The Relationship Between Education and Income in the US

Zhenyu Wang, Yi Ding, Yingbo Zhou, Longxiang Li

EC492 – Senior Thesis Econometrics

Adam Cai

The University of West Alabama

April 3, 2021

Contribution of work:
Zhenyu Wang: Data gathering, Sample selection, Literature review, Building empirical model and Data summary
Yingbo Zhou: Screening model, Literature review, Regression analysis, Interpret the results and Summary empirical model
Yi Ding: Data gathering, screening model, Building empirical model, Literature review, Regression analysis
Longxiang Li: Introduction part, Literature review, screening model, Regression analysis and Summary empirical model

Introduction

According to the World Bank, classical economics knew of the strong relationship between education and income. However, it was not until the 20th century that a strong relationship between education and income was ascertained in theory and practice. Other than education, income is determined by various issues. Such issues include racial and ethnic discrimination, social and economic policy, and gender discrimination. Across various societies in the world, education has always been regarded as a determinant of success. It widens an individual's mind and opens opportunities for better jobs. Education strongly relates to income levels among individuals. Across the world, it is common for households that have higher levels of education to have more income and thereby be able to withstand financial upsets (Beckles & Truman, 2011). Education makes people gain effective knowledge of managing their finance and building their wealth. Financial literacy is a key component in the wealth creation process. Wealth building and creation require shrewd financial decision-making that can only be obtained through education. This paper assesses the relationship between education and income, entails a literature review on the various literature that has explored the correlation between educational levels and income. Education can be regarded to be an investment in a person whose payment will be through income.

The rising and high-income inequality across the world has raised social concern. The United States is one such country that has experienced rising income inequality. A limitation in the economic and social solutions established to solve the situation is the failure to put the role of education in income disparities into consideration. Unemployment means that an individual cannot have a source of income. In the United States, better educational groups experience the lowest unemployment rates. Therefore, it is difficult to find highly educated individuals lacking a form of income. A similar trend is experienced in different countries where the highly educated attract the highest income while the low-income groups are mostly uneducated (Shu, 2010). Even if an assumption is made that the countries' labor markets are imperfect, the correlation between education and income cannot be ignored. Workers in demand represent highly trained and educated workers. On most occasions, the workers on-demand attract the highest income. The high income is largely influenced by the market conditions of demand and supply. Contrary, workers not in demand are usually given weak-to-flat wages. Again, the market forces of demand and supply determine their wages as their skills can easily be replaced. Therefore, the relationship between education and income is that individuals with higher educational levels experience higher income levels because they are employed at high rates.

Education can be regarded to be an investment in human capital. In economics, individual investments in human capital and financial assets have similar motifs, i.e., they are focused on making money. College wage premium refers to the higher incomes that can be directly associated with a college degree. Research in the US has shown that the premium has grown as the country's educational levels (Bennett & Vedder, 2015). Education imparts important skills to individuals. In most labor markets, the more skills an individual or worker possesses, the more employable he will become. Income is the benefit directly associated with employment. The relationship between education and income can also be examined from the point of view of wealth. A strong relationship exists between education and wealth.

The wealth creation process begins with higher earnings that ease saving abilities. Savings lead to wealth creation. Individuals with lower incomes do not make savings hence do not create wealth. On the contrary, individuals with more education can make sound financial decisions that directly contribute to the wealth creation process.

Literature Review

Abdullah et al. (2015) explore the role education plays in reducing income inequality in different countries. According to Abdullah et al. (2015), education affects two facets of distribution income: increasing the income of low earners and reducing the top earners' income. Abdullah particularly focuses on the relationship between education and income in Africa. The author asserts that education has been particularly helpful in decreasing income inequality in Africa. The author's stance is that income inequality has been a fundamental political, economic and social issue. Income inequality stifles political stability and economic growth. Income inequality may result in ethnic and class tensions. Existing literature has shown that factors such as developmental level, urbanization, government intervention, and the political regime directly result in income inequality. However, the author states that education is the most critical indicator of income inequality. Abdullah et al. conducted a meta-regression analysis to explore the relationship between education and income. The researchers conducted a comprehensive review of 64 econometrics literature on the correlation between education and income. The econometric studies resulted in interesting aspects of the relationship between education and income.

Abdullah et al. (2015) noted that education significantly raises the income of the poor. A comparison between the meta-regression analysis coefficient of the income share bottom and the income share top indicated that education has a lower effect on reducing the rich population's income share. The researchers found that education significantly reduces the income inequality gap between the high-income and low-income people. According to the meta-regression analysis, education is an efficient tool that can reduce income inequality, especially in developing regions. A limitation of the research is that the meta-regression analysis does not provide insights into education's opportunity costs. It does not highlight how income inequality will be affected if the emphasis is not made on education. This research has failed to conduct a cost-benefit analysis that would have provided information on the importance of policy alternatives in reducing various levels of inequality in a country. However, despite its shortcomings, the meta-regressions analysis in the research has provided a detailed analysis of how education relates to income and how it can reduce income inequalities. The phenomenon of income inequalities exists because there are inadequate social and economic structures. However, the issue can be solved if more people learn about wealth creation strategies through education.

Yang and Qui (2016) assessed the relationship between education and income from the perspective of how education can reduce income inequality. Besides, the researchers also focused on the impact of education on intergenerational mobility of income. Intergenerational mobility of income defines how wealth changes hand across generations. According to Yang and Qui (2016), economists and economic models have been more concerned with how random markets cause income inequalities. Intergenerational mobility of income is

determined by the different generations' key financial decisions that have resulted in the expansion of the family's income base. In terms of education, Yang and Qui (2016) posit that early education in China is jointly paid by family expenditure and government. The government's role in education entails paying textbook and tuition fees. On their side, parents pay sponsor fees, tutoring fee, among other miscellaneous fees. The researchers assert that higher private costs directly relate to better educational quality. Through education, it can be assumed that the parents care about the quality of life in the next generation. Therefore, parents take their children to school to assure them a better life in the future. In the study, a better life implies a life characterized by income and intergenerational wealth. The study was based on a calibrated model to explore the origins of intergenerational income mobility and income inequality.

In terms of education, this study explored several aspects of education, such as compulsory education, non-compulsory education, and the innate ability to study the phenomenon of intergenerational income mobility and income inequality (Gregorio & Lee, 2002). The study notes that wealthy families tend to invest more in their children's education. In their turn, the children learn more about financial spending and financial decision-making. Through education, the children gain a better understanding of financial sustainability than the previous generation. Therefore, they end up creating more wealth than the ones they receive. Through the wealth, they can easily earn income a return on their investments.

On the other hand, poor families may lack the necessary funds to finance their children's education. As a result, the children may end up not getting the ideal education to enable them to be competitive enough in the wealth creation process (Keller, 2010). As a result, they may end up being poor. A limitation of this study is its focus on intergenerational income mobility, income inequality, and education. It does not explore other means of creating wealth other than education. Therefore, it may lack objectivity when assessed with different factors other than the ones highlighted.

Income inequality is among the issues that have dominated the social and economic world order. Income inequality relates to the situation in which different individuals in an economic setting have different income levels. There are various social, political, and economic causes of income inequality in society. Capos et al. (2016) explores the relationship between education and income by exploring the effects of education on income inequality. The ethnic minorities may face systematic barriers that hinder them from accessing education opportunities. According to Capos et al. (2016), economists have noticed a causal link between education and poverty. As a result, the effects of education and training on income distribution have become more visible. About the relationship between education and income, the human capital theory asserts that the best way of enhancing income distribution is through long-term investments in human capital. Education defines such an investment in human capital. Economists believe that the economic structures will improve, and better balancing on income distribution will be achieved through an increase in learning and vocational training. Capos et al. (2016) suggest that an investment in education influences each person's income. As the investing rates in education increase, the outcome rate will increase, and consequently, personal income will increase.

Cody and Dizioli (2018) assess the correlation between education expansion and income inequality. The researchers found a large and statistically significant relationship

between education and income inequality. The study explores education from the perspective of inequality schooling and education expansion, especially in developing and emerging economies. According to Cody and Dizioli (2018), the existence of high-income inequality in different countries has been a concern for economists and policymakers. The researchers attribute the high levels of income inequality to factors such as the advent of skill-biased technology. Education expansion can be regarded as a fundamental policy in dealing with the rising levels of income inequality. Education expansion promotes economic growth and breaks the intergenerational linkage of poverty. A focus on reducing income inequality through education will limit the implementation of fiscal policies as a means of fiscal redistribution. This study utilizes a dynamic panel estimation method to determine the positive relationship that exists between education and income inequality. The paper's theoretical insight is that human capital expansion is the best framework for solving the income inequality issues within a society (Goldrick-Rab et al., 2014). Therefore, the dynamic panel estimation method in the study resulted in similar insights as to its theoretical understanding of the issue. Education expansion relates to making education more accessible to marginalized communities and ethnic minorities. Cody and Dizioli (2018) provide evidence that education expansion for the past one and a half-decade has reduced income inequality. This effect of education expansion has especially been experienced in developing and emerging economies. Therefore, rising education levels among ethnic minorities are key to increasing their income.

Saad et al. (2012) analyze how education has affected income among ethnic minorities in Asian Americans. According to the authors, educational achievements among Asian Americans can be correlated to their occupational aspirations. According to the researchers, education has a direct relationship with income in such a way that higher levels of education are equivalent to higher incomes. According to the research, higher education among ethnic minority groups usually elevates their lives by increasing their disposable income. According to Saad et al. (2012), minority ethnic groups often face various barriers when seeking to improve their economic opportunities. They may have weak infrastructural networks and may be neglected by the government. As a result, such ethnic groups are usually left with education as the only way of improving their economic well-being. Through education, members of the minority usually gain the requisite skills and knowledge to create their sources of income creation. After education, the ethnic minority groups may open investments that will ultimately become a source of income. Therefore, Saad et al. (2012) showed that the higher the level of education, the higher the income level, other things held constant.

Porter (2014) used a simplistic economic and logical model to show that more education equals income. According to Porter, the years between 1979 and 2012 experienced a rise in the wage gap between families of college graduates and high school graduates' families. Porter (2014) indicates that the wage gap between the years grew by $30 000. According to Porter (2014), the amount is huge and gives all the reasons that more education equals higher levels of income. The statistics indicate that the highly educated will continue experiencing an increase in income while the less educated incomes will remain constant. In the late 20th and early 21st century, individuals with a college degree experienced a significant rise in income because of the expansion in the economy. Besides, this period was

characterized by a rise in technology. Since the educated individuals had the requisite skills to operates technology-related tasks, they would acquire technology-related jobs. As a result, the educated experienced high incomes while the uneducated could receive income as they were not ideal for employment in firms that relied on technology. The income difference illustrated the importance of education on gaining employment opportunities across the United States. The period of technological boom has been the most beneficial to the educated as they could find employment opportunities to various employment opportunities. However, a setback in the study is that Porter (2014) considered education as the only factor that leads to employment. He failed to consider other factors, such as on-job skills that may not be gained through formal education.

Qazi et al. (2018) explored how an expansion of the educations system would have a direct impact on decreasing income inequalities. Their research was based in Pakistan and assessed various facets of the relationship between education and income. The study utilizes the yearly time-series statistics from 1973 to 2012. According to the study, there exists a long-run relationship between income inequality and higher education. The researchers conducted a cumulative sum that indicated no structural instability in the relationship between education and income levels. Qazi et al. (2018) also found that there is a unidirectional causal relationship between education and higher income. The study's results are more applicable in developing countries where the economic investments are significant that they cannot sustain the whole population. In such economies, the level of education is proportional to the amount an individual earns as processed from employment.

Arshed et al. (2019) assess how higher levels of education relate to income. Arshed et al. (2019) developed a quadratic relationship between income inequality and education levels among developed economies in Asia. The study applied an integration as the method of understanding the correlation between educational level and income. According to the study, an increase in educational investments has decreased income inequality in Asian countries. The researchers urge the government of various countries to increase their investment in both primary and tertiary education to enhance individual skills. Income is derived from undertaking economic roles with positive values (Bloom et al., 2014). Through education skills, an individual can open business opportunities or other ventures and investments with positive values. Therefore, there is a positive relationship between education and income in Asian economies, i.e., higher educational levels result in higher income. However, the relationship can only be true and objective when other factors that affect income are held constant.

There is a positive relationship between education and income. The various researchers explored in the literature review section have emphasized the connection between education and income levels. Most researchers have also illustrated that attainment of higher education levels can solve the income inequality issue that has existed in society for years. Education refers to the skills and knowledge an individual gains from a formal institution. Education is an investment in human capital. In economics, investments in human capital are often undertaken to derive money. College wage premium refers to the higher incomes that can be directly associated with a college degree. Porter et al. (2014) discovered that the capital wage between 1979 and 2012, college wage premium rose by approximately $30000. This figure reflects the amount college degree holders receive over and above those without

that level of education. Between 1979, the levels of technology have been rampantly advanced. Technology has increased employment opportunities and is among the sectors that hold the greatest potential for positive economic growth. However, involvement in the technology sector requires some basic education. Individuals with education have actively played a role and often get direct and indirect income sources from the sector. For the uneducated, the reverse is true. They cannot gain significant income from technology because they are unable to undertake simple tasks.

There is a positive relationship between education and income levels. The higher the education levels in a society, the lower the income inequality gaps in the society. Cody and Dizioli (2018) explored the association between education and income inequality. The researchers found out there exists a relationship between education and income inequality. The study explores education from the perspective of inequality schooling and education expansion, especially in developing and emerging economies. The existence of high-income inequality in different countries has been a concern for economists and policymakers. The researchers attribute the high levels of income inequality to factors such as the advent of skill-biased technology. Such technology has replaced several people who were previously employed (Jerrim & Macmillan, 2015). As a result, several people have been replaced with technology, thereby rendered jobless. However, the educated individuals always find a role to play in technology. For example, programmers are always in demand and are highly paid. The pay can be construed to be income for education acquired by the learned.

Education raises the income levels of ethnic minorities and other marginalized groups. These groups usually face systematic barriers and discrimination when seeking to improve their levels of income. Saad et al. (2012) explored how education has affected income among ethnic minorities in Asian Americans. Educational achievements among Asian Americans can be correlated to their needs to improve the quality of their lives. According to the researchers, education has a direct relationship with income in such a way that higher levels of education are equivalent to higher incomes. Higher education among ethnic minority groups usually elevates their lives by increasing their access to income. Therefore, education improves access to income among the ethnic minorities, therefore, improving their living standards.

When it comes to the statistical challenge introduced by endogeneity, Ebbs, Papies and Heerde (2016) introduce a method to eliminate the impact of endogeneity, they state that IV regression is one of the conventional data methodology to deal with problem incurred by endogeneity. One of the discoveries of this paper is that the endogeneity exists when relevance is high.

After reviewing so much of the literature, they all provide similar results, with a positive correlation between education and income level. Next, we will present a detailed data presentation and regression analysis of the relationship between education and income.

Data gathering

Our data is retrieved from Current Population Survey (CPS) including Age, Sex, Race, Marital Status, Occupation, State, Years, and Salary Income, Hours of Work, Educational Attainment. The unemployment rates were retrieved from BLS Beta Labs, the original data is monthly based, it contains 50 states of the US. It is worth to note that in law, the young people under the age of 18 are considered with not ability to act in integrity and the prejudice in the labor market is serious, which will cause the related bias on subsequent data analysis, thus, we remove the data which are collected from that group, make the samples and the results of the analysis more targeted. We annualized the data by taking the monthly average and fit in the CPS dataset. We specifically look at 3 years ahead of each individual unemployment rate. (See Table 1). For the variable of unemployment rate, we want to make an IV regression to make more accurately estimate the effect on of education level on income and study the extent to which the variable of unemployment rate can explain the variable of education level. We take salary and income as dependent variables and the other variables as independent variables. We will focus on the impact of education on income, and we will use the unemployment rate as our instrumental variable to replace education attainment.

Table 1

Summary statistics for all the variables

	mean	standard deviation	min	max
age	39.9924	11.90931	18	64
wage	48869.76	59119.31	1	1699999
white	.8002024	.3998483	0	1
black	.105154	.3067518	0	1
others race	.0946436	.2927222	0	1
male	.6289304	.4830913	0	1
female	.3710696	.4830913	0	1
occupation (1)	.3397193	.4736141	0	1
occupation (2)	.2759349	.4469844	0	1
occupation (3)	.1548633	.3617745	0	1
occupation (4)	.018672	.1353638	0	1
occupation (5)	.0960816	.2947032	0	1
occupation (6)	.1147289	.3186946	0	1
married	.6289304	.4830913	0	1
unmarried	.3710696	.4830913	0	1
Bachelor’s degree	.3487475	.4765741	0	1
high school degree	.2746821	.4463541	0	1
Less than high school degree	.3765704	.4845259	0	1
unemployment rate	5.924029	2.238044	2.3	13.5
hours work last week	39.56416	12.34889	1	99

N = 1,134,213

Sample selection

Among these variables, key variables are wage and education. Wage and Salary Income variable represents wages, and it collects each personal pre-tax wage and salary income from the previous year. And then Education represents citizens' educational attainment the data is annually distributed, and it is cross-sectional dataset.

The sample includes citizens in the U.S from 2004 to 2020 since our research topic is the relationship between income and education attainment in the U.S. For Education, 2004 to 2020 only include people age over 18. For income, it includes people aging over 18 from 2004 to 2020. Most of the people under 18 years old are in the study stage and have not reached the legal working age, and their income level is mostly 0. Hence, we are not going to study people under the age of 18. For a sample of people older than 65, most of whom are retired, the effect on income and education is small, and again not studied. Therefore, the sample range for age is 18 to 64 years. In this case, the collection of their data in the sample will have a great impact on the overall sample and the subsequent analysis. For the period, we only select data for 15 years period. Data from too long ago will affect the results because changes in the country's overall economy will also affect wages. After eliminating the missing data of each variable, the observations are 1,208,318.

When measuring the income, replacement values are used since in 2004, We converted income to consider inflation. Multiply each of their nominal incomes from 2004 to 2020 by the CPI index of that year to obtain the real income for each year from 2004 to 2020. Privacy of high-income citizens should be preserved. For the educational attainment, we divided it into three categories: those with a bachelor's degree, those with a high school degree, and those with less than a high school degree.(See Table 2&3)

Table 2

	(1) Bachelor's degree	(2) High school degree	(3) Less than school degree
	mean/sd	mean/sd	mean/sd
educ_att_dum1	.3513107 .47738		
educ_att_dum2		.273865 .4459407	
educ_att_dum3			.3748243 .4840777
N	1,208,318	1,208,318	1,208,318

Table 3

Education attainment	(1) Freq.	(2) Percent	(3) Cum.
Bachelor's degree	424,495	35.13	35.13
High school degree	330,916	27.39	62.52
Less than high school degree	452,907	37.48	100.00
Total	1,208,318	100.00	/

As you can see, in Table 3, it is the distribution of education attainment. The total number is 1,208,318. Bachelor's degree accounted for the proportion of 35.13%, and the number is 424,495. The proportion of high school education is 27.39%, and the number is 330,916. Bachelor's degree and high school degree together accounted for 62.52%. This is more than half of the total number. Less than high school degree is 37.48% which is the largest proportion, and the number is 452,907.

In this way, it is convenient for us to summarize these three groups in the subsequent regression analysis. we believe that in addition to the simple level of education, the choice of major and employment field will also have different influences on income. To make a more detailed and powerful analysis and explanation, we also classify the major and employment field and add them into the variables.

For the occupation variable, we divide it into six categories. To help explain the influence of different occupations on income in the subsequent regression analysis. The first category is managerial and professional, and the second category is professional technical, sales and administrative support. The third kind is service kind occupation, the fourth kind is agriculture, forest, fishery kind occupation, the fifth kind is precision production, craft and repair kind occupation, the sixth kind is operator, manufacturer and laborer.(See Table 4)

Table 4

Occupation	Freq.	Percent	Cum.
Management and professional expertise	412,337	34.12	34.12
Technology, sales and administration	332,327	27.50	61.63
Service occupation	186,263	15.42	77.04
Farming, forestry and fishing	22,564	1.87	78.91
Precision production, process, maintenance	115,704	9.58	88.49
Operators, manufacturers and laborers	139,123	11.51	100.00
Total	1,208,318	100.00	

For the variable of marital status, we simply divide it into two categories. One is married, the other is unmarried. In the following analysis, a dummy can be added to analyze whether getting married or unmarried will have an impact on income and to what extent it will have an impact. We will give the answer to this question in the regression analysis.(See Table 5&6)

Table 5

Marital status	Freq.	Percent	Cum.
Married	758,973	62.81	62.81
Unmarried	449,345	37.19	100.00
Total	1,208,318	100.00	

Table 6

My summary statistics

	(1) Married mean/sd	(2) Unmarried mean/sd
Married	.6281236	
	.4833059	
Unmarried		.3718764
		.4833059
N	1208318	1208318

For the variable of gender, we divide it into male and female, and the ratio of male and female in the sample is close to 1:1. Among them, boys accounted for 51.73 percent.(See table 7)

Table 7

Gender	Freq.	Percent	Cum.
Male	625,097	51.73	51.73
Female	583,221	48.27	100.00
Total	1,208,318	100.00	

For the variable of race, we divide it into three categories: white people, black people and other people. Whites account for 79.81 percent, black's 10.73 percent, and the rest 9.47 percent. (See table 8)

Table 8

race	Freq.	Percent	Cum.
White	964,337	79.81	79.81
Black	129,598	10.73	90.53
Other	114,383	9.47	100.00
Total	1,208,318	100.00	

Data summary

Overall, the average age of our sample is age 40, and the variance of age is 12 years. The average annual income of the masses is 48,869 dollars, and the income variance is 59,119, with large fluctuations. The highest income is as high as 1,699,999 dollars. In the sample, there are 412,337 people belonging to management and professionals. The number of people in science, sales and administration is 332,327. There are 186,263 in the service sector. There were 22,564 agricultural jobs. The number of jobs in precision manufacturing and machine production was 115,704. There are 139,123 production workers. Among them, 62.81 percent are married, while the remaining 449,345 are unmarried. Out of a total of 1,1208,318, 35 percent had a bachelor's degree, 27 percent had a high school degree but not a bachelor's degree, and the remaining 37 percent did not even have a high school degree. The average 15-year unemployment rate in the sample was 5.92 percent, with a variance of 2.23 percent. After preprocessing our dataset, we continue the regression analysis as follow.

Correlation coefficient

As we can see, educational attainment record is negatively related to wage and salary income. Age is positive related to the educational attainment record and negative related to wage and salary income. Sex has a weak correlation with other variables. Hours worked last week is negative related to the educational attainment record and age but positive related to the wage and salary income and sex. (See Table 9)

Table 9

	educ	incwage	age	sex	ahrswo~t
educ(educational attainment record)	1.000				
incwage(wage and salary income)	-0.834***	1.000			
age	0.620***	-0.694***	1.000		
sex	0.035***	-0.029***	0.040***	1.000	
ahrsworkt(hours worked last week)	-0.508***	0.492***	-0.195***	0.063***	1.000

Empirical method

To improve the unbiasedness of parameter estimation, we use two stage least square method. It allows us to illustrate the relevance of measuring education. Since the level of education is difficult to define, we use the unemployment rate of each state in the United States from 2004 to 2019 as an instrumental variable to study the impact on income. The unemployment rate is related to the level of education and is not affected by other variables. A higher education level corresponds to a lower unemployment rate.

OLS method

We use log wage instead of wage as the dependent variable. We use the ordinary least square method to analyze these models. We include marital status(marst), race, year, occupation on 1990 basis(occ1990) and state(statefip) these variebles as controls.

Let y represent the logarithm of wage, E represent education, E1 represents the educ_att_dum1 and, E2 represents the educ_att_dum 2, Xahr represents the ahrsworkt, Xage represents age, O1 represents Occ_dum1, O2 represents Occ_dum2, O3 represents Occ_dum3, O4 represents Occ_dum4 and O5 represents Occ_dum5, M represents Mar_dum1, Em represents empyearago_dum1.

Model 1:

$$y = \beta_0 + \beta_1 E_1 + \beta_2 E_2 + \beta_3 X_{ahr} + \beta_4 X_{age} + u$$

In Table 10, the independent variables of our first model, (1)***lwage, are educ_att_dum1, educ_att_dum2, ahrsworkt*** and ***age***. We set ***educ_att_dum1, educ_att_dum2*** these two dummy variables about education. The variable educ_att_dum1 is bachelor's degree. The variable ***educ_att_dum2*** is high school degree and the base group is less than high school degree. The coefficient of ***educ_att_dum1*** is 0.561 which means the wage of bachelor's degree is 56.1% higher than that of less than high school degree. It is significant at 5% level. The coefficient of ***educ_att_dum2*** is -0.039 which means the wage of high school degree is 3.9% lower than that of less than high school degree. It is still significant at 5% level. The variable ahrsworkt is hours worked last week. The coefficient is 0.028 which means when ***ahrsworkt*** increases by one unit, the wage will increase 2.8%. It is also significant at 5% level. The coefficient of age is 0.016 which means when age increases by one unit, the wage will increase 1.6%. In this model, bachelor's degree makes wages have a large positive increase and high school degree make wages a negative increase. Therefore, education is an important independent variable in our research. And age has the smallest effect on wages. The adjusted R square is 0.287. This seems relatively small, so we need to add more independent variables. We use F test to test whether ***educ_att_dum1*** and ***educ_att_dum2*** are jointly significant. The result shows that it is significant.

Model 2:

$$y = \beta_0 + \beta_1 E_1 + \beta_2 E_2 + \beta_3 X_{ahr} + \beta_4 X_{age} + \beta_5 O_1 + \beta_6 O_2 + \beta_7 O_3 + \beta_8 O_4 + \beta_9 O_5 + u$$

In the second model, we added the occupation variable based on first model. We set five

dummy variables about occupation. ***Occ_dum1*** is managerial and professional specialty occupations. ***Occ_dum2*** is technical, sales, and administrative support occupations. ***Occ_dum3*** is service occupations. ***Occ_dum4*** is farming, forestry, and fishing occupations. ***Occ_dum5*** is precision production, craft, and repair occupations. The base group is operators, fabricators, and laborers. The coefficient of ***educ_att_dum1*** changed from 0.561 to 0.411. It is a big change. It has dropped by about 15%, which shows that the coefficient of education has an upward bias in model 1. The inclusion of different occupations in income and education can also affect income. The coefficient of ***educ_att_dum2*** changed from -0.039 to -0.016. ***Ahrsworkt*** and ***age*** becomes 0.026 and 0.015. These two have not changed much. The adjusted R square is 0.323. Compared with model 1, the adjusted R square has improved. This shows that occupation has a certain impact on income. We use F test to test whether ***occ_dum1, occ_dum2, occ_dum3, occ_dum4*** and ***occ_dum5*** are jointly significant. The result shows that it is significant.

Model 3:

$$y = \beta_0 + \beta_1 E_1 + \beta_2 E_2 + \beta_3 X_{ahr} + \beta_4 X_{age} + \beta_5 O_1 + \beta_6 O_2 + \beta_7 O_3 + \beta_8 O_4 + \beta_9 O_5 + \beta_{10} M + u$$

In the third model, we added variable whether to marry based on second model. We set one dummy variable which is ***Mar_dum1***. ***Mar_dum1*** is married. The base group is unmarried. Compared with model 2, the coefficient of ***educ_att_dum2*** and ***ahrsworkt*** has not changed. ***Educ_att_dum1*** and age become 0.399 and 0.012. These two variables have not changed much. Adjusted R square is 0.331. It still has not changed much. Therefore, we believe that whether to get married has a very small impact on income. (See Table 10)

Two-stage least square

Model 4:

$$(1) \quad E = \delta_0 + \delta_1 Z + \beta_3 Xahr + \beta_4 Xage + \beta_5 X_{male} + \varepsilon$$

$$(2)\, y = \beta_0 + \beta_1 E + \beta_2 X_{ahr} + \beta_3 X_{age} + \beta_4 X_{male} + (u + \beta 1\varepsilon)$$

Let Z represents the unemployment rate of each state in the United States from 1973 to 2019.

We still include marital status(marst), race, year, occupation on 1990 basis(occ1990) and state(statefip) these variebles as controls. For the first stage regression, we regress educational attainment record on unemployment rate, age, gender of male and hours worked last week by using OLS. At this stage we can isolate the exogenous part of educational attainment record. The result shows that unemployment rate and educational attainment record are strongly negatively correlated. For the second stage regression, we use the fitted value of educational attainment record instead of the endogenous variable educational attainment record. So, we regress the lwage on fitted value of educational attainment record, age, gender of male and hours worked last week by using OLS. All variebles are significant at 5% level. The adjusted R square is 0.3866. Compared with the adjusted R square of OLS, it is significantly higher (See Table 11). The F test is also significant. Education does affect wage according to the analysis.

Table 10

	Model 1 lwage	Model 2 lwage	Model 3 lwage
Bachelor's degree	0.562***	0.411***	0.399***
	(0.002)	(0.002)	(0.002)
High school degree	-0.039***	-0.016***	-0.016***
	(0.002)	(0.002)	(0.002)
hours worked last week	0.028***	0.026***	0.026***
	(0.000)	(0.000)	(0.000)
age	0.016***	0.015***	0.012***
	(0.000)	(0.000)	(0.000)
managerial and professional specialty		0.312***	0.302***
		(0.003)	(0.003)
technical, sales, and administrative support occupations		0.065***	0.066***
		(0.003)	(0.003)
service occupations		-0.252***	-0.241***
		(0.003)	(0.003)
farming, forestry, and fishing occupations		-0.263***	-0.271***
		(0.006)	(0.006)
precision		0.263***	0.249***

production, craft, and repair occupations			
		(0.003)	(0.003)
Married			0.191***
			(0.002)
Constant	8.210***	8.251***	8.240***
	(0.003)	(0.004)	(0.004)
N	1208318	1208318	1208318
R^2	0.287	0.323	0.332

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11	first	second
VARIABLES	educ	Real wage
Educ		
(educational attainment record)		-0.000***
		(-2.91)
UnERate		
(unemployment rate)	7608.403***	
	(2526.441)	
age	551.6658***	0.090
	(3.905194)	(0.34)
male	11551.35***	0.064
	(93.63005)	(0.08)
Ahrsworkt		
(hours worked last week)	774.5996***	-0.001***
	(13.24941)	(-3.87)
Observations	2,949,274	2,949,274
r2_a	.	0.3866

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Conclusion

To conclude, research is conducted on the educational impact on income level. Therefore, income and salary wage are chosen to be dependent variable and educational attainment is chosen to be independent variable. Except two key variables, age, sex, race and job are chosen to be controlled variables. Except the education level, working ability also be consider as an important independent variable, but the capacity is a difficult to directly measure and quantify the indexes, we chose the unemployment to instead, although the degree of education will cover part of the corresponding work ability, but the learning ability can't completely represent the working ability, level of education may be the premise of a job, the unemployment rate is better able to show whether a person can suit to a job for a long time, so we chose the unemployment rate as an instrumental variables to process IV regression for comparison and find the optimal model. The samples were successively classified into 2 categories of gender, 6 categories of different occupations, 3 categories of groups with different educational levels, and 2 categories of marital status were married and unmarried respectively. From these models, you can see the effect of education level on income. In the scope of our analysis, the income of bachelor’s degree is the highest when other variables are controlled unchanged. After Instrumental Variable, education still being the most crucial factor on income level.

References

Beckles, G. L., & Truman, B. I. (2011). Education and income—the United States, 2005 and 2009. *MMWR Surveill Summ*, *60*(Suppl. l), 13-17.

Shu, X. (2010). Technical Progress Return to Education and Income Inequality in China [J]. *Economic Research Journal*, *9*.

Bennett, D. L., & Vedder, R. K. (2015). Public policy, higher education, and income inequality in the US: Have we reached diminishing returns. *Higher Education, and Income Inequality in the US: Have We Reached Diminishing Returns (June 16, 2014). Social Philosophy and Policy*, *31*(2), 252-280.

Keller, K. R. (2010). How can education policy improve income distribution? An empirical analysis of education stages and measures on income inequality. *The Journal of Developing Areas*, 51-77.

Bloom, D. E., Canning, D., Chan, K. J., & Luca, D. L. (2014). Higher education and economic growth in Africa. *International Journal of African Higher Education*, *1*(1), 22-57.

Goldrick-Rab, S., Kelchen, R., Harris, D. N., & Benson, J. (2016). Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion. *American Journal of Sociology*, *121*(6), 1762-1817.

Gregorio, J. D., & Lee, J. W. (2002). Education and income inequality: new evidence from cross-country data. *Review of income and wealth*, *48*(3), 395-416.

Jerrim, J., & Macmillan, L. (2015). Income inequality, intergenerational mobility, and the Great Gatsby Curve: Is education the key?. *Social Forces*, *94*(2), 505-533.

Arshed, N., Anwar, A., Hassan, M. S., & Bukhari, S. (2019). Education stock and its implication for income inequality: The case of Asian economies. *Review of Development Economics*, *23*(2), 1050-1066. https://doi.org/10.1111/rode.12585

Qazi, W., Raza, S. A., Jawaid, S. T., & Karim, M. Z. A. (2018). Does expanding higher education reduce income inequality in emerging economy? Evidence from Pakistan. *Studies in Higher Education*, *43*(2), 338-358. https://doi.org/10.1080/03075079.2016.1172305

Porter, E. (2014). A simple equation: More education= more income. *The New York Times*.

Saad, C. S., Sue, S., Zane, N., & Cho, Y. I. (2012). The relationship between education and ethnic minority factors in income among Asian Americans. *Asian American Journal of Psychology*, *3*(2), 66. https://psycnet.apa.org/doi/10.1037/a0026867

Coady, D., & Dizioli, A. (2018). Income inequality and education revisited: persistence, endogeneity and heterogeneity. *Applied Economics*, *50*(25), 2747-2761. https://doi.org/10.1080/00036846.2017.1406659

Campos, B. C., Ren, Y., & Petrick, M. (2016). The impact of education on income inequality between ethnic minorities and Han in China. *China economic review*, *41*, 253-267. https://doi.org/10.21315/ijaps2018.14.2.3

Yang, J., & Qiu, M. (2016). The impact of education on income inequality and intergenerational mobility. *China Economic Review*, *37*, 110-125. https://doi.org/10.1177%2F0973184915603171

Abdullah, A., Doucouliagos, H., & Manning, E. (2015). Does education reduce income inequality? A meta-regression analysis. *Journal of Economic Surveys*, *29*(2), 301-316.

https://doi.org/10.1111/joes.12056

Ebbs, P., Harald,H., and Papies, D.(2016). Dealing with Endogeneity: A Nontechnical Guide for Marketing Researchers. *Research Gate,*2-34.DOI:10.1007/978-3-319-05542-8_8-1