Mannat Sandhu

Gender Wage Gap Analysis

Problem Statement

Objective: To study differences in wages and hours worked by the gender of respondents in Org data and how these differences in employment outcomes change along the lifecycle of work, from college age to retirement.

Specifically, we estimate the following using Generalized Linear Model:

$$log(rw) = as.factor(age) * female + educ + wbho + u$$

where:
rw : real wage
educ : education factor variable
wbho : factor variable describing race of the respondent
female : takes on value 1 when respondent is a woman

Part (a)

We estimate the model described above for 1993 and 2013 (separately), restricting the models to only those respondents who are in the labor force.

I am using poisson regression because this will constrain predicted wages to be positive.

Here are the education coefficients for the year 1993 -

as.factor(age)90	[illegible]	[illegible]	[illegible]	< 2e-16	
female	-0.063323	0.026109	-2.425	0.015296	*
educHS	0.220502	0.003500	62.999	< 2e-16	***
educSome college	0.374241	0.003577	104.622	< 2e-16	***
educCollege	0.661744	0.003661	180.749	< 2e-16	***
educAdvanced	0.832202	0.003886	214.155	< 2e-16	***
wbhoBlack	-0.117953	0.002965	-39.784	< 2e-16	***
wbhoHispanic	-0.083121	0.003588	-23.168	< 2e-16	***
wbhoOther	-0.040707	0.003854	-10.561	< 2e-16	***

Since the regression we're running is for log(rw), we can interpret the coefficients as follows -

1. Being educated till HS increased the real wage by 22% w.r.t educ less than high school.
2. For further, education in some college, the real wage increases by (0.374 - 0.220)*100% = 15.4%.

3. And so on...

Moreover, the education coefficients are all statistically significant at the 95% level.

Now, running the same regression for 2013 -

as.factor(age)85	2.681538	0.031400	85.398	< 2e-16	***
female	-0.137686	0.037051	-3.716	0.000202	***
educHS	0.268778	0.004476	60.051	< 2e-16	***
educSome college	0.415207	0.004479	92.697	< 2e-16	***
educCollege	0.825885	0.004466	184.915	< 2e-16	***
educAdvanced	1.031921	0.004556	226.512	< 2e-16	***
wbhoBlack	-0.132618	0.002827	-46.913	< 2e-16	***
wbhoHispanic	-0.084240	0.002680	-31.429	< 2e-16	***
wbhoOther	-0.002117	0.002814	-0.753	0.451726	

Once again, all the coefficients of education are statistically significant at the 95% level.

1. Being educated till HS increased the real wage by 26.9% w.r.t educ less than high school.
2. For further, education in some college, the real wage increases by (0.415 - 0.269)*100% = 14.6%.
3. And so on …

If we compare the education coefficients of 1993 to the education coefficients of 2013, one clear distinction is that the coefficients of 2013 for all educ variables are higher in value than in 1993.

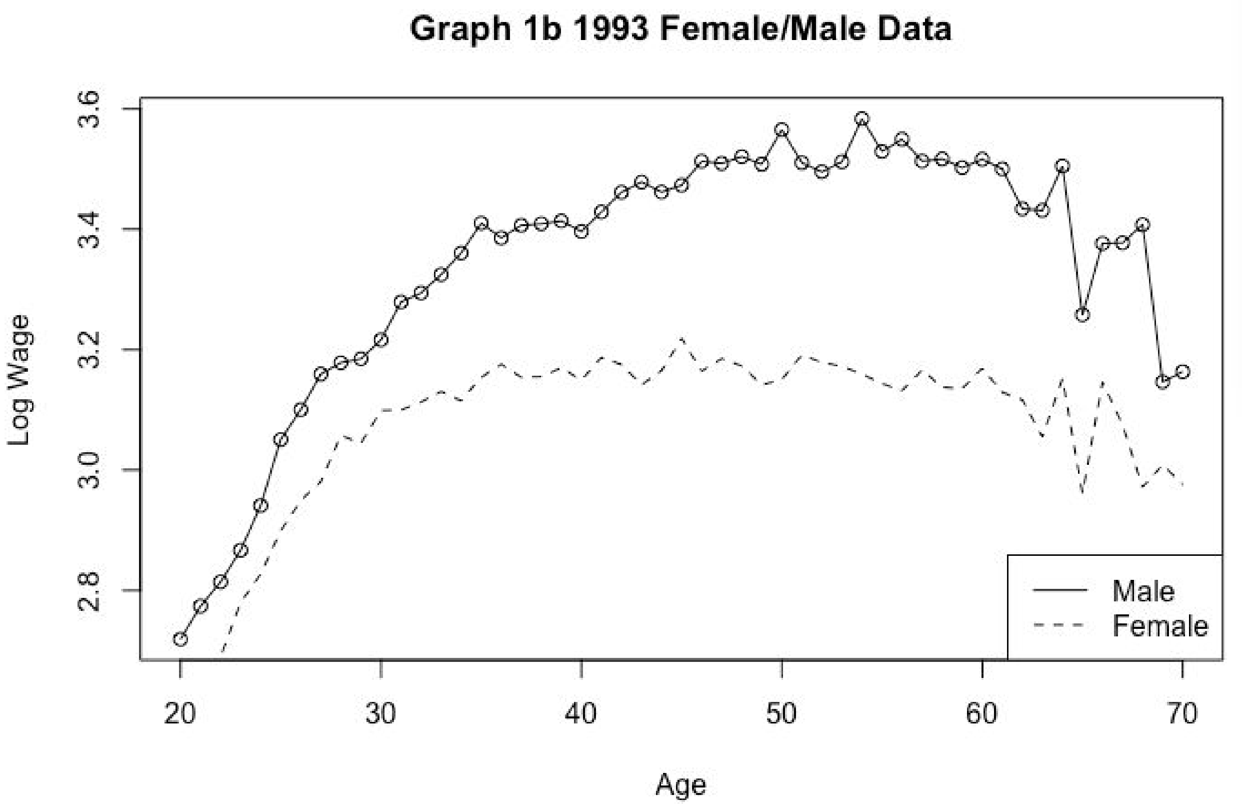
This means that education has more of an impact in the total value of the real wage in 2013 than in 1993.

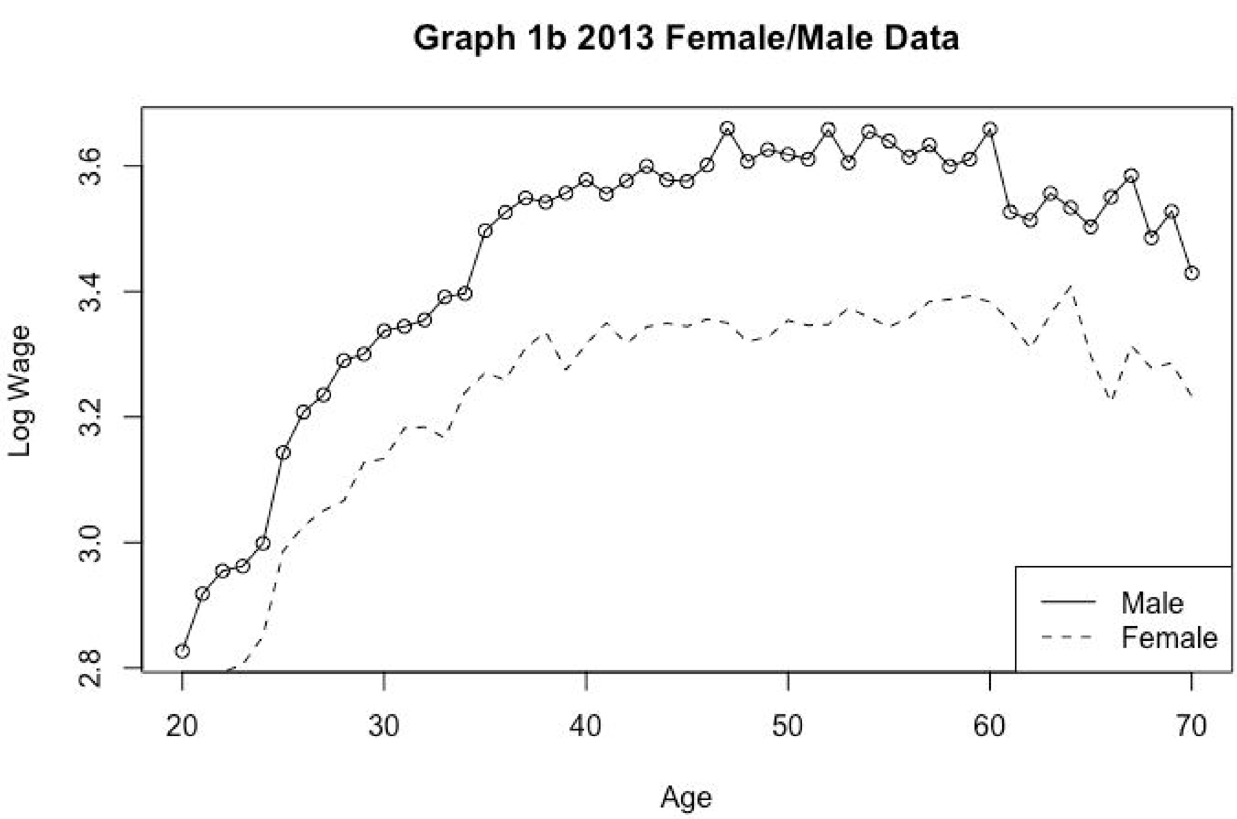
Part (b)

For each sample (1993 and 2013), we generate a predicted log wage for a white female with a college education for all ages 20-70, and then for a white male with a college education for the same ages. Then, we will illustrate the log differences in male-female wages on a plot, how the wage gap changes over the age of the respondent, and how the life cycle of the wage gap changes between 1993 and 2013.

The following graphs are for college educated white male/females.

Mannat Sandhu





Mannat Sandhu

The wage of males and females increases from age 20 to roughly till 58 years of age and then starts reducing. In both 1993 and 2013, the absolute value of the wage gap between males and females increase as the wage increases (~58 years), and then reduces.

(This is, in part, a result of the regression that we have chosen. Since we have age*female as the only term capturing the detail female, the wage gap is a scaled version of the age effect on wage.)

The absolute value of the wage gap is less in 2013 as in 1993. To give a concrete example, the wage gap between males and females in 1993 for 58 years of age was (3.516807 - 3.137453)*100% = 37.9% and in 2013 was (3.599531 - 3.387576)*100% = 21%.

Part (c)

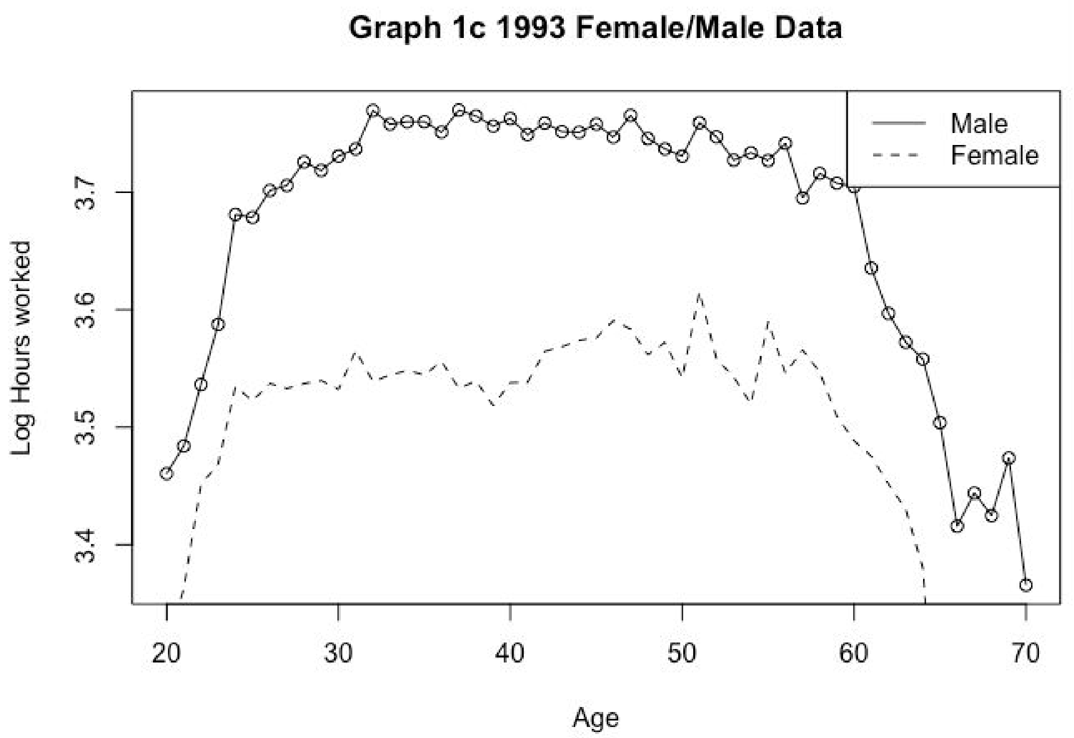
We use the same method as in (b), but propose and use a model evaluating hours worked.

Since we want to study the wage differences by gender, and how these differences change with age, I am considering the following model, without taking *rw* as an independent variable -

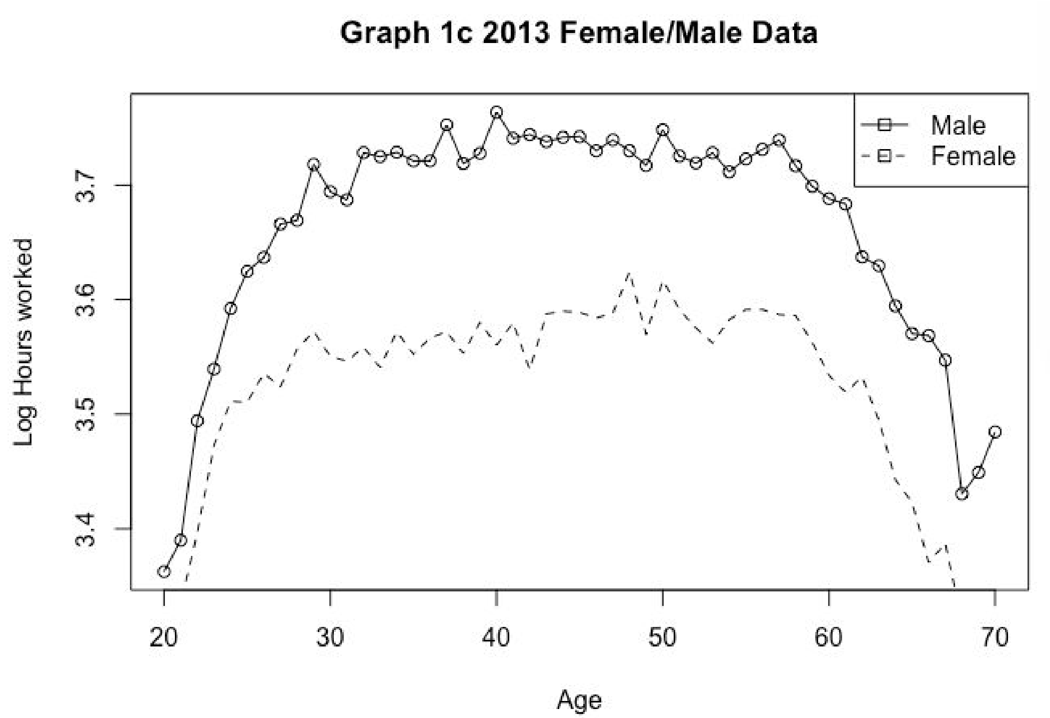
log(hours_worked) = as.factor(age)*female + educ + wbho + u

Once again, I am using poisson regression, which will constrain the number of hours worked to positive.

Mannat Sandhu



Graph 1c



The hours worked for males and females increases from age 20 to 32, stays same till the age of 55, and then starts reducing. This can probably be explained by the fact that we have taken the average number of hours worked for the prediction, and converted the N/A values to 0. As a result, not all of the people in the age group 20-32 are working (e.g., they may be in school).

As before the absolute value of the gap between the hours worked for males and females increase as the number of hours worked increases, and reduces as the hours reduce. During the period from 32 to 55 however, the gap is highest at the age 40, and then reduces a bit.

(This is, once again, in part, a result of the regression that we have chosen. Since we have age*female as the only term capturing the detail female, the hours worked gap is a scaled version of the age effect on hours worked.)

The absolute value of the hours worked gap is less in 2013 as in 1993. To give a concrete example, the wage gap between males and females in 1993 for 40 years of age was (3.762635 - 3.537862)*100% = 22.4% and in 2013 was (3.763563 - 3.560295)*100% = 20.3%.