Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/62/20231346

The Relationship Between Education Level and Wages: Based on Regression Model Analysis

Hao Zhou[1,a,*]

[1]*Bachelor of Business Administration, Baylor University, 1311 S 5th St, Waco, TX, United States*
a. Hao_Zhou2@Baylor.edu
**corresponding author*

Abstract: This paper focused on finding how a higher level of education affects people's wages in China. Based on a popular social phenomenon that citizens in China are more likely to graduate from college, this study was focused on seeking the relationship between higher levels of education and wages. With the assumption that one of the elements that gives people the incentive to receive a higher level of education. This study used the data set from the China Family Panel Studies (CFPS) in the 2020 individual survey. This research uses different approaches to conduct the variables that could affect wages and uses the multiple linear regression model to estimate the effect. The result indicates that the higher education degree people achieved in China should have a strong positive coefficient on their future wages. And the subgroup of their future confidence and their authority in the working field also demonstrate a positive relationship with wages.

Keywords: Higher level education, wage, multiple linear regression

1. Introduction

Education expenditure is always a visible contribution of the Chinese government years expenditure report. However, even though this percentage of education expenditure has gradually increased over decades, the outcomes did not indicate that the Chinese government will have more highly educated people. Yet, numerically, about 30 million students in China were seeking bachelor's degrees in 2019, while about 2.5 million were studying for master's degrees, and 420,000 were pursuing doctorate degrees, according to the country's Ministry of Education [1]. Short-cycle tertiary credentials make up 10% of the population of 25–64-year-olds in China, followed by bachelor's degrees at 8% and master's and doctorate degrees combined at 1%. This contrasts with the OECD average, which places bachelor's degrees (19%), master's degrees (14%), and short-cycle tertiary certificates (7%), as the most prevalent educational attainments [2]. So, knowing how individuals decide whether to achieve a higher education degree would help the government know how to produce the incentive for citizens to achieve a higher level of education that provides more talented people. Generally, people take action to seek for the well-being. The initial indicator, income, shows that when wage increases, well-being also increases but at a diminishing pace; other factors that might affect well-being include education, empowerment, and social norms. The desire for greater wealth also affects well-being. stronger wages are correlated with stronger goals and ambitions, but worse well-being. A person's social standing rises as their money increases [3]. Knowing whether higher education affects

DOI: 10.54254/2754-1169/62/20231346

someone's income would contribute to people's well-being or not should provide the government a direction to incentivize people to decide to receive a higher education.

There are lots of researchers focused on education in China, from both Marco and Mirco's perspectives. A significant development in Chinese higher education in the late 1990s and early 2000s was the enrollment's dramatic growth beginning in 1998 [4]. Reality does not support the anticipated short-term effects of enrolment growth on the Chinese economy. Expanding enrollment has also increased pressure on Chinese higher education to further change its administration, structure, and curriculum. More significantly, the increase in enrolment has highlighted the issue of equality [4]. Wage and education are highly correlated. Education is frequently referred to as "an investment in human capital." In the same way, that people invest in financial assets to generate income, they also do the same with human capital. Generally speaking, those with higher education earn more money [5]. However, one of this research is been done far age that losing the external representative or not applicable to Chinese. This research focuses on providing a more recent analytical correlation between individuals' education and income. That should fill the research gap for knowing the recent individual's education and income coefficient.

2. Data Description

The data that this research chose is the 2020-person category from the CFPS data set. The China Family Panel Studies (CFPS), a nationally representative biannual continuous survey of Chinese societies, households, and individuals was first conducted in 2010 by the Institute of Social Science Survey (ISSS) at Peking University in China. The CFPS is designed to collect longitudinal data in contemporary China along every stage of the person, family, and society [6]. The dependent variable this research conducts is the total income of individuals in the past 12 months. The key variables have been divided into 4 parts, first is the education level. This paper segregates the highest education level that once has been served. The rest are the dummy variables of whether someone archives a Bachelor's degree, Master's degree, or Doctorate. 1,347 observations have been dropped due to the unprovided information. According to table 1, all the potential variables for the four models this research is going to use have an observation number greater than 9000. Based on those four models' degree of freedom, more than 9000 observations is a large enough sample size for the model to be consistent and unbiased.

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Wage	9315	44523.429	43229.625	-9	700000
EduLevel	9315	3.469	2.082	-8	6
Bachelor	9315	.142	.349	0	1
Master	9315	.013	.113	0	1
FConfidence	9241	3.772	1.466	-8	5
Authority	9207	40.586	14.373	18	83
age	9315	39.111	12.171	15	83
Emp	9241	6.205	2.794	-8	10
status	9241	2.78	1.44	-8	5

The dependent variable chosen is Wage, which represents the total income individuals can get in the past 12 months. The EduLevel represents the artificially defined education level that is below the bachelor's degree. For those who do not receive any education, this study defined their education level as 0. As long as they receive a primary school education, their education level begins from 3 and increases with a value of 1. The bachelor and the master represent the dummy variable this research

DOI: 10.54254/2754-1169/62/20231346

reproduced whether someone is achieving that degree or not. The FConfidence represents the person's confidence in their future, the more optimistic the survey taker is about his future, the higher score he will have. The Authority represents people's expertise in their area so that others might value it more. Age represents how old the survey taker is. Usually, people interpret a positive coefficient between age and Wage. Emp is an index that people evaluate the problem of employing. Status represents the social status people believe they are located in.

The purpose of this research is to figure out how a higher level of education can people's income. This study uses Edu Level, Bachelor's, and Master's. Since the doctor is the exclusive group, it does not need to pull the doctor as a variable out separately. By comparing the difference between people with a lower education level, it can know how much their wage can be influenced by their degree. This paper also uses other variables to construct our model. Recent research supports that the employee's optimistic outlook has an impact on his or her pay both directly and indirectly through its impacts on happiness. The research provides proof that an employee's joy affects their pay in both direct and indirect ways. It does so by assuming that the optimistic mindset variable is endogenous and by simultaneously computing happiness and attitude equations using a two-step method [7].

The first elemental component in the analysis is the future confidence of the server taker. It is reasonable to interpret that people with high confidence should be more optimistic in their motivation to find an increasing trend in their salary. It can be inferred that they are already in a position in the increasing trend of their salary. The more optimistic a worker is, the increasing likelihood he or she will complete a task with higher quality that benefits them by increasing their salary. On the other way, if people always show their negative thoughts, that would not only influence them but also their coworkers in the result of deduction of their salary. A recent study has suggested this opinion, at work, employees with higher authority often exhibit greater flexibility than those with lower authority. According to the notion of compensating differentials, employees with flexible schedules will make less money than other employees [8].

The second elemental component in the analysis is the individual's authority in his working field. This research used the mean of two different measurement methods—ISEI and SIPOS. How others value someone's job skill can be appealed by the authority scores. Typically, the more working authority someone has in his working field, and if they can persuade the employer, this paper assumes he will also have a higher salary. This should also indicate a positive relationship between the authority and the salary.

3. Econometric analyses

3.1. Regression analysis

This paper sets up different multiple linear regressions, that help this research to find a better model that demonstrates the coefficient between higher-level education and the income someone might had in China. The different regressions are:

1. $Wage = \beta 0 + \beta 1 \times EduLevel + \beta 2 \times Bachelor + \beta 3 \times Master + u$
2. $Wage = \beta 0 + \beta 1 \times EduLevel + \beta 2 \times Bachelor + \beta 3 \times Master + \beta 4 \times FConfidence + u$
3. $Wage = \beta 0 + \beta 1 \times EduLevel + \beta 2 \times Bachelor + \beta 3 \times Master + \beta 4 \times FConfidence + \beta 5 \times Authority + u$
4. $Wage = \beta 0 + \beta 1 \times EduLevel + \beta 2 \times Bachelor + \beta 3 * Maste + \beta 4 \times FConfidence + \beta 5 \times Authority + \beta 6 \times age + \beta 7 \times status + u$

The first regression focused on the education level and the wages people could get which is the simplest regression. The second regression adds the coefficient of Future Confidence which can be used to compare with the first regression to check whether it is reasonable for us to add the first

DOI: 10.54254/2754-1169/62/20231346

elementary variable to our model which helps us predict the coefficient better. The third regression uses a similar thinking process by adding our second elemental variable Authority to our model, to see if it is the best fit. The last regression is adding other variables to the model will increase the efficiency of our model.

Table 2: Coefficients of Different Variables for Four Models

	(1).	(2).	(3)	(4)
Constant	17745.065(0)	11421.387(0)	601.537(.743)	2738.029(.309)
EduLevel	4916.098(0)	5191.804(0)	3586.803(0)	4078.861(0)
Bachelor	58971.236(0)	60027.212(0)	44855.404(0)	47897.387(0)
Master	102483.89(0)	103428.99(0)	86195.469(0)	89391.927(0)
FConfidence		1384.698(0)	1296.525(0)	1430.015(0)
Authority			470.994(0)	471.149(0)
Age			0.33(0)	104.281(.005)
Employ issue rate				-769.668(0)
Social Status				484.478(.235)
R-squared	0.1659	0.168	0.185	0.1876
Number of Observation	9315	9241	9207	9207
Prob > F	0.0000	0.0000	0.0000	0.0000
Root MSE	39488	39466	38999	38932

According to Table 2, the column's number indicates the number of our regression. It appears that all the coefficients for our 4 regressions have shown statistical significance with their p-values less than 0.01 except for the Social Status in the fourth regression. The r-square did not appear to have a significant increase. The RMSE also did not appear a dramatic decrease from regression three to regression four. However, from regression 2 to regression three, it can be seen that the r-square increased by a noticeable amount. The RMSE also has a similar approach, it is experiencing a noticeable decrease. With all the conditions that hold all the variables' coefficients significant, and regression three has a better r-square and RMSE, this study would like to interpret that regression three is the best fit for the model.

3.2. Omit variables examination

The lack of a variable that belongs in the model because it is important to the connection you are attempting to understand is known as omitted variable bias. This variable that has been left out could have an impact on the dependent variable as well as any of the independent variables that were provided. To test whether omit variable bias still exists This paper uses the third regression, and would like to run an omit variable test. Then this study used the Ramsey RESET test for omitted variables.

Omitted: Powers of fitted values of Wage

H0: Model has no omitted variables

F(3, 9308) = 25.23

Prob > F = 0.0000

The P-value shows it can reject the null hypothesis that this paper does have the omitted variable bias that can influence our model. However, as Table 2 analyzed, the relevant variables adding to our regression did not appear a huge amount of support for the model. So, this study assumes that omitting variable bias did not influence our model by a significant amount.

DOI: 10.54254/2754-1169/62/20231346

3.3. Multicollinearity Examination

The regression analysis might face issues as a result of multicollinearity. The particular impact of each independent variable on the dependent variable may be difficult to evaluate. Predictions of the coefficients may become inaccurate due to high multicollinearity. So, this paper would like to run a matrix function.

Table 3: Matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) EduLevel	1.000				
(2) Bachelor	-0.698	1.000			
(3) Master	-0.197	-0.047	1.000		
(4) FConfidence	-0.008	0.017	0.009	1.000	
(5) Authority	-0.116	0.424	0.158	0.025	1.000

According to Table 3, it can be seen that most of our variable is weakly correlated since their absolute value is lower than 0.3. The only median negative correlation was between EduLevel and Bachelor. However, since the Bachelor is a dummy variable that can be easily correlated with other variables, this study would like to assume that covariance will not affect our model's accuracy on a large scale.

3.4. Heteroskedasticity examination

Testing hypotheses incorrectly can be driven by heteroskedasticity. T-tests and F-tests for the significance of variables can generate incorrect outcomes because the standard errors of the coefficient estimates are biased. Variables can appear to be unnecessary when they are significant, and conversely. The regression model's predictions may also be impacted by heteroskedasticity. The model's predictions are less reliable since the prediction intervals might be too narrow in some locations and too broad in others. So, this paper would like to run a heteroskedasticity test using the Breusch-Pagan/Cook-Weisberg test for the heteroskedasticity method.

Assumption: Normal error terms

Variable: Fitted values of Wage

H0: Constant variance

chi2(1) = 3391.30

Prob > chi2 = 0.0000

The P-value here is 0.0000 which is smaller than the alpha, thus, this research rejects the null hypothesis. So, the author needs to use the robust method. The result is in Table 4.

Table 4: Robust Result

Wage	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
EduLevel	5191.804	252.219	20.58	0	4697.4	5686.208	***
Bachelor	60027.212	1937.954	30.97	0	56228.395	63826.029	***
Master	103428.99	9203.33	11.24	0	85388.431	121469.55	***
FConfidence	1384.698	223.916	6.18	0	945.773	1823.624	***
Constant	11421.387	1330.95	8.58	0	8812.43	14030.343	***
Mean dependent var		44727.091		SD dependent var		43262.925	
R-squared		0.168		Number of obs		9241	
F-test		275.995		Prob > F		0.000	
Akaike crit. (AIC)		221828.264		Bayesian crit. (BIC)		221863.921	

*** p<.01, ** p<.05, * p<.1

DOI: 10.54254/2754-1169/62/20231346

3.5. Endogeneity examination

One or more predictor variables and the unobserved elements that have an impact on the dependent variable are said to be endogenous. Endogeneity can provide biased coefficient estimates, which makes it difficult to infer solid causal relationships from the model. So, this study would like to run an Endogeneity test. This research chooses the age as the IV.

Tests of endogeneity

H0: Variables are exogenous

Durbin (score) chi2(1) =11.0228 (p = 0.0009)

Wu-Hausman F(1,9200) = 11.0276 (p = 0.0009)

As shown in Table 5, the p-value=0.0009<0.05=alpha. This research should reject the null hypothesis, to settle the endogeneity, this study will use the 2SLS approach.

Table 5: Instrumental variables 2SLS regression

Wage	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
EduLevel	567.963	967.089	0.59	.557	-1327.497	2463.423	
Bachelor	30249.994	4855.598	6.23	0	20733.197	39766.79	***
Master	71014.212	6087.405	11.67	0	59083.118	82945.306	***
FConfidence	1302.915	278.345	4.68	0	757.368	1848.462	***
Authority	590.812	49.97	11.82	0	492.872	688.752	***
Constant	8534.22	3027.088	2.82	.005	2601.236	14467.204	***
Mean dependent var		44668.551		SD dependent var		43175.556	
R-squared		0.176		Number of obs		9207	
Chi-square		1935.188		Prob > chi2		0.000	

*** p<.01, ** p<.05, * p<.1

4. Discussing

Intuitively, someone might think that receiving a higher education is not as important as it looks like in future jobs, and that is due to the early entry of the lower education receiver can getting more experience of living in the society that helps them success. The success of various examples like Steve Jobs gives those types of people the imagination that it is not important if someone is achieving a higher level of education. But that concept could be led by the survival processing effect. The survival processing effect, a memorial advantage, describes how individuals primarily recall information that is processed for its survival importance. Information connected with the possibility of reward is likewise skewed in memory [9]. what they recognized first was those examples who a huge success after they dropped out of school. And they ignored those who did not get success after they dropped out. At the same time dropping out could be an irrational behavior from the student's perspective.

To evaluate this concept this research, can find that holding all other factors the same, the higher education degree someone achieved in China should provide him or her a dramatic positive influence on their future wage. So, this paper would like to say the view held by those groups of people is led by the survival processing bias.

Some people have thought that it is not how education affects someone’s future wage, it is its family wealth condition that determines that person’s education also provides that person a better opportunity to get a higher salary job. There is research that appears similar to this concept, but the result indicates was not support this view very well. Some evidence links wealthy parents with a child's probability of landing employment and earning a living. However, neither the job market

DOI: 10.54254/2754-1169/62/20231346

results nor the job market output indicate an obvious distribution of family wealth, and this impact is greatly mitigated by the educational opportunities for the kids [10].

For the above type of view, this research was also dedicated to rejecting if this study assumes the social status in our regression four is similar to the family wealth condition. The coefficient of social status even in the fourth regression is similar to the coefficient of authority which is significantly lower than the education coefficient.

5. Conclusion

The contribution of the characters that affect someone's future wage is majority influenced by the education factor. And other variables also contribute positive correlations with the wage. After the testing of all those potential biases, this paper could conclude that the higher level of education influenced the wage the most significantly. Since the robust regression method will not affect the coefficient, it will find that for each level of lower education, the education level below the university education, increasing will increase that person's future wage by 567.089 if holding all other factors constant. If someone achieves his bachelor's degree, it is likely his wage will increase by 30249.994 holding every other factor constant. If someone has achieved a master's degree, his wage is likely to increase by 71014.212 compared to those who have just finished the lower level of education holding all other factors constant. The other two factors in this regression do not appear the same significant coefficient on someone's future wage. Holding all other factors constant, for each 1 score someone's authority is, the wage will increase by 590.812. And for future confidence in their life, under the same condition, for each 1 confidence level increase, the wage will respond to increase by 1302.915. So, this paper can say that people should go get a higher level of education if he or she want to get a higher wage in the future China.

References

[1] *BOFIT. "Chinese Increasingly Likely to Attend University, Nearly All Young Adults Are Literate." BOFIT, www.bofit.fi/en/monitoring/weekly/2021/vw202120_3*

[2] *Education GPS – China - Overview of the Education System (EAG 2023), gpseducation.oecd.org/CountryProfile?primaryCountry=CHN&treshold=10&topic=EO. Accessed 17 Sept. 2023.*

[3] *Liu, S.J., and Almas, H. (2022) "Relationship Between Education and Well-being in China." Journal of Social and Economic Development, 25(1), 123–51.*

[4] *Wan, Y.M., (2006) "Expansion of Chinese Higher Education Since 1998: Its Causes and Outcomes." Asia Pacific Education Review, 7(1), 19–32.*

[5] *Scott, A. Wolla and Jessica, S. (2017) "Education, Income, and Wealth," Page One Economics.*

[6] *CFPS. "Introduction." , www.isss.pku.edu.cn/cfps/en. Accessed 3 Sept. 2023.*

[7] *Mohanty, and Madhu, S. (2009) "Effects of Positive Attitude on Happiness and Wage: Evidence From the US Data." Journal of Economic Psychology, 30(6), 884–97.*

[8] *McCrate and Elaine. (2005) "Flexible Hours, Workplace Authority, and Compensating Wage Differentials in the US." Feminist Economics, 11(1), 11–39.*

[9] *Forester, Glen, et al. (2020) "Adaptive Memory: Independent Effects of Survival Processing and Reward Motivation on Memory." Frontiers.*

[10] *Karagiannaki and Eleni. (2017) "The Effect of Parental Wealth on Children's Outcomes in Early Adulthood - the Journal of Economic Inequality." SpringerLink.*