Gender Wage Gap

Introduction

This is a causal analysis of the wage gap between female and male workers. We investigating the wage gap as...

HERE COMES THE MOTIVATION FOR WHY THIS IS A MEANINGFUL PROJECT AND WHAT IS THE MAIN GOAL!

Data

The data are ... (US, which occupation, etc.) Further information is available at the National Bureau of Economic Research (US) - CPS Merged Outgoing Rotation Groups.

Some further information on the data is necessary to interpret the results.

Selecting the sample

We have restricted our attention to age cohorts between 40 and 60 with a graduate degree as the participants in the CPS data are better balanced along these dimensions. Figure 1 shows the age distribution of employees with a graduate degree by gender.

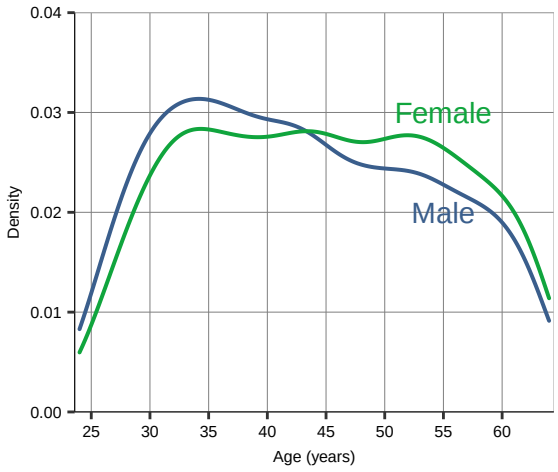

Figure 1: Age distribution of employees with graduate degree by gender

Descriptive statistics

Using this sample our parameter of interest is the expected difference between female and male workers' hourly wages. Table 1 describes the main characteristics of hourly wage and log wages for each gender.

Table 1: Descriptive statistics of hourly wage by gender

gender		Mean	Median	SD	P05	P95	Max	Min
female	wage	33.06	30.00	16.07	12.02	64.10	144.23	0.00
	log-wage	3.37	3.40	0.61	2.49	4.16	4.97	−8.29
male	wage	40.48	40.00	16.86	14.42	72.12	144.23	0.00
	log-wage	3.59	3.69	0.55	2.67	4.28	4.97	−7.20

The number of observations is 18241 for all of our key variables.

DESCRIPTION OF THE SUMMARY STATS: WHAT CAN WE LEARN FROM THEM?

As the focus is the price difference, Figure 2 shows the conditional distributions of the wage and log-wage.

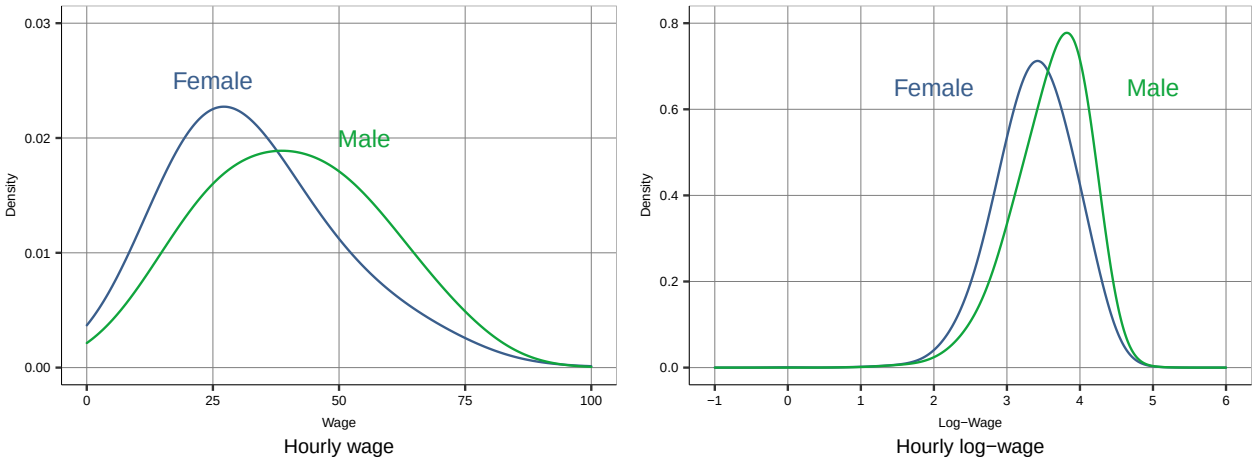

Figure 2: Density plot of hourly wage and log-wage by gender

DESCRIPTION OF THE FIGURE. WHAT DOES IT TELL US?

(May change the order of descriptive stats and graph or if not enough space put one of them into the appendix.)

How will you include this in your model?

Short description of the other variables: 2-10 sentences depending on the number of variables you have. You should reference your decisions on the graphs/analysis which are located in the appendix.

Model

I have run several models, from simply regressing female variables on log wages to more elaborate models, where I have controlled for several socio-economic and other variables.

The preferred model was the largest, where I have conditioned on ...

$$\log \text{ wage} = \alpha - 0.14 \; female + \delta Z$$

where Zs are standing for the controls, which includes controlling for ... . From this model we can infer:

- for the female worker, when controlling for the above-mentioned variables, the wage difference is 0.14.

I have compared multiple models to learn about the stability of the parameters. EXPLAIN BRIEFLY MODELS AND THE RESULTS OF IT

Table 2: Models to uncover gender wage gap

	(1)	(2)	(3)	(4)
Female	-0.224***	-0.212***	-0.153***	-0.142***
	(0.012)	(0.012)	(0.012)	(0.012)
Age and Education	No	Yes	Yes	Yes
Family background	No	No	Yes	Yes
Hours worked	No	No	Yes	Yes
Government or private	No	No	Yes	Yes
Union member	No	No	Yes	Yes
Not born in USA	No	No	Yes	Yes
Age in polynomial	No	No	No	Yes
Hours in polynomial	No	No	No	Yes
R2	0.035	0.043	0.183	0.196
Observations	9,816	9,816	9,816	9,816

Hypothesis testing

All the aforementioned models show that the results are statistically different from zero. To show that, I have run a two-sided hypothesis test:

$$H_0 := \beta_1 = 0$$

$$H_A := \beta_1 \neq 0$$

using or preferred model (4). I have the t-statistic as -11.46 and the p-value as 0, which confirms the conclusion: the probability of making a Type I is almost zero.

Robustness check / 'Heterogeneity analysis'

Task: run a model on the full sample. Report a similar table, with the model (4) as the first column.

Table 3: Models on the full sample

	(4)	(1-full)	(2-full)	(3-full)	(4-full)
Female	-0.142***	-0.195***	-0.172***	-0.115***	-0.110***
	(0.012)	(0.008)	(0.008)	(0.009)	(0.008)
Age and Education	Yes	No	Yes	Yes	Yes
Family background	Yes	No	No	Yes	Yes
Hours worked	Yes	No	No	Yes	Yes
Government or private	Yes	No	No	Yes	Yes
Union member	Yes	No	No	Yes	Yes
Not born in USA	Yes	No	No	Yes	Yes
Age in polynomial	Yes	No	No	No	Yes
Hours in polynomial	Yes	No	No	No	Yes
R2	0.196	0.028	0.055	0.204	0.230
Observations	9,816	18,241	18,241	18,241	18,241

INTERPRET THESE RESULTS IN 2-5 SENTENCES

Conclusion

HERE COMES WHAT WE HAVE LEARNED AND WHAT WOULD STRENGTHEN AND WEAKEN OUR ANALYSIS.

Appendix

Here comes all the results which are referenced and not essential for understanding the MAIN results.

A short explanation of the variables.

Table 4: Descriptive statistics of variables by gender

	female			male		
	Mean	Median	SD	Mean	Median	SD
wage	33.06	30.00	16.07	40.48	40.00	16.86
log-wage	3.37	3.40	0.61	3.59	3.69	0.55
age	49.35	49.00	6.03	49.73	50.00	5.95
afroamerican	0.10	0.00	0.30	0.06	0.00	0.24
hispanic	0.05	0.00	0.22	0.05	0.00	0.22
asian	0.07	0.00	0.26	0.11	0.00	0.32
other ethnicity	0.02	0.00	0.14	0.01	0.00	0.12
non US born	0.13	0.00	0.34	0.20	0.00	0.40
Professor	0.10	0.00	0.29	0.14	0.00	0.34
PhD degree	0.11	0.00	0.31	0.17	0.00	0.37
married	0.68	1.00	0.46	0.81	1.00	0.39
divorced	0.17	0.00	0.38	0.10	0.00	0.30
widowed	0.02	0.00	0.14	0.00	0.00	0.07
1 child	0.01	0.00	0.09	0.02	0.00	0.14
2 child	0.02	0.00	0.14	0.02	0.00	0.14
3 child	0.15	0.00	0.36	0.15	0.00	0.36
4 or more children	0.25	0.00	0.43	0.32	0.00	0.47
usual work hours	41.52	40.00	8.59	45.36	40.00	9.28
federal gov.	0.05	0.00	0.23	0.07	0.00	0.26
state gov.	0.15	0.00	0.36	0.10	0.00	0.30
local gov.	0.24	0.00	0.43	0.11	0.00	0.31
private sector	0.16	0.00	0.37	0.12	0.00	0.33
union	0.28	0.00	0.45	0.14	0.00	0.35

Table 5: Unit shares in each state and industry by gender

		female		male	
		N	Percent	N	Percent
state	AK	57	0.58	56	0.57
	AL	54	0.55	49	0.50
	AR	35	0.36	24	0.24
	AZ	60	0.61	43	0.44
	CA	396	4.03	458	4.67
	CO	124	1.26	135	1.38
	CT	157	1.60	141	1.44
	DC	171	1.74	168	1.71
	DE	79	0.80	62	0.63
	FL	164	1.67	175	1.78
	GA	117	1.19	110	1.12
	HI	76	0.77	51	0.52
	IA	70	0.71	61	0.62
	ID	50	0.51	44	0.45
	IL	161	1.64	178	1.81
	IN	66	0.67	53	0.54
	KS	73	0.74	79	0.80
	KY	61	0.62	44	0.45
	LA	54	0.55	39	0.40
	MA	119	1.21	138	1.41
	MD	185	1.88	168	1.71
	ME	84	0.86	67	0.68
	MI	84	0.86	118	1.20
	MN	128	1.30	96	0.98
	MO	71	0.72	70	0.71
	MS	54	0.55	31	0.32
	MT	48	0.49	41	0.42
	NC	87	0.89	78	0.79
	ND	49	0.50	43	0.44
	NE	71	0.72	66	0.67
	NH	115	1.17	100	1.02
	NJ	119	1.21	153	1.56
	NM	57	0.58	40	0.41
	NV	51	0.52	49	0.50
	NY	281	2.86	212	2.16
	OH	113	1.15	109	1.11
	OK	42	0.43	45	0.46
	OR	62	0.63	74	0.75
	PA	143	1.46	134	1.37
	RI	101	1.03	87	0.89
	SC	62	0.63	45	0.46
	SD	56	0.57	43	0.44
	TN	45	0.46	41	0.42
	TX	215	2.19	210	2.14
	UT	37	0.38	63	0.64
	VA	123	1.25	158	1.61
	VT	113	1.15	79	0.80
	WA	68	0.69	99	1.01
	WI	95	0.97	85	0.87
	WV	67	0.68	43	0.44
	WY	48	0.49	43	0.44
industry code	0	2951	30.06	2367	24.11
	1	1744	17.77	1845	18.80
	2	323	3.29	586	5.97

Discuss the age effect. Short description of the model

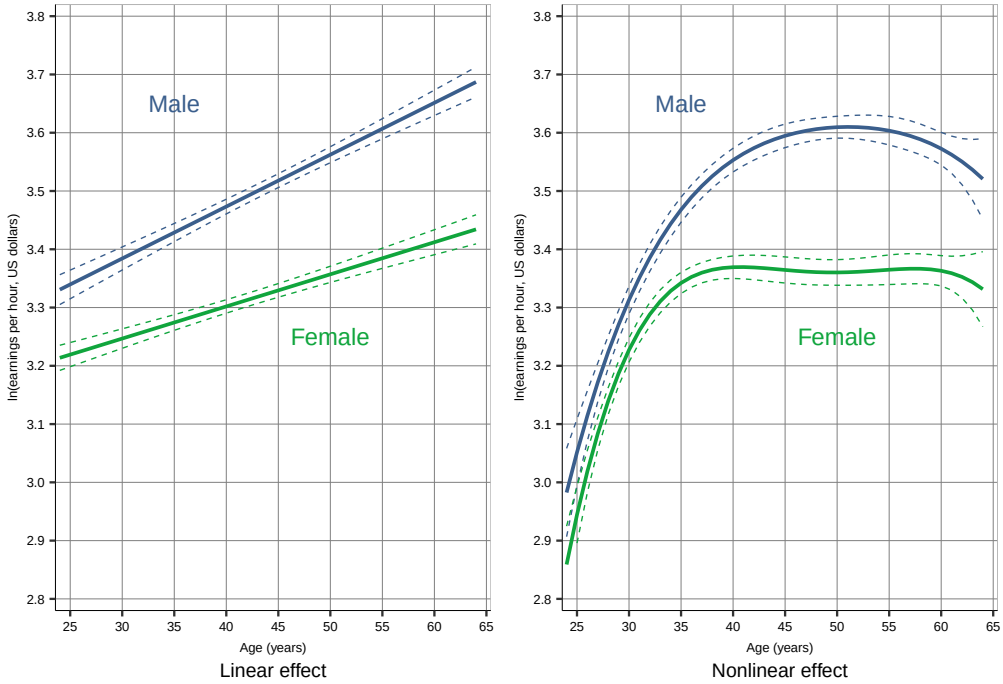

Figure 3: Earning differences by gender as function of age