Analysis of the Factors Affecting the Difference of Divorce Rate in Different Regions of China

-- Empirical Research Based on Factor Analysis

Yunqing Xiong

Department of Economics, Jinan University, Guangzhou, China

18702504208@163.com

Abstract

Freedom of marriage is the right that every law gives us to every citizen. This freedom contains two meanings: one is freedom of marriage; the other is freedom of divorce. With the development of social and economic and the change of people's family values, the divorce rate in China has also been rising, and the divorce rate in various regions of China has also shown great differences. In order to study and analyze the influencing factors of the divorce rate differences in different regions, this paper selects 13 original statistical indicators, and builds a factor analysis model to generate 4 statistical indicators, and conducts a multiple linear regression analysis with the divorce rate as the explained variable variable by using the generated statistical indicators as explanatory variables to 31 provinces in China. Furthermore, the entropy weight TOPSIS method and other relevant data are used for further statistical tests and analysis of the results.

Keywords

Divorce rate; Regional differences; Factor analysis; Entropy weight TOPSIS.

1. INTRODUCTION

At present, China is in the transition period from traditional society to modern industrialized society, and the probability of divorce rate tends to rise. Since the Second World War, the divorce rate in many developed countries has shown an increasing trend. Since the reform and opening up, and with the continuous advancement of China's industrialization and urbanization, China's socialist market economic system has gradually been established and improved. Under the double impact of modernization and globalization, Chinese society is gradually getting rid of the shackles of feudal traditional thinking. Our life philosophy and attitude towards life are constantly changing [9]. With the continuous adjustment of social order and population order, the divorce incident has shown a growing trend. The rising divorce rate is not conducive to the stability of family life, the establishment of a harmonious society and the steady growth of the economy. However, the increasing divorce rate indicates that the concept of freedom and gender equality is deeply rooted in the hearts of the people, and the society’s tolerance for divorce is increasing.

At present, sociologists and economists pay special attention to the reasons for the increase in divorce rate. sociologists recognize that the dissolution of marriage is closely related to the social costs and pressures of social integration, social exclusion, religious expulsion and job discrimination [5]. Economists interpret divorce based on economic theory and economic factors. They think that marriage is a function of welfare utility maximization, and there is potentially a competitive marriage market. The choice of the best partner is influenced by

World Scientific Research Journal Volume 5 Issue 9, 2019
ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

market conditions. Because of the gradual decline in marital happiness and the imbalance in the regulation of the marriage market, they finally choose to divorce. Judging from the current research results, although the research results of divorce in China are relatively rich, there are few studies on the factors affecting the regional differences in divorce rate [1]. Therefore, this paper starts from this perspective and analyzes the influencing factors of regional differences in divorce rate.

The main research contents of this paper include the theoretical basis of factor analysis and entropy weight TOPSIS method. It briefly explains the current situation of divorce rate in China and the divorce rate in different regions, as well as the factors affecting the regional divorce rate. Using factor analysis to solve the problem of high correlation between social cohesion and family structure influencing factors. Factor analysis is used to obtain the scores of new social aggregate factors and new family structure factors, and the new scores of these two indicators are included in the multiple regression analysis of divorce rate, and finally conclusions are made and relevant recommendations are made.

2. BACKGROUND AND RESEARCH OBJECTS

2.1. Background

The divorce rate in various regions of China basically shows a growing trend. As shown in Figure 1, from 2010 to 2016, the overall gross divorce rate in the country remains above 2‰, and the divorce rate is high. At the same time, the national and regional divorce rate is increasing year by year. The provinces with higher divorce rate are mainly Xinjiang, Heilongjiang, Jilin, Liaoning, Sichuan, Chongqing, Beijing and Tianjin [6]. From the geographical distribution, we know that most of these provinces and cities are located in the northwest, northeast, north China and Chuanxiong area. The provinces with low divorce rates are mainly in Tibet, Hainan, Guangdong, Gansu, Shanxi, Qinghai and Jiangxi.

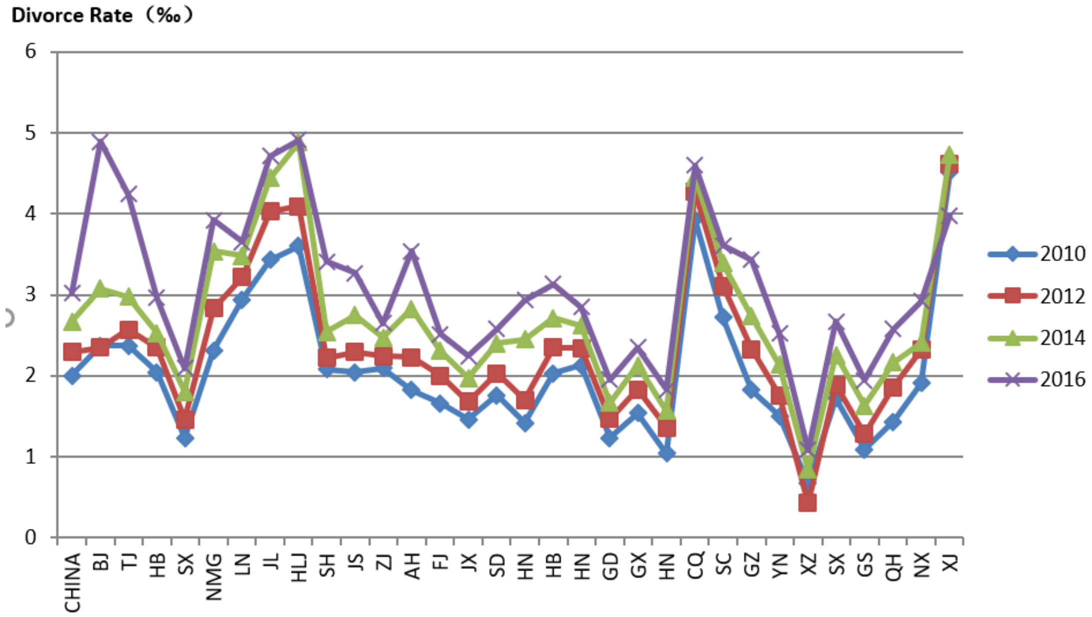

Figure 1. Divorce rate change figures in various regions of China

2.2. Indicator Selection

By consulting relevant literature, this paper finds that the influencing factors affecting the divorce rate difference can be roughly divided into four aspects: social cohesion, family structure, ethnic customs and population structure. This paper refers to the relevant literature,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

and selects a total of 13 original statistical indicators in four aspects [7]: social cohesion, family structure, ethnic customs and population structure, as shown in Table 1.

Table 1. Main statistical indicators

Target layer	Indicator layer
Social cohesion	The proportion of urban population to the total population
	The proportion of employees in the second and third industries to the total number of workers
	Per capita GDP
	Per capita use of foreign capital
	The proportion of higher education or above in the population over 6 years old
	Household consumption level
Family structure	Household size per household
	Birth rate
	Number of surviving children per woman aged 15-64
	Total population burden factor
National customs	The proportion of ethnic minorities in the total population of the region
Population structure	Sex ratio of population over 15 years old in the region

3. FACTOR ANALYSIS

Since the metrics of social cohesion and family structure are highly correlated, there are common factors. In order to exclude the influence of multicollinearity of these two variables on the statistical analysis of regression models, This paper use a factor analysis of the 10 indicators that represent two variables. In general, the impact of social cohesion, family structure, ethnic customs and population structure on the divorce rate in the region is time-delayed. Therefore, this paper uses the relevant data of 31 provinces in 2011 as an independent variable to conduct statistical analysis.

3.1. Theoretical Basis of Factor Analysis

The goal of Factor Analysis is to interpret the correlation of a set of observable variables by exploring a set of fewer, more fundamental, unobservable variables hidden under the data. These virtual, unobservable variables are called factors. (Each factor is thought to explain the variance common to multiple observed variables, so to be precise, they should be called common factors.)

The form of the model is:

$$\begin{cases} x_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1n}F_n + \varepsilon_1 \\ x_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2n}F_n + \varepsilon_2 \\ \quad\quad\quad\quad\quad\quad . \\ \quad\quad\quad\quad\quad\quad . \\ \quad\quad\quad\quad\quad\quad . \\ x_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pn}F_n + \varepsilon_p \end{cases}$$

This model is also known as the R-type orthogonal factor model, and its matrix form is:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

$$X = AF + \varepsilon$$

Where x_p is the pth observable variable, F_n is the common factor, and n<p. ε_p is a unique part of the x_p variable (cannot be interpreted by a common factor). a_{pn} can be thought of as the contribution of each factor to the composite observable variable.

3.2. Modeling Steps for Factor Analysis

3.2.1 Verify that the variables to be analyzed are relevant

To check whether there is correlation between each observed variable, calculate the correlation matrix of the data to be analyzed as follows:

Table 2. Correlation matrix of observed data

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.00	0.88	0.94	0.60	0.90	0.93	-0.76	-0.69	-0.79	-0.79
X2	0.88	1.00	0.89	0.65	0.76	0.88	-0.60	-0.57	-0.72	-0.80
X3	0.94	0.89	1.00	0.63	0.86	0.91	-0.70	-0.63	-0.79	-0.82
X4	0.60	0.65	0.63	1.00	0.41	0.68	-0.36	-0.39	-0.41	-0.48
X5	0.90	0.76	0.86	0.41	1.00	0.82	-0.67	-0.57	-0.72	-0.75
X6	0.93	0.88	0.91	0.68	0.82	1.00	-0.72	-0.60	-0.72	-0.74
X7	-0.76	-0.60	-0.70	-0.36	-0.67	-0.72	1.00	0.76	0.59	0.63
X8	-0.69	-0.57	-0.63	-0.39	-0.57	-0.60	0.76	1.00	0.74	0.78
X9	-0.79	-0.72	-0.79	-0.41	-0.72	-0.72	0.59	0.74	1.00	0.83
X10	-0.79	-0.80	-0.82	-0.48	-0.75	-0.74	0.63	0.78	0.83	1.00

Among them, X1 represents the proportion of urban population to the total population; X2 represents the proportion of employees in the second and third industries to the total number of workers; X3 represents the per capita GDP; X4 represents the per capita use of foreign capital; X5 represents the proportion of higher education or above in the population over 6 years old; X6 represents the household consumption level; X7 represents household size per household; X8 represents the birth rate; X9 represents the number of surviving children per woman aged 15-64 years; X10 represents the total population burden factor.

It can be seen from the correlation matrix of the observation data in Table 2 that there is a strong positive correlation between the six indicators representing the social cohesion and the proportion of the total population; There is also a strong positive correlation between the four indicators representing the family structure; there is a negative correlation between the six indicators representing social cohesion and the four indicators representing family structure. In summary, the metrics of social cohesion and family structure have a high correlation, which is basically consistent with the previous assumptions in this paper.

3.2.2 KMO test and Bartlett sphericity test for determining whether the factor analysis is feasible

The KMO statistic (Kaiser-Meyer-Olkin) test statistic is an indicator used to compare simple correlation coefficients and partial correlation coefficients between variables [3]. The KMO statistic is between 0 and 1. If the value of the KMO statistic is larger, the correlation between the observed original variables is stronger, and vice versa, the weaker the correlation. Whether the correlation between variables determines whether it is suitable for factor analysis of the original data, the stronger the correlation between the original variables, the more suitable for factor analysis, and the weaker the correlation between the original variables, the less suitable for factor analysis. If the KMO value exceeds 0.9, it means that it is very suitable for factor analysis; if the KMO value is between 0.8 and 0.9, it means that it is suitable; if the KMO value is between 0.7 and 0.8, it means that it is suitable; if the KMO value is between 0.6 and 0.7 Between,

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

it means that it is not suitable; if the KMO value is between 0.5 and 0.6, it means that it is barely suitable; and if the KMO value is less than 0.5, it means that it is not suitable.

In this paper, the KMO statistic of 10 indicators representing social cohesion and family structure is calculated by R language software psych package [3]. The overall KMO statistic of observation data is 0.85, which is between 0.8 and 0.9, indicating that it is suitable for factor analysis.

Bartlett's ball test is used to test whether the correlation matrix is a unit matrix, ie whether the variables are independent. It is based on the matrix of correlation coefficients of variables, and the null hypothesis: the matrix of correlation coefficients is a unit matrix. If the statistical value of the Bartlett's ball test is large and the probability value is less than the level of significance given by the user, the null hypothesis should be rejected; otherwise, the null hypothesis cannot be rejected, and the correlation coefficient matrix may be a unit matrix. So it is not suitable for factor analysis. If the hypothesis cannot be denied, it means that these variables may provide some information independently and lack common factors. Normally, if Bartlett's ball test shows a p-value of less than 0.05, then there is a correlation between the observed variables and factor analysis can be performed. As shown in Table 3, the data observed in this paper all pass the KMO test and the Bartlett's ball test, so factor analysis can be used.

Table 3. KMO and Bartlett's ball test results

The KMO statistic test	KMO value	0.85
Bartlett's ball test	Chi-Square	1344.327
	Degree of freedom	45
	P value	3.24146e-252

3.2.3 Determine the number of common factors to be extracted

Figure 2 shows the factor graph drawn by the R language psych package [4]. The factor graph not only can show the bending variation of the data feature value curve, but also can generate the random number matrix of the same dimension as the original matrix. The factor graph will display the principal component and the common factor simultaneously. Results of the analysis. The factor graph result shows that two factor numbers should be selected.

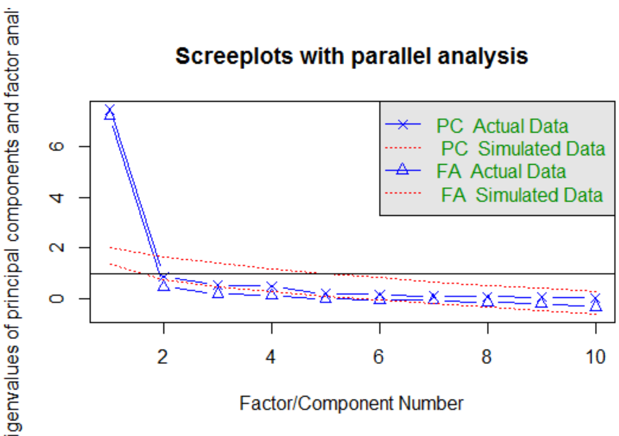

Figure 2. Factor graph

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

3.2.4 Extract common factor

Select p(p<n) principal factors (marked with F) so that the sum of the variance contribution rates of the p principal factors accounts for more than 85% of the total variance contribution rate, or the eigenvalue corresponding to the main factor is greater than 1, then It means that these main factors basically retain the information of the original analysis indicators, and reduce the original n analysis indicators to p factors, achieving the purpose of simplifying the analysis index optimization model.

There are many methods for extracting common factors, including maximum likelihood method (ml), spindle iteration method (pa), weighted least squares (wls), generalized weighted least squares (gls), and minimum residual method (minres). Since the maximum likelihood method has better statistical properties, this paper uses the maximum likelihood method to extract common factors. The results of extracting common factors are shown in Table 4. The factor variance and its contribution are shown in Table 5.

Table 4. Maximum likelihood factor load result without rotation

Indicators	Factor1	Factor2
The proportion of urban population to the total population	0.78	0.58
The proportion of employees in the second and third industries to the total number of workers	0.67	0.62
Per capita GDP	0.73	0.63
Per capita use of foreign capital	0.46	0.43
The proportion of higher education or above in the population over 6 years old	0.67	0.59
Household consumption level	0.7	0.64
Household size per household	-0.8	-0.19
Birth rate	-0.99	0.15
Number of surviving children per woman aged 15-64	-0.79	-0.3
Total population burden factor	-0.83	-0.29

Table 5. Factor variance and contribution rate

	Variance	Contribution rate	Cumulative contribution rate
Factor1	5.68	0.71	0.71
Factor2	2.30	0.29	1.00

3.2.5 Rotation method

The purpose of establishing a factor analysis model is not only to find the main factors, but more importantly to know the meaning of each main factor in order to analyze the actual problems. If the main factor is found, the typical representative variables of each main factor are not very prominent, and the factor rotation is needed to obtain a satisfactory main factor by appropriate means.

There are many methods for factor rotation. The orthogonal rotation and the oblique rotation are two kinds of methods of factor rotation. This article uses the oblique rotation method. Table 6 shows the result of the factor load matrix after the oblique rotation method.

Table 6. Maximum likelihood factor load result with the oblique rotation

Indicators	Factor1	Factor2
The proportion of urban population to the total population	0.87	-0.14
The proportion of employees in the second and third industries to the total number of workers	0.93	0.02
Per capita GDP	0.94	-0.03
Per capita use of foreign capital	0.65	0.02
The proportion of higher education or above in the population over 6 years old	0.89	-0.01
Household consumption level	0.96	0.01
Household size per household	-0.28	0.60
Birth rate	0.23	1.16
Total population burden factor	-0.45	0.46
Number of surviving children per woman aged 15-64	-0.42	0.51

It can be seen from the results of the factor loading matrix after the oblique rotation method that the six indicators representing the social cohesion variable have a higher load on the factor 1, and the factor 1 is the new social cohesion factor; The four indicators representing family structure have a higher load on factor 2, which can be called factor 2 is the new family structure factor. Through factor scores, we analyze these two newly obtained composite factors—new social cohesion factors and new family structure factors—and ethnic customs and demographics into multiple linear regression models.

4. MULTIPLE REGRESSION ANALYSIS

4.1. Multiple Regression Analysis of Divorce Rate in Different Regions

For the statistical test of the interpretation of differences in divorce regions, this paper uses a multiple linear regression model. In order to reduce the contingency and instability of the divorce rate in the same province, we used the average of the rough divorce rates for the three years from 2014 to 2016 as the explanatory variables of the regression model; using the new social cohesion factor, new family structure factor, ethnic customs and population structure (sex ratio and marriage rate) are used as explanatory variables.

Table 7. Divorce rate regional difference regression model

	coefficients	T statistic	P>|t|
_cons	1.71573	0.551	0.586603
social cohesion	-0.20205	-1.308	0.202669
family structure	-0.72538	-4.498	0.000137 ***
ethnic customs	0.5082	1.97	0.060066**
sex ratio	-0.01954	0.701	0.489688
marriage rate	0.31286	3.439	0.002055***
R2	0.5707		
F	6.647	***	

The results of the regression analysis are shown in Table 7. The family structure has the greatest impact on the regional divorce rate. The regression coefficient is -0.72538. The higher

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

the family structure index, the more buffering effect between relatives on the conflict between husband and wife. There are more embarrassments between husband and wife, and the intention and possibility of divorce will be relatively lower. Urbanization, modernization, and the improvement of the educational and cultural level may make both men and women more free from the shackles of the old-time arranged marriage, the combination of the two sides is more due to the pursuit of true love. Under certain other conditions, the probability of divorce in ethnic minority areas is relatively high; under other conditions, the marriage rate in previous years is directly proportional to the current divorce rate; however, social cohesion and sex ratio have little effect on the divorce rate [8].

4.2. Discussion on Regression Analysis Results Based on Entropy Weight TOPSIS

It can be seen from the above that the new family structure factor is the most important factor affecting the regional difference of divorce rate. This paper uses the entropy weight TOPSIS method to estimate Household size per household, the birth rate, the total population burden factor and the number of surviving children per capita for 15-64 years old women. These four indicators representing family structure are comprehensively evaluated and analyzed.

4.2.1 The basic principle of entropy weight TOPSIS

The entropy weight TOPSIS method is an objective weighting method that determines the index weight based on the size of the information provided by each observation. There are m evaluation plans and n evaluation indicators to form the original data matrix $X = (x_{ij})_{m*n}$ For a certain index x_j ,the larger the difference of the index value x_{ij}, The greater the role of this indicator in the overall evaluation; If the indicator values of an indicator are all equal, the indicator does not work in the comprehensive evaluation. Entropy is a measure of uncertainty. The larger the amount of information, the smaller the uncertainty and the smaller the entropy; According to the characteristics of entropy, we can judge the randomness and disorder degree of a scheme by calculating the entropy value. We can also use the entropy value to judge the degree of dispersion of an index. The greater the degree of dispersion of the index, The greater the role of this indicator in the overall evaluation. Therefore, according to the degree of variation of various indicators, the information entropy can be used to calculate the weight of each indicator, which provides a basis for multiple comprehensive evaluations.

4.2.2 A brief description of the steps of the entropy weight TOPSIS method

Firstly, the non-negative processing of data is performed. Since the entropy method calculates the proportion of a certain indicator of each scheme to the sum of the same index value, there is no dimension effect, and no standardization is needed. If there is a negative number in the data, data translation is required (X_{ij} is non-negative processing).

For larger indicators:

$$X_{ij}^{'} = \frac{X_{ij} - \min(X_{1j}, X_{2j}, \cdots, X_{nj})}{\max(X_{1j}, X_{2j}, \cdots, X_{nj}) - \min(X_{1j}, X_{2j}, \cdots, X_{nj})} + 1 \qquad ,i = 1,2,\cdots,n;\ j = 1,2,\cdots,m$$

For smaller indicators:

$$X_{ij}^{'} = \frac{\max(X_{1j}, X_{2j}, \cdots, X_{nj}) - X_{ij}}{\max(X_{1j}, X_{2j}, \cdots, X_{nj}) - \min(X_{1j}, X_{2j}, \cdots, X_{nj})} + 1 \qquad ,i = 1,2,\cdots,n;\ j = 1,2,\cdots,m$$

Then the proportion of the i-th plan under the j-th indicator is:

$$P_{ij} = \frac{X_{ij}}{\sum_{i=1}^{n} X_{ij}} \qquad (j = 1,2,\cdots m)$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

Secondly, calculate the entropy of the jth indicator $e_j = -k * \sum_{i=}^{n} p_{ij} \log(p_{ij})$, and k>0, Log is the natural logarithm, $e_j \geq 0$, the constant k is related to the number of samples m, Generally $k = \frac{1}{lnm}$, then $0 \leq e \leq 1$;

Then the coefficient of variation of the jth indicator is $g_j = 1 - e_j$. For the j-th index, the greater the difference of the index value X_{ij}, the greater the effect on the scheme evaluation, the smaller the entropy value, so the larger the g_j is, the more important the index is.

Third, calculate the weight $w_j = \frac{g_j}{\sum_{j=1}^{m} g_j}, j = 1,2 \ldots m$, Finally calculate the comprehensive score of each program $S_i = \sum_{j=1}^{w} W_j * P_{ij} (i = 1,2, \ldots n)$

4.2.3 Analysis of the relationship between family structure and divorce rate

From the structure of multiple regression analysis, family structure factors are the most important factors affecting regional differences in divorce rates. Through the entropy weight TOPSIS method, we obtained a comprehensive score of the family structure of 31 provinces. The relationship between family structure and divorce rates in various regions is shown in Figure 3.

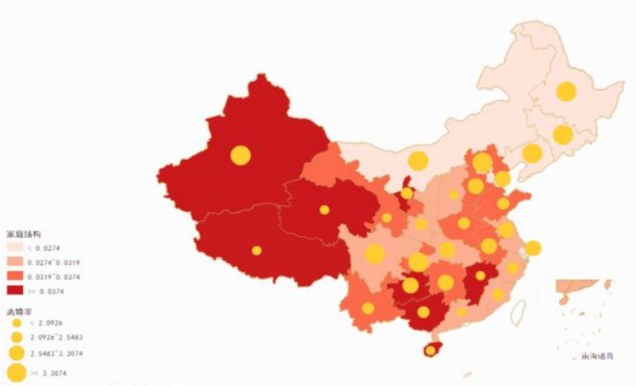

Figure 3. Family structure and divorce rate in various regions

First of all, the family structure has an important influence on the impact of divorce rates in various regions. The reasons for the low divorce rate in Ningxia Hui Autonomous Region, Tibet Autonomous Region, Yunnan Province, Hainan Province, Guangxi Province, Jiangxi Province, Qinghai Province and Gansu Province are due in large part to the fact that most of these areas are inherited from family values and fertility concepts, for example "multiple sons and more blessings" and "nurturing children to prevent old age".

The reasons for the high divorce rates in Beijing, Tianjin, Inner Mongolia Autonomous Region, Liaoning Province, Jilin Province, Heilongjiang Province, Chongqing City and Sichuan Province are that family cohesion in these areas is low. Due to the high level of economic development in Beijing, Tianjin, Chongqing and Sichuan provinces, the concept of fertility advocating more eugenics and family concepts, and the medical facilities and social security are relatively perfect, The risk of divorce between men and women is small, The divorce rate will increase accordingly; The three provinces of Northeast China (Liaoning Province, Jilin Province and Heilongjiang Province) have relatively open marriage concept, and low fertility rate influenced by Russian culture, These will have an impact on the marriage relationship between men and women.

5. CONCLUSION AND DISCUSSION

Family structure, especially children, has the greatest impact on the role of stable parental marriage. Areas with developed economics, advanced fertility concepts and better social security systems will have a higher probability of divorce because of fewer family restraints and

ISSN: 2472-3703 DOI: 10.6911/WSRJ.201909_5(9).0011

ties. The concept of fertility and family concept containing the traditional culture of "multiple sons and more blessings" and "supporting the elderly and bringing young children" retains the low probability of divorce [2]. Although in the current society, we are more tolerant of divorce, but this high divorce rate in China is not conducive to the stability of marriage, family and even society. Every adult should consider the education and growth of their children, marital ethics and family responsibilities when making the decision of divorce.

REFERENCES

[1] Bai Caiquan, Chen Yu, Ding Yifan: Temporal and spatial pattern and convergence evolution of China's provincial divorce rate, Surveying and Mapping Science, Vol. 11(2015), p.48-53. (In Chinese)

[2] Su Liyun, Liu Yang, Peng Xiangwu: Analysis of the spatial aggregation and time-space pattern of divorce rate in various provinces of China, Population Research, Vol.6 (2015), p.74-84. (In Chinese)

[3] Wang Binhui: Multivariate Statistical Analysis and R Language Modeling (Jinan University Press, China, 2010). (In Chinese)

[4] Wang Binhui: Data Statistical Analysis and R Language Programming (Peking University Press, China, 2014). (In Chinese)

[5] Wang Guohua: From the acquaintance society to the stranger society: the sociological perspective of the high divorce rate in society, Journal of Beijing University of Science, Vol.1 (2015), p.5-9. (In Chinese)

[6] Xu Anqi, Yan Yongfu: Analysis of the characteristics of high divorce rate in Xinjiang Uygur area and its causes, Chinese Population Science, Vol.2 (2001), p.25-35. (In Chinese)

[7] Xu Anqi, Ye Wenzhen: Analysis of Regional Differences in Divorce Rate in China, Population Research, Vol.4 (2002), p.28-35. (In Chinese)

[8] Ye Wenzhen, Lin Qingguo: Analysis of the divorce situation and causes in contemporary China, Population and Economy, Vol.3 (1998), p.22-28. (In Chinese)

[9] Zhang Yi: China's current marital situation and trends, Hebei Academic Journal, Vol.3 (2008), p.6-12. (In Chinese)