An Analysis of the Impact of the Types of Higher Education Institutions on graduates' Income —— Using PSM

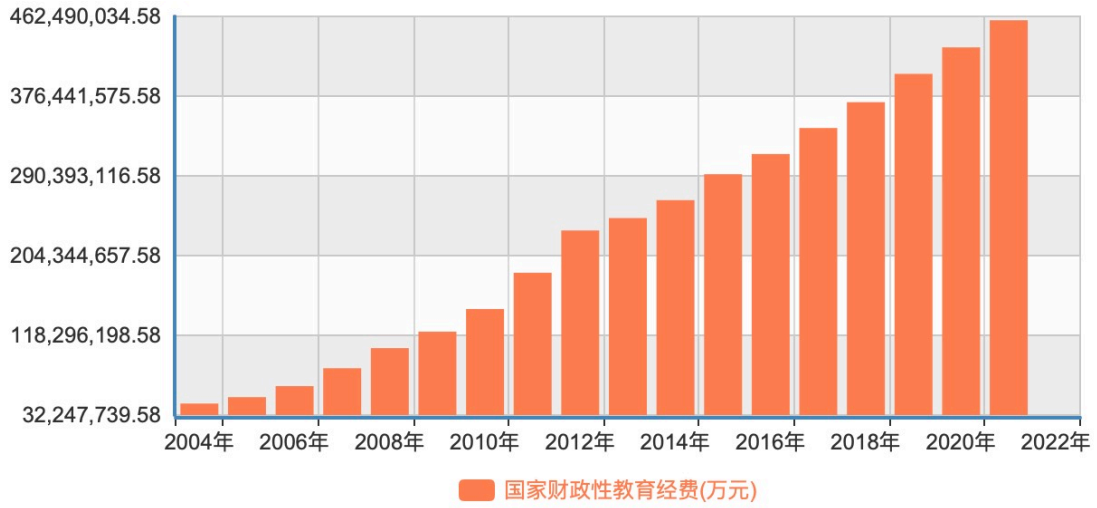
I. Research Background and Significance

1.1 Research Background

Higher education is a special stage. On the one hand, there are long-term supply-side constraints in higher education in China, which necessitates access to higher education through a very competitive selection mechanism. On the other hand, the higher education stage is directly linked to the labor market, and access to and the quality of higher education will affect labor market opportunities. Although higher education opportunities have expanded significantly since the mid to late 1990s, the competitive nature of the selection process remains. To gain a competitive advantage in this selection process, households often tend to competitively increase their educational expenditures, accordingly, including the choice of high-quality senior high schools.

So far, the monetary investment in children's basic education by the government and families in China has been increasing, while the return on education has been on a downward trend in recent years, and the investment in education has not achieved the expected results, both for national development and personal growth. The employment situation in recent years has been severe, and the problem of difficult employment for college students has also made some residents' confidence in education investment shaken.

In this paper, after studying the urban data we used, we found that high quality high school education, like key high schools, has a greater impact on the choice of whether to enter higher education institutions in the future, and the difference between types of institutions also has a greater impact on the expected income after graduation. This disparity in education can further lead to greater income disparities between urban and rural residents, as well as deprive students with high academic achievement but from poor families of access to quality higher education. For our nationals, we need to confront the positive effects of education on individual income and future growth, and gradually increase investment in education while also increasing the proportion of investment in high school education among the rest of education investment at all levels. For rural families, it is important to make the best use of the existing education enrollment system, and to cherish the educational resources they have and encourage and help poor students to study in institutions with better teaching quality.



1.2 Research Significance

Key high schools may, on the one hand, increase starting salaries through a selection effect, i.e., by being a signal of a student's intrinsic ability or family background, and on the other hand, they may increase future starting salaries by increasing a student's human capital and the probability of attending a good college (major).

From the perspective of economic research, the high demand for admission to a key high school necessarily comes from the high returns it is expected to give. Therefore, a very real question is how high the return to the job market after college for students graduating from a high school is if the student's personal ability and family background are fixed. Another question is: What are the payoffs from high schools, and do they only affect the ability to go to a good college? In other words, is a high school a steppingstone to college, or does it itself have a direct impact on post-college advancement? Neither of these questions has been answered by rigorous empirical research.

From the recruitment information of major employers and companies in recent years, we can find that companies focus on the following two aspects when recruiting fresh graduates: their graduation institutions and personal internship experience. The internship experience depends on the personal reserve, which starts in the junior year or even earlier, and the students' awareness of their future planning is a more subjective factor. Therefore, we focus on the aspect of the type of individual's graduation institution, considering the current situation in our country, we divide the educational institution difference into two aspects of research: the graduation institution difference during the undergraduate period and the education difference starting from high school. In general, we believe that the education of undergraduate institutions is better than that of colleges and universities, and the teaching quality of key institutions, i.e., 985 and 211 colleges and universities, is better than that of ordinary undergraduate institutions. Therefore, we discuss the differences in personal income under different education levels and the differences in income of graduates from different types of institutions, to illustrate the impact of differences in higher education institutions on graduates' income in China.

In this study, we found that the effect of the type of high school institution on the type of higher education institution received next is significant, and the comparison between county

key and other non-key high schools and key high schools can be clearly found that the undergraduate rate has a substantial increase. On this basis, it can be confirmed that high school education plays an important role in the development of an individual's mind and has an important impact on the choice of college immediately afterwards, and further influences the future income and development of an individual, so we believe that residents should invest more in high school education.

1.3 Research Content

This paper illustrates the importance of high school education by looking at the differences in the number of students in rural and urban areas for each high school institution type as an entry point, exploring in turn the impact of high school education quality on the choice of college-educated institutions of opportunity, and further exploring the income differences caused by differences in education institution type. One of the main aspects regarding the difference in the quality of college education is represented by the difference in the type of undergraduate college and college degree. In terms of college degrees, we select undergraduate and college degrees as comparisons to illustrate poor education quality, respectively. In terms of undergraduate institution type, since there is no detailed division of educational quality differences among institutions in China before, in this study, we borrowed the division in market recruitment and selected 211 project institutions (including 985 project institutions) as key institutions and other undergraduate institutions as non-key institutions. Based on the characteristics of the index classification, we use logistic regression models to explore the effect of differences in the quality of high school education on the size of individuals' choice of higher education institutions. Given that factors such as individual ability and geographic differences affect not only the size of residents' opportunity to choose a school but also their income, this paper adopts a propensity score matching model.

The rest of the paper includes: the second part is the research literature review; the third part introduces the data sources and provides a descriptive analysis of higher education opportunities and their distribution characteristics between urban and rural areas; the fourth part introduces the PSM model and its assumption; the fifth part presents the estimation results; the fifth part concludes the whole paper; the sixth part is the conclusion of this paper.

II. Research Literature review

The study in this paper is a specific example of the estimation of the rate of return on education. The return on education is a classical problem in labor economics research. There are established literatures on estimating the effect of years of education and educational attainment on labor market wages based on individual microdata of educated individuals, with a Mincerian income function approach. While this kind of literatures are innumerable, this paper focuses on the effect of education quality on labor market wages.

Some domestic studies on the labor market returns to the quality of education have examined the relationship between the quality of secondary school education and student achievement and college admissions, some using DID. One study, based on data from Gansu Province, found that being admitted to a key high school improved the points of

high school entrance exam results and the chance of being admitted to college. And others found that the quality of peers and teachers in junior high school had a significant positive effect on achievement and high school admissions.

After controlling for personal and family factors, this paper found that key undergraduate institutions (defined as "211" colleges and universities in our paper) can increase the starting salary of college students. The methodology in this paper is inspired by an analysis of the effect of type of higher education institution on graduate earnings and examines how the quality of education at a major high school affects labor market performance through different factors at different stages. It is beneficial to identify the impact of the returns to education at a major high school on the future development of educated individuals at different stages.

III. Data Description

3.1 Data Source

The data in this paper come from the National Household Income Survey conducted by the China Income Distribution Panel (CHIP) in 2018, which includes both urban and rural components for each year. The data used in the following study are from the urban part of this paper, and the samples of the survey are from the regular household survey sampling frame of the National Bureau of Statistics, covering provinces with different levels of economic development in the east, middle and west. In this paper, the education level of the target population is set at a range of undergraduate degree. Usually, people in this education level have completed high school education, and even if they did not receive high school education, the possibility of returning to school and choosing the college entrance examination again is very low.

3.2 Index Selection

In this paper, we refer to the studies of existing scholars, and on this basis, we explore the income differences caused by different education levels of urban residents. In this paper, we use individuals who graduated from undergraduate institutions as the intervention group and college as the control group, and we compare the differences between the two to show the impact of education level on expected income and the size of the income differences caused. For the difference in expected income caused by the difference in the type of graduation institutions, we consider all graduates from 985 and 211 institutions as key institutions as the intervention group, and other undergraduate institutions as ordinary undergraduates, and as the control group. We can use the difference in income between the two to illustrate the difference in income of urban residents in China caused by different types of institutions.

As for the effect of education on individual income, in labor economics, most existing studies use the Mincer income equation to explore the returns to education, which is based on human capital theory and uses years of service and years of education as two explanatory variables as the determinants of individual income. We generally believe that there is a positive correlation between productivity and human capital, while the education received in school before entering the workforce and the experience learned on the job are

the two main expressions of human capital, and the Mincer income equation uses years of education as an expression of an individual's educational attainment. Therefore, the variable of years of service is used as an expression of the knowledge learned at work by individuals.

3.3 Data Processing

Considering the differences in the ability of individuals to go to college, here we match individuals with similar abilities based on factors such as the size of the individual's choice of higher education institutions from different provinces and types of high school institutions as weights to eliminate heterogeneous differences among equation variables. The individual's choice of higher education opportunity is affected by many factors, first, one's learning ability depends largely on one's parents' education, and there are great differences in the entrance examination results of students from various types of high schools and colleges, and the admission policy of China's current college entrance examination varies greatly among provinces and regions. Therefore, in the following discussion, we use parents' education level, province and region of the college entrance examination, and type of high school institution as variables influencing individuals' access to higher education. The difference in the type of higher education institutions in this paper is mainly reflected in two aspects, one is the difference in individuals' education in college education: undergraduate and college, and the other is the difference in the type of institutions in which individuals receive undergraduate education: heavy and non-heavy institutions First, we simply convert the original data 2018 resident income into individual hourly income and use the logarithm of hourly income as the equation dependent variable in the model.

In terms of the selection of control variable indicators, as the type of work unit and work industry in the data source have detailed descriptions, for the aspect of the study, combined with the income differences in China on different unit types and industries, we next reclassify the two respectively. In terms of work unit types, the party and government bodies and institutions as the first category, while state-owned enterprises and the enterprises they hold as the second category, while foreign or Sino-foreign joint ventures are classified as the third category, some individual and private enterprises as private enterprises separately as the fourth category, and collective enterprises with other types as the fifth category.

In terms of industry types, here we have the real estate, mining, manufacturing and electricity, gas and water production as the first category, while finance, scientific research and technical services and information transmission, software and technical services are classified as the second category, and basic industries such as education, health and social work, culture, sports and entertainment and public administration, social security and organization are classified as the third category.

In discussing the effect of differences in the quality of high school education on the choice of receiving higher education institutions, whether to receive a heavy college or university degree was taken as the dependent variable, respectively. While the provinces in the control variables are divided according to the characteristics of high school admissions in China and the competitiveness of each province, here we take municipalities such as Beijing, Tianjin and Chongqing as the first category, provinces with less competitiveness such as Inner Mongolia, Jilin, Yunnan and Xinjiang as the second category, provinces with high acceptance rates of heavy undergraduate education such as Jiangsu, Anhui, Hunan and

Anhui as the third category, and relatively few higher education institutions such as Shanxi, Henan and Guangxi provinces as the fourth category. In terms of differences in high school education quality, we are exploring the impact of high school education quality on the choice of higher education institutions by categorizing national or provincial key high schools as the first category, regional (city) key high schools as the second category, county and other key high schools as the third category, and the remaining other types of institutions as the first category. The specific indicator variables used in this paper to down study are described as shown in Table 3-1.

Table 3-1

y	Individuals 2019 hourly revenue
lny	Natural logarithm of hourly revenue
exp	Working experience of individuals at the end 2019
male	=1, if individual is male; =0, if individual is female
edu1	=1, if individual comes from key institutions; =0, else
edu2	=1, if individual has a bachelor's degree; =0, else
sch1	=1, if individual high school is national or provincial key secondary school; =0, else
sch2	=1, if individual high school is regional key secondary school; =0, else
sch3	=1, if individual high school is country or other key secondary school; =0, else
sch4	=1, if individual high school is not key secondary school; =0, else
epro1	=1, if individual comes from Beijing, Shanghai, Tianjin, Chongqing; =0, else
epro2	=1, if individual comes from Nei Monggol Autonomous Region, Guangxi Zhuang Autonomous Region, Hainan, Guizhou, Yunnan, Tibet, Gansu, Qinghai, Ningxia Hui Autonomous Region, Xingjiang Uygur Autonomous Reign; =0, else
epro3	=1, if individual comes from Jiangsu, Zhejiang, Anhui, Fujian, Shandong, Hubei, Hunan, Guangdong, Sichuan, Shanxi; =0, else
epro4	=1, if individual comes from Hebei, Shanxi, Liaoning, Jilin, Heilongjiang, Jiangxi, Henan; =0, else
uni1	=1, if government agencies or institutions; =0, else
uni2	=1, if state-owned enterprises; =0, else
uni3	=1, if sino-foreign joint venture or wholly foreign-owned enterprise; =0, else
uni4	=1, if other joint-stock enterprises, individual or private enterprises, land contractors, other; =0, else
uni5	=1, if collective enterprises; =0, else
ind1	=1, if working in the first category industry; =0, else
ind2	=1, if working in the second category industry; =0, else
ind3	=1, if working in the third category industry; =0, else

3.4 Descriptive Statistics

Table 3-2

Type of Undergraduate Institution - Individual Income Statistics Analysis

Indicator variable	Key institutions		General institutions	
	count	average	count	average
male	235	42.69	2130	31.78
female	157	33.57	1815	25.99
uni1	131	34.80	1242	31.74
uni2	76	42.53	518	29.24
uni3	17	46.46	135	34.55
uni4	164	40.52	1969	27.44
uni5	4	18.81	78	19.86
ind1	4	28.33	36	26.61
ind2	77	43.71	763	31.97
ind3	311	38.01	3146	28.45
total	392	30.03	3945	29.11

A brief statistical description of the variables involved in this paper was made by organizing the original data according to the needs, and the results are shown in Table 3-2. It can be seen from this that after removing the missing observations from the data, the effective sample size is 4337, with 392 from key institutions and 3945 from general institutions. Among the graduates of key institutions, there are 235 males, with an average monthly income of ¥42.69 per hour, which is substantially higher than the ¥33.57 of females. The size of the difference may also be related to the number of the sample, and it can be seen from the table that the number of females is 157, and the sample size of females is about 62.5% of that of males. For the general institutions, the difference is smaller for both males and females with 54% and 46% respectively, while the average hourly income for females is lower than the average hourly income for males, which is in line with the proof of existing scholars and the current situation in China, that is, it is generally believed that female residents in China earn relatively lower wages than males. In addition, both men and women, the average income of individuals with key institution's degree is higher than that of ordinary bachelor's degree.

Also, in terms of enterprise type and industry classification, there is a large difference in per capita hourly income between both key institutions and general institutions. From the enterprise type, it can be found that the highest individual income is in the third category, i.e., employees in foreign or Sino-foreign joint ventures, while the two lowest income categories are the first category of state-owned unit employees and other categories (including collective enterprises, etc., which are not detailed here because the sample size is smaller compared to other variables). In terms of number distribution, this sample is mostly from the fourth category of other shareholding units, but compared to the first category of state-owned enterprises is not too different, the two accounted for up to 80% of the employees of these two types of work is characterized by their stable income and relatively low work intensity.

In addition, compared to other industries require higher education, similarly, it can also be seen that there are large differences between different industries, and the relatively highest income is the second category of industries such as finance, software and technical services, the lowest income is the first category of industries, mainly manufacturing, while the third industry is also the relatively largest number of industries, accounting for 80%. The

characteristics of these industries are mostly public service industries, i.e., mostly state-owned units, which also illustrates the common characteristic of our residents to pursue a stable life. However, no matter what industry they come from, there are significant differences in personal income among graduates from different types of institutions.

Table 3-3

Contribution of key high schools to key institutions

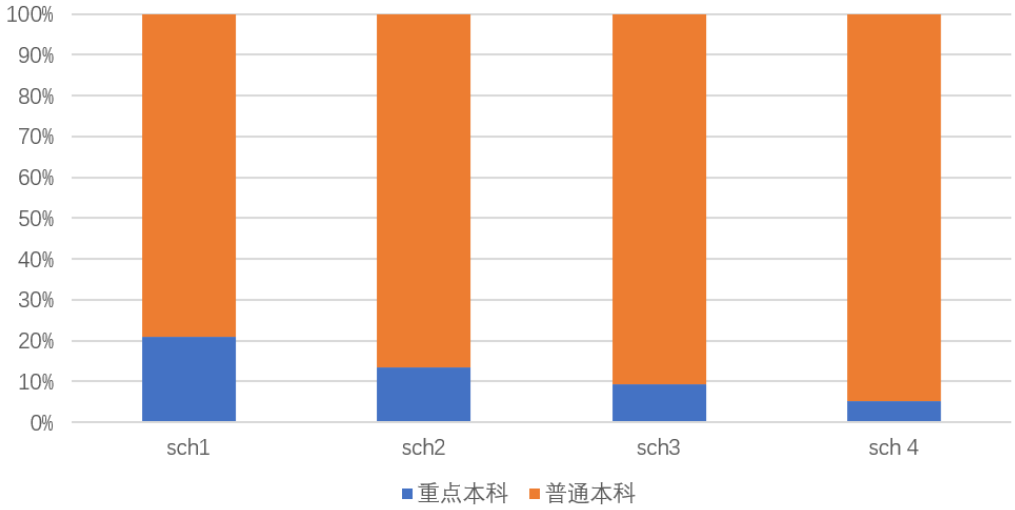


Table 3-3 indicates the distribution of graduates from each type of high school in terms of the type of university undergraduate degree. Where sch1, sch2, sch3 and sch4 denote national or provincial key high schools, regional (city) key high schools, county and other key high schools and other types of high schools, respectively. From the figure, we can see that the probability of entering a key undergraduate school becomes progressively smaller as the level of the high school institution decreases in order. In line with the reality of high school institutions in China, national or provincial key high schools have the highest rate of admissions to the key bachelor's degree, and about 20% of college graduates from national or provincial key high schools receive education from major bachelor's degree institutions, while the remaining 4/5 receive education from general bachelor's degree institutions. In contrast, about 10% of undergraduate graduates from high school institutions from county-level key high schools receive a key undergraduate education. This is because there is a big difference between each type of high school in terms of teacher resources, school environment and hardware facilities, and the student population at the time of enrollment. Generally, we believe that students with higher academic ability and more outstanding ability have a greater chance of receiving a relatively higher quality education in high school. Therefore, when considering individual differences in access to higher education institutions, we also consider the individual's high school institution type as an influential indicator variable.

Table 3-4

Undergraduate/college-Personal Income Statistics Analysis

Indicator variable	Bachelor's degree		College's Degree	
	count	average	count	average
male	1308	36.30	1057	28.62
female	1114	29.04	858	23.42

uni1	957	32.50	416	30.95
uni2	374	33.11	220	27.25
uni3	86	43.83	66	25.53
uni4	965	32.83	1168	24.82
uni5	38	21.82	44	18.07
ind1	17	27.60	23	26.17
ind2	425	37.03	415	28.95
ind3	1980	32.13	1477	25.54
total	2422	32.96	1915	26.29

Tables 3-4 show the results of descriptive statistics on the effect of university degree differences on graduates' income. Among them, there are 4337 valid samples, and there are 1915 college degrees, accounting for 44%. From the overall perspective, the average income of men is higher than that of women for both undergraduate and college degrees, and the income of undergraduates is higher than that of college graduates.

From the viewpoint of enterprise types, in the first category of state-owned units, the number of graduates with bachelor's degree accounts for 40%, while the number of graduates with college degree only accounts for 21%, which is because most state-owned units have restrictions on education when recruiting. The proportion of private enterprises with bachelor's degree is 40%, while the corresponding proportion of college degree is 60%, which is related to the nature of private enterprises. In terms of income differences, the difference in income between the types of enterprises with college degree is more obvious, the highest is still foreign-funded enterprises, while the best income among college degrees is state-owned enterprises, but in comparison, the difference between the types of industries is not too big. This is in line with the actual situation in our country, compared with the college degree, the difference in education will limit a person's development in the later stages, so it is not possible to pull away too much distance in each industry. From the point of view of the industry to which it belongs, the third industry category in the bachelor's degree in the number of 82%, in the college accounted for 77%, and this is because the third industry is education, functional common management, social security and organizations, such as most of the state-owned nature of public services. In terms of income, the highest average hourly income is the second industry, mostly finance, information transmission, software and technology services and other highly competitive industries, income is relatively low for the first industry, which also shows that although China's manufacturing industry is developed, overcapacity led to the manufacturing industry began to lay off or shut down, so the average wage is low.

Table 3-5

Different Levels of High School Graduates in Higher Education

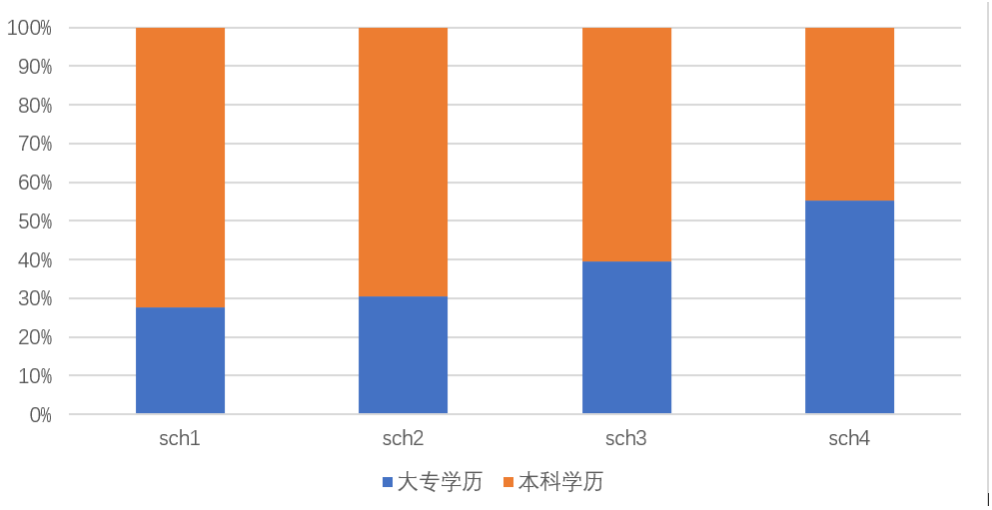


Figure 3-5 shows the difference in the percentage of college degrees by the type of high school, from which it is found that as the quality of high school education decreases, the percentage of graduates with bachelor's degree becomes progressively smaller. Among the graduates from non-key institutions, the number of students with bachelor's and college degrees is the same. Considering that the quality of high school education can, to some extent, reflect the learning ability of individuals, the type of high school institution of individuals is also used as an indicator variable in the following discussion of the opportunity choice of individual higher education institutions.

Through the above simple descriptive statistics on the data, we can find that education has a significant effect on individual income, both in terms of education and type of undergraduate institution. And there is also a significant difference in the percentage of graduates from different levels of high school institutions in different types of higher education institutions. Geographical differences in enrollment at the time of our college entrance exams also have a significant impact on the size of an individual's choice of access to higher education. It is with these factors in mind that in the next study of the effect of type of higher education institution on individual income, we will match individuals with similar access to higher education and compare them to illustrate the income differences brought about by the type of higher education institution.

IV. The Model Used in this Paper

4.1 PSM

The indicators selected in this paper are hourly income, gender, work experience and education, and the control variables selected include enterprise type, industry and region. From the attributes of variables, only hourly income is a quantitative variable, and the rest are qualitative variables. In many essays, researchers have used multiple linear regression, instrumental variable method, matching and other methods to study this problem. However, individual education and personal income interact with each other, which results in biased and inconsistent estimators of multiple linear regression. For the method of instrumental variables, it is not easy to find exogenous variables that only affect education but have no effect on personal income to be used as instrumental variables. Therefore, in

this paper, we adopt the method of propensity matching score (PSM) to study this problem.

4.1.1 PSM Model

For individual i, two results can be classified depending on whether a particular process is performed or not:

$$Y_i = \begin{cases} Y_{1i}, if\ D_i = 1 \\ Y_{0i}, if\ D_i = 0 \end{cases}$$

D_i : whether individual i is treated or not. If is treated, then equals to 1; If not treated, then equals to 0.

Y_{1i} : Results of the treated individuals

Y_{0i} : Results of the untreated individuals

Under the observable x_i , the probability of individual i being treated is:

$$p(x_i) = Pr(D_i = 1|x = x_i) = E(D_i|x_i)$$

Therefore,

$$\begin{aligned} ATT &= E\{y_{1i} - y_{0i}|D_i = 1\} \\ &= E[E\{y_{1i} - y_{0i}|D_i = 1, p(x)\}] \\ &= E[E\{y_{1i}|D_i = 1, p(x)\} - E\{y_{0i}|D_i = 0, p(x)\}|D_i = 1] \end{aligned}$$

4.1.2 PSM Assumption

i. Common Support Assumption

$$0 < Pr(D_i = 1|X_i) < 1$$

Individuals that appear in the treatment group can also be found in the control group corresponding to.

ii. Balancing Assumption

$$D_i \perp x_i|p(x_i)$$

Before receiving the treatment, there is no difference between the treatment and control group, and the effect produced by the treatment group came exclusively from the experimental treatment.

4.2 Logistics Regression Model

In this paper, the types of undergraduate institutions are divided into key and non-key, and university degrees are divided into undergraduate and college degrees. Both are dummy variables. Therefore, we can use Logistics regression model to estimate the probability of getting into key institutions and receiving undergraduate degrees.

A person's access to higher education is influenced by factors such as gender, location, and the type of high school attended, therefore, the logistics regression model can be written as:

$$e(z) = P(D_i = 1|Z_i = z_i) = \frac{1}{1 + e^{-Z}}$$

Z composes gender(male), province where individual has college entrance exam(epro) and the type of high school(sch)

4.3 Multiple linear regression model

In constructing the model, we draw on previous discussions of the effect of education on earnings. We chose the logarithm of individual hourly earnings as the dependent variable and

the choice of institution, gender, length of working, squared length of working, region, type of establishment and industry as the independent variables to construct the multiple linear regression equation. To eliminate the heteroskedasticity, we chose the variance weight of individual's chance of receiving higher education as the variance weight because individual's income is related to individual's learning ability in addition to industry.

Therefore, the multiple linear regression model is as follow:

$$lny_i = \alpha + \beta X_i + \tau T_i + w_i \varepsilon_i$$

y_i is the income of individual i. T_i is whether accept better college education. τ is the average effect of accepting better college education. X_i is a group of covariates consisting of male, exp, exp_2, pro, uni, ind. w_i is the weight.

V. Analysis of the effect of type of higher education institution on personal income

5.1 Effect of type of higher education institution on personal income

5.1.1 PSM Results

When we discuss income disparity due to college level, we set students from key institutions as the intervention group and students from the normal institutions as the control group. If individual graduates from key institutions, that means that edu_i=1, otherwise, edu_i=0. And the covariates we chose are the types of high school, which are sch1, sch2, sch3 in the regression, and the province where the individual take the college entrance exam, which are epro1, epro2, epro3 in the regression. We use logistic regression to estimate the probability of individual i getting into key institutions.

Table 5-1

```
Logistic regression                                     Number of obs =  4,316
                                                        LR chi2(7)    = 128.10
                                                        Prob > chi2   = 0.0000
Log likelihood = -1245.3105                             Pseudo R2     = 0.0489

------------------------------------------------------------------------------
        edu1 | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        male |   .2689379   .1099788     2.45   0.014     .0533834    .4844924
        sch1 |   1.560584   .1726849     9.04   0.000     1.222128     1.89904
        sch2 |   1.061534   .1421666     7.47   0.000     .7828926    1.340176
        sch3 |   .6480164   .1451472     4.46   0.000      .363533    .9324997
       epro1 |    .666706   .1859962     3.58   0.000     .3021601    1.031252
       epro2 |  -.2506494   .2126895    -1.18   0.239    -.6675133    .1662144
       epro3 |   .2163166   .1406007     1.54   0.124    -.0592558    .4918889
       _cons |  -3.241911    .163304   -19.85   0.000    -3.561981   -2.921841
------------------------------------------------------------------------------
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
lny	Unmatched	3.42960119	3.10154029	.328060899	.04175589	7.86
	ATT	3.42960119	3.14676466	.282836534	.040065055	7.06

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated	3,926	3,926
Treated	390	390
Total	4,316	4,316

In the above logistic regression, which is shown in table 5-1, the coefficient for male, sch1, sch2, sch3, and epro1 are both statistically significant at the level of 5%, which means that the probability of getting into key institutions is affected by gender, the types of high school and the province where the individual takes the college entrance exam.

Use odds ratio to further explain the economic meaning of this logistic regression.

$$ln\left(\frac{P(edu_1 = 1|Z_i = z_i)}{1 - P(edu_1 = 1|Z_i = z_i)}\right)$$

$$= -3.242 + 0.269male + 1.561sch1 + 1.061sch2 + 0.648sch3 + 0.667epro1 - 0.251epro2 + 0.216epro3$$

If an individual i is male instead of female, then the probability of getting into key institutions divided by the probability of not getting into key institutions increases by 26.8%. If an individual i graduated from national or provincial key high school, then the probability of getting into key institutions divided by the probability of not getting into key institutions increases by 156.1%. If an individual i graduated from regional key high school, then the probability of getting into key institutions divided by the probability of not getting into key institutions increases by 106.1%. If an individual i graduated from country or other key high school, then the probability of getting into key institutions divided by the probability of not getting into key institutions increases by 64.8%. If an individual i takes the college entrance exam in Beijing, Shanghai, Tianjin and Chongqing, then the probability of getting into key institutions divided by the probability of not getting into key institutions increases by 66.7%.

Besides, from the table 4-1, we find that the T-stat of ATT is 7.06, which is larger than 1.96, and that means that it is statistically significant at the level of 5%. And that means that the type of higher education institution(edu1) does have effect on the natural logarithm of income(lny)

We perform a weighted multiple linear regression of individual income using the weights

obtained above.

Table 5-2 Effect of type of higher education institution on personal income

	(1) OLS	(2) PSM_OLS
edu1	0.307***	0.237***
	(8.012)	(5.924)
male	0.188***	0.191***
	(8.006)	(2.746)
exp	0.017***	0.018***
	(11.134)	(4.381)
uni1	0.392***	0.379
	(4.374)	(1.274)
uni2	0.346***	0.453
	(3.778)	(1.509)
uni3	0.653***	0.810***
	(6.629)	(2.613)
uni4	0.372***	0.476
	(4.236)	(1.601)
ind1	-0.332**	-0.037
	(-2.007)	(-0.275)
ind2	0.098***	0.042
	(3.032)	(0.444)
_cons	2.424***	2.391***
	(27.805)	(8.136)
N	4316	4316
R^2	0.084	0.088
adj. R^2	0.082	0.086

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

From table 5-2, the coefficient we most care about is the coefficient of edu1. And the dependent variable is the natural logarithm of hourly earnings(lny). According to the above table, in the PSM_OLS regression, the coefficient of edu1 is statistically significant at the level of 5%, which means that if individual i graduate from key institutions, individual i's income will increase by 23.7%.

Besides, in the PSM_OLS regression, the coefficient of male, exp and uni3 is statistically significant at the level of 5%. If individual i is male, then his income will increase by 19.1%. If individual i have one more year of working experience, then his income will increase by 1.8%.

However, the coefficient of uni1, uni2, uni4, ind1 and ind2 in the PSM_OLS regression is not statistically significant even at the level of 10%, but these coefficients are both statistically significant in the OLS regression, we can probably draw the conclusion that the type of enterprises and industries also have effect on income.

5.1.2 Test of PSM Assumption

We perform a weighted multiple linear regression of individual income using the weights obtained above. And the result is as follow.

Table 4-3

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|	V(T)/ V(C)
male	U	.6	.53897	12.3		2.31	0.021	.
	M	.6	.58501	3.0	75.4	0.43	0.671	.
sch1	U	.16923	.06368	33.3		7.68	0.000	.
	M	.16923	.14076	9.0	73.0	1.10	0.272	.
sch2	U	.3	.19078	25.6		5.15	0.000	.
	M	.3	.28985	2.4	90.7	0.31	0.756	.
sch3	U	.2641	.2593	1.1		0.21	0.837	.
	M	.2641	.26477	-0.2	86.2	-0.02	0.983	.
epro1	U	.15385	.09832	16.8		3.44	0.001	.
	M	.15385	.14083	3.9	76.6	0.51	0.609	.
epro2	U	.08974	.14799	-18.1		-3.14	0.002	.
	M	.08974	.10021	-3.2	82.0	-0.50	0.619	.
epro3	U	.55897	.51987	7.8		1.47	0.140	.
	M	.55897	.55514	0.8	90.2	0.11	0.914	.

* if variance ratio outside [0.82; 1.22] for U and [0.82; 1.22] for M

Sample	Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
Unmatched	0.049	128.03	0.000	16.4	16.8	60.7*	1.14	.
Matched	0.003	2.74	0.908	3.2	3.0	11.8	1.02	.

* if B>25%, R outside [0.5; 2]

Table 5-4

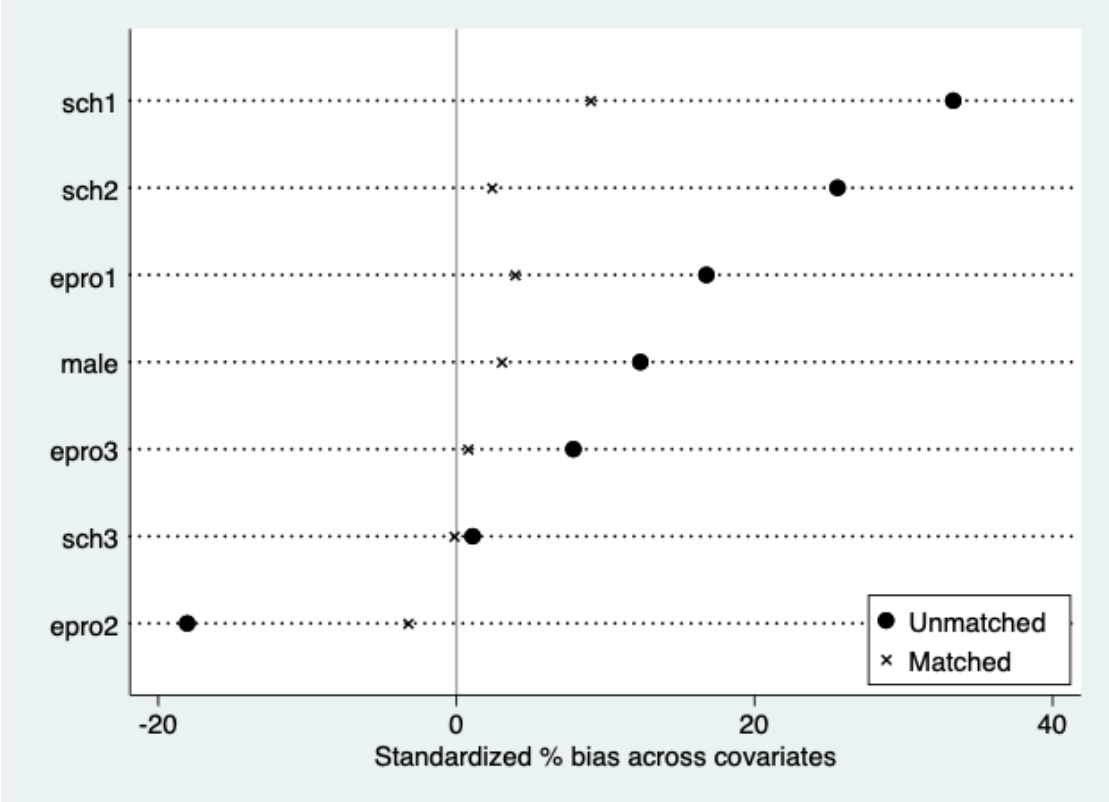
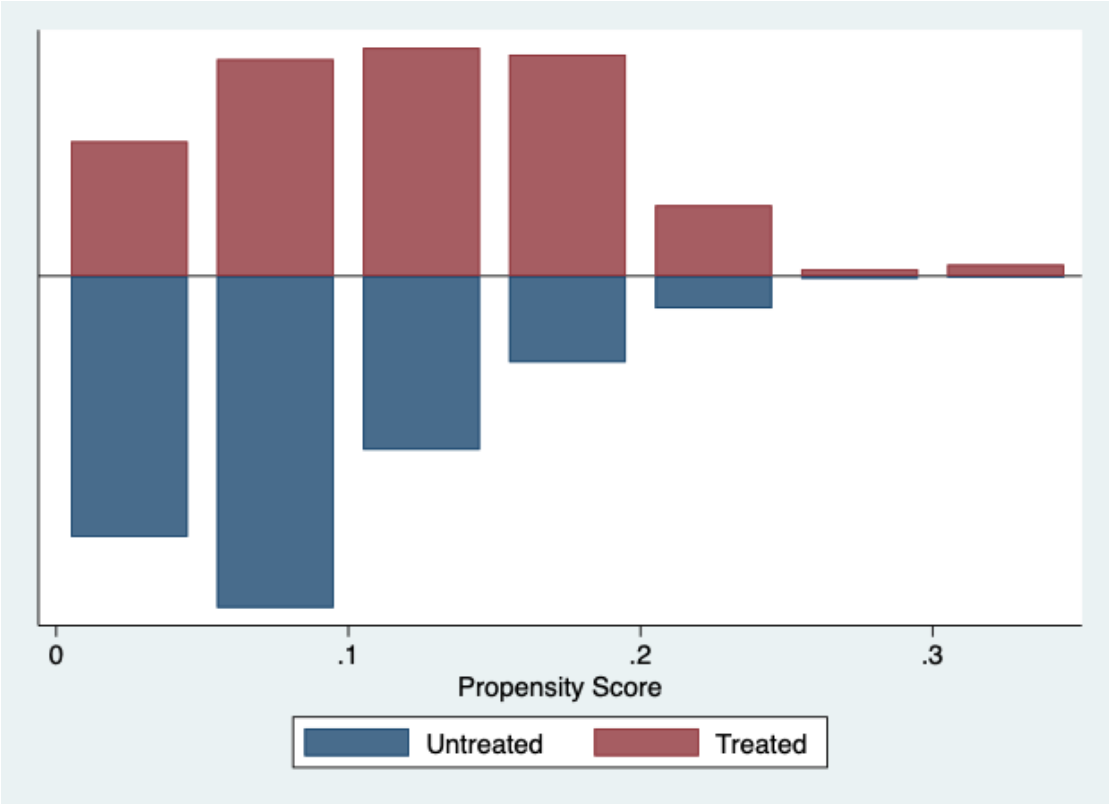


Table 5-5



From table 5-3 we can observe that both bias is less than 10%, which means that it passes the common support assumption.

5.2 Effect of type of undergraduate institution on income differences

5.2.1 PSM Results

When discussing the effect of type of undergraduate institutions differences on earnings, the intervention group we use is the sample with a bachelor's degree and the control group is the sample with a college degree. Therefore, when individual i has a bachelor's degree, then $T_i = edu2_i = 1$, otherwise $T_i = edu2_i = 0$. And the covariates we chose are the same as above. We use logistic regression to estimate the probability of individual i having bachelor's degree.

Table 5-6

```
Logistic regression                                     Number of obs =  4,316
                                                        LR chi2(7)    = 229.29
                                                        Prob > chi2   = 0.0000
Log likelihood = -2846.5376                             Pseudo R2     = 0.0387

------------------------------------------------------------------------------
        edu2 | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        male |  -.0300096   .0632759    -0.47   0.635     -.154028    .0940088
        sch1 |    1.21886   .1344549     9.07   0.000     .9553335    1.482387
        sch2 |   1.028868   .0865818    11.88   0.000     .8591705    1.198565
        sch3 |   .6514211    .076213     8.55   0.000     .5020463    .8007959
       epro1 |   .1700509     .11921     1.43   0.154    -.0635965    .4036982
       epro2 |   .1860266   .1068055     1.74   0.082    -.0233083    .3953616
       epro3 |  -.1426876   .0788599    -1.81   0.070    -.2972501     .011875
       _cons |  -.1646201   .0816246    -2.02   0.044    -.3246014   -.0046389
------------------------------------------------------------------------------
------------------------------------------------------------------------------
        Variable     Sample |    Treated     Controls   Difference          S.E.   T-stat
-----------------------------+------------------------------------------------------------
             lny  Unmatched | 3.26089586   2.96655575   .294340106   .023867842    12.33
                       ATT | 3.26089586   2.96241336   .298482502   .025566679    11.67
-----------------------------+------------------------------------------------------------
Note: S.E. does not take into account that the propensity score is estimated.

           | psmatch2:
 psmatch2: |   Common
 Treatment |  support
assignment | On suppor |     Total
-----------+-----------+----------
 Untreated |     1,902 |     1,902
   Treated |     2,414 |     2,414
-----------+-----------+----------
     Total |     4,316 |     4,316
```

From table 5-6, we can observe that the coefficient of sch1, sch2, sch3, epro2, epro3 are statistically significant at the level of 10%, which means that the type of high school and the province where individual i takes the college entrance exam do have effect on the probability of having bachelor's degree. However, the coefficient of male which means gender is not statistically significant, which means that gender do not have obvious effect on the probability of having bachelor's degree.

We can use odds ratio to further explain the economic meaning of this logistic regression.

$$ln\left(\frac{P(edu_2 = 1|Z_i = z_i)}{1 - P(edu_2 = 1|Z_i = z_i)}\right)$$

$$= -1.646 - 0.030male + 1.219sch1 + 1.029sch2 + 0.651sch3 + 0.170epro1 + 0.186epro2 - 0.143epro3$$

If an individual i graduated from national or provincial key high school, then the probability of having bachelor's degree divided by the probability of not having bachelor's degree increases by 121.9%. If an individual i graduated from regional key high school, then the probability of having bachelor's degree divided by the probability of not having bachelor's

degree increases by 102.9%. If an individual i graduated from country or other key high school, then the probability of having bachelor's degree divided by the probability of not having bachelor's degree increases by 65.1%. If an individual i takes the college entrance exam in Nei Monggol Autonomous Region, Guangxi Zhuang Autonomous Region, Hainan, Guizhou, Yunnan, Tibet, Gansu, Qinghai, Ningxia Hui Autonomous Region, Xingjiang Uygur Autonomous Region, then the probability of having bachelor's degree divided by the probability of not having bachelor's degree increases by 18.6%. If an individual i takes the college entrance exam in Jiangsu, Zhejiang, Anhui, Fujian, Shandong, Hubei, Hunan, Guangdong, Sichuan, Shanxi, then the probability of having bachelor's degree divided by the probability of not having bachelor's degree decreases by 14.3%.

Besides, from the table 4-6, we find that the T-stat of ATT is 11.67, which is larger than 1.96, and that means that it is statistically significant at the level of 5%. And that means that the type of undergraduate institution(edu2) does have effect on the natural logarithm of income(lny)

We perform a weighted multiple linear regression of individual income using the weights obtained above.

Table 5-7 Effect of type of undergraduate institution on income differences

	(1) OLS	(2) PSM_OLS
edu2	0.288***	0.297***
	(11.916)	(11.375)
male	0.194***	0.199***
	(8.358)	(7.831)
exp	0.018***	0.017***
	(11.636)	(10.299)
uni1	0.341***	0.289***
	(3.828)	(3.363)
uni2	0.323***	0.261***
	(3.557)	(2.956)
uni3	0.647***	0.582***
	(6.703)	(6.155)
uni4	0.389***	0.349***
	(4.470)	(4.203)
ind1	-0.284*	-0.291*
	(-1.688)	(-1.662)
ind2	0.103***	0.091**
	(3.228)	(2.577)
_cons	2.292***	2.333***
	(26.306)	(27.891)
N	4316	4316
R^2	0.102	0.100
adj. R^2	0.100	0.098

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In table 5-7, the coefficient we most care about is the coefficient of edu2. And the coefficient of edu2 is statistically significant at the level of 1%. It means that if individual i has a bachelor's degree, then his income will increase by 29.7%. Besides, the coefficients of male, exp, uni1, uni2, uni3 and uni4 is statistically significant at the level of 1%, and the coefficient of ind1 is statistically significant at the level of 10%, and the coefficient of ind2 is statistically significant at the level of 5%. These means that individual's type of undergraduate institutions, gender, working experience, types of enterprises and industries have effect on his income.

5.2.2 Test of PSM Assumption

Because we use the same covariates as above, so it passes the common support assumption.

VI. Conclusion

From the previous descriptive analysis of the data and the two empirical analyses above,

the following conclusions can be summarized.

First, from the above two logistic regressions, it was found that the coefficients of sch1, sch2, and sch3 were significant, which means that we can conclude that the type of individual high school institution has a significant effect on the choice of higher education institution opportunities. National or provincial key high schools have a greater chance of entering a key institution than regional key high schools. Regional key high schools have a higher chance of entering a key institution than county or other key high schools.

Second, a person's work experience, gender, type of enterprise, and type of industry have a significant effect on income.

Third, differences in institutions of higher education, indicated by both the type of undergraduate institution and the difference in college degrees, had a significant effect on personal income. Students graduated from key institutions earning about 23.7% more compared to students graduated from regular institutions. Undergraduate graduates earn about 29.7% more than college graduates.

In addition to the province where individual takes the college entrance exam, the quality of higher education is also related to the quality of education an individual receives in high school. Graduates from national or provincial key high schools, regional (city or district) key high schools, county key high schools and non-key high schools have decreasing chances to enter high quality higher education institutions. There is a significant difference in the choice of high school between urban and rural residents, i.e., rural residents are more likely to choose to attend county-level high schools, while urban residents could attend regional (city or district) or provincial key high schools according to their learning ability and environment. In addition, due to the difference in economic power, rural residents invest less in education than urban residents, so to obtain higher returns on education and narrow the difference in education choices with urban residents, rural residents should differentiate their investment in each stage of their children's education, considering their economic power, given the indirect effect of high school education on residents' income in China. We believe that we should invest more in high school education and encourage children to study in high schools of relatively high quality during the secondary school entrance exams, such as taking the entrance exams of provincial or municipal key high schools.