Difference in Difference Estimation

Bo Hou

Bo.hou@bthft.nhs.uk

- DiD originates from econometrics but now is widely used in other social sciences to evaluate policy and interventions.
- DiD is a quasi-experimental design that makes use of longitudinal data from treatment and control groups to obtain an appropriate counterfactual to estimate a causal effect.
- Aims to explain simple DiD and the rationale of this method.

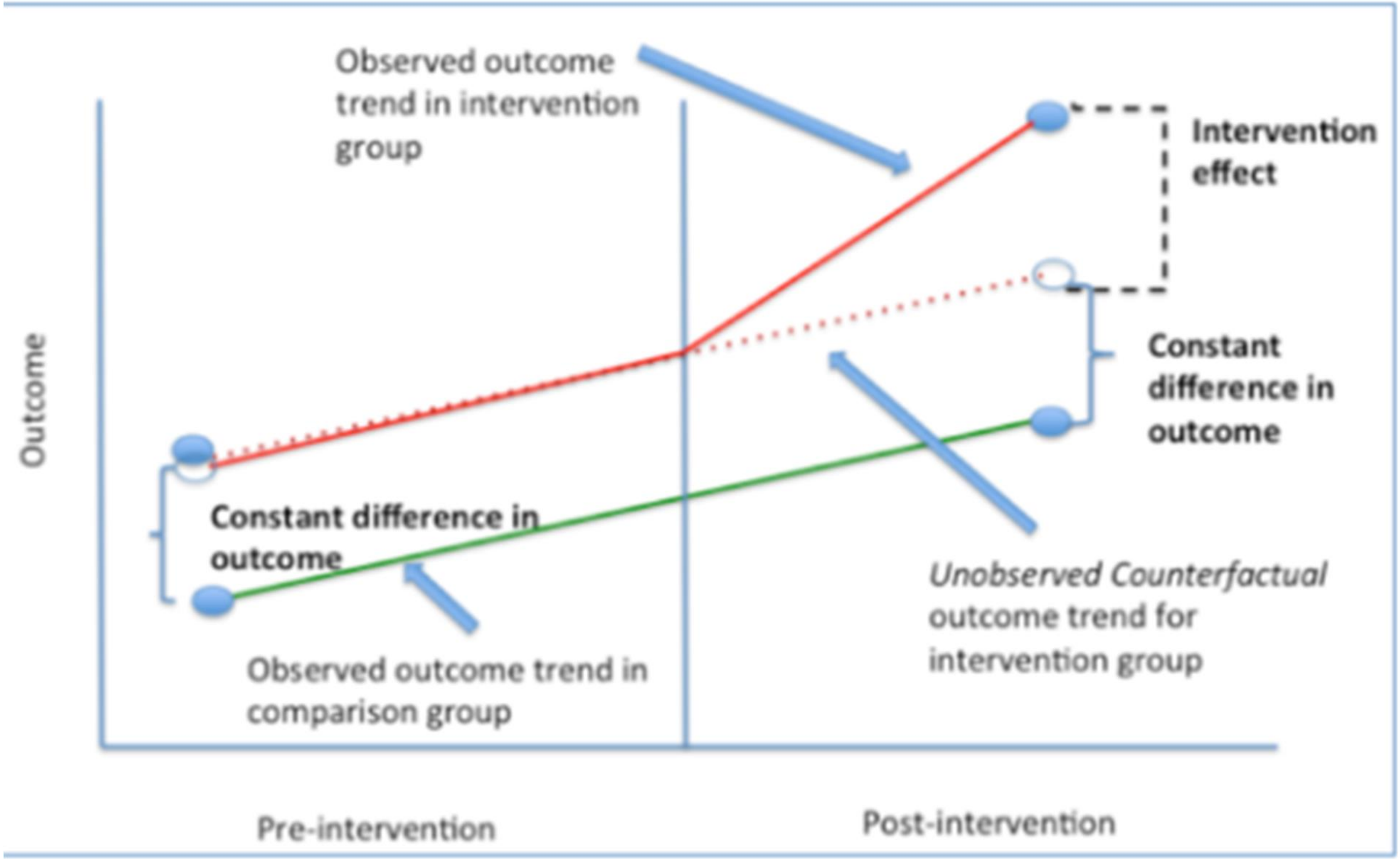

Figure 1. Difference-in-Difference estimation, graphical explanation

Assumptions

- exchangeability, positivity, and Stable Unit Treatment Value Assumption (SUTVA)
- treatment/intervention and control groups have **Parallel Trends in outcome**
- allocation of intervention was not determined by outcome

Strengths

- Intuitive interpretation
- Can obtain causal effect using observational data if assumptions are met
- Comparison groups can start at different levels of the outcome. (DID focuses on changerather than absolute levels)
- Accounts for change/change due to factors other than intervention

Limitations

- Requires baseline data & a non-intervention group
- Cannnot use if intervention allocation determined by baseline outcome
- Cannot use if comparison groups have different outcome trend
- Cannot use if composition of groups pre/post change are not stable

Simple DiD in details

The outcome Y_i is modeled by the following equation

$$Y_i = \alpha + \beta T_i + \gamma t_i + \delta (T_i \cdot t_i) + \varepsilon_i \qquad \text{(Outcome)}$$

where the coefficients given by the greek letters $\alpha, \beta, \gamma, \delta$, are all unknown parameters and ε_i is a random, unobserved "error" term which contains all determinants of Y_i which our model omits. By inspecting the equation you should be able to see that the coefficients have the following interpretation

α = constant term

β = treatment group specific effect (to account for average permanent differences between treatment and control)

γ = time trend common to control and treatment groups

δ = true effect of treatment

The purpose of the program evaluation is to find a "good" estimate of δ, $\hat{\delta}$, given the data that we have available.

2.1 Simple Pre versus Post Estimator

Consider first an estimator based on comparing the average difference in outcome Y_i before and after treatment *in the treatment group alone.*[1]

$$\hat{\delta}_1 = \bar{Y}_1^T - \bar{Y}_0^T \tag{D1}$$

Taking the expectation of this estimator we get

$$\begin{aligned} E\left[\hat{\delta}_1\right] &= E\left[\bar{Y}_1^T\right] - E\left[\bar{Y}_0^T\right] \\ &= [\alpha + \beta + \gamma + \delta] - [\alpha + \beta] \\ &= \gamma + \delta \end{aligned}$$

which means that this estimator will be biased so long as $\gamma \neq 0$, i.e. if a time-trend exists in the outcome Y_i then we will confound the time trend as being part of the treatment effect.

2.2 Simple Treatment versus Control Estimator

Next consider the estimator based on comparing the average difference in outcome Y_i post-treatment, between the treatment and control groups, *ignoring pre-treatment outcomes.*[2]

$$\hat{\delta}_2 = \bar{Y}_1^T - \bar{Y}_1^C \tag{D2}$$

Taking the expectation of this estimator

$$\begin{aligned} E\left[\hat{\delta}_1\right] &= E\left[\bar{Y}_1^T\right] - E\left[\bar{Y}_1^C\right] \\ &= [\alpha + \beta + \gamma + \delta] - [\alpha + \gamma] \\ &= \beta + \delta \end{aligned}$$

and so this estimator is biased so long as $\beta \neq 0$, i.e. there exist permanent average differences in outcome Y_i between the treatment groups. The true treatment effect will be confounded by permanent differences in treatment and control groups that existed prior to any treatment. Note that in a randomized experiments, where subjects are randomly selected into treatment and control groups, β should be zero as both groups should be nearly identical: in this case this estimator may perform well in a controlled experimental setting typically unavailable in most program evaluation problems seen in economics.

The **difference in difference** (or "double difference") estimator is defined as the difference in average outcome in the treatment group before and after treatment *minus* the difference in average outcome in the control group before and after treatment[3]: it is literally a "difference of differences."

$$\hat{\delta}_{DD} = \bar{Y}_1^T - \bar{Y}_0^T - \left(\bar{Y}_1^C - \bar{Y}_0^C\right) \tag{DD}$$

Taking the expectation of this estimator we will see that it is unbiased

$$\begin{aligned}
\hat{\delta}_{DD} &= E\left[\bar{Y}_1^T\right] - E\left[\bar{Y}_0^T\right] - \left(E\left[\bar{Y}_1^C\right] - E\left[\bar{Y}_0^C\right]\right) \\
&= \alpha + \beta + \gamma + \delta - (\alpha + \beta) - (\alpha + \gamma - \gamma) \\
&= (\gamma + \delta) - \gamma \\
&= \delta
\end{aligned}$$

This estimator can be seen as taking the difference between two pre-versus-post estimators seen above in (D1), subtracting the control group's estimator, which captures the time trend γ, from the treatment group's estimator to get δ. We can also rearrange terms in equation (DD) to get $\hat{\delta}_{DD} = \bar{Y}_1^T - \bar{Y}_1^C - \left(\bar{Y}_0^T - \bar{Y}_0^C\right)$ in which can be interpreted as taking the difference of two estimators of the simple treatment versus control type seen in equation (D2). The difference estimator for the pre-period is used to estimate the permanent difference β, which is then subtracted away from the post-period estimator to get δ.

References

- https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation
- https://eml.berkeley.edu/~webfac/saez/e131_s04/diff.pdf
- https://www.stata.com/news/conferences/arm/DID-slides.pdf
- #Stata command didregress