Proof that Difference-in-Differences Recovers ATT

Alexander Dean

March 20, 2025

In the two-period setting, we define:

- $t = 0$ (pre-treatment) and $t = 1$ (post-treatment).
- Two groups: Treated ($D = 1$) and Control ($D = 0$).
- Outcome of interest: Y_{it}.

The observed outcome can be written as:

$$Y_{it} = Y_{it}(0) + D_i \cdot (Y_{it}(1) - Y_{it}(0))$$

where:

- $Y_{it}(0)$ is the potential outcome without treatment.
- $Y_{it}(1)$ is the potential outcome under treatment.
- D_i is an indicator for treatment.

The Average Treatment Effect on the Treated (ATT) is defined as:

$$ATT = E[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1]$$

Under the parallel trends assumption, in the absence of treatment, the treated group would have followed the same trend as the control group:

$$E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0]$$

The Difference-in-Differences estimator is given as:

$$\hat{\delta}_{DiD} = (E[Y_{i1} \mid D_i = 1] - E[Y_{i0} \mid D_i = 1]) - (E[Y_{i1} \mid D_i = 0] - E[Y_{i0} \mid D_i = 0])$$

Substituting the decomposition:

$$\hat{\delta}_{DiD} = (E[Y_{i1}(1) \mid D_i = 1] - E[Y_{i0}(0) \mid D_i = 1]) - (E[Y_{i1}(0) \mid D_i = 0] - E[Y_{i0}(0) \mid D_i = 0])$$

Using parallel trends assumption:

$$E[Y_{i1}(0) \mid D_i = 1] - E[Y_{i0}(0) \mid D_i = 1] = E[Y_{i1}(0) \mid D_i = 0] - E[Y_{i0}(0) \mid D_i = 0]$$

After simplification:

$$\hat{\delta}_{DiD} = E[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1] = ATT$$

$$\hat{\delta}_{DiD} = ATT \quad \square$$

Thus, under the parallel trends assumption, the Difference-in-Differences estimator (with a canonical set-up!) consistently recovers the ATT.