PRIMARY RESEARCH

Human capital, marketization, and innovation: Evidence from China

Fanghan Liu [1*], Lingyu Qian [2], Tai Ren [3]

[1] Nanjing Foreign Language School, Nanjing, China

[2] Hefei No. 8 Middle School, Anhui, China

[3] Jin'an Middle School of Resort Attached to Kunming No.1 High School, Kunming, China

Abstract

This paper aims to analyze the relationship between the development of human capital and local innovation across China. Based on provincial panel data collected from 2000 to 2016, a multivariate regression model was constructed based on the OLS method to explore the correlation between human capital development and innovation level on a provincial level. Regression results display a significant positive correlation between the development of human capital and the improvement of local innovation level in the past 16 years. After taking into account other determinants of innovation, including the Gross Domestic Product, industrial structure, tax level, and regional openness, the positive relationship is still robust. On this ground, this paper further discussed regional heterogeneity that existed across China by dividing the provinces into Eastern, Middle, and Western regions according to geographical locations. The regression results showed that human capital generally facilitates the innovation progress across China, and its influence is especially significant in the Western region, followed by the East region. The researchers also used alternative indicators of innovation and human capital to increase the main findings' robustness. Meanwhile, the researchers also added interactive terms to the regression model and jointly discussed the influence of external factors such as marketization and high-tech industries on the beneficial effect of human capital. The results show that a highly market-oriented economy and a high proportion of high-tech industries' contribution to GDP will encourage local innovation. Our research indicates that the government should focus on the development of human capital and industrial upgrading in order to promote the local innovation level, especially in the eastern and western regions.

© 2020 The Author(s). Published by TAF Publishing.

INTRODUCTION

During the 40 years of reforming and opening, China has maintained a high pace of economic development. The Gross Domestic Product of China has soared from 458.76 billion Yuan in 1980 to 98.85 trillion Yuan in 2019. As the second largest economy in the world, China had been increasing its GDP at a surprising rate of over 6% during the last decade. Although tremendous achievement has been made, this country still faces various challenges in the new are where international competition is fierce and innovation plays an unprecedentedly important role in economic development. Considering that China is still falling behind internationally in the field of miscellaneous cutting-edge technologies, the need for innovation, as the momentum of traditional development weakens, innovation becomes a necessity for China.

Many recent studies have suggested that knowledge-intensive economic activities have been rising in a high pace recently. In many areas, innovation has become the main driven force of economic development. To promote the pace of innovation and cultivate high-tech industries, the Chinese government has published the National Innovation-driven Development Strategy Program in May, 2016. In this program, the government proposed the key idea that innovation is a "human-oriented" activity and confirmed the importance of talents. This program calls attention to China's

*corresponding author: Fanghan Liu
†email: 1679579394@qq.com

human capital situation: with the greatest population in the world, China has great potential in the field of human capital. In 2019, more than 7.58 million people graduated from universities in China. This number has increased over 794% since 1999, when Chinese universities began to increase their enrollments. However, according to researches done by Organization for Economic Co-operation and Development (OECD), in 2017, only 17% of the Chinese population from 25 to 64 have received tertiary education, while this proportion is higher than 40% in developed countries such as the United States, the United Kingdom, Japan and Korea. Another research of the World Bank in 2018 ranked countries according to their human capital level, and China ranked the 46th. Given that China is insisting on developing the country through science and education, we can expect China's ranking to keep rising in the future.

China has made an enormous amount of investment on education in order to improve its overall technological power. Under this circumstance, we believe that the relationship between human capital and innovation level is worth serious investigation. In this situation, we have concern with the following problems: to what extent can human capital boost innovation, and how does the situation change when the environment changes? Former studies mainly hold a positive attitude towards that relationship: most researchers believe that human capital has an overall positive influence on innovation level. Nevertheless, there is still much to explore on the factors that may influence the relationship between human capital and innovation. In a vast country like China, regional differences are prominent, and the specific effect of human capital across regions can be revealed with analysis of panel data. In this paper, we used linear regression models to discuss the relationship between human capital and innovation on the perspective of space. Furthermore, we discussed a few other possibly influential factors on innovation and finally proposed suggestions for the government to increase the innovation level and promote technological advancements.

Our research consists of 5 main parts. First, we define our variables, create the interaction terms and set up a multivariate regression model. Secondly, we collect the panel data of 31 Chinese provinces from 2000 to 2016. Thirdly, we perform a regression with the basic model using the OLS approach and summarize our findings on the relationship between human capital and innovation level. Fourthly, we divide China into 3 regions and perform regressions with the basic model for the three regions respectively. We then compare the regression results and summarize regional differences. Fifthly, we add interaction terms to our model and

perform regression with the model with interaction terms. We perform regression for all provinces and for three regions respectively, and then we compare all the regression results to summarize the new findings on the factors that influence the effect of human capital.

Our research includes the following main findings. Firstly, human capital does have a strong positive relationship with innovation level. Secondly, the effect of human capital varies across regions as the economic development level, industrial structure, and marketization level differs around the country: these factors decide whether talented workers can work in innovation-oriented environments and make greater contribution to innovation.

After consulting former studies on the topic of human capital and innovation, we believe that our research has the following contributions and novelties:

Firstly, we selected a medium scope of view and based our research on provincial panel data, while former studies were mainly based on a macroscopic view or a microscopic view and used data on the national level or focused on enterprises.

Secondly, we divided China into regions with specific economic environments and considered the relationship between human capital and innovation level for each region. We further utilized the regional heterogeneity to analyze the proper policies for each region and made our suggestions more pointed.

Thirdly, instead of taking human capital as a factor with fixed effects, we added special interaction terms to the model to figure out how we can improve the efficiency of human capital.

Finally, we included various controlled variables in our model to partly exclude influences of other variables and tell the relationship between human capital and innovation more accurately.

Overall, our study brings a new perspective which helps people consider the relationship between human capital and innovation in a mode detailed way and shed light on policy-making. By analyzing the relationship among human capital, innovation level and other possible explanatory variables, researchers can figure out the optimal sequence of development and help the factors reach high efficiency.

REVIEW OF LITERATURE

According to economist Claudia Goldin, human capital is "is the stock of habits, knowledge, social and personality attributes (including creativity) embodied in the ability to perform labor so as to produce economic value". Early

economists like Adam Smith and Alfred Marshall had developed their own view of human capital and confirmed the value of knowledge and skills in production. However, the modern concept of human capital was proposed in the early 1900s and popularized by economists including (Becker, 1962; Mincer, 1958; Schultz, 1961). In his famous speech in 1961, Investment in Human Capital, Theodore Schultz clearly explained the concept of human capital and its important properties. Furthermore, he discussed the relationship between human capital and economic development. His research on human capital theory shed light on later investments in human capital and helped him win the 1979 Nobel Prize in Economic Sciences. Mincer (1958) displayed the relationship between investment in human capital and personal income distribution in the 1958 article Investment in Human Capital and Personal Income Distribution. Becker (1962), the winner of the Nobel Prize in Economic Sciences of 1992, systematically investigated human capital on a large scale and made great contribution to the development of modern human capital theory, and his 1964 book Human Capital marked a milestone event in the revolution of human capital investment. After their work, many economists such as Romer (1989) proposed growth models that include human capital as an important factor.

Although there has been extensive study on the impact of human capital on economic growth, it should be noticed that human capital exerts its influence in a variety of ways. Some researchers regarded innovation as an intermediate stage in the process that human capital facilitates economic growth; for example, Acemoglu et al. (2012) that human capital can influence economic development directly or indirectly, in particular through the generation of technology. Similarly, Cinnirella and Streb (2017) claimed in their 2017 article that human capital enhances labor productivity indirectly through the channel of technological change. However, studies on the precise relationship between human capital and innovation is relatively scarce. Among former researches, Dakhli and De Clercq (2004) took an international view and pointed out that a positive correlation exists between human capital and innovation exists at the country level based on a multi-country study in 2004. Another multi-country study done by Teles and Joiozo (2011) used pooled co-integration tests analyzed pooled data of 27 countries from 1960 to 2000 to support the idea that education hierarchy and innovation are positively correlated; they further proposed the theory that tertiary education is the only level of education that is very significant when determining innovation. In 2019, Diebolt and Hippe (2019) took historical human capital levels since the 19th century

into consideration and revealed the long-run positive impact of human capital on innovation in Europe at a regional level. Researches focusing on the human capital situation in China included the study done by Mincer (1958). They based their research on spatial regression model and supported the idea that human capital has a significant positive correlation with innovation on a provincial level. However, their study cast doubt on the idea that human capital or innovation boosts economic development of China. The research of J. Li, Fu, and Wei (2016) specifically focused on the impact of R&D personnel among human capital and confirmed the positive relationship between the proportion of researchers and innovation level in enterprises. Also in 2014, Yang (2014) created a division of human capital and argued that human capital will increase the level of innovation and industrial upgrading overall. He further concluded that there is a negative relationship between human capital geographic Gini Coefficient and innovation level, which supports the importance of balanced development of human capital. The 2016 research of P. Li, Zhao, and Wan (2014) is one of the few papers in the field of human capital that include interaction terms in linear regression models. They discussed the influence of FDI level on the relationship between human capital and innovation and provided evidence that foreign investment will promote the ability of human capital to boost innovation.

From former papers, we conclude that existing theories mainly supported the positive relationship between human capital and innovation. Considering that former researches mainly focused on theoretical deduction instead of numerical analysis, to test whether this relationship is supported by empirical data on a provincial scale, we now propose our main hypothesis.

H1: There is a strong positive relationship between human capital and innovation level.

However, former studies considered this relationship as an intermediate step between the growth of human capital and that of economy. Therefore, the positive relationship between human capital is generally acknowledged but rarely investigated. To form a comprehensive understanding of the relevant factors that influences the precise relationship between human capital and innovation, we take other factors into consideration. According to Schultz (1961), human capital belongs to the means pf production; thus, we can expect a free market to benefit the flow of human capital and optimize the distribution of human capital. The research of Romer (1989) has pointed out that the development of non-state owned economy and that of the factor market have a significant positive effect on the output of hu-

man capital in an analysis with provincial panel data. In addition to that marketization will increase the income elasticity of rural inhabitants and therefore increase their income. Furthermore, the human capital has a significant promotion effect to industrial upgrading, while a higher marketization level will strengthen that effect. In accordance to these beneficial effects of marketization, we estimate that a higher marketization level will bolster the effect of human capital on innovation.

H2: A higher marketization level will increase the positive effect of human capital on innovation.

It is universally agreed that high-tech industries makes great contribution to innovation. Existing studies indicate that human capital and high-tech industries support each other: for example, the 2016 research of Pei demonstrated the interaction between talent aggregation and high-tech industries and provided evidence that the development of high-tech industries will create demand for technology talents, which will in turn bring momentum for further development. The study of Chen and Yang in 2014 also highlighted the fact that human capital plays a key role in industrial upgrading. At the same time, high-tech industries provide job opportunities for knowledge-oriented personnel and offers them a proper working environment to maximize their output and apply their professional skills in production. However, there is relatively few research that displays the positive relationship between high-tech industries and the output of human capital. Therefore, we expect that in a region where high-tech industry is well-developed, the knowledge and skills of labor (which is a key part of human capital) will have a greater positive influence on production, and we will use collected data to test our third hypothesis.

H3: A higher proportion of high-tech industry in the economy will increase the positive effect of human capital on innovation.

Researchers have long noticed the regional variation in the level of human capital. Apart from researches on the country level, many researches have discussed the within-country differences in value generated by human capital. For example, López-Bazo and Moreno (2007) and Moreno suggested that substantial spatial variations in returns to human capital existed among Spanish regions. Similarly, regional heterogeneity existed in the positive relationship between human capital accumulation and economic growth; they further proposed that the positive relationship was strongest in the West Region of China, and therefore encouraged the cross-region transfer of human capital. Following their experience, we conjecture that regional difference also exists on the relationship between human capital

and innovation. At the same time, according to the theory of marginal benefit, we suppose that the marginal benefit of human capital is higher in relatively underdeveloped regions, such as the Middle and West regions.

H4: The marginal benefit of human capital is higher in the Middle and West regions.

RESEARCH METHODS
Data
In our research, we mainly used data from the National Bureau of Statistics of China. Our data of GDP, population, the number of patent applications and approvals, the number of industrial firms above scale, the general financial revenue of the government, the gross value of import and export goods, the number of firms and the number of state-owned or collectively-owned firms for each province are collected from the National Bureau of Statistics of China and the China Statistical Yearbook. The data of labor education levels are collected from China Labor Statistical Yearbook. Moreover, we used data from economists Gang Fan and Xiaolu Wang when we used their market index to measure the level of marketization for each province.

Definitions of Variables
Patent is a form of intellectual property that gives its owner the legal right to exclude others from making, using, or selling an invention for a limited period of years in exchange for publishing an enabling public disclosure of the invention. Typically, when a firm or an individual makes innovative progress, they will apply for a patent to protect their intellectual property. Thus, the number of patents can closely resemble the innovation level in a region. Although this index is considered to be controversial as some researchers pointed out that patents can only represent innovation levels in certain fields, it is still the most commonly used standard. To exclude influence of population on the number of patents, we will choose the number of patent approvals per 10 thousand people to be our indicator for innovation level. This variable is denoted as PAT in our model, and it is the explained variable in our model.

In our research, we use the education level of labor as an indicator for human capital. The education level of labor is highly accessible, and this index has been adopted by many researchers as a measure for human capital. The major oppositions to this measure focus on the fact that education only represents the cognitive skills of labor, while specific skills in each industry also contributes to human capital. However, this indicator is still widely used in recent studies, and we will use the proportion of labor who received ter-

tiary education (any kind of education beyond high school) as our measure for education level. This variable is denoted as HEDU in our model, and it is considered the key explanatory variable in our model. Another widely used index is the average years of education; however, since the former study of Teles and Joiozo (2011) and concluded that tertiary education is the only level of education that has a significant positive influence on innovation, we choose to use HEDU as the explanatory variable.

Many other factors are also considered influential to the innovation level of a region. Therefore, we will include these as control variables in our model to better tell the relationship between human capital and innovation level.

GDP: The Gross Domestic Product measures the monetary value of final goods and services that are bought by the final user and are produced in a given country or region. GDP is the most commonly used indicator for economic development, and it is widely considered that a higher economic level will boost the innovation level in a region. Better-developed regions will provide more job opportunities in the secondary and the tertiary industries instead of the primary industry. Similarly, to exclude the influence of population, we will use the GDP per capita to be our indicator for economic development, and this variable is denoted as GDPN in our model. The unit of GDPN is 10 thousand Yuan.

Industrial Structure: The industrial structure of a region describes the composition of a country's economic activity. Industries are usually categorized into three basic types according to their stage within the production process. It is widely acknowledged that the proportion of tertiary industry output value in GDP can reflect the industrial structure of a region: the higher the proportion is, the more advanced the industrial structure is. We use the proportion of tertiary industry output value in GDP as an indicator of the industrial structure, and this variable is denoted as IND in our model.

Industrial Firms: Many technological breakthroughs happened in large industrial firms. The need for high efficiency in mass production encourages people to seek changes to their production process and make improvements. We will use the number of above-scale industrial firms (whose annual operating revenue exceeds 20 million Yuan) in each province as a control variable. Since this variable has large values, we will calculate the logarithm of this variable to fit it to our linear model better. This variable is denoted as LFIRM in our model.

Government Revenue: Government revenue is money received by a government from taxes and non-tax sources to enable it to undertake government expenditures. Although other sources exist, the major part of government revenue still come from taxes. A lower tax value is generally expected to encourage enterprises to take risks and make innovative breakthroughs. We will calculate the ratio of government revenue to GDP in each province as an indicator of the tax level of a region, and if the ratio is larger, the tax level will be considered higher. This ratio is denoted as TAX in our model.

Import and Export Value: The import and export value of a region can resemble its level of openness. We calculated the sum of value of imported and exported goods and services for each province, and we calculated the ratio of that sum to the provincial GDP. This ratio is chosen as an indicator for regional openness in our model. A higher degree of openness is generally expected to facilitate the exchange of ideas and boost innovation. This ratio is denoted as EXP in our model.

Apart from these variables, we also included alternative indicators of labor education level and innovation level check the robustness of our results.

Similar to PAT, the number of patent applications accepted per 10 thousand people is an alternative indicator of regional innovation level, and it is widely used in former studies. We will use PATN to denote it in our model.

As mentioned before, education level only partly reflect the knowledge and skills of labor. Vocational education or experience in specific industries also contribute to human capital. At the same time, the well-being of labor will influence human capital, so factors such as labor health and social welfare are related to human capital. Given these conditions, there may be other indicators that can reflect human capital more comprehensively. The Jorgenson-Fraumeni Method is extensively used to measure human capital internationally. This method is based on the principle that the value generated by human capital can be represented by the expected future earnings of labor. In our research, we collected data of human capital calculated by the J-F Method from researches done by China Center for Human Capital and Labor Market Research. This is an alternative indicator of human capital, which will be used later in our studies. This variable is denoted as JF in our model.

To find more specific details about the relationship between human capital and innovation, we created a few interaction terms.

Firstly, we take marketization into consideration. Marketization is a restructuring process that enables state enterprises to operate as market-oriented firms by changing the legal environment in which they operate. Marketization index is a concept introduced by economists Fan Gang and

Wang Xiaolu which reflects the level of marketization in one region in a comprehensive way. It takes into consideration the relationship between market and government, the development of non-state owned economy, the development of product markets and factor markets, and the market environment. We collected data of marketization index from 2000 to 2016 from Fan et al, and this variable is denoted as MAR in our model. In supplement of that, we introduces another variable NAT. NAT is the proportion of state-owned and collectively-owned enterprises in a province. This can be used as an alternative variable that represents the marketization level. We collected data of the total number of enterprises and the number of state-owned and collectively-owned enterprises from the National Bureau of Statistics of China.

At last, to find out how the effect of human capital on innovation is influenced by the industrial structure, we introduced a new variable HTECH. This variable denotes the proportion of output value of high-tech industries in Gross Domestic Production. High-tech industry refers to the sectors characterized by innovative or complex technology and knowledge-intensive labor force. Compared to IND, this variable focuses specifically on high-tech industries instead of the whole tertiary industry.

Summary Statistics

As the number of industrial firms is a large positive number, we have taken the logarithm of it to avoid extreme values and linearize the data. Table 1 below is the summary statistics of all variables that are going to be used in our models:

TABLE 1. Summary statistics of provincial data

Variable	Obs	Mean	Std Dev.	Min	Max
PAT	527	4.107551	7.091746	0.026119	46.28532
HEDU	496	11.73838	8.714051	0.2	55.87197
GDPN	527	2.878791	2.270101	0.2759	11.8198
IND	510	39.33765	10.04651	1.6	80.2
LFIRM	527	8.520512	1.342991	4.025352	11.08973
TAX	527	0.09014	0.031677	0.043906	0.22734
EXP	527	0.410758	0.477368	0.020233	2.256493
PATN	527	7.614362	12.66769	0.05597	87.0359
JF	527	109.4507	53.36798	38.0283	358.6689
MAR	527	5.851027	2.010834	-0.3	10.92
NAT	186	0.075973	0.045057	0.021851	0.347053
HTECH	372	0.1022029	0.119445	0.0024386	0.5061789

The variables in Table 1 has been explained in 3.2 Definition of Variables.

Empirical Model

We will use least-squares approximation with the following two linear regression models.

Basic model

$$Y_{it} = \alpha_0 + \alpha_1 \text{HEDU}_{it} + \alpha_2 X_{it} + \delta_1 + \gamma_t + \varepsilon_{it}$$

Model with interaction terms

$$Y_{it} = \alpha_0 + \alpha_1 \text{HEDU}_{it} + \alpha_2 X_{it} + \alpha_3 \text{INTER}_{it} + \alpha_4 \text{HEDU\_NTER}_{it} + \delta_i + \gamma_t + \varepsilon_{it}$$

In the models above, Y is the dependent variable (which is PAT in most our regression models), and HEDU, the indicator of human capital, is the explanatory variable. X represents the controlled variables, including GDPN, IND, LFIRM, TAX and EXP. INTER represents the interaction term in our model, which can be MARS, NAT and TIME. Finally, HEDU_INTER is the product of HEDU and the interaction term. These variables have all been explained in Chapter 3.2. The subscript i represents provinces; the subscript t represents years; δ_i stands for provincial fixed effects; γ_t refers to time fixed effects; and ε_{it} a stands for random disturbance.

EMPIRICAL ANALYSIS
Baseline Results and Analysis

To start with, we examined the correlations of HEDU with PAT using the basic linear regression model. Then, we added the controlled variables one at a time to our regression model. As shown in Table 1 below, the first column demonstrates the regression results with only the explanatory variable HEDU and no controlled variables. In the second column, the controlled variable GDPN is added to the model; in the third column, IND is added to the model; in the fourth column, LFIRM is added to the model; in the fifth column, TAX is added to the model; and in the sixth column EXP is added to the model.

TABLE 2. Baseline regression results with basic model

	(1) PAT **PAT**	(2) **PAT**	(3) **PAT**	(4) **PAT**	(5) **PAT**	(6) **PAT**
HEDU	0.9136***	0.3588***	0.3276***	0.2859***	0.3010***	0.2996***
	(0.0645)	(0.0704)	(0.0794)	(0.0829)	(0.0834)	(0.0833)
GDPN		3.4207***	3.3013***	3.3647***	3.3771***	3.4618***
		(0.2685)	(0.2834)	(0.2852)	(0.2850)	(0.2904)
IND			0.0950**	0.0809**	0.0836**	0.0729*
			(0.0372)	(0.0381)	(0.0381)	(0.0387)
LFIRM				-1.2015*	-1.2876*	-1.5113**
				(0.7066)	(0.7080)	(0.7234)
TAX					-23.2050	-24.1573
					(15.6895)	(15.6820)
EXP						2.0827
						(1.4209)
Constant	-1.6345	-11.5804***	-1.6176	9.1230	11.2435*	13.2255**
	(1.0639)	(1.3781)	(1.0111)	(6.3965)	(6.5466)	(6.6763)
Observations	496	496	480	480	480	480
R^2	0.792	0.847	0.849	0.850	0.851	0.852

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

After multiple regression analyses, our hypothesis 1 is strongly supported. The regression results indicate that the coefficients of HEDU are positive in all 6 regression models, and they are significantly positive at the 1% level. In regression model 6, the model with all controlled variables, we discover that the coefficient for HEDU is 0.2996, nearly 0.3. This means that a 1% increase in the proportion of labor that receives tertiary education will result in a 0.3 increase in the amount of annual patents approved per 10,000 people. This matches the hypothesis that human capital level is positively correlated with innovation. Moreover, some interesting results are revealed by the coefficients. We find strong evidence for a firm positive relationship between average GDP and innovation level, because the coefficients of variable GDPN is significantly positive in every regression model that includes it as an explanatory variable.

Moreover, the coefficient of IND is significantly positive on the 5% significance level in 3 models, and in the sixth model it is positive on the 10% significance level. Since IND rep-resents the proportion of tertiary industry output value in GDP, a higher IND value implies a more advanced industrial structure. So, our regression result corresponds to our expectation that innovation level will be higher in a region with more developed industrial structure. The coefficient of LFIRM is negative, and it is significant on the 10% level. This is partially out of our expectation. One possible explanation we find for this phenomenon is that more innovation happens, overall, in non-industrial enterprises or enterprises with relatively smaller scales. In the most developed regions of China, the tertiary industry contributes more than the secondary industry to GDP. In tertiary industry, firms are usually of smaller scales, but their innovation abilities can be significantly stronger than traditional industrial enterprises, where a system of mature production process has been formed in practice for a long time and the space left for innovation is scarce. The coefficients of the other two explanatory variables are not significant at the 10% level, and to figure out their effects on innovation more detailed

investigation is needed.

Results with the Basic Model on Regional Heterogeneity
To tell how the beneficial effect of human capital varies among regions with different geological positions, social background and degree of economic development, we created a partition of 31 Chinese mainland provinces with respect to their location. With the reference to standards set by the National Bureau of Statistics of China, 13 Provinces are divided into the East region: Beijing, Fujian, Guangdong, Hainan, Hebei, Heilongjiang, Jilin, Jiangsu, Liaoning, Shan-

dong, Shanghai, Tianjin, and Zhejiang. 6 Provinces are divided into the Middle region: Anhui, Henan, Hubei, Hunan, Jiangxi, and Shanxi. At last, 12 Provinces are divided into the West region: Gansu, Guangxi, Guizhou, Inner Mongolia, Ningxia, Qinghai, Shaanxi, Sichuan, Tibet, Xinjiang, Yunnan, and Chongqing. Table 3 shows the regression results after we apply the model to three regions respectively. Generally speaking, the East region is the most economically developed region in China, while the West region is the most underdeveloped; the Middle region lies in between.

TABLE 3. Regression results on regional heterogeneity with the basic model

	(1)	(2)	(3)	(4)	(5)	(6)
	East		Middle		West	
HEDU	0.3714***	0.3024**	-0.1339	-0.0325	0.1443**	0.3319***
	(0.1243)	(0.1461)	(0.0831)	(0.0731)	(0.0599)	(0.0639)
GDPN	3.9326***	4.7394***	1.0503**	0.2079	0.5801***	-0.0971
	(0.5576)	(0.6033)	(0.4959)	(0.3789)	(0.2170)	(0.2210)
IND		0.0287		-0.0357		0.0485
		(0.0702)		(0.0341)		(0.0327)
LFIRM		-2.2615		3.7814***		1.7333***
		(1.6157)		(0.5109)		(0.5398)
TAX		-75.8449**		9.9357		-0.7016
		(31.6406)		(10.4997)		(11.0366)
EXP		5.9966**		-9.8968***		10.8041***
		(2.4919)		(3.0254)		(1.3321)
Constant	-16.3715***	0.2733	5.1758**	-29.6061***	-1.3968	-14.0874***
	(3.3493)	(13.0343)	(2.3329)	(4.6644)	(1.4539)	(3.2232)
Observations	208	208	96	96	192	176
R^2	0.858	0.867	0.849	0.930	0.729	0.837

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Results from multiple regression analyses suggest us that the way our explanatory variables correlates to innovation level varies substantially. In the models with all controlled variables, the coefficients of our key variable HEDU is significantly positive at 5% level in the East region, and in the West region it is even more significant, reaching the level of 1%. In the models with only HEDU and GDPN as variables, the coefficients still stay positive at the significance level of 1% and 5% respectively. These results provide convincing evidence that our Hypothesis 1 stays true for the East and West region. However, the results for the Middle region appears interesting, as the coefficients of HEDU are not significant in the Middle region.
To explain this phenomenon, we checked our raw data and looked for clues among the other coefficients in the regression model. Together, they suggest some reasons for our

hypothesis to fail in the Middle region. Some former researches have pointed out that human capital cannot directly exert its influence on economic development; instead, there is an intermediate stage where human capital first facilitates management level, innovation and knowledge spillovers. Then, these factors will boost economic development. Similarly, we consider the possibility that a middle stage exists between the increase of human capital and the increase of innovation. The definition of variable HEDU was given in Chapter 3.2, and it focused on the labor who received tertiary education. However, former researches on human capital theory indicates that cognitive skills only comprise part of all knowledge and capabilities owned by labor; their working experience, relevant vocational training and even physical well-being are also factors that influence human capital. On the other hand, innovation

activity focuses on the application of knowledge. In certain industries, such as the Internet industry, innovation takes place at a much higher pace than it does in other industries. Thus, we deduce on this basis that the industrial structure of the Middle region causes the coefficient of HEDU to be not significant. Over the 17 years under investigation, the average PAT value of Shanxi Province is about 1.14, ranking the last in all 6 Middle provinces (any other province has a value of at least 1.47). However, its average value of HEDU is 11.40, ranking the first in all 6 Middle provinces (any other province has a value less than 10). Such an outlier exerts a powerful influence on our linear regression model, especially under the condition that the Middle region has only 6 provinces and we are applying our model to a small sample. Shanxi Province is well-known for its coal industry, which partially explains its low innovation level—coal industry has a long history and does not leave much space for innovation. Simultaneously, we notice that the coefficient of LFIRM, despite being negative in the overall model, is significantly positive at 1% level in the Middle region. This strong contrast indicates that innovation in the Middle region and East region may be driven by very different forces: in the Middle region, most innovation happens in large industrial firms and in the traditional industry, where practical working experience overweighs theoretical knowledge learnt in universities in making technical breakthroughs. In the East region, emerging industries such as the environmental industry, information technology industry, bio-industry and new-energy industry contribute an important part to innovation. These industries are closely connected to tertiary education and cognitive skills, which explains the positive coefficient of HEDU in the East region. For the Western region, the positive effect of human capital on innovation is especially obvious, as a key controlled variable GDPN does not have a significant positive coefficient here. In provinces such as Tibet and Qinghai, the value of HEDU is too low to allow the existence of innovative enterprises, and their number of patents remains at a low level among the whole 17 years. To summarize, our hypothesis 4 is partly supported by regression data here.

In addition to analysis of our key variable HEDU, some coefficients above provides support for our hypothesis. For example, the coefficients of LFIRM in the Middle and West regions are significantly positive at 1% and 5% level respectively. This fits our intuition better than the results in the baseline model, since it is hard to imagine how an increase in the number of large firms will harm innovation. The truth is that large industrial firms still do great contribution to innovation levels in the Middle and West regions.

However, since the East region has a great lead in innovation level and large industrial firms contribute a relatively smaller part to innovation there, the linear model underestimates the contribution of above-scale industrial firms. The coefficients of TAX further validates our idea. In the Eastern region, much innovation comes from small enterprises who are much more sensitive to tax levels and risks than large industrial firms. Therefore, the decrease in tax level has a significant positive effect (at 5% level) on innovation in the Eastern region, while in the other two regions it has no significant influence on innovation level.

Overall, the regional regression results indicate that the relationship between human capital and innovation depends greatly on the industrial structure of a region. Two things relies on each other: A basic level of human capital is needed to allow innovation-driven enterprises to exist, while an advanced industrial structure is needed to let human capital fully achieve its potential.

FURTHER STUDY
Model with Interaction Terms and Analysis

In Part 4, we discussed the general relationship between human capital and innovation. From the regression results, we noticed that the key explanatory variable, human capital, has different effects on innovation in different regions. To explore the factors that influence the effect of human capital on innovation, we introduced interaction terms to our model and created a new model, which can be seen in Chapter 4.1.

In traditional economic opinions, an open market will enable production factors to flow freely and achieve a high production efficiency. Human capital, as an important factor in production, is no exception. In an open market where competition is fierce, the subjects of market are encouraged by potential rewards to make more innovation attempts. In Chapter 3.2, we have introduced the concept of marketization index as well as the variable MAR. Before adding MAR to our model as an interaction term, we first need to notice a possible multi-collinearity that exists between MAR and our key explanatory variable, HEDU. If the two variables are highly correlated, then a product of them in a linear regression model can greatly harm the effectiveness of our model. First, we make a test of the collinearity of MAR and HEDU. We first construct a linear regression model of MAR and HEDU . The linear regression results proves that the two variables are strongly positively correlated. To address this problem, we introduce a new variable MAR_R. With this model, we define that , and MAR_R represents the residual of each MAR value. This variable can better reflect how the

marketization index of a region compares to the expected value. If a region has a negative MAR_R value, we consider its marketization level to be relatively low, while if the value is positive we consider its marketization level to be rela-

tively high. Then, we define our interaction term . Table 4 below shows the regression results after we add MAR_R and HEDU_INTER1 to our model.

TABLE 4. Regression results of model with MAR_R*HEDU as interaction term

	(1) All	(2) East	(3) Middle	(4) West
HEDU	0.5245***	0.5798***	0.0029	0.4363***
	(0.0846)	(0.1620)	(0.0867)	(0.0540)
HEDU_INTER1	0.0975***	0.0796***	0.1443***	0.0621***
	(0.0140)	(0.0247)	(0.0314)	(0.0118)
GDPN	2.5124***	3.8442***	-1.1139**	-0.2604
	(0.2993)	(0.6340)	(0.4555)	(0.1828)
IND	0.0711*	0.0192	-0.0048	0.0090
	(0.0363)	(0.0683)	(0.0313)	(0.0272)
LFIRM	-1.8695***	-3.4887**	2.7300***	0.9682**
	(0.6809)	(1.5980)	(0.5224)	(0.4533)
TAX	-3.0150	-73.1041**	6.4360	24.3496**
	(14.9396)	(31.1488)	(10.2431)	(9.6946)
EXP	1.2130	5.4865**	-15.2547***	5.6517***
	(1.3479)	(2.4488)	(2.9219)	(1.3033)
MAR_R	-0.1701	0.6751	-1.0959**	0.0430
	(0.3724)	(0.7296)	(0.4554)	(0.2716)
Constant	14.1684**	23.2675	-20.0789***	-8.6870***
	(6.2609)	(17.3235)	(4.6557)	(2.7409)
Observations	480	208	96	176
R^2	0.871	0.877	0.947	0.892

Standard errors in parentheses* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The regression results validate our model with interaction terms. They reveal that the coefficient of HEDU_INTER1 is positive in all four models, including one overall model and three respective models for 3 regions. This clearly indicates that a higher market index will strengthen the positive effect of human capital on innovation. To be specific, if we use the regression model for all provinces, we can expect an 1 unit increase in market index to bring about a 0.0975 increase in the coefficient of HEDU, or an increase of (0.0975/0.0846) = 1.15 Standard Deviations, which is a remarkable increase in the efficiency of human capital. In the East region, a 1 unit increase in market index can lead to an increase of (0.0796/0.1620) = 0.49 Standard Deviations, while this number is 1.66 in the Middle Region and 1.15 in the West region. This difference suggests us that market index has the greatest influence on the effect of human capital in the Middle region, and the West region ranks the second. Despite the differences, this strengthening effect of market

index seems to be significant across regions of different levels of economic development. This suggests that increasing the marketization level can be a generally applicable way for the government to improve the effect of human capital and promote the innovation level, and our Hypothesis 2 is strongly supported. Apart from this new discovery, we find the coefficients of HEDU to stay significantly positive at 1% in all three models except the one of the Middle Region, which corresponds to our results in Chapter 4. The coefficients and significance levels of other controlled variables were also similar to that in the baseline model, which provides evidence that our new model still fits the overall situation.

Next, to further examine our hypothesis that industrial structure is one main deciding factor of the effect of human capital on innovation, we add HTECH into consideration. To examine our former assumptions that high-tech industry contributes more to the economy in the East region than

it does in the Middle and West regions, we summarized the data of HTECH for every region, and the results are shown below. The mean value of HTECH is significantly higher in the East region (mean = 0.18) than it is in the Middle (mean = 0.05) and West (mean = 0.04) regions, which supports our former assumptions.

Similarly, according to our hypothesis before, HEDU and HTECH facilitates each other. We expect to see a signifi-

cant positive relationship between HEDU and HTECH. The regression results sustains this relationship:

Therefore, to avoid multi-collinearity, we define a HTECHR to be the residual of HTECH in the regression model of HTECH and HEDU: . Then, we define that Table 5 shows the regression results after we add HTECHR and HEDU_INTER2 to the baseline model.

TABLE 5. Regression results with HTECHR*HEDU as interaction term

	(1)	(2)	(3)	(4)
	All	East	Middle	West
HEDU	0.0796	0.0834	0.0537	0.1675***
	(0.0814)	(0.1432)	(0.1069)	(0.0471)
HEDU_INTER2	-0.1261	-0.2598	3.6127***	0.9301***
	(0.1627)	(0.2269)	(0.9692)	(0.2526)
GDPN	2.7755***	3.9968***	1.3505*	0.0258
	(0.3003)	(0.6288)	(0.6996)	(0.1337)
IND	0.0659**	0.0482	-0.0019	-0.0303**
	(0.0258)	(0.0459)	(0.0211)	(0.0144)
LFIRM	-1.0107	-2.0545	1.7783***	0.2274
	(0.6208)	(1.4147)	(0.3800)	(0.2371)
TAX	-5.4321	-46.3975	7.7907	27.8413***
	(14.5711)	(33.0449)	(7.7596)	(5.6909)
EXP	2.0970	2.0310	-8.0069**	-0.6391
	(1.5520)	(2.5821)	(3.1082)	(1.0359)
HTECHR	9.9552*	20.1377**	-36.2313**	2.1345
	(5.8229)	(9.0328)	(13.5444)	(3.9448)
Constant	7.1710	6.9536	-15.3028***	-2.4602
	(5.5363)	(10.5762)	(3.4800)	(2.1894)
Observations	330	143	66	121
R^2	0.834	0.853	0.881	0.862

Standard errors in parentheses* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

From the regression results, we failed to find a strong positive relationship between HTECHR and the effect of human capital on innovation in the Total model and in the model of the East region. This is because the coefficients of HEDU_INTER2 are not significantly positive in the Total and East models. In order to figure out the underlying causes of this unusual phenomenon, we looked into the raw data and drew a scatterplot of the HTECH and HEDU data. Graph 1 shows their relationship. From graph 1 we can see that the relationship between HTECH and HEDU is roughly linear when HEDU is below 20. However, a few outliers on the right of the graph severely influences the accuracy of the linear model. There are a few data points with HEDU level greater than 30, but their HTECH level are only

around 2, which is far less than the expected value. We found these outlier in the graph and soon discovered that these points all represented data of Beijing. This rationalizes the ourliers: Beijing is the capital and educational center of China. Across China, the density of universities is the highest in Beijing, which gives Beijing a leading HEDU level in the whole 17 years from 2000 to 2016. In addition to that, Beijing owns a great number of scientific institutions, including the Chinese Academy of Sciences, one of the leading scientifi research forces in China. Furthermore, many large enterprises in the Internet industry as well as other industries choose Beijing as the location of their headquarters in order to enjoy more resources and attract talented workers. These factors combines to give Beijing a strong po-

sition in the PAT level: from 2000 to 2016, Beijing ranks top 5 in PAT level for every year, and has the leading position for 7 years. However, Beijing does not rely heavily on manufacturing as a capital city, and much of its GDP comes from the service industry. Thus, Beijing has a relatively low HTECH levels compared to other cities with high HEDU levels. As

a result, the value of HEDU_INTER2 for Beijing are negative numbers with large absolute values. These values affect the coefficients of HEDU_INTER2 in the regression model. To deal with this situation, we temporarily excluded Beijing from regression analysis and acquired new regression results, which are shown in Table 6 below.
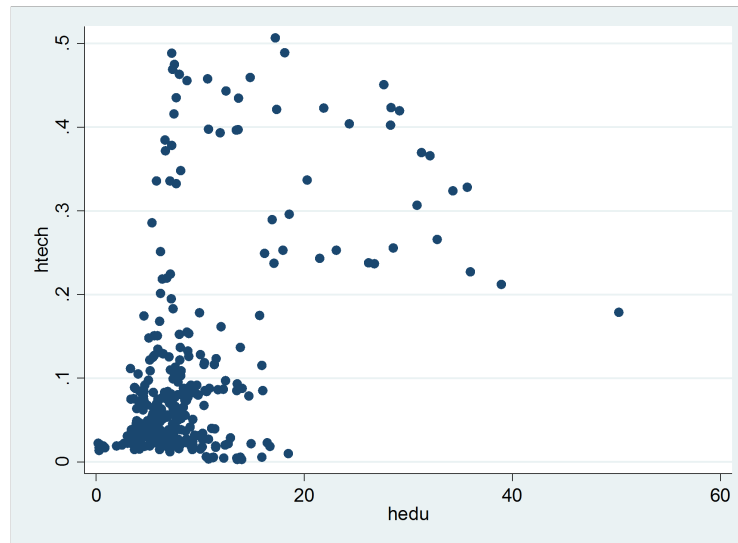

FIGURE 1. Relationship between HEDU and HTECH (Scatterplot)

TABLE 6. Regression results with HTECHR*HEDU as interaction term (Beijing excluded)

	(1) Total Without BJ	**(2)** East Without BJ	**(3)** Middle	**(4)** West
HEDU	0.0035	-0.1267	0.0537	0.1675***
	(0.0890)	(0.1946)	(0.1069)	(0.0471)
HEDU_INTER2	1.5974***	1.2042**	3.6127***	0.9301***
	(0.3558)	(0.5837)	(0.9692)	(0.2526)
GDPN	2.8335***	4.2575***	1.3505*	0.0258
	(0.2896)	(0.6460)	(0.6996)	(0.1337)
IND	0.0855***	0.0982*	-0.0019	-0.0303**
	(0.0292)	(0.0589)	(0.0211)	(0.0144)
LFIRM	-0.6448	-1.6457	1.7783***	0.2274
	(0.5987)	(1.4508)	(0.3800)	(0.2371)
TAX	-2.7730	-32.5692	7.7907	27.8413***
	(14.0764)	(34.2266)	(7.7596)	(5.6909)
EXP	0.3938	1.0118	-8.0069**	-0.6391
	(1.5206)	(2.8342)	(3.1082)	(1.0359)
HTECHR	-9.9433	0.4626	-36.2313**	2.1345
	(7.0030)	(12.6418)	(13.5444)	(3.9448)
Constant	-1.1552	8.6959	-15.3028***	-2.4602
	(5.6703)	(16.3329)	(3.4800)	(2.1894)
Observations	319	132	66	121
R^2	0.827	0.845	0.881	0.862

Standard errors in parentheses* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The new regression results in the total model and East model fits our expectations better. The coefficients of HEDU_INTER2 are significantly positive at 1% level in the Total model, the Middle model and the West model, and it is significantly positive at 5% level in the East model. After excluding abnormal values from the model, we get a linear regression result that fits our intuition better. This model points out that the proportion of output value of high-tech industries in GDP will significantly affect how human capital facilitates innovation across regions. In regions where high-tech industries contribute more to GDP, human capital will better boost innovation. From the numerical perspective, in the total model, a 1% increase in HTECH will lead to a 0.016 increase in HEDU, which is 0.18 times the Standard Deviation. This number is 0.06 in the East Region, 0.34 in the Middle Region and 0.20 in the West region. The marginal benefit of HTECH is higher in relatively underdeveloped regions, which accords to traditional economic views: in the East region, high-tech industries are better developed, so the marginal benefit of HTECHR on the effect of human capital will be relatively lower.

The new regression results in the total model and East model fits our expectations better. The coefficients of HEDU_INTER2 are significantly positive at 1% level in the Total model, the Middle model and the West model, and it is significantly positive at 5% level in the East model. After

excluding abnormal values from the model, we get a linear regression result that fits our intuition better. This model points out that the proportion of output value of high-tech industries in GDP will significantly affect how human capital facilitates innovation across regions. In regions where high-tech industries contribute more to GDP, human capital will better boost innovation. From the numerical perspective, in the total model, a 1% increase in HTECH will lead to a 0.016 increase in HEDU, which is 0.18 times the Standard Deviation. This number is 0.06 in the East Region, 0.34 in the Middle Region and 0.20 in the West region. The marginal benefit of HTECH is higher in relatively underdeveloped regions, which accords to traditional economic views: in the East region, high-tech industries are better developed, so the marginal benefit of HTECHR on the effect of human capital will be relatively lower.

Robustness Checks

In order to validate our findings and better understand the positive effect of human capital on innovation, we performed a few robustness checks. Table 7 below shows the results after we change the key explanatory variable from HEDU to JF in the baseline model. Apart from HEDU, JF is another variable that reflects human capital level, as we have introduced in Chapter 3.2. The advantage of JF is that it covers more aspects of human capital, while HEDU only takes the education level of labor into consideration.

TABLE 7. Regression results with baseline model (JF as explanatory variable)

	(1) PAT Total	(2) PAT East	(3) PAT Middle	(4) PAT West
JF	0.1258***	0.1037***	0.0690***	0.1003***
	(0.0127)	(0.0237)	(0.0136)	(0.0160)
GDPN	2.0184***	3.3041***	0.6167*	-0.3225
	(0.2824)	(0.6343)	(0.3234)	(0.2165)
IND	0.0351	0.0229	-0.0311	-0.0066
	(0.0345)	(0.0642)	(0.0264)	(0.0297)
LFIRM	-1.4491**	-0.3007	2.7118***	-0.1248
	(0.5908)	(1.4218)	(0.4411)	(0.4873)
TAX	-28.0358**	-55.4294*	0.7473	11.4022
	(13.5974)	(29.0002)	(8.5233)	(9.7950)
EXP	0.0044	4.0011*	-11.1488***	6.0755***
	(1.2270)	(2.2283)	(2.3070)	(1.3632)
Constant	5.7925	-6.8190	-23.6601***	-3.5930
	(5.7951)	(15.4313)	(4.0147)	(2.9322)
Observations	510	221	102	187
R^2	0.870	0.874	0.947	0.842

Standard errors in parentheses* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In Table 7, the coefficients for JF in all four models are significantly positive on 1% level, including the Middle Region, which was a counterexample in the baseline model. After taking more factors into consideration, the results became more significant: this fact provides convincing support for our main hypothesis that human capital and innovation level are positively correlated.

Next, we changed the explained variable in the model from PAT to PATN. As we have defined in Chapter 3.2, PATN rep-

resents the number of patent applications accepted, and it is another index that reflects the innovation level of a region. The main difference between PAT and PATN is that PAT will focus more on the quality of all patent applications, as it represents the number of applications ratified. PATN focuses more on innovative attempts made by individuals or groups. Table 8 below shows the regression results with the basic model and with PATN as the explained variable.

TABLE 8. Regression results with baseline model (JF as explanatory variable)

	(1) **Total**	**(2)** **East**	**(3)** **Middle**	**(4)** **West**
HEDU	0.5336***	0.4288*	0.0032	0.6737***
	(0.1350)	(0.2212)	(0.1902)	(0.1185)
GDPN	6.3005***	9.2392***	0.2480	-0.4802
	(0.4704)	(0.9137)	(0.9863)	(0.4097)
IND	0.0904	0.0577	-0.0831	0.0401
	(0.0627)	(0.1062)	(0.0887)	(0.0606)
LFIRM	-1.5620	-2.6135	8.8418***	2.7449***
	(1.1718)	(2.4469)	(1.3301)	(1.0006)
TAX	-4.1614	-82.6358*	36.9380	2.3585
	(25.4029)	(47.9178)	(27.3346)	(20.4581)
EXP	0.0228	4.9839	-26.3040***	21.4347***
	(2.3016)	(3.7738)	(7.8762)	(2.4693)
Constant	13.2166	-12.9075	-69.8999***	-23.1224***
	(10.8147)	(19.7396)	(12.1432)	(5.9747)
Observations	480	208	96	176
R^2	0.878	0.902	0.894	0.856

Standard errors in parentheses* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results shown in Table 8 are similar to the results of the Baseline Model. The coefficients of HEDU are significantly positive at 1% level in the Total model and the West model, while in the East model the coefficient is significantly positive at 10% level. In the Middle model, the coefficient of HEDU is again not significant. The similarity between Table 8 and Table 1 provides support for the stability of our model.

In Chapter 5.1 we talked about the level of marketization and took it as an interaction term in our model. At that time, we used MAR as the variable that denotes marketization level, and after making some adjustments, we used MAR_R in our model. To test the stability of our findings above, we used another variable NAT to represent marketization level. As explained in Chapter 3.2, NAT is the proportion of state-owned and collectively-owned enterprises in a province. We introduced NAT to our model, and we introduced , to be an interaction term in the model. The regres-

sion results were shown in Table 9 below.

From the regression results we can see that the coefficient of HEDU_INTER3 is significantly negative in the Total model and the East model. Normally, a region with a higher proportion of state-owned and collectively-owned enterprises will have a relatively low marketization level. Thus, the negative coefficients of HEDU_INTER3 again provides support for our hypothesis that a higher marketization level will have a positive influence on the effect of human capital on innovation. In the Middle and West regions, the coefficients of HEDU_INTER3 are not very significant. This phenomenon can be partly explained by our previous assumptions: in the Middle and West regions, large industrial firms contribute a great part to innovation, and a large proportion of these firms are owned by the state. Under this circumstance, a high proportion of state-owned enterprises will not pose a strong weakening effect on innovation.

TABLE 9. Regression results with NAT*HEDU as interaction term

	(1) **Total**	**(2)** **East**	**(3)** **Middle**	**(4)** **West**
HEDU	0.5721***	0.6382***	-0.0097	0.0846
	(0.1275)	(0.2172)	(0.3020)	(0.1708)
GDPN	1.5516**	1.2471	-1.7975	2.2733**
	(0.7061)	(1.3985)	(1.2867)	(1.0629)
IND	-0.1537	-0.3427	-0.1998	0.1373*
	(0.0979)	(0.2656)	(0.1362)	(0.0811)
LFIRM	-0.5611	-0.8918	4.3512*	2.4869
	(1.5555)	(3.4019)	(2.1557)	(1.7154)
TAX	-4.3407	16.6064	-12.8321	-39.4903*
	(21.6256)	(41.8272)	(28.3879)	(23.4209)
EXP	0.5226	1.7092	-9.4685	4.3553
	(2.7844)	(6.7449)	(11.4192)	(3.3241)
HEDU_INTER3	-7.7687***	-8.1564***	-1.0246	1.2742
	(1.1016)	(1.8341)	(3.7435)	(1.8941)
NAT	93.1681***	72.0293	40.1494	-1.8250
	(27.6816)	(74.2334)	(74.1356)	(26.6035)
Constant	5.3625	45.0221	-27.1753	-27.6047
	(16.3604)	(47.9840)	(22.4746)	(17.7479)
Observations	180	78	36	66
R^2	0.968	0.967	0.960	0.926

Standard errors in parentheses* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Another possible reason is that NAT only represents one aspect of marketization, so there may be some extent of disparity between the model with NAT and the model with MAR_R (the latter provides a holistic view of the problem.) To summarize, our model passed the three robustness checks above and produced reasonable regression results. At the same time, our two hypotheses and a few other findings are further supported in the checks.

MAIN CONCLUSION

In this paper, we discussed the relationship between human capital and innovation level in China on a provincial level. After collecting empirical data from 2000 to 2016, we proposed a linear regression model to analyze the data. Then, to further investigate regional heterogeneity, we did regression respectively for the East, Middle and West regions of China and compared the differences among the results. In order to explore the economic indicators that will influence the relationship between human capital and innovation level, we introduced interaction terms to our regression model and made more findings. Overall, our main conclusions include the following: firstly, there is a strong positive relationship between human capital level and innovation level in China. This finding is also clearly supported by

our results in the East and West regions, and is partially supported by our results in the Middle region. Secondly, there is a strong positive relationship between the marketization level and the positive effect of human capital on innovation in China. If one province has a higher marketization level, then generally the positive effect of human capital on innovation will be stronger there. This relationship is generally stronger in the Middle and West regions than in the East region. Thirdly, there is a strong positive relationship between the proportion of high-tech industry in GDP and the positive effect of human capital on innovation in China. If one province has a higher proportion of high-tech industry in its economy, then generally the positive effect of human capital on innovation will be stronger there. This relationship is generally stronger in the Middle and West regions than in the East region.

Apart from these major findings, we have a few findings with our controlled variables in the model. Firstly, there is a moderate negative relationship between tax level and innovation level in China. This negative relationship is strongly supported by regression data in the East region, but does not seem to be significant in the Middle and West regions. Secondly, there is a strong positive relationship between the number of above-scale industrial firms and innovation level

in the Middle and the West regions, while the relationship seems to be insignificant in the East region.

IMPLICATIONS AND LIMITATIONS

Following the conclusions above, we have a few suggestions for the Chinese government to help them improve the innovation level in the country. Firstly, the government should increase the spending in education and promote tertiary education across the country, since the positive relationship between human capital and innovation is strongly supported. Secondly, the government should consider to relax market restrictions, increase market openness, and encourage private enterprises to enter the economy. These approaches will increase the marketization level and therefore help human capital better facilitate innovation. Thirdly, the government should encourage high-tech industry to grow and offer subsidies for high-tech enterprises in regions where human capital level is high. These approaches will increase the proportion of high-tech industry in the economy and provide more opportunities for workers who have received tertiary education to make innovation. At last, the government should consider to make more government investments on education and high-tech industries in the Middle and West regions, since the marginal benefit of human capital is higher in the relatively underdeveloped areas. For the East region, the government should consider to lower the level of taxes in order to boost innovation.

Meanwhile, although our investigation provides convincing support for the beneficial effect of human capital on innovation, we recognize that our research are still subject to a few limitations. Firstly, as we used tertiary education as a main indicator of human capital, other important factors including vocational skills and the workforce's economic or physical conditions are absent from this model. By considering human capital on a comprehensive view, future research may demonstrate how other components of human capital may contribute to innovation and to what extent will they facilitate this process. Secondly, due to a relatively small sample size and a great number of control variables, our model may receive influence from random disturbance and outliers. To increase the strength of data analysis, further investigations may be conducted no a city level or similarly to increase the sample size or be based upon a more concise, stabilized model. This may also address the subtle cultural differences that exists across China's provinces, which is another factor that we haven't considered in this research.

REFERENCES

Acemoglu, D., et al. (2012). What does human capital do? a review of Goldin and Katz's the race between education and technology. *Journal of Economic Literature*, *50*(2), 426-63. doi:https://doi.org/10.1257/jel.50.2.426

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5), 9-49. doi: https://doi.org/10.1086/258724

Cinnirella, F., & Streb, J. (2017). The role of human capital and innovation in economic development: Evidence from post-Malthusian Prussia. *Journal of Economic Growth*, *22*(2), 193-227. doi:https://doi.org/10.1007/s10887-017-9141-3

Dakhli, M., & De Clercq, D. (2004). Human capital, social capital, and innovation: A multi-country study. *Entrepreneurship & Regional Development*, *16*(2), 107-128. doi:https://doi.org/10.1080/08985620410001677835

Diebolt, C., & Hippe, R. (2019). The long-run impact of human capital on innovation and economic development in the regions of Europe. *Applied Economics*, *51*(5), 542-563. doi:https://doi.org/10.1080/00036846.2018.1495820

Li, J., Fu, J., & Wei, P. (2016). Foreign direct investment spillover, human capital threshold and regional innovation ability—an empirical study based on China's provincial panel data. *Journal of Guizhou University of Finance and Economics*(1), 10-18.

Li, P., Zhao, L.-x., & Wan, J.-b. (2014). The impact of innovation factors on industry innovation performances: An empirical analysis based on Chinese manufacturing and high technology industries. *Studies in Science of Science*, *32*(4), 604-612.

López-Bazo, E., & Moreno, R. (2007). Regional heterogeneity in the private and social returns to human capital. *Spatial Economic Analysis*, *2*(1), 23-44. doi:https://doi.org/10.1080/17421770701232459

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*(4), 281-302. doi:https://doi.org/10.1086/258055

Romer, P. M. (1989). *Human capital and growth: Theory and evidence* (Technical report). National Bureau of Economic Research, New York, NY.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, *51*(1), 1-17.

Teles, V. K., & Joiozo, R. (2011). Human capital and innovation: Evidence from panel cointegration tests. *Applied Economics Letters*, *18*(17), 1629-1632. doi:https://doi.org/10.1080/13504851.2011.556584

Yang, F. (2014). *Human capital heterogeneity and regional industrial upgrading: Labor market segmentation and neg perspective* (Unpublished master's thesis). Zhejiang University, Zhejiang, China.