Empirical Project

Machine Learning in Econometrics

due on July 31*

1 Goal

Germany's government is interested in reducing gender inequality, especially gender wage gaps and gender gaps in employment. They are considering the introduction of a set of policies that incentivize firms to shrink gender inequality in working conditions. First, the policy forces firms to internally publish salaries of all workers, so discrepancies in salaries can be detected by workers themselves. Second, firms are incentivized to encourage salary negotiations. Third, firms are incentivized to offer childcare where needed for their employees to fulfill their duties.

The government has been made aware that, in parts of the United States, exactly these policies have been introduced and now asks you to evaluate the effectiveness of these policies in reducing gender inequality in wages and employment. The dataset `genderinequality` (provided in `RData` and `csv` formats) contains data on individuals in the U.S., some working in firms to which the new policies apply (these are the treated workers) and some working at firms to which the new policies do not apply (these are the untreated workers). The policies have been introduced in 2007 and we have panel data on workers for the years 2005 and 2010, i.e. before and after the introduction of the new policies. The table below describes the variables in the dataset.

*to be submitted on moodle by July 31, 23:59

variable name	description
year	year: either 2005 or 2010
id	identifier for a worker, every worker occurs twice in this dataset: once in each of the two years
wage	weekly wage
hours	hours worked per week
emp	employment (1=employed and 0=unemployed)
treat	treatment indicator (1=treated worker and 0=untreated worker)
female	gender (1=female and 0=male)
IQ	result on an IQ test
KWW	score on a test of occupational knowledge
educ	years of education
exper	years of work experience
tenure	years with current employer
age	age in years
married	1=married and 0=not married
black	1=black and 0=non-black
south	location of home (1=in the South and 0=not in the South)
urban	location of home (1=in an urban area and 0=not in an urban area)
sibs	number of siblings
brthord	birth order
meduc	mother's years of education
feduc	father's years of education

Your task is to (1) perform a statistical analysis evaluating the effectiveness of the new policies in reducing gender inequality in wages and employment, (2) raise potential difficulties in transferring the lessons learned for the U.S. to the German context, (3) and explain which parts of the populations are expected to gain the most/least from the new policies. Your report must be convincing, precise and concise so that important decisions can be based on it. In particular, your statistical analysis must be (a) based on an approach that works well under assumptions that can be justified, (b) transparent in the sense that your tuning and modelling choices are clearly spelled out, (c) and robust in the sense that results don't dramatically change when some of these choices are changed.

2 Instructions

2.1 General Instructions

It is very important that your report is convincing, precise and concise. So, at every step of your analysis think about how you can justify your choices and assumptions. Clearly define objects of interest, why you estimate them, under which conditions they are identified, how you estimate them, why the chosen estimator is a "good one". Provide sensitivity analyses.

2.2 Permissible Ressources

You must work on this project on your own, no group work is allowed. You must produce the report as well as code on your own. You may use any "non-human" resources available to you, i.e. books, information on the internet, the lecture materials. However, you are not allowed to use interactions with other humans, i.e. no discussions with your classmates about the project, no discussions with others through the internet, no soliciting of help in online forums etc.

2.3 Format of the Report

- Length: max. 10 pages including graphs, tables, references
- File has to be a pdf document
- Filename: `[lastname]_[firstname]_[studentnumber].pdf`
- Take time to polish the text, so it is easy to read, precise and concise.
- The report must include the following sentence at the end:

  > "I confirm that this report is based on my own work. In preparing this report, I have not received any help from another human nor have I discussed any aspects of the empirical project with others."

  Then sign the report just below this statement.

2.4 Code

You must submit all code used in your statistical analysis in a separate file or multiple files. Make sure your code is readable and well-documented.

2.5 Expected Workload

Expected amount of time spent on the entire project: about 15 hours of empirical analysis + 2-3 hours of carefully writing up the report.