Effect of Educational Attainment and Race on Labor Market Earnings in the United States

Tarun Shrivas[1], 2nd Author[1], 3rd Author [1]

[1] Seattle University, Seattle, WA, USA

Introduction

Education plays a key role in modern labor markets. Several studies in different countries have looked at the impact of educational attainment on income, unemployment, and occupation type. Tiffany Julian and Robert Kominski (2011) determined earnings gaps across the spectrum of educational levels [1]. Angrist and Kruger (1991) used the quarter of birth as a benchmark and found that an additional year of education results in 7.5 percent higher earnings on an average [2]. Similar results have been noted in developing countries too. Duflo (2000) examined the effect of education policy reforms in Indonesia to increase the number of primary schools. She found that the earnings of the individuals increase by 6.8-10.6 percent for each additional year of education [3]. Given the vast outlays of money spent on education in the United States, it's important to assess the effect of educational attainment levels on earnings in the labor market. While it's generally accepted that higher educational attainment confers a premium on earnings, what is the increase in earnings with an additional level of educational attainment in the United States? Furthermore, are there differences in earnings between races across educational attainment levels? These research questions are important as they help us understand the inequities in earnings by races for a given educational level.

In this paper, we estimate the earnings gap by educational attainment and by race/ethnicity sub-samples. We find that after controlling for gender, marital status, age, and race/ethnicity, higher educational attainment levels confer statistically significant ($p < 0.001$) premia on earnings. However, controlling for educational attainment levels, age, gender, and marital status, the non-Hispanic Whites earn statistically significantly more *($p < 0.001$)* than the Blacks and Hispanics. We also find that the earnings gap between non-Hispanic Whites and Asians are neither statistically nor economically substantively significant.

Econometric Model and Estimation Method

Econometric Model

Following labor economics convention, we have used natural logarithm of annual earnings as the dependent variable in our econometric model [4][5]. This functional form, known as a log-linear model, is appropriate to use here for statistical and substantive considerations. The log-linear model was selected considering the skewness in the earnings distribution. The log-linear model is also appropriate as multiplicative earnings gaps are more substantive than additive earnings gaps. For example, the additive gap between earnings of $40,000 and $50,000 is the same as that between the earnings of $150,000 and $160,000; these multiplicative gaps are 25% and 7% respectively and are more substantive.

While educational attainment and race are the two key independent variables of interest, our model also includes age, gender and marital status as other independent variables. In our model, individuals are distinguished amongst eight levels of educational attainment: No high school degree (the reference category), a high school degree, some college, an associate's degree, a bachelor's degree, a master's degree, a professional degree, or a doctorate degree. Further, individuals are distinguished amongst four race/ethnicity groups: non-Hispanic White (the reference category), Black, Asian, and Hispanic. We have included an individual's age as a proxy for his/her experience in the labor market. Following conventional

cross-sectional profile of earnings by age [6], our model includes age in a quadratic form, adding an age-squared term to age. Also, given the possibility that gender [7] and marital status [8] may have a significant effect on the estimated earnings, our model distinguished between females and males (the reference category) and between married and non-married (the reference category) individuals.

To summarize, we estimated a model of the form:

$$\ln(\text{Earnings}) = \alpha + \mathbf{A}'\beta + \mathbf{B}'\gamma + \mathbf{C}'\theta + \mathbf{D}'\delta + \varepsilon$$

where **A** represents a vector of dummy variables indicating the educational attainment, **B** represents a vector of dummy variables indicating the race/ethnicity, **C** represents a vector of variables indicating age and age squared, **D** represents the vector of dummy variables indicating gender and marital status.

Estimation Method

We estimated the log earnings equations with the Ordinary Least Squares (OLS) method. However, a diagnosis of the residual terms and using the Breusch-Pagan-Godfrey test, we rejected the null hypothesis of homoscedasticity and concluded heteroskedasticity exists in our model. To adjust for heteroskedasticity, we report robust standard errors with our results.

Data

Sample

The data used in this paper is a subset extracted from the 2011 American Community Survey (ACS) [9]. The 2011 ACS survey is a cross-sectional, nationally representative data of 2,128,104 households. The ACS survey, largely a self-administered survey, includes questions pertaining to each household member's demographic and labor market characteristics. Through the ACS, we have data regarding jobs and occupations, educational attainment, veteran status, whether people own or rent their homes, and other topics. Public officials, planners, and entrepreneurs are encouraged to use the ACS survey data to assess the past and plan for the future.

Although the ACS survey is conducted at a household level, the sampling unit in our model is for individuals. For our model, we extracted data in a two-step process. As a first step, we generated a random sample of 64,999 individuals, aged between 18 to 64 years. In the next step, we restricted the sample to exclude the following individuals, to get the final sample size for our model:

i. This sample included a wide range of self-reported annual earnings. Some individuals reported their annual earnings as zero. Others reported their annual earnings as less than $15,080, an earning level attainable in a typical 40 hours/ 52 weeks/year, at $7.25/hour, the 2011 minimum federal wage rate. Individuals who reported annual earnings less than $15,080 were excluded
ii. Individuals who reported working less than 40 weeks were excluded
iii. Individuals who reported working less than 35 hours in a typical week
iv. Hawaiian, Biracial, and Native American individuals were excluded as they were severely underrepresented in the sample

The final sample size thus obtained includes 28,432 individuals.

Descriptive Statistics

As we can observe in **Table 1** on the following page, across all race/ethnicity groups, the median earnings among individuals with no high school degree range from $25,000 to $31,000. As we move from no high school degree to a professional degree, the median earnings continue to rise across all race/ethnicity groups except for the Black individuals. Amongst Black individuals, the median earnings of individuals with associate's degree are $2,628 lesser than those with some college educational attainment. It is also interesting to note that the median earning among individuals with a doctorate degree is lesser than that among individuals with a professional degree, across all the four race/ethnicity groups. Amongst individuals with a bachelor's degree, master's degree, professional degree or doctorate degree, Asian individuals lead the other three race/ethnicity groups in median earnings. Amongst all individuals with educational attainment up to an associate's degree, non-Hispanic White individuals lead the other three race/ethnicity groups in median earnings. On average, these earnings disparities between race/ethnicity groups widens as we move up along the educational attainment levels.

Table 1: Descriptive Statistics of Earnings (in $) by Race–Educational Attainment sub-groups

		No High School	High School	Some College	Associate's	Bachelor's	Master's	Professional	Doctorate
Non-Hispanic White (N_t = 20954)	***N_s***	**704**	**5065**	**4568**	**2099**	**5297**	**2277**	**593**	**351**
	Mean	37,975	43,240	49,980	51,984	74,162	86,793	154,121	06,359
	Std. Dev.	21,398	29,818	34,965	34,607	6,228	7,482	119,486	83,225
	Min.	15,400	15,100	15,200	15,500	15,200	18,000	22,500	16,300
	Median.	31,000	36,000	41,000	45,000	58,000	66,000	110,000	81,000
	Max.	152,000	507,000	507,000	398,000	577,000	577,000	507,000	507,000
Black (N_t = 2561)	***N_s***	**173**	**694**	**715**	**240**	**438**	**242**	**30**	**29**
	Mean	34,374	35,209	46,150	43,522	56,016	70,474	171,861	76,683
	Std. Dev.	32,854	18,891	21,773	21,489	39,510	48,149	72,452	31,533
	Min.	15,400	15,600	15,200	16,800	15,100	18,000	22,500	34,800
	Median.	27,000	30,000	34,000	37,150	50,000	59,500	89,500	70,000
	Max.	400,000	230,000	250,000	127,000	491,000	399,000	318,000	150,000
Hispanic (N_t = 3362)	***N_s***	**789**	**910**	**735**	**260**	**434**	**166**	**56**	**12**
	Mean	30,264	35,973	40,952	47,984	56,598	72,228	107,173	92,833
	Std. Dev.	17,118	18,372	27,713	30,305	43,949	52,073	110,894	78,555
	Min.	15,100	15,400	15,400	15,600	15,600	18,000	15,700	18,000
	Median.	25,000	30,000	33,300	40,000	46,000	60,000	70,000	69,000
	Max.	285,000	150,000	360,000	340,000	398,000	360,000	507,000	330,000
Asian (N_t = 1555)	***N_s***	**107**	**198**	**185**	**112**	**521**	**269**	**83**	**80**
	Mean	32,782	35,316	46,150	51,621	73,969	93,926	171,561	112,263
	Std. Dev.	24,810	23,142	29,570	31,794	57,705	66,345	125,795	85,981
	Min.	15,400	15,600	15,200	15,400	15,600	18,000	22,500	24,000
	Median.	25,000	30,000	40,000	43,600	62,000	80,000	120,000	93,000
	Max.	210,000	241,000	301,000	217,000	507,000	491,000	450,000	450,000

N_t: Total sample size of the race/ethnicity group, across all educational attainment levels

N_s: Sample size of the race–educational attainment sub-groups

Results

Parameter Estimates

Table 2 below, presents the parameter estimates of our log (Earnings) equation for the overall sample of 28,432 individuals that includes all the race/ethnicity groups. As expected, our estimates suggest that the earnings are positively associated with educational attainment after controlling for gender, age, marital status and race/ethnicity.

Table 2: Estimated log (Earnings) equation by Age, Gender, Marital Status, Educational Attainment and Race/Ethnicity

Dependent Variable:	**Log(Earnings)**
Parameter Estimate:	Estimates (*Std. Errors*)
(Intercept)	8.897*** (0.0401)
Age	0.063*** (0.0020)
Age-squared	-0.001*** (0.0000)
Female[1]	-0.265*** (0.0060)
Married[2]	0.087*** (0.0064)
Educational Attainment[3]	
High school	0.161*** (0.0121)
Some college	0.317*** (0.0125)
Associate's Degree	0.387*** (0.0143)
Bachelor's degree	0.664*** (0.0131)
Master's degree	0.812*** (0.0152)
Doctorate degree	0.951*** (0.0303)
Professional degree	1.240*** (0.0288)
Race/Ethnicity[4]	
Asian	0.0004 (0.0147)
Black	-0.114*** (0.0096)
Hispanic	-0.170*** (0.0093)
Sample Size	**N = 28432**
Multiple R-squared:	34.8%
Adjusted R-squared:	34.7%
F-Statistics	1082 p-value: < 0.000***

[1] Female: Dummy variable with for gender; Female =1, Men = 0
[2] Married: Dummy variable for marital status; Married = 1, Non-married = 0
[3] Educational Attainment: Dummy variable with No high school degree as the reference category
[4] Race/Ethnicity: Dummy variable with Non-Hispanic White as the reference category

Note: Parameter estimates are obtained with the Ordinary Least Squares (OLS) method. Robust standard errors in () parentheses.
***p* < 0.05. ****p* < 0.01. *****p* < 0.001**

For example, individuals with a master's degree earn, on average, about 125% more than the individuals with no high school degree, after controlling for gender, age, marital status and race/ethnicity. This gap in earnings is statistically significant ($p < 0.001$). Likewise, controlling for other variables in the model, individuals with a high school degree earn 17.5% more, on average, than individuals with no high school degree, the gap being statistically significant ($p < 0.001$). Controlling for other variables in the model, on average, a professional degree confers the highest premium on earnings. Individuals with a professional degree, on average, earn about 246% more than the individuals with no high school degree, the gap in earnings being statistically significant ($p < 0.001$).

Further, controlling for gender, marital status, educational attainment and race/ethnicity; earnings increase with age, at a decreasing rate peaking at 51.6 years and then decrease at an increasing rate. While gender earnings gap is not a key focus of this paper, we find that after controlling for age, marital status and race/ethnicity and educational attainment, women earn 23% less than men, the earnings gap is statistically significant ($p < 0.001$). Marriage widens the earnings gap between men and women, married -women, on average earnings 30% less than married-men, controlling for other variables in the model. The gap in earnings being statistically significant ($p < 0.001$).

Table 3 on the following page, presents the parameter estimates of our log (Earnings) equation for the four sub-samples by race/ethnicity groups. As expected, our estimates suggest that within each sub-sample, the earnings are positively associated with educational attainment, after controlling for gender, age, and marital status. However, the premia on earnings for a given educational attainment level is different across ethnic sub-samples.

For example, Black individuals with a bachelor's degree earn, on average, about 70% more than the Black individuals with no high school degree, after controlling for gender, age, and marital status. Likewise, among Non-Hispanic White individuals with a bachelor's degree, the estimated premium on earnings is 89% over no high school degree. -

Table 3: Estimated log (Earnings) equation by Age, Gender, Marital Status, Educational Attainment for Race/Ethnicity sub-samples

Dependent Variable:	**Log (Earnings)**			
Race/Ethnicity:	**Non-Hispanic White**	**Black**	**Asian**	**Hispanic**
Parameter Estimate:				
(Intercept)	8.766*** (0.0486)	9.012*** (0.1147)	8.938*** (0.2206)	9.278*** (0.0956)
Age	0.070*** (0.0023)	0.055*** (0.0056)	-0.060*** (0.0109)	0.038*** (0.0050)
Age-squared	-0.001*** (0.00003)	-0.001*** (0.00007)	-0.001*** (0.00013)	-0.0003*** (0.00006)
Female[1]	-0.294*** (0.0071)	-0.151*** (0.0184)	-0.197*** (0.0281)	-0.183*** (0.0164)
Married[2]	0.102*** (0.0076)	0.074*** (0.0185)	0.045 (0.0317)	0.050*** (0.0167)
Educational Attainment3:				
High school	0.129*** (0.0181)	0.079** (0.0395)	0.108 (0.0548)	0.208*** (0.0197)
Some college	0.285*** (0.0183)	0.233*** (0.0395)	0.383*** (0.0566)	0.328*** (0.0229)
Associate's Degree	0.345*** (0.0198)	0.313*** (0.0463)	0.447*** (0.0687)	0.488*** (0.0316)
Bachelor's degree	0.639*** (0.0187)	0.532*** (0.0432)	0.798*** (0.0522)	0.618*** (0.0297)
Master's degree	0.778*** (0.0207)	0.748*** (0.0477)	1.013*** (0.0558)	0.819*** (0.0434)
Professional degree	1.213*** (0.0339)	1.055*** (0.1376)	1.517*** (0.0901)	0.972*** (0.1104)
Doctorate degree	0.910*** (0.0376)	0.863*** (0.0849)	1.108*** (0.0824)	0.942*** (0.1685)
Sample Size	**N = 20954**	**N = 2561**	**N = 1555**	**N = 3362**
Multiple R-squared:	33.0%	29.7%	38.2%	28.1%
Adjusted R-squared:	33.0%	29.4%	37.8%	28.0%
F-Statistics	938.74 p-value: < .001***	97.73 p-value: < .001***	86.75 p-value: < .001***	119.50 p-value: < .001***

[1] Female: Dummy variable with for gender; Female =1, Men = 0
[2] Married: Dummy variable for marital status; Married = 1, Non-married = 0
[3] Educational Attainment: Dummy variable with No high school degree as the reference category

Note: Parameter estimates are obtained with the Ordinary Least Squares (OLS) method. Robust standard errors in () parentheses
***p* < 0.05. ****p* < 0.01. *****p* < 0.001**

The premia on earnings associated with a high school degree and higher educational attainment levels are presented in **Figure 1**, below.

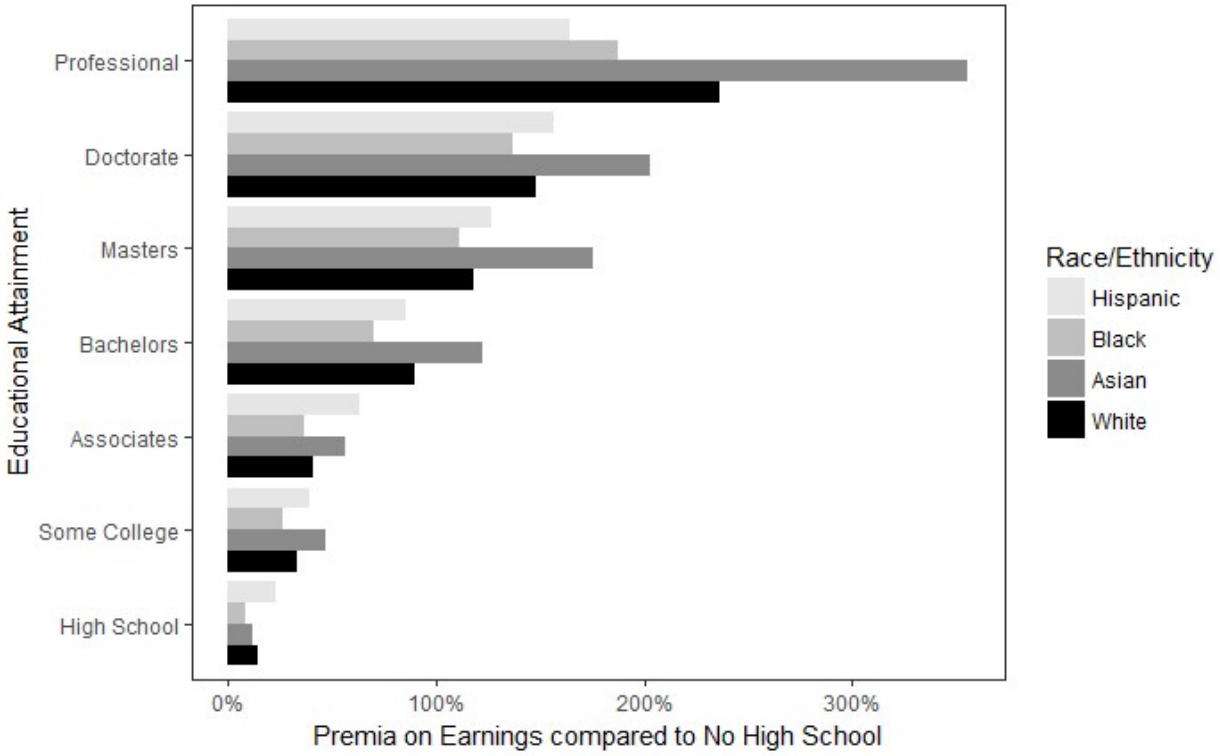

Figure 1: Estimated premia on earnings with educational attainment levels

In viewing the premia on earnings compared to no high school, Hispanics lead in premia followed closely by Asians from high school to some college and associates degree. Once the effect of bachelor's degrees is measured, this lead is lost to Asians and non-Hispanic Whites. Additionally, the premia amongst Asians is economically significant for doctorate and professional degrees with earnings, on average, being 202% and 355% respectively, compared to no high school. While the corresponding premia amongst non-Hispanic Whites are only 148% and 236% respectively.

Conclusion

The earnings and educational attainment level have a strong positive association. Controlling for gender, age, marital status, and race/ethnicity, subsequently higher educational attainment levels confer premia on earnings. The effect of educational attainment level on earnings within race/ethnicity groups, follow the directional effect, higher educational attainments conferring premia within all the groups. We find evidence for a systematic gap in earnings between non-Hispanic Whites and Blacks and Hispanics. At all educational attainment levels, statistically significant and economically substantive gaps are observed in the earnings premia estimated for non-Hispanic Whites and those estimated for Blacks and Hispanics, after controlling for gender and marital status. Our findings suggest that approximately 35% of the overall earnings gap is explained by the difference in educational attainments and age (a proxy for age), gender, marital status and race/ethnicity. The unexplained portions of the earnings gap not captured by our model may relate to other variables that were not present in the data. One possible explanation is discrimination in the labor market on gender and/or race/ethnicity. These findings provide an insight into possible labor

market discrimination on race/ethnicity but the results from these findings should be interpreted with caution. As discussed earlier, the ACS data on which our model is based is a household level survey having individual-level data. Our final sample may have more than one individuals from the same household. In such cases, the unobserved characteristics that influence earnings may have a high correlation, therefore the parameter estimates may not be efficient.

Also, as ACS is a self-administered survey, the response quality and hence the data quality may have suffered. Our sample includes some seemingly non-random earnings levels such $507,000 and $577,000 reported by many individuals. It is counter-intuitive, as only a very few of such outliers were expected. Self-selection bias, which is a type of non-sampling error, could have biased the ACS data. Respondents who completed the ACS survey could be demographically and/or behaviorally very different than those who did not.

Nevertheless, the paper finds evidence of demographic differences on earnings when exploring their relationship with educational attainment and race/ethnicity in the U.S and provides insights on potential policy level issues.

References

1. https://www.census.gov/prod/2011pubs/acs-14.pdf
2. http://faculty.wwu.edu/kriegj/Econ445/Papers/Angrist%20and%20Krueger.pdf
3. https://www.isid.ac.in/~tridip/Teaching/DevelopmentMicroeconomics/Spring2012/Readings/04Education/02Duflo-AER2001.pdf
4. https://www.sciencedirect.com/science/article/pii/S1574069206010075
5. http://journals.sagepub.com/doi/pdf/10.1177/1044207316658752
6. http://www.statcan.gc.ca/pub/75-001-x/2009101/charts-graphiques/10779/c-g000a-eng.htm
7. http://www.pewresearch.org/fact-tank/2017/04/03/gender-pay-gap-facts/
8. http://epc2014.princeton.edu/papers/140782
9. https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size/index.php

Contributions

Technical Appendix, tables and graphs was generated by Tarun Shrivas and 2nd Author. Initial paper was drafted by 3rd Author and Tarun Shrivas after discussing model and conclusions together. Tarun Shrivas[1], 2nd Author[1], 3rd Author [1] proofread and updated the final version.