2 The counterfactual framework

1. Causal parameters and the selection problem.

   (a) Explain, in your own words, the interpretation of (and difference across) the causal parameters discussed in the lecture: the average treatment effect ATE, the average treatment effect on the treated ATT, the average treatment effect on the untreated $ATUT$, and the naive estimate Δ^N.

   (b) Show that the ATE can be written as a weighted average of the ATT and the ATUT. Discuss why identification fails in general, i.e. discuss which of the involved quantities are observable and which are not observable from data on $\{y_i, d_i\}_{i=1}^N$. Carefully discuss the assumption(s) needed in order to identify the ATE.
   Hint: Use the Law of Iterated Expectations.

   (c) Show that the ATT (ATUT) can be written as the sum of the ATE and a bias term.
   Hint: Note that Y_{di} can always be written as $Y_{di} = \mathbb{E}(Y_d) + (Y_{di} - \mathbb{E}(Y_d))$.

   (d) Show that the Δ^N is the sum of the ATT and a bias term. Discuss the condition which is sufficient for $\Delta^N =$ ATT. Give a substantive example in which this assumption is *unlikely* to hold.

2. Self selection.

   Assume that you want to evaluate the effects of a training program. There is no randomization and thus individuals are allowed to self-select into the program. Assume that potential earnings are given by:

   $$Y_{0i} = \beta_0 + \epsilon_{0i} \tag{1}$$
   $$Y_{1i} = \beta_0 + \beta_1 + \epsilon_{1i} \tag{2}$$

   Assume that $\beta_1 > 0$. Also assume that $\mathbb{E}(\epsilon_{di}) = 0$ (which is only a normalization) and that $Cov(\epsilon_{0i}, \epsilon_{1i}) = 0$.

   Further assume that people self-select into the training program based on the following rule:

   $$D_i = \mathbf{1}(Y_{1i} > Y_{0i}), \tag{3}$$

   That is, people enter the program if earnings with training are larger than earnings without training.
   Hint: Use the results from exercise (1) above.

   (a) Derive the ATE.

   (b) Derive the ATT.

   (c) Which one is larger? What is the intuition for this result?

3. Randomization at work.

   For the subsequent practical exercises, use the data from the file `randomization.dta`. The file contains part real, part hypothetical data (you actually see the *counterfactual*).

   (a) Explain (show) why randomization of the treatment solves the selection problem.

   (b) How would you implement randomization in practice?

   (c) Explain how you would use a software to randomize the treatment across individuals. Implement your algorithm to the sample data.

(d) How could you check whether randomization worked?

(e) Estimate the ATE and the ATT using the analogy principle. (In the end, estimate the ATE on your statistical software. You would need to create a variable y containing the realized outcome (y_0 or y_1) for each observation.)

(f) Estimate the ATE using a simple regression model, i.e. run the following simple regression model:

$$y_i = \alpha + \beta d_i + u_i \tag{4}$$

where d_i is the randomized treatment variable you created in (c).
Show that the two parameters of the regression model can be written as follows:

$$\hat{\alpha} = \overline{y}_0 \tag{5}$$

$$\hat{\beta} = \overline{y}_1 - \overline{y}_0 \tag{6}$$

Hint: This exercise is not as straightforward as it looks (if you want to derive the result in detail). Start with the usual formula for $\hat{\beta}$ and then use the fact that d_i is a binary variable in combination with the properties of the summation operator (you'll also need the LIE along the way).

(g) What happens if you estimate the causal parameter by a linear regression model including additional regressors? Will the estimate for the causal parameter differ from the simple estimator from above? Why (or why not)?
Hint: Don't show this formally. Think about omitted variable bias.

(h) Why may it nonetheless be useful to include additional variables in the regression model? Explain.
Hint: Just give the intuition for this result. What happens to the residual variance if you include additional regressors?