

Impact of Higher Education Resources Input on Population Agglomeration: Evidence from China

Jiang Nan*

School of Finance, Capital University of Economics and Business, Beijing, China

*Corresponding author: jiangnanxileng@sina.com

Abstract. The development of regional economy has always been the research hotspot of economists. The distribution of population is an important factor affecting regional economy. Based on provincial panel data, this paper studies the relationship between higher education resource input and population agglomeration in different provinces. At the same time, the number of health institutions, per capita GDP and urban population ratio are controlled for explanatory variables and binary variables by region. The results show that higher education resources input has a significant positive effect on population agglomeration. From the perspective of the 31 provinces divided into eastern, central and western regions, the higher education resources input in the eastern and central regions has an obvious positive effect on population agglomeration, and the effect in the central region is higher than that in the eastern region. Based on this, the paper puts forward some suggestions, such as paying attention to education equity, promoting the balance of higher education resources, enhancing the independent development ability of higher education in central and western China, improving the matching degree between the specialty setting of higher education and the demand of labor market, and strengthening the policy guidance of employment and entrepreneurship.

Keywords: Higher Education Resources; Population Agglomeration.

1. Introduction

1.1 Research Background

For China's urbanization development and regional economic growth, population agglomeration is of great significance. Economic development is inseparable from the labor force with high level and high matching degree. A large number of studies show that the rational migration and agglomeration of population is an important driving factor of China's economic development, and human capital plays a positive role in promoting economic growth.

Theoretically speaking, stable population migration and agglomeration can help narrow the gap between regions. Due to the change of per capita resources and the limited population carrying capacity of regions, rational people will naturally migrate from places with relatively large populations to places with relatively small populations, so that the relationship between supply and demand in various regions tends to balance. However, China's current migration path is relatively immature. After the migration, the macro-economy does not present a balanced state, but expands the development gap between regions. Since the reform and opening up, China's economy has achieved rapid and steady development. However, due to China's vast territory, in the process of sustained economic development, the great differences in factor endowment, culture and public governance between regions are the main reasons for the obstruction of reasonable migration of population and the imbalance of regional development. It is a common phenomenon of human social development that population gathers in cities and more developed regions. In the process of population gathering, technology advances and regional economy grows, which further widens the economic gap between regions. The agglomeration of population also brings great challenges to cities with rapid development. Excessive urbanization, such as congestion and environmental pollution, hinders the high-quality development of cities, while relatively backward areas face a substantial loss of basic labor force and high-level talents, leading to a vicious circle of stagnant development and gradually expanding talent gap.



Fig. 1 China's population distribution in 2018

1.2 Literature Review

According to the poverty Alleviation Plan for 2020, the CPC Central Committee has proposed that "poverty alleviation through education" can be regarded as the root of poverty alleviation and ignorance, whether from the perspective of educated individuals or poor families, and is a key measure to eradicate the root of poverty and break the intergenerational transmission of poverty. It plays a leading, fundamental and basic role in the task of poverty alleviation, among which the return of higher education is the highest. As a result, people attach more importance to higher education. However, there is a certain degree of imbalance in the development of higher education among different regions in China. The unbalanced spatial pattern of higher education will lead to misallocation of factor resources under the action of the market. Therefore, this paper holds that the answer to population agglomeration can be found from the input of higher education resources, and the rational guidance of population agglomeration lies in the scientific allocation of higher education resources.

Education is an important way to promote population migration and change population burden into national wealth. Many scholars have studied the relationship between population agglomeration and education. On the one hand, some scholars analyze the migration decision of the subject from the perspective of education level and explain the conclusion that education promotes migration [1]. With the continuous improvement of citizens' years of education, their ability to search for information is also constantly enhanced. More information will help reduce the risk of migration, thus improving the willingness and probability of migration, and providing a stable information environment for population agglomeration. The improvement of education level also plays a positive role in the improvement of individuals' comprehensive quality. Education improves individuals' adaptability, employability and acceptance of different cultures, and reduces the psychological cost of migration. The impact of education on population migration is further amplified in the stage of higher education, and the expansion of higher education promotes population migration[2]. On the other hand, some studies show that educational resources attract population agglomeration, and the main reason why people migrate to large and medium-sized cities is that the rich educational resources in large and medium-sized cities are extremely attractive. Wang Zhiyong's (2017) research shows that the scale of basic education has a significant impact on population agglomeration, and education migration has become a new driving force to promote population agglomeration. In the process of building new urbanization and urban population distribution, full attention should be paid to the guiding role of basic education[3]. Tong Yufen and Liu Hui (2018) found that high-end medical facilities, quality of compulsory education resources, agglomeration level of tertiary industry and knowledge spillover play a centripetal role to attract highly educated population based on relevant data of Beijing-Tianjin-

Hebei region[4]. In addition to compulsory education, Liu Ye (2019) found in her research that the agglomeration level of institutions of higher learning is an important factor determining the spatial distribution of labor force[5].

2. Theoretical Mechanism

2.1 Theoretical Analysis

The impact path of higher education resources on population agglomeration is mainly reflected in the following aspects:

First, families hope their children can get more and better educational resources and growth environment to make migration decisions. In the process of economic transformation and development, education is playing an increasingly important role in personal income and social status. Chinese families pay more and more attention to education investment, and gradually change the concept of education as a consumption to education as an investment that benefits the whole life. For the working population in the period of choosing the city to settle down, among all the necessary elements of the city, more and more attention is paid to the educational resources of the city, so as to make early plans for the future education of their children. People are more intuitively aware that the richer and higher quality of higher education resources in a region, the stronger the foundation of basic education, and the regional tilt of college enrollment. They have easier access to higher education and a wider choice of schools nearby. For the eastern developed regions, although there are abundant higher education resources, a large number of people have gathered, and the per capita access to higher education is not high. With the great pressure of housing price and living cost, it is not the optimal choice for people to move to the city. For economically backward regions, if higher education resource allocation is improved, it will have an advantage to attract the population in terms of the educational competition environment of the offspring, economic cost and the difficulty for children to obtain high-quality higher education opportunities.

Second, the regional destination preference of college graduates has a certain stickiness. From the Angle of graduates, the university entrance exam candidates invariably hope I can go to the best within their power level or accept the ideal university of higher education, but education resources in less developed areas is less number of brand-name colleges and universities, undergraduate college total amount is not high, make students when choosing to other abundant education resources flow area is economically developed regions. During their higher education, these students have adapted to the local life and the speed of the city. Faced with the choice of working and settling places, they are more inclined to stay in the region after graduation due to the continuous updating of their vision and experience and the existing connections. On the other hand, due to the scarcity of higher education resources, fewer students from the eastern region flow into the western region, which makes the central and western regions more isolated in terms of education and exchange development and lack of development potential.

Thirdly, the spatial agglomeration of higher education resources gives the region the advantage of human capital accumulation, which affects the long-term income of individual migration and then affects the decision of population migration. Areas with more resources of higher education have more talents. By working with talent groups, it is easier to improve their productivity through face-to-face "learning" and imitation, which is called the "learning effect". In modern cities where the importance of science and technology becomes increasingly prominent, knowledge and ideas are more likely to be generated and spread in high-density cities, and the role of knowledge spillover will become more obvious. In the process of knowledge spillover, "learning effect" is its important manifestation. No matter for low-skilled workers or high-skilled talents, the learning effect of a region attracts people to gather. It is particularly important for the economic development of relatively backward areas to make clear under what circumstances population agglomeration can be realized and sufficient labor supply and advanced knowledge and technical personnel can be obtained, which is a exploring way to remedy the imbalance of economic development between regions in China.

2.2 Variable Declaration

According to the latest development statistics bulletin of the Ministry of Education, by 2019, the total number of higher education students in China had reached 40.2 million, with a higher education gross enrollment rate of 51.6 percent. There were 2,688 regular institutions of higher learning in China, 25 more than the previous year, an increase of 0.94 percent. Among them, 1,265 are undergraduate universities, 20 more than last year; There were 1,423 higher vocational colleges (specialized colleges), 5 more than the previous year. There were 268 adult institutions of higher learning in China, 9 fewer than the previous year. There were 828 graduate training institutions, including 593 institutions of higher learning and 235 research institutes. The average size of regular institutions of higher learning is 11,260, of which 15,179 are from undergraduate colleges and 7,776 are from higher vocational colleges. Taking into account the explanatory ability of indicators and the availability of data, and exploring from the perspective of higher education resource investment, this paper intends to select the following explanatory variables, target explanatory variables and control variables of 31 provinces, municipalities and autonomous regions in China from 2013 to 2018.

Total population (ten thousand) : the number of permanent residents in each region at the end of the year calculated from the census. The report represents the "number" of population agglomeration, which is the explained variable.

Public financial expenditure and infrastructure expenditure of colleges and universities (100 million yuan) : This target variable is the sum of educational expenses and infrastructure expenditure of public financial budget of higher education institutions in China Education Statistical Yearbook. In the report, represents the government's investment in education and is introduced as a target explanatory variable. However, this variable only analyzes the regional education level from the perspective of "quantity" of investment, and does not involve the impact of "quality" of investment on education level, such as the number of senior teachers in colleges and universities.

(Note: Higher education institution = ordinary higher education institution + adult higher education Institution + other private higher education institution)

Proportion of urban population (%) : By region, the proportion of people living in urban areas for more than 6 months to the total urban population. The urbanization rate reflects the local economic level and the degree of population aggregation, and may affect the increase of investment in education, thus acting as a control variable.

Number of health institutions (number) : Number of medical and health institutions include hospitals, health centers, sanatoriums, out-patient departments, clinics, health centers (rooms), first-aid stations, village clinics, family planning technical service institutions and private clinics. This variable usually represents the infrastructure of a region, which may have a crowding out effect with the government's investment in education. At the same time, the number of health institutions can indirectly reflect the economic situation and population of a region, so it acts as a control variable.

Per capita GDP (one hundred million yuan/ten thousand) : will a regional accounting period (usually one year) to realize the regional GDP than the region's resident population (or census register population), is a measure of national people's living standards, one of the most important macroeconomic indicators, often used as a measure of economic development in development economics. This variable is obviously related to the explained variable (calculated by it), and also represents the economic strength of a region, which may indirectly affect the amount of education investment in the region, so it is used as a control variable introduction.

2.3 Economic Intuition

Table 1 The economic intuition of variables

Variable types	variable name	Symbolic representation	Economic intuition
Target explanatory variable(+)	Public financial Expenditure and infrastructure Construction Expenditure of colleges and universities (100 million Yuan)	Expenditure	At the regional level, the higher the financial investment in higher education, the stronger the ability to attract talents, the higher the degree of population agglomeration, there is a significant positive correlation between the two
Control variables(+)	Proportion of urban population(%)	Proportion_urban	At the regional level, the larger the proportion of urban population is, the better the regional economic development is and the higher the degree of population agglomeration is. There is a positive correlation between the two
Control variables(+)	Number of health institutions (number)	Health_Institutions	At the regional level, the more health institutions there are, the better the infrastructure and the higher the degree of population agglomeration are. There is a positive correlation between the two
Control variables(+)	Per capita GPD (100 million yuan / 10 thousand people)	Capita_GDP	At the regional level, the higher the PER capita GDP, the better the regional economic development, the higher the degree of population agglomeration, there is a positive correlation between the two

Control variables(+)	Eastern region (binary variable)	east	Divided into central, eastern and western regions, the eastern region on average has more people than the southern region. From a simple analysis of the data, we believe that the order from strong to weak of education investment, economic development and infrastructure construction in eastern, central and western regions is east > Central > west, so the coefficient is predicted to be positively correlated
Control variables(+)	Central region (binary variable)	middle	Divided into central, eastern and western regions, the eastern region on average has more people than the southern region. From a simple analysis of the data, we believe that the order from strong to weak of education investment, economic development and infrastructure construction in eastern, central and western regions is east > Central > west, so the coefficient is predicted to be positively correlated

3. Data Declaration

3.1 Data Collection and Collation

Higher education belongs to the advanced stage of education system, which plays an important role in the process of China's economic reform and development. Since 1999, when the Ministry of Education issued the Action Plan of Revitalizing Education for the 21st Century, China's higher education has entered a period of rapid development. Represented by college students, who account for more and more of the total social population, China's higher education is no longer a rare elite education, but gradually become the public education is chosen by people. The rapid expansion of higher education has greatly promoted the development of China's economy and society.

Considering the explanatory ability of indicators and the availability of data, the data in this paper are mainly from the annual China Education Statistics Yearbook, National Bureau of Statistics, Ministry of Education and Ministry of Finance. In the variable selection of higher education resources, educational resources, also known as "educational economic conditions", refer to the human, material

and financial resources occupied, used and consumed in the education process. Therefore, this paper selects the educational expenses and capital construction expenditure of public financial budget of higher education schools as explanatory variables of the target to reflect the quantity of higher education resources input. The higher education institutions mentioned above refer to the total of ordinary institutions of higher learning, adult institutions of higher learning and other private institutions of higher education. According to document No. 33 of Guofa (2000), the eastern, central and western regions are distinguished. The eastern region refers to Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan. The central region refers to Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei and Hunan; The western regions are Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang, Guangxi and Inner Mongolia.

Table 2 Collection and sorting of variables

Variable types	Variable name	Symbolic representation	Data declaration	Data sources
/	Year	Year	T =6, 6 years of data from 2013 to 2018	/
/	Provinces	Province_id	31 provinces, municipalities and autonomous regions (provinces) in mainland China were collated into panel data form by Stata command	/
Explained variable	Total population (ten thousand)	Population	The number of permanent residents in each region at the end of the year is calculated from the census	China Statistical Yearbook 2013-2018
Target explanatory variable	Public financial Expenditure and infrastructure Construction Expenditure of colleges and universities (100 million Yuan)	Expenditure	Public financial budget of institutions of higher education education expenses + capital construction expenses (note: institutions of higher education = ordinary institutions of higher learning + adult institutions of higher learning + other institutions of higher	China Statistical Yearbook 2013-2018

			education run by the people)	
Control variables	Proportion of Urban Population (%)	Proportion_urban	Population living in urban areas for more than 6 months/total urban population by region	China Statistical Yearbook 2013-2018
Control variables	Number of health institutions (Number)	Health_Institutions	Medical and health institutions include hospitals, health centers, sanatoriums, out-patient departments, clinics, health centers (rooms), first-aid stations, village clinics, family planning technical service institutions and private clinics	China Statistical Yearbook 2013-2018
Control variables	Per capita GPD (100 million yuan / 10 thousand people)	Capita_GDP	Gross domestic product by region/total resident population by region	China Statistical Yearbook 2013-2018
Control variables	Eastern region (binary variable)	east	Eastern region: Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan. Command "East =1" for the above regions and "East =0" for the rest.	/

Control variables	Central region (binary variable)	middle	Central region: Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei and Hunan. Set the preceding areas to "middle=1" and other areas to "middle=0".	/
-------------------	----------------------------------	--------	---	---

3.2 Descriptive Analysis

Table 3 Descriptive statistics

Variables	Obs	Mean	Std. Dev.	Min	Max	Skew.	Kurt.
Population	186	4437.21	2795.705	312	11346	0.627	2.613
Expenditure	186	9329.556	5599.196	1464.513	32980.109	1.293	5.126
Proportion urban	186	57.244	12.755	23.71	89.6	0.589	3.805
Health Institutions	186	31503.753	21912.348	4140	81070	0.852	2.955
Capita GDP	186	5.55	2.53	2.309	14.076	1.329	4.149

3.3 Scatter Plots and Fitting Lines

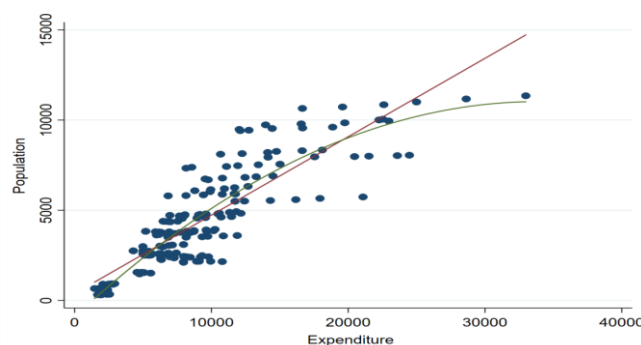


Fig. 2 Population and Expenditure

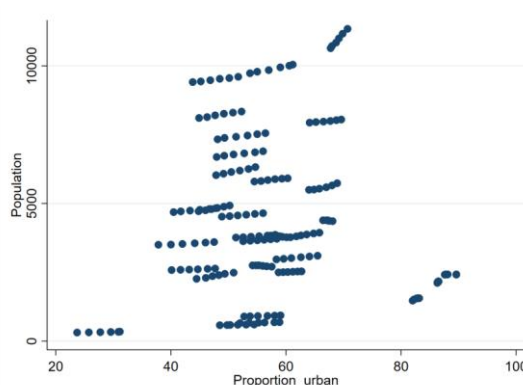


Fig. 3 Population and Proportion_urban

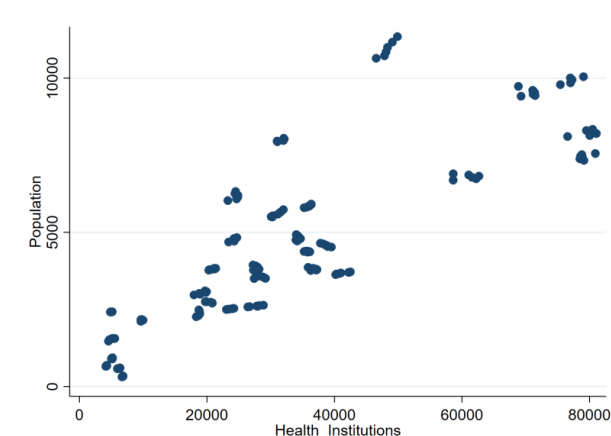


Fig. 4 Population and Health_Institutions

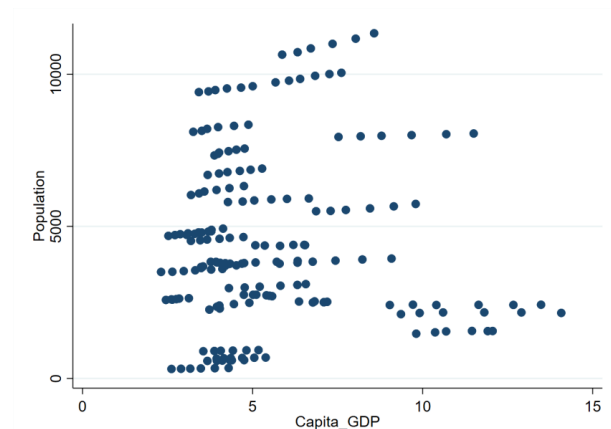


Fig. 5 Population and Capita_GDP

According to the scatter diagram, the target explanatory variable and the explained variable are fitted linearly and nonlinear, while the other control variables have no good practical significance for linear or nonlinear fitting due to the obvious fixed effect, so they are only displayed as scatter diagram. It also shows that the function of control variables can be more reasonably reflected after the data is processed by panel data. Based on the analysis of the images and the descriptive statistics table, we can preliminarily conclude that this set of data meets the third assumption of OLS: there is no outlier. (We are not strictly following the general principle of "less than 4 standard deviation" here)

3.4 Preliminary Judgment on the Standard Deviation of Heteroscedasticity Robust Form

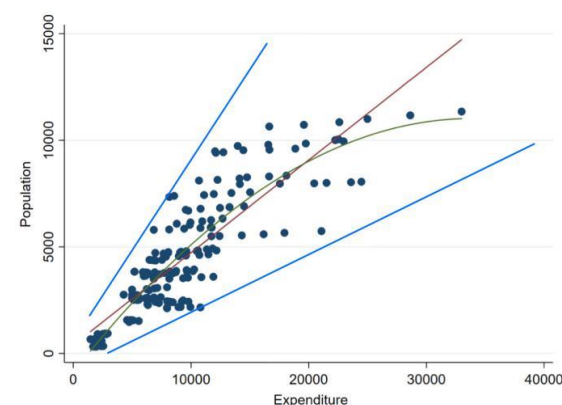


Fig. 6 Heteroscedasticity and homoscedasticity tests

Here we only make a preliminary judgment of heteroscedastic robustness. As shown in the figure (total population and expenditure), our explained variables have different degrees of dispersion in the case of given target explanatory variables. The quality of meet the heteroscedasticity $E(u | x) = f(x)$, residual conditional expectation is very few, therefore, after the return to command, we all take different variance solid form, namely in the Stata command to add the "r" and "cluster".

4. Independently Pooled Cross Section

4.1 Selection of Control Variables in Regression Equation Functions

Table 4. Choice of regression function

Explained variable:Population				
	(1)	(2)	(3)	(4)
Expenditure	0.435*** (0.0233)	0.454*** (0.0222)	0.282*** (0.0174)	0.317*** (0.0198)
Proportion_urban		-41.24*** (7.426)	-1.800 (4.531)	43.16*** (8.535)
Health_Institutions			0.0628*** (0.00429)	0.0556*** (0.00487)
Capita_GDP				-287.3*** (53.17)
Constant	382.2** (192.1)	2,566*** (507.9)	-65.79 (296.1)	-1,143*** (292.7)
Observations	186	186	186	186
R-squared	0.758	0.792	0.908	0.919
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Based on relevant literature review, the group only selected "public financial expenditure and infrastructure expenditure of colleges and universities" as explanatory variables, and the other variables were control variables. Before selecting the function form of regression equation, we first test and analyze the control variables selected by this group. With the addition of control variables one by one in columns (1) to (4), the coefficients in front of the target explanatory variables gradually approach, and the coefficients in front of the target explanatory variables in the subsequent addition of fixed effects become more real. At the same time, with the addition of control variables, the model is gradually improved, which can better explain the data. Although in column (3), the coefficient of the control variable urban population proportion is no longer significant, since we finally choose the functional form of column (4), it does not affect the preciseness of our practical steps.

4.2 Selection of Function Form of Regression Equation

Table 5 Choice of regression function

Explained variable: Population						
	(1)	(2)	(3)	(4)	(5)	(6)
Expenditure	0.317*** (-0.0198)	0.298*** (-0.0692)		0.283*** (-0.0179)	0.273*** (-0.0327)	0.325*** (-0.019)
Proportion_urban	43.16*** (-8.535)	45.13*** (-9.646)	5.596 (-12.13)	-0.638 (-5.797)	0.202 (-6.116)	30.76*** (-10.81)

Health Institutions	0.0556*** (-0.00487)	0.0522*** (-0.00554)	0.0604*** (-0.00544)	0.0627*** (-0.00437)	0.0636*** (-0.00541)	0.0511*** (-0.00448)
Capita_GDP	-287.3*** (-53.17)	-305.8*** (-55.34)	-64.12 (-66.11)			-246.9*** (-50.98)
exp2		5.24E-06 (-5.43E-06)				
exp3		-1.71E-10 (-1.20E-10)				
lnexp			2,363*** (-228.7)			
D_GDP				-49.21 (-185.1)	-155.1 (-262.3)	
exp_DGDP					0.0111 (-0.0313)	
east						283.8 (-188.3)
middle						689.9*** (-172.9)
Constant	-1,143*** (-292.7)	-1,163*** (-249)	- 18,588*** (-1,869)	-122.7 (-326.5)	-121.6 (-327.2)	-860.5** (-392.2)
Observations	186	186	186	186	186	186
R-squared	0.919	0.921	0.867	0.908	0.908	0.928
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						

Our group first conducted multiple regression for all explanatory variables. As shown in Table (1), all explanatory variables were significant at the significance level of 1%, and the goodness of fit reached 0.919, indicating strong explanatory power for data and good regression results.

In columns (2) and (3), our group wants to explore whether the nonlinear model can better explain the data under the condition of the original control variables. In column (2), polynomial regression is adopted. The coefficient estimator of the primary term of the objective explanatory variable (Expenditure) is still significant at the significance level of 1%, while the quadratic and cubic terms are not significant, and the goodness of fit is not significantly improved. Therefore, polynomial regression model is not suitable. Column (3) adopts log-linear model, and the logarithmic term (LnexP) is significant at the significance level of 1%, and the regression results are good. However, the goodness of fit ratio (1) of the multiple linear regression model decreased slightly.

In column (4), our group divided the continuous variable PER capita GDP (Capita_GDP) into discrete binary variable (D_GDP) based on the average per capita GDP of 31 provinces. In other words, provinces are divided into high and low economic development areas, and then the influence of the objective explanatory variable (Expenditure) on the total population is explored to make the results more stable. However, the regression result of D_GDP is not significant, which may be due to the rough division of regional economic conditions into two categories, which may lose a large number of individual characteristic differences. As a result, the explanation of face change contains too little information and the regression result is inaccurate. In column (5), interactive effects are added, that is, our group assumes that the impact of the target explanatory variable (Expenditure) on the total population depends on whether the region is an area of high economic development.

However, the regression result of the interaction term (exp_DGDP) is not significant, which may be the same as that of column (4). The binarization of continuous variables loses a lot of information.

In column (6), we added east and MIDDLE variables to our group. That is, our group divided 31 provinces into eastern, central and western regions according to geographical location. Our group assumes that migration is also related to the geographical location of the region. Therefore, as a control variable, the influence of geographical location on the total population of the explained variable is excluded. The regression results verify that when other conditions remain unchanged, the total population of the region increases by 0.325 million for every 100 million yuan increase in the public financial expenditure and infrastructure construction expenditure of colleges and universities.

Although the regression results of column (6) are also very good, considering that variables East and MIDDLE do not change with time, we adopt the column (1) multiple linear regression model in the panel processing below.

4.3 VIF Multicollinearity Test

Table 6 VIF multicollinearity test

	(1)	(2)	(3)	(4)	(5)	(6)
Expenditure	2.498	57.173		2.336	7.338	2.655
Proportion_urban	5.042	5.548	4.941	2.277	2.477	5.465
Health_Institutions	2.369	2.878	2.691	2.139	2.893	2.725
Capita_GDP	5.977	6.351	5.331			7.013
exp2		242.731				
exp3		87.877				
lnexp			2.691			
D_GDP				2.442	10.888	
exp_DGDP					6.975	
east						2.289
middle						1.387
Mean VIF	3.972	67.093	3.925	2.298	6.114	3.589

The multicollinearity problem of regression results is discussed. Although the Mean VIF = 67.093 in the regression of column (3) is significantly higher than the test standard VIF = 10 required by the empirical rule, the function form of this regression equation is not selected due to the lack of explanatory power. However, for the regression of column (1) we finally selected, its Mean VIF = 3.972, lower than the test standard VIF = 10 required by the empirical rule. Therefore, it is reasonable to believe that there is no multicollinearity problem in the regression equation function form of column (1).

5. Panel Regression Analysis

This panel includes a total of 186 pieces of relevant data from 31 provinces in 6 years. The following is the panel data regression by adding dummy variables and using the panel commands provided by Stata.

5.1 Add Dummy Variables for Fixed Effect Estimation

Table 7 Add dummy variables for fixed effect estimation

Explained variable: population					
	(1)	(2)	(3)	(4)	(5)
Expenditure	0.317*** (0.0282)	0.334*** (0.0270)	0.0297*** (0.00938)	0.0300*** (0.00963)	0.0317*** (0.00858)
Proportion_urban	43.16*** (15.01)	36.69** (14.95)	4.301 (4.591)	3.621 (7.505)	2.503 (7.241)
Health_Institutions	0.0556*** (0.00807)	0.0534*** (0.00792)	0.00319 (0.00367)	0.00285 (0.00394)	0.00575 (0.00539)
Capita_GDP	-287.3*** (93.71)	-240.9** (91.54)	-5.468 (7.844)	-7.103 (13.34)	-14.12 (16.03)
Constant	-1,143** (540.9)	-1,124** (542.3)	3,844*** (242.6)	1,618** (687.5)	3,887*** (457.6)
Province effects	NO	NO	YES	YES	YES
Time effects	NO	YES	NO	YES	YES
Clustered standard errors	YES	YES	YES	YES	YES
The P-value of the F statistic	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	186	186	186	186	62
R-squared	0.919	0.933	0.999	0.999	0.844
Robust standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

Column (1) shows the OLS regression results of educational Expenditure and capital construction Expenditure from public financial budget of institutions of higher learning on population without considering the individual fixed effect of time, and its coefficient estimation is significant at the significance level of 1%. It is proved that the educational Expenditure and capital construction Expenditure in public financial budget of institutions of higher learning have positive influence on population. It verifies our group's view that there is a positive correlation between the input of higher education resources and the total population. The more higher education resources a region has, the more positive it is to its own population agglomeration. In addition, the coefficient estimation is 0.317, which means that under the condition that other conditions remain unchanged, the total population will increase by 0.317 million for every 100 million yuan increase in educational Expenditure and capital construction Expenditure in the public financial budget of institutions of higher learning.

Column (2) in the table shows that when only the time-fixed effect is considered, some missing variables that only change with time are controlled, such as the country's overall economic situation and population policy. Under this condition, its coefficient estimator is positive (0.334), which is not much different from column (1).

The regression in column (3), which contains only the fixed effect of province, has a significantly higher explanatory power than that in column (2). It rose from 0.933 to 0.999. Coefficient estimator decreased to 0.0297. Individual fixation eliminates differences between regions that do not change over time. Column (4) has both time fixed effect and province fixed effect, and has little change compared with column (3). The coefficient estimators differ by only 0.0003 and remain the same. Clearly, the provincial fixed effect explains most of the variation. Therefore, it can be seen that the omitted fixed factors (such as regional culture, etc.) have a greater impact on the total population.

In the other column (5), the difference method is used, and only the data from 2013 and 2018 are used. The goodness of fit of the model decreases slightly compared with the other three, and the coefficient estimation may have a large deviation.

5.2 Stata Comes with Fixed Effect Estimates for Panel Data

Table 8 Stata comes with fixed effect estimates for panel data

Explained variable: population			
	(1)	(2)	(3)
Expenditure	0.334*** (0.0270)	0.0297*** (0.00857)	0.0300*** (0.00877)
Proportion_urban	36.69** (14.95)	4.301 (4.193)	3.621 (6.835)
Health_Institutions	0.0534*** (0.00792)	0.00319 (0.00335)	0.00285 (0.00359)
Capita_GDP	-240.9** (91.54)	-5.468 (7.165)	-7.103 (12.15)
Constant	-1,124** (542.3)	3,844*** (221.5)	3,895*** (398.3)
Province effects	NO	YES	YES
Time effects	YES	NO	YES
Clustered standard errors	YES	YES	YES
The P-value of the F statistic	0.0000	0.0000	0.0000
Observations	186	186	186
R-squared	0.933	0.799	0.800
Number of Province_id	31	31	31
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

It can be seen that the coefficient estimators of the results of dummy variables are the same as those of fixed effect regression using Stata's own estimation command, and there is only a slight difference in the standard error.

6. Model Selection and Model Evaluation

6.1 Internal Effectiveness Threat Analysis

(1) Omitted Variable Bias

In our multivariate regression, the educational expenses and capital construction expenses of the public financial budget of colleges and universities are the explanatory variables. The control variables of per capita GDP and urban population ratio control the degree of urban modernization (social development) and economic growth. The number of health facilities controls public environmental problems in the area. However, some missing variables, such as the fiscal policy on education expenditure, vary from province to province and are difficult to observe and quantify. And it has an impact on both the explanation and the explained variable, which will produce bias. Our team only selected 6 years, during which it can be approximated that the fiscal policies of each province did not change significantly. Therefore, this variable does not change with time, and the fixed effect can be eliminated in panel regression.

(2) Measurement Errors and Variable Measurement Errors

All the data in our group come from the official China Statistical Yearbook, and there is no measurement error of variables in theory.

(3) Missetting of the Form of a Regression Function

In our group's regression, the analysis of explanatory variables started from multivariate. The main explanatory variable (Expenditure) and three control variables in this model are significant, with a

value of 0.919, indicating that our model has a strong explanatory power to the changes of the total population. Moreover, the final regression form VIF tests that there is no complete multicollinearity. So the multiple linear model fits the actual regression function.

In addition, our team also tried nonlinear regression models, including polynomial regression, logarithmic regression, adding interaction terms, etc. East and MIDDLE variables are also introduced to divide the regions of the sample and explore whether the regions have an impact on the explained variable Y.

Finally, it is concluded that the multiple linear regression model with regional variables has the strongest explanatory power. Therefore, there is no possibility of deviation caused by missetting of function form.

(4) Bidirectional Causality

We believe that there may be a two-way causal relationship between education expenditure and population agglomeration and migration. In the process of urban economic transformation and development, education is playing an increasingly important role in personal income and social status. Families pay more and more attention to education investment, and gradually change the idea that education is a kind of consumption to education is a kind of lifelong investment. In other words, cities with a large concentration of population are likely to have more prosperous economic development and people will pay more attention to education. Hence the increase in education spending.

In the period of choosing to settle down in the city, the labor force population is also paying more and more attention to the educational resources of the city in measuring all aspects of the essential elements of the city. People intuitively realize that the richer and higher quality of higher education resources in a region, the better the foundation of basic education will be. Because of the regional skew of college enrollment, access to higher education seems to be easier. Therefore, educational resources will also have a positive impact on population migration and agglomeration.

(5) Sample Missing and Sample Selection

Our group selected the annual data of 31 provinces from 2013 to 2018 as samples. Excluding Hong Kong, Macao and Taiwan, sample selection covers all provincial administrative regions in mainland China, which means there is no sample selection bias.

6.2 Model Selection

(1) Model Results

After establishing multiple linear regression, polynomial regression, linear logarithmic regression, adding interaction terms, individual and time fixed effect regression and other models to analyze, and then related to the actual economic meaning. We finally chose to add time and individual fixed effects to the model.

$$\text{Population} = \beta_0 + \beta_1 \text{Expenditure}_{it} + \beta_2 \text{Proportion_urban}_{it} + \beta_3 \text{Health_Institutions}_{it} + \beta_4 \text{Capita_GDP}_{it} + \alpha_i + Z_t + \mu_{it} \quad (1)$$

Table 9. Descriptive statistics of function expressions

Population	Coef.	Robust Std.Err	t	P> t	[95% Conf. Interval]	
Expenditure	0.030028	0.0087732	3.42	0.002	0.0121107	0.0479452
Proportion_urban	3.620853	6.835123	0.53	0.600	-10.33833	17.58004
Health_Institutions	0.002867	0.0035889	0.79	0.434	-0.0044829	0.0101763
Capita_CDP	-7.103382	12.14627	-0.58	0.563	-31.90938	17.70261
Individual fixation effect	control	control	control	control	control	control
Time fixed effect	control	control	control	control	control	control

The coefficient estimator of the target explanatory variable was 0.03 and significant at the 1% significance level. The economic significance of the coefficient is that under the condition that other conditions remain unchanged, the total population will increase by 0.03 million people when 100 million yuan of educational Expenditure and capital construction Expenditure is increased in the public financial budget of institutions of higher learning.

Through regression, it can be concluded that only the coefficient estimation of per capita GDP is negative among these control variables. However, per capita GDP is a control variable that controls social and economic growth, and it is not significant, so it does not violate the hypothesis of our group. The other factors are all positive. This result reflects that when other conditions remain unchanged, the better regional economic development or infrastructure (medical level) is, the easier it is to attract population, which is consistent with the previous prediction of our group.

7. Research Conclusion

The input of educational resources has positive influence on population migration. The temporal fixed effect alone has little influence on the target explanatory variable (Expenditure), but the coefficient estimation of Expenditure becomes original after the provincial fixed effect is added. It can be seen that the fixed effect of provinces explains most data changes and is the main source of omission bias. The three control variables selected by our group were not significant in panel regression. However, the control variables only control economic conditions and social development. The coefficients don't mean much.

Allocation of educational resources: After China's economic society enters the new normal of transformation and upgrading, it is of great practical significance to continue to increase the financial investment of higher education in central and western regions for the regional human capital stock and long-term economic and social development. Starting from promoting higher education fair chance, you will first need to expanding the scope of the quality of higher education to share, not one of the important task is to continue to strengthen the construction of the higher education school in the Midwest, support and oversee parallel, focus on supporting and fostering a number of distinctive, high level of local universities, promote the Midwest higher education school teaching and scientific research practice ability, Gradually make it consistent with the level of universities in developed regions, in line with the international advanced education concept, increase the number of high quality colleges and universities in the central and western regions, and realize the balanced distribution of high quality higher education resources in different regions.

Improvement of education quality: the overall revitalization of higher education in relatively backward areas is still on the way, and a large part of the obstacle is the ideological constraint of self-denial. We should emancipate our minds and re-understand the equality of regional higher education. We should set high standards for ourselves, establish the consciousness of developing higher education with high aspirations and first-class standards, stimulate the spiritual power of unremitting struggle and pursuit of excellence, effectively improve the positioning of higher education in this region, and comprehensively enhance our "hematopoietic" function. The central and western regions should constantly improve their own scientific research, professional and vocational education and training capabilities, and explore more of the disciplines in which they have advantages, so as to achieve the domestic first-class level in different fields, increase cross-regional cooperation and exchanges with high-quality universities, and constantly improve each other. At the same time, we should gradually abandon the prejudice and social atmosphere of questioning their education level due to geographical location factors. We will gradually achieve equal access to basic public services within the scope of national or local financial resources. Break the shackles that economically developed areas are unwilling to go to more backward areas to receive higher education even if education is degraded.

Guide the employment market, strengthen employment entrepreneurship policies to guide the local government should attach importance to institutions of higher learning in the aspect of "production,

study and research" with the local industrial structure of form a complete set of guidance, can the region over the next five to 10 years of the Labour market demand investigation and forecast, provide the results as the index to colleges and universities as a professional set of reference. On the one hand, the local government should not only build and invest in industries with potential and practical feasibility, but also tap the untapped market demand and cooperate with the talent training plan of colleges and universities. On the other hand, colleges and universities should be encouraged to arrange the number of students in each major reasonably, so that the major setting can match the needs of local industries, so that the talents with higher education can find satisfactory jobs in the region and avoid being forced to move due to the lack of posts and opportunities. Central and local governments can introduce more preferential policies to attract and retain talent. We will encourage senior intellectuals to conduct scientific research and teach in the central and western regions so as to gradually accumulate teachers. Increase the local employment subsidies for college students, preferential policies for talents to buy houses, and actively promote regional confidence, so that excellent talents in economic ability and ideology in the central and western regions can "settle down" and join in the long-term construction of the central and western regions.

References

- [1] Wu Keming. Education and Labor Mobility [M]. Beijing: Beijing Normal University Press,2009.
- [2] Xu Chao. The impact of higher education expansion on labor mobility: an empirical analysis based on provincial panel data [J]. Northwest Population,2015(4):40-45.
- [3] Wang Zhiyong. Basic education and population agglomeration: An analysis based on panel data of prefecture-level cities. Population and Development,2017,(6):14-25.
- [4] Tong Yufen, Liu Hui. Spatial agglomeration and influencing factors of educated population in Beijing-Tianjin-Hebei region [J]. Population Journal,2018(3):5-17.
- [5] Liu Ye, Wang Ruoyu, Xue Desheng, Zeng Yuan. The spatial pattern and influencing factors of high-skilled labor force and general labor force in China [J]. Geographical Research,2019(8):1949-1964.