

## Problem Set 3: Causal Inference

1. This problem pertains to simulating a data set from the scratch.

There are 3 groups: i) early treated, ii) late treated, and iii) untreated and 10 periods.

- a. Start with a sample of 10,000 for each group; so 1,000 observations for group-period cell.
  - b. Label the groups: “group1”, “group2”, and “group3”.
  - c. Group 1 is treated in the 3rd period (early treated), group 2 is treated in the 7th period, and group 3 is untreated throughout the sample. Generate these treatment indicators.
  - d. Simulate the outcome. Both group1 and group2 has a treatment effect of 20. Be sure to add the following:
    - Group 2 has a starting point of 5 units higher than group 1.
    - Group 3 has a starting point of 2 units higher than group 1.
    - time effect is linear.
  - e. Next, collapse the outcome at the group-period cell and name it a new dataframe. Plot the figure depicting the outcome by groups over time.
  - f. Estimate the TWFE model specification. Discuss the twfe in relation to the ATT estimate.
2. This is doing the same simulation as 1, but treatment effect is heterogeneous by group. Group 1 has an ATT of 10, whereas group 2 has an ATT of 20.
    - a. Re-simulate the data and estimate the TWFE.
    - b. Estimate the TWFE using: i) group1 and untreated group; ii) group2 and untreated group.
    - c. Discuss the implications of using TWFE estimate as an estimate representing the ATT in this case.
  3. This is doing the same simulation as 1, but treatment effect is heterogeneous by period. Group 1’s ATT starts at 3 and increases over time by 1 unit each period. Group 2’s ATT starts at 10 and increases by 1 unit each period.
    - a. Re-simulate the data and estimate the TWFE.
    - b. Discuss the implications of using TWFE estimate as an estimate representing the ATT in this case.
    - c. Estimate the event study model capturing the treatment dynamics.