# Environmental Econometrics: Causality Homework Assignment

## Eric Strobl

### October 13, 2022

**Instructions:** This homework assignment is due on the $10^{th}$ of October. For every (starting) week late after that, one grade will be deducted. Please send a .pdf file of your solution to eastrobl13@gmail.com with joelhuesler@gmail.com in copy, and the subject headline "Problem Sheet 1".

## 1 Neyman-Rubin Model

*Context:* Heat waves are associated with decreases in general population well-being and with increases in mortality and morbidity, especially in vulnerable population groups. Importantly, the number of heat extremes has substantially increased across Europe in recent decades. As a matter of fact, it is widely believed that these heat waves have caused tens of thousands of premature deaths in Europe since 2000. For example, the heat wave in the summer of 2003 is estimated to have killed over 70,000 in Southern Europe.

You are interested in whether individuals above 70 years of age that live in a place where there is a heat wave would die. Let $D$ be a binary variable where $D=1$ indicates if an elderly person lives where there is a summer heat wave, and $0$ otherwise. Let $Y=1$ indicate whether the person dies, and $0$ otherwise

Consider the following sample of 6 individuals, and that assignment to treatment is random:

| $i$ | $D_i$ | $Y_{1i}$ | $Y_{0i}$ |
|-----|-------|----------|----------|
| 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 0 |
| 3 | 1 | 1 | 0 |
| 4 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 |
| 6 | 0 | 1 | 0 |

(i) Calculate (a) the average treatment effect (ATE), (b) the average treatment effect on the treated (ATT), (c) the average treatment effect on the controls (ATC), and (d) the naive estimator of the ATE; Why might these differ?

(ii) Provide examples of situations when a policy maker might be interested in the ATE, ATT, and ATC

(iii) Now assume that treatment is not random. (a) Looking at the table above, in what sense do you expect a selection bias? What might be reasons for this? (b) Calculate by how much the naive estimator will differ from the ATT and the ATC.

(iv) Assume again that treatment assignment is random, but now that the effect of heat waves in affected areas spills over to those in non-affected areas because health spending is increased in the former and decreased in the latter, and this reduces the health in the latter by 1. Calculate how would this affect your expected ATE.

## 2 Regression Analysis

Let's say you are interested in estimating whether the mortality rate of 70+ year olds in France will increase if heat waves increase.

*Data:* To this end you collect data for communes with over 20,000 inhabitants in the South of France on the number of days with above 35 degrees during July and August of 2003, as well as the mortality rate of people of 70 years and older. You also have data on the per capita number of doctors of each commune.

Assume you estimate the following regression:

$$MortalityRate_i = \alpha + \beta_{HeatDays}HeatDays_i + \beta_{Doctors}Doctors_i + \epsilon_i$$

where $Mortality$ is the per capita mortalty rate of 70+ year olds, $HeatDays$ is the number of days with at least 35 degrees, $Doctors$ is the doctor per capita rate, and $\epsilon$ is an iid error term.

(i) (a) Do you expect any attenuation bias? (b) Why? (c) What implications will this have? (d) How could one avoid such attenuation bias?

(ii) (a) Do you expect omitted variable bias? (b) Why? (c) Provide two examples of possible omitted variables and speculate as to what direction this bias will take (and why)

(iii) In terms of the *external validity* of your study, discuss what layers may stand between your findings and the question you sought to answer

(iv) Assume that you want to inform policy makers about what your study found in terms of the relationship between 70+ year old mortality rate and heat days during the summer, and that your estimate on $\beta_{HeatDays}$ is 0.03. Taking account all the possible drawbacks of your estimated effect, explain your finding in one sentence.