Labor Market Analysis

District of Columbia

Barkha Sharma
Sri Lakshmi Mallipudi
Jiayang Wu
Amrapali Samanta

Introduction:

There has been a long-standing discussion about the relationship between gender and wages, and most of them drew similar conclusions based on factual evidence: Female workers usually get lower wages than male workers on average. The fact is that occupations that have a larger proportion of female labor will have lower average wages than those with more male labor (Albak and Thomsen 2014), and work done by women will be recognized as less valuable than work done by men which directly caused partiality of wages between two genders of workers(Cohen and Huffman 2007).

The importance of understanding the earnings gap by gender lies in the fact that it could help researchers foresee the effects of occupational segregation, or the guiding of women and men into different types of industries and jobs based on gender norms and expectations. The gap has a compounding effect that results in a woman's reduced earning capacity over her lifetime, less progress as compared to men, and in turn less accumulated money while retiring. Several factors may contribute to disparities in earnings, and in this paper, we aim to discuss how the earnings gap is being affected by gender and educational attainment, while also considering the marital status and the class of workers as controlling factors.

Research questions:

1. In the state District of Columbia, how does a person's wage vary by gender?
2. In the state District of Columbia, how does a person's wage vary by educational attainment?

Based on the data from the 2021 American Community Survey (ACS), we estimate the gender gap in wages of the past 12 months as well as the effect of educational attainment on the labor force in the Columbia district of Washington D.C. It is observed that based on a multi-linear estimated regression model, controlling all other factors, females tend to earn less than men on average. However, if females are government employees they tend to earn more on average. Also, in general, people having at least a bachelor's degree earn more on average. Based on the estimated results, there could be some essential policy changes like employing more women in the government sectors or helping people in the district to earn at least a Bachelor's degree which could possibly help to minimize the earnings gap between men and women and ensure better economic security for all workers.

Econometric Model and Estimation Method:

For an economic Model the wages for the past 12 months can be derived as a function of various factors as below:

Wages in past 12 months($) = f(Age, Gender, Marital Status, Weekly hours worked, Educational attainment, Type of Employment)

The below specific econometric model is considered based on different parameters determined by the economic theory and data considerations, with 'u' representing unobserved factors.

$\ln(\text{Wages past 12 months}(\$)) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Female} + \beta_3 \text{Married} + \beta_4 \text{Weekly hours worked} + \beta_5 \text{High School or some college} + \beta_6 \text{Associates or Bachelor's Degree} + \beta_7 \text{Master's Degree} + \beta_8 \text{Professional and Doctorate Degree} + \beta_9 \text{Age} * \text{Age} + \beta_{10} \text{Female.government employee} + \beta_{11} \text{Private Employee} + \beta_{12} \text{Government employee} + u$

The econometric model above is a semi-elastic multiple regression model using the ordinary least squares(OLS) method. The model attempts to explain the logarithmic value of wages past 12 months of an individual based on various demographic, gender, and employment-related factors. The response variable is the natural log of wages past 12 months while the predictor variables chosen for the model specification include age, gender, marital status, weekly hours worked, education level, and class of worker. The categorical factors like marital status, sex, education level of the individual, and the class of worker are included in the model as a set of dummy variables, as we want to estimate the wage gap by gender and educational attainment where the marital status and type of employment could play a major role.

The model uses a linear regression approach with a combination of non-linear terms such as taking the logarithmic transformation of the dependent variable wages past 12 months to track a single unit change of independent variables with a constant change in wages, and including squared term of Age to capture any non-linear relationship with wages. The inclusion of non-linear and interaction terms allows for a more nuanced and accurate understanding of the factors that influence an individual's wage based on gender. To get an accurate understanding of the change in the estimated slope, we estimate the interaction of females with Government employees as a new interaction variable called Female.Government. In this analysis, the model provides a comprehensive approach to the understanding of the relationship between the independent and logarithmic value of Wages past 12 months. The use of linear, non-linear, and interaction terms and OLS regression methodology provides a robust and flexible framework while explaining the relationship between the dependent and independent variables.

Data:

The data used is sourced from the **2021 American Community Survey (ACS)** 1-year PUMS.

The number of observations under analysis is **4281.**

Sample selection criteria of dataset: Based on the preliminary exploratory data analysis, sample selection criteria have been imposed to do meaningful regression analysis. We have restricted the **age** of a person in the dataset above 16 years to include only people who are eligible to work, which helps us evaluate the variation in the earnings of a person by gender. The predictor variable '**Wages or Salary in the past 12 months**' are restricted to people with wages greater than zero because we want our model to evaluate the variation in the wages of both part-time and full-time workers. The other quantifying variable '**Usual hours worked per week past 12 months**' by a person is restricted to having only greater

than zero hours to consider people who are actually eligible to earn and are currently working in the past 12 months with non-zero hours.

Categorization of qualitative variables: The categorical variables are grouped as a dummy set of variables to obtain comparisons that are most meaningful for testing the hypotheses. **Sex** is categorized as Female and Male. **Marital Status is** categorized as 'Married' and 'Unmarried' where Unmarried is grouped as Widowed, Divorced, Unmarried, or Separated. **Educational Attainment** is categorized as before the High School diploma level (i.e., 1- 15) under the category 'No High School Diploma,' from High school diploma to one/more years of college credit but no degree level (i.e., 16 - 19) under the category 'High School Diploma or Some college', from Associate degree to Bachelor's degree(i.e., 20 - 21) under the category 'Associate or Bachelor's degrees, Master's (i.e, 22) under the category 'Master's Degree' and from Professional to Doctorate degree(i.e, 23 - 24) under the category 'Professional or Doctorate degree'. The final categories considered for our analysis are High School Diploma or Some College, Associate degree or Bachelor's Degree, Master's Degree, Professional Degree or Doctorate Degree. The **class of workers is** grouped as local, state, and federal government employees together as 'Government Employees', Employees of Private for-profit and non-profit organizations as 'Private Employees', and Self-employed in their own business or family business as 'Self-employed' (reference category).

Descriptive Statistics by gender after imposing sample selection criteria:

Female: Number of observations: 2257

Variable	Minimum	Maximum	Mean	Median	Standard Deviation
Wage past 12 months(in US Dollars)	30	787000	86347.23	70000	98936.11
Age(in years)	16	83	38.11	35	14.20
Usual Weekly Hours worked (in hours)	1	99	38.52	40	13.08
Married	0	1	0.31	0	0.46
Associate or Bachelor's Degree	0	1	0.74	1	0.44
Master's Degree	0	1	0.42	0	0.49
Professional or Doctorate Degree	0	1	0.15	0	0.36
Government Employee	0	1	0.28	0	0.45
Female and government employee	0	1	0.257	0	0.437

Male: **Number of observations:** 2024

Variable	Minimum	Maximum	Mean	Median	Standard Deviation
Wage past 12 months(in US dollars)	30	787000	104417.31	80000	115122.78
Age(in years)	16	94	39.83	37	14.05
Usual Weekly Hours worked(in hours)	1	99	40.37	40	12.52
Married	0	1	0.41	0	0.49
Associate or Bachelor's Degree	0	1	0.75	1	0.44
Master's Degree	0	1	0.42	1	0.49
Professional or Doctorate Degree	0	1	0.16	0	0.37
Government Employee	0	1	0.28	0	0.45

Based on the descriptive statistics, we get to see a striking result that the average wage of the female is $20,000 less than male, however maximum and minimum values of wage are identical which excluded the possible impact of extreme value to averages. Other interesting facts from the sample are that on average males worked till the age of 94 whereas the females worked till 83 years and the marriage rate of females is 10% lower than males by 2021. In the sample, males worked approximately two hours more per week than females.

Results:

Table - Estimated Multiple Regression Model

Dependent Variable: Wages or salary in the past 12 months(in US Dollars)

Coefficients	Estimate		Std. error
Intercept	4.953	***	0.155
Age(years)	0.131	***	0.006
Age * Age(years)	-0.001	***	0.000
Female	-0.091	**	0.033
Married	0.160	***	0.032
Weekly hours worked(hours)	0.046	***	0.001
Educational attainment			
High School Diploma or Some College	0.195	*	0.082
Associate or Bachelor's Degree	0.887	***	0.041
Master's Degree	0.086	*	0.038
Professional or Doctorate Degree	0.168	***	0.045
Class of worker			
Private Employee (For Profit, Non Profit)	0.189	**	0.073
Government employee (Local, State, Federal)	0.217	**	0.082
Female.Government employee	0.104	.	0.062
Sample size	4281		
Degrees of freedom	4268		
R-Squared	0.609		
Adjusted R-Squared	0.608		
F-statistic	555.058***		

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

From the above regression model results, we can notice that a person's wages vary by gender and educational attainment. Males earn more than females, and persons with an associate's or bachelor's degree earn more than persons with a lower-level degree. But surprisingly, we noticed that in Washington DC, the magnitude of wages of people having a Master's degree is lower and less significant, Professional or Doctorate degree also being lower in magnitude than people with an associate's or bachelor's degree. The reason could be that most demanded jobs in the D.C. area usually do not require a Master's/Doctorate to qualify for the job position. Once the job is achieved, advancement in the career can be obtained through tenure and experience without a need for post-graduation. Interestingly, from the female government employee interaction term, we noticed that the wages of female government employees were more than male government employees, narrowing the gender wage gap.

All the variable coefficients are statistically significant at the 5% level except the interaction term female government employee, which is statistically significant at the 10% level. We were specifically interested in including female government employee interaction since it addresses the central aspect of our research question focusing on the gender wage gap.

Critical variables in our model addressing the research question are female, married, educational attainment, and government employee, and all of them are positively correlated with wages except for gender. As shown in the table, there is a negative relationship between wages and gender, which means that by controlling for age, marital status, educational attainment, weekly hours worked, and the worker class, female workers earn about 8.66% less than males resulting in a wage gap by gender. The coefficient on 'Age*Age' indicates that as the person's age increases, predicted wages increase at a decreasing rate until 52 years of age approx. and then decrease at an increasing rate. The coefficient on 'married' indicates that holding all else equal, married people earn approximately 17.36% more than unmarried people. The coefficient on 'Associate or Bachelor's degree' indicates that holding all else equal, on average, people with an associate or bachelor's degree earn approx. 142.88% more than those without a High school diploma, which intuitively makes sense because the higher the education level, the higher the wages a person is likely to earn. The coefficient on 'Government Employee' indicates that holding all else equal, an employee of local/state/federal governments earns approximately 24.18% more than those who are self-employed in their own business or family business. The coefficient on 'Female Government employee' indicates that holding all else equal, female government employees earn approximately 37.77% more than self-employed women in their own business or family business and approximately 1.31% more than male government employees.

The Adjusted R-squared (coefficient of determination) for the model is 0.608, which indicates that the estimated regression model explains about 60.8% of the variation in the natural log of wages in the past 12 months with age, gender, marital status, weekly hours worked, educational attainment and the class of worker as independent variables. The overall significance of the model(F-statistic) is highly statistically significant at a 0.1% level.

Conclusions:

From the demographic survey in the DC Washington area, we observe that women are still being paid less at every education level compared to men in the 2020s, as an indication of the structural problem concealed inside the gender distribution of occupations. Despite the consideration of educational attainment for the wage gap, the analysis shows that education does not appear to make a major difference. However, a possible change that could be established to bridge this gap is to increase the proportion of female workers in the government sector. There could be various reasons for this gap, like females focusing more on household work than careers, taking maternity breaks, or putting more hours into childcare. We can also incorporate these factors for future estimations.

The conclusions above are subject to some limitations. We cannot generalize these estimated results for the other states in the US. There could still be some other variables in the error terms that influence the dependent variable and correlate to other independent variables leading to omitted variable bias. For example, we believe that experience of an individual influences the wage and is correlated to, say, age in the model. Still, the same is not included in our model due to the unavailability of that variable in the population. In future research, we can include more data samples to produce results closer to an actual population. Also, it needs to be clarified how the ACS has sourced the data considered for analysis.

It is still evident that the gender wage gap is real. Sometimes people might assume that this gap is not evidence of discrimination, but could be a statistical failure to adjust for various factors. But in conclusion, we can say though the earnings gap has decreased over the years, there is room for improvement as every corporate sector should work on policy strategies to narrow the gender wage gap.

Contributions:

Worked as a team to finalize our research question, variables of interest, sample selection criteria, regression model analysis, and results.

Specific Contributions are as follows:
Barkha Sharma worked on the R coding part to get regression results and on the 'Econometric model and Estimation method section' for the final paper.
Sri Lakshmi Mallipudi on establishing the sample data by imposing meaningful sample selection criteria based on exploratory data analysis and worked on the 'Results' section for the final paper.
Jiayang Wu worked on the effective literature review focusing on the research question and the 'Conclusions and limitations' section for the final paper.
Amrapali Samanta ensured the regression analysis conformity with the research topic, worked on the 'Introduction' and 'Data' sections, and proofread the final paper.

References:

Albæk, K., & Thomsen, L. B. "Decomposing Wage Distributions on a Large Data Set - A Quantile Regression Analysis of the Gender Wage Gap." June. 2014, https://pure.vive.dk/ws/files/232486/WP_06_2014. Pdf

Cohen, P. N., & Huffman, M. L. "Working for the Woman? Female Managers and the Gender Wage Gap." American Sociological Review, 72(5), 681–704. 2007. https://doi.org/10.1177/000312240707200502

Word count: Excluding tables and references, the total word count is 2272.