DiD Session Discussion, or How I Learned to Stop Worrying and Love Untestable Assumptions

Eli Ben-Michael

CMU

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

. . . but sometimes people get weird with it

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

... but sometimes people get weird with it

Empirical researchers are doing X, is that ok?

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

... but sometimes people get weird with it

Empirical researchers are doing X, is that ok?

3 sub-questions behind the main one

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

... but sometimes people get weird with it

Empirical researchers are doing X, is that ok?

3 sub-questions behind the main one

1. What is the estimator actually estimating, under the assumptions I **think** I'm making

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

... but sometimes people get weird with it

Empirical researchers are doing X, is that ok?

3 sub-questions behind the main one

1. What is the estimator actually estimating, under the assumptions I **think** I'm making
2. What do we want to estimate?

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

... but sometimes people get weird with it

Empirical researchers are doing X, is that ok?

3 sub-questions behind the main one

1. What is the estimator actually estimating, under the assumptions I **think** I'm making
2. What do we want to estimate?
3. Under what conditions are they the same?

What the heck are we estimating!?

DiD/panel data: popular identification/estimation strategies

. . . but sometimes people get weird with it

Empirical researchers are doing X, is that ok?

3 sub-questions behind the main one

1. What is the estimator actually estimating, under the assumptions I **think** I'm making
2. What do we want to estimate?
3. Under what conditions are they the same?

Follow up: can we find an estimator that estimates what we want, under the assumptions that we're willing to make?

A core result

A core result

For a 2x2 DiD

$$\hat{\tau}_{11}^{\text{DiD}} = \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 1] - \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 0] \quad \text{(difference in post ATTs)}$$
$$- (\mathbb{E}[Y_{\text{pre}}(1) - Y_{\text{pre}}(0) \mid G = 1] - \mathbb{E}[Y_{\text{pre}}(1) - Y_{\text{pre}}(0) \mid G = 0]) \quad \text{(difference in pre ATTs)}$$

A core result

For a 2x2 DiD

$$\hat{\tau}_{11}^{\text{DiD}} = \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 1] - \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 0] \quad \text{(difference in post ATTs)}$$
$$- (\mathbb{E}[Y_{\text{pre}}(1) - Y_{\text{pre}}(0) \mid G = 1] - \mathbb{E}[Y_{\text{pre}}(1) - Y_{\text{pre}}(0) \mid G = 0]) \quad \text{(difference in pre ATTs)}$$

LS: No anticipation, only group 1 treated $\Rightarrow$ all ATTs are 0 except for post group 1

$$\hat{\tau}_{11}^{\text{DiD}} = \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 1]$$

A core result

For a 2x2 DiD

$$\hat{\tau}_{11}^{\text{DiD}} = \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 1] - \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 0] \quad \text{(difference in post ATTs)}$$
$$- (\mathbb{E}[Y_{\text{pre}}(1) - Y_{\text{pre}}(0) \mid G = 1] - \mathbb{E}[Y_{\text{pre}}(1) - Y_{\text{pre}}(0) \mid G = 0]) \quad \text{(difference in pre ATTs)}$$

LS: No anticipation, only group 1 treated $\Rightarrow$ all ATTs are 0 except for post group 1

$$\hat{\tau}_{11}^{\text{DiD}} = \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 1]$$

XZD: No anticipation, both groups treated $\Rightarrow$ pre-ATT is 0 for both groups

- insight: useful to define add'l layer of POs based on group $Y_t(g, z)$

$$\begin{aligned}\hat{\tau}_{11}^{\text{DiD}} &= \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 1] - \mathbb{E}[Y_{\text{post}}(1) - Y_{\text{post}}(0) \mid G = 0] \\ &= E[Y_{\text{post}}(1,1) - Y_{\text{post}}(1,0) \mid G = 1] - \mathbb{E}[Y_{\text{post}}(0,1) - Y_{\text{post}}(0,0) \mid G = 0]\end{aligned}$$

What do we want to estimate?

LS: In typical DiD setting, want a notion of average treatment effect on the treated

$$\text{effect} = \sum_{\text{times } t} \sum_{\text{ever trt groups } g} \text{weight}_{gt} \times \text{ATT}_g(t)$$

- For some aggregation strategy, what would be the average difference in outcomes under treatment and control across all groups and times?

What do we want to estimate?

LS: In typical DiD setting, want a notion of average treatment effect on the treated

$$\text{effect} = \sum_{\text{times } t} \sum_{\text{ever trt groups } g} \text{weight}_{gt} \times \text{ATT}_g(t)$$

- For some aggregation strategy, what would be the average difference in outcomes under treatment and control across all groups and times?

XZD: The **causal interaction** between group status and treatment

$$\text{interaction effect} = \mathbb{E}[(Y_{\text{post}}(1,1) - Y_{\text{post}}(1,0)) - (Y_{\text{post}}(0,1) - Y_{\text{post}}(0,0))]$$

- What would be the average difference in effects, if everyone was in group 1 vs. group 0?

Are we estimating what we want under the assumptions we're making?

Are we estimating what we want under the assumptions we're making?

LS: Under staggered adoption, the static and dynamic DiD estimators are estimating a weird thing that is really just a description of heterogeneity

Are we estimating what we want under the assumptions we're making?

LS: Under staggered adoption, the static and dynamic DiD estimators are estimating a weird thing that is really just a description of heterogeneity

XZD: Under generalized DiD, the 2x2 DiD estimator is estimating a weird thing that is really just a description of heterogeneity

Are we estimating what we want under additional assumptions?

Are we estimating what we want under additional assumptions?

Are we estimating what we want under additional assumptions?

LS: Restrictions on heterogeneity

Are we estimating what we want under additional assumptions?

LS: Restrictions on heterogeneity

- Static DiD
  - No time-varying effects within groups (Assumption 4), or
  - Same calendar time effects across groups (Assumption 5)

Are we estimating what we want under additional assumptions?

LS: Restrictions on heterogeneity

- Static DiD
  - No time-varying effects within groups (Assumption 4), or
  - Same calendar time effects across groups (Assumption 5)
- Dynamic DiD
  - Same relative time effects across groups (Assumption 6)

Are we estimating what we want under additional assumptions?

LS: Restrictions on heterogeneity

- Static DiD
  - No time-varying effects within groups (Assumption 4), or
  - Same calendar time effects across groups (Assumption 5)
- Dynamic DiD
  - Same relative time effects across groups (Assumption 6)
- These assumptions are unpalatable (and non-nested)

Are we estimating what we want under additional assumptions?

LS: Restrictions on heterogeneity

- Static DiD
  - No time-varying effects within groups (Assumption 4), or
  - Same calendar time effects across groups (Assumption 5)
- Dynamic DiD
  - Same relative time effects across groups (Assumption 6)
- These assumptions are unpalatable (and non-nested)

XZD: Restrictions on heterogeneity

- Generalized parallel trends assumptions (Assumption 4)
- Parallel trends across both levels of both treatments ($2 \times 2 = 4$ sets of constraints)
- What this buys us: conditioning on group doesn't matter, and so description = causal

Are we estimating what we want under additional assumptions?

LS: Restrictions on heterogeneity

- Static DiD
  - No time-varying effects within groups (Assumption 4), or
  - Same calendar time effects across groups (Assumption 5)
- Dynamic DiD
  - Same relative time effects across groups (Assumption 6)
- These assumptions are unpalatable (and non-nested)

XZD: Restrictions on heterogeneity

- Generalized parallel trends assumptions (Assumption 4)
- Parallel trends across both levels of both treatments ($2 \times 2 = 4$ sets of constraints)
- What this buys us: conditioning on group doesn't matter, and so description = causal
- Is this assumption palatable?
  - Possible diagnostic with multiple periods and add'l (strong!) assumption of no effect carryover

Is there a way to get us what we want without restricting heterogeneity?

Is there a way to get us what we want without restricting heterogeneity?

LS: Review of heterogeneity-robust estimators

- Several options to address this in the literature
- And it makes a difference in applications!

Is there a way to get us what we want without restricting heterogeneity?

LS: Review of heterogeneity-robust estimators

- Several options to address this in the literature
- And it makes a difference in applications!

XZD: Unclear

- Are there opportunities to use multiple periods?
- Are we just out of luck?

XZD: Exclusion restriction and cannonical DiD

GDID setting recovers typical DiD setting under **exclusion restriction** (Assumption 5)

$$Y_{\text{post}}(0,1) = Y_{\text{post}}(0,0)$$

$\Rightarrow$ ATT for group 0 is 0, and they are "clean controls"

XZD: Exclusion restriction and cannonical DiD

GDID setting recovers typical DiD setting under **exclusion restriction** (Assumption 5)

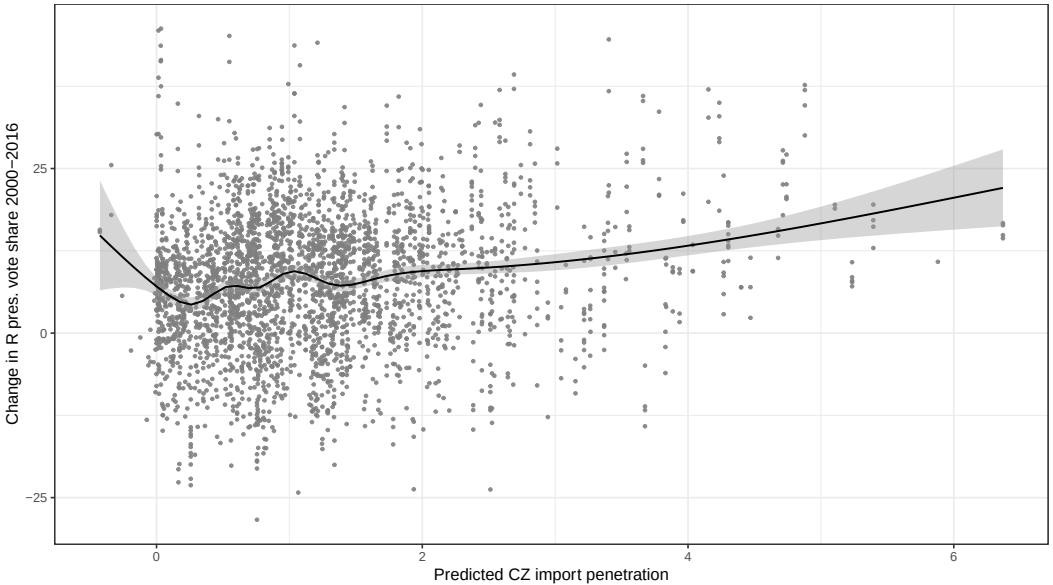
$$Y_{\text{post}}(0,1) = Y_{\text{post}}(0,0)$$

$\Rightarrow$ ATT for group 0 is 0, and they are "clean controls"

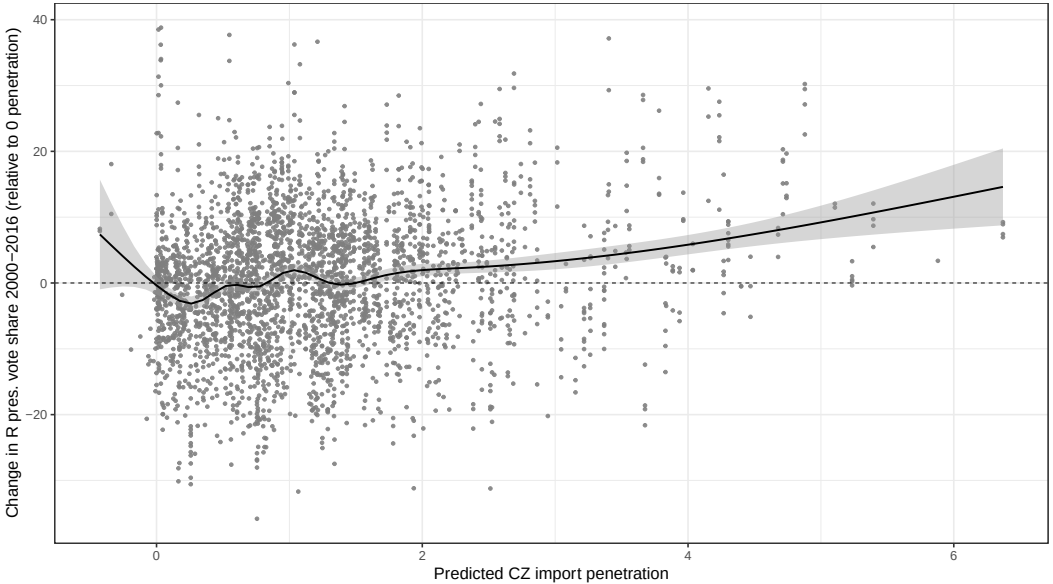
Are Bartik instruments/shift-share designs the other side of the coin?

- Groups defined by G have different level of exposure to Z
- e.g. the "China shock" [Autor et al., 2013, 2020]
- IV-style identification, but also fixed-effects/first differencing
  [Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022; Borusyak and Hull, 2023]
- Focus is on estimating effect of Z, not interaction, but does the GDID-style of analysis have something to say about shift-share designs, and vice versa?

Heterogeneity in impacts of “China shock” on presidential elections?



Heterogeneity in impacts of “China shock” on presidential elections?



LS: Miscelleneous notes

LS: Miscelleneous notes

To what extent are these issues are a matter of degree?

- Discussion about omitted relative time period
  - "Omitting a baseline period that does not appear for all units... places a much greater weight on the effect homogeneity assumption" (pg 28)
  - "Omitting pre-treatment relative time indicators that are too far from the start of treatment risks placing greater weight on the effect homogeneity assumption" (pg 35)

LS: Miscelleneous notes

To what extent are these issues are a matter of degree?

- Discussion about omitted relative time period
  - "Omitting a baseline period that does not appear for all units... places a much greater weight on the effect homogeneity assumption" (pg 28)
  - "Omitting pre-treatment relative time indicators that are too far from the start of treatment risks placing greater weight on the effect homogeneity assumption" (pg 35)

Including never-treated units vs sometimes-treated units

- Paglayan and Hall & Yoder examples: evidence of pre-trends with sometimes-treated units
- Q in paper: does including never treated units make it *worse*?
- Regardless, implies we shouldn't include sometimes-treated units

LS: Miscelleneous notes

To what extent are these issues are a matter of degree?

- Discussion about omitted relative time period
  - "Omitting a baseline period that does not appear for all units... places a much greater weight on the effect homogeneity assumption" (pg 28)
  - "Omitting pre-treatment relative time indicators that are too far from the start of treatment risks placing greater weight on the effect homogeneity assumption" (pg 35)

Including never-treated units vs sometimes-treated units

- Paglayan and Hall & Yoder examples: evidence of pre-trends with sometimes-treated units
- Q in paper: does including never treated units make it *worse*?
- Regardless, implies we shouldn't include sometimes-treated units

Strong form of parallel trends $\Rightarrow$ need to check for pre-trends across all pairs of groups?

- Omnibus diagnostic for strong parallel trends?
- Removing comparison pairs w/o parallel trends?
- But also pre-testing in problematic [Roth, 2018]

Your turn: Q&A

References I

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2020). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *American Economic Review*, 110(10):3139–3183.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103(6):2121–2168.

Borusyak, K. and Hull, P. (2023). Nonrandom Exposure to Exogenous Shocks. *Econometrica*, 91(6):2155–2185. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA19367.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies*, 89(1):181–213.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Roth, J. (2018). Should we condition on the test for pre-trends in difference-in-difference designs? *arXiv preprint arXiv:1804.01208*.