MOEC.0021

Empirical Methods

University of Zurich HS 2018

PROF: Greg Crawford / TA: E. Dicarlo, M.R. Greco, S. Bagagli, A. Jenni Handout 2

Exercise 2

---

1 Theory

1. **The gender wage gap.**

   Suppose you want to test whether in your country women are discriminated against relative to men in terms of wages. You decide that you want to test whether men and women have different salaries. Suppose you are able to gather data on the whole working population in your country. For each individual you have the following information.

   - *monthly wage*
   - *gender*
   - *years of education*

   (a) Suppose *years of education* have the same effect on wages for both men and women. Propose a simple regression model to test your hypothesis.

   (b) Provide a graphical representation of the conditional expectation function (i.e. the part of wages that we can explain with our covariates) and show if and how it differs for men and women.

(c) In retrospect, you decide that *years of education* might have a different marginal effect on men compared to women. How would you modify your regression model to account for this differential effect?

(d) Provide a graphical representation of the conditional expectation function (i.e. the part of wages that we can explain with our covariates) and show if and how it differs for men and women.

2 Empirical Application

1. **The Gender Wage Gap.** In this exercise we will try to explore some discrimination theories analyzing a subsample from the US CPS2015. Many politicians, institutional observers, and researchers still claim today the existence of discrimination against female workers in the labor market. They base their claims looking at the *gender wage gap*, i.e. the difference between men's and women's wages. As many other things in economics, this wage gap can be generated both from the demand side (employers who discriminate against women) and from the supply side (women having different skills or preferences for specific jobs or for entering the labor market at all). In this exercise we will try to learn more about the gender wage gap, while testing you on your econometric toolkit. For this question, assume that Assumption 2 (Mean-zero Error) holds so that you can make causal statements in your answers.[1]

   Download the dataset *sampleUScens2015.csv* from OLAT and import it into Stata or R. The dataset includes prime age individuals (i.e. $age \in [25, 54]$) active in the labor market (i.e. either employed or looking for job), and working in the private sector. There are seven relevant variables:

   - age, the age of the individual in 2015
   - $education$, years of completed education
   - $incwage$, income from wages in 2015 in USD
   - $female$, dummy for female
   - $childrenly$, dummy if had a children in the last year
   - $degfield$, field of degree
   - $occupation$, sector of occupation

[1] Note that this is a very strong assumption that is unlikely to hold, but we want to focus on other aspects of econometrics for the moment.

(a) Generate a new variable called $wage = incwage/1000$. Also, generate lw taking the log of *wage*. Generate a dummy named *university* which is equal to 1 if $education \geq 16$. First regress *wage* on education, then regress *wage* on education and the university dummy. How does the coefficient on education change? How do you interpret it in both specifications?

(b) Drop the university dummy. Now regress *wage* on education and age. Also, regress log wages (lw) on education and age. What are their coefficients? How do you interpret them? How do they compare? [Note: be sure you compare approximately equivalent objects from each specification.]

(c) Now regress log wages on education, age, and the female dummy. You get the following model:

$$lw_i = \beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 female_i + \epsilon_i$$

What is the coefficient on $female$? How do you interpret it? Is it economically significant in your opinion? Test both in R/Stata and "by hand" the hypothesis that $\beta_4 = 0$. Should you use a one-sided or two-sided test? Do the one you think most appropriate.

(d) Use R/Stata to get β_4 (the coefficient on $female$) using partitioned regression as we did in lecture.

(e) Use R/Stata to show that $\hat{\beta}_1 = \bar{y} - \bar{X}'_{-1}\hat{\beta}_{-1}$

(f) Include in the model in (1c) the interaction between $female$ and *education*, together with the interaction between $female$ and *age*. So your model is now:

$$lw_i = \beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 female_i +$$

$$\beta_5 female_i * educ_i + \beta_6 female_i * age_i + \epsilon_i$$

Test in R/Stata the individual hypotheses that $\beta_4 = 0$, $\beta_5 = 0$, and $\beta_6 = 0$. Test "by hand" and in R/Stata the joint hypothesis that they are all zero.

(g) Run again the model in (1c) **separately** for males and females. How do the coefficients for *educ* and *age* in the males regression compare to the coefficient estimates in part (1f)? How do the coefficients for *educ* and *age* in the females regression compare to the coefficient estimates in part (1f)? What does this tell you about the impact of interacting a dummy variable with *all* the other variables (including the constant) in a regression?

(h) Generate a dummy for each occupation category. Can you include all of them in your model? Why or why not?

(i) Now test the model in part (1c) for each occupational subsample (i.e. perform the regression in part (1c) each occupation at a time). Comment on the pattern of your wage gap estimates across occupations. Is the gender wage gap statistically different across occupations? Provide support for your conclusions.

(j) Drop all the males from your dataset.

i. Regress log wages on *educ*, *age*, and *childrenly*. Test in R/Stata $H1$: $childrenly < 0$ for workers in technology. Is the effect negative in every occupation? Provide support for your conclusion.

ii. Regress log wages on *educ*, *age*, *childrenly*, and occupation dummies (exclude the dummy for "other"). Following the "p-value" path, test whether the gender wage gap is the same in business and science (i.e. test $\beta_{business} = \beta_{science}$). Do the test both in R/Stata and "by hand". How does your answer compare to Stata's/R's?

iii. How do occupations' dummies compare to point (1i)?

(k) Throughout this question we have assumed that Assumption 2 holds. What do you think about this assumption? Can you think about other factors we did not

take into consideration in our model that could bias the conclusion that we are measuring the true gender wage gap?