

**AIL 7310: Machine Learning for Economics**  
**AY 2025-26 Semester I, Assignment 3**

---

**I. Difference-in-difference.** Use the provided dataset titled 'did\_data.csv'

It contains information on different economic regions (region\_id) which receive a certain subsidy from the government shown in 'treatment' in a given year. We want to know the causal impact of receiving this subsidy on the average wage in this region.

- a) Construct a variable 'treated' which takes the value 1 for treated observations and 0 for control observations.
- b) Construct a variable 'post' which takes the value 1 for all post-treated time periods i.e. 2010 or later and 0 for all pre-treated time periods.
- c) Plot the average y over time separately for treated and control groups. Describe whether the parallel trends assumption appears plausible before treatment (2006–2009).
- d) Using the previously defined 'treated' and 'post' variables to estimate the Difference-in-Difference regression model and interpret the causal estimate.
- e) Now add the other control variables 'population', 'unemployment', 'gdp\_per\_capita', 'exports\_per\_capita', 'fdi\_inflow' and re-estimate the causal parameter.
- f) Explore heterogeneity: estimate the DiD separately by the three sectors : Agricultural, Manufacturing and Services. How much do they vary? [This is called 'Heterogeneous treatment effects' and we will come back to this in the last module.]

**II. Regression Discontinuity.** Use the dataset titled 'rdd\_data'.

It contains information on difference students and their normalized score in a test in 5<sup>th</sup> standard. Those who score more than 0 get a scholarship. We want to estimate the causal effect of getting this scholarship on their 10<sup>th</sup> standard test scores.

- a) Create the treatment variable D based on the rule above.
- b) Plot the 'hours\_studied' and 'mother\_edu' against '5<sup>th</sup>\_score' to illustrate continuity around the cutoff.
- c) Plot the '10<sup>th</sup>\_score' versus '5<sup>th</sup>\_score' to illustrate the discontinuity.
- d) Estimate the regression discontinuity model including all the other covariates as discussed in class and estimate the causal parameter.