Econometrics 1 - Problem Set 2

LMEC, Fall 2021

Prof. Matteo Barigozzi

TA: Nektaria Glynia

October 20, 2021

In groups (as they have been decided), use the software Stata (and your brain), to answer the questions below. Please return one zip-file (one for each group) to nektaria.glynia2@unibo.it by **October 27, 2021 (11:59pm, 23:59)**; write as object of the email PS2-Solutions'. The zip-file must contain (i) a document with answers to each specific question (pdf format); (ii) the Stata log-file; (iii) the Stata do-file. Name all the files as surname1_surname2_surname3, and in any case remember to write name, surname and id number (matricola) of each student in the document.

Concise answers to all questions must be included in the pdf (either theoretical answers, or Stata output) but Stata commands and code can be contained in your do and log files. Good luck!

Questions

Acemoglu and Restrepo (Journal of Political Economy, 2020) study the effects of industrial robots on US labor markets. They show that robots may reduce employment and wages and that their local impacts can be estimated using variation in exposure to robots—defined from industry-level advances in robotics and local industry employment. The empirical approach followed in the paper is to estimate the changes in employment and wages on a variable measuring *exposure to robots*, by relying on the following model:

$$d\_labor\_mkt_{ci} = \beta_0 + \beta_1 exposure\_to\_robots_i + \boldsymbol{X}_{ci}\boldsymbol{\gamma} + u_{ci} \tag{1}$$

where $d\_labor\_mkt_{ci}$ is a measure of changes in labor market outcomes over a period of time (changing according to your group number) in community zone c and industry i, and $exposure\_to\_robots_i$ is a measure combining industry-level variation in the usage of robots and employment shares, and can be interpreted as the number of robots per 1,000 workers in each industry. $\boldsymbol{X}_{ci}$ contains a number of demographic characteristics measured in 1990 and including log population, the share of female in the population, the share of population

over 65, the share of population with different level of education, and the share of Whites, Black, Hispanics and Asians in the population. Community zones in the USA are a good proxy for local labor markets and for this reason it is used as the unit of observation in the first part of the exercise.

A. Preliminary Tasks

For this part, use the dataset_1 provided to your group. Use as outcome variable *d_empl_***years**, expressing the change in employment occurred at community zone level over the period indicated in the variable name (please check carefully the label attached to the variable in order to fully understand its meaning. Keep in mind that 'empman' means employment in manufacturing sector; 'empnpub' employment including self employment and public employment).

1. Compute the average value of *exposure_to_robots* by state and indicate the states that registered the highest and lowest variation in the variable (it is enough to look at the average values).

2. Did these states also registered the highest and lowest variation in the outcome variable?

3. Discuss the average value and the minimum and maximum values of the outcome variable computed at state level.

4. Make a scatter with a fitted line of change in employment over exposure to robots and discuss what it suggests.

B. Regression Analysis - Employment

1. List the minimal set of assumptions required to make large sample inference by OLS.

2. Estimate a simple univariate regression model on the exposure to robots through OLS using heteroskedasticity robust standard errors and comment your results.

3. Add the list of demographic controls as indicated above. Which problem do you encounter by adding *all* controls on race and education level? Adjust the list of control and discuss their coefficients.

4. Now add a polynomial of order two of exposure to robots to the previous model and comment the results.

5. Decide trough a suitable test if the marginal effect of exposure to robots in the model estimated in B.4 is equal to zero.

6. Present brief evidence for/against the presence of heteroskedastic errors in you regression at point B.4.

7. "Huber-White robust standard errorr should always be used to preserve the validity of your inference in case there is heteroskedasticity". Defend this position against using classical OLS standard errors, with reference to consistency of the estimator and validity of the t-test.

C. Regression Analysis - Wage

Now consider the following model:

$$d\_wage_{cg} = \beta_0 + \beta_1 exposure\_to\_robots_c + \boldsymbol{X}_{cg}\boldsymbol{\gamma} + u_{cg} \tag{2}$$

where d_wage_{ci} is a measure of changes in wage over a period of time (changing according to your group number) in community zone c and group g, where g identifies groups of individuals according to gender, education level and race. $exposure\_to\_robots_c$ is a measure combining community zone-level variation in the usage of robots and employment shares, and can be interpreted as the number of robots per 1,000 workers in each community zone. $\boldsymbol{X}_{cg}$ contains information on a dummy variable, called *female*, equal to 1 for the group of women, and 0 otherwise; the variable *education* which defines the group's education level and the variable *race*, which defines the race of the group.

Now, use the dataset_2 provided to your group. Use as outcome variable $d\_wage\_$**years**, expressing change in wage occurred over the period indicated in the variable name (please check carefully the label attached to the variable in order to fully understand its meaning. Keep in mind that 'wk' means weekly; 'yr' means yearly).

1. Define a dummy variable equal to 1 for the group of whites and 0 otherwise.

2. Run a regression on change in wage over exposure to robots and female, white and education. Comment the results.

3. Perform a Chow test in order to test if there is *any* difference between whites and non-whites.

4. Decide trough a suitable test if marginal effect of exposure to robots differs between men and women.

5. Run a new regression adding dummy variables for each level of education.What is the the wage premium for the group of individuals with college degree with respect to the group of individuals with less than high school?

6. Run a regression with an interaction term between exposure to robots and the set of dummies for education. What is the marginal effect of exposure to robots for the group of individuals with master degree? Is it statistically different with respect to the marginal effect for the group of individuals with less than high school?