Econometrics II - Workgroup 4

Due date: Sunday, January 29, 11.59 pm. Hand in your solutions as a **single .pdf file** including your code via Canvas. Include your R (or any other language) code by using R Markdown (preferred) or by using the package "minted" in your .tex file (see a template on Canvas). Each team has to come up with a unique name (without names or student numbers). Both teammates have to submit solutions via Canvas. Each subquestion has an equal number of points.

Problem 1: Judges and Prison Sentences

A researcher is interested in the effect of prison sentence on later criminal behavior. The researcher collects data from one particular court, which has two active judges. The researcher learns that cases are assigned randomly to the judges. However, judge Jones convicts someone to a prison sentence in 70% of the cases while judge Smith only convicts someone to a prison sentence in 40% of the cases. Next, the researcher collects information on whether or not an individual is arrested in the three years after the trial.

Judge	Jones		Smith	
Sentences	Prison	Other	Prison	Other
Cases	70%	30%	40%	60%
Future arrests	40%	60%	20%	50%

(*i*) Use the Wald estimator to compute the causal effect of a prison sentence on the probability of being arrested later.

(*ii*) What is the interpretation of the estimated effect? And for which fraction of the population does this causal effect hold?

(*iii*) Explain what an always taker is in this setting and which fraction of the population are always takers?

Problem 2: Eating and exams

A teacher is interested in how important eating and drinking is for academic achievement. Therefore, the teacher decides to run a field experiment. The teacher randomizes students into two groups. The control group is allowed to have breakfast and snacks before the exam. The treatment group is only allowed to have a glass of milk in the morning and drink water during the remaining time before the exam. The key outcome variable of interest is passing the exam. In past years about 50% of the students passed the exam. The teacher thinks that not eating or drinking will reduce the passing rate with 10%-points. The teacher is aiming for 70% power in the field experiment and a significance level of 5% (recall $t_{0.3} = -0.524$ and $t_{0.975} = 1.960$). The teacher randomizes students equally over the treatment and control group.

(*i*) Perform a power calculation for the number of students that the teacher should include in the field experiment.

The teacher is afraid that some people in the treatment group will actually have breakfast. Furthermore he assumes that all people in the control group will have breakfast.

(*ii*) The teacher assumes that 20% of the students randomized in the treatment group will actually have breakfast. How does this change the number of students required to participate in the field experiment?

Problem 3: Mother's labor supply

An extensive literature addresses gender differences in labor market outcomes. This literature often points towards a substantial child penalty for mothers, i.e. female labor supply is reduced after birth of a child which may be harmful for the future career. A common problem in the empirical analysis of the child penalty is that fertility decisions are likely endogenous. To deal with this endogeneity, Angrist and Evans (AER, 1998) have proposed an instrumental variable strategy. Their idea is that the gender of a child is random, but if the first two children of a mother have the same sex, the mother is more likely to have a third child. Within the econometric framework, having a third child is the treatment variable and the instrumental variable is if the first two children have the same sex. The outcome variables are hours worked and income of the mother.

The 1980s sample of Angrist and Evans (AER, 1998) will be used to answer the questions below. From this sample mothers are excluded that had fewer than two children or a twin birth. The dataset contains the following variables:

Variable	Description
SEXK	1 if the first born child is a girl
SEX2ND	1 if the second born child is a girl
SAMESEX	1 of the first and second born child have the same gender
CHILD3	1 if a third child is born
KIDCOUNT	total number of children born
HOURSM	weekly hours of work of the mother
INCOME1M	annual labor income of the mother

(i) Compute the fractions of girls among the first and among the second born children. And regress the gender of the second child on the gender of the first child.

(ii) Regress having a third child on whether the first two children have the same gender (first-stage regression). Is the instrumental variable sufficiently strong? To get some idea about what the treatment of having a third child entails also regress the total number of children on whether the first two children have the same gender.

(iii) What is the share of always takers, compliers and never takers in this approach?

(iv) Use IV regression to estimate the local average treatment effects of having a third child on working hours and income of the mother.

(v) Compute mean working hours and mean income for the four subgroups that can be defined by having a third child and the first two children having the same sex. Use these means to compute the mean outcomes of the always takers, never takers and compliers with and without the first two children having the same gender. (Hint: use Imbens and Rubin (1997) discussed in the slides.) Use these means to say something about the preference of having a third child.

(vi) Does this selection concur with the bias of an OLS regression of the mother's labor market outcomes on having a third child?

(vii) Stratify the sample by the gender of the first child and perform the instrumental variable regression again. Use the first-stage results to explain the heterogeneous effects.