Gender Wage Gap Study

ECO480

May 2022

The gender wage gap is still an issue today. According to the U.S. Census Bureau, full-time working women earned 82% of what their male counterparts earned. Although a large part of this gap has already been explained by measurable factors like educational attainment, occupational segregation and work experience, other factors which are difficult to measure such as gender discrimination also contribute to the wage gap. The gender wage gap is very much real but how much of it depends on other demographic characteristics like race, occupations, education, and others? After running many multiple regression models and adding many needed dummy variables, I concluded that on average, females made significantly less than males.

The first thing I did was set some constraints on the data sample. I dropped all data without sex as either male or female because in a study on gender, this data is not applicable. Another restriction I made on the data was to drop all data with no income because in a study on income and wages, this data is also irrelevant. On top of this, I dropped all data with an income of less than $15,000 being that it is the annual minimum wage earnings. I also added restrictions for age, dropping all data under 18, being that the base legal age in the United States is 18. When I ran my regression model, I also set constrains for only including full-time workers, and people who worked for a wage or salary. This meant I did not include people who worked part-time or was self-employed. I decided not to include such data as part-time and self-employed workers' wages are volatile and not a good basis for a study on income/wage differences.

After all these restrictions, I had a sample with 51,690 people. The average income/wage was $68,085.24, the average age was 42.9715, and the average hours worked per week was 42.7936.

```
. sum incwage age uhrswork1 if inrange(classwkr,21,28) & inrange(uhrswork1,40,99) & age >=18

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
     incwage |     51,690    68085.24    75153.06      15001    1599999
         age |     51,690     42.9715     12.4969         18         85
   uhrswork1 |     51,690     42.7936    6.443367         40         99
```

```
. sum incwage age uhrswork1 if inrange(classwkr,21,28) & inrange(uhrswork1,40,99) & age >=18 & se
> x1 == 1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
     incwage |     29,180    75484.74    83990.17      15100    1599999
         age |     29,180     42.9416    12.51452         18         85
   uhrswork1 |     29,180    43.51189    7.202483         40         99
```

```
. sum incwage age uhrswork1 if inrange(classwkr,21,28) & inrange(uhrswork1,40,99) & age >=18 & se
> x1 == 0

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
     incwage |     22,510    58493.17    60515.57      15001    1314999
         age |     22,510    43.01026    12.47418         18         85
   uhrswork1 |     22,510    41.86246     5.15311         40         99
```

However, when looking at the statistics while controlling for gender, we can see that the average male income/wage is $75,484.74. On the other hand, the average for females was $58,493.17 while average age and hours worked remained similar. Are employers being blatantly sexist? For the most part, we know this to be untrue. So, what could be the reason behind this large wage gap? To try to understand this further, I used many different variables which could affect this outcome and ran a multiple regression model.

Before I could run my regression model, I had to create dummy variables for gender, race, education levels, class of worker (i.e. private, government, etc.), and firm sizes. I also decided on creating a dummy variable for region instead of something like state or city as the latter two would be too time intensive. Similarly, I only created dummy variables for the races with the 4 highest frequencies being White, Black, American Indian/Aleut/Eskimo, and Asian while leaving the constant to be all other mixed races.

One problem I thought of immediately regarding omitted variable bias, was that males and females might tend to gravitate towards different jobs and industries resulting in different wages. To control for this, I added dummy variables for occupations with the 5 highest frequency differences between males and females. These occupations were drivers, elementary and middle school teachers, registered nurses, secretaries and administrative assistants, and construction laborers. Finally, I created variables for age after 18 and overtime hours worked for easier interpretation of the regression table.

The model I came up with was a regression on income/wage while controlling for gender, age, race, hours worked, education level, region, class of worker, firm size, and jobs with the highest male/female differences. As mentioned earlier, I also added constraints on the data for excluding self-employed and part-time individuals.

```
. reg incwage sex1 age18 white black amind asian overtime highschoolgrad somecollege assdocc assd
> ap bachelor master prof doctor northeast south west private federalgov localgov medfirm largefi
> rm malejob1 malejob2 femalejob1 femalejob2 femalejob3 if inrange(classwkr,21,28) & inrange(uhrs
> work1,40,99) & forces !=1 & age >= 18

      Source |       SS           df       MS      Number of obs   =    51,690
-------------+----------------------------------   F(28, 51661)    =    365.94
       Model |  4.8319e+13        28  1.7257e+12   Prob > F        =    0.0000
    Residual |  2.4362e+14    51,661  4.7157e+09   R-squared       =    0.1655
-------------+----------------------------------   Adj R-squared   =    0.1651
       Total |  2.9194e+14    51,689  5.6480e+09   Root MSE        =     68671
```

incwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
sex1	18417.61	642.9937	28.64	0.000	17157.34	19677.89
age18	631.5733	24.41024	25.87	0.000	583.729	679.4176
white	4210.413	1984.491	2.12	0.034	320.7904	8100.036
black	-5670.269	2170.628	-2.61	0.009	-9924.721	-1415.816
amind	-2308.757	3366.166	-0.69	0.493	-8906.476	4288.962
asian	5207.886	2261.163	2.30	0.021	775.9843	9639.787
overtime	1245.121	48.09766	25.89	0.000	1150.849	1339.393
highschoolgrad	11857.7	1376.148	8.62	0.000	9160.437	14554.96
somecollege	18784.02	1472.688	12.75	0.000	15897.54	21670.5
assdocc	20620.1	1852.4	11.13	0.000	16989.38	24250.83
assdap	24957.28	1760.979	14.17	0.000	21505.75	28408.82
bachelor	45893.09	1401.275	32.75	0.000	43146.57	48639.6
master	64794.46	1567.647	41.33	0.000	61721.86	67867.07
prof	122919	2672.893	45.99	0.000	117680.1	128157.9
doctor	92922.4	2302.036	40.37	0.000	88410.39	97434.41
northeast	6460.69	1031.489	6.26	0.000	4438.962	8482.418
south	1374.803	858.0466	1.60	0.109	-306.9768	3056.583
west	5043.158	906.0233	5.57	0.000	3267.344	6818.973
private	19214.74	1367.5	14.05	0.000	16534.42	21895.05
federalgov	20979.65	1984.26	10.57	0.000	17090.48	24868.82
localgov	6691.137	1695.344	3.95	0.000	3368.246	10014.03
medfirm	6196.992	863.8407	7.17	0.000	4503.856	7890.128
largefirm	9987.28	733.2241	13.62	0.000	8550.153	11424.41
malejob1	-10957.88	1983.924	-5.52	0.000	-14846.39	-7069.369
malejob2	-10708.82	2877.01	-3.72	0.000	-16347.79	-5069.853
femalejob1	-21701.33	1913.851	-11.34	0.000	-25452.5	-17950.16
femalejob2	1959.703	2245.325	0.87	0.383	-2441.157	6360.563
femalejob3	-5144.45	2299.093	-2.24	0.025	-9650.696	-638.2051
_cons	-22304.55	2895.757	-7.70	0.000	-27980.26	-16628.84

The full regression I used was “reg incwage sex1 age18 white black amind asian overtime highschoolgrad somecollege assdocc assdap bachelor master prof doctor northeast south west private federalgov localgov medfirm largefirm malejob1 malejob2 femalejob1 femalejob2 femalejob3 if inrange(classwkr,21,28) & inrange(uhrswork1,40,99) & forces !=1 & age >= 18”.

I left out a variable for region (Midwest), class of work (state government), and firm size (small firm) to avoid multicollinearity as these were dummy variables and would be included as the constant. The constant being a female, 18 years old, not white/black/americanindian/asian, did not work overtime, did not graduate high school, lives in the Midwest, works for the state government, works for a small firm, and does not work at a job with a high male/female difference.

After running the regression, the dummy variable for gender (sex1), had coefficient of 18417.61. This meant that males made an estimated $18,417.61 more than females just because of gender. Age had a coefficient of 631.5733, meaning income/wage increased by an estimated $631.57 per year of age. For the race variables, Asians had the highest expected increase of income/wage being $5207.886 more than the constant and Blacks had the highest expected decrease of income/wage being $5670.269 less than the constant. Comparatively, increase of income/wage by race from highest to lowest was Asian, White, Other (constant), American Indian, and Black. Overtime had a coefficient of 1245.121 meaning each hour worked after 40 had an expected increase of $1245.12. For education levels, high school graduates made $11,857.70

more than non-high school graduates, bachelor's degree graduates made $45,893.09 more than non-high school graduates, master's degree graduates made $64,794.46 more than non-high school graduates and professional school degree graduates had the highest increase being $122,919 more than non-high school graduates. For regions, living in the northeast had the highest increase in income/wage with $6460.69 and west, south, and midwest following in descending order. In terms of class of work, federal government workers made $20,979.65 more than state government workers while private firm and local government workers were second and third. Lastly, when it came to firm size, large firms had the highest increase in income/wage being $9987.28 more than small firms.

In conclusion, the gender wage gap can be seen in the difference of $18,417.61 for males and females. Considering the average income/wage in this sample was $68,085.24, this gap is larger than the 82% which the US Census Bureau published. However, this is expected as there were many other variables I could not control for in my regression model.

<u>Works Cited</u>

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 7.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D030.V7.0. See, https://cps.ipums.org/cps/citation.shtml (viewed June 29, 2020)

Barroso, Amanda, and Anna Brown. "Gender Pay Gap in U.S. Held Steady in 2020." *Pew Research Center*, Pew Research Center, 25 May 2021, https://www.pewresearch.org/fact-tank/2021/05/25/gender-pay-gap-facts/.