ECON 5300
Labor Market Analysis Paper
Dr. Bridget Hiedemann
Wei Li, Ziyu Jin, Florence Jiahui Yan

1. **Introduction**:

There is a general agreement that higher education pays. U.S. Labor Department statistics show that overall, college graduates today earn roughly 98% more per hour than people without a degree. For this paper, to get a more precise estimate, we will analyze women's earnings through raising two research questions as below:

1) How do women's earnings vary by educational attainment in Washington State?

2) Does the premium for higher educated women vary by race?

The purpose of this paper is to determine the correlation between women's earning, their educational levels, and earning's variation based on different races in Washington State.

Understanding the effect of educational attainment and race to women's earnings is important at least for two reasons. First, a prospective female student or a parent can use this information to plan a future career. Based on the result of this research, individuals will be able to analyze whether it is a good investment to spend time and money for bachelor's degrees or even higher educational attainments. Especially for those who heavily rely on students' loans for their educations, this research will provide a reasonable estimate for them to plan the future. Second, since the assumption of variance in earnings due to race has been in public mind for years, understanding the statistical relationship between women's earning and different races can help individuals pay attention to the possible discriminations that exist in the current workplace. For example, if pay inequality does exist between different races, an argument for fixing such disparity should be raised.

We use individual-level data in Washington State from the 2016 American Community Survey (ACS) to examine earning gaps between female individuals with different educational attainments and races. We divide educational levels into five groups, and group five categories for races. The dependent variable in each model is total personal earnings for women in WA. We run both linear and logarithmic regressions on the selected sample, as well as recalculating them based on the robust standard error we achieve from BPG Test after realizing the heteroscedasticity in the original model. For the result, we find that the correlation for educational attainment is positive to women's personal earning after controlling all other

variables. This finding is statistically significant and it explains that higher education can lead to higher pay for women in WA since the correlation of each educational attainment varies by races. For the second question, we discover like Asian women in WA have the highest premium for higher education within four races, and White women are estimated to have the lowest.

2. **Econometric Models and Estimation Methods**

2.1 Econometric Models

In our labor market analysis, the dependent variable in our model is the natural logarithm of total person's earnings of women in WA. We use the natural logarithm on the dependent variable for the following reasons: first, the distribution of the earnings is skewed to the right and logarithm can help correct; second, it makes more sense to talk about percentage differences in the dependent variable across categorical groups than unit (e.g., dollar) differences.

There are eight independent variables included fitting the multiple linear regression. Firstly, for educational attainment, we generalize educational attainment into six categories which are:

a. less than a high school degree (the reference category)
b. with a high school degree
c. with a college degree
d. with a professional degree
e. with a master degree
f. and with a doctor degree

Secondly, considering that earnings might be impacted by individuals' citizenships, we also include it as an explanatory variable (US-citizen as the reference category and non-US-citizen as the other). Given the evidence of the different impact on earnings brought by race and ethnicity, we distinguish the group of race and ethnicity as four categories which are:

a. non-Hispanic White (as the reference category)
b. non-Hispanic Black
c. non-Hispanic Asian
d. Hispanic observations

Besides, since females' earnings are also related to the marriage statutes regarding social researches, we include it as an independent variable in our model by using married statutes as the reference category. Additionally, as the employment status can also impact the earnings, we

distinguish this group as the present civilian employed work (reference category) and the present armed forces work.

We assume that experience can be an important independent variable to fit the earnings model. However, we don't have the actual working experience measurement. As an alternative, we utilize age as a measure of experience. Although age as a proxy might differ from the real experience, however, the latter is considered as an endogenous variable which might cause a violation of the linear regression assumption. Furthermore, we use both the quadratic form of age and age in years for our model. This could reduce the exponential effect of age brought to earnings. As we know in the workplace, the longer experience an individual has, the higher his wage rate is.

The last categorical variable is the work type since different corporations offer different pays. Usually, employees who work for private businesses get relatively higher pays compared to the same-level workers in other places. Therefore, we include three groups of work type which are:

a. Work for private business (reference category)
b. Work for government
c. Work for family business

Based on the analysis above, our model uses the form:

$$\text{Ln Earnings} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \varepsilon$$

In the equation, X represents all the independent variables, including educational attainment, citizenship, race and ethnicity, marital status, employment status, age, squared age, and work type.

To answer the second research question, we subset the data into 4 groups: White, Black, Asian and Hispanic observations. For each group, we use the same model but exclude other races.

2.2 Estimation Method

To explore the first research question, we use a general model for female individuals. It includes educational attainment, citizenship, race and ethnicity, marital status, employment status, age, squared age, and work type, the eight independent variables. These data are gained from the available population data set. As stated in the previous part, these variables are highly possible to impact earnings. So by putting them into the right side of the equation, we can see how the values in each variable change and affect the earnings.

However, there are some unobservable factors that might be correlated with some of the independent variables, such as age, work type. For example, the choice of work type might be correlated with an individual's family reason which is unable to see. So the expectation of the error term in the equation might not equal to zero. So parametric estimates of the coefficients might not work well and the standard errors of the coefficients may be underestimated. So we use robust standard errors to replace the ones gained from our model and then calculate the t statistics and p-values accordingly. This transformation gives us a better parametric estimate.
For the second research question, we estimate earnings separately for four races and ethnicity: non-Hispanic White, non-Hispanic Black, non-Hispanic Asian and Hispanic individuals. This approach allows us to easily compare the difference between the premium of the education level in different races and ethnicity. For example, a Hispanic individual might benefit less with a college degree than a non-Hispanic Asian, which means the premium of a college degree brings fewer earnings to a Hispanic, but more to an Asian.

3. **Data**

We use Washington State's female data from the 2016 ACS. This data set includes housing records as well as personal records. We will use personal data. There are 72383 records in this data set. Since the data set is large enough, we exclude the records that contain missing values. To better reflect the impact brought by educational attainment, we exclude individuals who cannot work due to age under 16 or not in the labor force. Since we are exploring earnings which means wage times hours worked, this definition indicates that earnings should not be negative, so we exclude the negative records. Also, we only focused on the full-time worker. In light of the possibility that a person who has self-care difficulty or any disabilities could negatively affect earnings, we exclude the data that have these records. Considering there might be people who chose not to work resulting in no earnings, we exclude the people who didn't work. For the races, we only focus on the following four races: non-Hispanic White, non-Hispanic Black, non-Hispanic Asian and Hispanic. After data cleaning, we have 8366 observations.

Based on the original educational data, we classified the individuals into 6 groups: with degrees lower than high school, with high school degree, with a college degree, with a professional degree, with master degree, and with doctor degree. Besides, we'd like to explore the impact of citizenship, marriage status, employment types on earnings. So we keep this information in our data set. Finally, we have a data set including 8366 observations on individuals and 8 columns containing useful information of these individuals. Below is the earnings distribution with different education/in different races.

Table 1: Earnings Distribution in Educational Attainment

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
College	3,347	63,827	55,071	20	35,000	78,000	850,000
Doctor	140	94,248	64,687	1,500	59,500	111,500	467,000
High School	3,081	41,865	34,229	4	25,000	50,000	483,000
Master	1,193	77,961	57,727	500	50,000	88,000	467,000
Professional	217	127,200	105,441	1,500	60,000	150,000	467,000
Under High School	388	29,525	28,106	350	16,000	35,000	383,000

Table 2: Earnings Distribution in 4 Races

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Asian	895	64,473	64,006	20	30,000	80,000	850,000
Black	203	48,075	35,257	300	26,000	60,000	214,000
Hispanic	725	37,842	34,897	4	19,500	45,000	467,000
White	6,543	60,061	53,566	50	32,000	72,000	583,000

By observing the data, we get the following facts.

1) The median of earnings is $45800, which is smaller than the mean $58316. So we can tell the earnings distribution is skewed to the right.
2) From Table 1, people with a professional degree has the highest earnings mean among the six educational attainments; also, the distribution of earnings of people with a professional degree spreads widest; the maximum of earnings is located in the group of people with a college degree.
3) From Table 2, Asian females averagely earn more than others; they also have the widest range of earnings; Black females earns least compared with other groups; all the four groups' earnings are skewed to the right, indicating there are extreme values in the right.

4. **Results**

4.1 WA women's earnings and their educational attainment

The related regression results of our model are presented in Table 3. Here we only present the coefficients of variables of our interests. And since all the coefficients are positive, we can say that the earning of a woman with a college degree is higher than those who have a high-school diploma. To have a more intuitive understanding of the results, we estimate the percentage

difference in earnings(in dollars) using $(e^{\beta} - 1) * 100$, where β is the coefficient of each variable in the table.

Table 3: Regression Results for Q1

Dependent variable: *ln*(earnings(in dollars))							
Intercept	Highschool Diploma	Bachelor's Degree	Professional Degree (beyond Bachelor's)	Master's Degree	Doctoral Degree	R-squared	Adjusted R-squared
8.104 (0.113)***	0.319 (0.049)***	0.714 (0.049)***	1.333 (0.07)***	0.922 (0.053)***	1.113 (0.073)***	0.256	0.256

Number of observations is 8366.
Standard errors are in parentheses.
*** significant at 0, ** significant at 1%, * significant at 5%

The computation results indicate that, on average, women in Washington state who have a high school diploma, a bachelor's degree, a professional degree beyond a bachelor's, a master's degree and a doctoral is associated with 37.58%, 104.21%, 279.24%, 151.43% and 204.35% more earnings in dollars than women who do not have a high school diploma respectively, controlling their citizenship, marriage status, work type and age. These results are all statistically significant, and the coefficient of determination shows that 25.6% of the variation in earnings is explained by our model.

4.2 Premium for higher educated women in WA and their race/ethnicity

To estimate the premium for higher education among each women's race and ethnicity, we summarize the regression results of our models in Table 4. Notice that we only presented the coefficients of variables of our interests, and the coefficient for variable Doctoral Degree is missing in that there is no record of a black woman who has a doctoral degree in our data set. Table 5 shows the premiums for higher education among each race and ethnicity that is

calculated in a similar way as in *4.1*.

Table 4: Regression Results for Q2

Dependent variable: *ln*(earnings(in dollars))				
	White	Black	Asian	Hispanic
Intercept	8.217	9.893	8.049	7.691
	(0.135)***	(0.625)***	(0.409)***	(0.419)***
Highschool Diploma	0.086	0.266	0.441	0.313
	(0.079)	(0.384)	(0.127)***	(0.081)***
Bachelor's Degree	0.474	0.8	0.951	0.602
	(0.079)***	(0.371)*	(0.131)***	(0.088)***
Professional Degree	1.075	1.488	1.566	1.44
(beyond Bachelor's)	(0.097)***	(0.453)***	(0.189)***	(0.377)***
Master's Degree	0.652	1.125	1.257	0.956
	(0.081)***	(0.364)**	(0.148)***	(0.14)***
Doctoral Degree	0.851		1.44	1.078
	(0.101)***		(0.159)***	(0.428)**
R-squared	0.232	0.268	0.257	0.269
Adjusted R-squared	0.231	0.226	0.247	0.257

Number of observations is 8366.
Standard errors are in parentheses.
*** significant at 0, ** significant at 1%, * significant at 5%

White women

On average, White women in Washington state who have a high school diploma, a bachelor's degree, a professional degree beyond a bachelor's, a master's degree and a doctoral is associated with 8.98%, 60.64%, 193%, 91.94% and 134.2% more earnings in dollars than women who do not have a high school diploma respectively, controlling their citizenship, marriage status, work type and age. These results are all statistically significant except for high school diploma(not significant) and bachelor's degree(only at 95% confidence level), and the coefficient of determination shows that 23.2% of the variation in earnings is explained by our model.

Black women

On average, Black women in Washington state who have a high school diploma, a bachelor's degree, a professional degree beyond a bachelor's, and a master's degree is associated with 30.47%, 122.55%, 342.82%, and 208.02% more earnings in dollars than women who do not

have a high school diploma respectively, controlling their citizenship, marriage status, work type and age. These results are all statistically significant except for high school diploma, and the coefficient of determination shows that 26.8% of the variation in earnings is explained by our model.

Asian women

On average, Asian women in Washington state who have a high school diploma, a bachelor's degree, a professional degree beyond a bachelor's, a master's degree and a doctoral is associated with 55.43%, 158.83%, 378.75%, 251.59% and 322.07% more earnings in dollars than women who do not have a high school diploma respectively, controlling their citizenship, marriage status, work type and age. These results are all statistically significant, and the coefficient of determination shows that 25.7% of the variation in earnings is explained by our model.

Hispanic women

On average, Hispanic women in Washington state who have a high school diploma, a bachelor's degree, a professional degree beyond a bachelor's, a master's degree and a doctoral is associated with 36.75%, 82.58%, 322.07%, 160.13% and 193.88% more earnings in dollars than women who do not have a high school diploma respectively, controlling their citizenship, marriage status, work type and age. These results are all statistically significant, and the coefficient of determination shows that 26.9% of the variation in earnings is explained by our model.

Table 5: Premium for Higher Education

Percentage difference of earnings (in dollars) relative to not having a high school diploma				
	White	Black	Asian	Hispanic
Highschool Diploma	8.98%	30.47%	55.43%	36.75%
Bachelor's Degree	60.64%	122.55%	158.83%	82.58%
Professional Degree (beyond Bachelor's)	193%	342.82%	378.75%	322.07%
Master's Degree	91.94%	208.02%	251.49%	160.13%
Doctoral Degree	134.2%		322.07%	193.88%

If we compare the results in Table 5, it is clear that Asian women are estimated to have the highest premium for higher education after the attainment of all degree types and White women are estimated to have the lowest. In between are Black women and Hispanic women, where

Black women have a relatively higher premium than Hispanic women except for the attainment of high school diploma and doctoral degree.

5. Conclusion

The analysis in this paper shows that in Washington State, higher educational attainment within women has a positive effect on personal earnings. The effect remains statistically significant after controlling all other variables. According to this, we can draw a conclusion that promoting higher education can have a positive impact on the overall earnings in WA.

Professional degree and doctoral degree have the strongest effect on improving earnings, which in logical in reality that people who accept higher education or have stronger techniques in a specific industry are supposed to earn more. At the same time, the result also suggests that education might be the cause of earnings disparities and hence it indicates a need for policies that are more effective in eliminating educational inequality. With regards to the second question, based on our analysis, we conclude that in Washington State, Asian women benefit the most from higher education in earnings, Black women the second, Hispanic women the third, and White women the least. Whether such situation is an indication of earning gaps among races and ethnicities is unable for our model to determine, but a possible explanation to it can be that White women without a high school degree earn more relative to women of other races, causing the premium of higher education for them to be lower. In addition, although we did not detect the effect of the attainment of a doctoral degree to Black women's earnings, we expect it to be positive as well since it shows so for women of all other races. Overall, pursuing higher education is a good strategy for women of all races in Washington State to improve their earnings, and gaining a professional and a doctoral degree make the most significant difference.

Although this paper provides the impact on earnings brought by educational attainment, races and ethnicity, etc, the conclusion above is subject to a number of limitations. First, the data are self-reported and hence the information we get might be subjective. Second, the data may not include enough samples to represent every woman in Washington State. Lastly, due to the nature of regression models, our model is not a perfect fit for the entire dataset, and therefore errors are predictable.

Contribution

Florence: Data Preparation and Cleaning, EDA, Sample Criteria, Paper Writing(Introduction, Econometric Models, and Conclusion)
Wei: Data Preparation and Cleaning, Sample Selection, New Variable Generation, Models Testing and Detection, Paper Writing(Econometric Models, Estimation models, Data)
Ziyu: Data Preparation and Cleaning, Sample Selection, Model Specification and Diagnosis, Code Modification and Finalization, Paper Writing(Results and Conclusion)