DA2 - Gender Wage Gap Analysis

Adam Kovacs, Nam Son Nguyen

11/22/2021

Contents

Filtering for occupation

Create new variables

Descriptive statistics

Table 1: Wage Metrics of Advertising and Sales Managers

	Mean	Median	SD	Min	Max	P05	P95
Weekly earnings	1489.93	1346.15	776.55	1.00	2884.61	440.00	2884.61
Weekly hours worked	44.50	40.00	9.05	2.00	99.00	35.00	60.00
Hourly wage	33.33	30.00	16.56	0.03	100.00	11.52	64.10

What we can infer from the descriptives table is that there are couple fo individuals who do overtime beyond the average 40 hours per week (99 at maximum, which is strangely specific), thus them being at the long tail part of the distribution makes the sample distribution right-skewed. Regarding the standardized wage KPI, hourly wage, we can observe a relatively high average dispersion between data points (standard deviation is USD 16.56). Anomalies can also be detected after taking the range of hourly wages into account, as according to our sample, there can be the case that someone earns USD 0.03 per hour. We hardly believe that its possible in the US, therefore we will exclude extreme datapoint having lower than USD 1 as their hourly wage. Nothing extremely unusual can be detected when observing the 5th and 95th percentiles.

Gender Wage Gap

Now that everything is in order, we can start to run our regressions to show whether the gender wage gap exists in this particular profession, and if so, what is the magnitude of the gap, what variables are the

confounders in this relationship. As a baseline model, we have fitted a simple linear semi-elasticity regression with log wage being at the LHS and female dummy on the RHS.

Then, we started to extend our model with first including the level of education factor (7th or 8th grade being the baseline), followed by the age and squared value of age.

We've decided to present our results with the summary table below.

A female is expected to earn on average 25.5% less than a male in these professions This wage gap is significant at 5% significance level (even at 1%).

This gap tend to be slightly narrower if we include the level of educations, but the difference remains highly statistically significant. The additional variable also decreased the standard error of our female coefficient, all else being equal and multiplied our goodness-of-fit (17.2% of the log hourly wage's variance can be explained by the RHS). Therefore, we can say that the level of education is meaningful in the setup, as all of its coefficients are significant at the 5% level.

Moving onto Model 2 where we also included the age of employees, which had a marginal part in mediating the difference between sexes below 20%, ceteris paribus, while being a significant confounder. We also managed to further improve the R2 to 0.203.

Lastly, our last model have taken into consideration the possibility that the relationship between age and log hourly wake is non-linear. Hence, we included age as a polynomial of order two in the RHS. It seems as thought it was a reasonable decision by just looking at the significance of the variable and how much further it improved our fit (R2=0.235). As far as gender wage gap is concerned, a female tend to have averagely 19% lower hourly wage than a male working as a manager in Advertising & Sales.

We could go on-and-on in including possible mediators to our model, but considering that we have a sample just greater than 1,000 observations, we should carefully evaluate how much further we want to extend what we have. There are also unobservable variables such as effort, education quality, etc. which education level can only partly proxy for. Our models indicate that there is indeed a significant relationship between gender and wage, but the gap narrows as we control for other variables.

Table 2: Gender Wage Gap | Level of Education

	Unconditional	Level-Level	Conditional	Interaction
(Intercept)	3.50 (0.02)**	50.29 (7.40)**	3.79 (0.22)**	3.57 (0.30)**
Female	−0.26 (0.03)**	−6.69 (0.93)**	−0.23 (0.03)**	0.33 (0.37)
No high school		−32.50 (7.89)**	−1.03 (0.27)**	−0.97 (0.36)**
High school graduate		−22.70 (7.48)**	−0.61 (0.22)**	−0.36 (0.31)
College no degree		−20.17 (7.49)**	−0.54 (0.22)*	−0.27 (0.31)
Associate degree vocational		−25.46 (7.68)**	−0.71 (0.24)**	−0.41 (0.32)
Associate degree academic		−17.59 (7.70)*	−0.44 (0.23)	−0.16 (0.32)
Bachelor's degree		−11.64 (7.42)	−0.23 (0.22)	−0.01 (0.30)
Master's degree		−8.20 (7.48)	−0.12 (0.22)	0.08 (0.30)
Professional school		−4.46 (9.17)	−0.01 (0.25)	0.43 (0.31)
F x No high school				−0.14 (0.44)
F x High school graduate				−0.62 (0.39)
F x College no degree				−0.64 (0.39)
F x Associate degree vocational				−0.72 (0.42)
F x Associate degree academic				−0.65 (0.41)
F x Bachelor's degree				−0.55 (0.38)
F x Master's degree				−0.51 (0.38)
F x Professional school				−1.10 (0.39)**
Num.Obs.	1090	1090	1090	1090
R2	0.057	0.150	0.171	0.178
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust

Prediction

Predicted Log Hourly Earnings

Fitted values from our three regression models