# Econometrics II. Assignment 5: Difference-in-differences design

**Due date: Sunday, February 6, 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.

### Question 1: Parallel trends assumption

Consider a treatment and control group. Looking only at the pre-treatment period, they have exactly the same outcomes (zero gap between them in each period).

*(i)* Despite having exactly the same pre-treatment outcomes, it happens to be the case that parallel trends assumption is violated. How is this possible? Explain what it means for parallel trends assumption to be violated, and give an example of how it could be violated.

*(ii)* What would be the main problem of applying the difference-in-difference estimator in this case? What would this depend on?

### Question 2: Mortality and medicine

In this exercise you are going to work with historical data on mortality rates and medical innovations. In the mid-1930s, an important event in the history of modern medicine took place: sulfa drugs were introduced. Sulfa drugs were the first medicine to effectively treat a range of potentially fatal bacterial infections, most notably scarlet fever. By as early as 1937, sulfa drugs were widely available and dispensed in the United States. There is strong evidence from clinical trials that sulfa drugs were effective against some infectious diseases, but not others, for instance, tuberculosis was not treatable with sulfa drugs. Thus, it is possible to estimate the effect of sulfa drugs on mortality using a difference-in-differences approach that compares mortality differences between treated (scarlet fever) and control (tuberculosis) diseases before and after 1937.
The dataset `assignment5.csv(\.dta)` contains the following variables:

| Variable | Description |
|---|---|
| lnm_rate | log of the mortality rate in state $i$ for disease $d$ at year $t$ |
| treated | 1 if a disease is scarlet fever and 0 if tuberculosis |
| state | state variable |
| year | year variable |

*(i)* Plot unconditional means of log mortality rates by the type of disease for all years. Comment on your graph.

*(ii)* Using only data for the years 1936 and 1937, make a table with the mean log mortality rate for treated and control diseases before and after the introduction of sulfa drugs. Use the numbers from the table to calculate the difference-in-differences estimator.

*(iii)* Using only data for the years 1936 and 1937, estimate a difference-in-difference regression. Comment on your results.

*(iv)* Using all years, estimate a difference-in-difference regression. To do that, you need to create an indicator variable equal to 1 for the years 1937–1943 and equal to 0 for the years 1925–1936. What is the interpretation of the difference-in-differences coefficients? What do you conclude about the effect of sulfa drugs on mortality rates?

*(v)* Estimate an event-study specification. Comment on your results.

*(vi)* Argue at which level you need to cluster standard errors. Implement your suggested cluster-robust standard errors. Comment on your results.

*(vii)* Do a test of whether the prior trends differ between the treated and control groups. What do you conclude?