did2s

Two-Stage Difference-in-Differences

Kyle Butts

March 10, 2022

TWFE Model

Unit i at time t has outcome y_{it} given by the standard TWFE model:

$$y_{it} = \mu_i + \eta_t + \tau D_{it} + \varepsilon_{it},$$

η_t are common time shocks, μ_i are time-invariant unit characteristics, and D_{it} is a treatment dummy.

Researchers care about the average treatment effect, τ.

TWFE Model

Unit i at time t has outcome y_{it} given by the standard TWFE model:

$$y_{it} = \mu_i + \eta_t + \tau D_{it} + \varepsilon_{it},$$

η_t are common time shocks, μ_i are time-invariant unit characteristics, and D_{it} is a treatment dummy.

Researchers care about the average treatment effect, τ.

If the model is correctly specified and parallel trends holds, then OLS is fine!

Thank you!

TWFE Model

Problem #1

τ is not constant.

TWFE Model

Problem #1

τ is not constant.

- Treatment effects may depend on when you start treatment
  - → *e.g.*, groups that benefit more from a policy implement it earlier
- Treatment effects may depend on treatment duration (event study!)
  - → *e.g.*, policy doesn't affect everyone right away

TWFE Model

New and Improved

Our TWFE model clearly needs to be enriched:

$$y_{it} = \mu_i + \eta_t + \tau_{gt} D_{it} + \varepsilon_{it}$$

Now, we have a group-time average treatment effect, τ_{gt}, which allows treatment effect size to depend on when you are treated g and how many periods it has been since treatment determined by $t - g$.

TWFE Model

New and Improved

Our TWFE model clearly needs to be enriched:

$$y_{it} = \mu_i + \eta_t + \tau_{gt} D_{it} + \varepsilon_{it}$$

Now, we have a group-time average treatment effect, τ_{gt}, which allows treatment effect size to depend on when you are treated g and how many periods it has been since treatment determined by $t - g$.

We want to know the overall average treatment effect:

$$\tau \equiv 1/N_{gt} \sum \tau_{gt}$$

Estimating overall average treatment effect

If we knew μ_i and η_t, then we could move terms around in our model and have:

$$y_{it} - \mu_i - \eta_t = \tau_{gt} D_{it} + \varepsilon_{it}$$

Then, if we ran *this* regression of a variable on an indicator variable, OLS algebra that $\hat{\tau}$ will take the average of τ_{gt} for us and estimate τ.

Too bad we don't know μ and η

Since we don't know μ and η, we have to estimate them. Using the FWL theorem:

$$y_{it} - \hat{\mu}_i - \hat{\eta}_t = \tau_{gt} \tilde{D}_{it} + u_{it},$$

where $\tilde{D}_{it}$ is the *residualized* treatment dummy.

TWFE Model

Problem #2

$$y_{it} - \hat{\mu}_i - \hat{\eta}_t = \tau_{gt} \tilde{D}_{it} + u_{it},$$

The left-handside is a good estimate for τ_{gt}, so what's the problem??

TWFE Model

Problem #2

$$y_{it} - \hat{\mu}_i - \hat{\eta}_t = \tau_{gt} \tilde{D}_{it} + u_{it},$$

The left-handside is a good estimate for τ_{gt}, so what's the problem??

All of the modern diff-in-diff problems (yes all of them) show different ways to interpret the same problem: $\tilde{D}_{it}$.

Since we have residualized D_{it}, OLS no longer computes the simple average of τ_{gt}'s.

TWFE Model

Problem #2

$$y_{it} - \hat{\mu}_i - \hat{\eta}_t = \tau_{gt} \tilde{D}_{it} + u_{it},$$

Instead, OLS estimates $\hat{\tau} = \sum w_{gt} \tau_{gt} \neq \tau$

TWFE Model

Problem #2

$$y_{it} - \hat{\mu}_i - \hat{\eta}_t = \tau_{gt} \tilde{D}_{it} + u_{it},$$

Instead, OLS estimates $\hat{\tau} = \sum w_{gt} \tau_{gt} \neq \tau$

The weights w_{gt} can be super weird. In extreme cases, $\hat{\tau}$ can be the opposite sign of the overall average treatment effect, τ.

Fixing the problem

$$y_{it} - \hat{\mu}_i - \hat{\eta}_t = \tau_{gt} \tilde{D}_{it} + u_{it},$$

Okay, so why don't you just regress $y_{it} - \hat{\mu}_i - \hat{\eta}_t$ on D_{it}? Great idea!

Two-stage difference-in-differences

Stage 1: Estimate μ_i and η_t using untreated/not-yet-treated observations ($D_it = 0$). Don't include $D_{it} = 1$ since the treatment effects will be partially absorbed by fixed effects.

Stage 2: Regress $y_{it} - \hat{\mu}_i - \hat{\eta}_t$ on D_{it}.

Two-stage difference-in-differences

Stage 1: Estimate μ_i and η_t using untreated/not-yet-treated observations ($D_it = 0$). Don't include $D_{it} = 1$ since the treatment effects will be partially absorbed by fixed effects.

Stage 2: Regress $y_{it} - \hat{\mu}_i - \hat{\eta}_t$ on D_{it}.

Inference is done in `did2s`.

Conclusion

Thank you!!

for real this time...