PS 6800: Problem Set 9

sub Professor: Heather Huntington Teaching Assistant: Donald Moratz

Problem 1

On average, women earn less than men in almost every country – this is referred to as the 'gender gap.' We know less about the mechanism behind this gap. Employers may discriminate against women by not hiring them for high-paying jobs or by paying them less than men for comparable jobs. Women may have different levels of education. Women may also choose occupations that generally have lower levels of monetary compensation – for example, due to social expectations of performing "feminine" jobs, or because these careers offer greater work-life balance.

This exercise employs a sample from a 1994 survey of South African workers. The data are contained in a file called safrica.dta. Variable definitions are as follows:

- wage: average hourly earnings (1994 RZA; 1 US dollar = 3.5 RZA in 1994).
- age: age in years.
- educ: number of years of education (note: 10=completed secondary school, 13=completed university degree)
- exper: years of work experience (age-education-7)
- union: = 1 if working a union job, = 0 otherwise.
- female: = 1 if female, = 0 otherwise.
- married: = 1 if married with spouse present, = 0 otherwise.
- urban: = 1 if lives in an urban area, = 0 otherwise.

a) Run a regression of wage on female. Interpret the coefficient. Do you believe this regression meets the assumptions to causally interpret the model? What kind of inferences would you be able to make given your assessment of the validity of the assumptions?

b) More education is associated with increased earnings. Run a regression of wage on female and years of education. Interpret the coefficients, explaining how they have changed from a) and why. Does controlling for education explain the gender gap? Why or why not? Does including education change whether the model meets the OLS assumptions? Why or why not?

c) Run a regression of wage on female, educ, age, union, marital status, and urban variables in the regression. Does the gender gap persist? Based on this, can you conclude that the mechanism behind the gender gap is discrimination?[1]

d) Run a regression of wage on age. Interpret the coefficient. Interpret the statistical and practical significance, too.

e) Run a regression of wage on age and exper. Interpret the coefficient estimate on age, explaining how it has changed from d) and why. Does this change strike you as a problem? If so, do you have any suggestion to cure the problem?

[1] Note that there are other research designs - like audit experiments using fake resumes - that have compellingly shown discrimination! This question is about what we can get from observational data.

Problem 2

An article by Ebonya Washington in the *American Economic Review* argues that having a daughter (as opposed to a son) might affect how politicians vote on women's issues. In particular, she argues that having a daughter causes Congressional representatives to vote more liberally on women's issues. We will use Washington's data (available on the website as `washington.csv`). The key variables are:

- ngirls: Number of daughters
- totchi: Total number of children
- party: Indicator for democrats (1), republicans (2), or other (3)
- aauw: Legislator's voting score as assigned by the American Association of University Women - higher scores indicate being more liberal on women's issues.

a) Comparing only Republicans and Democrats (dropping the 'Other' category), examine whether legislators of different parties have the same number of children. Do the results surprise you? Why or why not?

b) Regress a representative's aauw score on the number of female children and report your results. What is the relationship between the number of female children and the AAUW score? Do you think this relationship estimates the causal effect of having a female child?

c) Now regress a representative's aauw score on the number of female children and total number of children and report your results. What is the relationship between the number of female children and the AAUW score after controlling for the total number of children? Do you think this relationship accurately estimates the causal effect of having a female child? If you believe a causal claim is possible, discuss what assumptions you have to make. If you believe it is not possible, explain why.

d) Now add party fixed effects to your regression in part c). How does this change the results? Why? Is it a good idea to add this control?

e) Regress the aauw score on the number of female children, controlling for the total number of children, for Republicans, Democrats, and the full sample. Plot the coefficients and 95% confidence intervals. Interpret your results.