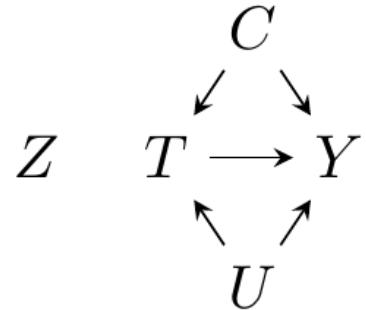


Estimating BHET Policy Impacts

Sam Harper

2022-03-14

Thinking about research design



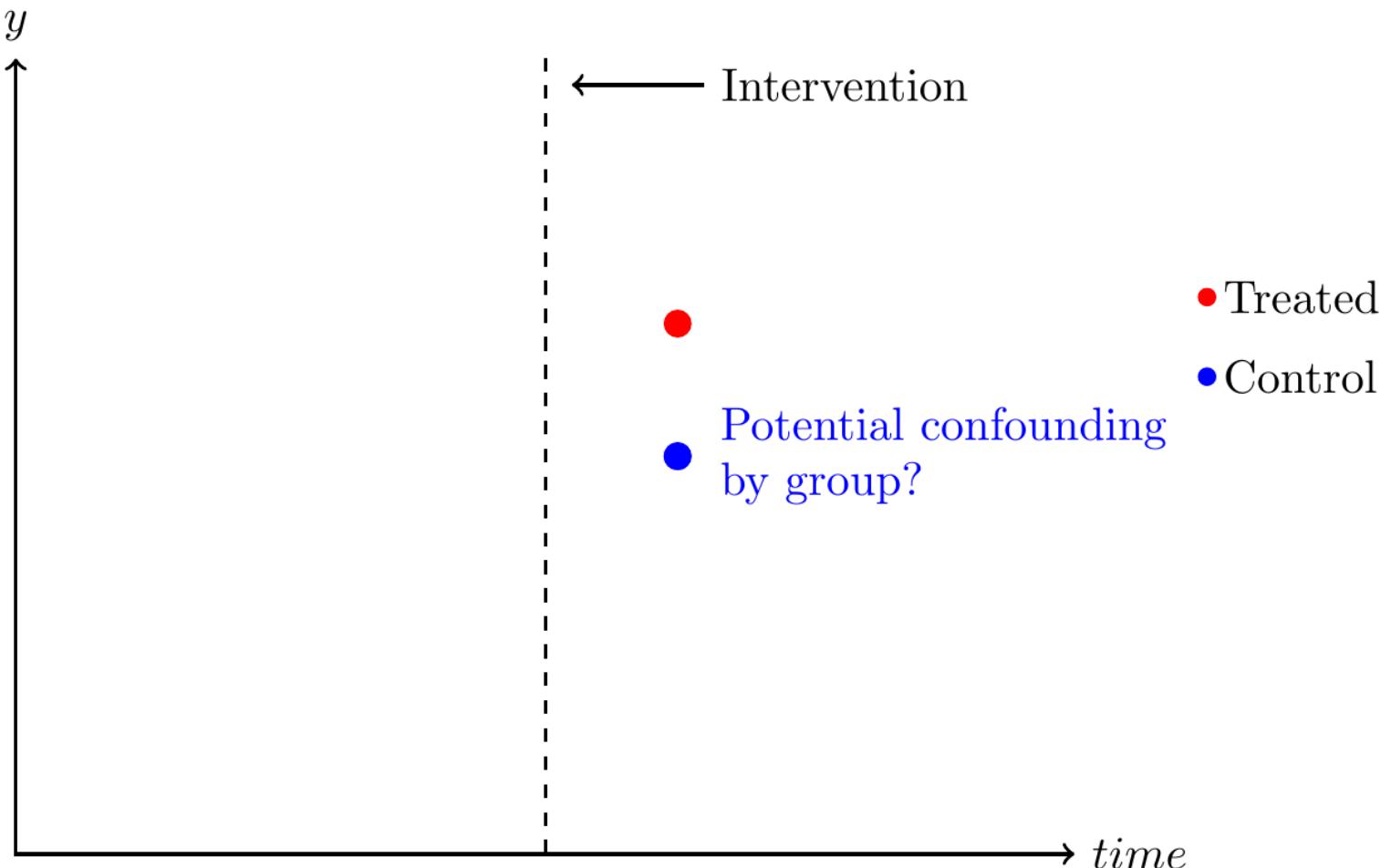
- Goal: estimate treatment [ban] effect (T) on our outcomes (Y).
- Recall the potential outcomes framework. We want to estimate what *would have happened* in ban villages without the ban:

$$E[Y^1 - Y^0] = E[Y^1|T = 1] - E[Y^0|T = 0]$$

- Where should we get our counterfactual?
- Ideally randomize (Z), but not feasible.
- Instead we can control for covariates (C), but still worry about unmeasured confounders (U).

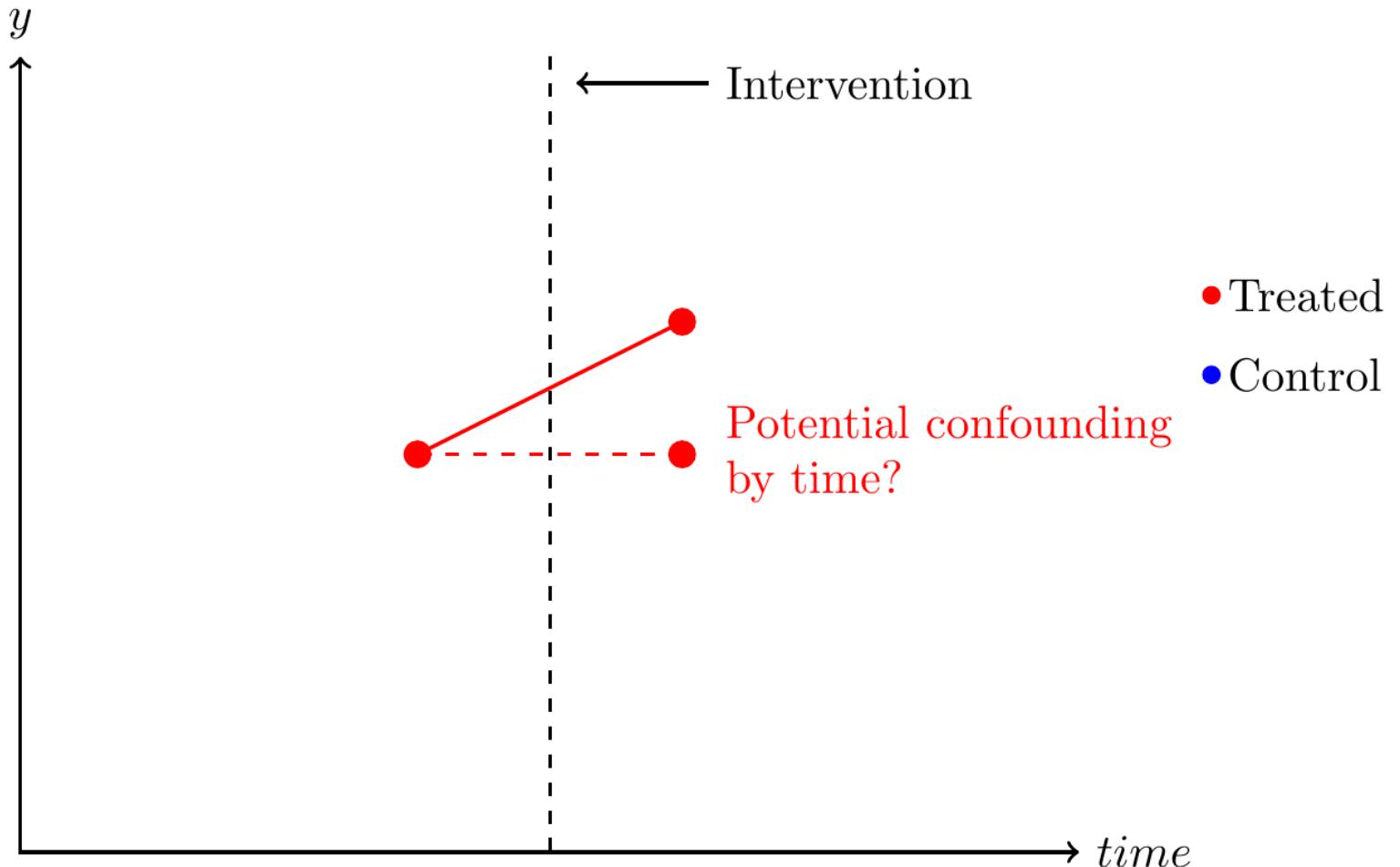
T vs. C after ban?

- Could compare "ban" and "no ban" villages after policy.
- Treated and control villages may differ in ways that are hard to measure.



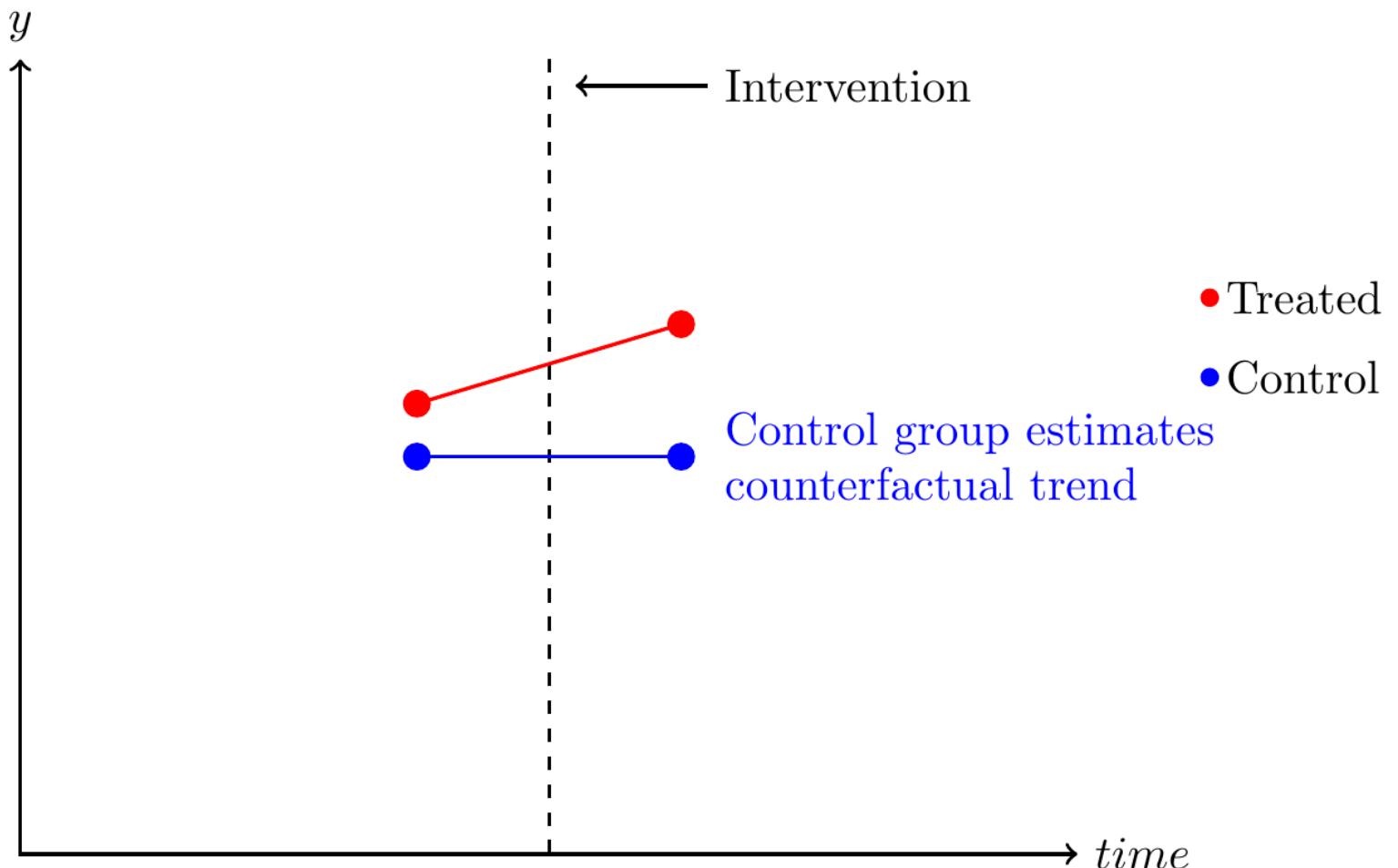
Pre-post in treated?

- Could compare "ban" villages before vs. after policy.
- Other factors affecting (Y) can lead to bias.

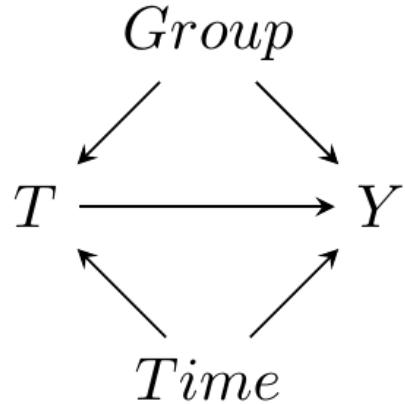


Add pretests for both

- Compares pre/post in treated and untreated.
- Captures all time-invariant group diffs.
- Captures all group-invariant time trends.



Difference-in-Differences: Basic Idea



- The average change over time in the control group is subtracted from the change over time in the treated group.
- Double differencing removes biases in second period comparisons between the treatment and control group that could result from:
 - Fixed differences between those groups.
 - Comparisons over time in the treatment group that could be the result of time trends unrelated to the treatment.

Difference-in-Differences without Regression

- DD is just differences in means!

Let $\mu_{it} = E(Y_{it})$

- $i = 0$ is control group, $i = 1$ is treatment.
- $t = 0$ is pre-period, $t = 1$ is post-period.
- One 'difference' estimate of causal effect is: $\mu_{11} - \mu_{10}$ (pre-post in treated)

- Differences-in-Differences estimate of causal effect is:
$$(\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00})$$

Area	Before	After	Difference
Treated	135	100	-35
Control	80	60	-20
T - C	55	40	-15

Difference-in-differences regression: How? The 2x2 case

- Single treated and control group, two periods
- β_1 = Treated group
- β_2 = Post period
- β_3 = Product term

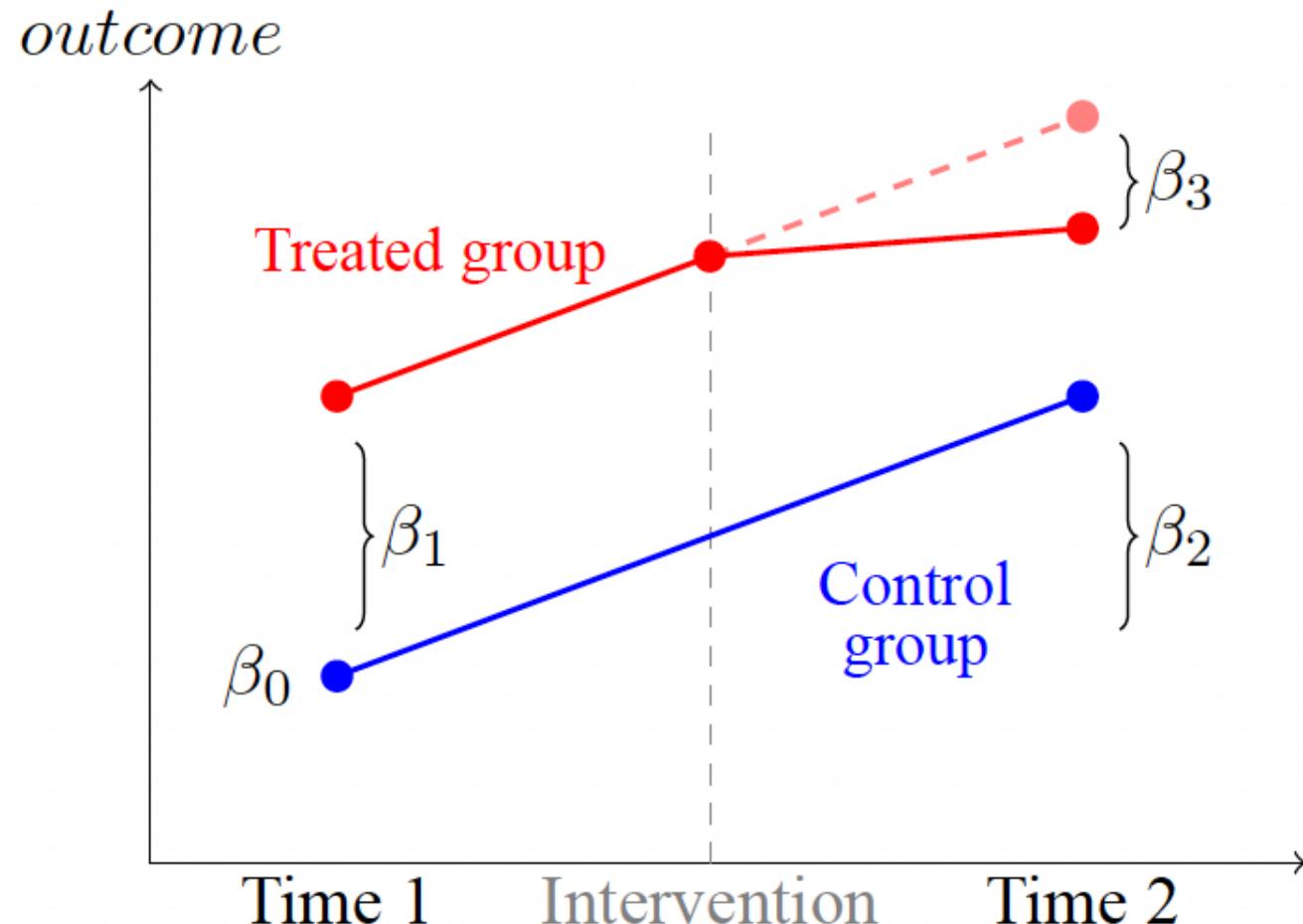
y	group	time	treat?	post?	treatXpost
:	1	1	0	0	0
:	1	2	0	1	0
:	2	1	1	0	0
:	2	2	1	1	1

$$Y = \beta_0 + \beta_1 * treat + \beta_2 * post + \beta_3 * treat * post$$

Visual representation

- β_0 : control pre
- β_1 : group difference pre
- β_2 : pre-post diff in control
- β_3 : DD estimate

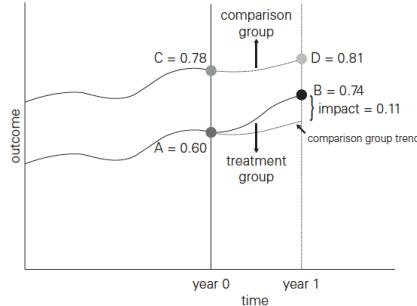
$$Y = \beta_0 + \beta_1 Treat + \beta_2 Post + \beta_3 Treat * Post + \varepsilon_t$$



Difference-in-differences (usually) estimates the ATT

- Our DD model is: $Y = \beta_0 + \beta_1 Treat + \beta_2 Post + \beta_3 Treat * Post + \varepsilon$.
 - we showed that β_3 is the DD estimate in the linear model case.
- In the (possibly counterfactual) absence of intervention, the expected outcome is:
 - $E(Y_i^0 | T = 1, P = 0) = \beta_1 + \beta_2$
- In the (possibly counterfactual) presence of intervention, the expected outcome is:
 - $E(Y_i^1 | T = 1, P = 1) = \beta_1 + \beta_2 + \beta_3$
- ATT is the expected difference in $Y_i^1 - Y_i^0$ for those treated in the post-period:
 - $ATT = E(Y^1 - Y^0 | T = 1) = \beta_3$

Key Assumption: Parallel Trends



- Basic DD controls for any time invariant characteristics of both treated and control groups.
- Does not control for any **time-varying** characteristics.
- If another policy/intervention occurs in the treated (or control) group at the same time as the intervention, we cannot cleanly identify the effect of the program.
- DD main assumption: in the absence of the intervention treated and control groups would have displayed similar **trends**.
- This is called the *parallel trends* assumption.

Note: Impossible to verify. See recent work by [Rambachan, et al. \(2019\)](#) and [Roth, et al. \(2020\)](#) for ways to investigate.

Reformulation of the model using 'fixed effects'

Express our earlier model using 'fixed effects':

- Dummy for Group
- Dummy for Time
- *Time-varying policy indicator*

$$Y = \beta_0 + \beta_1 * Group2 + \beta_2 * Time2 + \beta_3 * policy$$

y	group	time	treat?	post?	treatXpost	Group 2	Time 2	policy
:	1	1	0	0	0	0	0	0
:	1	2	0	1	0	0	1	0
:	2	1	1	0	0	1	0	0
:	2	2	1	1	1	1	1	1

- β_3 still estimates the 'difference-in-differences' parameter.

What about multiple treated groups?

- Easy to rewrite our earlier model for multiple groups **treated at the same time.**
- 3 units and 3 time periods.
- Groups 1 and 3 implement policy at **T2**.
- **g2** and **g3** are dummies for groups 2&3
- **t2** and **t3** are dummies for times 2&3.
- **policy** is a time-varying dummy when policy is in place.

y	group	time	policy	g2	g3	t2	t3
:	1	1	0	0	0	0	0
:	1	2	1	0	0	1	0
:	1	3	1	0	0	0	1
:	2	1	0	1	0	0	0
:	2	2	0	1	0	1	0
:	2	3	0	1	0	0	1
:	3	1	0	0	1	0	0
:	3	2	1	0	1	1	0
:	3	3	1	0	1	0	1

Basic setup for DD with variable timing ("two way fixed effects")

More generally, you could write the basic equation with multiple group (γ_g) and time (τ_t) fixed effects as:

$$Y_{gt} = \gamma_g + \tau_t + \delta^{DD} p_{gt} + \varepsilon_{st}$$

where

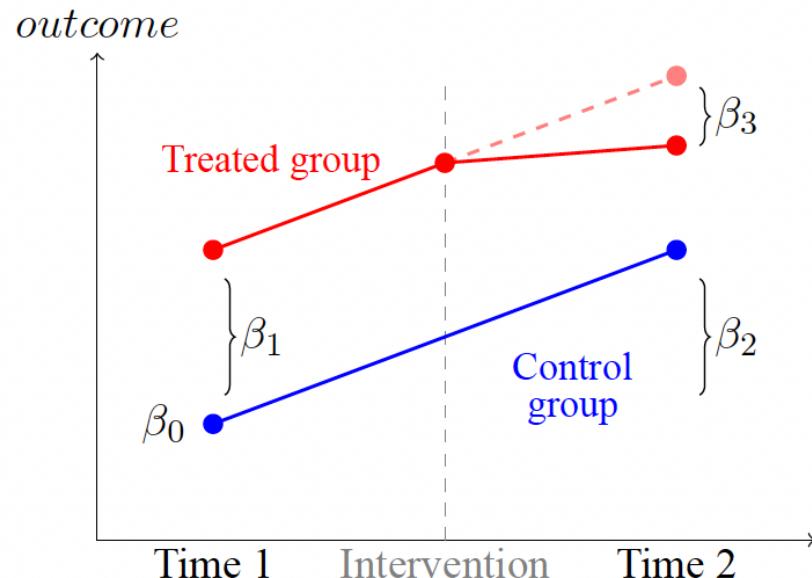
- y_{gt} is the outcome for unit i at time t .
- γ_g are group-specific fixed effects.
- τ_t are fixed effects for each time period.
- p_{gt} is a time-varying treatment indicator.
- δ^{DD} is the difference-in-differences estimate.

Our plan

lar trends affecting all villages that are unrelated to the policy. We will use basic two-way fixed effects models to estimate the total effect of the CBHP policy. The mean outcome will be defined using a set of linear predictors:

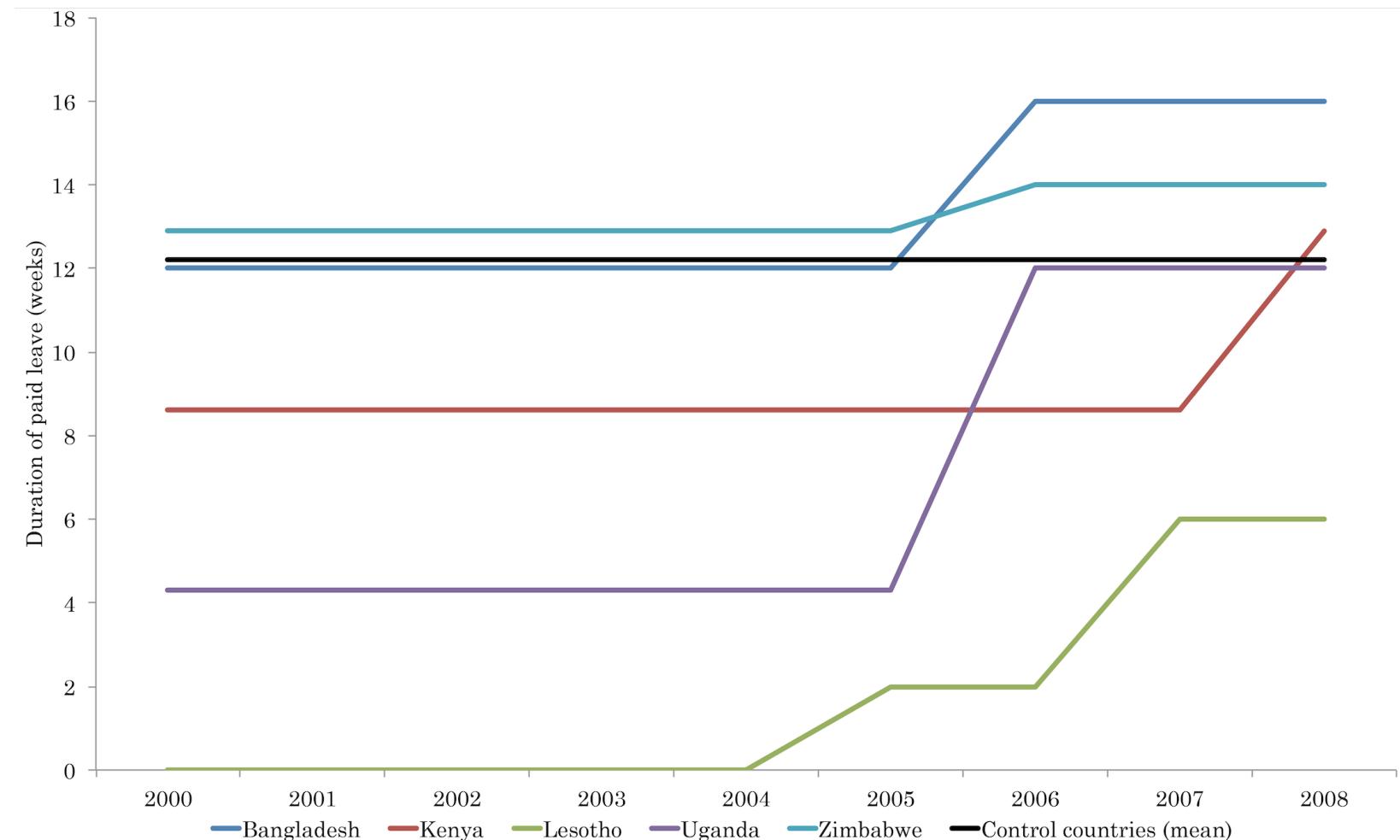
$$Y_{ijt} = g(\mu_{ijt}) = \beta_0 + \beta_1 V_j + \beta_2 P_t + \beta_3 T_{jt} + \delta \mathbf{Z}_{jt} + \eta \mathbf{W}_{ijt} + u_j + \varepsilon_{ijt}$$

where Y_{ijt} is the outcome for individual i in village j at time t . For binary or count outcomes we will use an appropriate link function $g(\bullet)$, which is the identity link in the model above. In this specification, V_j is an indicator for treated villages, P_t is an indicator for the period after the intervention, and T_{jt} is an indicator equal to 1 for treated villages in the post-intervention period (equivalent to a product term between V_j and P_t). In this specification, β_0 is the mean outcome in untreated villages ($V_j = 0$) in the pre-period ($P_t = 0$), β_1 captures differences between treated and untreated areas in the pre-period ($V_j = 1, P_t = 0$) and β_2 captures any changes over time in the outcome in the untreated



What about staggered treatments?

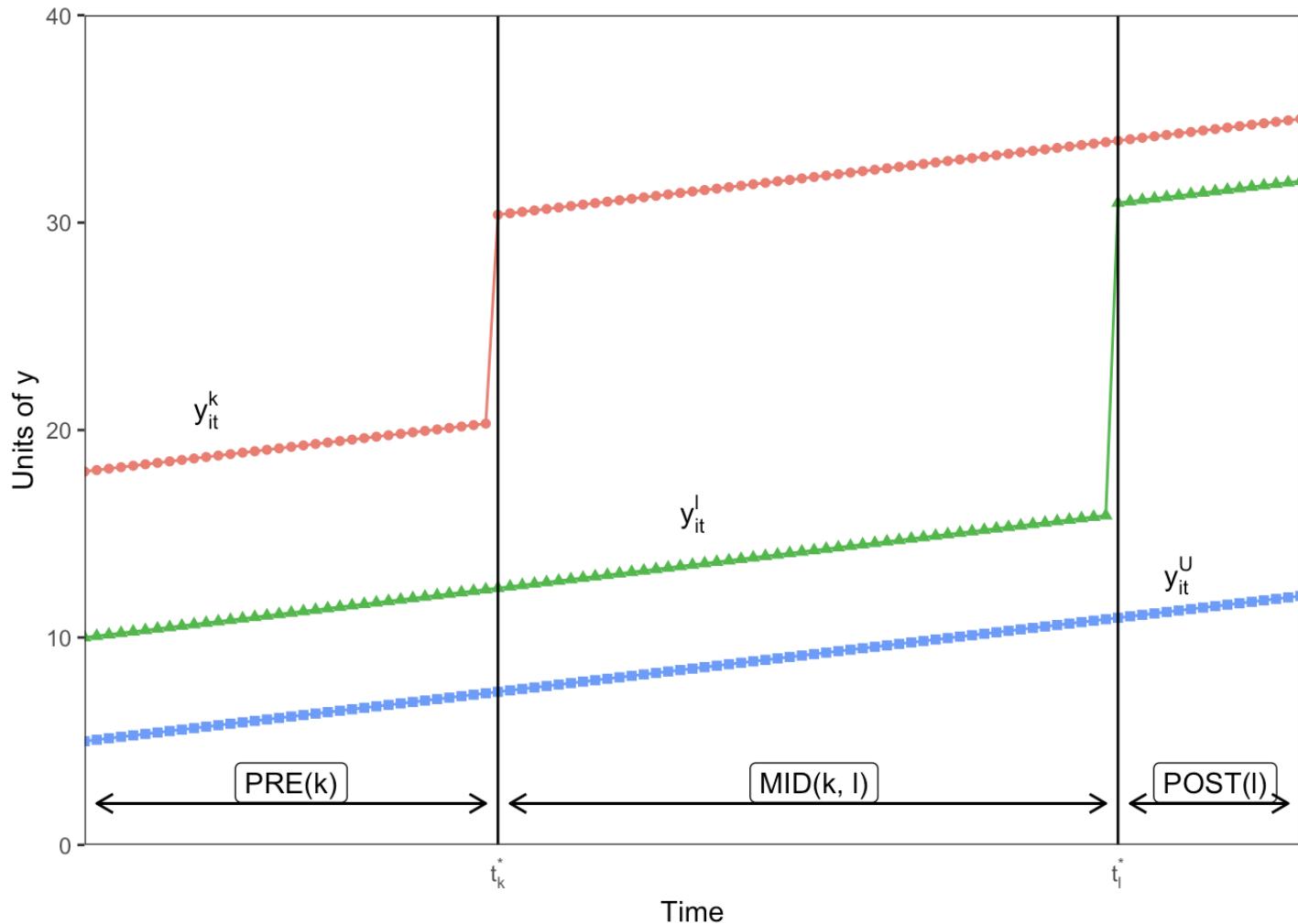
- Different groups adopt treatments at different times.
- Creates many 2x2 DDs.



Key points from Goodman-Bacon (2019)

- With OLS, DD with treatment timing is a variance-weighted average of many 2x2 ATTs.
- Weights are a function of both group sizes *and* variances.
- Can lead to δ^{DD} that is a poor summary of group-specific effects.

1. Early-adopters (k) vs. never treated (U)
2. Later-adopters (l) vs. never treated (U).
3. Early (k) vs. later (l) adopters.
- 4. Later (l) vs. earlier (k) adopters.**



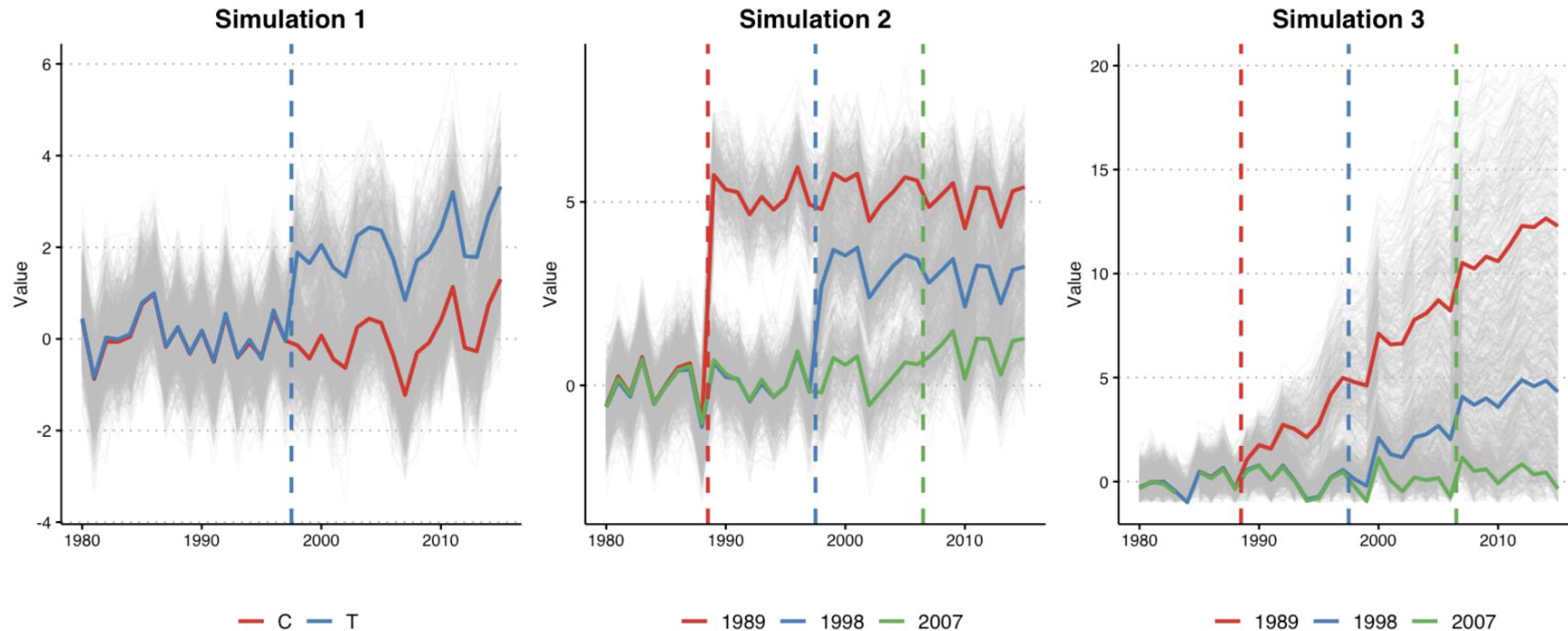
Graph from <https://andrewcbaker.netlify.app/2019/09/25/difference-in-differences-methodology/>

What is the problem?

- Using earlier treated groups as controls only 'works' if the treatment effects are:
 - Homogeneous across groups at a given time; and
 - Homogeneous over time (no dynamic effects).
- This adds any changes in treatment effects in the early group, which get **subtracted from the DD estimate**.
- Can lead to β^{DD} that is a poor summary of group-specific effects if there is heterogeneity.

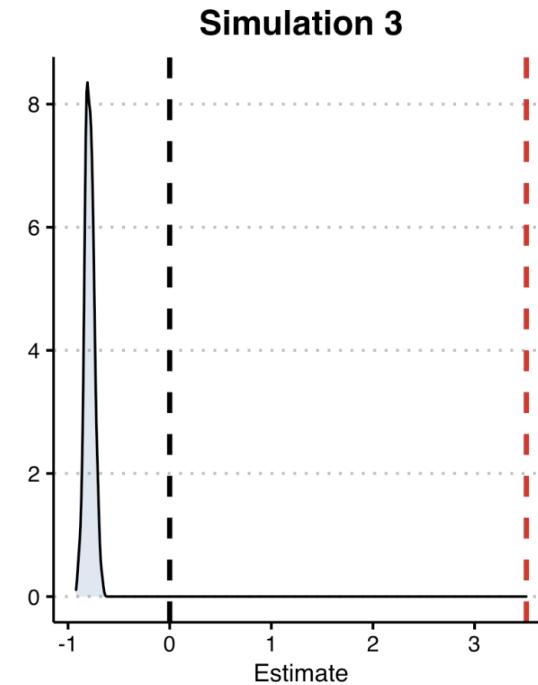
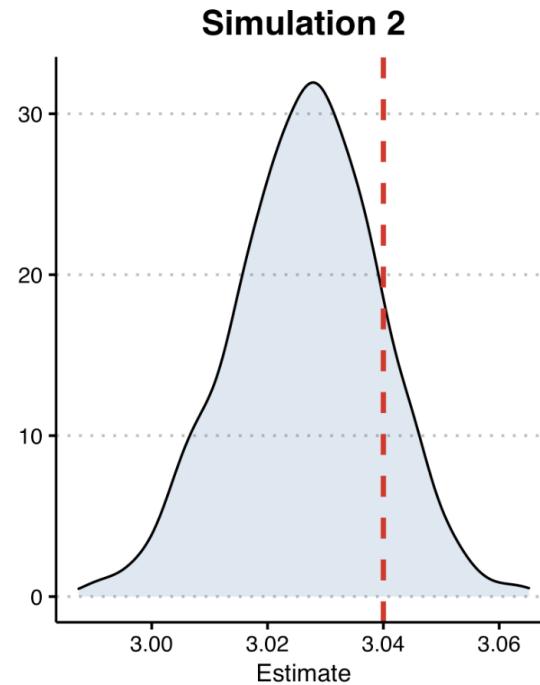
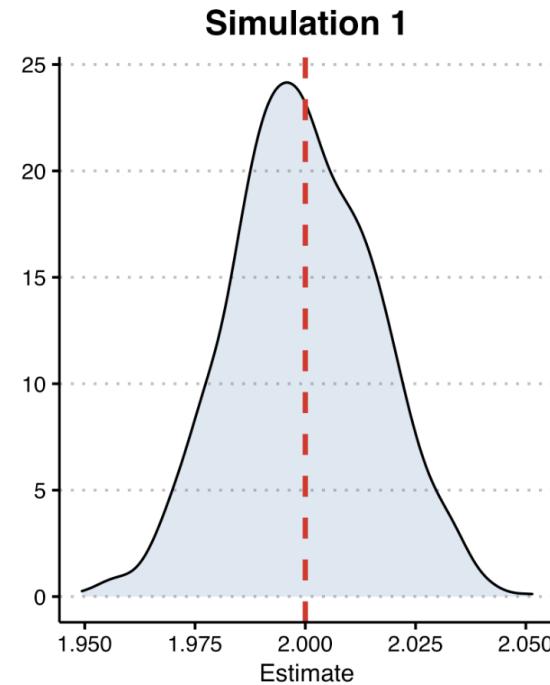
Simulated evidence from Baker et al (2022)

- 1 treated group at 1 time, constant effect.
- Multiple groups, staggered timing, constant effects.
- Multiple groups, staggered timing, heterogeneous effects.



Impact on estimates

1. No bias.
2. Little bias.
3. Wrong sign and magnitude!



See [Baker, et al. \(2022\)](#) for complete description.

What are potential solutions?

- All basically involve **not allowing** early treated groups to serve as controls later.

Callaway and Sant'Anna (2021)

Use non-parametric group-time ATTs (+ covariates).

Abraham and Sun (2021)

Use saturated fixed effects to ensure that prior treated units are not used as controls

Cengiz, Dube, Lindner, and Zipperer (2019)

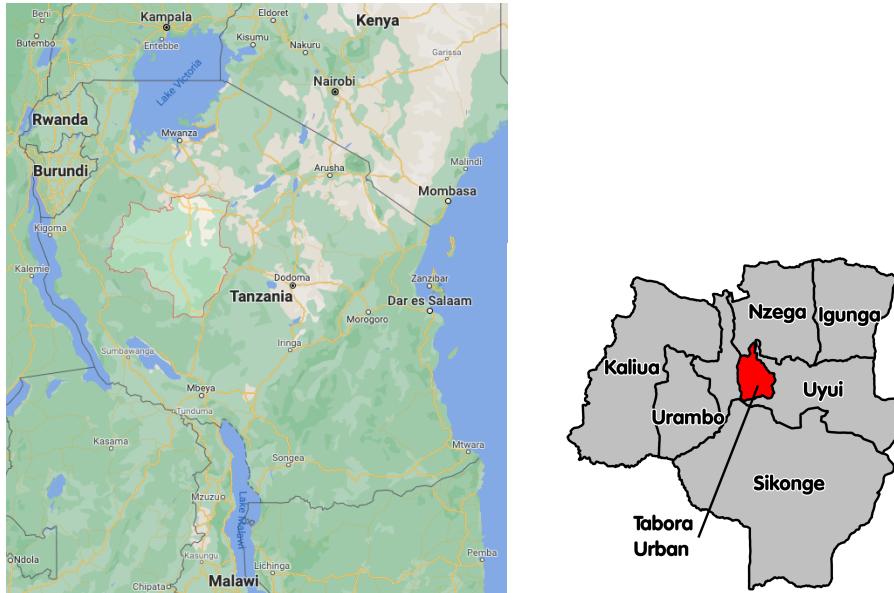
Create group-event-specific panel datasets and calculate event-specific estimates using separate regressions for each group-event.

Many new papers on this. See recent papers by [Callaway, et al. \(2021\)](#), [Goodman-Bacon \(2021\)](#), [Cengiz, et al. \(2019\)](#) and [Sun, et al. \(2021\)](#)

Example
(time permitting)

Tabora Maternal Newborn Health Initiative (TAMANI)

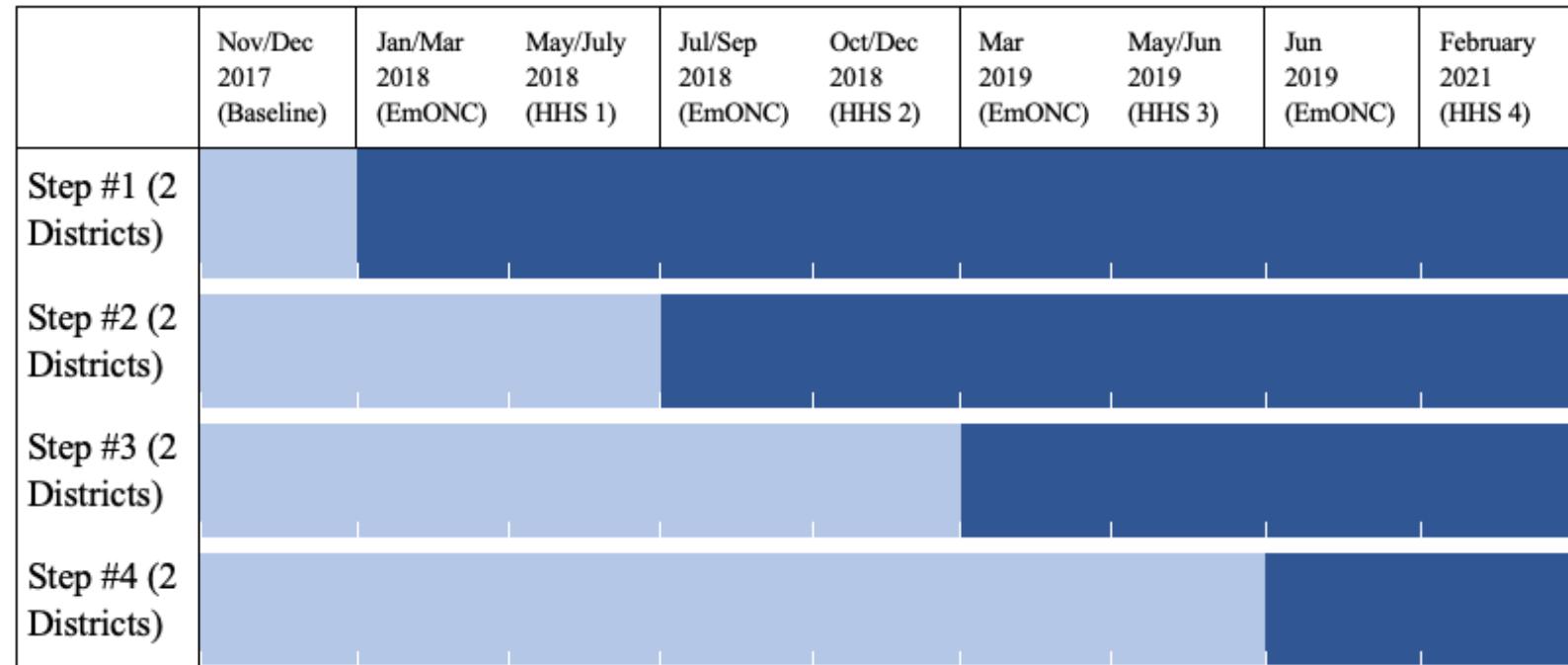
- Implemented by CARE Canada/Tanzania in 8 districts in Tabora region, Tanzania



Source: <https://commons.wikimedia.org/w/index.php?curid=47130439>, Google Maps

Initial Design: Stepped Wedge

- Political constraints led to breaking random timing.
- Switching of order of two districts.
- Analyze as DD?



Note: EmONC=Emergency Obstetric and Newborn Care;
HHS=Household Survey



Data structure: Individual-level

- sba_birth = SBA present
- txdel = treated
- time = survey wave
- pid = person ID
- group = time when group first treated

district	sba_birth	txdel	time	dist_id	p_id	group
Kaliua DC	0	0	1	1	237	2
Kaliua DC	1	1	5	1	2168	2
Kaliua DC	0	1	5	1	6723	2
Kaliua DC	0	0	1	1	2269	2
Kaliua DC	1	1	5	1	5084	2
Kaliua DC	1	1	5	1	2058	2

Data structure: Pooled by district

- tsba = total SBA births
- tpop = total pop (births)
- txdel = treated
- time = survey wave
- group = time when group first treated

district	dist_id	time	tsba	tpop	psba	txdel	group
Kaliua DC	1	1	409	705	0.580	0	2
	1	2	49	73	0.671	1	2
	1	3	63	77	0.818	1	2
	1	4	9	10	0.900	1	2
Nzega DC	1	5	94	114	0.825	1	2
	2	1	465	634	0.733	0	3
	2	2	60	76	0.789	0	3
	2	3	43	57	0.754	1	3
Nzega DC	2	4	9	10	0.900	1	3

TWFE models (OLS)

- Individual and aggregate basically identical.
- Clustered SEs approximately the same.
- Intervention increased the Pr(SBA) by 6 pp (95% CI -1.3 to 13.3).

$$y_{sba} = \alpha + \beta * txdel + \gamma_{district} + \delta_{time} + \epsilon$$

	Individual	Aggregate
txdel	0.060 (0.037) [-0.029, 0.148]	0.060 (0.040) [-0.035, 0.154]
Num.Obs.	5555	40
Std.Errors	by: district	by: district
FE: district	X	X
FE: time	X	X

Callaway-Sant'Anna Approach: Group-Time cohorts

- The CS approach starts with defining Group-Time cohorts.
- Groups defined by when they were *first* treated.

Group	P(SBA)					Total Pop				
	Time1	Time2	Time3	Time4	Time5	Time1	Time2	Time3	Time4	Time5
2	0.631	0.711	0.814	0.846	0.826	898	90	102	13	144
3	0.678	0.770	0.664	0.800	0.784	1455	165	125	30	232
4	0.778	0.807	0.673	0.750	0.868	1128	140	101	16	167
5	0.854	0.939	0.854	0.800	0.855	556	66	41	10	76

- Different aggregation schemes for ATTs are possible.
- Can allow for covariates via regressions adjustments, IPW and DR.

Callaway-Sant'Anna implementation

- Includes options for different structure, SE calculation, weights, etc.

```
# Use not-yet-treated as comparison group
did::att_gt(yname = "sba_birth", # name of the LHS variable
             tname = "time", # name of the time variable
             idname = "p_id", # name of the id variable
             gname = "group", # name of the first treatment period
             data = d_ind, # dataset
             xformula = NULL, # conditional parallel-trends
             weightsname = NULL, # can add weights
             est_method = "reg", # estimation method
             control_group = "notyettreated", # set the control group
             bstrap = TRUE, # compute bootstrapped SE
             biters = 1000, # bootstrap iterations
             print_details = FALSE, # if TRUE, print detailed results
             panel = FALSE, # panel or repeated cross-sectional
             clustervars = NULL) # cluster ID
```

See <https://bcallaway11.github.io/did/articles/multi-period-did.html> for R, <https://econpapers.repec.org/software/bocbocode/S458976.htm> for Stata

Estimates from CS approach

- Note there is no *overall* estimate.
- Each treatment group has an ATT at each time period.
- Can be combined to produce different aggregate ATTs.

term	group	time	estimate	std.error	conf.low	conf.high
ATT(2,2)	2	2	0.011	0.056	-0.098	0.120
ATT(2,3)	2	3	0.260	0.055	0.153	0.367
ATT(2,4)	2	4	0.269	0.169	-0.063	0.601
ATT(3,2)	3	2	0.045	0.044	-0.041	0.132
ATT(3,3)	3	3	0.018	0.062	-0.102	0.139
ATT(3,4)	3	4	0.170	0.152	-0.128	0.467
ATT(4,2)	4	2	-0.062	0.044	-0.149	0.024
ATT(4,3)	4	3	-0.048	0.082	-0.209	0.113
ATT(4,4)	4	4	0.130	0.197	-0.256	0.516

Note:

P-value for pre-test of parallel trends assumption: 0.31

Re-creating the Group-Time ATTs

- $\text{ATT}(2,2)$ means estimating ATT *at time 2* for the group *first treated at time 2*
- For $\text{ATT}(2,2)$ we are comparing $\text{Pr}(\text{SBA})$ between:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g22	time1	time2	Long diff	ATT_2_2
0	0.745	0.814	0.069	NA
1	0.631	0.711	0.080	0.011

This estimate says that intervention increased the probability of an SBA birth by 0.01 for Group 2 at Time 2.

- For ATT(2,3) the groups being compared are:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- ATT(2,3) means estimating the ATT *at time 3* for the group *first treated at time 2*
- Need to exclude any group treated at time 3 to avoid bias.
- The group first treated at time 3 is excluded.
- Provides an estimate of the lagged impact of intervention.

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g23	time1	time3	Long diff	ATT_2_3
0	0.803	0.725	-0.078	NA
1	0.631	0.814	0.182	0.26

Etc., etc., etc...

- For ATT(4,2) the groups being compared are:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- ATT(4,2) means estimating the ATT *at time 2* for the group *first treated at time 4*
- Need to exclude any group treated at time 2 to avoid bias.
- The group first treated at time 2 is excluded.
- Provides an estimate of the lead effects or non-parallel trends.

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g42	time1	time2	Long diff	ATT_4_2
0	0.727	0.818	0.091	NA
1	0.778	0.807	0.029	-0.062

Our 'hand-calculated' ATT(4,2) of -0.062 is the same as the regression-based CS estimate:

term	group	time	estimate	std.error	conf.low	conf.high
ATT(2,2)	2	2	0.011	0.057	-0.100	0.123
ATT(2,3)	2	3	0.260	0.057	0.148	0.373
ATT(2,4)	2	4	0.269	0.181	-0.085	0.623
ATT(3,2)	3	2	0.045	0.044	-0.042	0.132
ATT(3,3)	3	3	0.018	0.070	-0.118	0.155
ATT(3,4)	3	4	0.170	0.159	-0.143	0.482
ATT(4,2)	4	2	-0.062	0.043	-0.147	0.022
ATT(4,3)	4	3	-0.048	0.090	-0.224	0.128
ATT(4,4)	4	4	0.130	0.197	-0.255	0.516

"Final" ATT(4,4) compares only 2 groups (4 clusters):

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g44	time3	time4	Long diff	ATT_4_4
0	0.854	0.80	-0.054	NA
1	0.673	0.75	0.077	0.13

Dynamic Effects

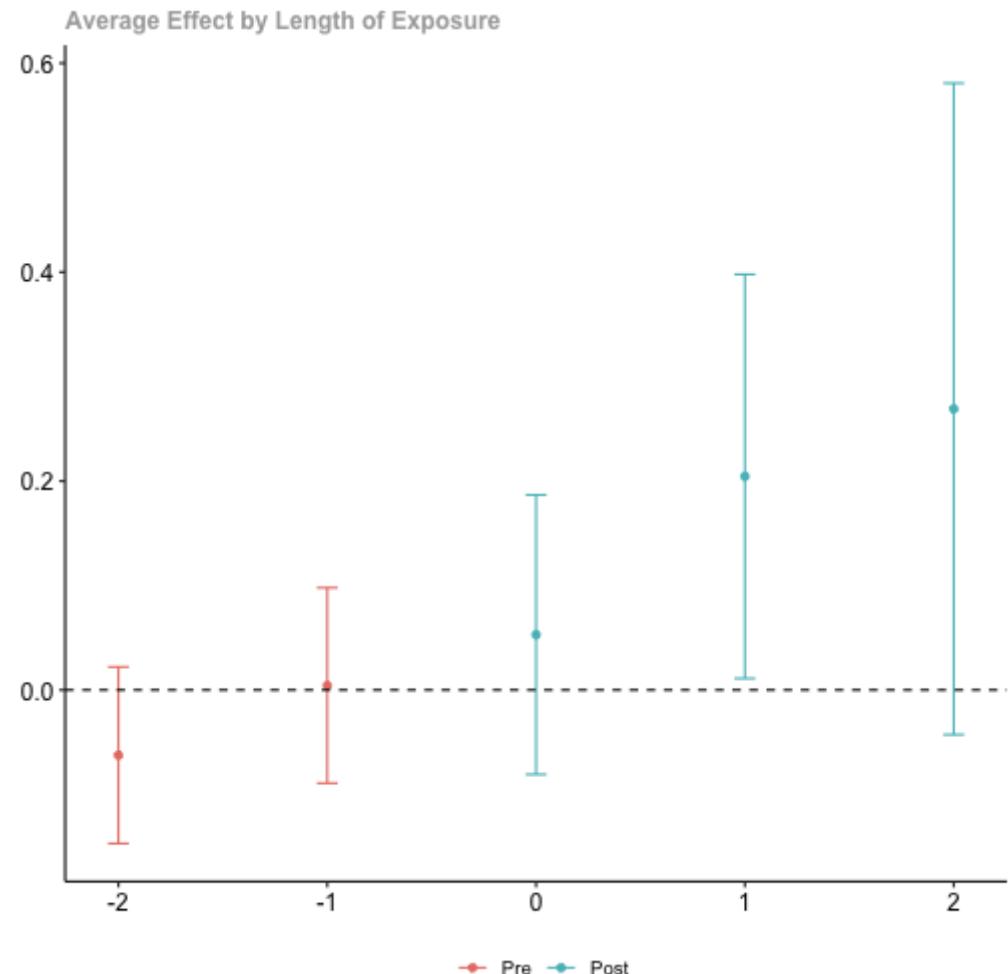
- Estimates by length of exposure
- Overall summary:

$$\overline{ATT} = 0.176 (-0.02, 0.37)$$

event.time	estimate	std.error	conf.low	conf.high
-2	-0.062	0.043	-0.147	0.022
-1	0.004	0.048	-0.089	0.098
0	0.053	0.068	-0.081	0.187
1	0.204	0.099	0.011	0.398
2	0.269	0.159	-0.043	0.581

- TWFE estimate was

$$\overline{ATT} = 0.06 (-0.013, 0.133)$$



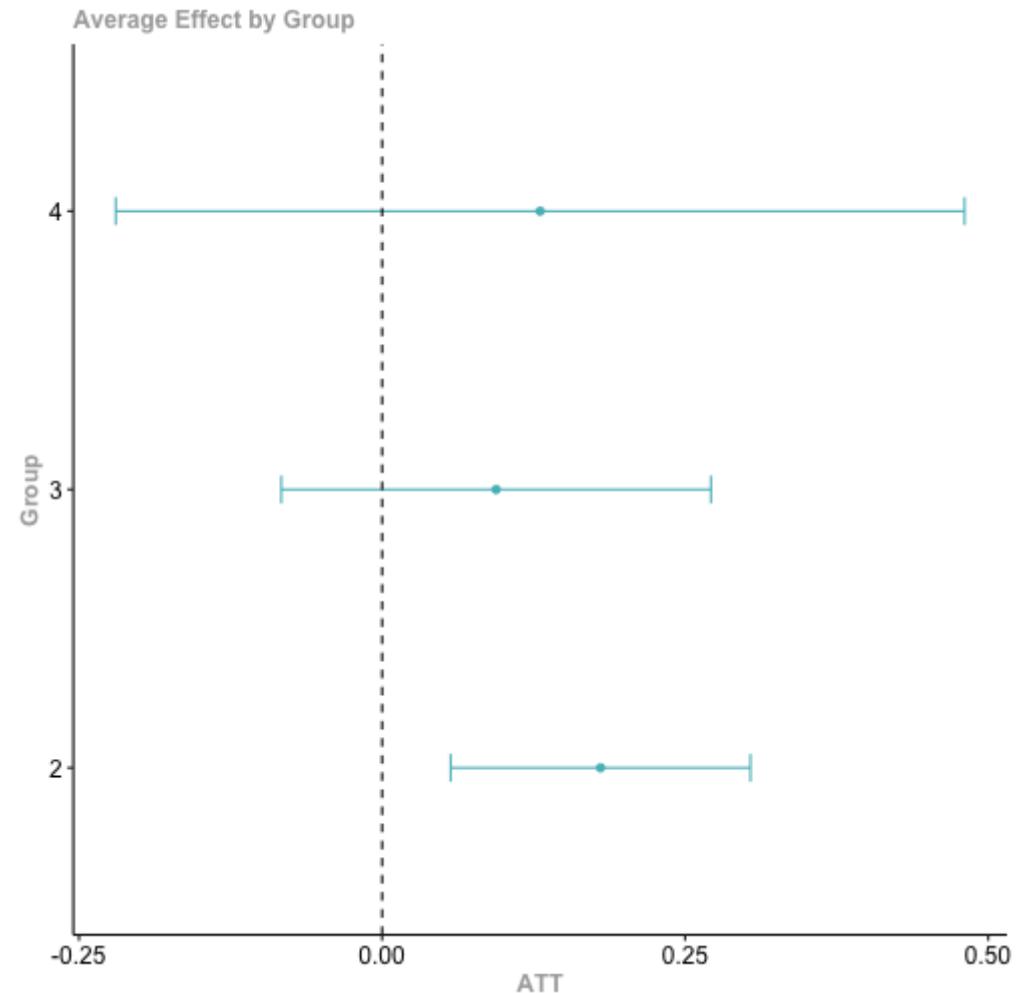
Group-specific ATTs

- Group-specific estimates
- Weighted avg of each group-time ATT for group g

group	estimate	std.error	conf.low	conf.high
Average	0.128	0.096	-0.059	0.316
2	0.180	0.063	0.057	0.304
3	0.094	0.090	-0.083	0.271
4	0.130	0.179	-0.220	0.481

- TWFE estimate was

$$\overline{ATT} = 0.06 (-0.013, 0.133)$$



Calendar Time ATTs

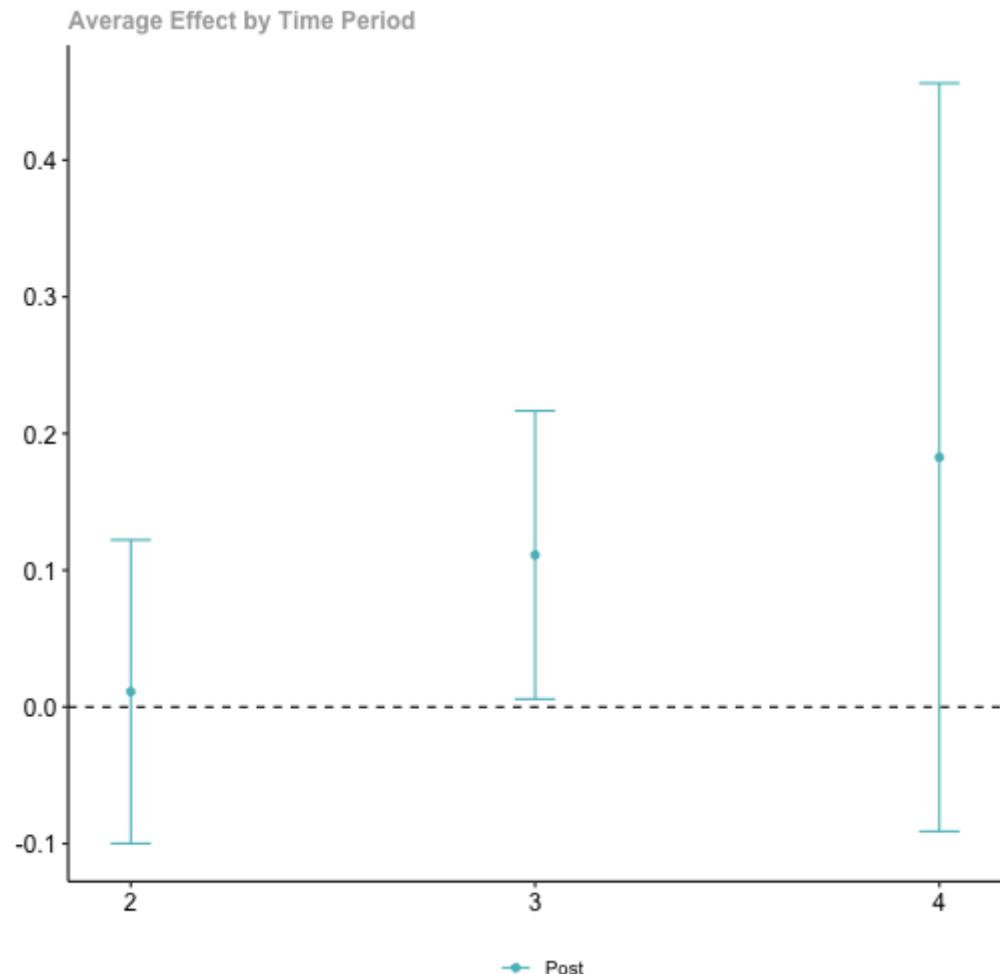
- ATT in time period t for groups that have participated in the treatment by time period t.
- Overall summary:

$$\overline{ATT} = 0.102 (-0.02, 0.22)$$

time	estimate	std.error	conf.low	conf.high
2	0.011	0.057	-0.100	0.122
3	0.111	0.054	0.006	0.217
4	0.183	0.140	-0.091	0.456

- TWFE estimate was

$$\overline{ATT} = 0.06 (-0.013, 0.133)$$



Summary and Implications



- BHET has multiple groups and multiple times, but only 2 times and little room for dynamic effects.
- TWFE may not lead to serious bias, but still conceptually problematic.
- We don't want to include villages treated in 2019 as controls for villages treated in 2020 or 2021.
- Should use alternative methods.
- More to discuss re: standard errors, mediation analysis.

References

- Baker, A. C. et al. (2022). "How much should we trust staggered difference-in-differences estimates?" In: *Journal of Financial Economics* 144.2, pp. 370-395.
- Callaway, B. et al. (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200-230.
- Cengiz, D. et al. (2019). "The effect of minimum wages on low-wage jobs". In: *The Quarterly Journal of Economics* 134.3, pp. 1405-1454.
- Goodman-Bacon, A. (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254-277.
- Rambachan, A. et al. (2019). "An honest approach to parallel trends". In: *Unpublished manuscript, Harvard University*.
- Roth, J. et al. (2020). "When is parallel trends sensitive to functional form?" In: *arXiv preprint arXiv:2010.04814*.
- Sun, L. et al. (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175-199.