Basic DiD

- Notation:
  - $T = \mathbf{1}\{\text{Post Treatment Period}\}$.
  - $G = \mathbf{1}\{\text{In the Treatment Group}\}$.
  - $D = GT$.
- Question: Can we identify ATT?
- What is ATT: $E[Y_1 - Y_0|D = 1] = E[Y_1 - Y_0|T = 1, G = 1]$.
- What we can identify:
  - $E[Y_1|T = 1, G = 1]$
  - $E[Y_0|T = 0, G = 0]$
  - $E[Y_0|T = 1, G = 0]$
  - $E[Y_0|T = 0, G = 1]$.
- But no $E[Y_0|T = 1, G = 1]$

Time Trend and Regression

- Need additional assumption to identify the ATT
- Most common one: Parallel Trend Assumption
- Time trend for $G = 0$:
  $E[Y_0|G = 0, T = 1] - E[Y_0|G = 0, T = 0]$ equals to time trend for $G = 1$:
  $E[Y_0|G = 1, T = 1] - E[Y_0|G = 1, T = 0]$.
- Note: Only $E[Y_0|G = 1, T = 1]$ not directly observed. Identified with assumption.
- In regression, people regress Y on $1, G, T, GT = D$.

Pre-Trend Evaluation

- To evaluate the parallel trend assumption, people look at pre-period trends
  - If in pre-period treatment and control evolve similarly
  - then the counterfactual will hopefully be the same
- Eye-ball tests are the most common
- In theory no reason to require the trend to be flat at 0
- Logically a straight trend should also be acceptable
- But in practice most people want things to be flat

Implementation

- Let post period dummy be P_t
- Let treatment group dummy be T_i
- $Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 T_i + \beta_3 P_t \times T_i + \epsilon_{it}$
- Parameter of interest is β_3
- Usually people put control variables as well
- $Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 T_i + \beta_3 P_t \times T_i + \gamma X_{it} + \epsilon_{it}$
- With multiple periods, put time and individual fixed effects instead
- $Y_{it} = \delta_t + \delta_i + \beta^{DD} P_t \times T_i + \gamma X_{it} + \epsilon_{it}$

Remark

- DiD fundamentally relies on parametric assumption
- eg., Parallel trend for y and $log(y)$ will not hold at the same time
- No pre-period trends do not guarantee the identification of ATT
- Treatment should still not be a result of selection

Extension to Multiple Treatment

- Usually multiple periods e.g., staggered difference-in-difference
- Many people used to run two-way fixed effects
  - $y_{it} = \alpha_i + \alpha_t + \beta^{DD} D_{it} + \epsilon_{it}$.
  - $y_{it} = \alpha_i + \alpha_t + \sum_r \beta_r \mathbf{1}\{R_{it} = r\} + \epsilon_{it}$ where R_{it} is relative time.
- Econometricians find problems on this around 2020

Assumptions Extensions

Parallel Trend

- If treatment had not occurred, avg. outcome for all groups would evolve in parallel

No Anticipation

- Units don't act on the knowledge of future treatment dates before treatment starts

Does TWFE Work?

$$Y_{it} = \alpha_i + \alpha_t + \beta^{DD} D_{it} + u_{it}$$

- Suppose there is only heterogeneous treatment effect in time (denoted τ_s)
- i.e., the effect 1 year after v.s. 2 years after are different ($\tau_1 \neq \tau_2$)
- $\beta^{DD} = \sum_s \omega_s \tau_s$, where ω_s could be negative
- Issue: τ_s could be all positive but $\beta_{\text{post}} < 0$
- Source of problem: using the treated as control
- So TWFE does not work!

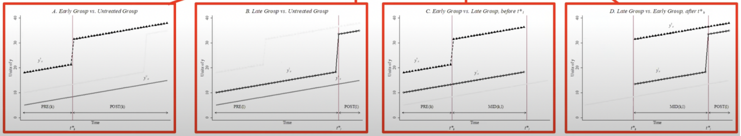
Illustrations (Baker 2019)

What is $\widehat{\beta}^{DD}$?

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = s_{kU} \boxed{\hat{\beta}^{DD}_{kU}} + s_{\ell U} \boxed{\hat{\beta}^{DD}_{\ell U}} + \left[s^{k}_{k\ell} \boxed{\hat{\beta}^{DD,k}_{k\ell}} + s^{\ell}_{k\ell} \boxed{\hat{\beta}^{DD,\ell}_{k\ell}} \right]$$



Illustrations continued (Baker 2019)

What is $\hat{\beta}^{DD}$?

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}^{DD}_{kU} + s_{\ell U}\hat{\beta}^{DD}_{\ell U} + \left[s^{k}_{k\ell}\hat{\beta}^{DD,k}_{k\ell} + s^{\ell}_{k\ell}\hat{\beta}^{DD,\ell}_{k\ell}\right]$$

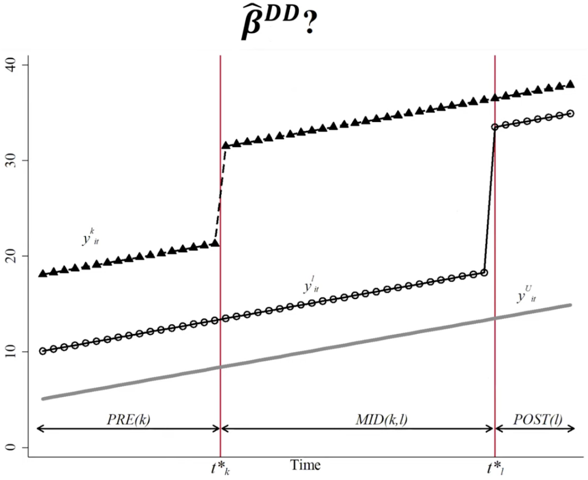
$$s_{kU} = \frac{(n_k + n_U)^2 n_{kU}(1 - n_{kU})\bar{D}_k(1 - \bar{D}_k)}{V(D_{it})}$$

$$s^{k}_{k\ell} = \frac{\left((n_k + n_\ell)(1 - \bar{D}_\ell)\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell}\frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}{V(\tilde{D}_{it})}$$

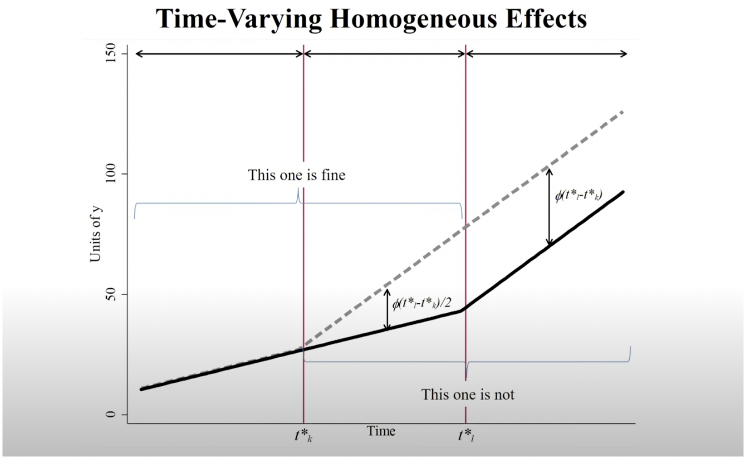
$$s^{\ell}_{k\ell} = \frac{\left((n_k + n_\ell)\bar{D}_k\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}\frac{\bar{D}_\ell}{\bar{D}_k}}{V(\tilde{D}_{it})}$$

Subsample variance of treatment

Illustrations continued (Baker 2019)



When things go wrong (Baker 2019)



Does Dynamic TWFE Work?

$$Y_{it} = \alpha_i + \alpha_t + \sum_{r \neq 0} \mathbf{1}\{R_{it} = r\}\beta_r + u_{it}$$

- where R_{it} is the relative event time
- This would work if only heterogeneity in time
- But not if different cohort has different cohort effect
- Again use the treated as control
- Does not work!

What to Do?

- Two Approaches
  - Diagnosis: Check if any weights are negative
  - New Estimators: Avoid using the treated as controls

New Estimators

- Idea: we still know how to do cohort-specific simple DiD
- Pick the valid comparisons and use those only
- A never-treated group could be useful
- Can also use later-treated groups
- Many packages now available

Parallel Trends Assumptions

- Potentially you need to control for observables to make trends parallel
- Assess your pre-period trends with a plot
- The common pre-trend tests have low power though, need to be careful
- Many sensitivity methods available now