**Labor Market Analysis**

**Introduction**

Education. It’s a common theory for people to believe that the higher the education an individual attains the higher the earnings that individual will earn. Is that a fact? Common knowledge amongst individuals is that education is important because it provides the basic information for individuals to provide improvement of innovation, as well as the ability for others to gain the knowledge to utilize innovation. Education allows the mind to develop further and help individuals to support the economy. Part of supporting the economy is protecting and securing exchanges with the new innovational advancement of financial technology. With even more benefits of education beyond those, education is seen as beyond important for society to function efficiently. However, does the change of earnings, increase the same or greater than the previous level of education’s change of earnings. From this information, we can analyze whether gaining higher education does lead to better earnings. Or, if there is a situation where attaining just your bachelor’s degree would have been more bang for your buck so to say, a higher increase in earnings compared to the amount of education that was taken.

One important factor that has become very prevalent in society in recent decades is gender equality. If both a male and a female have the same education and age and potential to give in a job, are they going to be attaining the same earnings. If they were to both attain the same earnings that would be known as gender equality. Gender equality is a key to societal development, taking in account which countries have issues with gender equality and which do not and how their economies are developing and flourishing. With gender equality upheld, equal rights for everyone are upheld and protected. However, does the educational level that two opposite genders attain have the same earnings difference percentage as other types of educational levels? Would attaining a master’s degree mean a smaller difference percentage between genders compared to the difference percentage between two individuals of opposite sex attaining nothing higher than a high school degree. From this information, we can analyze whether society has a bias amongst gender, and how much of a bias they have per education level. Although this bias is perceived as morally unethical by society, is it present in society at that time the census was taken.

To explore these two subjects, we will be gathering data and information relating to two research questions that we will focus on. In the state of Florida, what is the overall gender earnings gap? And how does it vary by educational attainment? We will approach these questions by gathering data from a census in Florida, cleaning the data, and using criteria, that is mentioned later in the paper, to develop an econometric model and estimated method section. If the earnings of individuals show a positive correlation to educational attainment that would indicate that more educational attainment leads to higher earnings. If the earning of individuals categorized by gender reveals a discrepancy between them that would indicate one of the genders is consistently attaining a higher earning than the other. Also separating by the educational factors as well, for a more in-depth analysis of the relationship between gender and educational attainment and earnings. This will lead to the analysis of earnings and educational attainment positively correlating and a premium for males with the same education level as females.

**Econometric Model and Estimation Method**

For our model, we will be using the ordinary least squares method (OLS) to estimate the unknown parameter in our equation for both male and female to best estimate the difference between observed values and predicted values. The Gauss-Markov theorem shows that, if our data fulfill certain requirements, OLS is the best linear unbiased estimator available. By using OLS and satisfying the necessary assumptions, we will be able to minimize the sum of squared residuals from the data. The model will then allow us to examine the strength of the relationship between our dependent variable (annual earnings) and our independent variables (gender, race/ethnicity, educational attainment).

To test the relationship between male and female’s annual earning in relation to their level of educational attainment, we added an interaction term to the model. Interaction terms allow the partial effect of an explanatory variable to depend on the level of another variable. Through the model, we are given the coefficient of the interaction terms. This allow us to observe the effect of changes in male and female’s annual earning in relation to their education attainment level.

To examine whether earnings vary by educational attainment and the premium for higher education vary by gender, we develop and estimate a log-level model which gave us the natural logarithm of annual earnings as our dependent variable. We will be using the log-level model to adjust the skewness in earnings distribution. Since we will be looking at earning distribution with percentage as the measurement, the log-level model will give us a more meaningful and reasonable interpretation since it is more appropriate to use as a measurement of percentage changes in the dependent variable given a change in 1 unit in the independent variable.

To assure that our model does not violate the assumption of homoskedasticity, because when homoskedasticity, which is one of the assumptions for OLS estimator, is violated, the OLS estimators may no longer be best linear unbiased estimators; hence, we would not be able to use regression analysis. Therefore, by running Breusch-Pagan-Godfrey Test or BPG Test, it helps us to test for heteroskedasticity of errors in regression, which is the violation of the assumption of homoskedasticity.

We used one measure of human capital, which is the level of educational attainment. To be more specific, our models distinguish among nine levels of educational attainment: HS represents individuals who have attained a high school degree, GED represents individuals who have attained an equivalent to a high school degree, SOME college represents individuals who have earned less than a year of college credit, ONE or more years of college credit represents individuals who have earned over a year of college credit, AA represents individuals who have attained their associates degree, BA represents individuals who have attained their bachelor’s degree, MS represents individuals who have attained their master’s degree, Pro beyond BA represents individuals who have attained professional experience after attaining a bachelor’s degree, PhD represents individuals who have attained their PhD. The models are provided below in the results. It is highly significant to include this variable in our model to help determine if and to what extent educational attainment was a premium for annual earnings. The results will allow us to see on average how certain educational attainments levels affect earnings. Additionally, given evidence of difference between educational attainment and annual earnings based on employment levels, our models include civilians who are employed at work. Additionally, researches show that marital status influences earnings, so our model controls for marital status to capture that relationship to better estimate the premiums on other variables. We also separate our model into subsamples by race/ethnicity. Similarly, to marital status, race/ethnicity influence the outcome of the labor market, therefore, our model also controls for race/ethnicity to better estimate the premiums on other variables.

**Data**

The data comes from the 2017 Census Bureau, from a survey conducted by the American Community Survey (ACS) and include most population and housing characteristics. Our analysis relies on data collected from the population characteristics rather than the housing characteristics. The total data collected for Florida Population Records was a sample number of 199,644 individuals, 286 variables were then measured for the data. The population survey includes questions pertaining to each individual’s earnings, labor market activity, and health status. For our project, we have imposed a sample selection criterion based on our research questions which will help reduce our sample size. We first limited our sample’s educational attainment level by individuals who completed high school, have GED credentials, or attained higher education after. This concludes that we did not included any educational attainment level lower that high school graduate in our sample. We then selected individuals who are civilian who are full-time employed at work and those who have positive earning. We identify full-time workers by selecting individuals who work over 20 hours per week. This sample selection criteria help us focus on working individuals and their earnings. We moved on to limit our sample by race by only including White, Black and Asian. This conclude that we did not included other race variables because their subsample sizes were too small for a proper analysis. Lastly, we excluded from our sample individuals that omitted incomplete information on all our selected variables. By applying these sample selection criteria, our sample size was narrow down to 77,957 observations with 36,778 male and 41,179 females.

**Table 1**. presents the descriptive statistics by gender for the individuals who fitted our sample selection criteria. It shows that female’s average earning of $44,570.90 which clearly lower than male average earning of $64,319.84. Even accounting for the new sample size through our sample selection criteria, the number of males with educational attainment level high school degree or higher that are employed at work and have positive earnings (41,179) is still higher than female (36,778).

**Table 1. Descriptive Statistics for Female and Male (Year 2017)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Female** | | **Male** | |
| **Parameter estimate** | | **Mean** | **St. Dev** | **Mean** | **St. Dev** |
| **Earnings** | | $44,570.69 | $47,819.41 | $64,319.84 | $78,409.76 |
| **Educational Attainment** | |  |  |  |  |
|  | **GED** | 0.03 | 0.17 | 0.04 | 0.2 |
|  | **High school** | 0.19 | 0.39 | 0.22 | 0.41 |
|  | **Some College** | 0.07 | 0.26 | 0.07 | 0.25 |
|  | **One or more college credits** | 0.14 | 0.35 | 0.14 | 0.35 |
|  | **Associate's Degree** | 0.13 | 0.34 | 0.1 | 0.3 |
|  | **Bachelor's Degree** | 0.24 | 0.43 | 0.21 | 0.41 |
|  | **Master's Degree** | 0.1 | 0.3 | 0.08 | 0.27 |
|  | **Professional beyond Bachelor's** | 0.02 | 0.15 | 0.03 | 0.18 |
|  | **Doctorate Degree** | 0.01 | 0.12 | 0.02 | 0.12 |
| **Race/Ethinicity** | |  |  |  |  |
|  | **White** | 0.78 | 0.42 | 0.81 | 0.39 |
|  | **Black** | 0.14 | 0.35 | 0.11 | 0.31 |
|  | **Asian** | 0.03 | 0.18 | 0.03 | 0.17 |
| **Married** | | 0.51 | 0.5 | 0.59 | 0.49 |
| **Never married** | | 0.26 | 0.44 | 0.27 | 0.44 |
| **Sample size** | | 36,778 | | 41,179 | |
|

**Breusch-Pagan-Godfrey Test**

**Figure 1.** BPG test result (Interaction with Female)

**Studentized Breusch-Pagan test**

data: logEarnings.Equation.interaction

BP = 4208.4, df = 21, p-value < 2.2e-16

From Figure 1, we could interpret a small p-value of (2.2e-16) which means at 1% significant level we reject the null hypothesis of homoskedasticity. Therefore, the statistical evidence implies that heteroskedasticity is present. A weakness of the BPG test is that it assumes the heteroskedasticity is a linear function of the independent variables. Failing to find evidence of heteroskedasticity with the BPG doesn’t rule out a nonlinear relationship between the independent variable(s) and the error variance. Additionally, the BPG test isn’t useful for determining how to correct or adjust the model for heteroskedasticity. Therefore, we will be generating robust standard error of the data. This analysis will give us the new standard error value that we can use to find the corrected t-value to help us adjust the model for heteroskedasticity. We will then compare our corrected t-values to critical value of t0.05 (1% significant level of 2 tail test).

**Results**

By examining **Table 2** of the Logarithmic Model for Genderbelow, the R-squared is equal to 0.2131, which can be interpreted as 21.31% of the variance in the earnings can be explained by educational attainment, race, and marital status. Referring back to our introduction section, we stated that there is a premium benefiting individuals with higher educational attainment. According to the results of the Logarithmic Model for Gender, the assumption we made holds true. For example, when controlling for other factors in the model, the earnings of an individual with a high school degree, on average, will increase by 24.29% compared to an individual without a high school degree. While an individual with a higher educational attainment level, for example, a bachelor's degree has an increase of 82.03% compared to an individual who does not have a bachelor’s degree. To go one further, an individual with a PhD will earn 125.07% more compared to an individual without a PhD. These results were highly statistically significant given the small value of their p-value.

These same types of increases in earnings with higher levels of education is also shown in Table 3 of the Logarithmic Model of the Estimate Earning Equation for Interaction with Female. To examine whether premium for higher education varies by gender, we developed and estimated a Logarithmic model with Female being the interaction term. From the result [(e-0.1837-0.002-1)\*100], we interpreted that controlling for all the other variables, on average we would expected a female with Bachelor’s Degree to have 16.95% lower earning than a male with Bachelor’s degree. We also expect that on average, a female with high school diploma would earn 7.75% less than a male with high school diploma, controlling for other variables.

The same results are observed for female with Professional Degree beyond Bachelor's where they earn on average 25.34% less than male, controlling for other variables. When we compare the gender earnings gap of individuals with Masters’ degree and Professional Degree beyond Bachelor, we could see that female are at a disadvantage compare to male. Female with Masters’ Degree earns 12% less than male with the same level of education. However, when female hold Professional Degree beyond Bachelor’s, the earning gap between female and male with the same level of education increase (from 12% to 25.34% earnings gap). This indicates that having a Professional Degree beyond Bachelor’s benefits male more than female. Additionally, we interpret that female with Doctorate Degree earns only 0.53% less than male with Doctorate. This shows us that the earning gap between female and male at the highest level of education is significantly small. **Table 4** provides a summary of Gender Earnings Gap for each level of education attainment, which demonstrates that there is clearly a pattern of gender earning gap between male and female throughout all educational attainment levels, with no high school degree being the reference category. We conclude that the earnings gap is in favor of male.

**Table 2. Logarithmic Model Estimated Earning Equation for Gender (Question 1)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficients** | | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |  |
| **(Intercept)** | | 9.759377 | 0.016837 | 579.622 | < 2e-16 | \*\*\* |
| **Marital Status (Not Married)** | |  |  |  |  |  |
|  | **Married** | 0.300062 | 0.006009 | 49.936 | < 2e-16 | \*\*\* |
| **Gender (Male)** | |  |  |  |  |  |
|  | **Female** | -0.271519 | 0.005943 | -45.69 | < 2e-16 | \*\*\* |
| **Educational Attainment (Less than High School)** | | |  |  |  |  |
|  | **High school** | 0.242974 | 0.012681 | 19.16 | < 2e-16 | \*\*\* |
|  | **GED** | 0.237901 | 0.018665 | 12.746 | < 2e-16 | \*\*\* |
|  | **Some College** | 0.390298 | 0.015489 | 25.199 | < 2e-16 | \*\*\* |
|  | **One or more college credits** | 0.423956 | 0.013411 | 31.614 | < 2e-16 | \*\*\* |
|  | **Associate degree** | 0.52752 | 0.013954 | 37.804 | < 2e-16 | \*\*\* |
|  | **Bachelor's Degree** | 0.820304 | 0.012576 | 65.227 | < 2e-16 | \*\*\* |
|  | **Master's Degree** | 1.047555 | 0.014825 | 70.662 | < 2e-16 | \*\*\* |
|  | **Professional beyond Bachelor's** | 1.450243 | 0.020555 | 70.554 | < 2e-16 | \*\*\* |
|  | **Doctorate Degree** | 1.25073 | 0.026536 | 47.133 | < 2e-16 | \*\*\* |
| **Race/Ethinicity (Some other race alone)** | |  |  |  |  |  |
|  | **White** | 0.19655 | 0.013727 | 14.319 | < 2e-16 | \*\*\* |
|  | **Black** | 0.028706 | 0.015738 | 1.824 | 0.0682 | . |
|  | **Asian** | 0.116045 | 0.021144 | 5.488 | 4.07E-08 | \*\*\* |

**Signif. codes**: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Residual standard error:** 0.8188 on 77942 degrees of freedom

**Multiple R-squared**: 0.2131, **Adjusted R-squared:** 0.213

**F-statistic:** 1508 on 14 and 77942 **DF, p-value:** < 2.2e-16

**Table 3. Logarithmic model of the Estimate Earning Equation for Interaction with Female after corrected for Heteroskedasticity (Question 2)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Coefficient** | **Estimate** | **Robust SE** | **t value** | **Pr(>|t|)** |  |
| **(Intercept)** | 9.8500 | 0.0149 | 660.3074 | < 2e-16 | \*\*\* |
| **Married** | 0.4294 | 0.0086 | 50.0658 | < 2e-16 | \*\*\* |
| **High school** | 0.2371 | 0.0163 | 14.5860 | < 2e-16 | \*\*\* |
| **GED** | 0.2147 | 0.0244 | 8.7820 | < 2e-16 | \*\*\* |
| **Some College** | 0.3695 | 0.0208 | 17.7596 | < 2e-16 | \*\*\* |
| **One or more college credits** | 0.4289 | 0.0176 | 24.3520 | < 2e-16 | \*\*\* |
| **Associate Degree** | 0.4815 | 0.0183 | 26.2639 | < 2e-16 | \*\*\* |
| **Bachelor's Degree** | 0.8480 | 0.0168 | 50.5378 | < 2e-16 | \*\*\* |
| **Master's Degree** | 1.0438 | 0.0207 | 50.5298 | < 2e-16 | \*\*\* |
| **Professional beyond Bachelor's** | 1.5117 | 0.0313 | 48.2266 | < 2e-16 | \*\*\* |
| **Doctorate Degree** | 1.1885 | 0.0374 | 31.7814 | < 2e-16 | \*\*\* |
| **Female** | -0.1837 | 0.0249 | -7.3910 | 3.27E-15 | \*\*\* |
| **I(Female \* Married)** | -0.2473 | 0.0120 | -20.6603 | < 2e-16 | \*\*\* |
| **I(Female \* GED)** | 0.1030 | 0.0393 | 2.6193 | 0.0643 | . |
| **I(Female \* High school)** | 0.0483 | 0.0272 | 1.7781 | 0.0072 | \*\* |
| **I(Female \* Some College)** | 0.0868 | 0.0326 | 2.6625 | 0.0059 | \*\* |
| **I(Female \* One or more college credits )** | 0.0241 | 0.0285 | 0.8454 | 0.3813 |  |
| **I(Female \* Associate's Degree)** | 0.1248 | 0.0288 | 4.3321 | 1.22E-05 | \*\*\* |
| **I(Female \* Bachelor's Degree)** | -0.0020 | 0.0271 | -0.0729 | 0.9390 |  |
| **I(Female \* Master's Degree)** | 0.0560 | 0.0310 | 1.8097 | 0.0635 | . |
| **I(Female \* Professional beyond Bachelor's)** | -0.1085 | 0.0484 | -2.2423 | 0.0101 | \* |
| **I(Female \* Doctorate Degree)** | 0.1784 | 0.0545 | 3.2739 | 0.0009 | \*\*\* |

**Signif. codes:** 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Residual standard error:** 1.871 on 77935 degrees of freedom

**Multiple R-squared:** 0.002572, **Adjusted R-squared:** 0.002303

**F-statistic:** 9.569 on 21 and 77935 **DF, p-value:** < 2.2e-16

**Table 4. Gender Earnings Gap for each level of Education Attainment**

|  |  |
| --- | --- |
| **Educational Attainment** | **Gender Earnings Gap** |
| **No High School Degree** | 16.78% |
| **High school** | 7.75% |
| **GED** | 12.66% |
| **Some College** | 9.24% |
| **One or more college credits** | 14.75% |
| **Associate's Degree** | 5.72% |
| **Bachelor's Degree** | 16.95% |
| **Master's Degree** | 12% |
| **Professional beyond Bachelor's** | 25.34% |
| **Doctorate Degree** | 0.53% |

**Conclusion**

The earnings per education level of an individual in the state of Florida, have a strong positive correlation.

Through our findings, we were able to conclude the overall gender earnings gap in Florida and how it varies by educational attainment. The results above allow us to determine that there is indeed a gendersgender earnings gap due to the strong correlation between annual earnings and the increase in educational attainment.

We also found evidence of gender biasness in the state of Florida, when we analyzed the difference in earnings per educational level based on an individual’s gender. Female individuals are earning less than male individuals who have attained the same educational level. We can infer the gender of an individual does affect the premium given for attainment of higher education in favor of male individuals in the state of Florida.

We also recognize that these results are limited to trusting that individuals answered the census honestly, and that the model did not capture the effect of all the variables that could influence earnings as there are so many factors that affect one’s earnings. Using fixed effects to control for the type of education level an individual has attained, we find that the growth between level of education attainment to earning is exponential. This indicates that the higher the education level, the higher of an increase in earnings is gained by the individual. Highschool education increases an individual’s earnings by an estimate of 6,564.65 dollars. Similarly, an associate degree increases by 16,549.40 dollars, a bachelor’s degree increases by 38,451.75 dollars, and PhD increases by 75,018.65 dollars. The increase of an individual’s earnings is a positive exponential relationship based on education level attainment. We can infer the level of education does affect the earnings of that individual in a positive manner in the state of Florida.

Highschool educational level attainment will lead to an estimated increase in earnings of 5,431.2 dollars for females and 7,698.1 dollars for males. After the attainment a four-year degree the average difference in earnings between female and male individuals is approximately 30,000 dollars in favor of male individuals.

**Contributions**:

**Yahya** wrote the R codes that consisted in uploading the data, filtering the data, running tests and creating models with their results summaries.

**Tiffany** imported data from R and formatting it into Excel. She also contributed to the writing sections of data, Econometric Model and Estimation Method, and editing all the paper’s writing.

**Allex** wrote the introduction, results, and conclusion including analysis of data and interpretation of econometrics concepts.