**How Does Higher Education Impact Earnings Potential?**

**Does Attaining Higher Education Represent an Equally Worthwhile Investment for both Whites and Blacks?**

**Prepared for:**

Professor Bridget Hiedemann

Applied Econometrics

ECON 5300

**Prepared by:**

Kyler Adams

Thomas del Zoppo

Ryan Loong

**February 16th, 2017**

**Introduction**

Benjamin Franklin once said, “An investment in knowledge pays the best interest.” Since the founding of the United States, Americans have intuitively understood that the attainment of higher education should lead to higher earnings. However, as the costs of higher education in America continue to rise, one must remain confident that a prospective investment in higher education will remain economically, as well as personally, worthwhile. Not everyone with the same level of education will receive the same earnings. Other factors may influence earnings as well, and unfortunately, with a dark history of racial injustice in America, a person’s race may influence their earnings if racial discrimination remains present. So, discerning the true benefit of higher education requires a comprehensive econometric analysis that isolates the effects of educational attainment from other factors that influence earnings.

This paper conducts an econometric analysis of earnings in America to determine, ceteris paribus, the average economic value (or premium) of attaining additional levels of higher education, compared to the value of not attaining even a high school diploma. We also seek to determine how these earnings premiums may differ between the White and Black races. Starting with random survey data, we carefully subset the respondents (primarily by race) into three subsamples and performed multiple linear regression analyses to determine the average earnings premiums resulting from the higher education attained by the respondents of each subset, compared to the subset’s non-high school graduates. From these analyses, we conclude that, on average, all additional investments in the attainment of progressively higher education do produce progressively higher earnings premiums, but these premiums may be less valuable for Blacks than for Whites.

**Econometric Model and Estimation Method**

In order to examine the determination of earnings, we were compelled to create econometric models that positioned earnings as the dependent variables in our regression analyses. However, unadjusted earnings seemed inappropriate for our analysis. As earnings are often skewed across a population, we considered, for example, that the effect of a small earnings premium may be less measurable for someone earning $200k than for someone earning $30k. Thus, as the dependent variable in each of our model specifications, we chose instead to use the natural log of the respondent’s total earnings over the preceding 12 months. The independent variables in each of our models included various levels of educational attainment, in addition to age (a partial proxy for work experience), sex, and marital status. The variable of not having a high school diploma was not included in our modeling as to serve as a reference group. We ultimately produced a total of three subsetted models in order to effectively introduce White and Black races as independent variables. We chose to analyze White and Black races individually in two of our subsets because we thought there might be diverging earnings profiles between Whites and Blacks. The variable “White” was included in the third subset that combined both White and Black samples. In this subset, Blacks were thus the racial reference group. We applied the ordinary least squares (OLS) method to estimate the parameters in each of our multiple linear regression models. We chose the OLS method for its interpretability in measuring the estimators of each population parameter. For all percentage estimates in interpreting these estimator coefficients, we used the formula: 

**Data**

The original data set was obtained from the results of the American Community Survey (ACS) provided by the US government on their website (<http://www.census.gov/acs/www/>). The original sample was taken in 2011 by assessing approximately 3.3 million US housing units, randomly sampled, based upon their physical addresses registered with the government. From here, a separate random sample of 64,999 observations was extracted for this project.

Upon our preliminary exploration of the data, and with the formulation of our research question, we decided to further subset the data for various reasons. In short, we wanted to look only at respondents who worked full-time and were fully employed for the entire year. It would have been more difficult for us to research our particular questions with confounding variables present such as part-time vs. full-time workers, or temporary vs. year-round workers. With this in mind, the following filtering was applied to the data set:

1. Earnings in the preceding 12 months were greater than or equal to $10,150 - These were determined to be the lowest possible earnings still relevant to our research question (an amount equivalent to earnings from working at least 40 weeks \* 35 hours a week \* $7.25 for the 2011 federal minimum wage).
2. Worked 35+ hours a week - Like the reasoning above, this further ensured a focus on only full-time workers.
3. Worked 40+ weeks - We assumed that working more than 40 weeks in a year meant that the participant was employed for the full year.
4. Usual weekly hours greater than 0 - Our previous filtering should have ensured this requirement, but to address any potential data entry errors we applied this redundant filter.

Table 1 shows the descriptive statistics of each variable for our two primary subsets, Whites and Blacks who are all fully employed. The first row shows that the mean earnings for White and Black participants are $60,216.32 and $43,166.64 respectively. The second row shows the standard deviation for White and Black earnings as $56,663.18 and $32,751.55 respectively. These standard deviation values are quite large in relation to the mean, suggesting there is a large variance and thus a large range of values of overall earnings.

The third row shows the median earnings for Whites and Blacks as $45,000 and $35,000 respectively. Note that the means for both samples are much higher than the medians, suggesting that the data is skewed to the right by a small number of participants with extremely high earnings. This is typical and expected of wage data since there often exists a small number of individuals with extremely high earnings, but there cannot exist a small number of individuals with equally low earnings (i.e. it is impossible for someone to earn -$300,000 in a year).

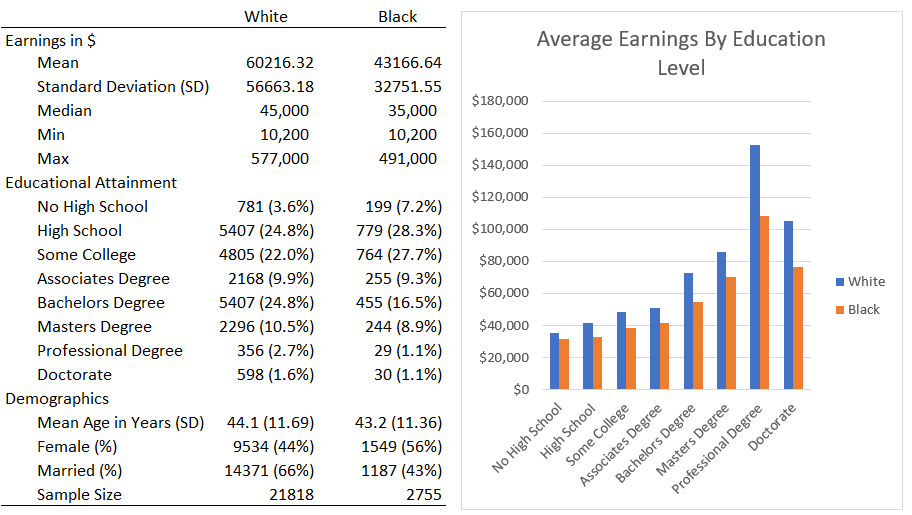
The fourth and fifth rows show the minimum and maximum earnings values for each subset. Note that the minimum is $10,200 for both samples, as this was what was determined to be the least possible wage that was applicable to the population parameters of fully employed workers that we wished to estimate.

The Educational Attainment section shows the number of respondents observed at each level of their highest educational attainment. Thus, there is no double counting. A person who has received a Bachelor’s Degree (and undoubtedly a high school diploma) is not also included in the High School group.

The demographic section shows the mean age of each subset, as well as the number and proportional percentages of female and married members of the subsets. Interestingly, the majority of Whites were male at 56%, while the majority of Blacks were female at 56%. 66% of the White subset was married and 43% of the Black subset was married. Finally, the sample sizes of each subset are given, with the White subset being approximately eight times larger than the Black subset. The total number of observations in the two subsets is 24,573, which is substantially reduced from our initial number of 64,999 observations, yet still large enough for effective statistical analysis.

Graph 1 visualizes the average earnings of the respondents at each level of educational attainment in the White and Black subsets. It shows that average earnings tend to increase with each higher level of educational attainment. At the two highest levels of attainment, the average earnings for those with doctoral degrees are lower than for those with professional degrees. However, doctoral earnings are still higher than the average earnings for those with master’s degrees. As defined by our survey, a person does not typically attain both a professional and a doctorate degree. They are separate options for attaining additional education beyond that of the master’s level, and it is open to interpretation as to which of these may be the “higher” level of education. Although this graph depicts a strong correlation between earnings and educational attainment, it does not demonstrate that higher education causes higher earnings. To investigate that hypothesis, we have conducted statistical hypothesis testing in the course of our regression analyses. The empirical results of those tests are explored in our next section.

Table 1: Descriptive Statistics (Year 2011) Graph 1: Earnings Breakdown



**EMPIRICAL RESULTS**

As stated above, we estimated three model specifications through multiple linear regression analysis with ordinary least squares estimation. The dependent variable in each model was the natural log of earnings over the preceding 12 months. In all three specifications, all coefficients were statistically significant at the 1% level, except for high school degree in the Black only subset, which was statistically significant at the 5% level. We discovered evidence of heteroskedasticity in our models, and continued to see this even after re-specifying to log-linear models. As heteroskedasticity causes standard errors to be invalid, to fix this problem, we decided to use (and report) robust standard errors in order to calculate valid t-statistics and p-values. For the “full model,” and its parceled White and Black subsets, we found adjusted R-squares of 0.347, 0.342 and 0.295, respectively. These values indicate that the percentage of variation in the dependent variable that can be explained by our models is 34.7%, 34.2%, and 29.5%, respectively. This also means that about 70% of the observable variation in earnings cannot be explained by our modeling. This percentage acknowledges that many other factors undoubtedly influence the determination of earnings, many of which are impossible to measure quantitatively, such as work ethic.

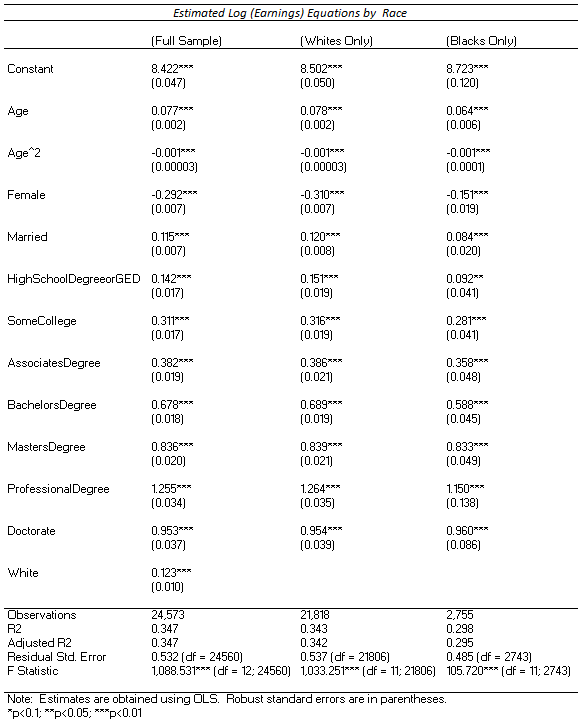
Our first model specification analyzes all White and Black workers who are fully employed. Its coefficients for educational attainment levels are 0.142 for a high school diploma, 0.311 for some college, 0.382 for an associate’s degree, 0.678 for a bachelor’s degree, 0.836 for a master’s degree, 1.255 for a professional degree, and 0.953 for a doctorate. We interpret these coefficients to estimate, all else being equal, an average earnings premium of 15.26% for a high school diploma, 36.48% for some college, 46.52% for an associate’s degree, 96.99% for a bachelor’s degree, 130.71% for a master’s degree, 250.78% for a professional degree, and 159.35% for a doctorate. Additionally, White has a coefficient of 0.123 which we interpret as an average earnings premium of 13.09% for Whites when compared to Blacks, all else being equal.

Our second model specification analyzes the subset of White workers who are fully employed. The coefficients for educational attainment levels are 0.151 for a high school diploma, 0.316 for some college, 0.386 for an associate’s degree, 0.689 for a bachelor’s degree, 0.839 for a master’s degree, 1.264 for a professional degree, and 0.954 for a doctorate. We interpret these coefficients to estimate, all else being equal, an average earnings premium of 16.30% for a high school diploma, 37.16% for some college, 47.11% for an associate’s degree, 99.17% for a bachelor’s degree, 131.41% for a master’s degree, 253.96% for a professional degree, and 159.61% for a doctorate.

Our final specification analyzes the subset of Black workers who are fully employed. The coefficients for educational attainment levels are 0.092 for a high school diploma, 0.281 for some college, 0.358 for an associate’s degree, 0.588 for a bachelor’s degree, 0.833 for a master’s degree, 1.15 for a professional degree, and 0.96 for a doctorate. We interpret these coefficients to estimate, all else being equal, an average earnings premium of 9.64% for a high school diploma, 32.45% for some college, 43.05% for an associate’s degree, 80.04% for a bachelor’s degree, 130.02% for a master’s degree, 215.82% for a professional degree, and 161.17% for a doctorate.

In all three specifications, the educational coefficients and their associated estimates of earnings premiums increased with each additional level of educational attainment. Professional and doctoral degrees had the highest premiums overall (both being typically terminal degrees). The magnitudes of the coefficients, and thus the estimates of earnings premiums, were smaller in the Black subset than in the White subset. This appears to suggest that the earnings premiums due to higher education are different and less valuable for the population of Blacks than for the population of Whites. However, upon further testing, we could not verify this condition with statistical significance. When testing an exploratory model complete with interaction terms between race and the levels of education, we could not estimate any of these interactions with statistical significance. The White variable did, however, remain statistically significant at the 5% level with a similar magnitude of 10.1%. Together, this implies that Whites have somewhat of a “head-start” on earnings over Blacks, but that their earnings profiles may not actually diverge more with additional academic attainment.

Table 2: Regression Results



**Conclusion**

With our results, we have strong statistical evidence to conclude that for both Whites and Blacks, there does exist a progressively higher earnings premium for attaining additional levels of education (compared to the value of not attaining a high school diploma). Nor can we conclude with any statistical significance that these educational premiums are any less valuable for Blacks than for Whites. However, we do conclude that, all else being equal, White workers earn an average of 10-13% more than Black workers. Unfortunately, this earnings gap could possibly be explained by racial discrimination or by hiring practices that influence a Black person’s earnings in a disparately impactful way. But our modeling was obviously limited in that it failed to explain about 70% of the variation in earnings. There are undoubtedly many other relevant variables and other modeling methodologies that could offer a more accurate explanation of the earnings gap between White and Black workers.

In addition to the limitations of our econometric modeling, our research was also subject to the limitations of our data. For example, our data did not include any information on specific job types or individual degree fields. If two individuals were identical in every aspect, and both had the same doctoral degree, but one continued working in academia while the other performed consulting services at a private corporation, the latter would typically earn more. Substantial variations in earnings might also be explained by factors relating to self-selection bias that are impossible or unethical to measure and include in a practical data analysis. A more theoretical analysis would encompass and measure the variety of individuals who have attained higher degrees, but who are strongly influenced by personal factors such as intelligence, interpersonal skills or work ethic that may better explain the variation of their earnings.