Example Assignment: The Effect of Alcohol Withdrawal on Wage Incomes

Identification and Causal Inference

January 30, 2019

1 Introduction

We examine the effect of an alcohol withdrawal treatment on wage income in a sample of 5'000 alcoholics in Switzerland. We observe that only a subgroup of our sample takes the treatment (goes to rehab). Since each alcoholic above the age of 40 is in principle entitled to the treatment but decides by herself whether or not she wants to take it, we are worried that selection into the treatment is not random. Most likely, more motivated alcoholics would rather go to rehab, but would also have higher wage incomes compared to non-treated alcoholics. This is the case even in the absence of the treatment because of the individual's personal characteristics. Our causal question most likely suffers from omitted variable bias, since we cannot measure personal traits such as motivation.

(a) In an experiment, we would assign alcoholics randomly to the alcohol withdrawal programme and could capture the causal effect by comparing wage incomes of treated with untreated alcoholics.

(b) The optimal dataset would reveal information on all personal traits that affect selection into treatment and labour market outcomes such as motivation, ability, etc. simultaneously.

(c) The most important threat to our identification strategy is that selection into treatment in the real world is not random and might thus bias the results. The second most important threat is the validity of our instrument.

2 Data Generation and Model Specification

We consider a sample consisting of 5'000 observations. The outcome variable $income_i$ represents monthly labor income of individual i. The binary treatment variable $D_i \in \{0,1\}$ indicates whether individual i has participated in an alcohol withdrawal programme. Other covariates are age_i, $educ_i$ and $motiv_i$. The age of individual i is given by age_i, which is uniformly distributed between 25 and 65. Education follows a discrete uniform distribution with values between 1 and 5, which means that each individual is assigned one out of five education levels. There is an unobserved variable $motiv$ following a uniform distribution between 0 and 1, which determines the level of intrinsic motivation for each individual.

$$\begin{aligned} age &\sim U(25, 65) \\ motiv &\sim U(0, 1) \\ educ &\sim U\{1, .., 5\} \end{aligned}$$

In the true model, given by equation (1), the dependent variable $income$ is a linear function of a age, age squared, education, motivation and treatment.

$$income_i = \beta_0 + \beta_1 * age_i + \beta_2 * age_i^2 + \beta_3 * educ_i + \beta_4 * motiv_i + \Delta_i * D_i + v_i \tag{1}$$

The intercept and coefficients for each variable are selected as shown in equation (2) such that the data generated is reasonable in a stylized real life scenario.

$$income_i = \mu(0) + 12 * age_i - 0.05 * age_i^2 + 30 * educ_i + 120 * motiv_i + \Delta_i * D_i + v_i \tag{2}$$

where $\mu(0) = 4000$, the treatment effect is given by $\Delta_i = \mu(1) - \mu(0) = 4300 - 4000 = 300$ and the error term v is normally distributed, $v \sim N(0, 50)$. Because motivation is unobserved, we can only estimate equation (3).

$$income_i = b_0 + b_1 * age_i + \beta_2 * age_i^2 + \beta_3 * educ_i + \Delta_i * D_i + e_i \tag{3}$$

where $e_i = v_i + b_4 * motiv_i$. The following section will show that self-selection into treatment depends on the motivation of each individual. As a result, the model suffers from omitted variable bias because the regressors are correlated with the error term, $\text{Cov}(D_i, e_i) \neq 0$. Hence, the conditional independence assumption (CIA) is not fulfilled and a causal interpretation of the coefficients of equation 3 is not possible.

3 Assignment into Treatment

Assignment into treatment occurs through two different rules. The first rule is a hypothetical scenario in which treatment is assigned randomly to about half the population. The second rule is a stylized real life scenario in which people self-select into treatment and where we have in addition a discontinuity because only people aged 40 or older are allowed into treatment due to funding restrictions.

- **Selection Rule 1: Random Assignment**
  In an experiment, treatment is randomly assigned to individuals irrespective of their characteristics.

$$D_i^* \sim U(-1, 1)$$

$$D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0 \\ 0 & \text{if } D_i^* < 0 \end{cases}$$

- **Selection Rule 2: Self-Selection** Alcoholics select into treatment according to their individual characteristics. These are age, education, unobserved motivation and distance to the next rehab center, where $dist \sim U(0, 100)$. Because of poor health care funding and scientific evidence that older patients are less likely to relapse, treatment can only be made available to alcoholics that are older than 40 years.

$$D_i^* = -150 - 1 * dist_i + 1 * age_i + 10 * educ_i + 40 * motiv_i + u_i$$

where $u \sim N(0, 20)$. Because treatment is only available to alcoholics aged 40 or older, the decision rule is different from the random case.

$$D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0 \quad \& \quad age_i \geq 40 \\ 0 & \text{else} \end{cases}$$

Table I shows the mean values for the full sample, the treated and non-treated alcoholics given they were selected randomly for treatment and the treated and the non-treated alcoholics given that they were able to self-select into treatment.

Table I: Summary Statistics

	Total	Random Assignment		Self-Selection	
		Treatment	No Treatment	Treatment	No Treatment
Income		4973	4674	5072	4626
Age	45	45	45	53	41
Education	3.0	3.0	3.0	3.4	2.8
Motivation	0.5	0.50	0.49	0.56	0.46
Distance	50.0	50	50	32	58
Observations	5000	2531	2469	1633	3367

If treatment had been assigned randomly to about half of the entire population, there is no obvious difference in age, education, distance to the next rehab center or unobserved motivation to undergo such a treatment. The raw difference in income is CHF 299, which is very close to the causal effect of the treatment of CHF 300. Hence, we can already conclude that a random assignment allows us to identify the treatment effect by calculating the difference in the mean outcomes between the treated and the non-treated. In reality, only alcoholics above 40 are allowed to undergo treatment and there is self-selection. This means that alcoholics that are more motivated, higher educated and live closer to a rehab center are more likely to enter rehab. As a result, the difference in income is larger than in the random case. This is clearly visible in the income distributions in figures 1 and 2.

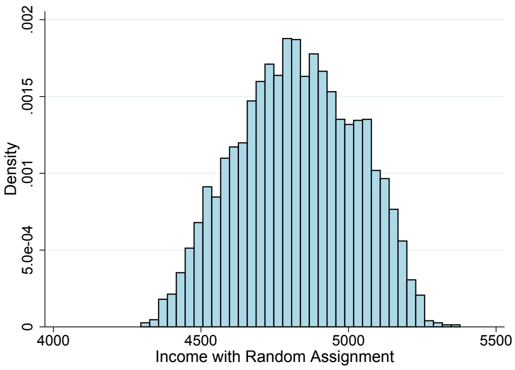

Figure 1: Randomly Assigned Treatment

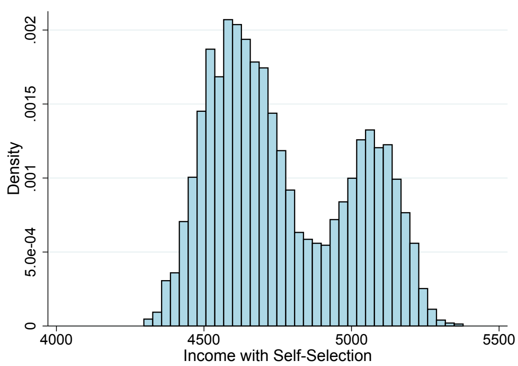

Figure 2: Self-Selected Treatment