surround the straight line and are concentrated at (0, 0). According to the results of Shapiro-Wilk normal test, P value is 0.06805. At the significance level of

0.05, we do not refuse the original hypothesis that the residuals are normal distribution, so the residuals obey normal distribution. Therefore, it is reasonable to select FSLM model.

5 Conclusions

Through the empirical results of the least square regression model and the spatial regression model, we can get that the birth rate has significant correlation with economic development, social security and cultural life. Economic development has a significant hindering effect on the rise of birth rate, which is in line with China's current national conditions. At the same time, the social security aspect promotes the birth rate to rise. However, cultural life has a strong negative effect on the birth rate. Among them, aging rate has a significant hindering effect on the growth of birth rate. The higher the aging rate is, the lower the birth rate is. In general, the spatial regression model is a good and innovative application to the birth rate data with spatial correlation.

Acknowledgments. The Humanities and Social Sciences Research Program Foundation of the Ministry of Education in China (19XJA840002) and Humanities and Social Sciences Research Project of Chongqing Municipal Commission of Education of China in 2020 (20SKGH174).

References

1. L. Li, "Research on the Impact of High Birth Rate and Economic Growth," Modern Business Trade Industry, vol. 23, pp. 19–20, September 2011.
2. Z. Li, "Family Upbringing Costs and the Adjustment Mode of the Macro Employment Market on the Birth Rate," East China Economic Management , vol. 26, pp. 110-147, October 2012.
3. X.T. Liu, J.S. Zhang, Y. Hu, "The Impact of Rising Housing Prices on the Birth Rate—— an Empirical Study Based on China's 1999–2013 Data," Journal of Chongqing University of Technology(Social Science), vol.20, pp. 53–61, April 2016.
4. Y. Kang, G. J. Wang, X.J. Zhao, "An Empirical Study on the Feasibility of Adjusting the Birth Rate by the Social Security Level," Research on Financial and Economic Issues, vol. 10, pp. 111–115, October 2013.
5. L. Anselin, Spatial Econometrics: Methods and Models, Boston: Kluwer Academic Publishers, 1988, pp.32–40.
6. National Data of the National Bureau of Statistics of the People's Republic of China. Available online: http://www.stats.gov.cn/
7. L. Y. Su, X. H. Yang, G. Y. Bai, F. L. Li, "Analysis of spatial disequilibrium and regional difference of birth rate in China," Journal of Chongqing University of Technology (Natural Science), vol. 32, pp. 249–258, March 2018.
8. L. Y. Su, C.H. Liang, X.H. Yang, Y. Liu, ".Influence Factors Analysis of Provincial Divorce Rate Spatial Distribution in China," Discrete Dynamics in Nature and Society, vol. 2018, pp. 1–12, July 2018.