

Final Project

Hedgehogeconomists

April 4, 2018

Title: The Relationship Between Income Level and Inequality Da Yeon Lee, Susan Cossa, Clara Rosenberg, Cas Sweeney SDS 291 Multiple Regression Professor Capistrant 5.11.2018

Abstract: This paper examines whether there is a statistically significant relationship between educational achievement and income, and how this relationship differs between different races. We found that while not all combinations of race and education level result in a lower expected average income, there is a general trend in this direction.

Introduction

Throughout U.S. history, people have claimed that the reason behind disparity among unemployment levels across races is not because of racial discrimination, but rather because people of color were supposedly not as qualified as white people. In this paper, we define qualification as education level. Our main focus is to demonstrate that with the same level of education, a worker of color is estimated to, on average, earn less than a white worker. Existing scholarly works which discuss the disparity in how education affects earnings between white workers and workers of color include Erik Wright's Race, Class, and Income Inequality; and Randall Weiss's The Effect of Education on the Earnings of Blacks and Whites. The works that we have studied and cited provide empirical data that demonstrate the disparity in income between white workers and workers of color with the same educational achievements. For example, Wright observed the disparity between black and white workers in 1960: at less than elementary education black workers earn \$700 less; at the high school level, black workers earn \$1,400 less, and at the college level, black workers earn \$3,800 less (Write 1368). We are examining the extent to which this disparity has changed in 2016. The question that we've examined in this research is: does there exist a statistically significant relationship between educational achievement and income? Furthermore, we seek to determine in what way race affects this relationship. In other words, does income respond differently to educational achievement depending on the individual's race, and to what extent? Our primary hypothesis is that the higher the level of education, the higher the level of income will be. Our first model examines the relationship between individual income (response variable) and two explanatory variables, metropolitan status and education level: *Individual Income=??1Education Level+??2Metro+??3Age+??4Sex+??5School Type+??0*. The secondary hypothesis will look at the difference in wealth ratios within racial groups of people, using the following model: *Individual Income=B1Education Level + B2WhiteRace + B3(Education Level * WhiteRace) + B0*.

Methods

This data is from the American Community Survey 2016 Sample, a 1-in-100 weighted random sample of the U.S. population. This survey can be completed online, by mail, in person, or by phone. Our primary hypothesis is that as education increases, income will also increase. Our secondary hypothesis is that people of minority races will have lower income across all levels of education and that the increase in income across education levels will be smaller. We will adjust for gender, age, school type, and metropolitan area. This research project will use data from IPUMS U.S. Census Data for Social, Economic and Health Research. The selected variables requested from IPUMS include the following: ? Race: This is broken down into various racial categories that each have values of 1 for No or 2 for Yes. We then recategorized this into a factor variable with values of White, Black, Asian, and Other. ? Income: a quantitative variable, INCWAGE. For this, we filtered out missing data because it was assigned a high value, which would have skewed the results. ? Education: originally a variable called EDUC, we transformed this into a factor variable with new categories of less than high school, finished high school, and finished college. ? School type: the variable

SCHLTYPE, which indicates whether the person went to public or private school. We included this because we thought that it might have some association with an individual's family income growing up, which might have association with income later in life. ? Sex: the variable SEX, indicating 1 for male or 2 for female. We included this because the gender wage gap is well-documented. ? Age: a quantitative variable, AGE. We included this because we thought that people might make more money as they get older (up to retirement) due to having more experience. ? Employment status: the variable EMPSTATUS, requested so that we could filter for only those who were employed. ? Metropolitan area: the variable METRO, tracking where people live in relation to a large city, in the following categories: 0 (not identifiable); 1 (not in metro area); 2 (in metro area, central city); 3 (outside central city); 4 (status unknown). We included this because we thought that different kinds of jobs, and therefore different incomes, might be available in different regions. Our hypotheses are tested with a multiple regression model, and we will consider the p-values of each coefficient in the model to determine whether these factors are significant. For each of our hypotheses, we checked for the assumptions of linearity, normal distribution of residuals, and equal distribution of errors, all of which were met.

Result

```
stargazer(m1, m2, type="text", ci=TRUE, single.row=TRUE, digits = 0, title="The comparison of education and income")

##
## The comparison of education and income
## =====
##                               Dependent variable:
## -----
##                                     INCWAGE
##                                     (1)          (2)
## -----
## Educationfinished college      52,913*** (52,531, 53,294)    55,702*** (55,231, 56,173)
## Educationfinished high school 13,718*** (13,349, 14,086)    15,417*** (14,958, 15,876)
## METRO                         2,114*** (2,050, 2,177)
## SEX                            -20,209*** (-20,393, -20,025)
## AGE                            490*** (484, 497)
## SCHLTYPE                      -9,730*** (-9,974, -9,485)
## RaceAsian                      1,921** (333, 3,509)
## RaceBlack                      -1,146* (-2,429, 137)
## RaceOther                      666 (-263, 1,596)
## Educationfinished college:RaceAsian 4,207*** (2,528, 5,887)
## Educationfinished high school:RaceAsian -6,736*** (-8,456, -5,017)
## Educationfinished college:RaceBlack -17,786*** (-19,220, -16,352)
## Educationfinished high school:RaceBlack -5,361*** (-6,706, -4,015)
## Educationfinished college:RaceOther -17,385*** (-18,630, -16,141)
## Educationfinished high school:RaceOther -7,841*** (-8,885, -6,797)
## Constant                        36,294*** (35,647, 36,941)    21,682*** (21,245, 22,119)
## -----
## Observations                    1,459,560                  1,459,560
## R2                             0                          0
## Adjusted R2                     0                          0
## Residual Std. Error            56,404 (df = 1459553)    58,018 (df = 1459548)
## F Statistic                     47,193*** (df = 6; 1459553) 17,052*** (df = 11; 1459548)
## -----
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

Our primary hypothesis is that an increase in education is correlated with an increase in income. This is

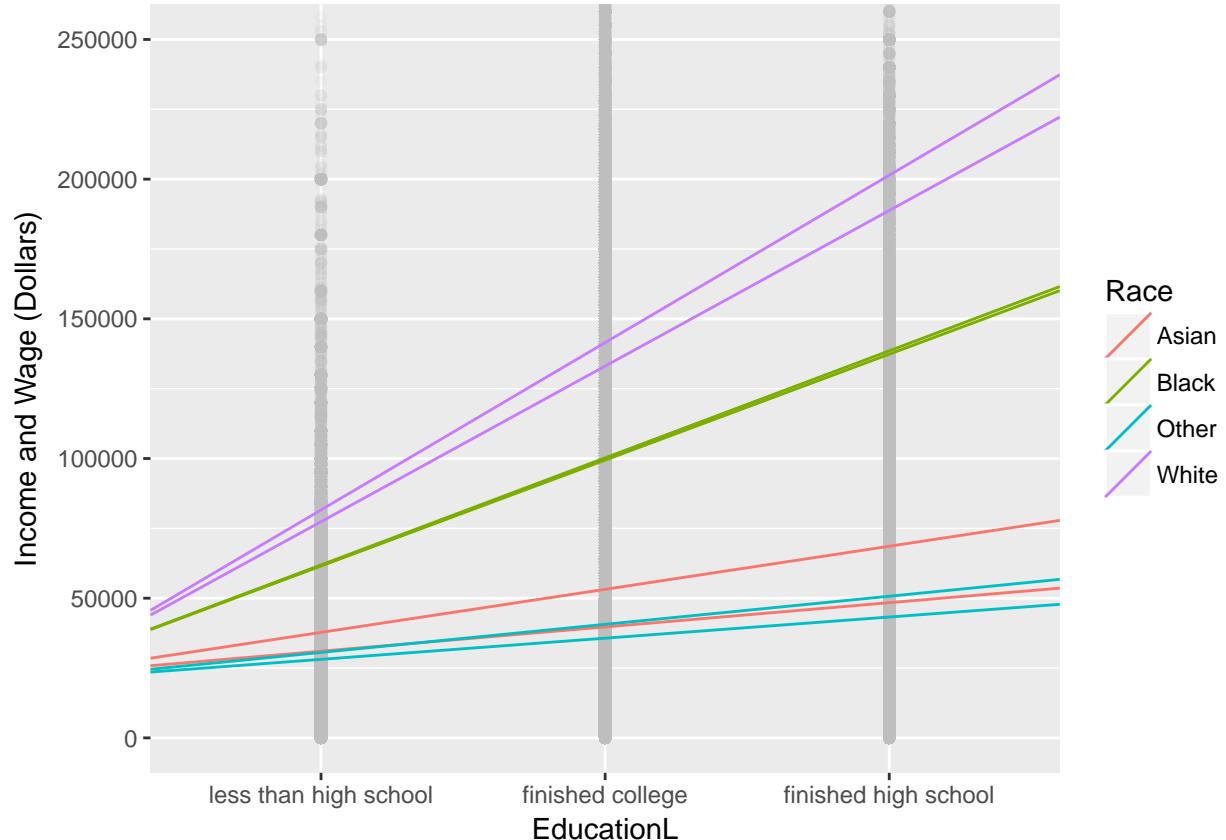
tested with a null hypothesis that the beta value associated with education is equal to zero. We included the confounding factors of metro status, sex, age, and school type in our model. This model shows that on the individual level, compared to not finishing high school, finishing high school is associated, on average, with a \$13,718 increase in income, and finishing college is associated, on average, with a \$52,913 increase in income, adjusting for other factors in the model. Even though we narrowed “Education” variable into 3 groups, this is not an easily interpretable number because a one-unit increase in education does not always correspond to the same level of increase according to the IPUMS data.

According to our model, all the confounders that we have chosen are important for the model. We can arrive at this conclusion because the beta values all resulted in p-values of less than 0.05, representing statistically significant differences. Therefore, we reject the null hypotheses that the beta values are equal to zero and conclude that they are each statistically significantly different from zero, significantly impacting income.

Our secondary hypothesis is that people of color tend to earn less than white people. This is tested with a null hypothesis that the beta value associated with race is equal to zero. We also included an interaction term between education and race. The model takes white and less than high school as the default and displays beta values for all other indications of race. In this model, all of the interaction terms between education and race are statistically significant, but the terms associated with race alone are not. This suggests that the interaction between race and education has more of an impact on income than race alone. The only statistically significant term that has a positive value, meaning that income is, on average, increasing at a greater rate than the default groups, is the interaction between finishing college and RaceAsian. The existence of a positive coefficient means that we cannot reject our null hypothesis entirely that people of color make statistically significantly less than zero. However, the large number of negative coefficients mean that this bears further inquiry. Furthermore, we did not include confounding variables in this hypothesis.

`usaplot1`

```
## Don't know how to automatically pick scale for object of type labelled. Defaulting to continuous.
```



This graph shows different regression lines for different races. Based on this figure, income for all race groups increases when their education levels increase. The slope is the largest for the variable “white”. In the real world context, this means that an increase in the education level causes the biggest increase in income and wage for white people. On the other hand, an increase in the education level causes the smallest increase in income and wage for people in “other race, nec” variable. Therefore, changes in education levels influence white people’s income and wage the most.

Discussion

We are looking at the association between education level and income across race in the US. We believed that as education level increases, so does income, but that increase would not be the same across different races. We found that there is a positive association between education and income, and that the interaction between race and education is also significant. For white people, the increase in income is the largest and the intercept representing income with less than a high school education is greater.

This data is only from 2016. Looking at more years would add another level to the analysis. An additional question would be whether and how the income gap has changed across races over time, for each education level.

References

Weiss, Randall. “The Effect of Education on the Earnings of Blacks and Whites.” Wright, Erik Olin. “Race, Class and Income Inequality.”