Report on gender wage gap in legal jobs

Péter ENDES-NAGY

Introduction

This is a report on *gender wage gap in legal jobs*. I investigated the magnitude of the gender wage gap in legal occupations and how other features, like occupation, level of education and age influences the wage gap.

Data

I used data from the 2014's Current Population Survey (CPS), that is a monthly survey of about 60,000 households in the US. A detailed description of the Survey and the methodology is available here: https://osf.io/uqe8z/

I was interested in gender wage gap in legal occupations, so I narrowed down my sample to this field, using `2100` (Lawyers, Judges, magistrates, and other judicial workers), `2105` (Judicial law clerks), `2145` (Paralegals and legal assistant) and `2160` (Miscellaneous legal support workers) census codes. It seems that there is a hierarchy among these occupations, the `2105-2145-2160` occupations have a rather supporting role, so I created an `occupation` variable that takes 1 for `2100` and 0 for the rest.

The legal field requires a degree, especially the `2100` occupations (at least a Professional degree), so the level of education was also filtered for Bachelor's degree and above.

During the preparation phase, hourly wages were also constructed for the sake of comparability, as well as the log hourly wage.

First, I investigated the descriptive statistics of the key variables:

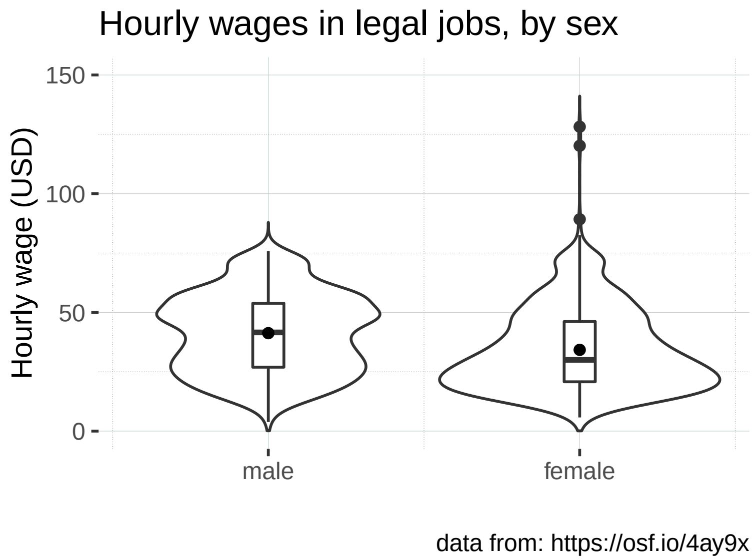
Table 1: Descriptive statstics of hourly wages and age

	Mean	SD	Median	Min	Max	Range	P05	P95	N
wage	37.85	17.39	35.47	3.83	128.20	124.38	14.00	72.12	1333
ln(wage)	3.52	0.51	3.57	1.34	4.85	3.51	2.64	4.28	1333
age	42.00	11.29	41.00	19.00	64.00	45.00	26.00	62.00	1333

data from: https://osf.io/4ay9x

The number of observations is 1333 for all of our key variables. The mean and median of the hourly wage are surprisingly close to each other which usually isn't the case since the wage follows lognormal distribution.

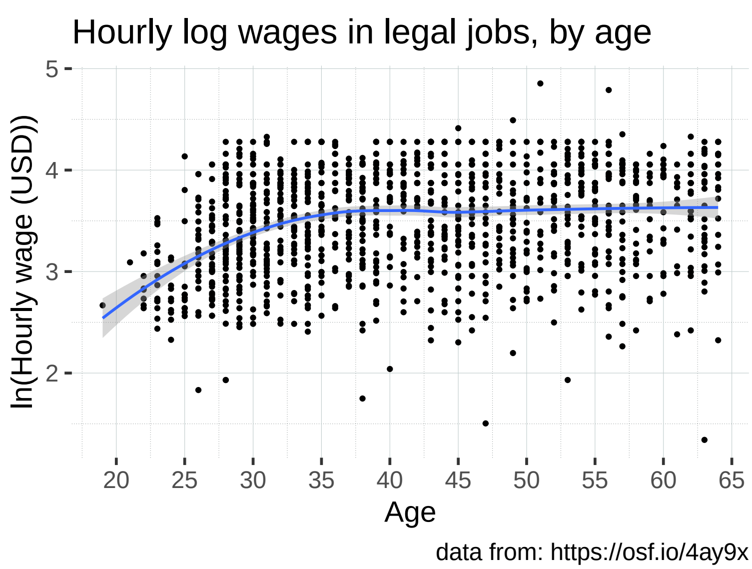
To further investigate the distribution of wages, the next figure shows the violin plots of hourly wages by gender.



The wage distribution of women follows lognormal distribution as expected. Among men, there is a disruption in the distribution, most men earn around 50 USD/hour, while another large group earns around 25 USD/hour. It is also interesting that extreme values can be observed among women only.

This odd looking distribution may raise some concerns about the log transformation of wages, but as we look at the women's wages, we are still better off with log.

As I'm planning to use age in the models, the next Figure displays the relationship between hourly wages and age.



The relationship clearly isn't linear. It flattens around 35 years, so using a spline is recommended in the model. The figure also draws our attention to another interesting feature of the sample: the wages seem to be capped! Most probably because a large share of respondents are employed by the public sector, so their hourly wages are defined by pay scales.

Modelling gender wage gap

For modeling the wage gap, I run 4 log-level regressions: 1) gender 2) gender and occupation 3) gender, occupation and level of education 4) gender, occupation, level of education and age (spline).

The following table shows the Models's results:

Table 2: Modelling wage gap

	Model 1	Model 2	Model 3	Model 4
(Intercept)	3.6210**	3.2645**	3.1969**	2.0931**
	(0.0189)	(0.0328)	(0.0354)	(0.1474)
female	−0.2169**	−0.1056**	−0.0929**	−0.1003**
	(0.0272)	(0.0270)	(0.0269)	(0.0262)
occ		0.4034**	0.1980**	0.1430**
		(0.0312)	(0.0497)	(0.0485)
MA_degree			0.1771**	0.1411*
			(0.0575)	(0.0559)
Prof_degree			0.2950**	0.2710**
			(0.0537)	(0.0522)
PhD_degree			0.2315**	0.1927**
			(0.0599)	(0.0583)
lspline(age, 35)1				0.0349**
				(0.0048)
lspline(age, 35)2				0.0015
				(0.0016)
R2	0.046	0.152	0.171	0.223

* $p < 0.05$, ** $p < 0.01$

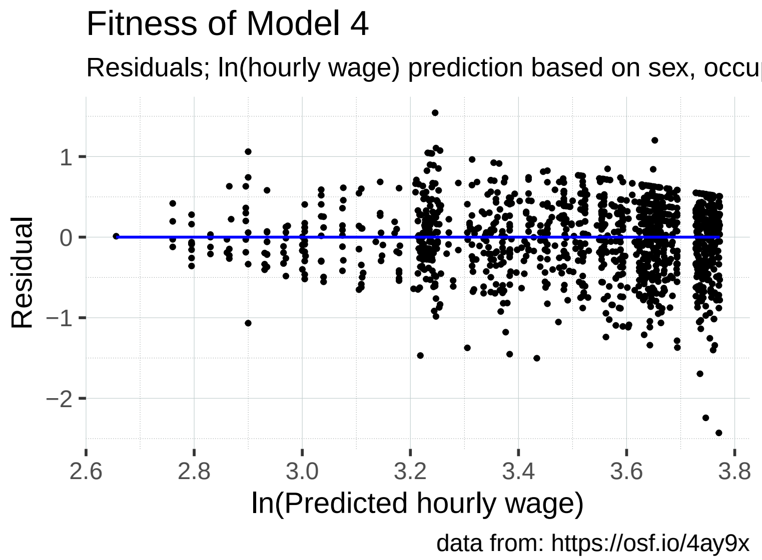
The first model shows the unconditional gender gap. Men earn 37 USD/h on average, while women earn -22% less on average. The Model isn't too strong, gender only explains 4.5675559% of variation in log wages.

Including occupation, the model becomes stronger, explaining already 15.2190808% of variation in log hourly wages. It lowers the effect of gender on hourly wages, woman working in supporting legal occupations earn -11% less on average compared to man in same occupation. Man working as lawyers, judges, magistrates or as other juridical workers earn 40% more on average. It implies that woman are likely concentrated in these supporting jobs.

By including level of education, the model becomes slightly stronger. The gender gap remains relatively the same, the level of education takes over occupation - not that suprising, since given degrees are necessary for certain occupations in the legal field - but occupation remains significant. Compared to Bachelor's degree holder man working in supporting jobs, Master's degree yields 18% higher hourly wage in average, Professional degree holders earn 29% more and PhD holders have 23% higher wage on average.

As we saw on Figure 2, the relationship between age and hourly wage isn't linear, a spline at age of 35 is recommended to be introduced in Model 4. It makes our model stronger, the model explains 22.2877769% of variation in log hourly wages. As expected based on Figure 2, age loses it's importance after 35, the coefficient isn't significant nor at 99% nor at 95%. The other variables' coefficients remain relatively the same - except for Master's Degree that loses it's significance at 99%.

The next Figure shows the fitness of our best Model 4:



Since only the age is quasi-continous, the other variables are binary, the predicted values are more dense around given values. The distribution of residuals suggest that our model is more likely to preduct higher wages in some "extreme" cases - see 2 predicted hourly wages are more than -2 away from the actual value.

Conclusion

In this report, I investigated the gender wage gap in legal occupations. I build 4 Models that included gender, occupation, level of education and age.

As I introduced more variables, the gender gap decreased, but it remained relatively stable and considerable: -22%, -11%,-9% and -10% respectively. All the other variables are positively and significantly associated with hourly wages.