ECO375 Project

Instructor: Courtney Ward

Authors: Huilin Xie huilin.xie@mail.utoronto.ca

Yichun Zhang kimberlyzyc.zhang@mail.utoronto.ca

November 23, 2022

Abstract

The causal relationship between education and earnings has long been a parameter of interest to governments, as well as to people themselves and labour economists. Existing literature put heavy weight on studying the relationship between education and wages, from different perspectives in different countries. David Card (1999), surveyed literature on causal relationship between education and earnings, and concluded that the average return to education, consistent with other literature, has a causal effect. We mainly focus on the relationship between monthly earnings and the length of education in our study. It fills the gap of the original research for our data source which only paid attention to the unobserved ability as well as interindustry wage differentials and overlooked other important factors influencing wages. We also investigate difference of the education effect on monthly earnings between different groups. In conclusion, we find a sizable association between monthly earnings and years of education, no matter controlling for other factors or not, and the association varies between different demographic groups.

Context & Data

The data in this analysis were from M. Blackburn and D. Neumark (1992)'s paper and follows a cross-sectional structure. The data covers 935 individuals in the US in 1980 and measured 17 variables of study, including wage, years of education, years of working experience, IQ score and KWW (knowledge of world work) score, as well as dummy variables like married/unmarried, urban/non-urban, etc. We use years of education as a proxy of education level, which is a commonly used method, so the treatment variable is years of education, shortened as education in this analysis. The outcome is the monthly wage of selected individual in dollars, shortened as wage in this paper. A summary of statistics is shown in ***Table 1.*** Years of education ranges from 9 years to 18 years, with a mean of 13.47 years. Wage ranges from 115 to 3078 dollars per month, meaning that the data includes both high-income and low-income individuals. The average wage of sample is 957.95, and the standard deviation is 404.36.

In the original data set, there are some strange observations which have years of experience larger than ages and years with current employer larger than total years of work experience. We deleted all the observations with the noticeable inconsistency. After cleaning the data, the average wage is 932.71, and the standard deviation is 404.62. Considering the mean and standard deviation of wage, the variability in wage is sizable. The cleaned data's average length of education is 13.22 years with a standard deviation of 2.10, suggesting that there is some variations in length of education, but its distribution of sample's length of education is more clustered than wage. We drew a scatterplot that shows the relationship between education and wages as in ***Figure 1***, and the trend seems linear and positive .

Regression Analysis

Using an ln(wage) is usually preferred by researchers when studying the effect on wage. Nonetheless, we found a decrease in the correlation between monthly earnings and education level if we do a log transformation to the wage, so we keep using wage as the main outcome instead of the ln(wage).

Baseline

To study the relationship between years of education and monthly wage, we first used simple linear regression specification, ***(Table 3, specification (1))***. Our model of assumption is: wage = $\beta_0 + \beta_1\ educ + \epsilon_i$, where wage is the monthly earnings of employees in the US in 1980 measured in dollars, and educ is the length of education in years. In this linear model, β_0 is the intercept, which is mean wage earned when the person receives no education, but it should not have a reasonable interpretation since it is out of the data range given the minimum education is 9 years in the sample. β_1 is the coefficient (the marginal increase in wage given one more year of education), and ϵ_i is the error. In order to deal with heteroskedasticity, we use robust standard error. Our null hypothesis is $H_0: \beta_1 = 0$, and the alternative hypothesis is $H_1: \beta_1 \neq 0$. The regression results are in ***Table 3.*** The results show that estimated slope is 63.97, suggesting that comparing two individuals with different years of education, the person who received one more year of education, on average, earns about 63.97 more dollars per month. The standard deviation of years of education in the data is 2.10, so a one standard deviation change in the length of education is associated with about 134.34 dollars difference in the monthly wages. Meanwhile, the standard deviation in monthly wages is 932.71, so a 1 standard deviation difference in education is associated with 14.4% (about one seventh) of a standard deviation in monthly wages. This standardized change allows us to see a large increase in wage with increase in education. The estimated slope is sizable in terms of the variation in education.

The large t-statistics of 8.76 leads us to reject the null hypothesis ($H_0: \beta_1 = 0$), that monthly earning is unassociated with the length of education for an individual. This is a highly statistically significant result even at 1% significance level. We expect the slope of education in other samples lies between 49.64 and 78.30 dollars for 95 percent of the time (with one more year of education, a person's monthly earnings are found to increase between 49.64 to 78.30 dollars). Simply, more years of education is associated with higher wages. This finding is economically significant, and the results are consistent with economic literatures. The standard deviation of slope is 7.3, which is relatively large, representing a sizable variation in education effect.

We continue to use robust standard errors for the rest regressions. The result, nevertheless, might also be influenced by other factors. The main findings of interest may be distorted if additional factors are not controlled. The uncontrolled variables may be responsible for the changes in the outcomes rather than our treatment variables. We think other variables that might correlates with education. IQ and KWW scores are related to intellectual capability and individual's motivation respectively, which may contribute to a longer education; as age goes up, people have more time to be educated; years of employment and years with current employer also relate to age and education. Marriage status, living in cities and being black also associated with different education level, as stated in many literatures. Therefore, we control for these variables in our study to deal with the endogeneity problem. We found that the correlation between tenure1 (the measure of years working with the current employer) and exper1 (the measure of total years of work experience) has a high correlation of 0.44, so we first consider adding each of them as a controlling variable to get more statistically significant result and then including them together to reduce more of confounding variable issue. By running a regression analysis with adding tenure1 as control variable but not exper1 (***Table 3 specification (2)***), the slope estimator on length of education becomes 44.79, which is lower than ***specification (1)*** but still large. The estimator is still statistically significant as the P value remains small.

We then added exper1 but not tenure1 for controlling variables (***Table3 specification (3)***), and the coefficient increased slightly to 51.96. The P value remains close to 0 so the results are statistically significant. The t-statistics is 6.04, which is also large enough to reject the null hypothesis.

Now with both tenure1 and exper1 added for controlling ***(Table 4 Specification (4))***. As suggested by Card (1999), people with higher education levels are more likely to work more hours per week, which may lead to a bias captured by the slope estimator of education. Thus, we include the average weekly working hour as a control variable to address this issue. After controlling for those other factors, the slope estimator of education became 50.77, which is also highly statistically significant and is a noticeable association economically speaking. The t-statistics is 5.93, the P-value remained near 0, and we are 95 percent confident that the slope of education in other samples is between 33.98 and 67.57.

Specification (1)-(4)'s F values are 76.77, 31.35, 31.29 and 26.65, respectively. The value of F decreases since not all control variables are highly correlated with the monthly earnings. But adding them is appropriate in order to reduce confounding bias. All these F values are larger than critical value, meaning all models are statistically significant. The standard deviation of slope, remain roughly unchanged. Looking at R-Square, the four specification's R-Square are 0.11, 0.24, 0.236 and 0.25 respectively. R-Squared is a statistical measure of fit. Higher R square indicates more variation of dependent variable is explained by treatment variable. The four specifications all suggest an important education effect on wage.

Extension

After controlling for many other factors such as IQ, KWW scores, work experience and some demographic characteristics, the effect of longer education went down to some degree, but the association between monthly earnings and length of education is still significant. This is consistent with intuition and many existing literatures about the relationship between wage and education. For instance, Morgan and David found a 4 to 6 percent increase in hourly earnings as the effect of one unit of higher education after controlling for other factors, using the 1960 data in the US. For the extension model, we add interaction terms to see the association of education and earnings within different subgroups.

First, we separate the education-ethnic effect by including the variable *bla_educ*, which is generated by the dummy variable of being black times the years of education. The model of assumption for extended model is $Wage = \beta_0 + \beta_1 educ + \beta_2 black + \beta_3(educ * balck) + \cdots + \beta_k X_k + \epsilon_i$, where β_3 is the increment effect to the effect of education, w hen $black = 1$, and "$\ldots + \beta_k X_k$" are the controlled variables and their coefficients. After this, the main education effect (for people not being black) ***(Table 4 Specification (2))*** does not change much, with a slope of 54.57 ($\beta_1 = 54.57$), and the resulting model's β_2 is 449.1 and β_3 is -47.10, holding constant IQ, KWW score, years with current employers, total work experience, age, average weekly working hours, and among people with same marriage status and living in cities/noncities status. This means an individual not being black who received one more year of education on average earns 54.57 dollars higher monthly earnings after controlling for the factors mentioned above. The slope estimator on education is both highly statistically significant (with a t-test statistic of 6.14) and economically significant.

However, the slope on education after controlling for other factors is much lower for black people***, (Table 4 Specification (3)).*** The estimator of 7.47 means holding other factors constant, a black person with one more year of education has only 7.47 dollars higher monthly wages on average. Although it is statistically significant for the group difference at a 1% significance level, it is not a big effect of education on income for this group economically speaking. We can see a large difference in the education effect within groups of different ethnic characteristics. The F-statistic for this specification is 24.62, not very different from that of previous regression results, indicating this model still fits the data well overall and the overall result does not change much. Our finding is consistent with the finding of Assari (2018) that the return of education is relatively lower for black people compared to white. This difference in education effect indicates black people may be unfairly treated in the job market, which should a be concern of policy making.

In addition, we generate 5 new dummy variables measuring the order of birth using variable *brthord.* The order of birth varied from 1 to 10 in the data set, with a majority of observations born in the first (254 observations) or second (190 observations). 139 individuals in the data set were born in the fifth or later. The dummy variables *br1, br2, br3, br4, and br5_* equals 1 when the person is birth in first, second, third, fourth, and fifth or more respectively. We choose being born first as the omitted category, so we derive and include interactions of the other four dummies with education respectively. The base variable used to create the interaction terms should have to be included as well. The results are shown in ***Table 4 Specification (4)***. Holding the other factors constant (same as in the previous part except for the birth order) the person born as the first child of the family on average earns 65.82 dollars higher if he/she receives one more year of education. This is still a significant result both statistically and economically. The standard error becomes even larger for this model, resulting in a wider 95% confidence interval between 42.29 and 89.36 dollars.

There are large differences in education effect on monthly wages between groups with different birth orders. Although most of the differences are not statistically significant, this might be the result of not big enough sample size and need further investigation of this interactive effect. This may also shed light on policy designing. People born later, especially in the forth order in the family may need extra support for their jobs. With one more year of education, people born in the second on average have a 27.86-dollar lower increase in monthly earnings than people born in the first. This is statistically significant only at a 10% significance level but not a 5% significance level. Besides, people born in the fourth on average have a 55.24-dollar lower increase in monthly earnings associated with one more year of education than people born in the first. This is a very big change and statistically significant at a 5% significance level. However, we can still conclude that a significant increase in monthly earnings is associated with education in general. The F test statistic is 12.3, only half of the previous F. The performance of this model in fitting the data is worse than previous models, but it is still a large F, and it is appropriate to use this model as an explanation of the education effect on income.

Limitations

We use the years of education as an indicator of education level in our study, but the quality of education may matter as well. When analyzing the interactive effects of birth order, the small sample size of groups with different birth orders can result in a large standard error and thus statistically insignificant differences. Besides, there may be other confounding variables that can lead to bias in the estimated effect of education. For instance, some demographic characteristics such as the number of siblings and parental education are not controlled in our research since the many missing values and thus small sample size may weaken the statistical significance. There is also a chance that the observations missing this information are different in nature, so deleting the missing values may lead to a selection bias in this case. Finally, since we use data from more than 40 years ago, there may be some change in the education effect on income not included in this project.

Conclusion

This project analyzes the association between education and wage. We found that years of education have a large effect on monthly earnings using survey data in 1980. After we control for many factors measuring an individual's ability and demographic characteristics, the effect is still large. This effect differs a lot in different races and different birth orders. Although adding interaction terms lead to a small education effect in some subgroup, the overall effect is significant with no doubt. The policy implications are that the government should promote more education of individuals so that they can earn more money and lead to more social wealth. Encouraging and subsidizing higher education for citizens will lead to higher individual income, for both black and non-black population. Holding IQ, KWW score, years with current employers and years of working, age, average weekly working hours, and among people with same marriage status and living in cities/noncities status, more years of education can still booth individual wages.

Reference

Assari, Shervin. “Blacks’ Diminished Return of Education Attainment on Subjective Health; Mediating Effect of Income.” *Brain Sciences*, vol. 8, no. 9, 2018, p. 176., https://doi.org/10.3390/brainsci8090176.

Blackburn, McKinley, and David Neumark. “Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials.” 1991, https://doi.org/10.3386/w3857.

Card, David. “The Causal Effect of Education on Earnings.” *Handbook of Labor Economics*, 1999, pp. 1801–1863., https://doi.org/10.1016/s1573-4463(99)03011-4.

Morgan, James, and Martin David. “Education and Income.” *The Quarterly Journal of Economics*, vol. 77, no. 3, 1963, p. 423., https://doi.org/10.2307/1879570.

Appendix

Table 1

Summary of wage and education, before data cleaning					
Variable	**Observations**	**Mean**	**Standard Deviation**	**Min**	**Max**
Wage	935	957.9455	404.3608	115	3078
Education	935	13.46845	2.196654	9	18

Table 2

Summary of wage and education, after data cleaning					
Variable	**Observations**	**Mean**	**Standard Deviation**	**Min**	**Max**
Wage	731	932.7114	404.6243	115	3078
Education	731	13.22025	2.098121	9	18

Table 3

	(1) Wages	(2) wages	(3) wages	(4) wages
years of education	60.21***	44.79***	47.86***	50.77***
	(9.78)	(5.91)	(6.64)	(5.93)
IQ score		2.789**	3.240***	2.652**
		(2.70)	(3.50)	(2.60)
KWW Score		4.980*	5.009*	5.167*
		(2.17)	(2.44)	(2.28)
Years with current employer		10.32***		8.170*
		(3.46)		(2.36)
Age		12.64**	6.739	7.739
		(2.79)	(1.39)	(1.18)
Married/unmarried		165.8***	168.2***	165.5***
		(4.51)	(4.93)	(4.50)
Black/non-black		-124.0***	-123.1***	-135.2***
		(-3.54)	(-3.74)	(-3.85)
Live in SMSA/not in SMSA		153.9***	165.3***	155.7***
		(5.42)	(6.73)	(5.48)
Years of working experience			11.20**	6.893
			(3.03)	(1.07)
average weekly hours				-3.585
				(-1.56)
Constant	147.0	-828.3***	-799.3***	-652.4**
	(1.83)	(-4.61)	(-4.84)	(-3.25)
Observations	935	731	935	731
Adjusted R-squared	0.106	0.231	0.231	0.236

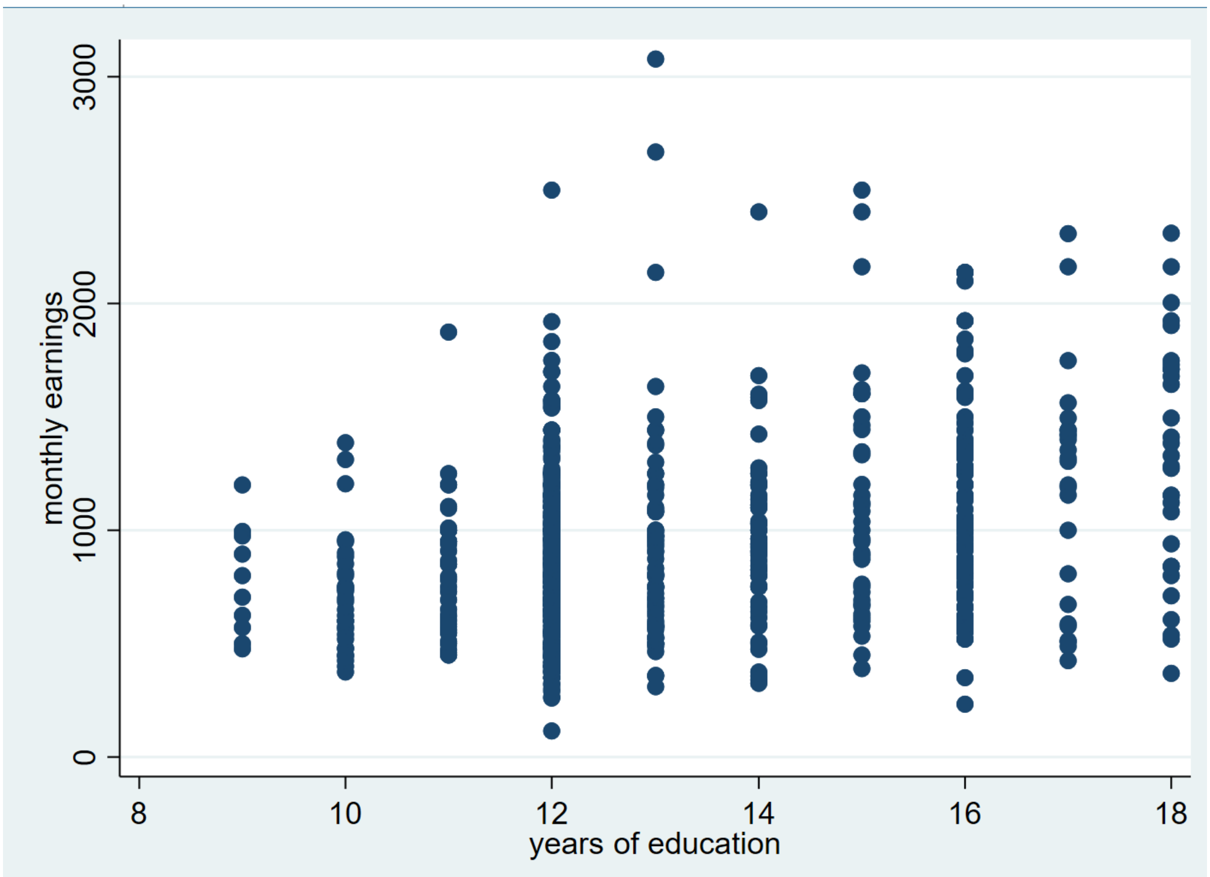
Note: Unit of observation is men is US in 1980. The t-statistics are in the parentheses.
* p<0.05, ** p<0.01, *** p<0.001

Table 4

	(1) monthly wage	(2) monthly wage	(3) monthly wage	(4) monthly wage
years of education	50.77***	54.57***	47.82***	65.83***
	(5.93)	(6.14)	(5.32)	(5.15)
IQ score	2.652**	2.598*	2.516*	2.477*
	(2.60)	(2.56)	(2.41)	(2.37)
KWW score	5.167*	5.583*	4.526	4.709
	(2.28)	(2.46)	(1.85)	(1.93)
Years with current employer	8.170*	8.141*	8.538*	8.909*
	(2.36)	(2.35)	(2.37)	(2.50)
Years of working experience	6.893	6.785	3.427	4.470
	(1.07)	(1.05)	(0.49)	(0.64)
age in years	7.739	6.812	10.05	9.393
	(1.18)	(1.04)	(1.45)	(1.33)
average weekly hours	-3.585	-3.354	-3.446	-3.233
	(-1.56)	(-1.45)	(-1.31)	(-1.22)
Married/Unmarried	165.5***	160.8***	164.8***	159.7***
	(4.50)	(4.39)	(3.78)	(3.67)
Black/non-black	-135.2***	449.1*	-157.6***	-157.4***
	(-3.85)	(2.18)	(-4.12)	(-4.14)
In SMSA/not in SMSA	155.7***	154.8***	178.9***	178.9***
	(5.48)	(5.44)	(5.91)	(5.92)
Black*education		-47.10**		
		(-2.84)		
Birth-order: first			-39.41	337.2
			(-1.14)	(1.62)
Birth-order: second			12.05	367.9
			(0.28)	(1.40)
Birth-order: third			-83.22	653.5**
			(-1.91)	(2.61)
Birth-order: forth			-68.48	351.2
			(-1.48)	(0.94)
Born in second * education				-27.86
				(-1.76)
Born in third * education				-26.43
				(-1.32)
Born in forth * education				-55.24**
				(-2.96)
Born in fifth * education				-32.49
				(-1.01)
Constant	-652.4**	-686.0***	-608.4**	-855.3***
	(-3.25)	(-3.41)	(-2.82)	(-3.41)
Observations	731	731	658	658
Adjusted R-squared	0.236	0.239	0.234	0.236

Note: The t statistics are in parentheses, shows the t statistics for each coefficient. The unit of observations are men in the US in 1980. Variables with * represent interaction terms. For instance, black*education is the interaction term of being black and education.
* p<0.05, ** p<0.01, *** p<0.001

Figure 1



monthly earnings
0
1000
2000
3000
8
10
12
14
16
18
years of education