Rescuing impact measurements

Differential treatment timing and duration in d-i-d estimation

2024-08-30

Outline of presentation

- Background
- Problem
- Solutions
- Case study - MISTI
- Final thoughts

Bottom line up front

- In certain settings, beware the Two-Way Fixed Effects Estimator!
- Don't conflate your modeling approach (TWFE) with your estimation strategy
- Examine the different groups created by differential timing
- Use event study designs
- Specify a fully flexible model (Two-way Mundlak)

- **Background**
- Problem
- Solutions
- Case study - MISTI
- Final thoughts

What is the canonical d-i-d setup?

where..

β_0 is the comparison group at baseline

δ_0 is the secular change from baseline to endline, unrelated to treatment

β_1 is the difference between the treatment and comparison groups at baseline, and

δ_1 is the treatment effect, the interaction of treatment and time

Algebraically, δ_0 can be expressed as the difference between the pre/post difference in each of the treatment and comparison groups

$$\delta_1 = (\bar{y}_{POST,TREAT} - \bar{y}_{PRE,TREAT}) - (\bar{y}_{POST,COMPARISON} - \bar{y}_{PRE,COMPARISON})$$

hence, difference-in-differences (d-i-d or DiD or DD)

Canonical d-i-d, 2x2

$$y_{it} = \beta_0 + \delta_{0,t} Post_t + \beta_{1,i} Treat_i + \delta_{1,it} Post_t * Treat_i + \epsilon_{it}$$

Canonical d-i-d 2x2 setup

	Pre	**Post**	**Post - Pre**
Comparison	β_0	$\beta_0 + \delta_0$	δ_0
Treatment	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treatment - Comparison	β_1	$\beta_1 + \delta_1$	δ_1

How does the canonical d-i-d generalize to multiple time periods and/or groups?

When we generalize the two-period setup to multiple time periods and/or groups, we have the two-way fixed effect (TWFE) estimator

$$y_{it} = \alpha_i + \alpha_t + \beta^{DD} it + \epsilon_{it}$$

where..

α_i are group fixed effects

α_t are time fixed effects

B_{it}^{DD} indicates whether group *i* in period *t* is treated

TWFE is a workhorse in program evaluation

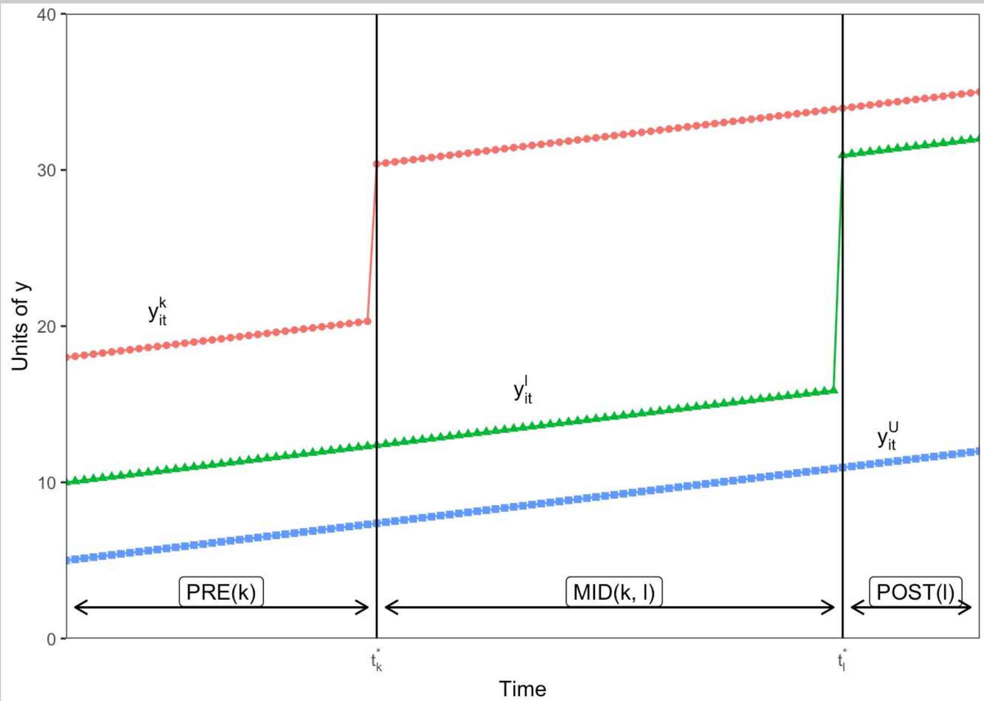
- 744 d-i-d studies across ten journals in finance and accounting, 2000-2019 [Baker 2022]
- 19 percent of all empirical articles published by the American Economic Review (AER) between 2010 and 2012 used TWFE [de Chaisemartin and D'Haultfoeuille 2020]

- Background
- **Problem**
- Solutions
- Case study - MISTI
- Final thoughts

But what is β_{it}^{DD} actually telling us?

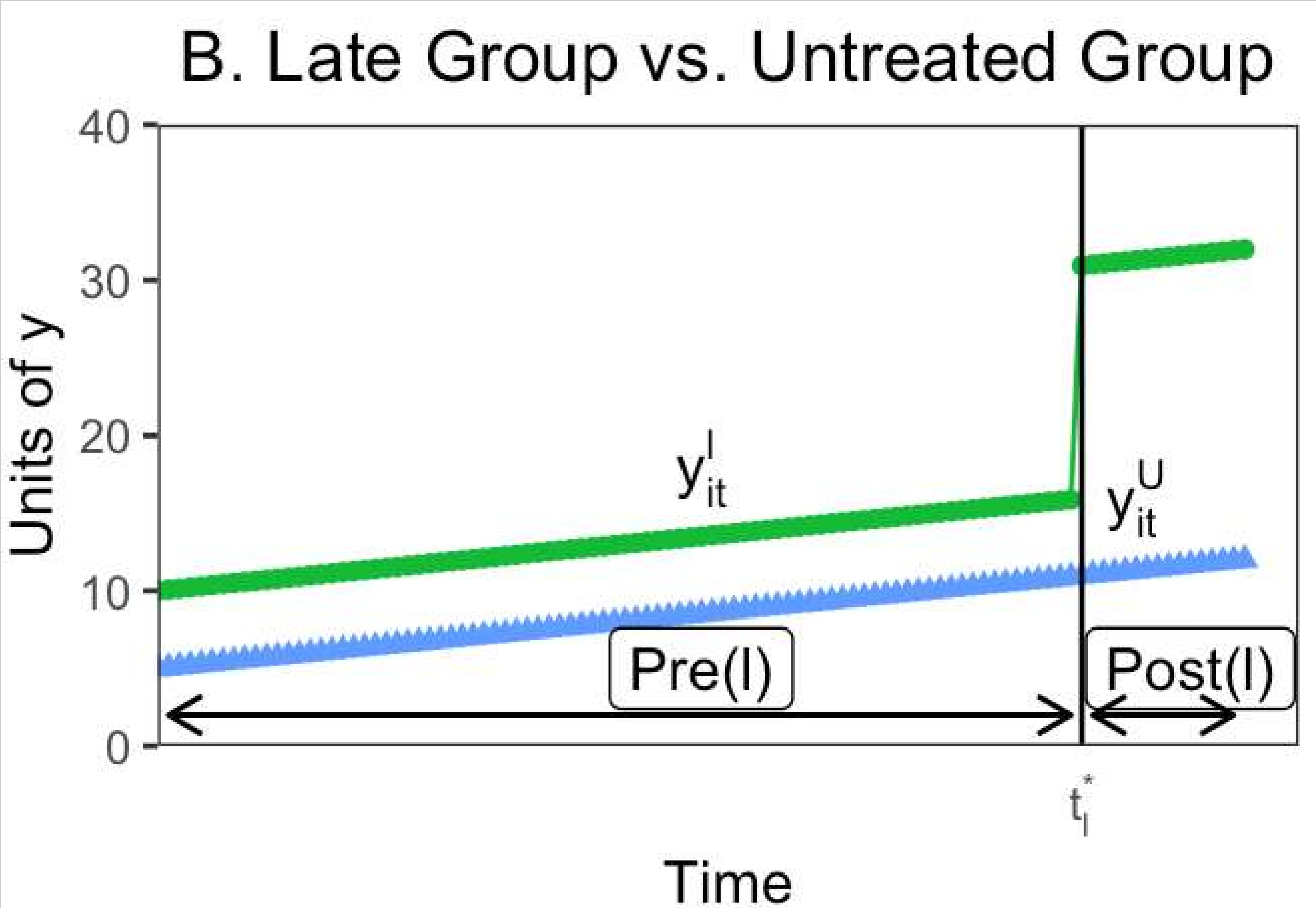
- For the canonical 2x2, we know exactly what we are estimating
- For *i* groups and *t* time periods, we are getting some average of multiple 2x2s
- But how does this work, exactly?
- Goodman-Bacon (2021) decided to work it out

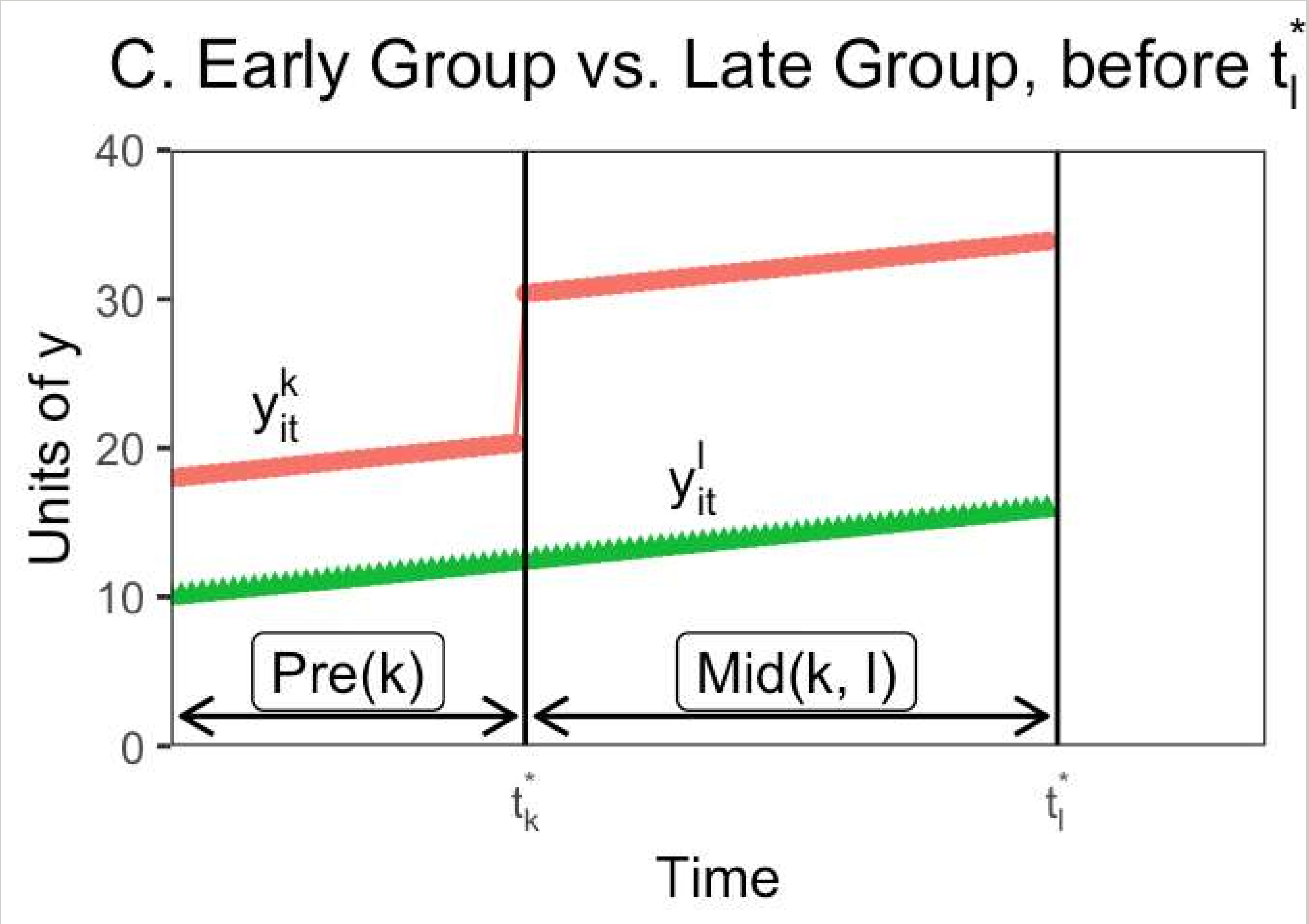
Let's take a single step from two time periods to three, where treatment can be adopted at either t = 2 or t = 3

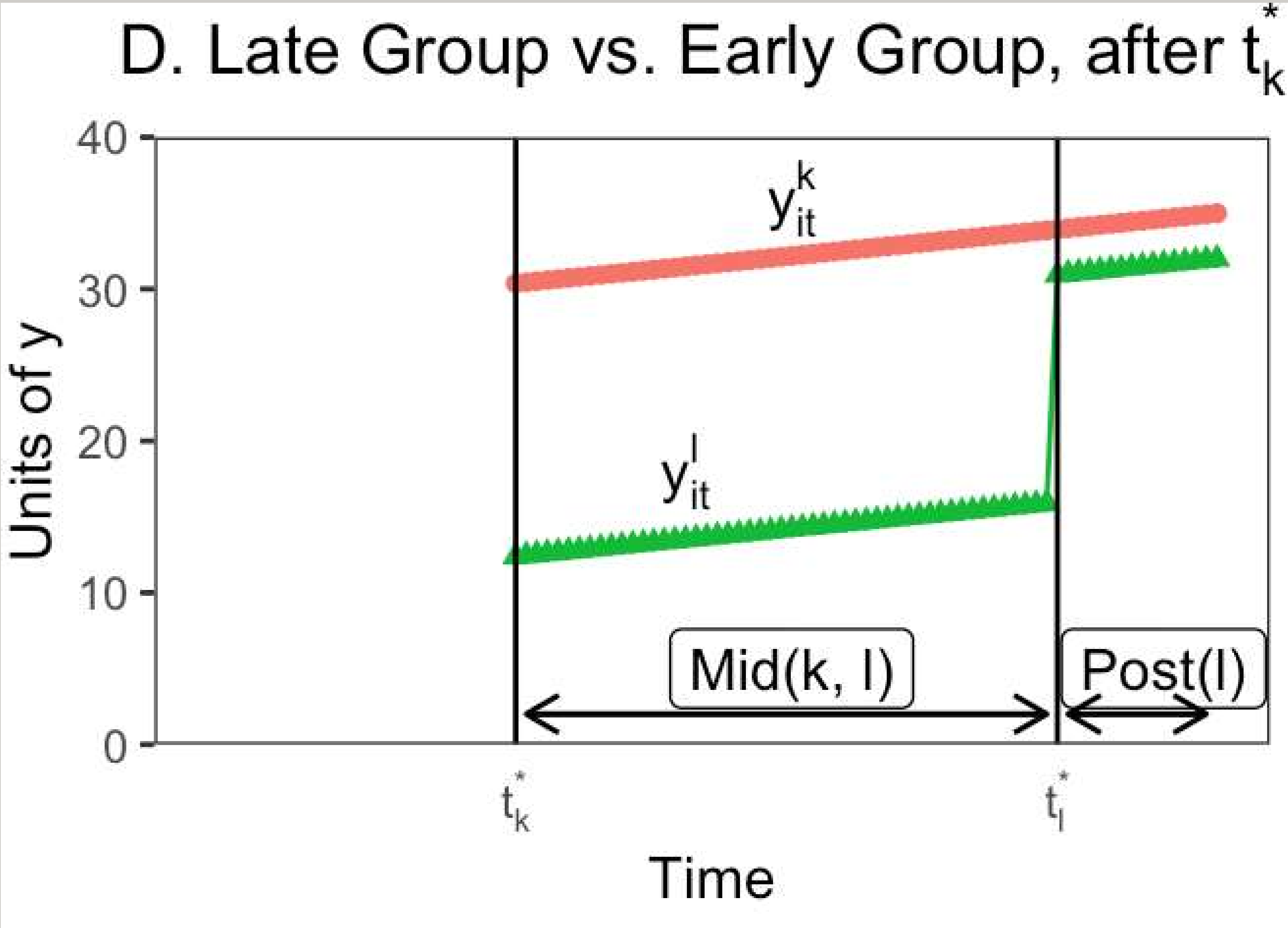


Any design with multiple treatment timings will have k^2 groups, where *k* is the number of timings.

```
Error in include_graphics("Baker panel a.png"): Cannot find the file(s):
"Baker panel a.png"
```


B. Late Group vs. Untreated Group
Units of y
40
30
20
10
0
y_{it}^{l}
y_{it}^{U}
Pre(l)
Post(l)
t_l^*
Time


C. Early Group vs. Late Group, before t_l^*
40
30
20
10
0
Units of y
y_{it}^k
y_{it}^l
Pre(k)
Mid(k, l)
t_k^*
t_l^*
Time


D. Late Group vs. Early Group, after t_k^*
40
30
20
10
0
Units of y
y_{it}^k
y_{it}^l
Mid(k, l)
Post(l)
t_k^*
t_l^*
Time

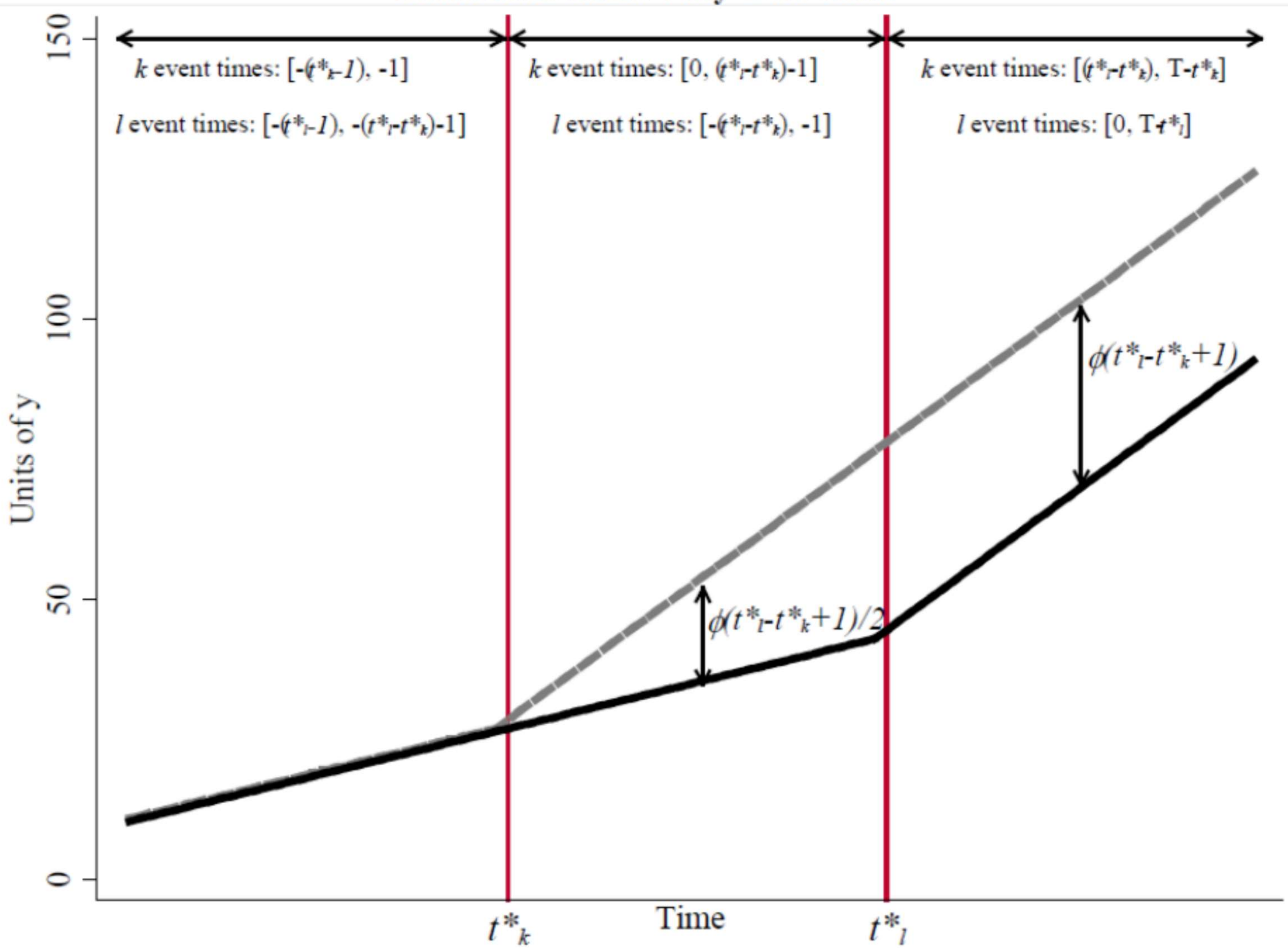
Where does this leave us?

- TWFE treats some data that is under treatment status as comparison!
- Not an issue under constant treatment effect
  - stable unit treatment value (SUTVA)
  - no variation in treatment effect for any reason

But TWFE fails under following conditions:

- different treatment groups have different treatment effects
- treatment effects are dynamic over post-treatment periods
- heterogeneous treatment effects across sub-groups within a treated group

Figure 3. Difference-in-Differences Estimates with Variation in Timing Are Biased When Treatment Effects Vary Over Time



An example of failure

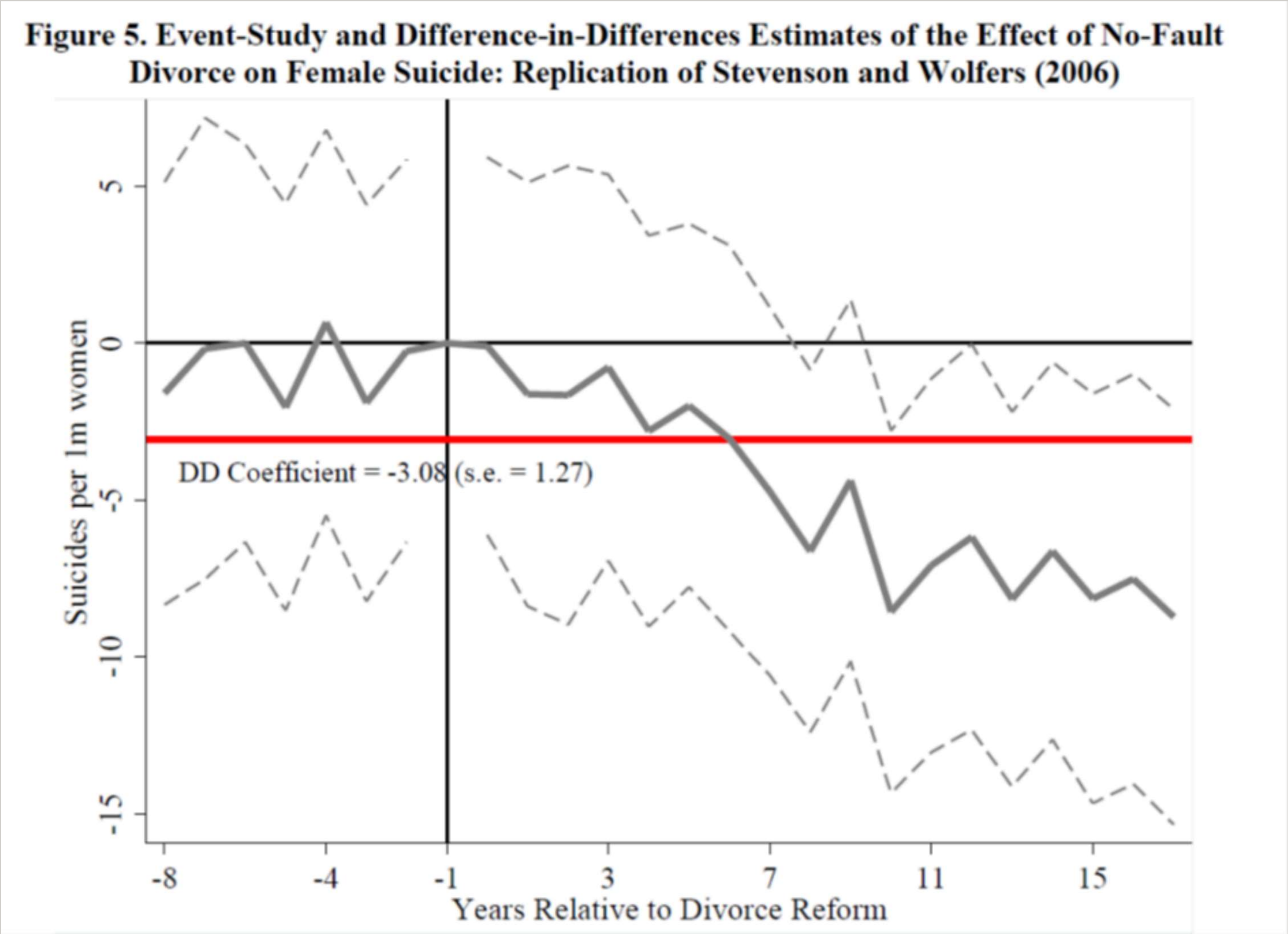
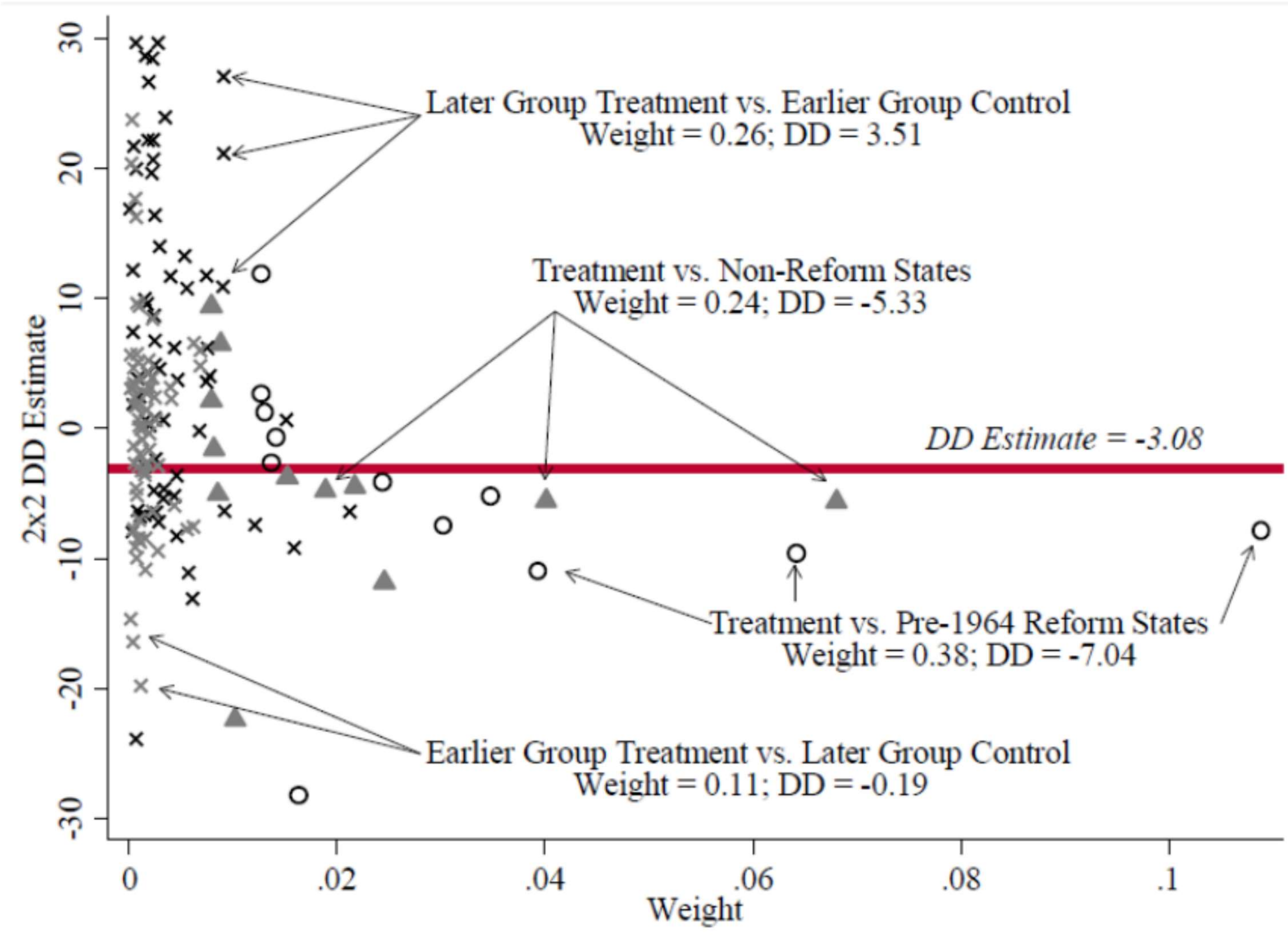


Figure 6. Difference-in-Differences Decomposition for Unilateral Divorce and Female Suicide



- Background
- Problem
- **Solutions**
- Case study - MISTI
- Final thoughts

Diagnostic: the Bacon decomposition

- The Bacon decomposition will take a TWFE model and decompose it into the full array of 2x2 d-i-d cells used to construct the overall estimate
- The decomposition will also calculate the variance-weights used in regression to see which 2x2 cells are powering the overall estimate

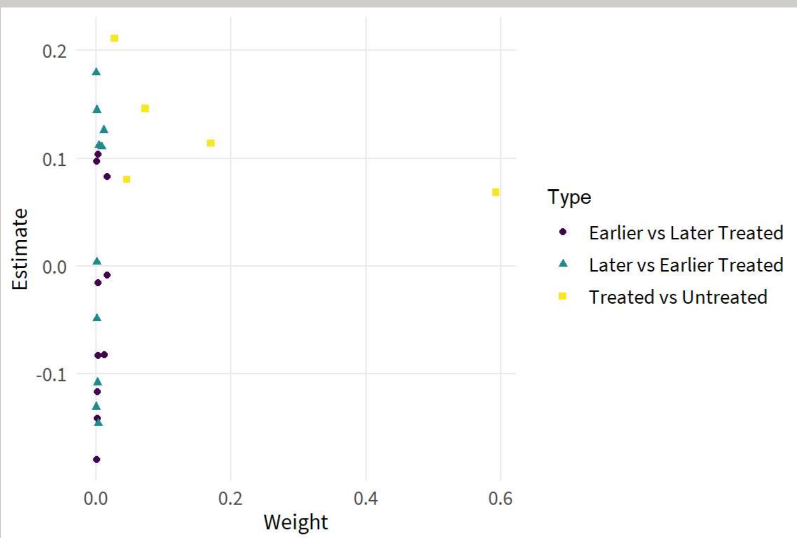
```r
castle_bacon <- bacon(l_homicide ~ post,
                      data = bacondecomp::castle,
                      id_var = "state",
                      time_var = "year") %>%
  arrange(treated, untreated)
```

```
                      type weight  avg_est
1 Earlier vs Later Treated 0.0598 -0.00554
2 Later vs Earlier Treated 0.0319  0.07032
3     Treated vs Untreated 0.9083  0.08796
```

treated	untreated	estimate	weight	type
2005	2006	-0.08313	0.003405	Earlier vs Later Treated
2005	2007	-0.11672	0.002095	Earlier vs Later Treated
2005	2008	-0.14123	0.001571	Earlier vs Later Treated
2005	2009	0.09714	0.001048	Earlier vs Later Treated
2005	99999	0.08017	0.045569	Treated vs Untreated
2006	2005	-0.14607	0.003405	Later vs Earlier Treated
2006	2007	0.08302	0.016342	Earlier vs Later Treated
2006	2008	-0.00848	0.016342	Earlier vs Later Treated
2006	2009	-0.08226	0.012256	Earlier vs Later Treated
2006	99999	0.06824	0.592395	Treated vs Untreated
2007	2005	-0.10806	0.001676	Later vs Earlier Treated
2007	2006	0.12596	0.010895	Later vs Earlier Treated
2007	2008	0.10372	0.002933	Earlier vs Later Treated
2007	2009	-0.01598	0.002933	Earlier vs Later Treated
2007	99999	0.11406	0.170124	Treated vs Untreated
2008	2005	-0.04898	0.000943	Later vs Earlier Treated
2008	2006	0.11069	0.008171	Later vs Earlier Treated
2008	2007	0.14479	0.001257	Later vs Earlier Treated

treated	untreated	estimate	weight	type
2008	2009	-0.17989	0.000838	Earlier vs Later Treated
2008	99999	0.14605	0.072910	Treated vs Untreated
2009	2005	0.17952	0.000419	Later vs Earlier Treated
2009	2006	0.11210	0.004085	Later vs Earlier Treated
2009	2007	0.00373	0.000838	Later vs Earlier Treated
2009	2008	-0.13078	0.000210	Later vs Earlier Treated
2009	99999	0.21108	0.027341	Treated vs Untreated

```r
ggplot(castle_bacon) +
  aes(x = weight, y = estimate, shape = factor(type)
  labs(x = "Weight", y = "Estimate", shape = "Type",
  geom_point() +
  scale_color_viridis_d()
```



Adjustment: new estimators

R packages for new d-i-d estimators

Reference	R package	Description
Callaway Sant'Anna (2020)	did	Compare treatment only to never treated, or never-treated + not-yet-treated. Also propensity score weights with covariates.
Sun Abraham (2020)	fixest	Fully saturate relative time indicators with treatment initiation; equivalent to Callaway Sant'Anna
Chaisemartin D'Haultfoeuille (2020)	DIDmultiplegt	Applies time-unit adjustments for a more general range of settings than just staggered adoption
Wooldridge (2021)	fixest	Dummies for all group, time, time-to-treat, time-since-treatment units

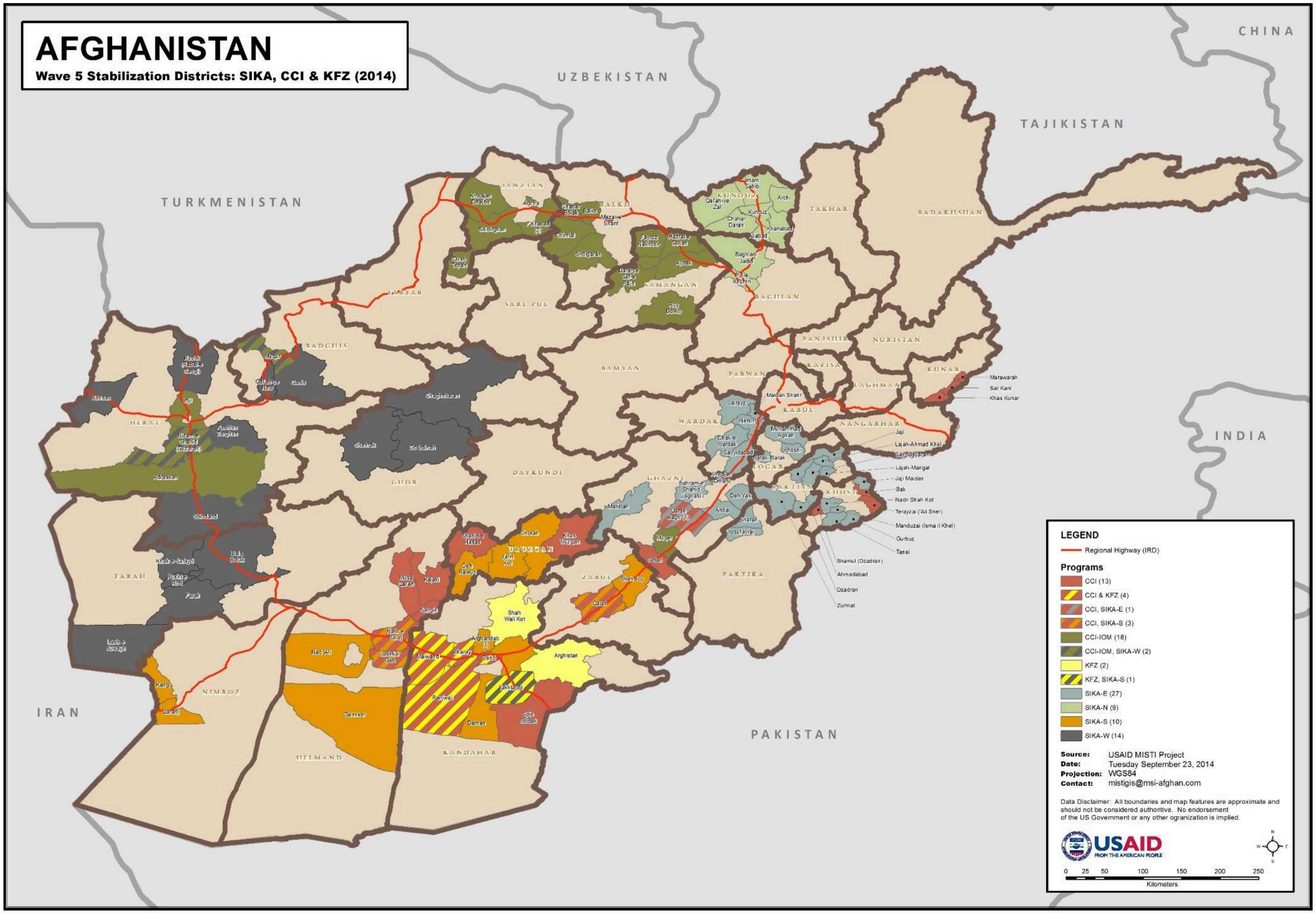
- Background
- Problem
- Solutions
- **Case study - MISTI**
- Final thoughts

Measuring Impact of Stabilization Initiatives (MISTI)

Can small scale, community-driven development activities build local government legitimacy in a kinetic conflict-affected environment?

MISTI

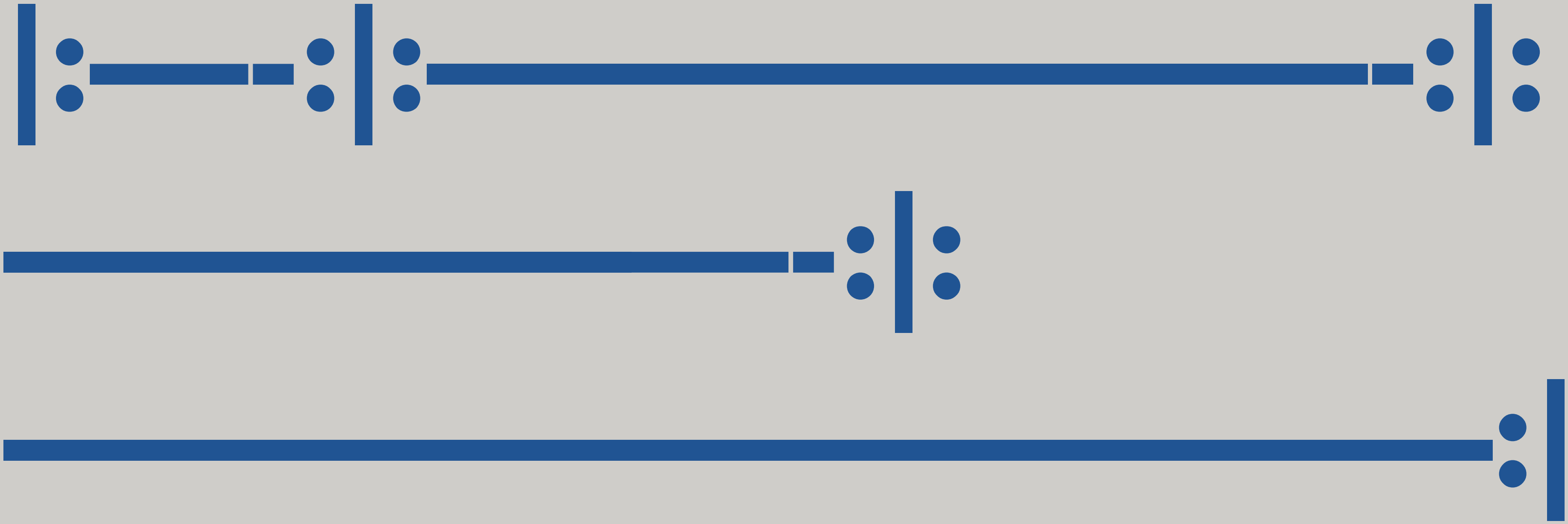
- Village panel survey in five waves, Sep 2012 - Nov 2014
- ~5,000 villages surveyed across 130 districts and 23 provinces
- ~ 30,000 household interviews per wave
- 860 treated villages at any wave (17%)
- 355 villages surveyed in all five waves
- 85 villages treated (24%)

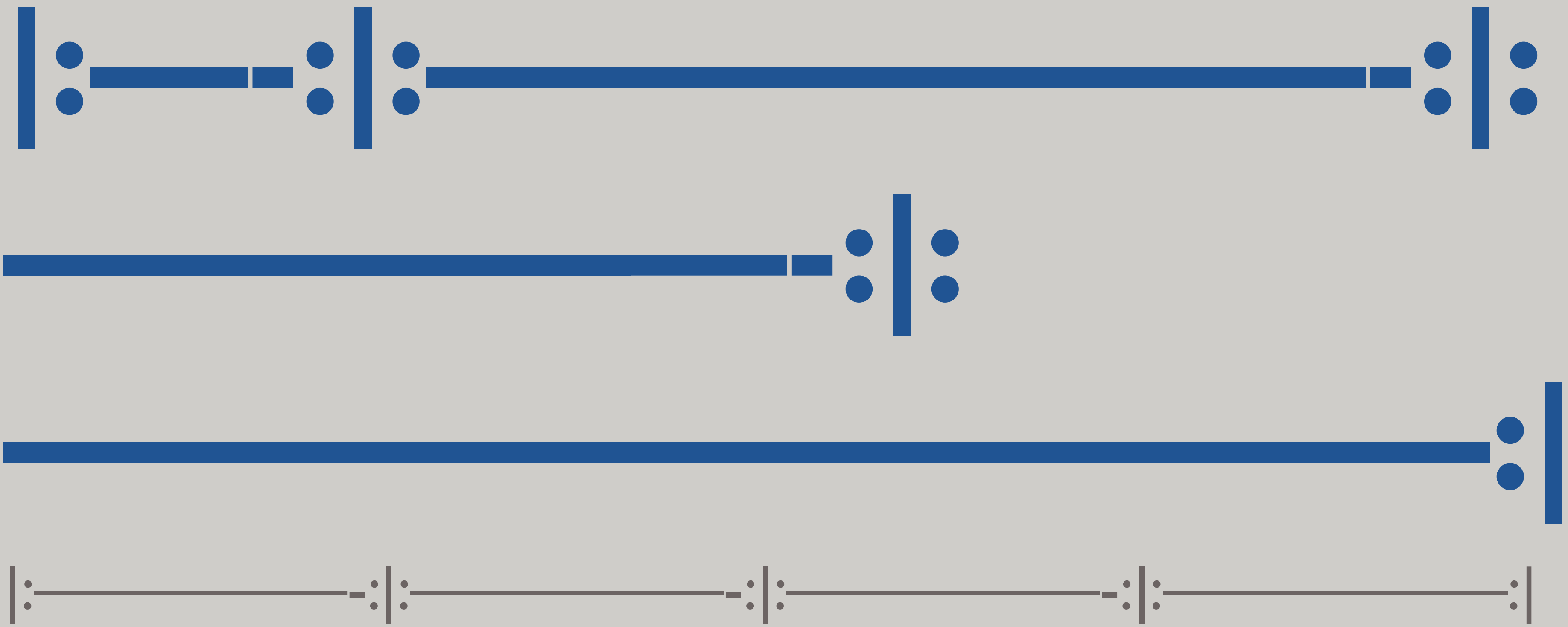


AFGHANISTAN
Wave 5 Stabilization Districts: SIKA, CCI & KFZ (2014)
UZBEKISTAN
TURKMENISTAN
TAJIKISTAN
CHINA
INDIA
IRAN
PAKISTAN
BADAKHSHAN
TAKHAR
KUNDUZ
BALKH
JAWZJAN
SAMANGAN
SARI PUL
BAGHLAN
FARYAB
BADGHIS
HERAT
GHOR
BAMYAN
PARWAN
PANJSHIR
NURISTAN
KAPISA
KUNAR
LAGHMAN
KABUL
NANGARHAR
WARDAK
LOGAR
PAKTIYA
KHOST
GHAZNI
PAKTIKA
DAYKUNDI
URUZGAN
ZABUL
FARAH
NIMROZ
HELMAND
KANDAHAR
Marawarah
Sar Kani
Khas Kunar
LEGEND
Regional Highway (IRD)
Programs
CCI (13)
CCI & KFZ (4)
CCI, SIKA-E (1)
CCI, SIKA-S (3)
CCI-IOM (18)
CCI-IOM, SIKA-W (2)
KFZ (2)
KFZ, SIKA-S (1)
SIKA-E (27)
SIKA-N (9)
SIKA-S (10)
SIKA-W (14)
Source: USAID MISTI Project
Date: Tuesday September 23, 2014
Projection: WGS84
Contact: mistigis@msi-afghan.com
Data Disclaimer: All boundaries and map features are approximate and should not be considered authoritive. No endorsement of the US Government or any other ogranization is implied.
USAID
FROM THE AMERICAN PEOPLE
0 25 50 100 150 200 250
Kilometers

MISTI treatment timings

Wave | Comparison villages | Treated villages | Treated villages (cumulative) |





1 | 355 | 0 | 0 |

2 | 341 | 14 | 14 |

3 | 322 | 19 | 33 |

4 | 302 | 20 | 53 |

5 | 270 | 32 | 85 |

Single-wave analysis

- Final reporting of MISTI relied on a series of single-wave estimations

```
Error: `path` does not exist: 'MISTI tables.xlsx'
Error in eval(expr, envir, enclos): object 'wav' not found
```

MISTI TWFE

```
twfe <- lm(stab_std ~ treat_event + as.factor(villag
              data=mistifull) %>%
  tidy()

twfe[1:2,]
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.0581	0.359	-0.162	0.871
treat_event	-0.0389	0.0947	-0.411	0.681

MISTI bacondecomp

```r
mistibacon <- bacon(stab_std ~ treat_event,
                    data=mistifull,
                    id_var="village",
                    time_var="wave") %>%
  arrange(treated, untreated)
```

```
                      type weight avg_est
1 Earlier vs Later Treated 0.0578  0.3911
2 Later vs Earlier Treated 0.0466 -0.3465
3     Treated vs Untreated 0.8956 -0.0506
```

```r
mistibacon_coef <- sum(mistibacon$estimate * mistiba
```

```
mistibacon_coef
```

```
[1] -0.0389
```

MISTI bacondecomp 2x2 cells

Plot of 2x2 cells

treated	untreated	estimate	weight	type
2	3	0.7473	0.00211	Earlier vs Later Treated
2	4	0.7093	0.00444	Earlier vs Later Treated
2	5	0.6868	0.01066	Earlier vs Later Treated
2	99999	0.8232	0.11998	Treated vs Untreated
3	2	-0.1216	0.00633	Later vs Earlier Treated
3	4	0.4973	0.00603	Earlier vs Later Treated
3	5	0.2976	0.01930	Earlier vs Later Treated
3	99999	0.1964	0.24425	Treated vs Untreated
4	2	-0.2707	0.00889	Later vs Earlier Treated
4	3	0.2135	0.00603	Later vs Earlier Treated
4	5	0.1182	0.01524	Earlier vs Later Treated
4	99999	-0.0291	0.25710	Treated vs Untreated
5	2	-0.5680	0.01066	Later vs Earlier Treated
5	3	-0.5686	0.00965	Later vs Earlier Treated
5	4	-0.5375	0.00508	Later vs Earlier Treated

treated	untreated	estimate	weight	type
5	99999	-0.6729	0.27424	Treated vs Untreated

Callaway and Sant'Anna (2020)

This estimation gives you multiple outcomes

- Treatment by treatment group
- An overall treatment effect
- Overall dynamic effects / event study
- Treatment effects by calendar time

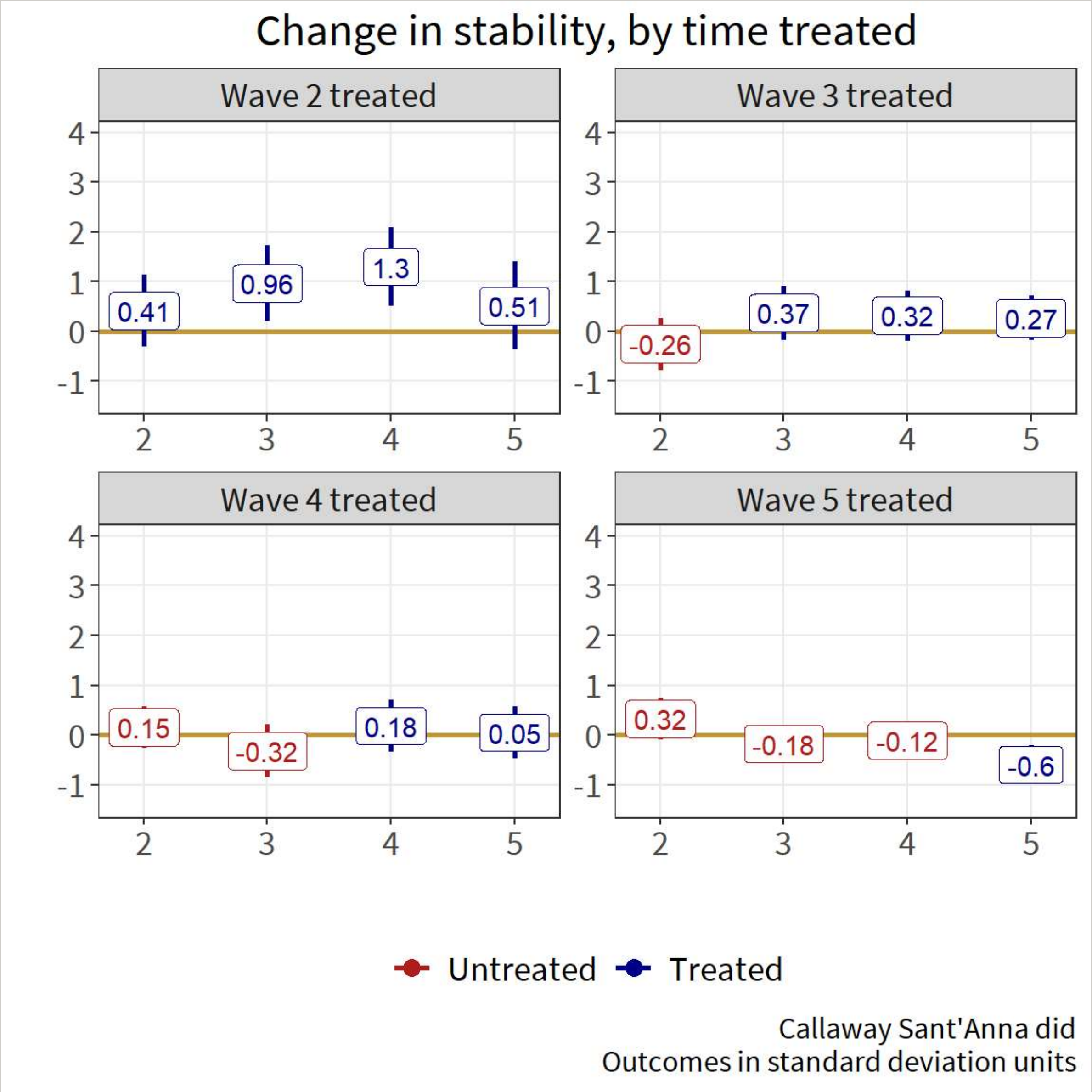
Callaway and Sant'Anna replication

```r
cal <- att_gt(yname="stab_std",
              tname="wave",
              idname="idname",
              gname="first.treat",
              xformla= ~ nsp + ln_dist,
              data=mistifull)
summary(cal)
```

```
Call:
att_gt(yname = "stab_std", tname = "wave", idname = "idname",
    gname = "first.treat", xformla = ~nsp + ln_dist, data = mistifull)

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-
Differences with Multiple Time Periods." Journal of Econometrics, Vol. 225,
No. 2, pp. 200-230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
<https://arxiv.org/abs/1803.09015>
```

```
Group-Time Average Treatment Effects:
 Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
     2    2   0.4105      0.369       -0.5857      1.4066
     2    3   0.9595      0.388       -0.0886      2.0077
     2    4   1.2952      0.404        0.2056      2.3847 *
```


Change in stability, by time treated
Wave 2 treated
Wave 3 treated
Wave 4 treated
Wave 5 treated
0.41
0.96
1.3
0.51
-0.26
0.37
0.32
0.27
0.15
-0.32
0.18
0.05
0.32
-0.18
-0.12
-0.6
Untreated
Treated
Callaway Sant'Anna did
Outcomes in standard deviation units

```r
cal_simple <- aggte(cal, type="simple")
cal_simple
```

```
Call:
aggte(MP = cal, type = "simple")

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-
Differences with Multiple Time Periods." Journal of Econometrics, Vol. 225,
No. 2, pp. 200-230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
<https://arxiv.org/abs/1803.09015>


  ATT    Std. Error     [ 95%  Conf. Int.]
 0.26         0.149    -0.0327       0.552


---
Signif. codes: `*' confidence band does not cover 0
```

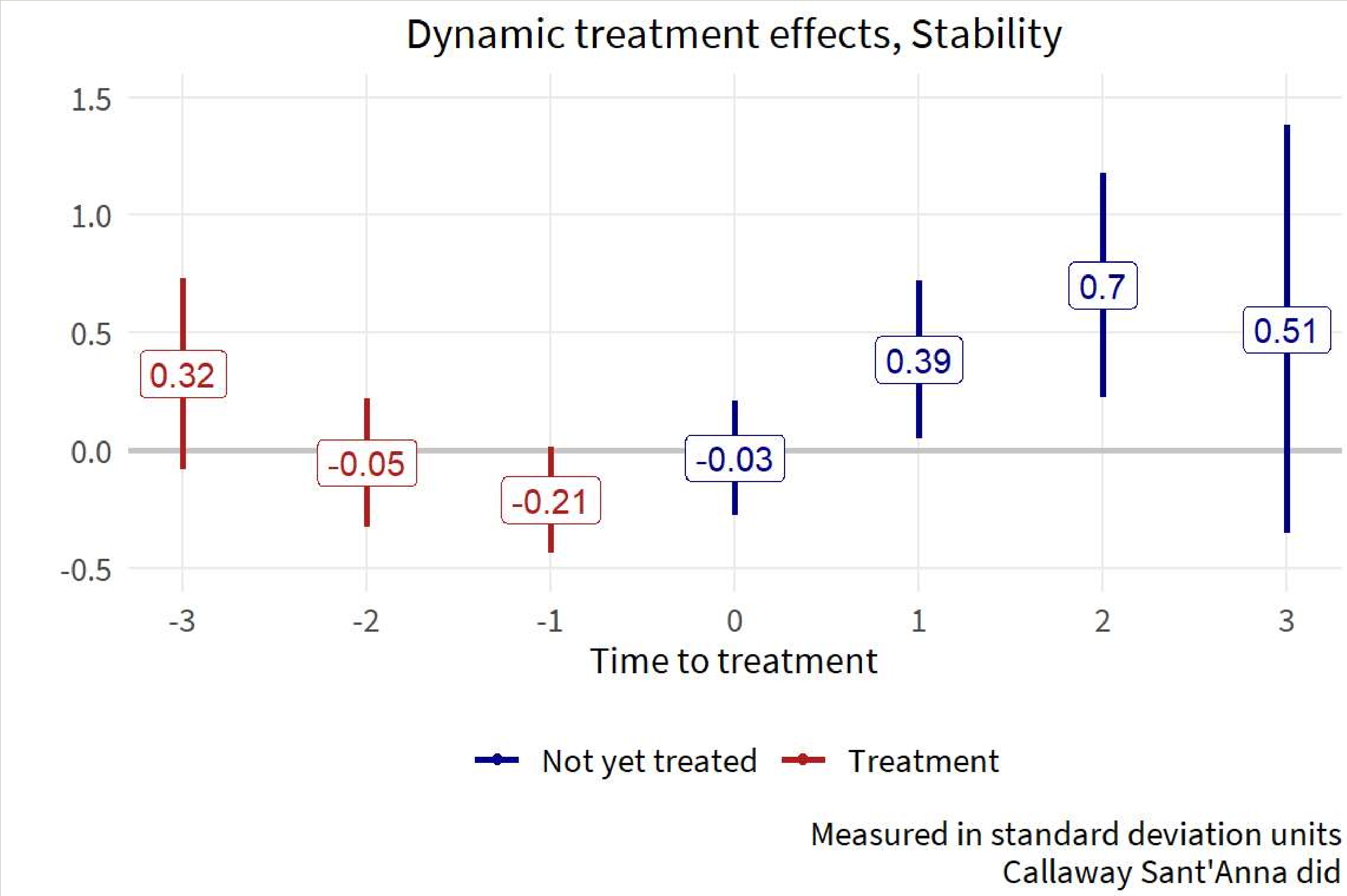
```r
cal_dyn <- aggte(cal, type="dynamic")
cal_dyn
```

```
Call:
aggte(MP = cal, type = "dynamic")

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-
Differences with Multiple Time Periods." Journal of Econometrics, Vol. 225,
No. 2, pp. 200-230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
<https://arxiv.org/abs/1803.09015>


Overall summary of ATT's based on event-study/dynamic aggregation:
   ATT    Std. Error     [ 95%  Conf. Int.]
 0.392        0.199      0.003       0.781 *


Dynamic Effects:
```


Dynamic treatment effects, Stability
1.5
1.0
0.5
0.0
-0.5
0.32
-0.05
-0.21
-0.03
0.39
0.7
0.51
-3
-2
-1
0
1
2
3
Time to treatment
Not yet treated
Treatment
Measured in standard deviation units
Callaway Sant'Anna did

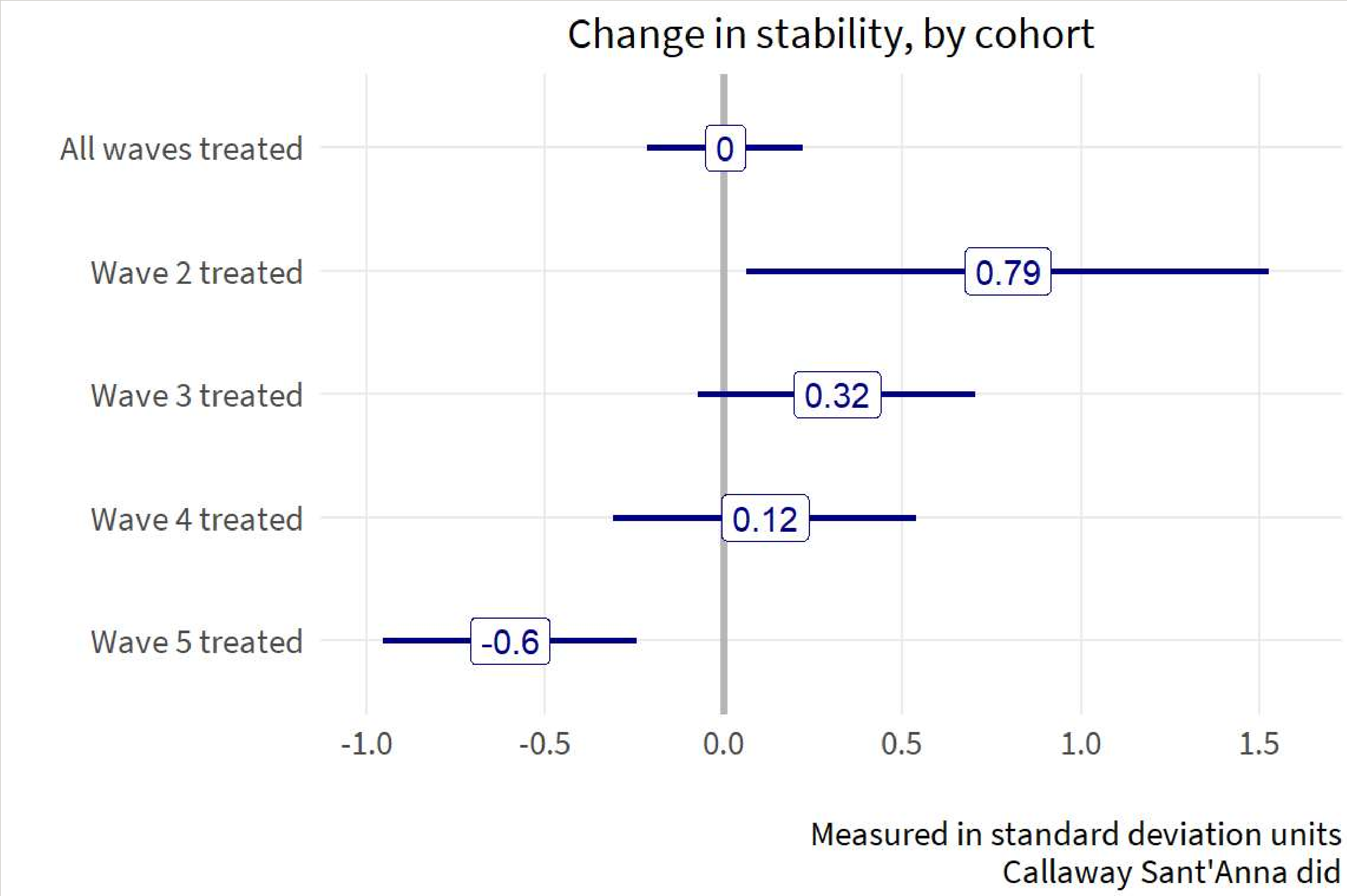
```r
cal_grp <- aggte(cal, type="group")
cal_grp
```

```
Call:
aggte(MP = cal, type = "group")

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-
Differences with Multiple Time Periods." Journal of Econometrics, Vol. 225,
No. 2, pp. 200-230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
<https://arxiv.org/abs/1803.09015>


Overall summary of ATT's based on group/cohort aggregation:
    ATT    Std. Error     [ 95%  Conf. Int.]
 0.0037        0.112     -0.215      0.223


Group Effects:
```


Change in stability, by cohort
All waves treated
Wave 2 treated
Wave 3 treated
Wave 4 treated
Wave 5 treated
0
0.79
0.32
0.12
-0.6
-1.0
-0.5
0.0
0.5
1.0
1.5
Measured in standard deviation units
Callaway Sant'Anna did

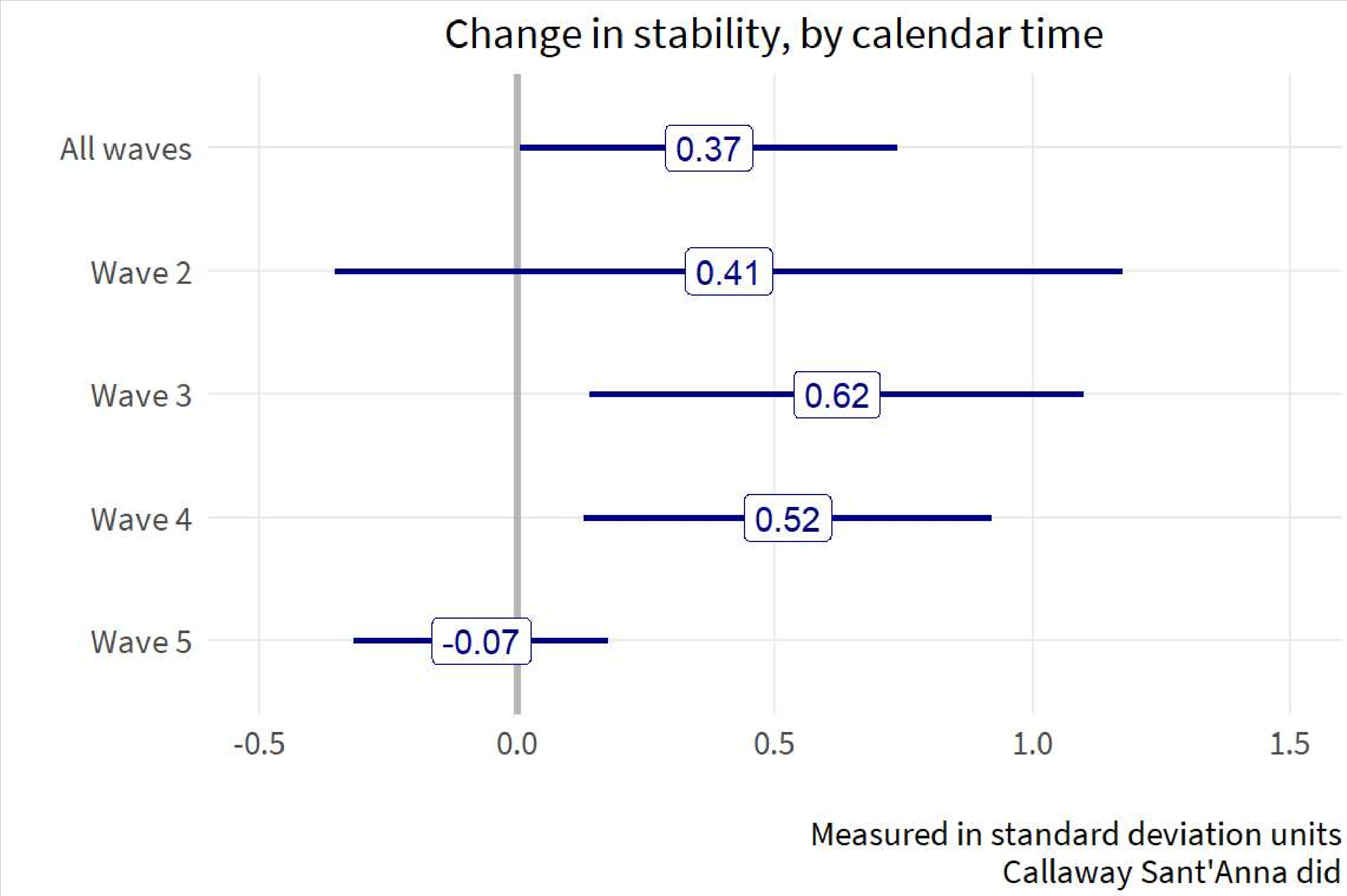
```r
cal_cal <- aggte(cal, type="calendar")
cal_cal
```

```
Call:
aggte(MP = cal, type = "calendar")

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-
Differences with Multiple Time Periods." Journal of Econometrics, Vol. 225,
No. 2, pp. 200-230, 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
<https://arxiv.org/abs/1803.09015>


Overall summary of ATT's based on calendar time aggregation:
   ATT    Std. Error     [ 95%  Conf. Int.]
 0.371        0.187     0.0048       0.737 *


Time Effects:
```


Change in stability, by calendar time
All waves
Wave 2
Wave 3
Wave 4
Wave 5
0.37
0.41
0.62
0.52
-0.07
-0.5
0.0
0.5
1.0
1.5
Measured in standard deviation units
Callaway Sant'Anna did

- Background
- Problem
- Solutions
- Case study - MISTI
- **Final thoughts**

What have we learned?

- In certain settings, two-way fixed effects estimation is biased in ways that we only recently came to realize
- We have to carefully think through the data generating process (logic modeling) for each individual setting
- As we get more granular data and ask deeper questions, econometric tools are starting to provide better insight into treatment dynamics

What should we do?

- For any two-way fixed effects setting, use the Bacon decomposition to diagnose any problems
- Use stacked d-i-d to remove problematic 2x2 cells, or apply any of the new estimators
- Go back to your old evaluations!!

Thank you!