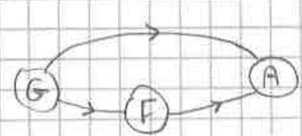
# ECONOMETRICS II - PS6

EXERCISE 1

a) Yes, it can be considered a binary treatment: a dummy = 1 if male, = 0 if female.
Being male is a characteristic assigned only to a part of the population, hence we have a control group, i.e. being female (or viceversa, switching =1, =0)

b) ATE with full independence = 0.141645 (on the excel)

c) If we are interested in the overall effect of being male or female for admission, you don't care whether being female or male causes selection in a particular field, because you are anyway interested in both channels without distinguish them. If gender is randomly assigned (plausible assumption), then you are estimating the effect of being born male on admission. Independence holds.



If we are interested to test whether given any observable characteristics a female is more or less likely to get admitted (e.g. 2 candidates with the same CV one male, one female, what's the difference in admission probability between the two?) then it's not enough to match based on field of study. And independence would not hold.

If we are interested in the effect of being born male on the admission in university controlling for the fact that people may self select into fields where they have higher chance to get in, then full independence does not hold.

Gender is random, but the choice of field given gender is not random.

d) Assume that treatment and potential outcome are independent conditional on the field of study (CIA)

(on the excel)

$ATE_{CIA} = -0.042637$

$ATT_{CIA} = -0.070969$

e) the weights given to each "cell" $X=x$ change the estimated effect:

- unconditional ATE gives the same weight to all the cells
- conditional ATE gives ≠ weights to each "cell" (set of characteristics, in our case the department) depending on how many observations are there in that particular cell $X=x$ given the sample population. This gives higher weight to department A, which has a lot of applicants (especially males) and a big negative difference in outcome between male and females $Y_1 - Y_0 = -0.2$, hence the conditional ATE is smaller than the unconditional ATE

f) on dofile "exercise_1f.do" $\hat{\beta}_{OLS} = -0.0184$

g) regression and matching estimates are different because they use ≠ weighting schemes

- The matching combines estimated effects in each cell $X=x$ (in our case ≠ departments) weighted by the probability of the observation to fall in cell $X=x$
- Regression produces a variance weighted average of these effects to minimize sum of squared residuals: higher variance observations have more weight.

Since the treatment effect varies for each $X=x$ the weighting scheme used changes the estimation of the average effect

## EXERCISE 2

$$Health_i = \beta_1 + \beta_2 Educ_i + \beta_3 X_i + \mu_i$$

a) this formulation assumes linearity

↳ for consistency we need $\mu \perp$ Education or $Cov(\mu, Educ) = 0$ → this does not hold if there is omitted variable (in error term) correlated with education and not included as regressor in $X_i$ ↓ e.g. unobservables cannot, for sure, be included in X

$$* \begin{cases} \mathcal{H} = \beta_1 + \beta_2 Educ_i + \beta_3 X + \mu_i \\ Educ = \gamma_1 + \gamma_2 Z + \gamma_3 X + \nu_i \end{cases}$$

b) if instruments for Educ are $\perp$ to $\mu$ (exclusion restriction)
instruments for Educ are $\perp$ to $\nu$ (exogeneity) } ⇒ problem in a) is solved

c) STRUCTURAL FRAMEWORK

- EXOGENEITY : covariance of instrument for education and the error term in 1st stage regression is equal to 0
- EXCLUSION RESTRICTION : covariance of instrument for education and the error term is 2nd stage regression is equal to 0
  ↳ The instrumental variable has an effect on Health only through education.
- RELEVANCE : Covariance of instrument for education and education is different from 0

POTENTIAL OUTCOME FRAMEWORK

* under homogeneity of treatment

- SUTVA : Non-interference ⇒ the treatment status of a unit does not affect the potential outcomes of the other units
  No variation in treatment ⇒ the treatment for all units are comparable

(I used my notation *) → Education $(Z)$ = Education$_i(Z_i)$ and $\mathcal{H}$(Education, $Z$) = $\mathcal{H}$(Education$_i$, $Z_i$)

- EXOGENEITY : $Cov(Z, \nu) = 0$
- EXCLUSION RESTRICTION : $Cov(\mathcal{H}, Z | Educ) = Cov(Z, \mu) = 0$
- RELEVANCE : $Cov(Z, Educ) \neq 0$
- HOMOGENEITY : $\beta_{2,i} = \beta_2$

* heterogeneous treatment effect
  we can use LATE THEOREM

- SUTVA : same as before
- EXOGENEITY $(Educ | Z) \perp Z$ $\quad \mathcal{H}(Educ, Z) \perp Z$
- EXCLUSION RESTRICTION $\mathcal{H}(Educ, Z) = \mathcal{H}(Educ)$
- RELEVANCE Education and the instrument for education are correlated
- MONOTONICITY : defiers are not present

d) <u>compliers</u> are the ones that being born early in the year drop school with lower educational attainment
and being born late in the year drop school with higher educational attainment

<u>defiers</u> are the ones that being born early in the year they stop studying with higher educational attainment
and being born late in the year stop studying with lower educational attainment

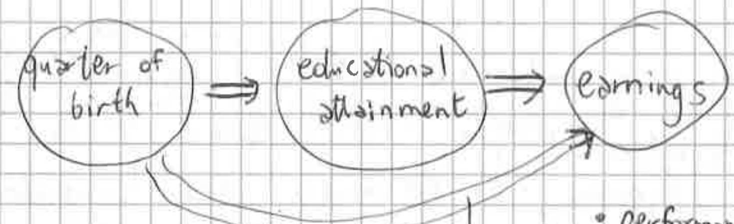
<u>always takers</u> are the ones that independent of their quarter of birth stop studying later, accomplishing more years of education

<u>never takers</u> are the ones that independent of their quarter of birth they always get a lower number of years of education

in c) I assumed defiers do not exist (MONOTONICITY)

e) <u>EXCLUSION RESTRICTION FAILS</u> if quarter of birth affects income <u>not</u> only through education

<u>BJ1996</u> :




- performance in school
  - → behavior
  - → attendance
  - → performance reading, writing, arithmetic ...
- health differences
  - → mental health
  - → multiple schlerosis
  - → dyslexia
  - → autism ...
- birth seasonality related with
  - regional patterns
  - race
  - family income
  - personality

<u>BJ1995</u> : • WEAK INSTRUMENT PROBLEM

→ association between quarter of birth and education is very weak

↓

BIAS if present is inflated

↓

even a small correlation between quarter of birth and wages is likely to badly bias the effect of education on earnings

f) exclusion restriction :
law does influence income only through education

exogeneity :
error term in 1st stage uncorrelated with law, but only to education

g) through information about behaviour good for health : healthier life style and food

i) Education $= \gamma_0 + \gamma_1$ quarter of birth $+ \gamma_2$ married $+ \gamma_3$ SMSA $+ \gamma_4$ region 1 $+ \dots + \gamma_{m+4}$ region $m$ $+ \mu_i$

Health $= \delta_0 + \delta_1$ Education $+ \delta_2$ Married $+ \delta_3$ SMSA $+ \delta_4$ region 1 $+ \dots + \delta_{m+4}$ region $m$ $+ v_i$

no : inclusion of income violates exclusion restriction ; income and health are correlated and income is also correlated with education

K) The 1st stage is very strong in all specifications (67.04, 203.37, 316.81, 362.97)
Finite sample bias might be an issue if exclusion restriction does not hold
We reject the null of no correlation between quarter of birth and health outside the one mediated by education

$$\begin{tikzcd} QB \arrow[r] \arrow[rr, bend right] & E \arrow[r] & H \end{tikzcd}$$

do not reject null this is 0

p value $= 0.0004$

l) if instrument valid I expect
OLS to be upward bias, bigger than 2SLS

income is for sure an omitted variable in the OLS regression between education and health, income increases both education and good health and probably it's likely to be associated with lower physical risk jobs

m) the difference in income caused by finishing or not one particular year of schooling because of your birth quarter. You should expect 2SLS to be smaller than the OLS. We should expect the instrument to impact only on people at the margin between completing or not the school-year when they reach the minimum age to leave school.

n) No, I'm not convinced. Exclusion restriction is likely not to hold.