Causal Panel Analysis under Parallel Trends: Lessons from A Large Replication Study

American Political Science Review

Albert Chiu Xingchen Lan Ziyi Liu Yiqing Xu

B. Markdown Files

Beazer and Reuter (2022)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

A First Look at Data

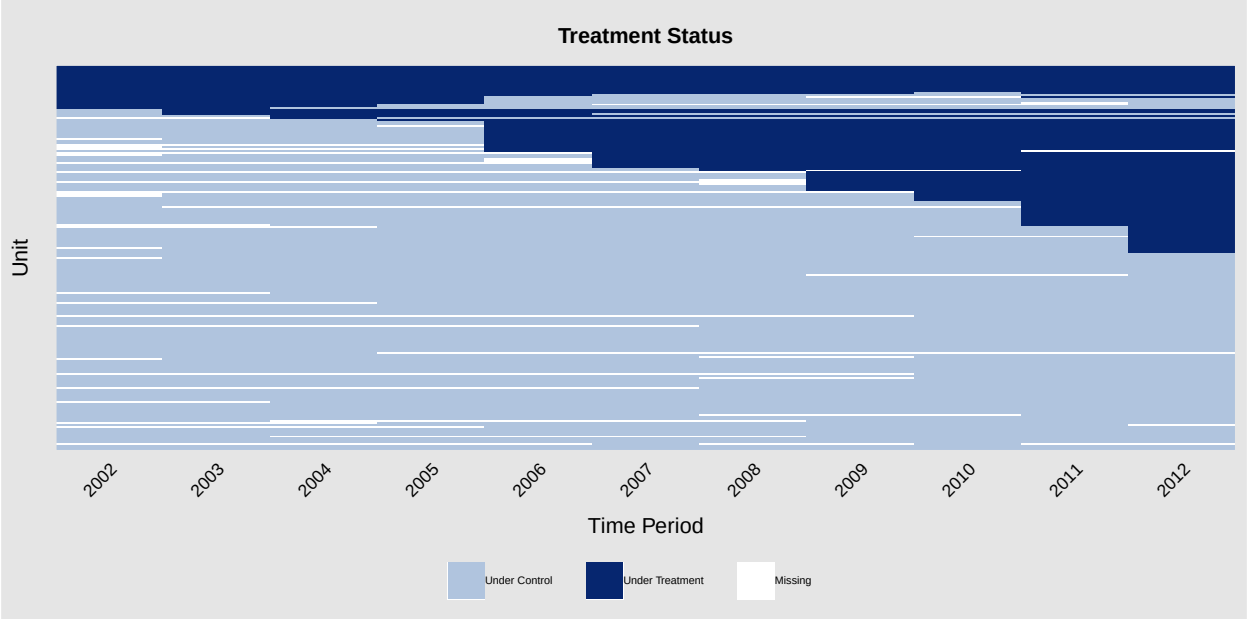
The paper investigates the effects of appointed mayors (v.s. elected mayors) on public goods provision (the maintenance of old and unsafe housing), using Russian city-year level panel data between 2002 and 2012. One of the main findings of this paper is that "replacing mayoral elections with appointments is associated with an average increase of just over 20,100 m^2 of unsafe housing." (p444-445, Table 1)

Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the city level.

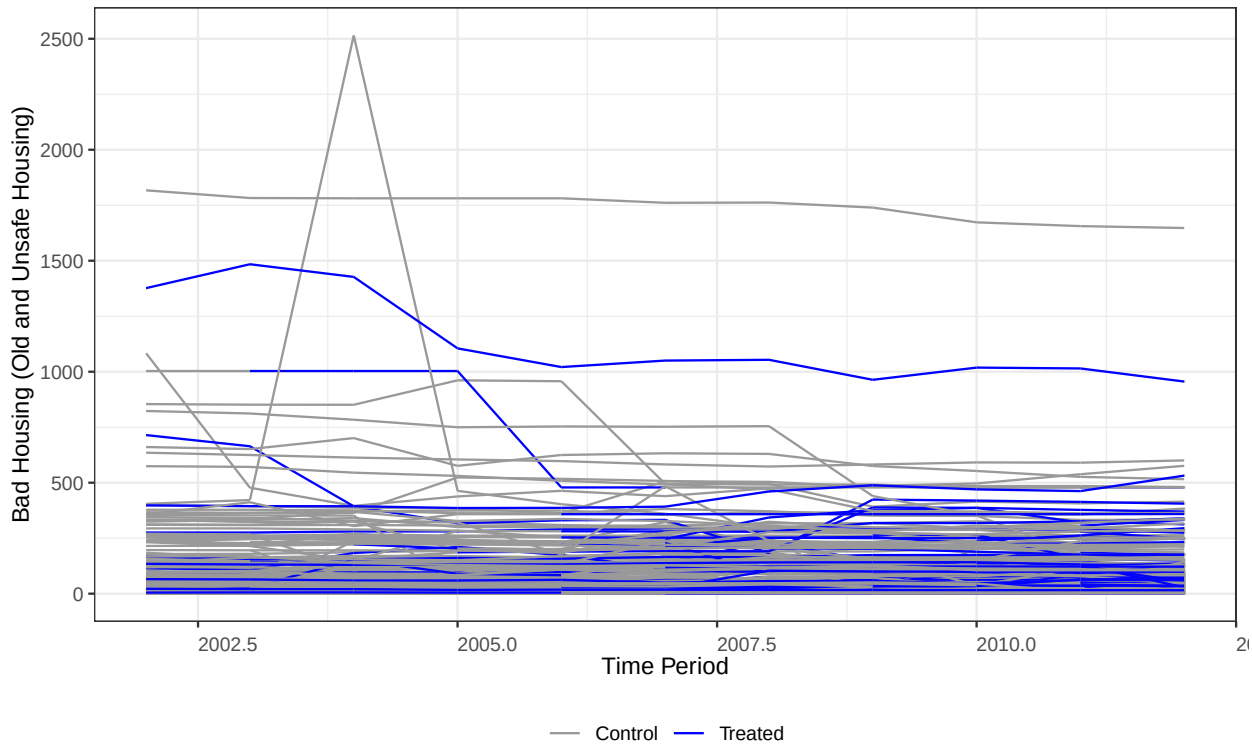
Table 1: Replication Summary

Unit of analysis	City × year
Treatment	Appointed mayor
Outcome	Old and unsafe housing in 1,000 squaremeters
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a city). The treatment has reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

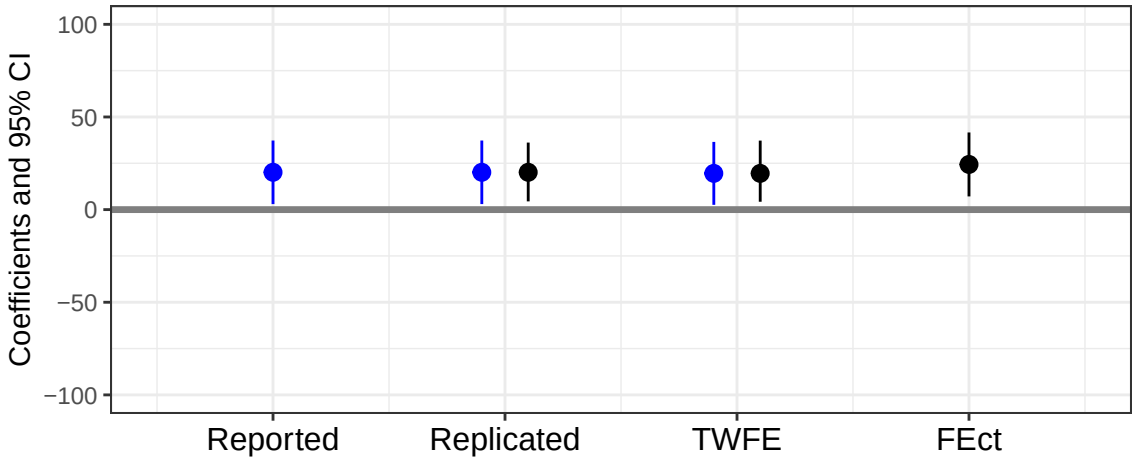
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and imputation estimator (FEct). The point estimates and

their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

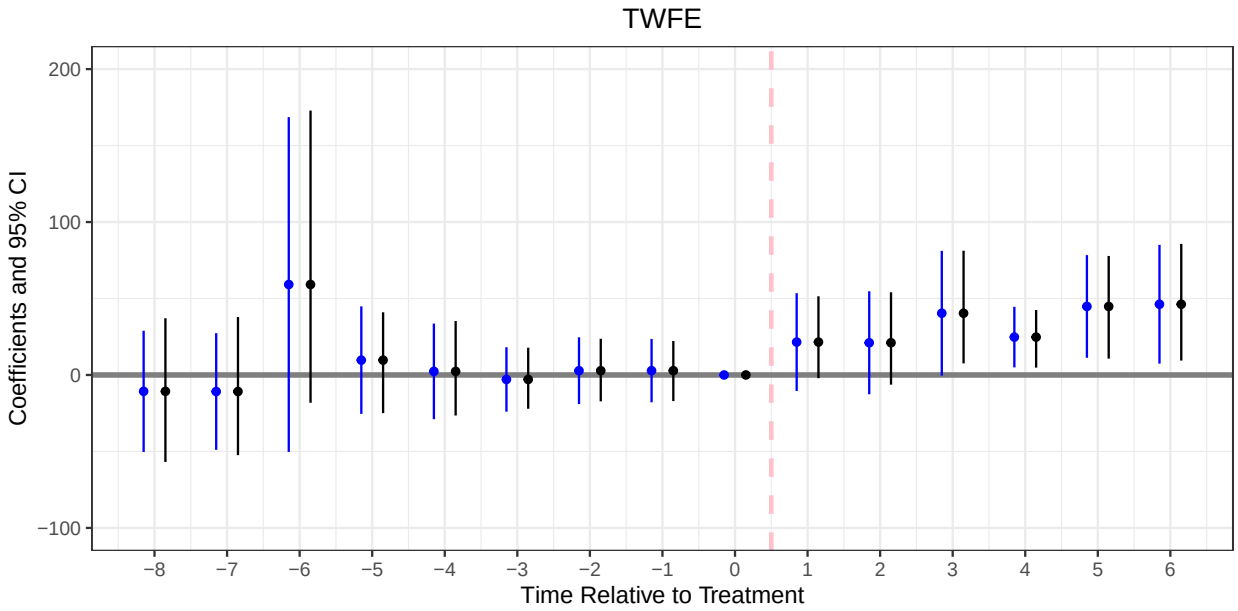
```
sol <- feols(badzhil~appointed+zhiltot+pressfreedom+
               workpop+AvgSalary1+regdem+birthrate+regprom+pop|city_id+year,
             data = df,cluster = "city_id")
summary(sol)
```

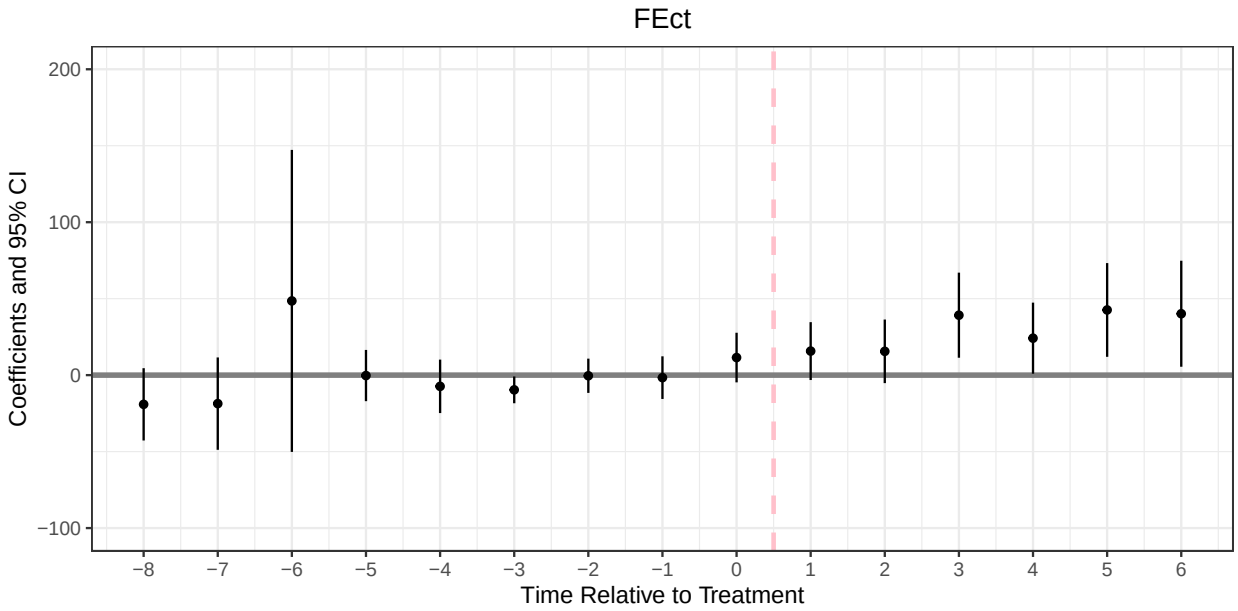
```
## OLS estimation, Dep. Var.: badzhil
## Observations: 2,027
## Fixed-effects: city_id: 199,  year: 11
## Standard-errors: Clustered (city_id)
##                 Estimate Std. Error   t value Pr(>|t|)
## appointed     20.135595   8.753984  2.300163 0.022481 *
## zhiltot       -0.020352   0.011345 -1.793922 0.074351 .
## pressfreedom -13.991886   6.892482 -2.030021 0.043692 *
## workpop        0.351233   0.326577  1.075497 0.283461
## AvgSalary1    -0.448646   0.262603 -1.708459 0.089118 .
## regdem         0.093511   1.309676  0.071400 0.943151
## birthrate      4.058484   2.825439  1.436408 0.152464
## regprom      266.600825 386.692739  0.689438 0.491355
## pop            0.306841   0.599405  0.511909 0.609286
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 65.8     Adj. R2: 0.88893
##               Within R2: 0.031025
```



Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct.

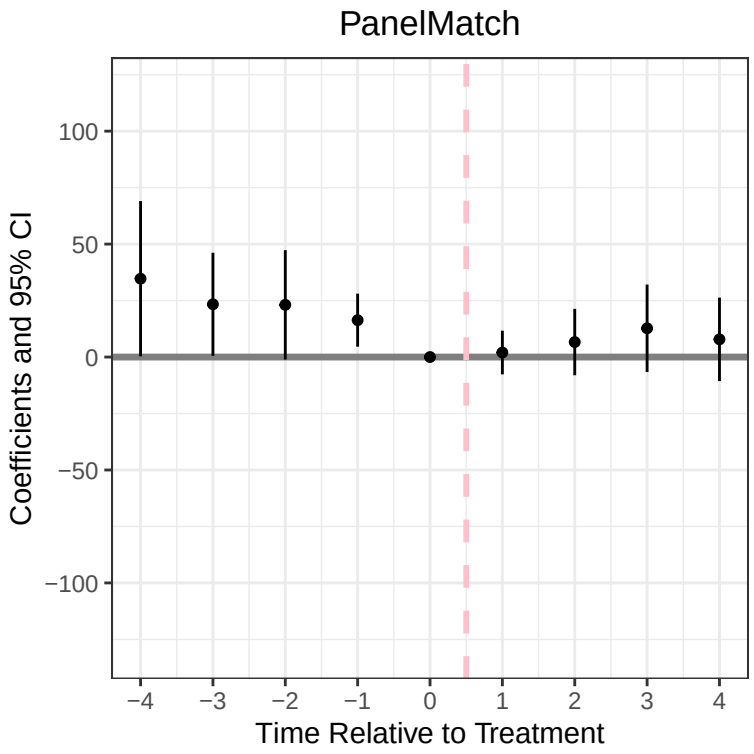




The TWFE and FEct estimator are consistent with each other. The estimated DTEs are all positive in the plot.

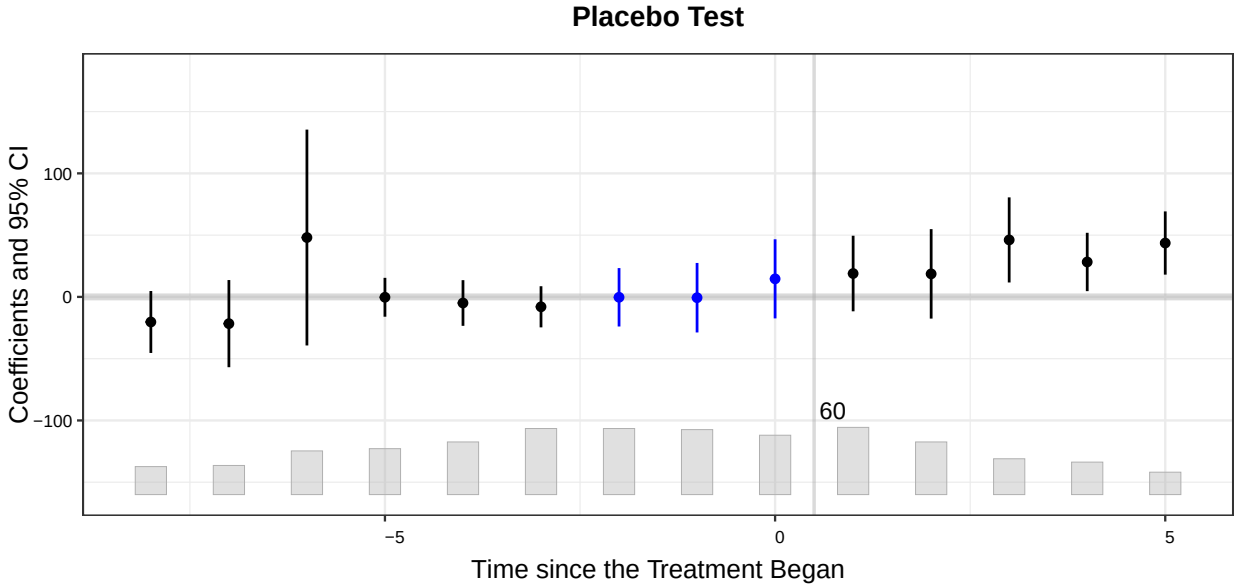
DTE of a Balanced Subsample

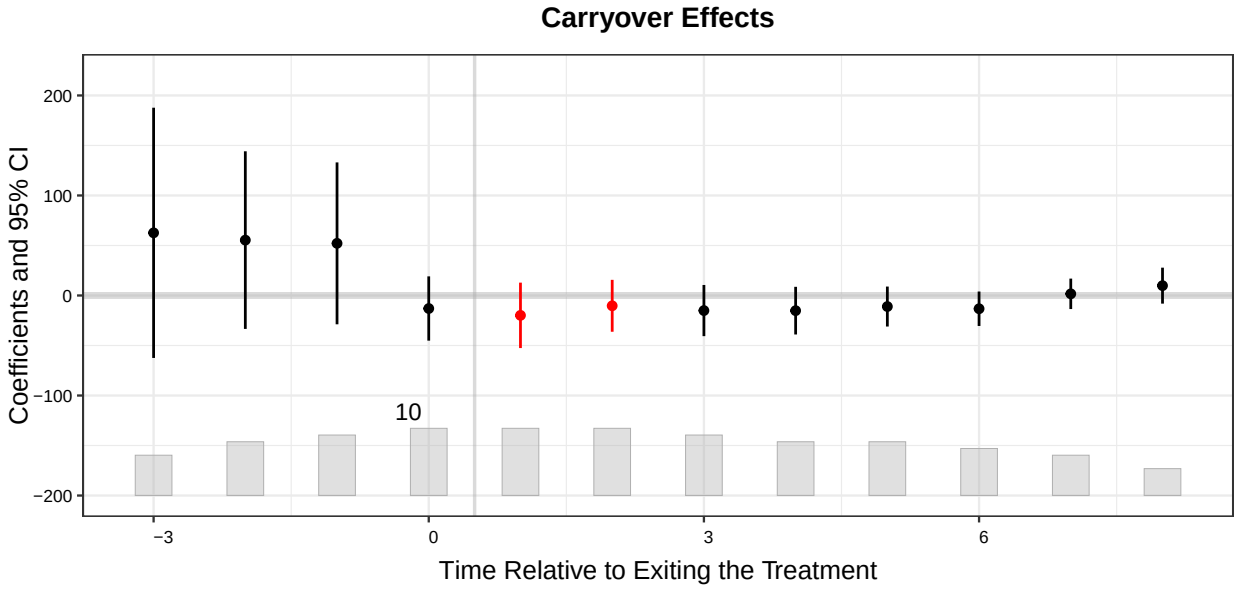
We also estimate the DTE using PanelMatch (lead = 4 and lag = 5). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





Test Statistics

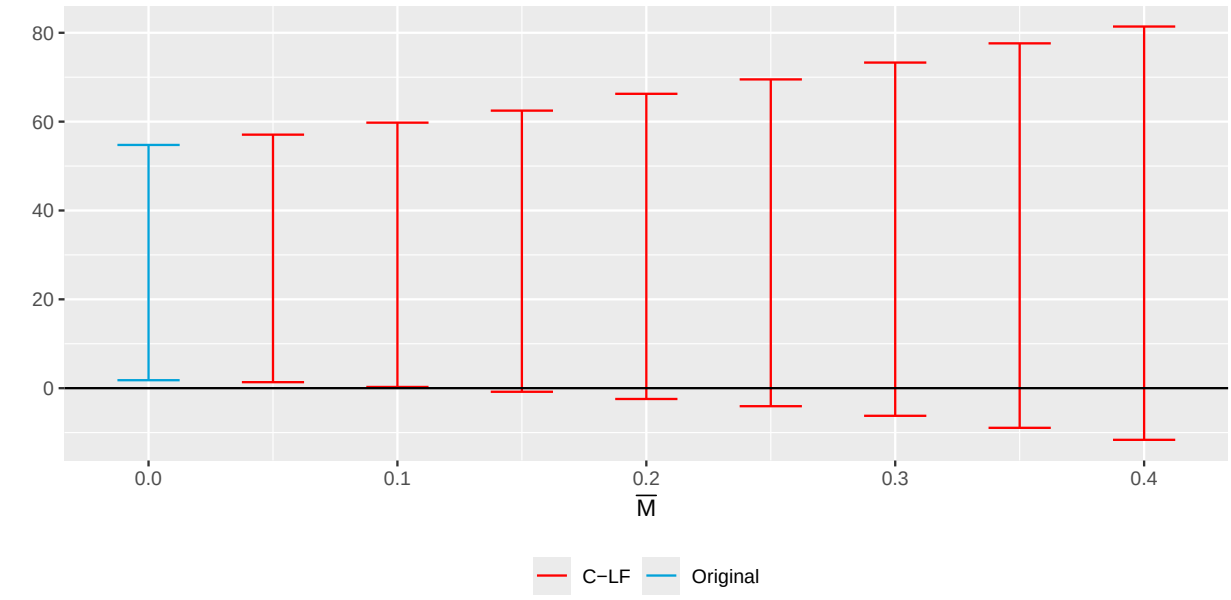
```
##                                    p-value
## F test                               0.297
## Equivalence test (default)           0.679
## Equivalence test (threshold=ATT)     0.694
## Placebo test                         0.748
## Carryover effect test                0.296
```

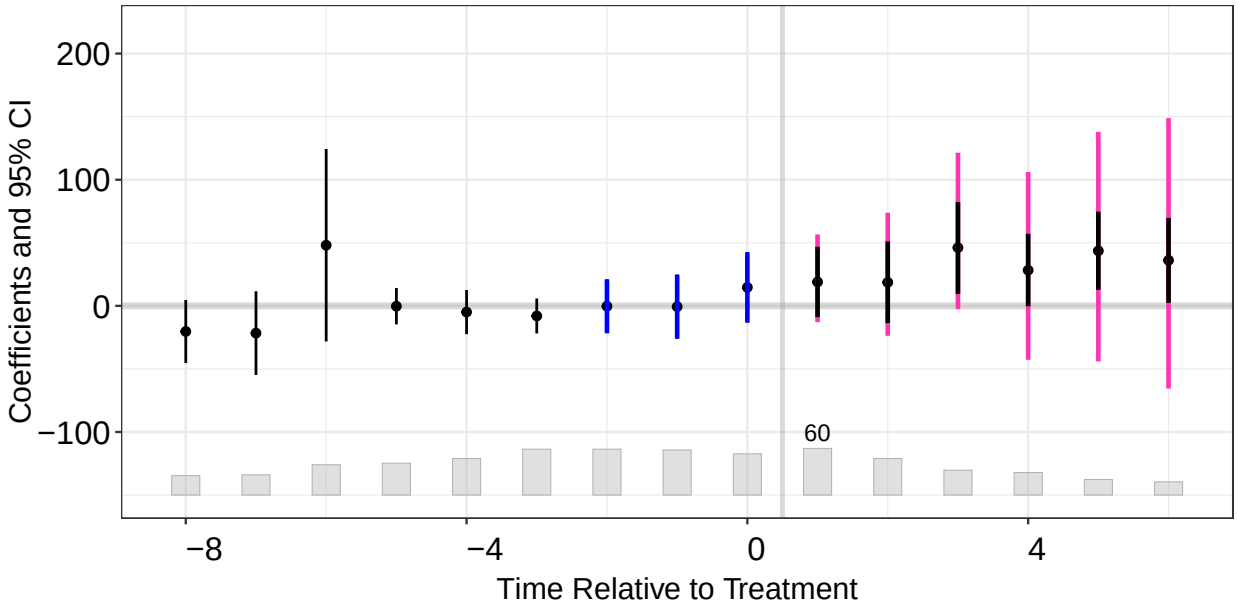
We find little evidence of violations of the parallel trends (PT) assumption, anticipation effects, or carryover effects. However, the equivalence test fails to reject the null hypothesis that the residuals in the pre-treatment periods exceed the estimated ATT, possibly due to limited statistical power.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.1$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.1 times the maximum deviation observed between consecutive periods in the placebo periods.

We further evaluate the robustness of the estimated DTE under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, accounting for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model seems to be robust to HTE-robust estimators. There is no strong evidence for potential PT violations, anticipation effects, or carryover effects. However, due to limited power, we cannot reject the null that the ATT is different from zero under mild PT violations benchmarked against the pre-treatment placebo estimates.

Bischof and Wagner (2019)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 6
Sensitivity Analysis 7
Summary 8

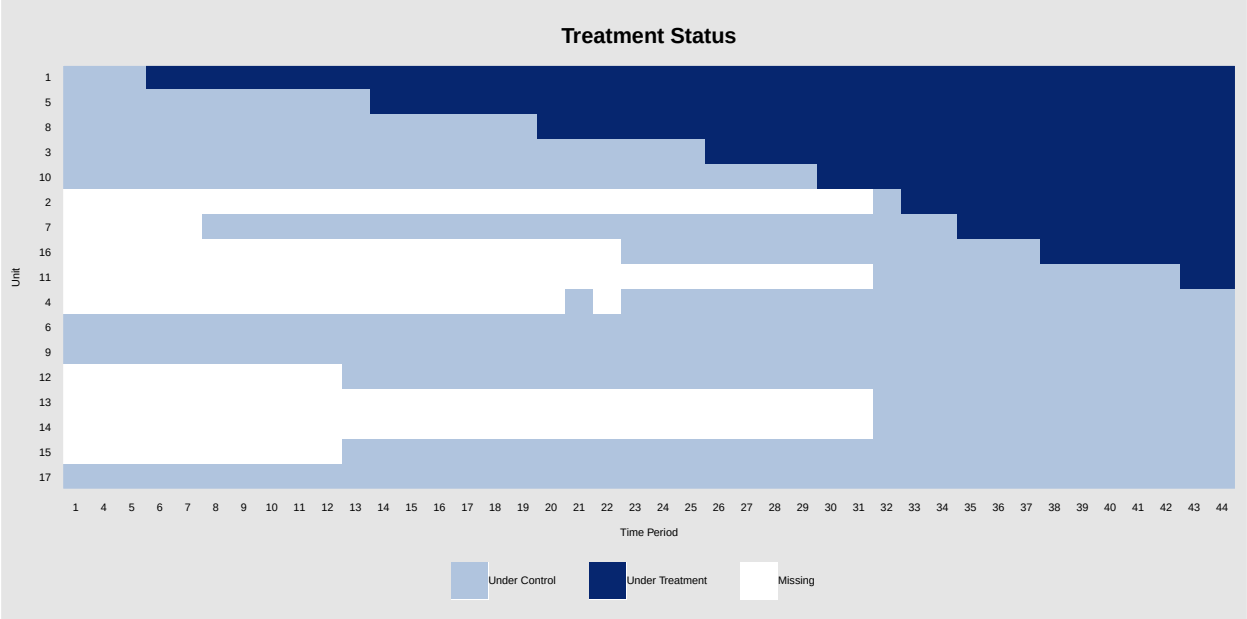
A First Look at Data

This paper investigates the effect of radical-right party entrance on polarization, using country-year panel data from 17 European countries during 1973-2016. One of the main findings of this paper is that "The effect of radical-right party entrance is also substantively large, resulting in an increase in polarization of about half a standard deviation" (p898).

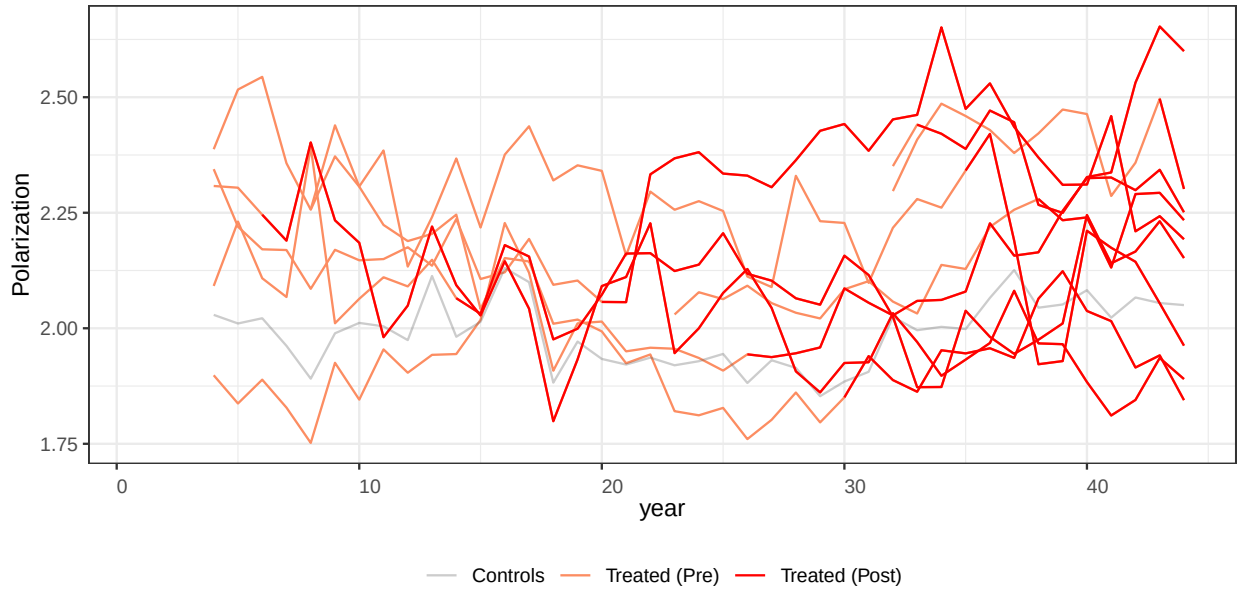
Model. We focus on **Model 2 of Table 3** in the paper. The authors use country and decade fixed effects model and report robust standard errors clustered at the country level. The decade fixed effects are less granular than the year fixed effects.

Replication Summary	
Unit of analysis	Country × year
Treatment	Entrance of Extreme Right Party
Outcome	Polarization
Treatment type	Staggered
Outcome type	Continous
Fixed effects	Unit+Higher-level Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each colum represents a time period (a year) and each row represents a unit (a country). There is some missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each country. The trajectories of the control units are depicted in gray. For the ever-treated units, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

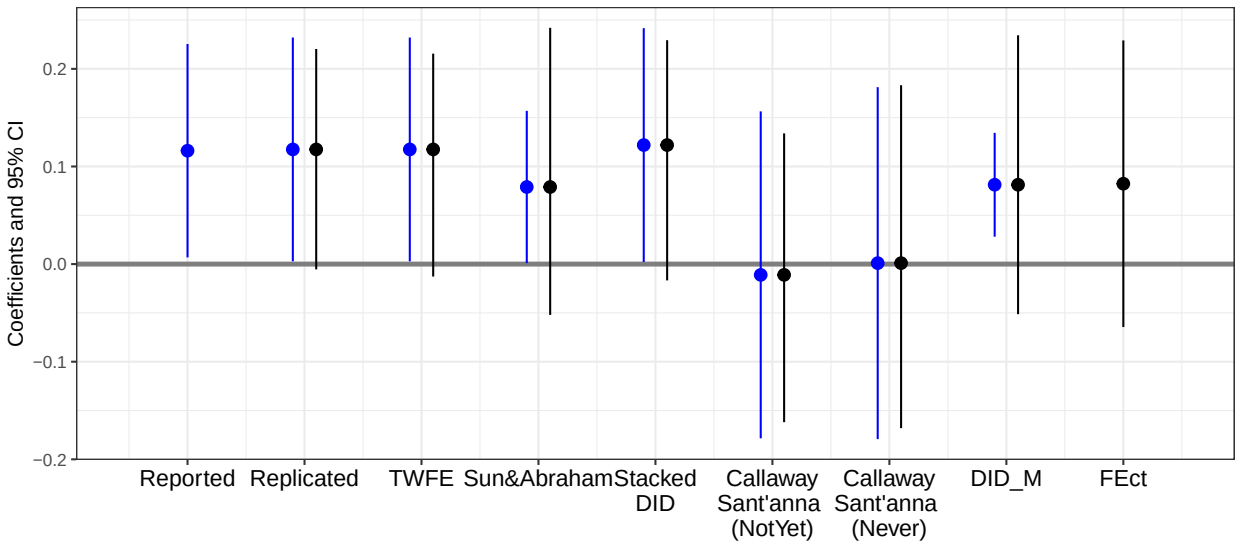
We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this case) and apply the TWFE (with country and year fixed effects), Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DID multiple estimator (DIDM) and imputation estimator (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

```
sol <- feols(polarization~rtreatment|country+decade,data = df,cluster = "country")
summary(sol)
```

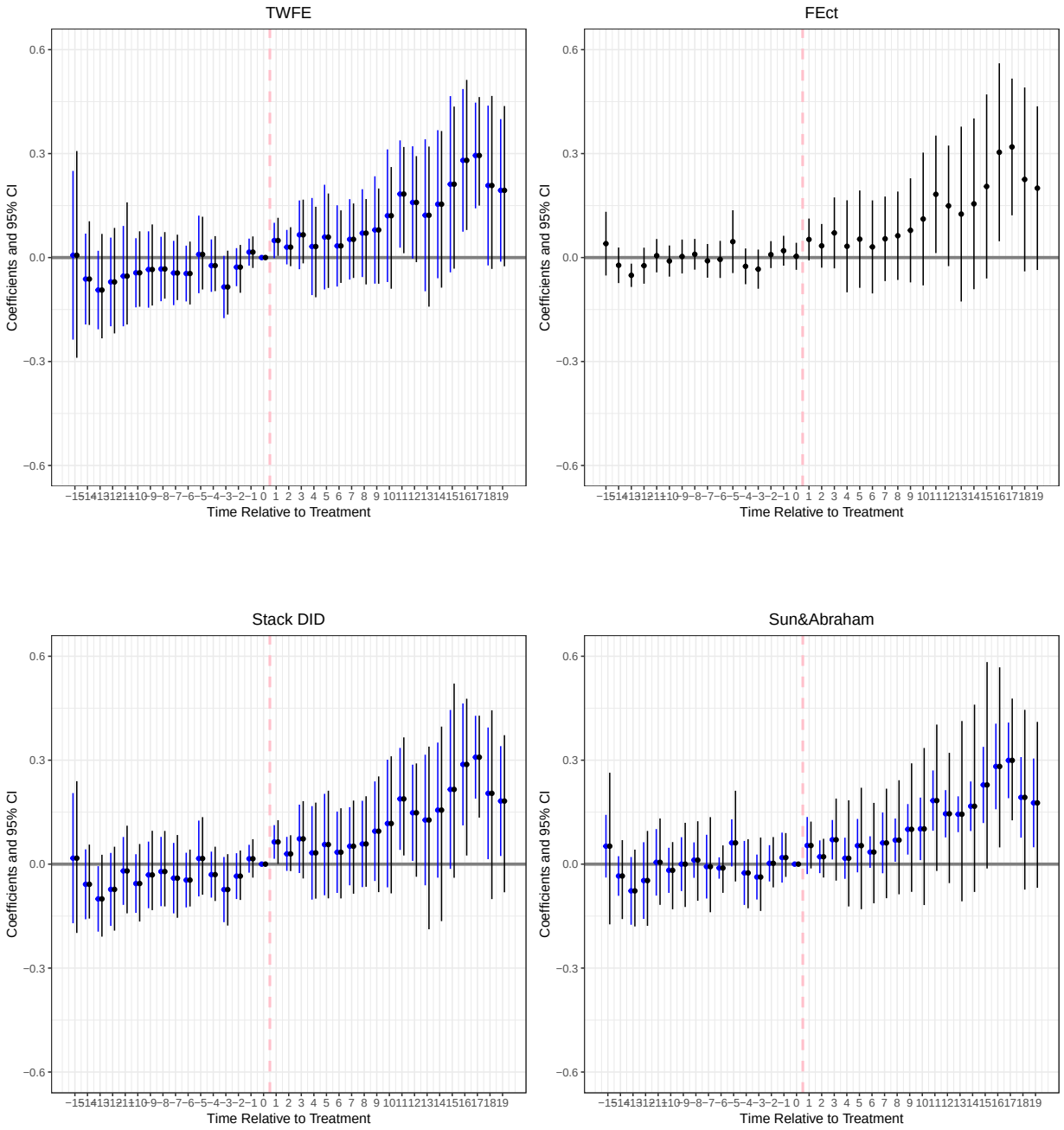
```
## OLS estimation, Dep. Var.: polarization
## Observations: 534
## Fixed-effects: country: 17,  decade: 5
## Standard-errors: Clustered (country)
##             Estimate Std. Error t value Pr(>|t|)
## rtreatment 0.116163   0.055776 2.08268 0.053685 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.125562     Adj. R2: 0.660432
##                  Within R2: 0.058052
```

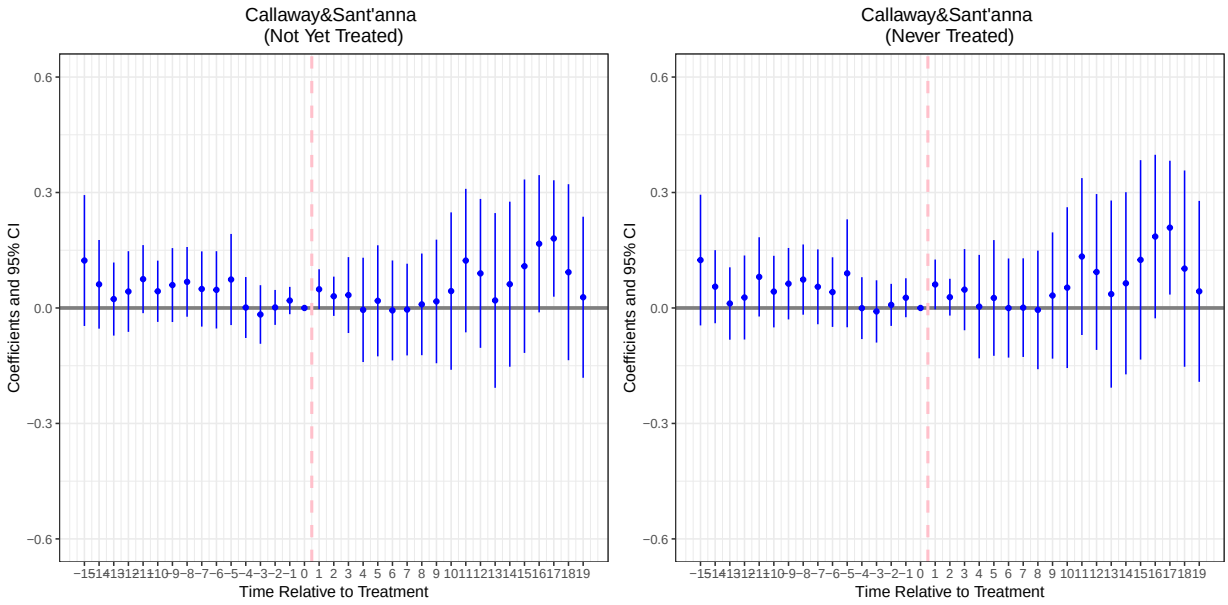
```
##   |                                                                      |
```

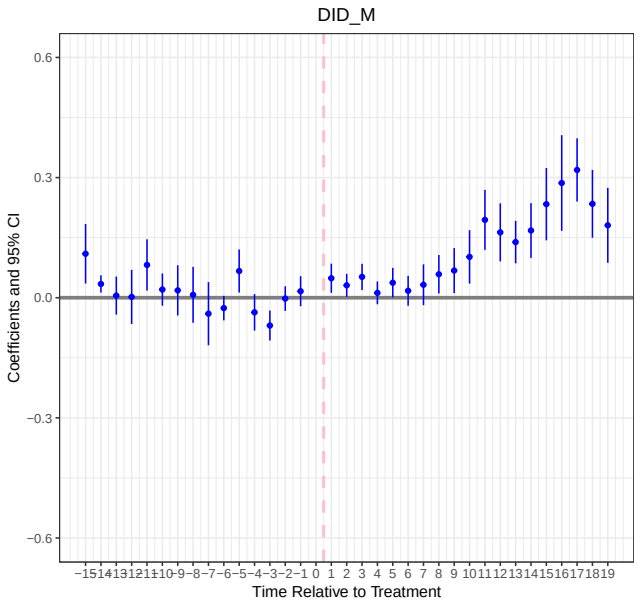


Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DID_M, and FEct. The results are shown below.


TWFE
FEct
Stack DID
Sun&Abraham
Coefficients and 95% CI
0.6
0.3
0.0
−0.3
−0.6
−15 −14 −13 −12 −11 −10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Time Relative to Treatment

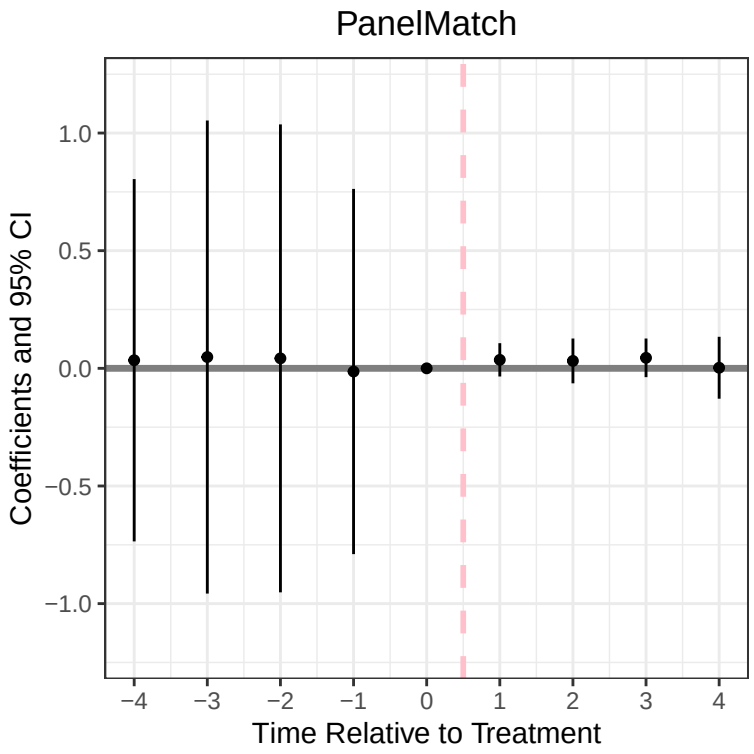




All HTE-robust estimators and TWFE yield similar estimated DTE. The estimated DTE are mostly positive but insignificant. The analytical standard error of the DIDM estimator is under-estimated.

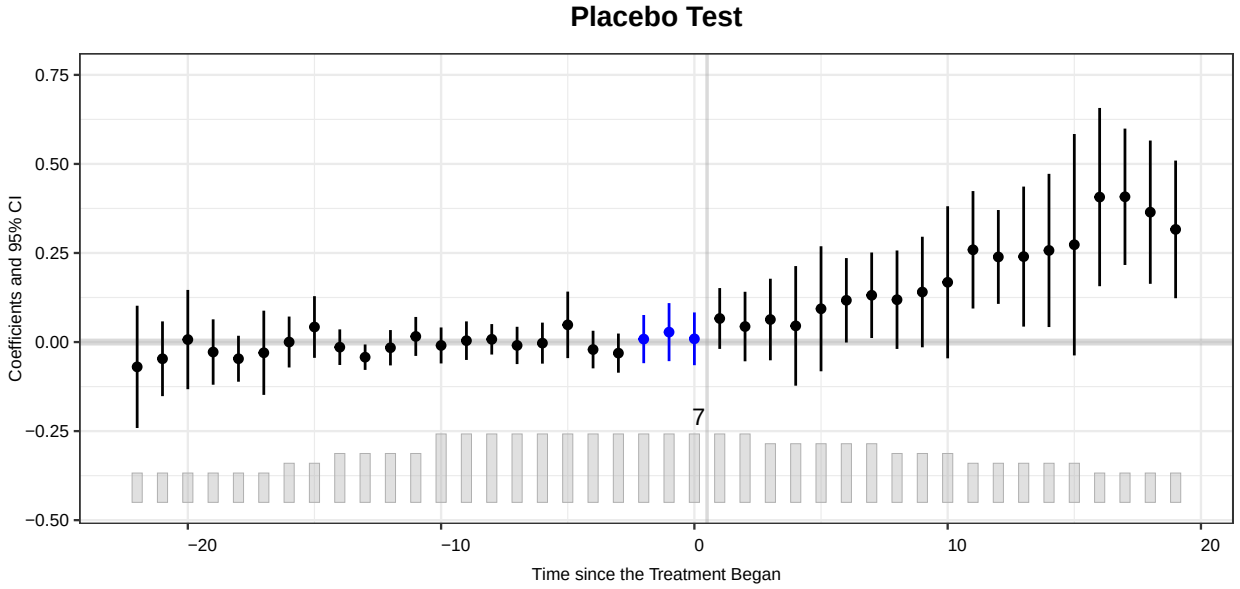
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 4 and lag = 5). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



Test Results

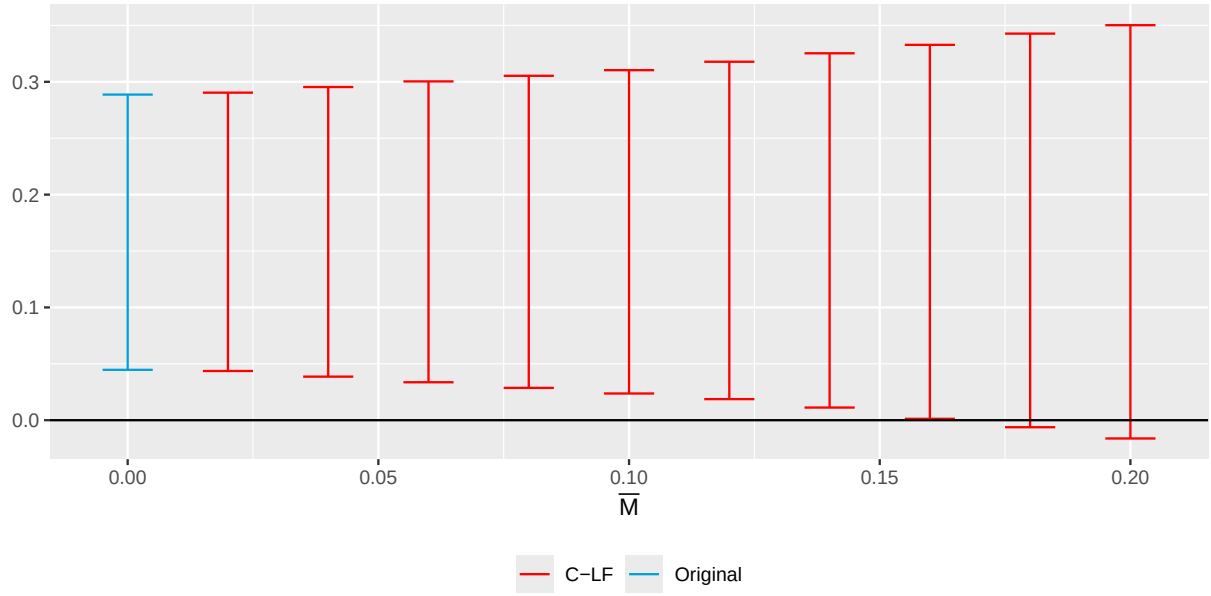
```
##                                   p-value
## F test                              0.957
## Equivalence test (default)          0.544
## Equivalence test (threshold=ATT)    0.215
## Placebo test                        0.665
## Carryover effect test                  NA
```

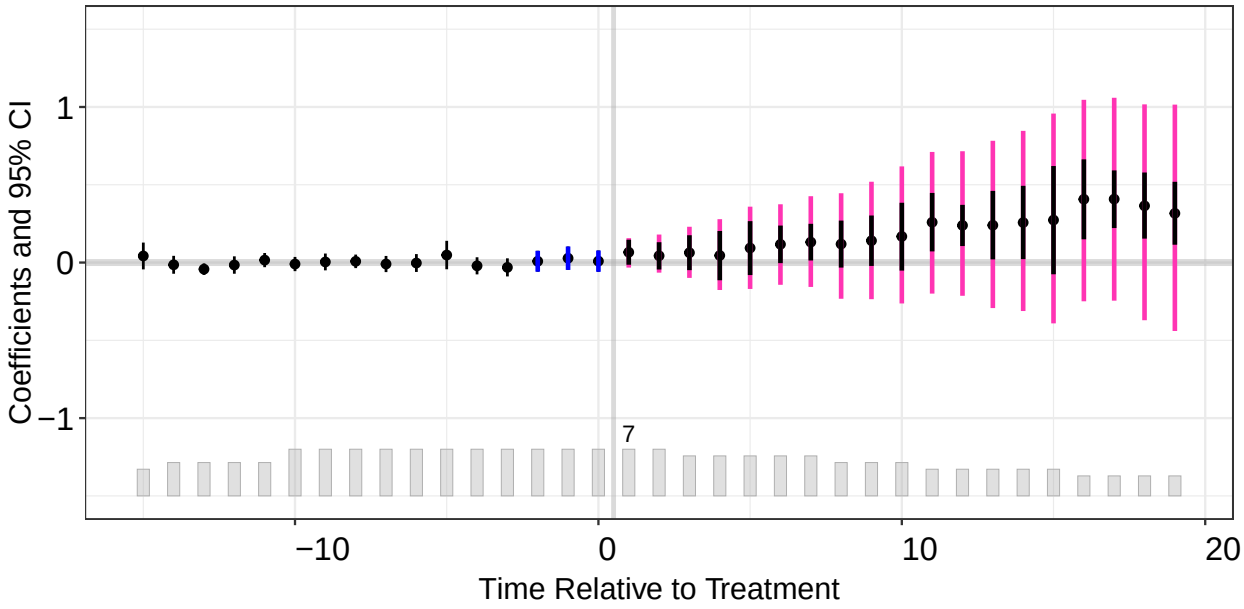
We find little evidence of violations of the parallel trends (PT) assumption, anticipation effects, or carryover effects. However, the equivalence test fails to reject the null hypothesis that the residuals in the pre-treatment periods exceed the estimated ATT, possibly due to limited statistical power.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.16$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.16 times the maximum deviation observed between consecutive periods in the placebo periods.

We further evaluate the robustness of the estimated DTE under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, accounting for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Most HTE-robust estimators (except for CSDID) produce point estimates similar to those of the TWFE model; however, these estimates are no longer statistically significant. Our DTE estimation suggests a positive treatment effect following the intervention, but the average effect lacks statistical significance, likely due to limited power. There is no strong evidence of potential PT violations or anticipation effects. However, due to limited power, we cannot reject the null hypothesis that the ATT differs from zero under mild PT violations benchmarked against the pre-treatment placebo estimates.

Bisgaard and Slothuus (2018)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Summary 4

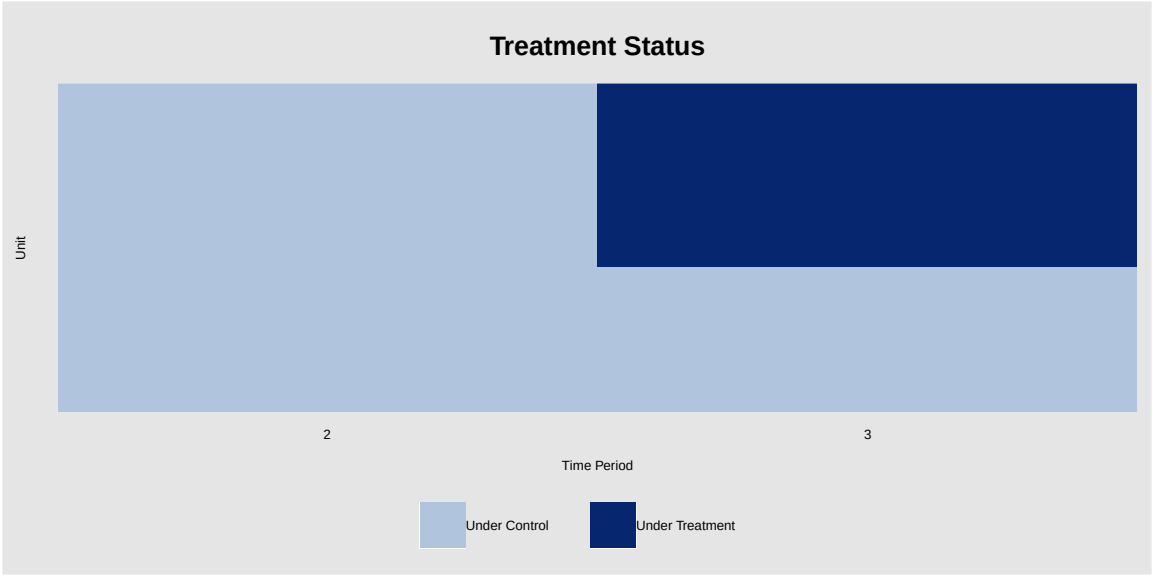
A First Look at Data

This paper investigates the effect of party cues on partisan perceptual gaps using individual-period panel survey data from Denmark. The paper finds that "Partisans do appear to form more positive perceptions of economic reality if their party is in office." (p463).

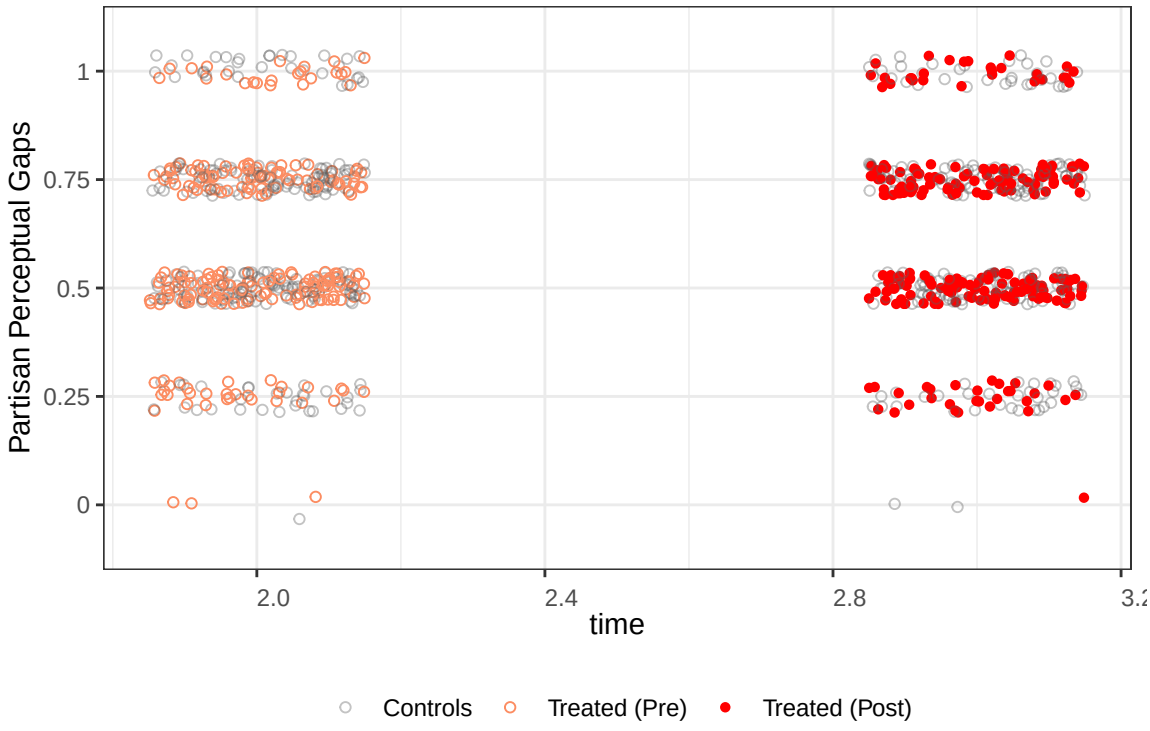
Model. We focus on **Table 1** in the paper. The authors use a a two-way fixed effects (TWFE) model and report robust standard errors clustered at the individual level.

Replication Summary	
Unit of analysis	Individual × period
Treatment	Party cues
Outcome	Partisan percenptual gaps
Treatment type	Classic
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period and each row represents a unit (an individual).



Plotting the outcome variable. We plot the outcomes for each individual. The observations of control units are represented in gray, while the untreated observations of ever-treated units are shown in light blue. The treated observations of ever-treated units are highlighted in deep blue.



Point Estimates

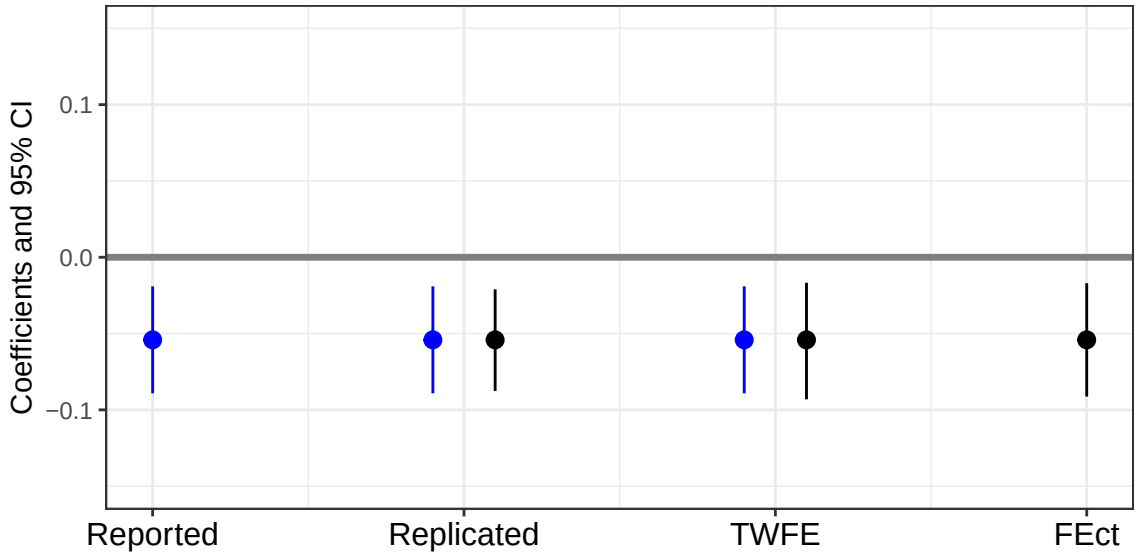
We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and imputation estimator (FEct).

The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(bi~treatment|id+time,data = df,cluster = "id")
summary(sol)
```

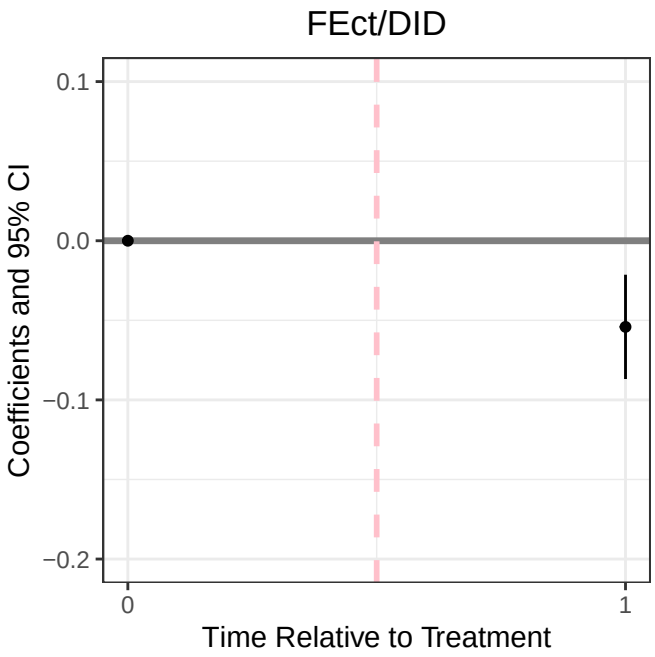
```
## OLS estimation, Dep. Var.: bi
## Observations: 1,140
## Fixed-effects: id: 570,  time: 2
## Standard-errors: Clustered (id)
##            Estimate Std. Error  t value  Pr(>|t|)
## treatment -0.054114   0.017889 -3.02497 0.0025988 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.10734     Adj. R2: 0.525692
##                 Within R2: 0.01543
```



As this data poses a classic difference-in-differences pattern, the TWFE and FEct estimator give the same result. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plot using FEct. Because there is one pre-treatment period, TWFE is equivalent to FEct and DID.



Summary

Overall, the main result of the chosen model appears to be robust to bootstrap uncertainty measures. Because there is one pre-treatment period, we cannot evaluate whether the parallel trends (PT) assumption is plausible.

Blair, Christensen and Wirtschafter (2022)

Contents

A First Look at Data 1

Point Estimates 2

Dynamic Treatment Effects 3

Diagnostic Tests 5

Sensitivity Analysis 6

Summary 7

A First Look at Data

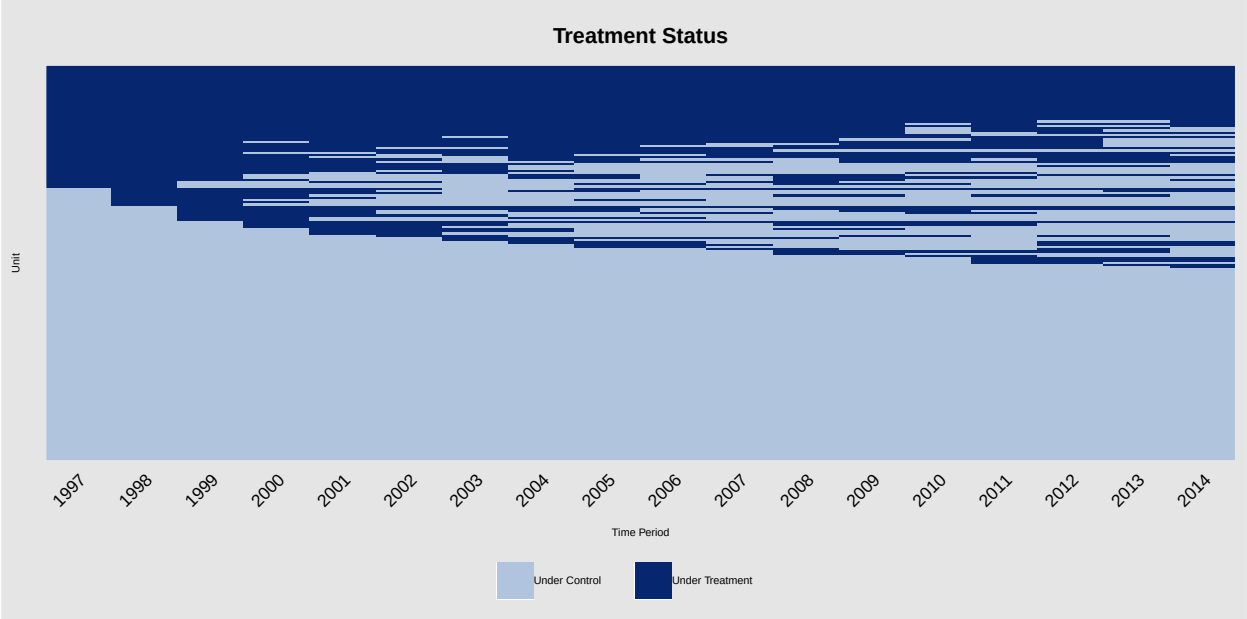
The paper investigates the effects of armed conflicts on investment, using country-year level panel data between 1997 and 2014. One of the main findings of this paper is that "the incidence of at least one fatal armed conflict in the current or previous year reduces aggregate investment by 0.77 log points" (p128, Table 5).

Model. We focus on **Model 1 of Table 5** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the country level.

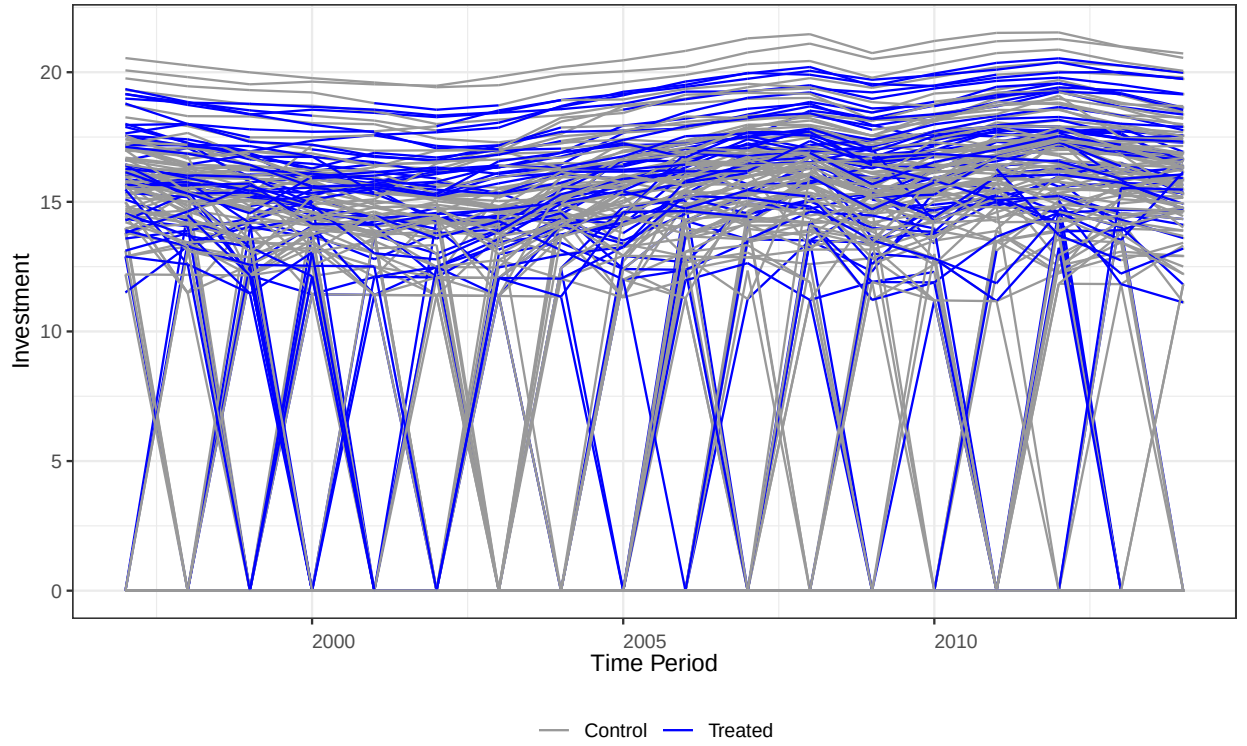
Table 1: Replication Summary

Unit of analysis	Country × year
Treatment	Armed conflict
Outcome	Investment
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a country).



Plotting the outcome variable. We plot the trajectory of the outcome variable for each country. The observations under treated status are marked in blue.



Point Estimates

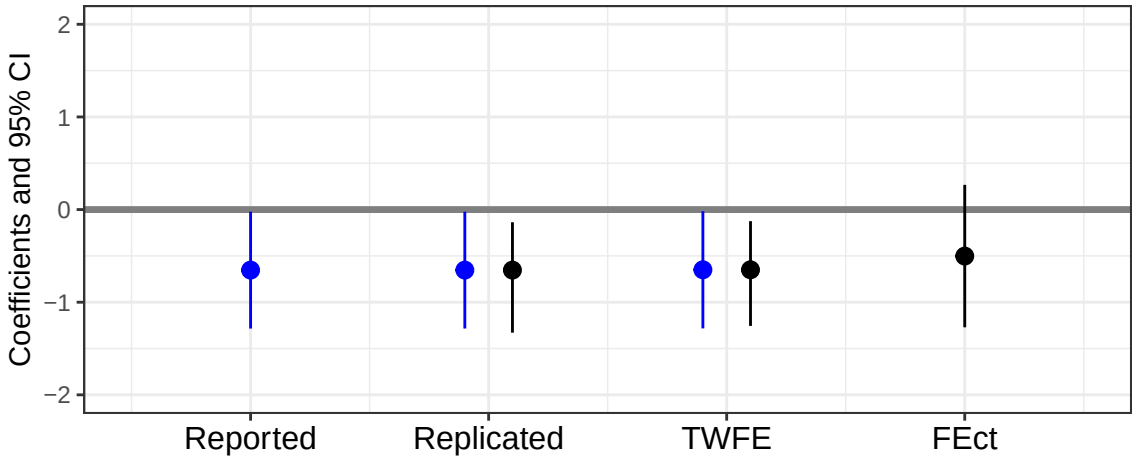
We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and imputation estimator (FEct).

The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

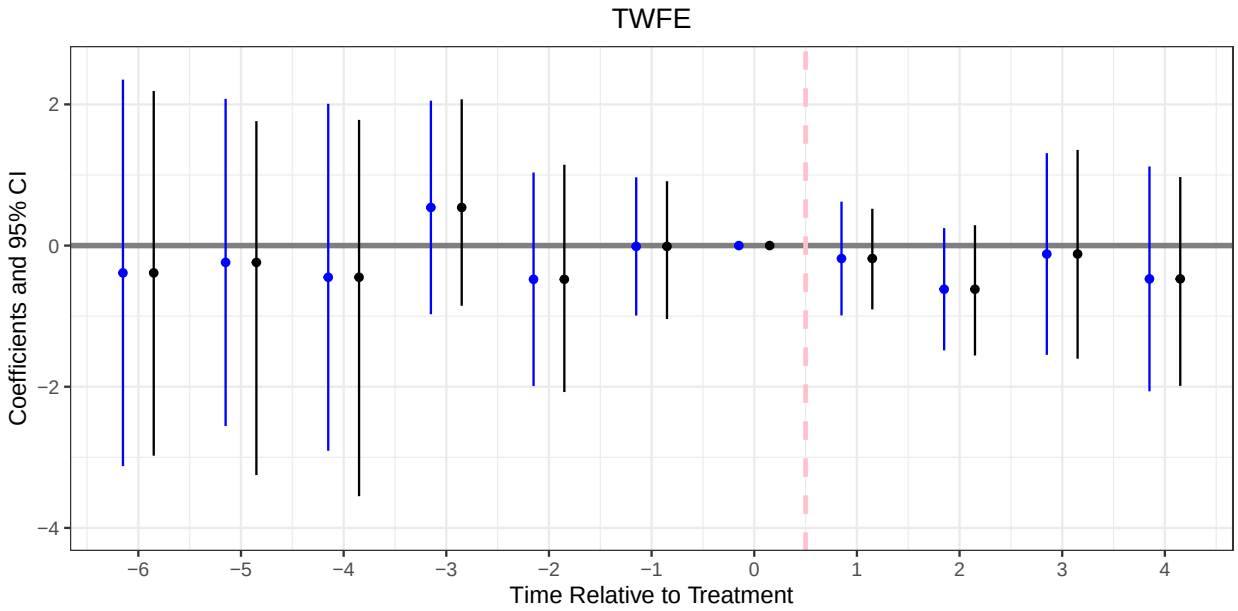
```
sol <- feols(dv~ucdp_lead|country+year,
             data = df,cluster = "country")
summary(sol)
```

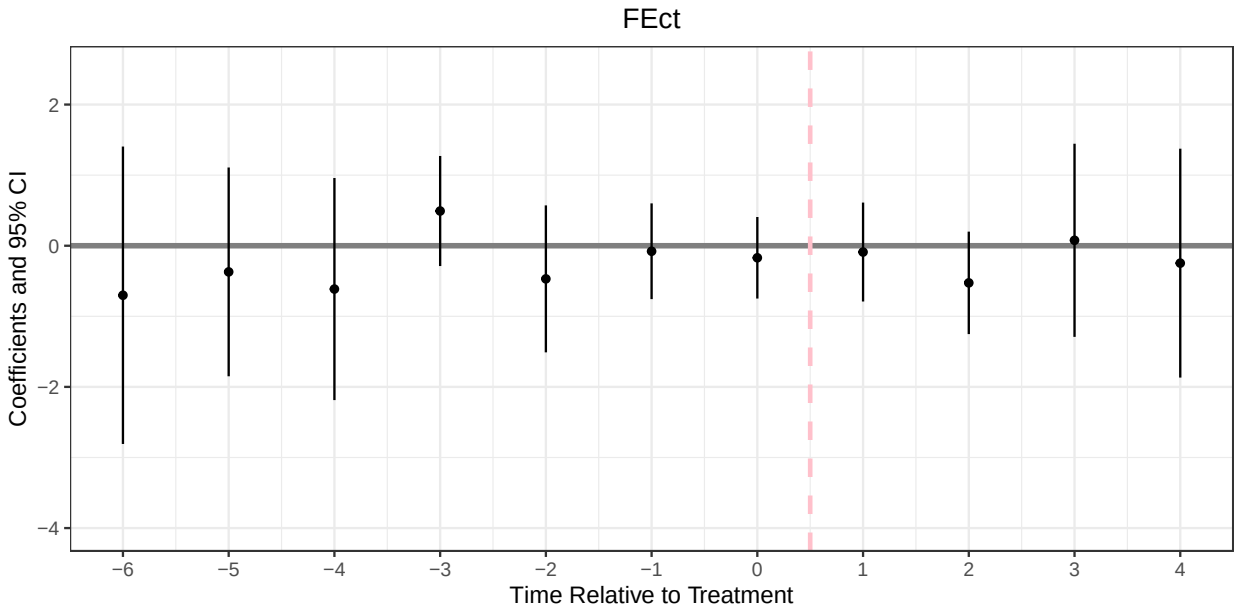
```
## OLS estimation, Dep. Var.: dv
## Observations: 3,186
## Fixed-effects: country: 177,  year: 18
## Standard-errors: Clustered (country)
##             Estimate Std. Error  t value Pr(>|t|)
## ucdp_lead -0.653507   0.321751 -2.03109  0.04375 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 3.54449     Adj. R2: 0.779161
##                 Within R2: 0.00209
```



Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects and obtaining corresponding event study plots. We use two estimators, TWFE and FEct. The results are shown below.

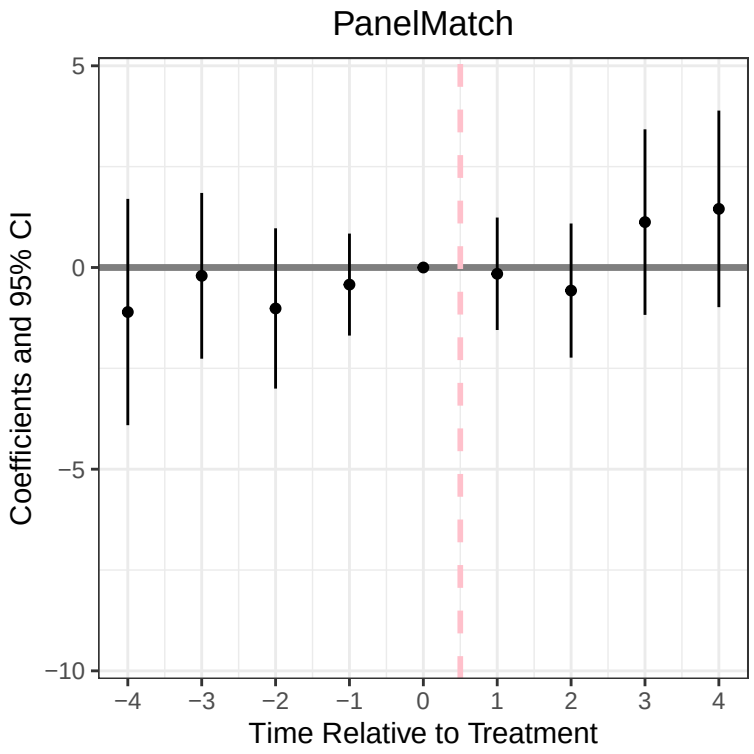




TWFE and FEct are broadly consistent with each other although the study appears to be under-powered. The estimated DTE are close to 0 on post-treatment periods in the plot.

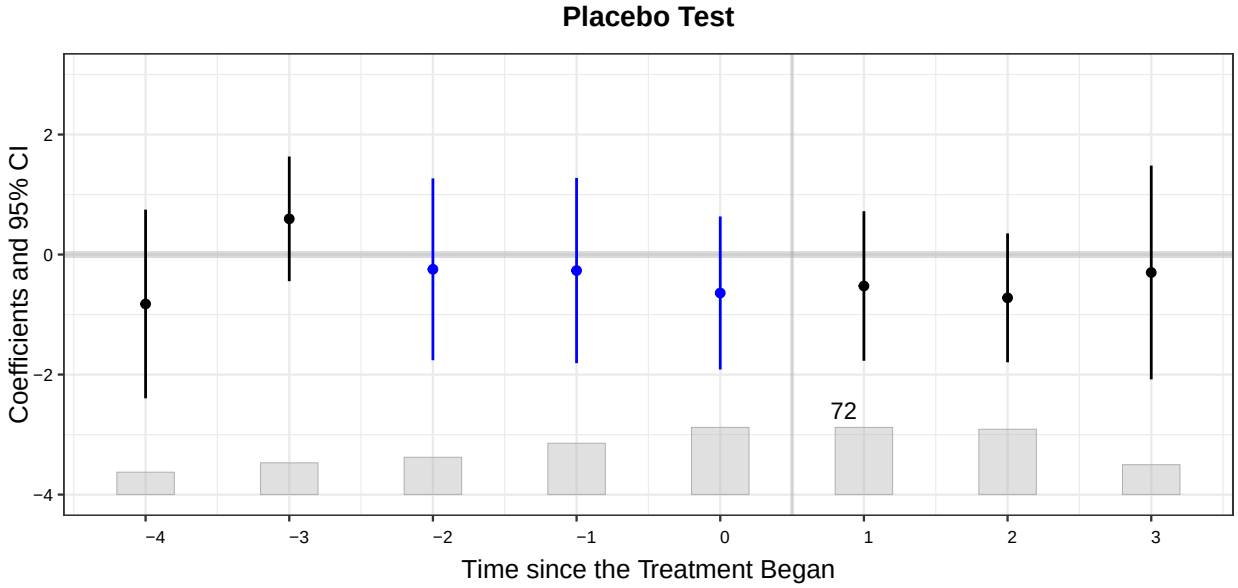
DTE of a Balanced Subsample

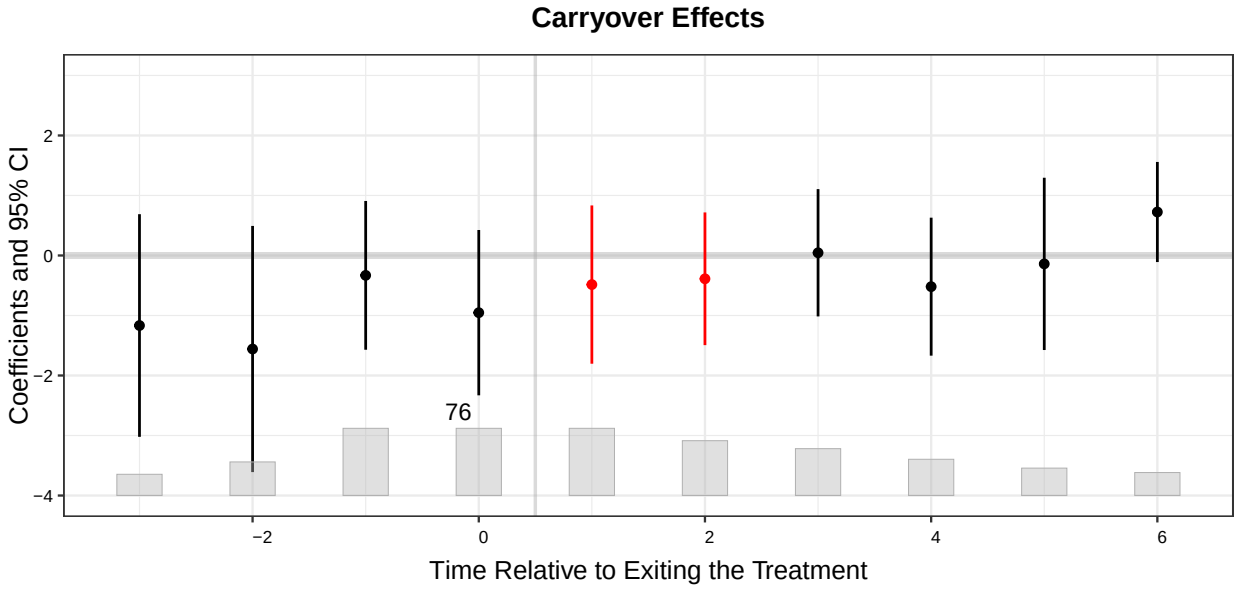
We also estimate the DTE using PanelMatch (lead = 4 and lag = 5). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.



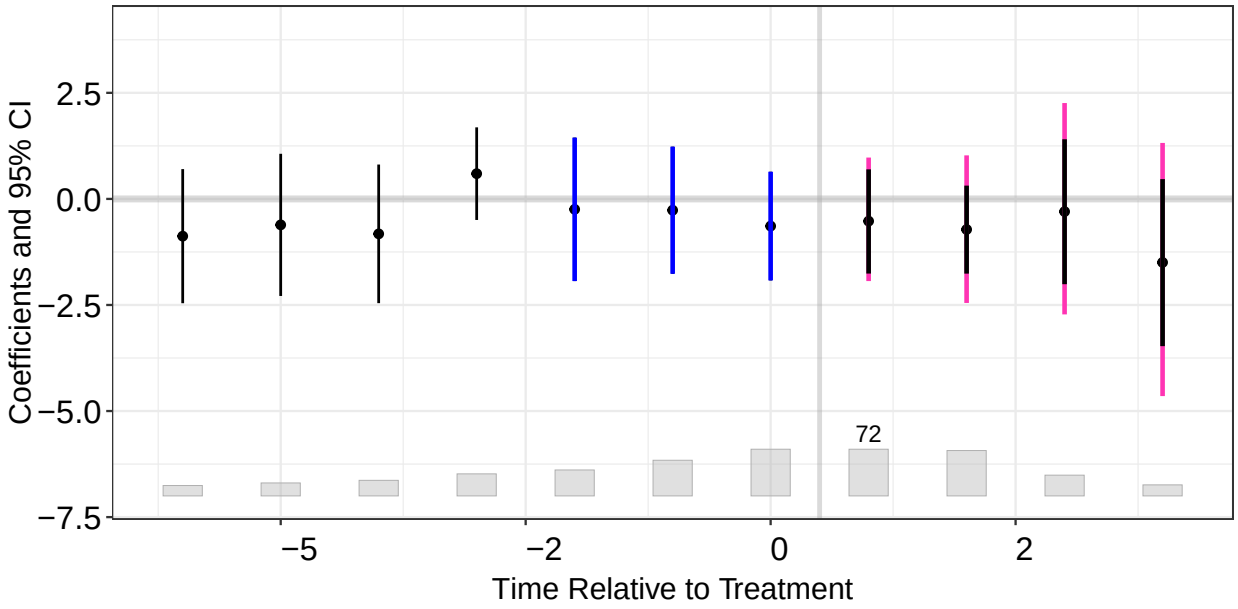


Test Statistics

```
##                                    p-value
## F test                               0.500
## Equivalence test (default)           0.060
## Equivalence test (threshold=ATT)     0.490
## Placebo test                         0.518
## Carryover effect test                0.438
```

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

FEct produces a point estimate similar to it of the TWFE model; however, this estimate is no longer statistically significant. There is no strong evidence of potential PT violations, anticipation effects or carryover effects. However, due to limited power, we cannot reject the null hypothesis that the ATT differs from zero under mild PT violations benchmarked against the pre-treatment placebo estimates.

Bokobza et al. (2022)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

A First Look at Data

The paper investigates the effects of failed coups on the purge of cabinet ministers in autocracies, using country-year level panel data between 1967 and 2016. One of the main findings of this paper (H1) is that "the yearly replacement rate in autocracies increases by 10 percentage points following a failed coup attempt" (p1447).

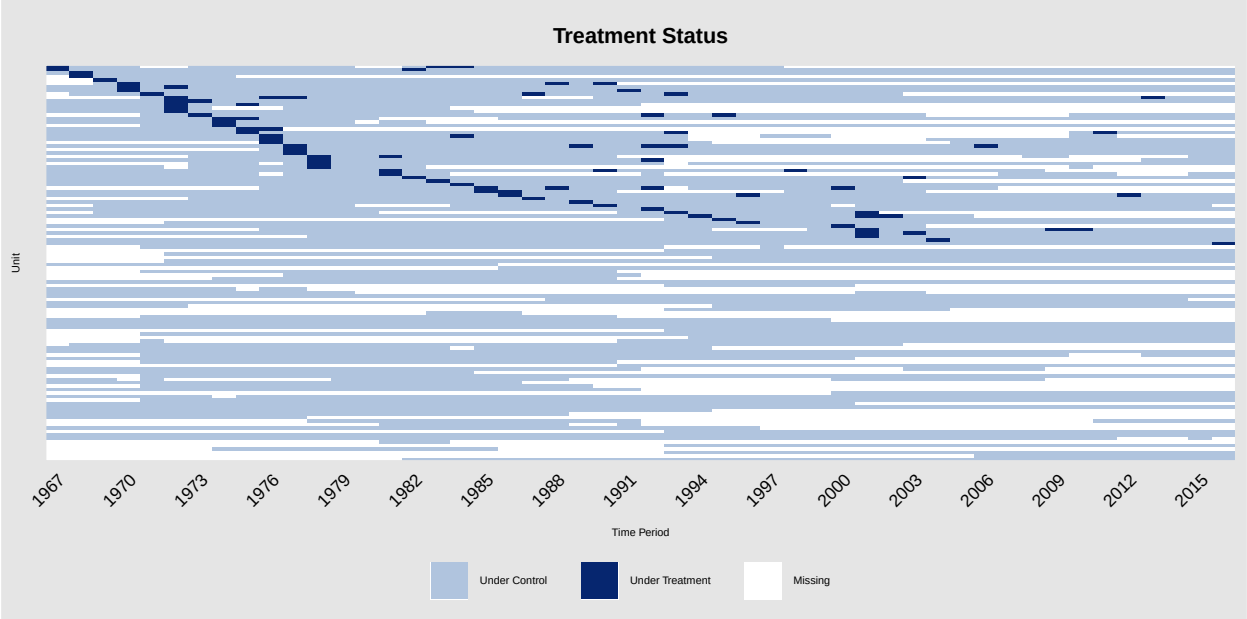
Model. We focus on the "main model" of **Figure 2** in the paper (model 2 in Table A1 of the Appendix), which contains the highest point estimate and t-statistics among all models testing H1. In this "main model", The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the country level. In the Appendix of this paper, it should be noted that authors also use a lagged-dependent-variable model for robustness check.[1]

Table 1: Replication Summary

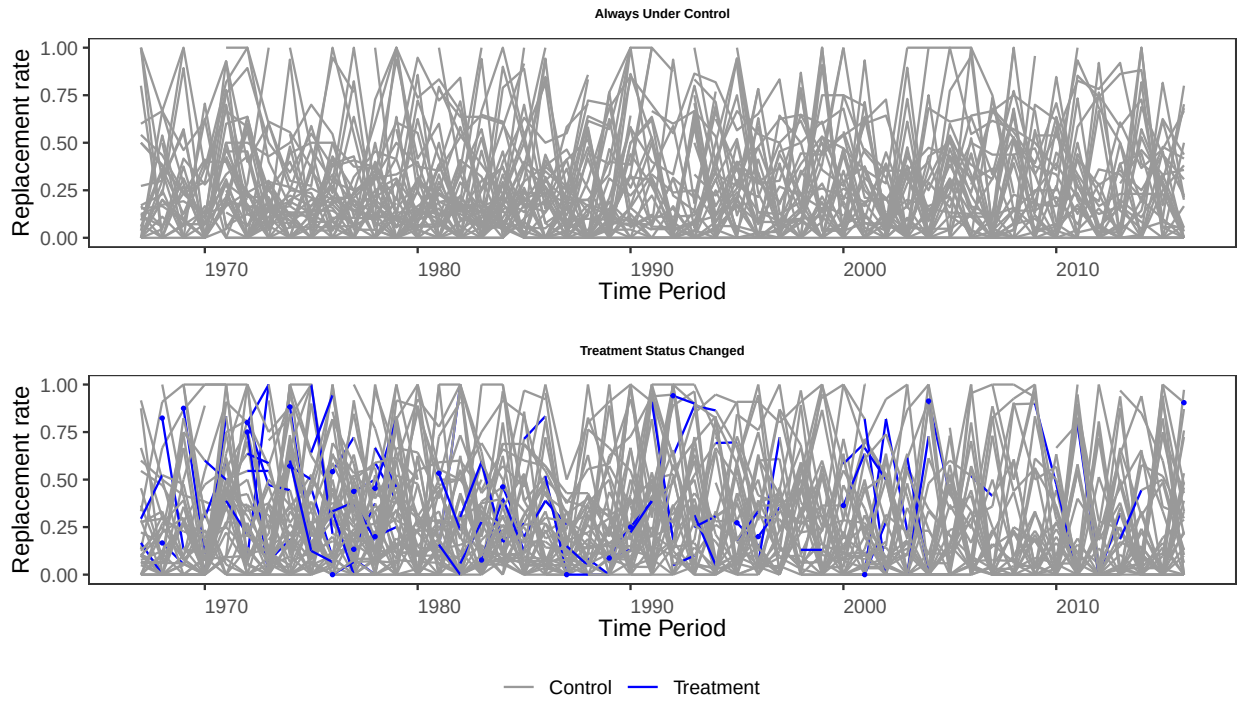
Unit of analysis	Country × year
Treatment	Failed coup attempt
Outcome	Replacement rate of cabinet ministers
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a country). There is some missingness.

[1] We recognize that this article tests six different hypotheses. We focus on the first hypothesis for two reasons. First, the first hypothesis is tested using the country-level TSCS data with a TWFE model, which is consistent with our replication criteria. Second, while the remaining five hypotheses tested mainly through individual-level TSCS data are important, the equation (p.1445) used to test it is an interaction model, which is beyond the scope of our analysis.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each country. The observations under treated status are marked in blue.



Point Estimates

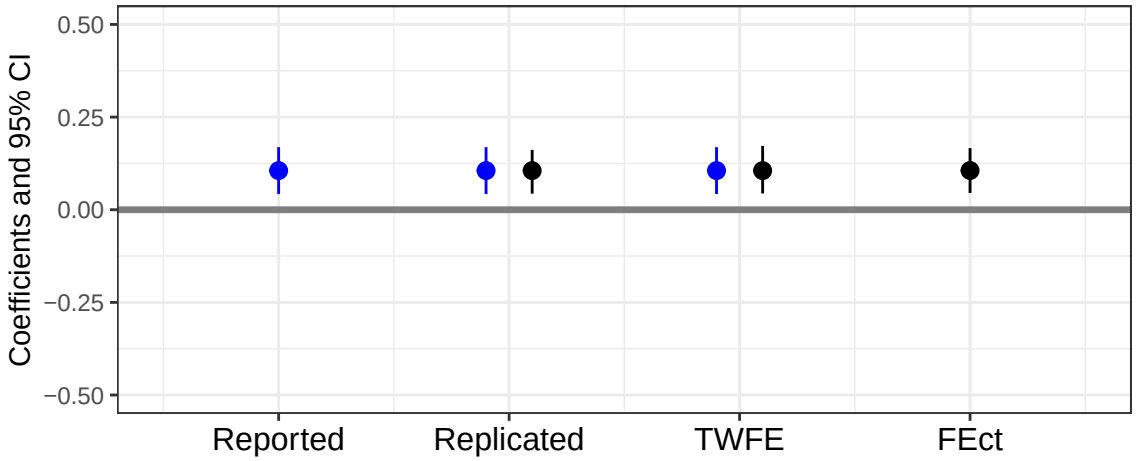
We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and imputation estimator (FEct).

The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

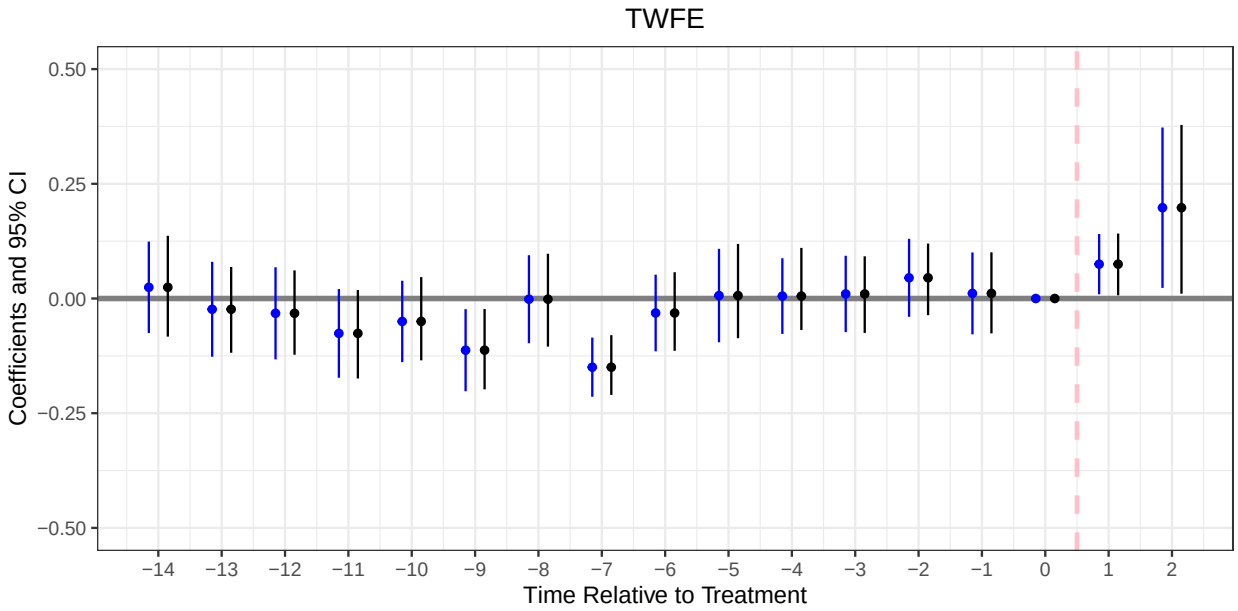
```
sol <- feols(replacement_rateadj_minister~coupattempt_whogov +
               gdp_cap_pwt_ln_c  + pop_pwt_ln_c + military_c_lag + monarchy_c_lag +
               party_c_lag|COUNTRY_ID+YEAR_DATA,
             data = df,cluster = "COUNTRY_ID")
summary(sol)
```

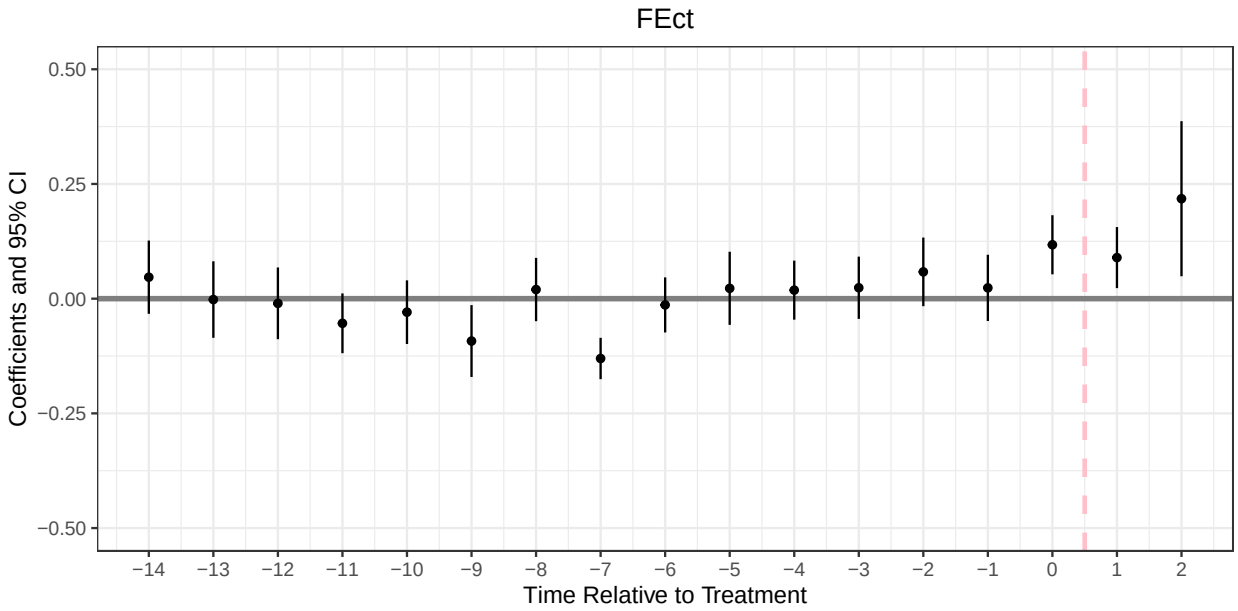
```
## OLS estimation, Dep. Var.: replacement_rateadj_minister
## Observations: 3,676
## Fixed-effects: COUNTRY_ID: 114,  YEAR_DATA: 50
## Standard-errors: Clustered (COUNTRY_ID)
##                      Estimate Std. Error  t value  Pr(>|t|)
## coupattempt_whogov  0.105549   0.032266  3.27121 0.0014203 **
## gdp_cap_pwt_ln_c   -0.033091   0.020522 -1.61249 0.1096454
## pop_pwt_ln_c       -0.120178   0.066496 -1.80730 0.0733763 .
## military_c_lag     -0.033072   0.029281 -1.12946 0.2610945
## monarchy_c_lag      0.030770   0.021088  1.45911 0.1473083
## party_c_lag        -0.063919   0.032053 -1.99417 0.0485411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.252742     Adj. R2: 0.191561
##                  Within R2: 0.008979
```



Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.

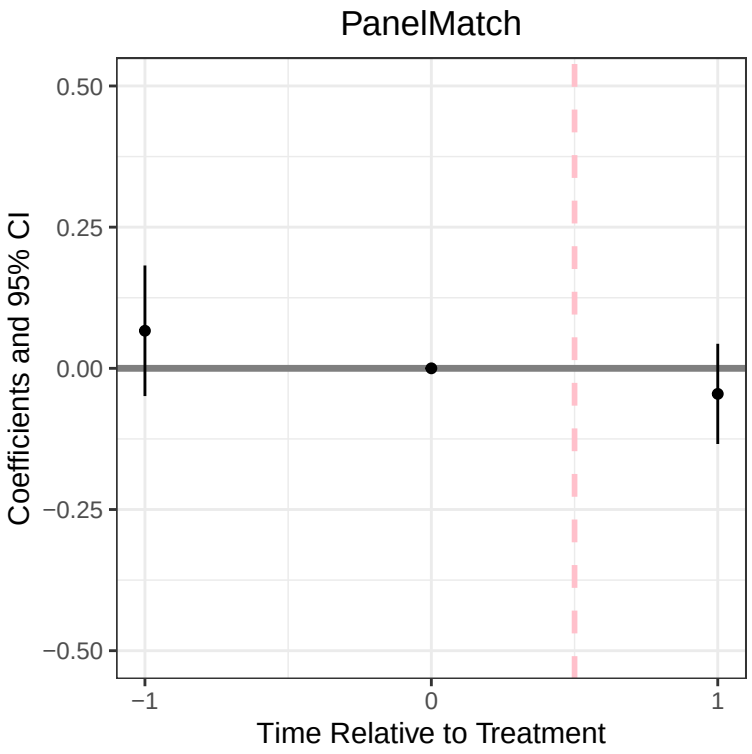




TWFE and FEct are broadly consistent with each other. The estimated DTE are positive on all two post-treatment periods in the plot.

DTE of a Balanced Subsample

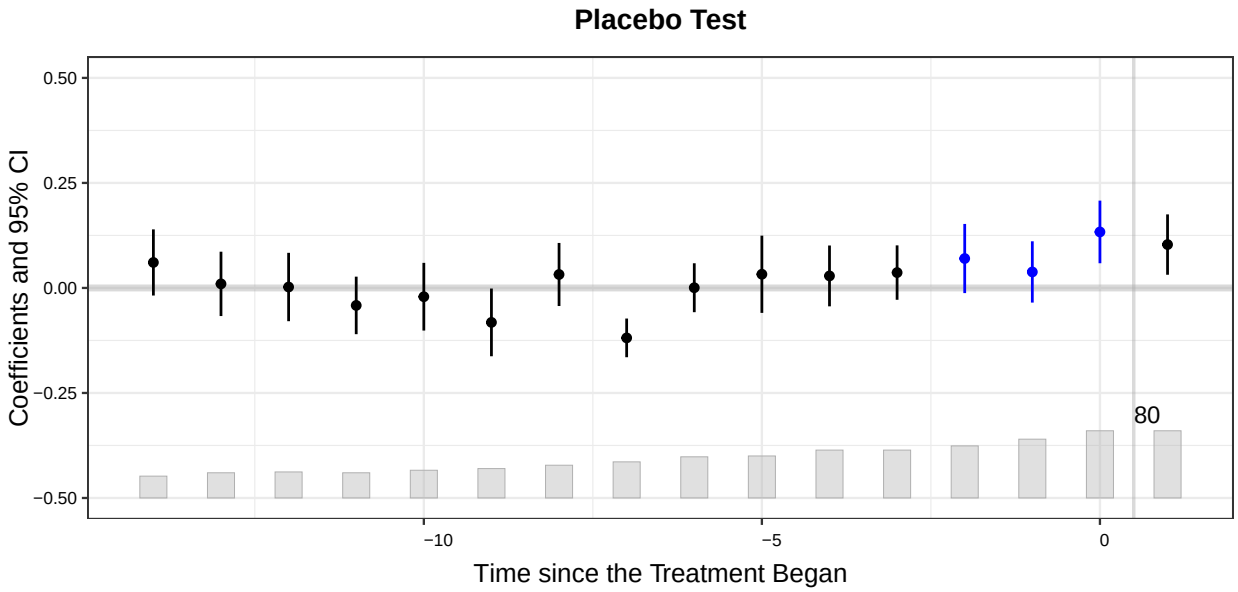
We also estimate the DTE using PanelMatch (lead = 1 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.

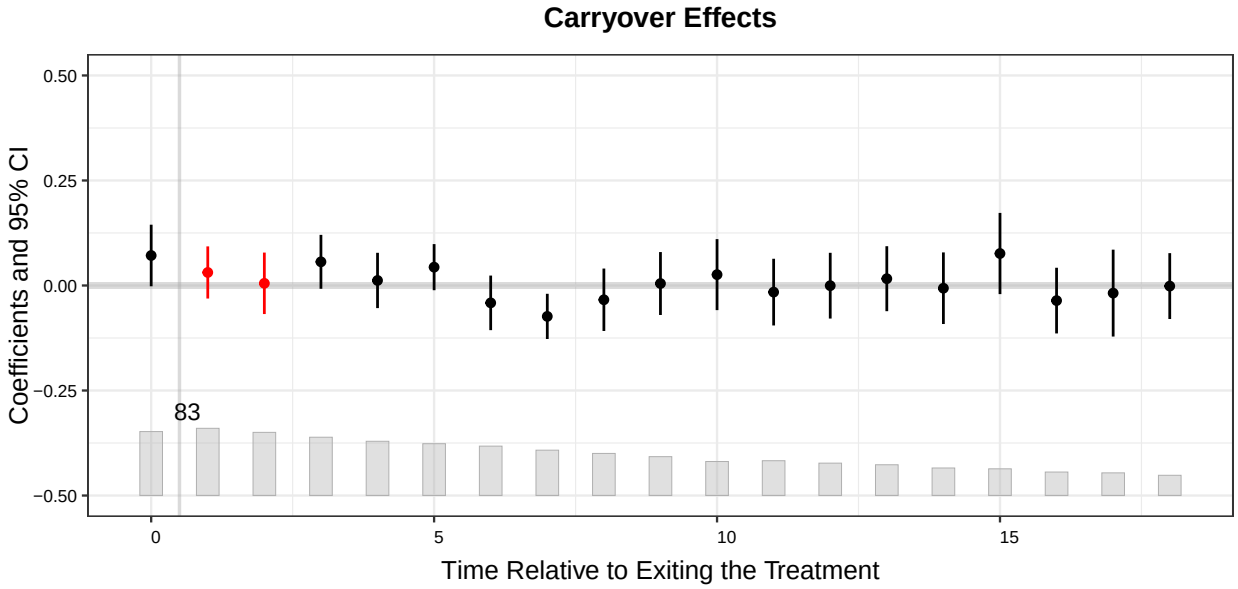


The PanelMatch estimate is different from the FEct estimate on a balanced subsample.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





Test Statistics

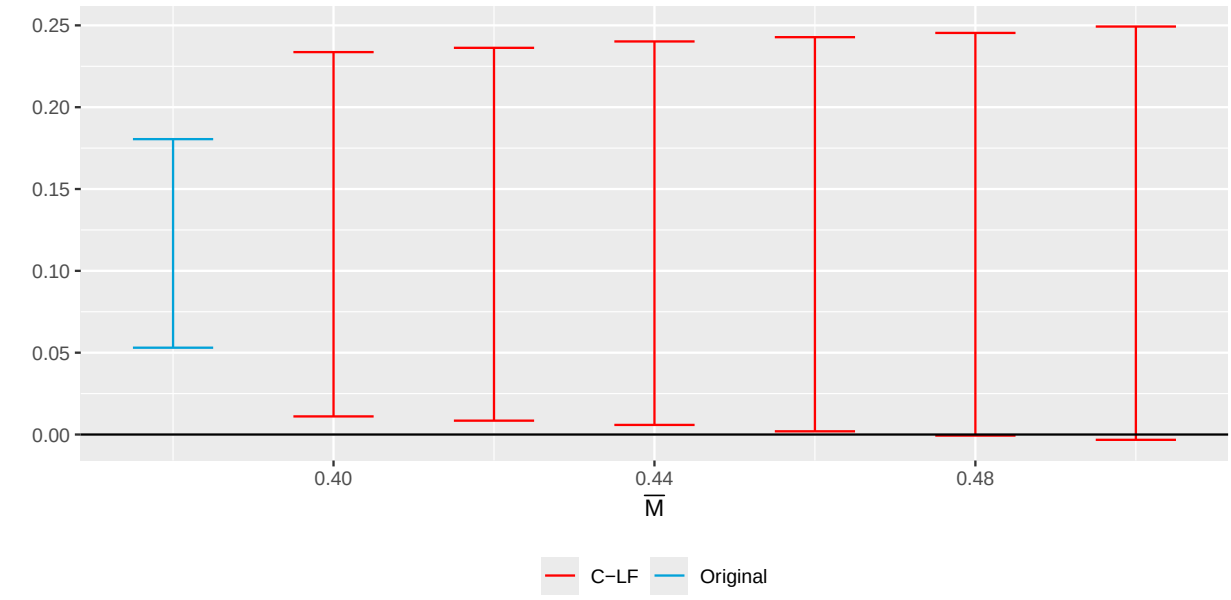
```
##                                    p-value
## F test                               0.000
## Equivalence test (default)           0.949
## Equivalence test (threshold=ATT)     0.858
## Placebo test                         0.002
## Carryover effect test                0.491
```

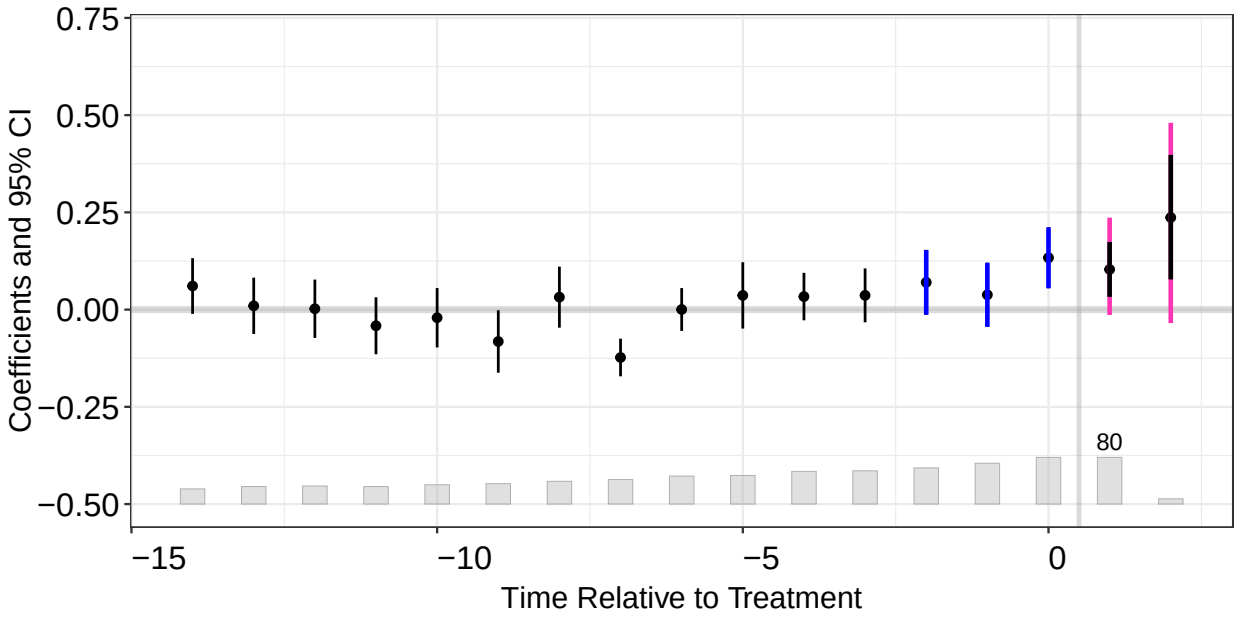
We find some evidence for a potential violation of parallel trend assumption and no-anticipation assumption with the F test and the placebo test. The equivalence test also fails to reject the null that the residual averages in pre-treatment periods exceed the estimated ATT possibly due to limited power. We find no evidence for a potential violation of the no-carryover-effect assumption.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.48$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.48 times the maximum deviation observed between consecutive periods in the placebo periods.

We further evaluate the robustness of the estimated DTE under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, accounting for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model remain statistically significantly at the 5% level when the imputation estimator is applied. We find some evidence of potential PT violations or anticipation effects, as indicated by the F test and placebo test. However, the estimated ATT appears to be robust to mild PT violations in the sensitivity analysis, with breakdown threshold $M = 0.48$.

Caughey, Warshaw, and Xu (2017)

Contents

A First Look at Data 1

Point Estimates 2

Dynamic Treatment Effects 3

Diagnostic Tests 5

Sensitivity Analysis 6

Summary 7

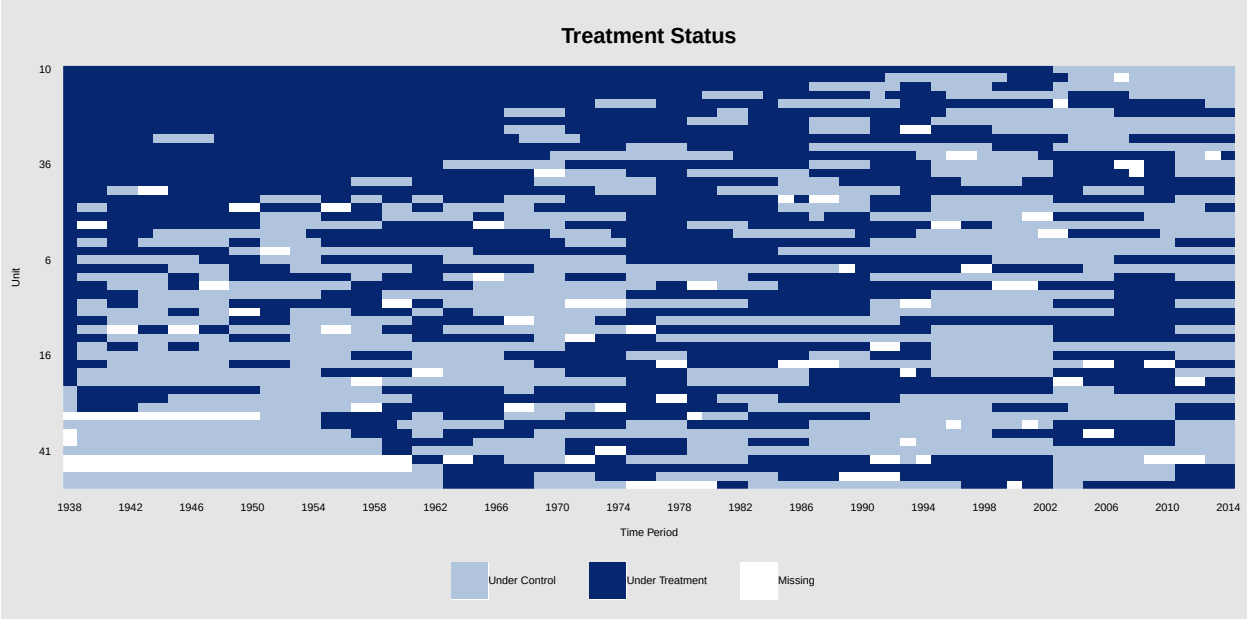
A First Look at Data

This paper investigates the effects of electing a democratic governor on policy liberalism, using US state-year level panel data, between 1936 and 2014. One of the main findings of this paper is that "throughout the 1936-2014 period, electing Democrats has led to more liberal policies" (p1342).

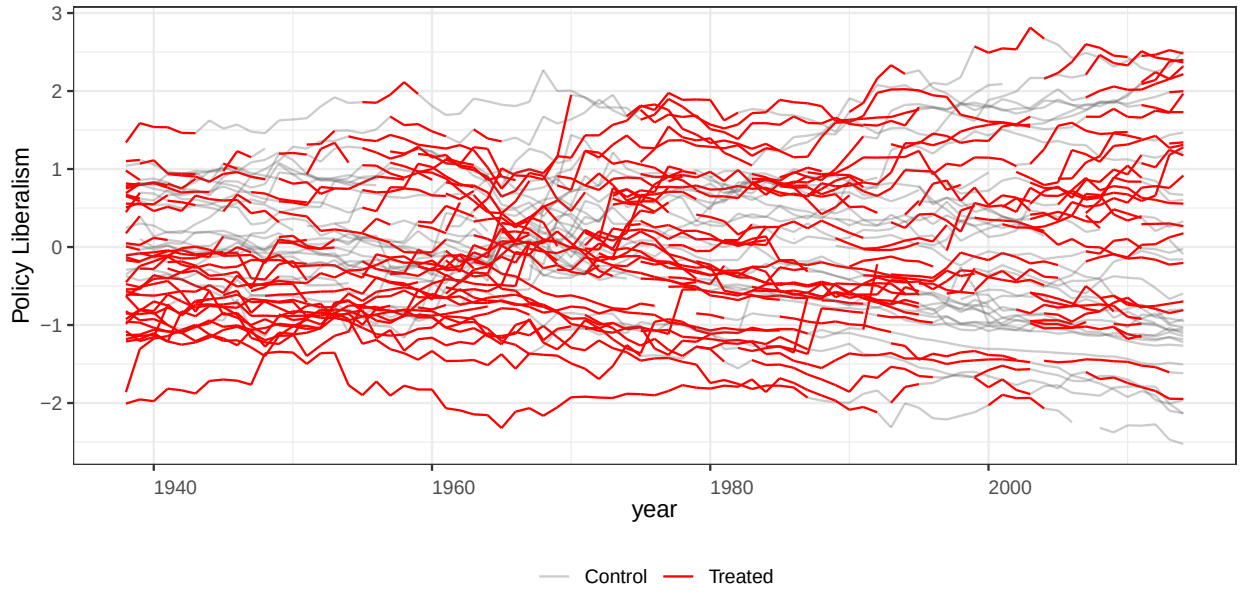
Model. We focus on **Model 2 of Table 2** in the paper, which includes a lagged dependent variable (LDV) with 1 and 2 lagged terms. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the state level.

Replication Summary	
Unit of analysis	State × year
Treatment	Electing a democratic governor
Outcome	Policy Liberalism
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a state). There are treatment reversals and some data missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each state. The observations under treated status are marked in blue.



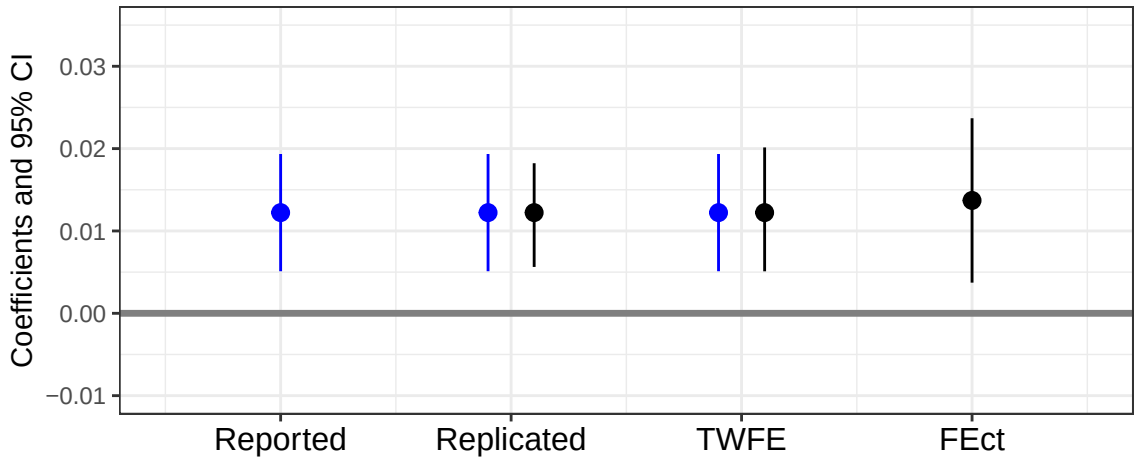
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and imputation estimator (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

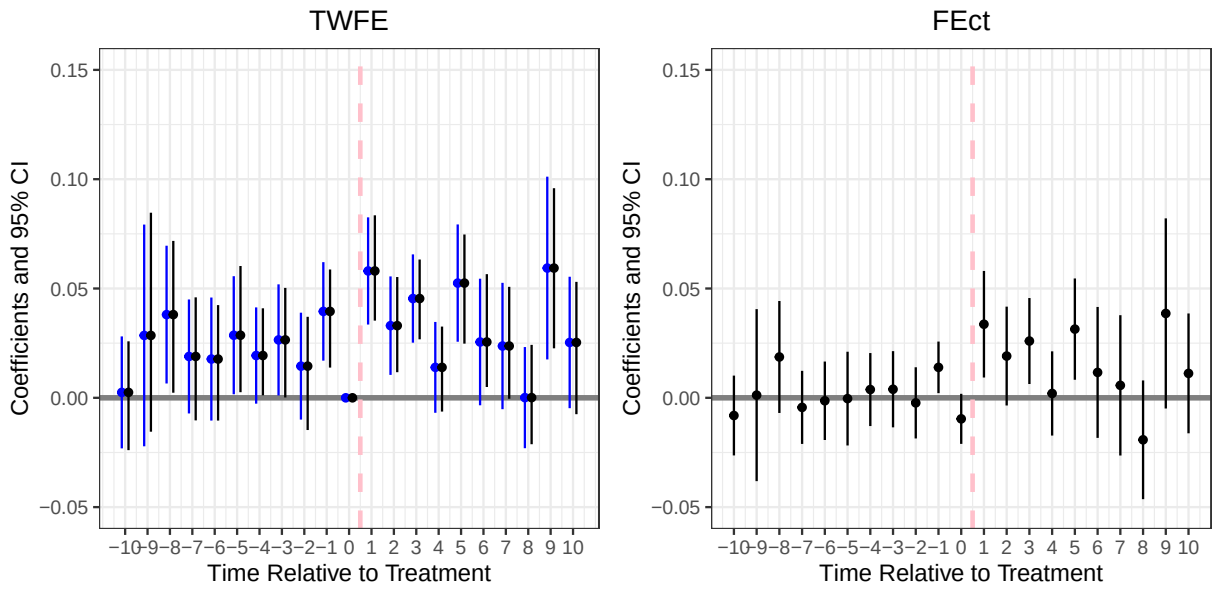
```r
sol <- feols(Policy~gov_dem + hs_dem_control + sen_dem_control + Policy_L1 + Policy_L2|state+year,
             data = df,cluster = "state")
summary(sol)
```

```
## OLS estimation, Dep. Var.: Policy
## Observations: 3,586
## Fixed-effects: state: 49,  year: 77
## Standard-errors: Clustered (state)
##                  Estimate Std. Error  t value   Pr(>|t|)
## gov_dem          0.012227   0.003632  3.36678 1.5050e-03 **
## hs_dem_control   0.030349   0.005886  5.15617 4.7266e-06 ***
## sen_dem_control 0.020818   0.005577  3.73304 5.0129e-04 ***
## Policy_L1        0.866231   0.016314 53.09646  < 2.2e-16 ***
## Policy_L2        0.087902   0.015812  5.55926 1.1728e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.109298     Adj. R2: 0.987303
##                   Within R2: 0.917128
```



Dynamic Treatment Effects

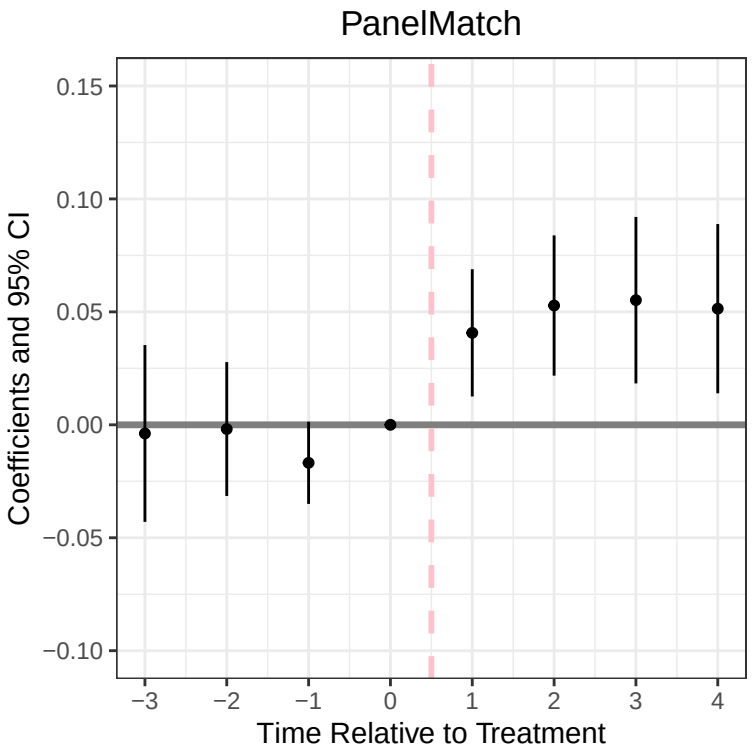
We then move onto estimating dynamic treatment effects and obtaining corresponding event study plots. We use two estimators, TWFE and FEct. The results are shown below.



The shape of the estimated DTEs using TWFE and FEct are similar. However, the estimated DTE using TWFE consistently appear larger. This difference might be attributed to our selection of the last pre-treatment period (period 0) as the reference period for TWFE.

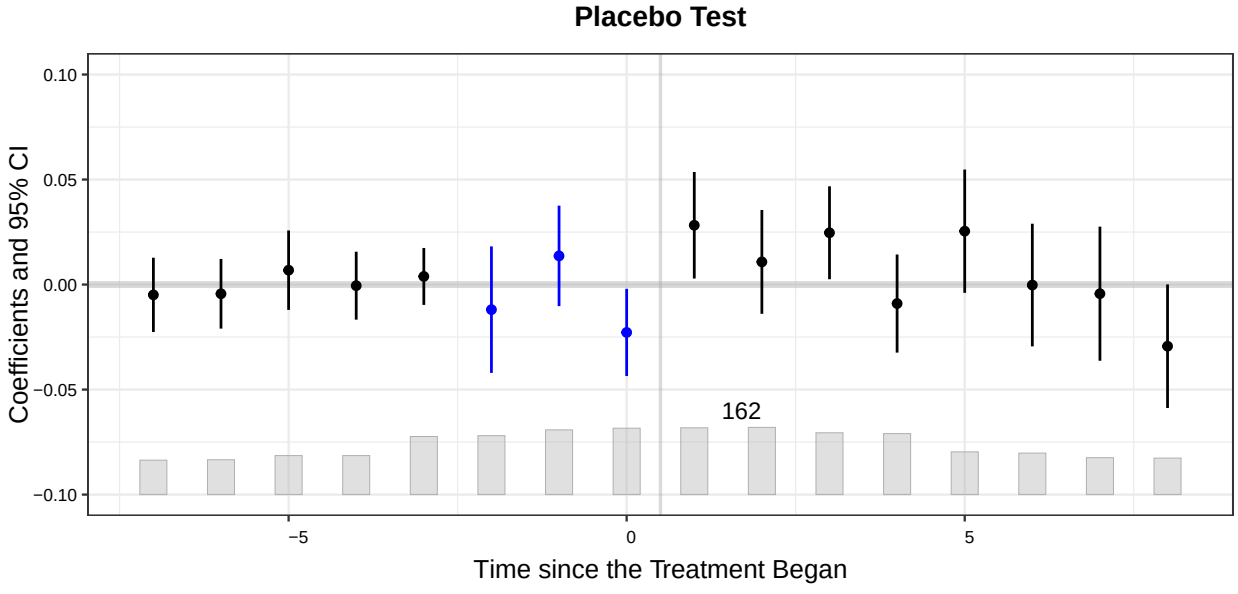
DTE of a Balanced Subsample

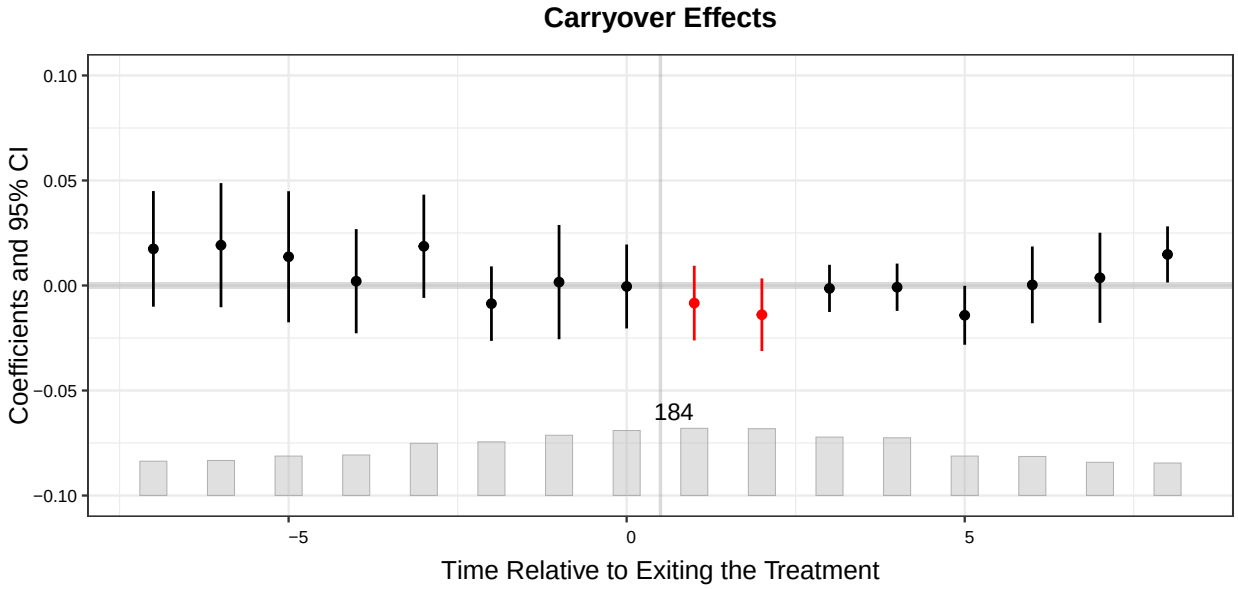
We also estimate the DTE using PanelMatch (lead = 4 and lag = 4). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





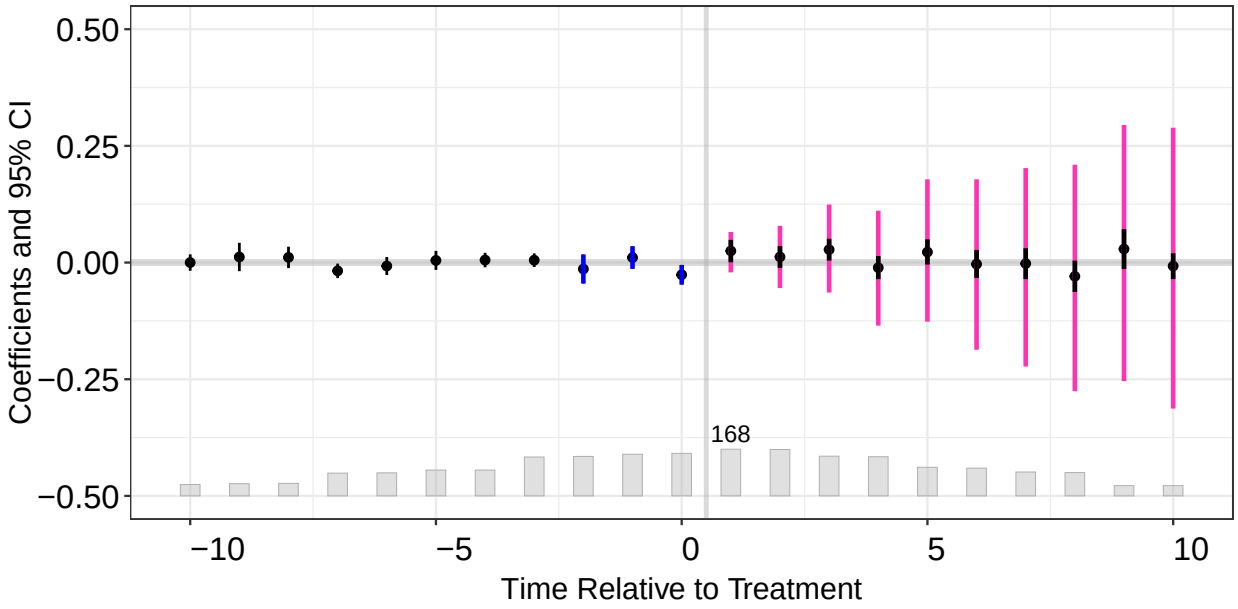
Test Statistics

```
##                                    p-value
## F test                               0.043
## Equivalence test (default)           0.000
## Equivalence test (threshold=ATT)     0.515
## Placebo test                         0.443
## Carryover effect test                0.090
```

We find some evidence for violations of the parallel trend (PT) assumption and the no-carryover-effect assumption. The equivalence test also fails to reject the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, the main result of the chosen model seems to be robust to the HTE-robust estimator. We find some evidence of violations to the identifying assumptions using diagnostic tests. The significant ATT is not robust to the removal of placebo periods and under mild PT violations benchmarkded against the pre-treatment placebo estimates, possibly due to limited power. It is worth noting that the estimated DTE exhibit a drop right before the onset of the treatment, which may indicate the presence of some anticipation effect.

Christensen and Garfias (2021)

Contents

Summary . . . 1
Point Estimates . . . 2
Dynamic Treatment Effects . . . 4
Diagnostic Tests . . . 7
Sensitivity Analysis . . . 8
Summary of Findings . . . 8
Subsample Summary . . . 9
Subsample Point Estimates . . . 10
Subsample Dynamic Treatment Effects . . . 11
Sensitivity Analysis . . . 15
Summary of findings . . . 15

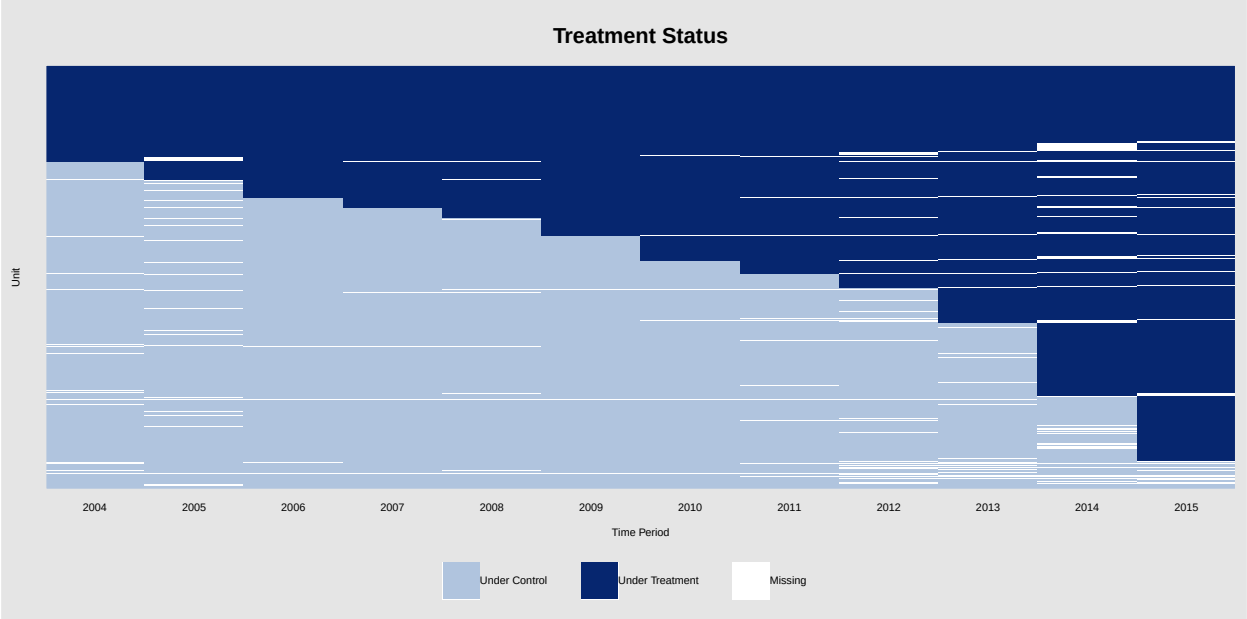
Summary

This paper compares changes in property tax revenue in municipalities that update their cadastre, relative to the change in municipalities that do not. The paper finds property tax revenues rise by over 10% in municipalities that update their cadastre.

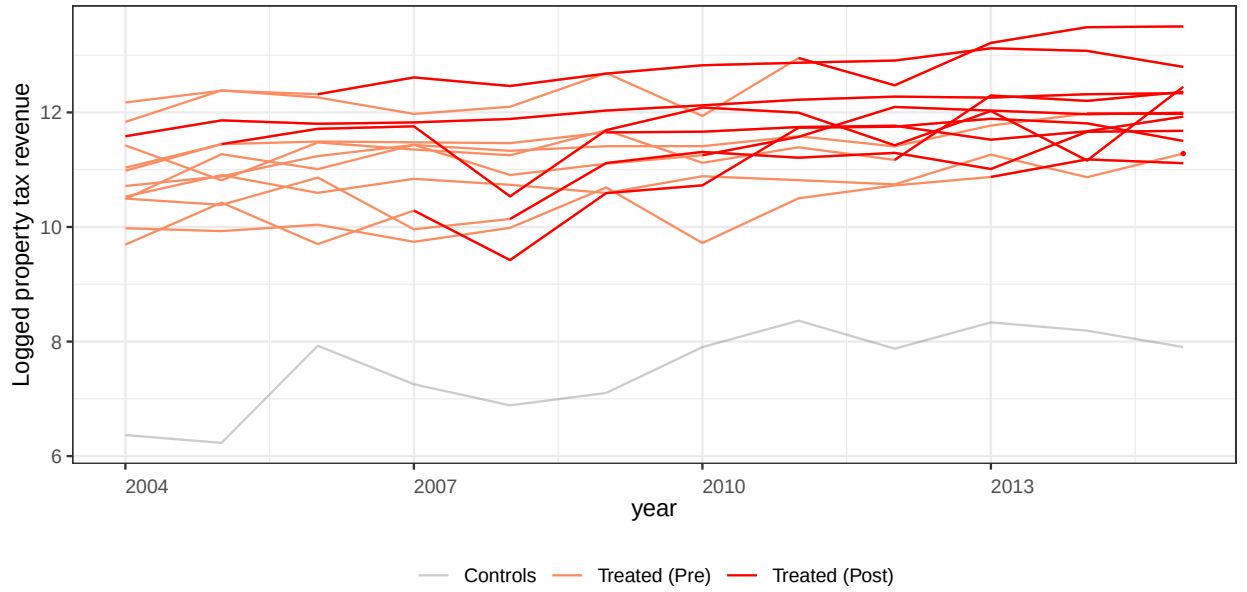
Model. We focus on **Colum 1 of Table 2** in the paper. The basic specification is $\log(\text{IPTU}_{it}) = \beta \text{ Cadastre Update }_{it} + \lambda_t + \gamma_i + \varepsilon_{it}$. The outcome is the log of property tax revenues and the treatment is whether the cadastre is updated. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the municipality level.

Replication Summary	
Unit of analysis	Municipality
Treatment	Cadastre
Outcome	Logged property tax revenue
Treatment type	Staggered
Outcome type	continuous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a municipality). There exists a large ratio of always-treated units.



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

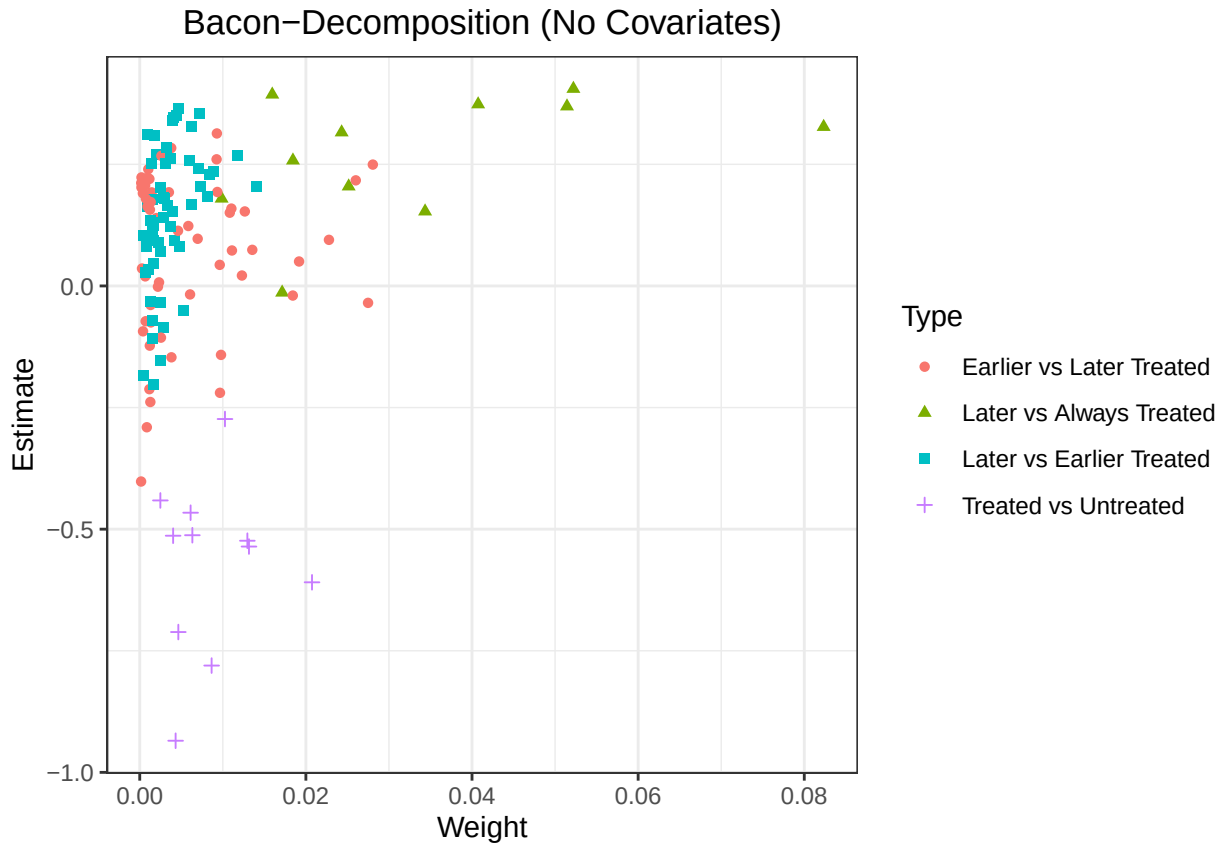
We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DIDM estimator, and imputation estimator (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

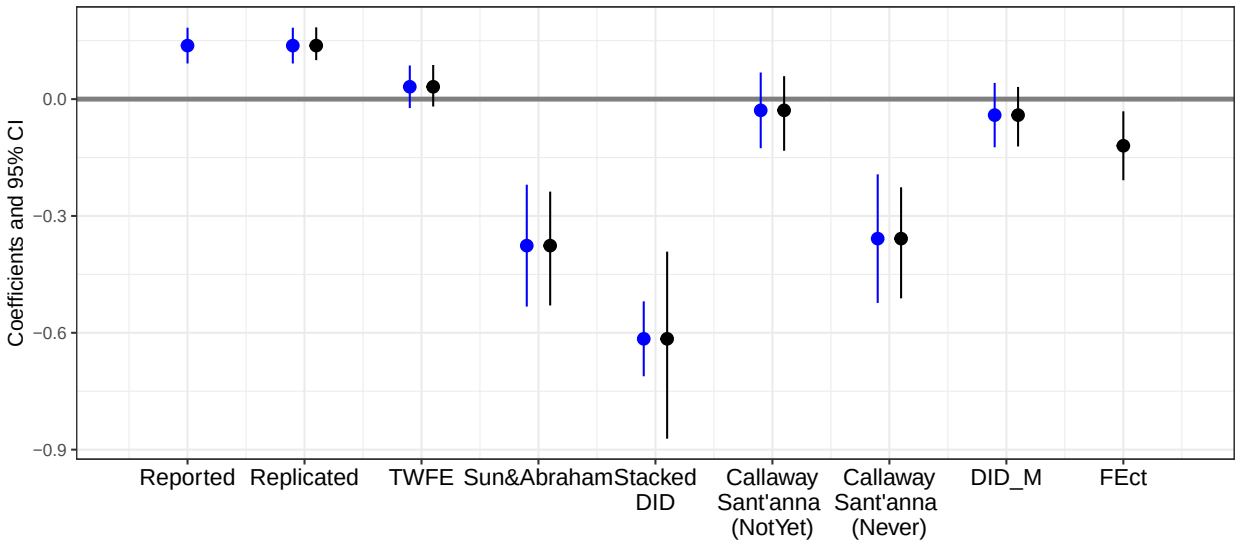
```
sol <- feols(logiptu~cad_update|c6_ibge+year,data = df,cluster = "c6_ibge")
summary(sol)
```

```
## OLS estimation, Dep. Var.: logiptu
## Observations: 62,161
## Fixed-effects: c6_ibge: 5,401,  year: 12
## Standard-errors: Clustered (c6_ibge)
##             Estimate Std. Error t value   Pr(>|t|)
## cad_update 0.137385   0.023447 5.85938 4.9209e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.20574     Adj. R2: 0.827612
##                 Within R2: 8.784e-4
```

Goodman-Bacon Decomposition In the Goodman-Bacon decomposition, green triangles represent the estimates of 2-by-2 DIDs that use the always-treated units as "controls," which are all positive and have large weights on the original estimated ATT.



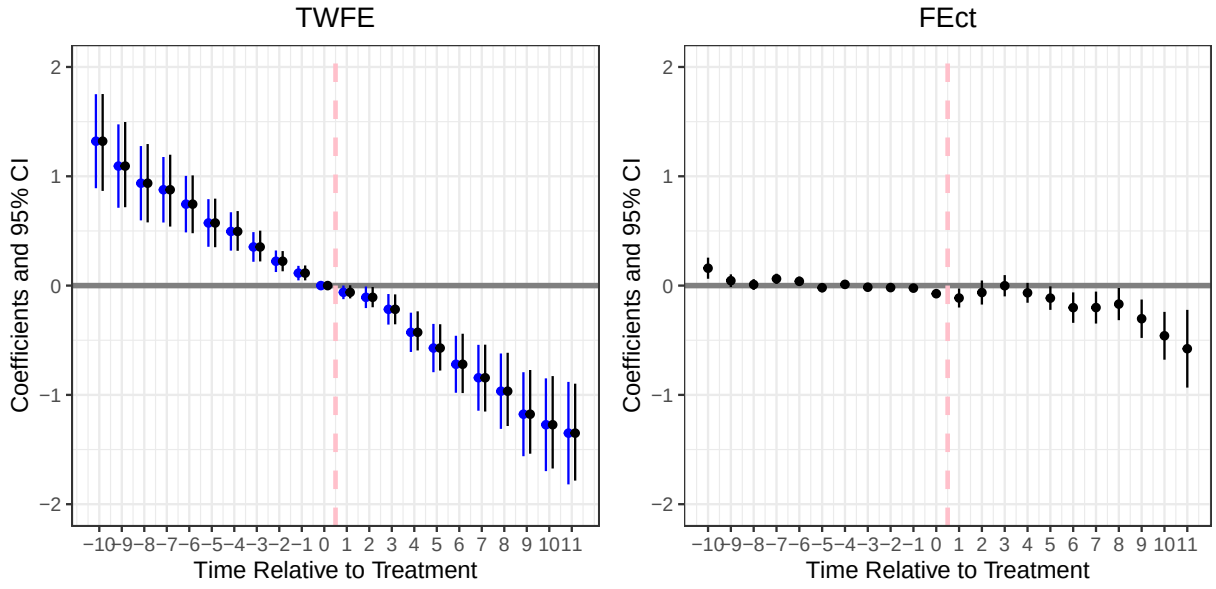

```
##   |                                                                      |
```

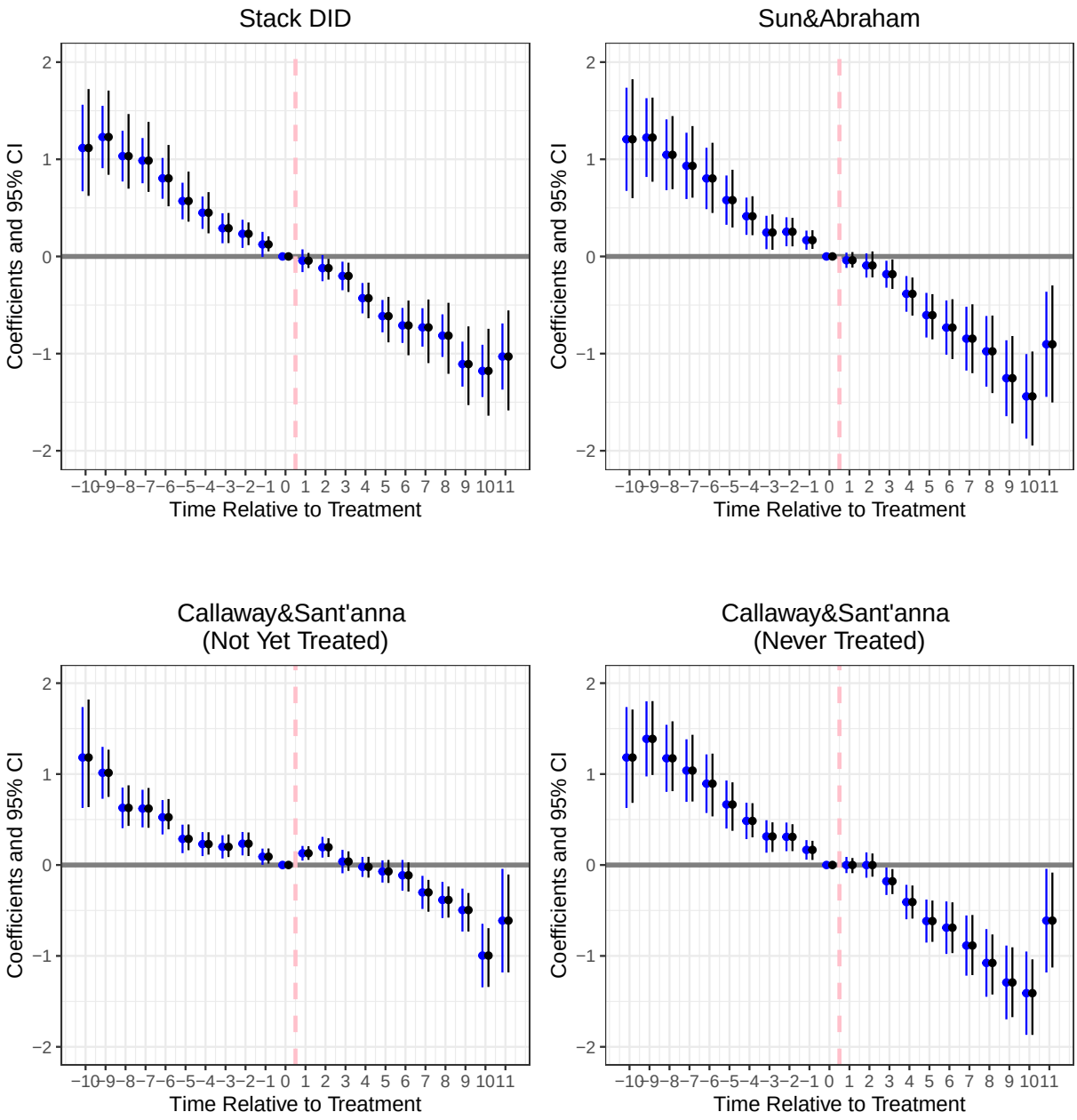


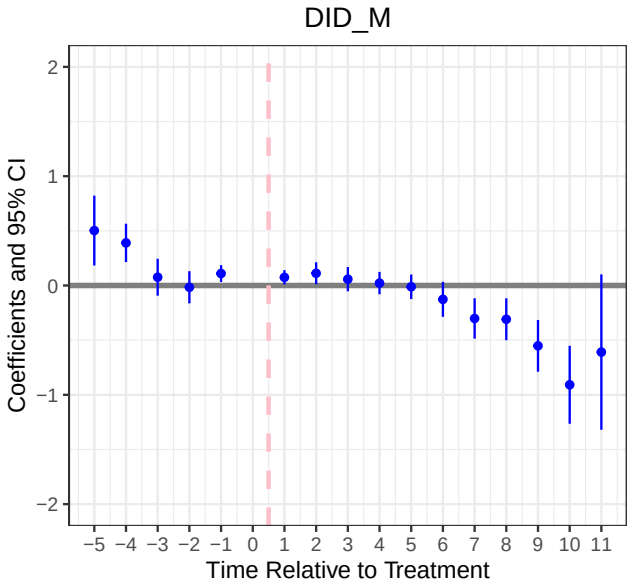
After excluding the always-treated units, the statistical significance of the TWFE estimate diminishes, though it remains positive. The Stacked DID, IW estimator, CS estimator, and FEct all produce negative results. Only the CS estimator utilizing not-yet treated units as the control group yields result that is not statistically significant. There are also notable differences in the magnitude of estimated ATT across these estimators.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining following event study plots.



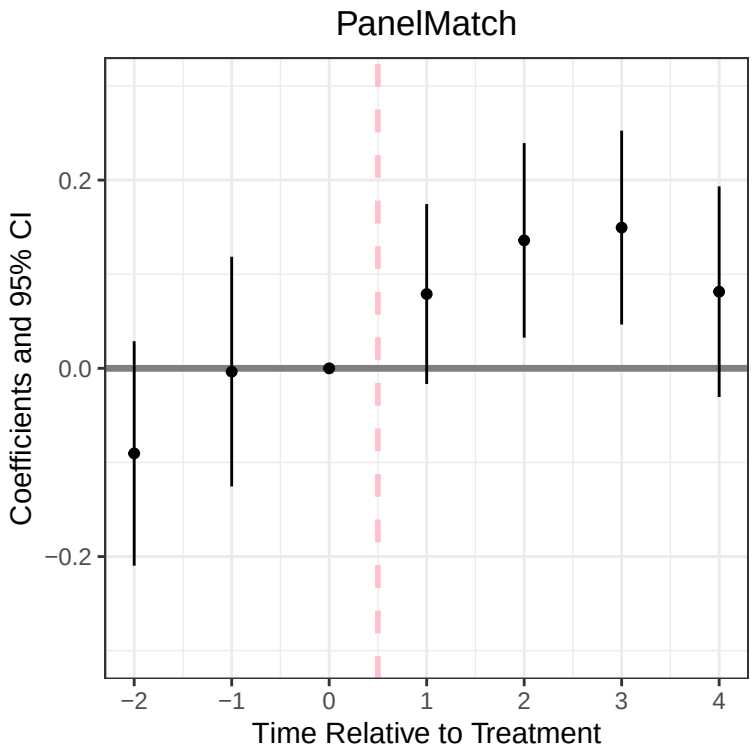

Stack DID
Sun&Abraham
Callaway&Sant'anna
(Not Yet Treated)
Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
Time Relative to Treatment
2
1
0
−1
−2
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10 11



With the exception of the DTE estimated using FEct, all other estimated DTEs display a significant downward trend. This indicates a substantial violation to the parallel trend (PT) assumption. Considering the average outcome figure above, this violation can be attributed to the notably different trends between the pure-control cohort and ever-treated cohorts.

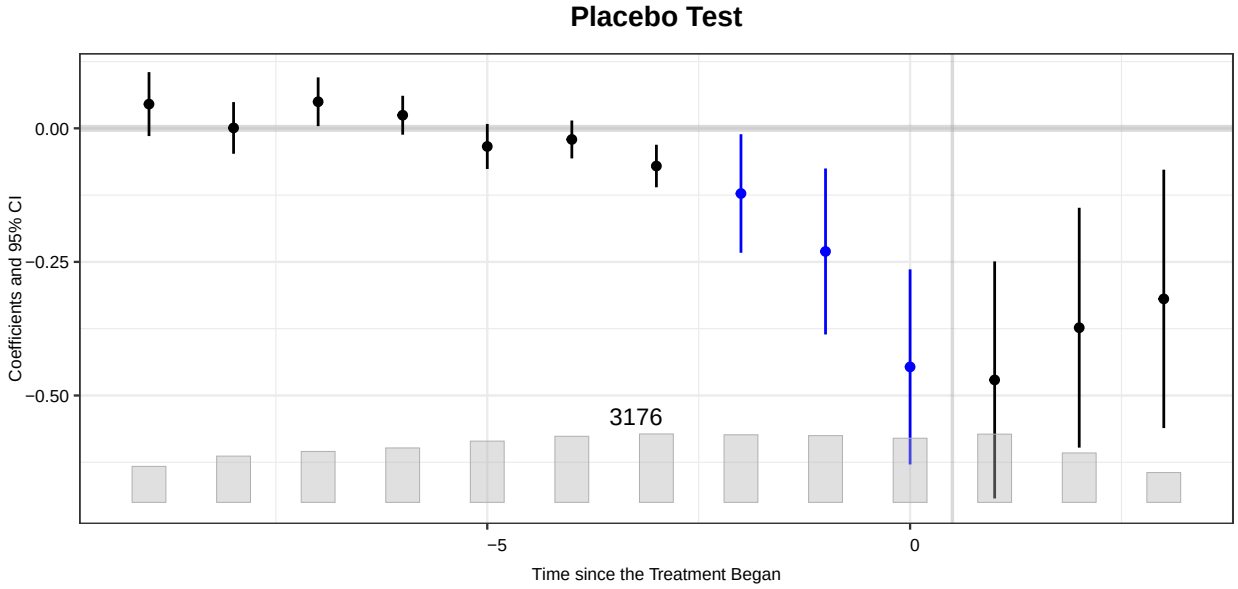
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 4 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



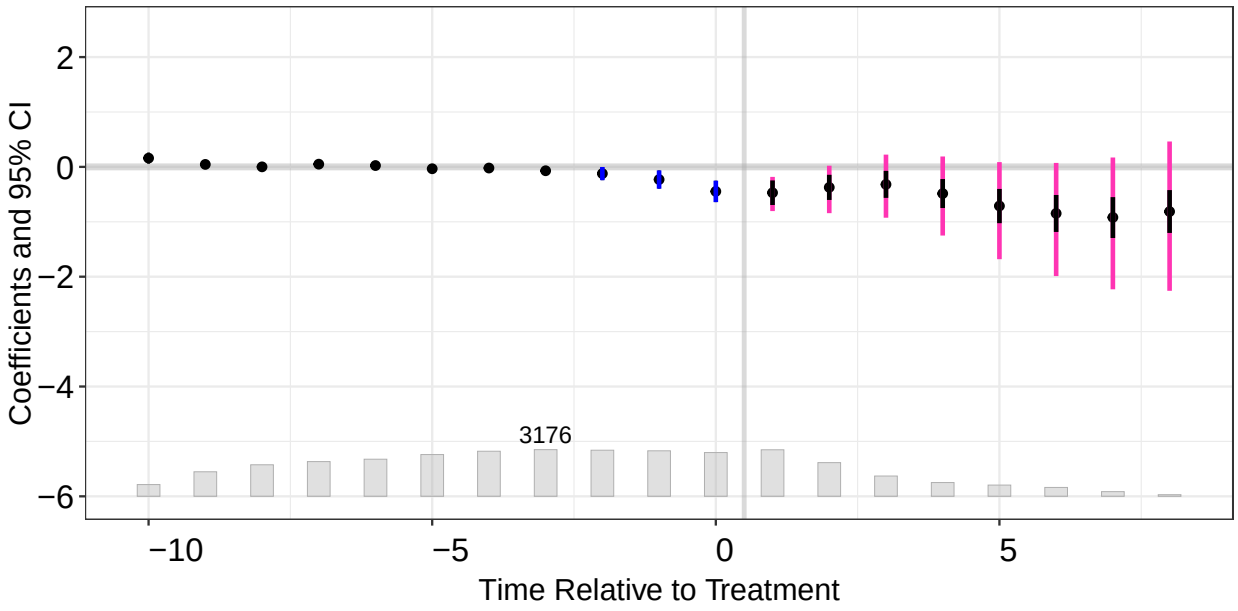
Test Results

```
##                                  p-value
## F test                             0.000
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.008
## Placebo test                       0.000
## Carryover effect test                 NA
```

We find evidence for a violation of the PT with the placebo test and the F-test for no pretrend. However, the equivalence test rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

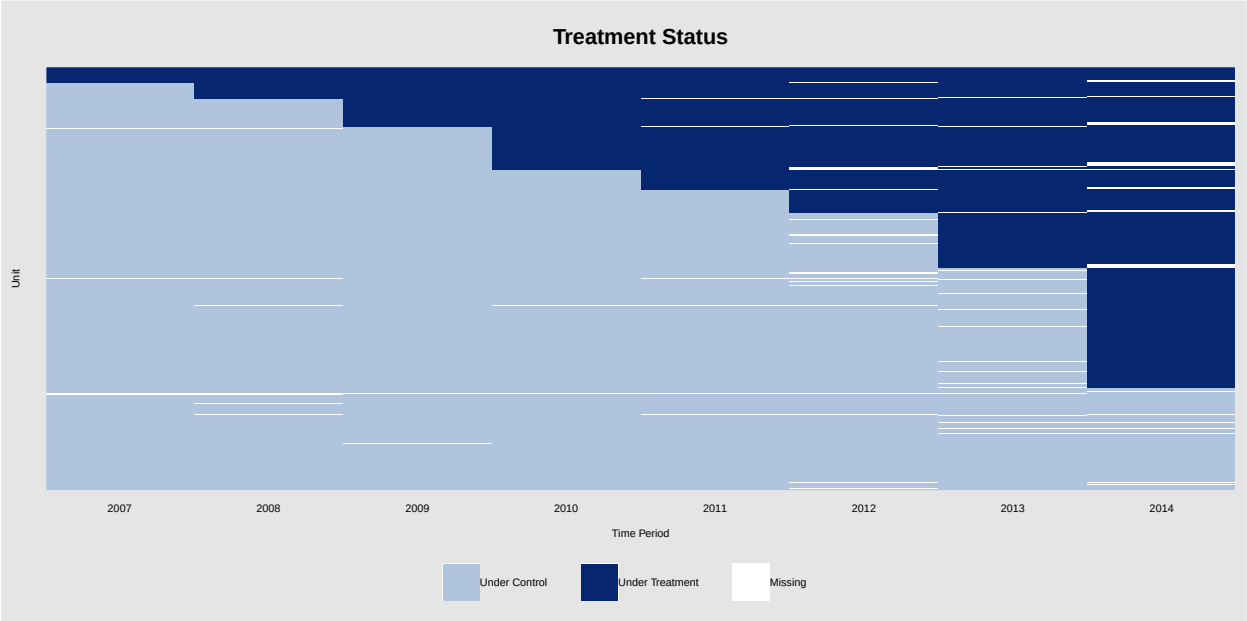


Summary of Findings

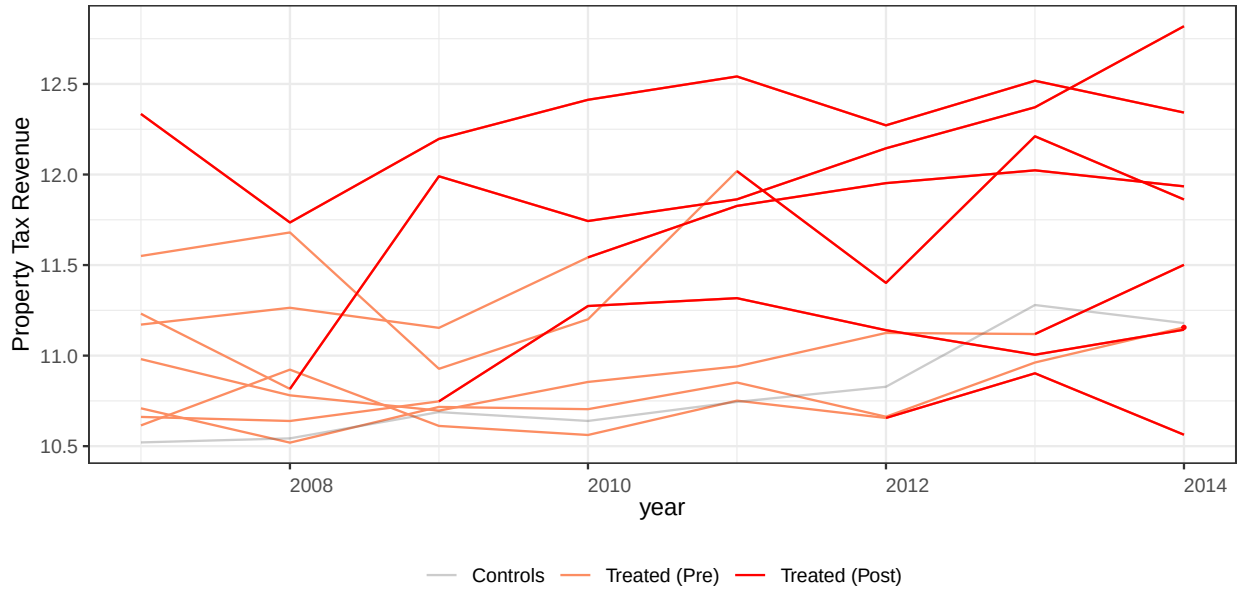
First, the presence of a substantial proportion of units that are always treated noticeably biases the estimated ATT upward under the original TWFE specification. Second, the primary finding of the chosen model does not hold when HTE-robust estimators are used, suggesting the initial conclusion's vulnerability to heterogeneity in treatment effects. Third, it is highly likely that the PT is violated due to significant divergence in outcome trends between the ever-treated cohorts and pure-control cohorts. The substantial disparity among various HTE-robust estimators can be attributed to their varying degrees of reliance on the control cohort within their estimation strategies.

Subsample Summary

We also examine the subsample used by the authors to generate the DTE/event-study plot Figure 1(B). This subsample differs from the original dataset in two key aspects. First, it excludes the always-treated units, focusing solely on the ever-treated cohorts. Second, it uses data between 2007 and 2014, designating the cohort treated in 2007 as the always-treated cohort and the cohort treated in 2015 as the control cohort within this subsample.



View the outcome



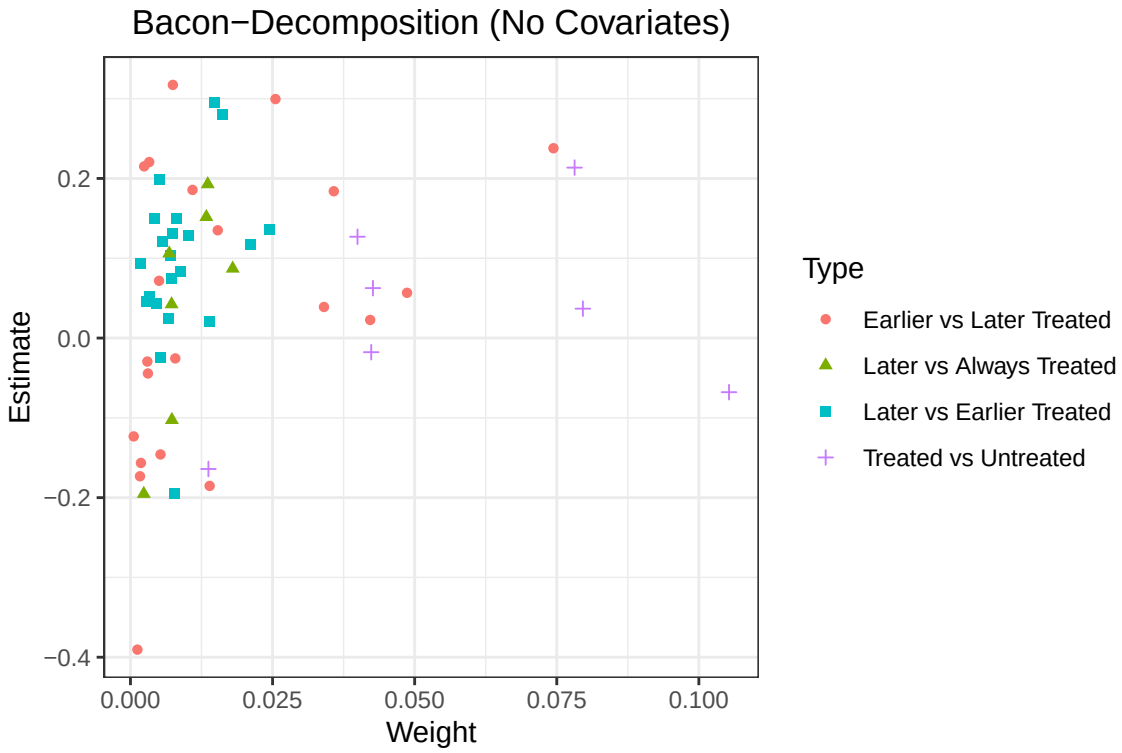
Subsample Point Estimates

Original Results

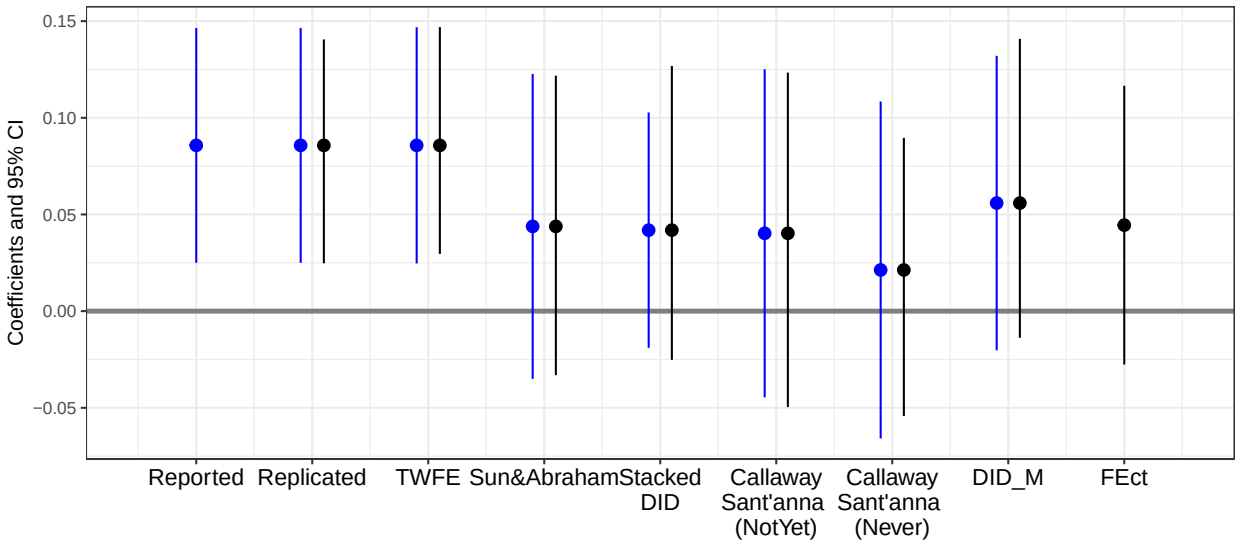
```
sol <- feols(logiptu~cad_update|c6_ibge+year,data = df,cluster = "c6_ibge")
summary(sol)
```

```
## OLS estimation, Dep. Var.: logiptu
## Observations: 25,536
## Fixed-effects: c6_ibge: 3,289,  year: 8
## Standard-errors: Clustered (c6_ibge)
##             Estimate Std. Error t value  Pr(>|t|)
## cad_update 0.085739   0.031001 2.76566 0.0057125 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.09035     Adj. R2: 0.836085
##                  Within R2: 4.387e-4
```

Goodman-Bacon Decomposition

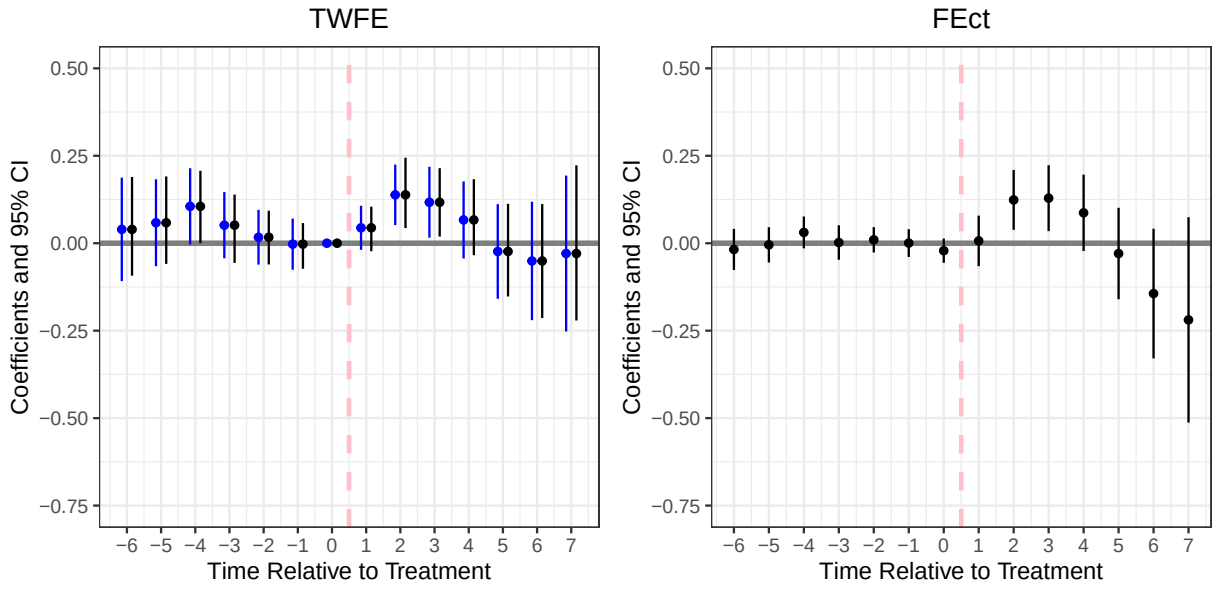


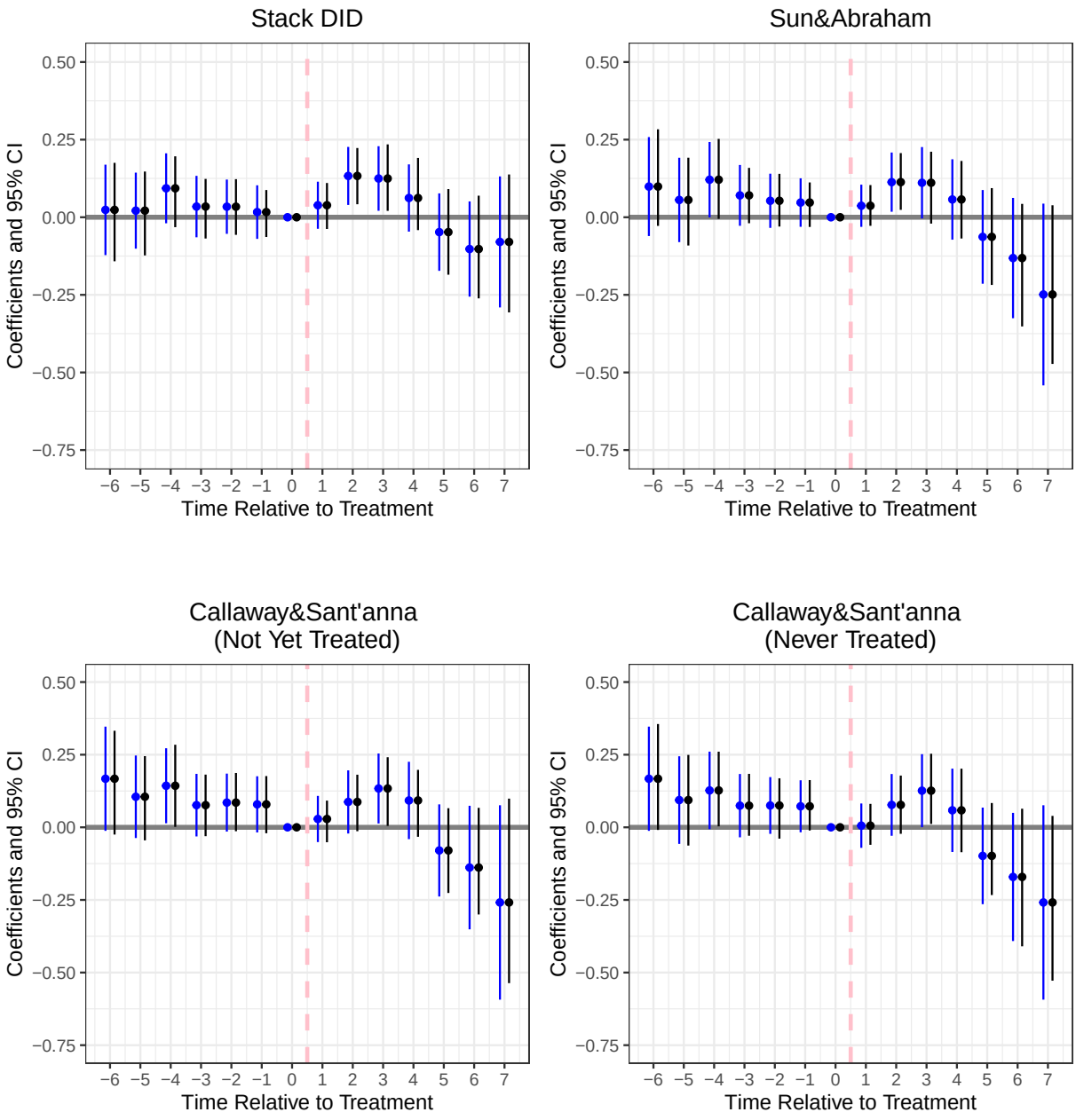

```
##   |                                                                      |
```

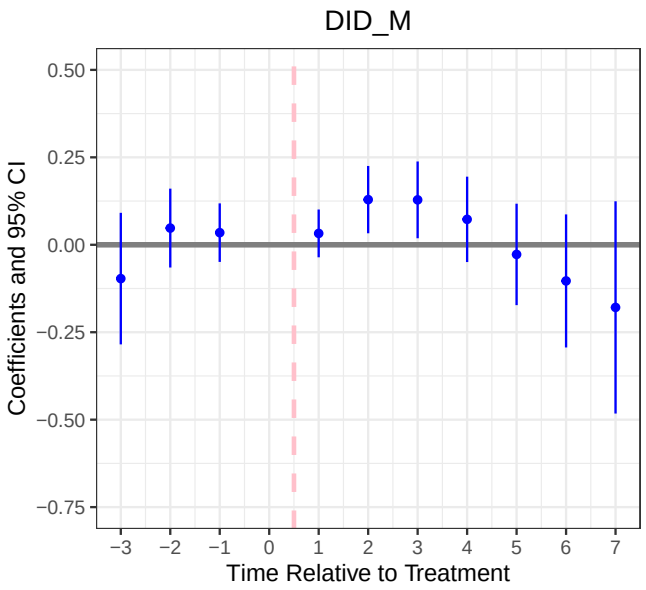


Using the subsample, all HTE-robust estimators, as well as TWFE, yield similar estimated ATT, although many estimates are statistically insignificant.

Subsample Dynamic Treatment Effects



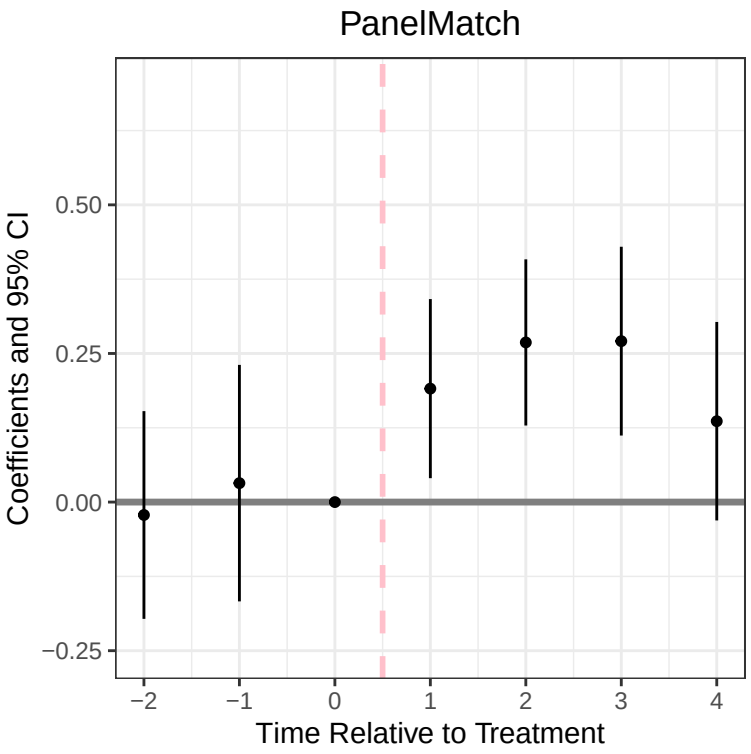

Stack DID
Sun&Abraham
Callaway&Sant'anna
(Not Yet Treated)
Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
0.50
0.25
0.00
−0.25
−0.50
−0.75
−6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7
Time Relative to Treatment



All HTE-robust estimators as well as TWFE yield similar estimated DTEs. These estimated DTE exhibit a consistent pattern across the post-treatment periods. Initially, the estimated DTEs are positive during the first several post-treatment periods. They shrink over time and become negative in subsequent periods.

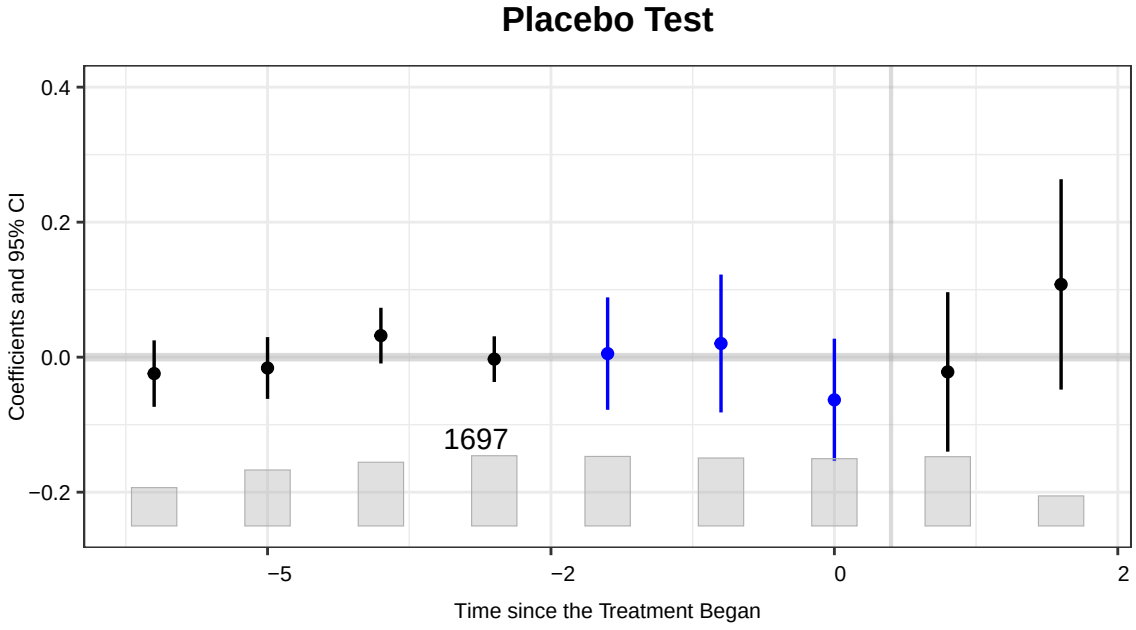
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 4 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Placebo Test

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



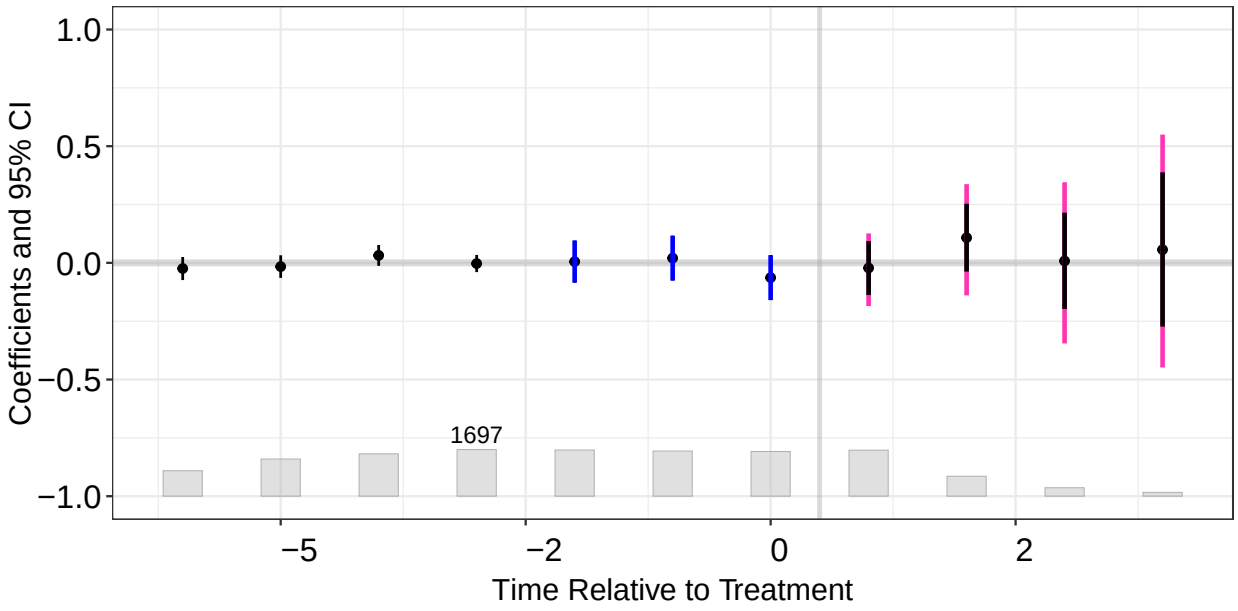
Test Results

```
##                                  p-value
## F test                             0.719
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.279
## Placebo test                       0.752
## Carryover effect test                 NA
```

We find little evidence for violations of the PT assumption in the subsample. However, the equivalence test fails to reject the null that the residuals in pre-treatment periods exceed the estimated ATT possibly due to limited power.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary of findings

We find strong evidence for the violation of the PT in the full sample, mainly caused by the pure-control group. As a result, different estimators drastically disagree with each other. In the subsample, however, we find little evidence for violations of the PT and all estimators yield qualitatively similar results (though some of which are underpowered). The estimated ATT in both samples are not statistically significant.

Clarke (2020)

Contents

A First Look at Data 1
Point Estimates 4
Dynamic Treatment Effects 5
Diagnostic Tests 6
Sensitivity Analysis 7
Summary 8

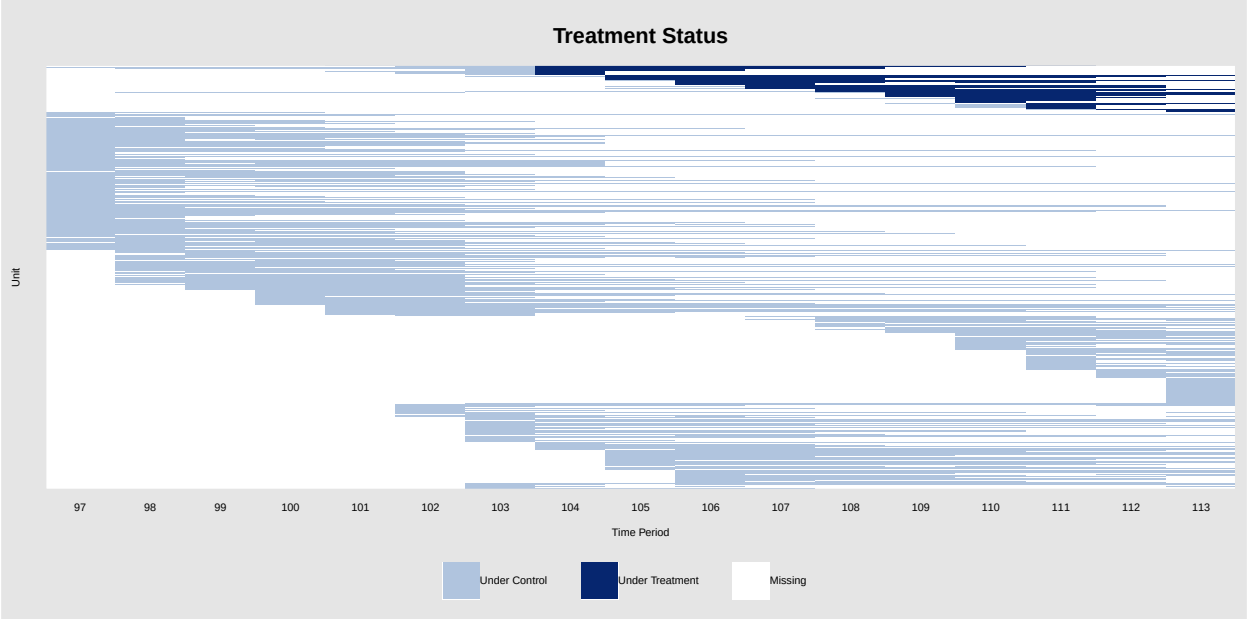
A First Look at Data

The paper investigates the effects of faction affiliation on average donor ideology, using US House candidate-congress year level panel data between the 97th and 113rd congress. One of the main findings of this paper is that "Joining the Blue Dog Coalition leads to a more conservative donor base, as reflected by the average contributor" (p463).

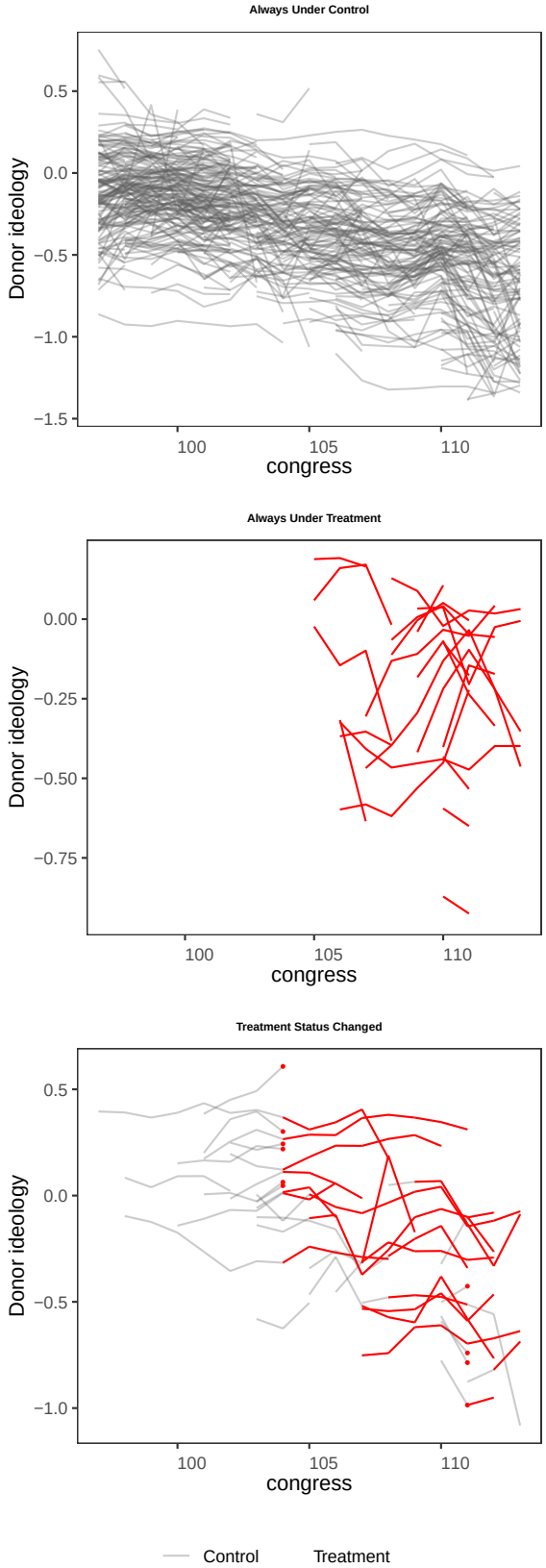
Model. We focus on **Figure 4** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors (not clustering).

Replication Summary	
Unit of analysis	House candidate × congress year
Treatment	Faction affiliation
Outcome	Donor ideology
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a candidate). There are treatment reversals and a large amount of data missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each candidate. The observations under treated status are marked in blue.


Always Under Control
Donor ideology
0.5
0.0
−0.5
−1.0
−1.5
100
105
110
congress
Always Under Treatment
Donor ideology
0.00
−0.25
−0.50
−0.75
100
105
110
congress
Treatment Status Changed
Donor ideology
0.5
0.0
−0.5
−1.0
100
105
110
congress
Control
Treatment

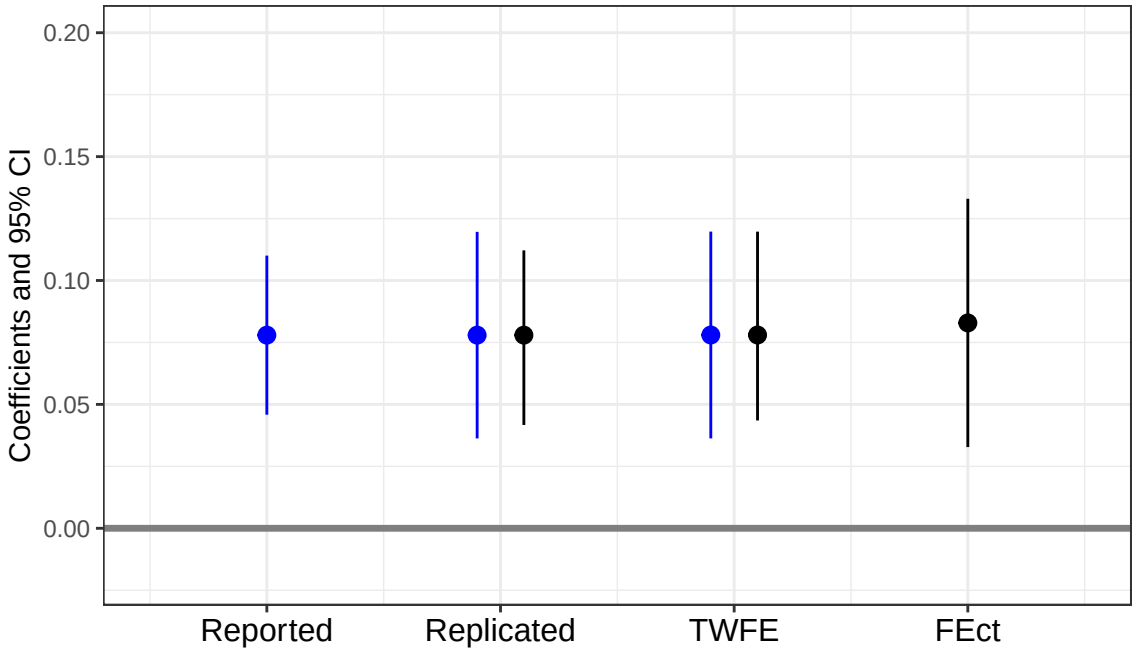
Point Estimates

We first present the regression result following the authors' original specification (not clustering). We then drop the always-treated units and apply two estimators: TWFE and imputation estimator (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

```
sol <- feols(mean_donor_current~bdc + general_percent + primary_percent +
             districtpres + nokken_poole_dim1|icpsr+congress,
             data = df,vcov = 'hetero')
summary(sol)
```

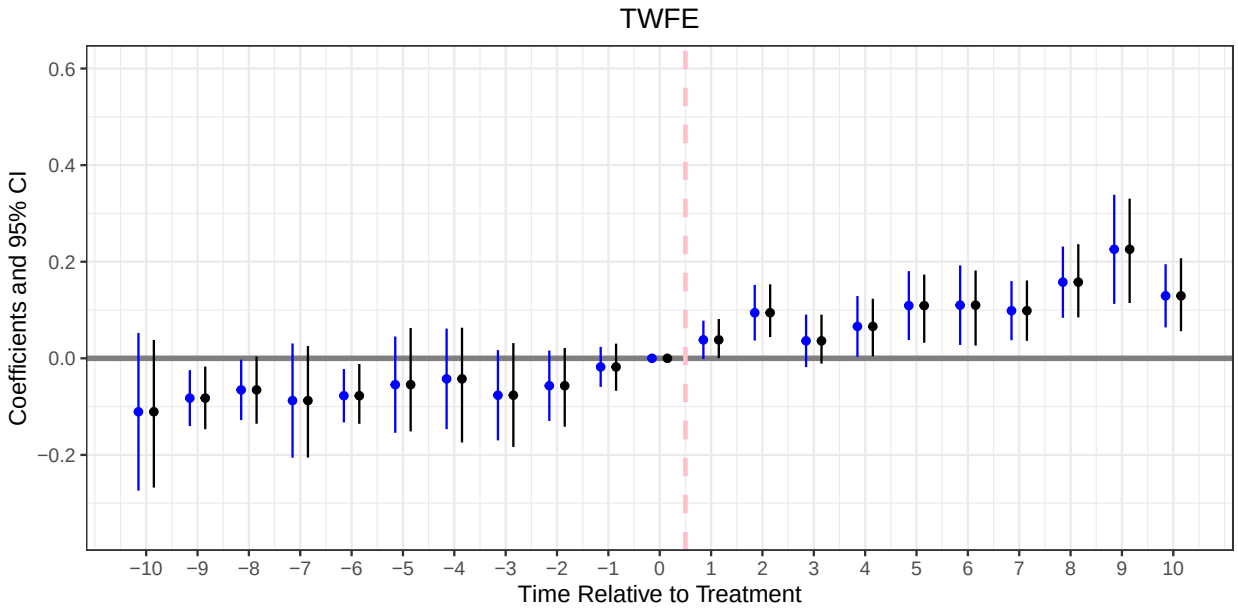
```
## OLS estimation, Dep. Var.: mean_donor_current
## Observations: 3,603
## Fixed-effects: icpsr: 702,  congress: 17
## Standard-errors: Heteroskedasticity-robust
##                   Estimate Std. Error  t value   Pr(>|t|)
## bdc               0.077952   0.016377 4.759753 2.0340e-06 ***
## general_percent   0.000878   0.000182 4.816331 1.5378e-06 ***
## primary_percent   0.000337   0.000145 2.323990 2.0195e-02 *
## districtpres      0.000075   0.000410 0.182130 8.5549e-01
## nokken_poole_dim1 0.065863   0.044430 1.482393 1.3835e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.089028     Adj. R2: 0.910214
##                  Within R2: 0.022808
```

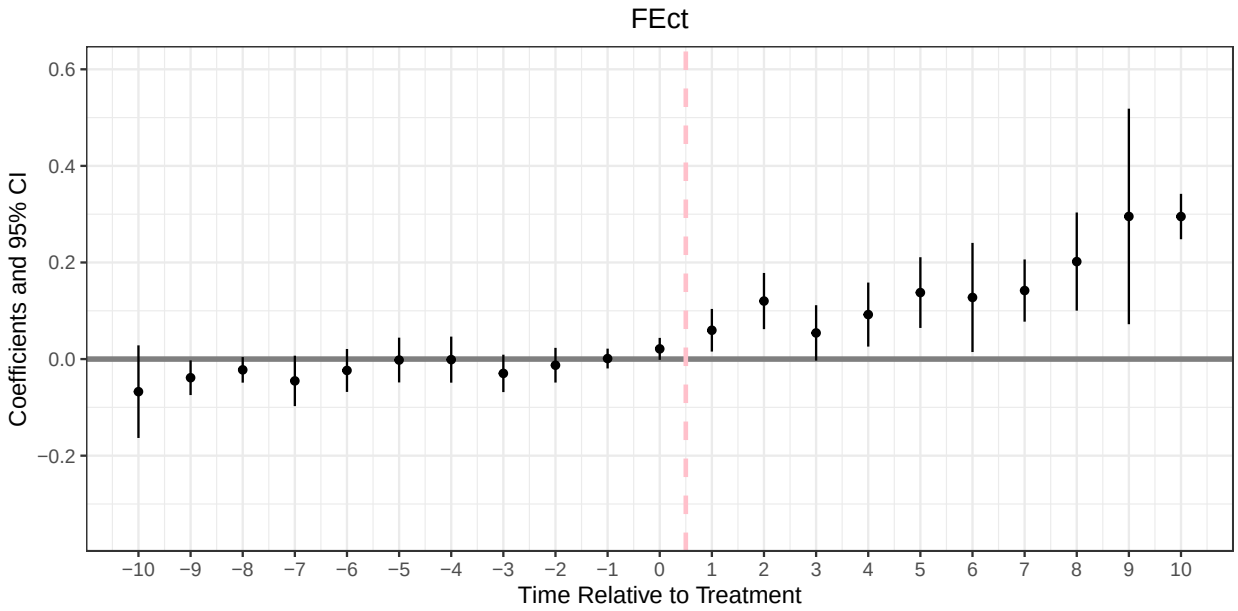


The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding DTE/event-study plots using TWFE and FEct.



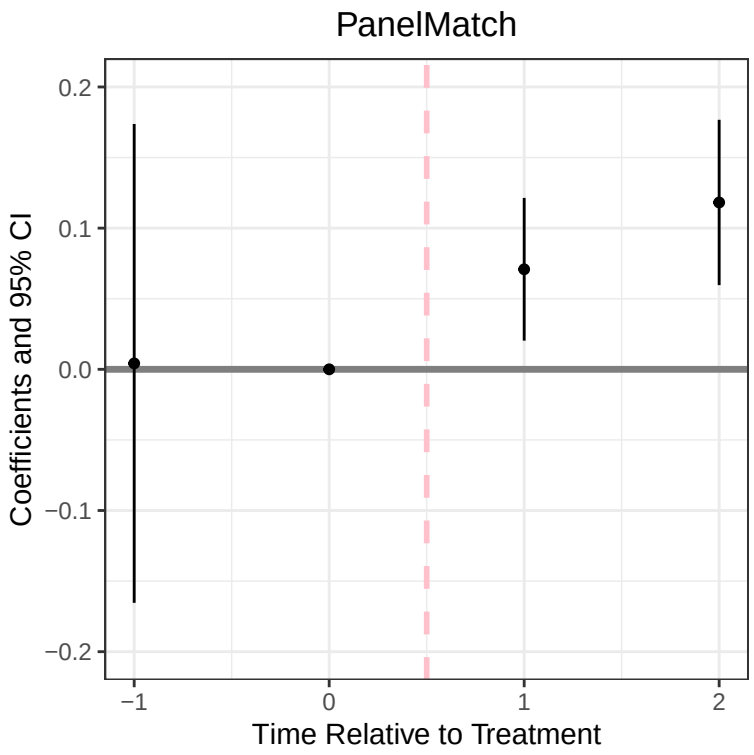


The TWFE and FEct estimators generally align with each other. The estimated DTEs display positive values and demonstrate increasing patterns during post-treatment periods.

DTE of a Balanced Subsample

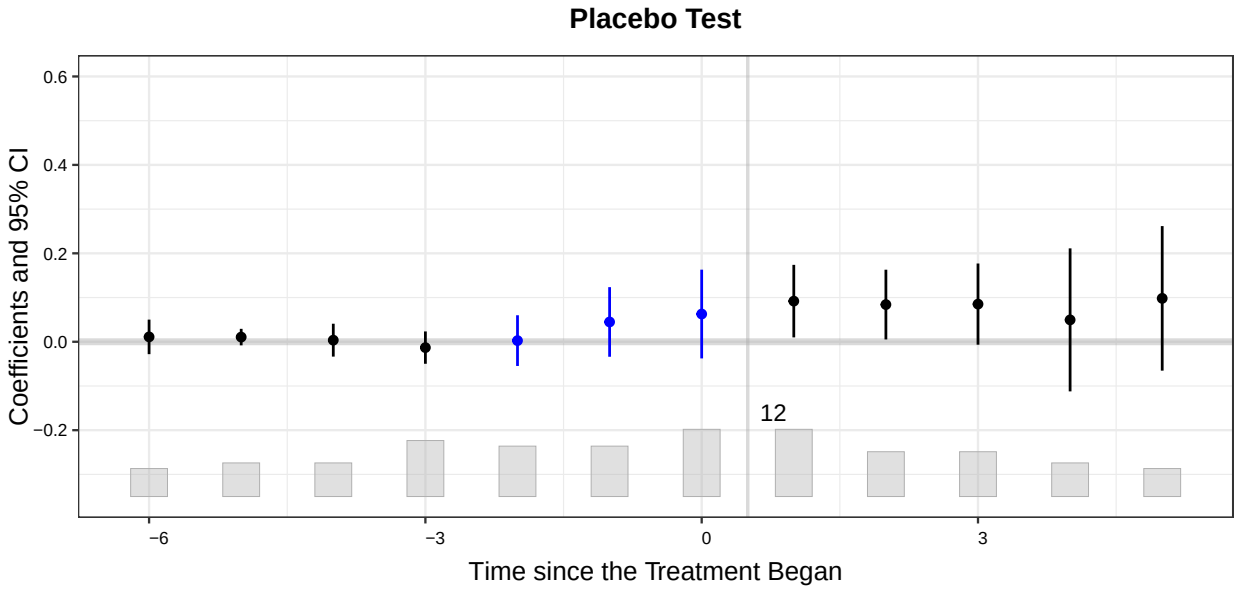
We also estimate the DTE using PanelMatch (lead = 2 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate

time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.

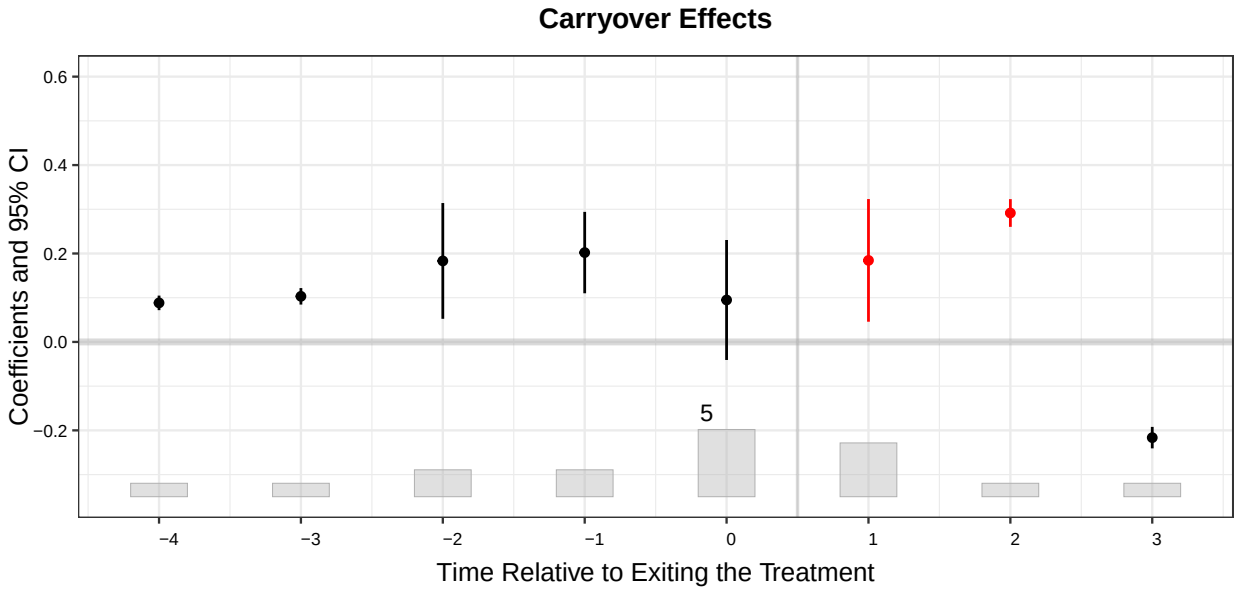


Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.



Carryover Test



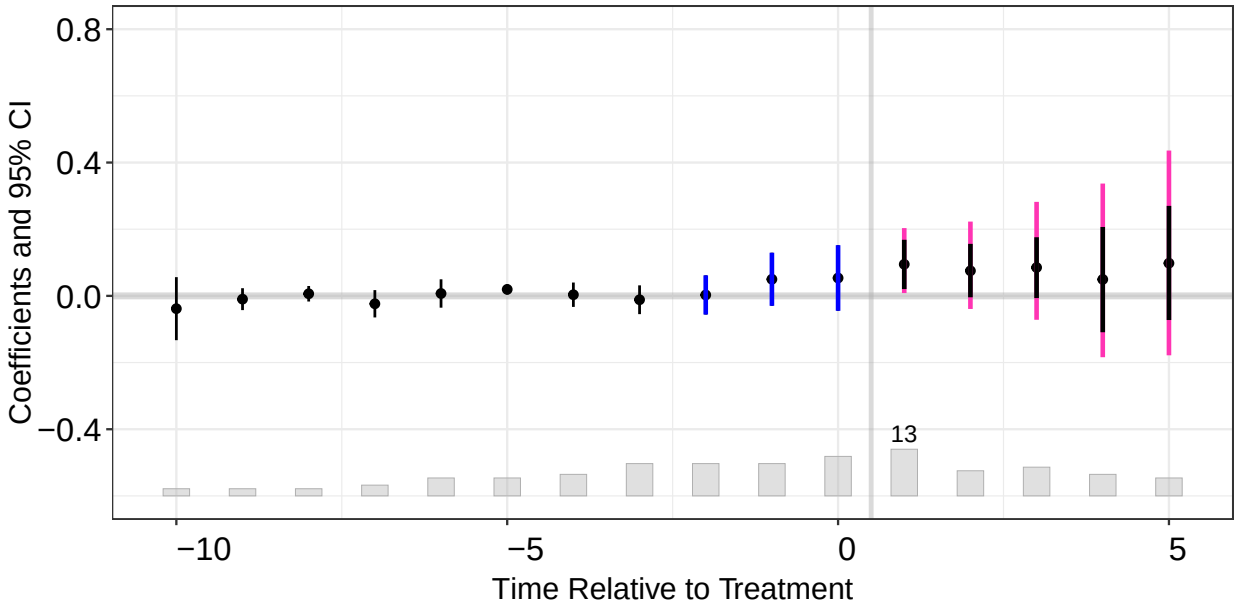
Test Statistics

```
##                                  p-value
## F test                             0.378
## Equivalence test (default)         0.104
## Equivalence test (threshold=ATT)   0.000
## Placebo test                       0.271
## Carryover effect test              0.000
```

We find little evidence of violations of the parallel trends (PT) assumption and anticipation effects, though there a few DTE estimates that are significant prior to the onset of the treatment. The equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT. However, the test indicates a potential violation of the no-carryover-effect assumption.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

The amount of missing data poses a challenge. However, the results appear robust to HTE-robust estimators, and we find little evidence of violations of the PT. While the issue of carryover effects exists, it can be addressed by removing a few post-treatment periods. The significant ATT is sensitive to the removal of placebo periods and to PT violations benchmarked against the pre-treatment placebo estimates, likely due to the prevalent missingness in the data.

Clayton and Zetterberg (2018)

Contents

A First Look at Data 1

Point Estimates 2

Dynamic Treatment Effects 5

Diagnostic Tests 7

Sensitivity Analysis 8

Summary 9

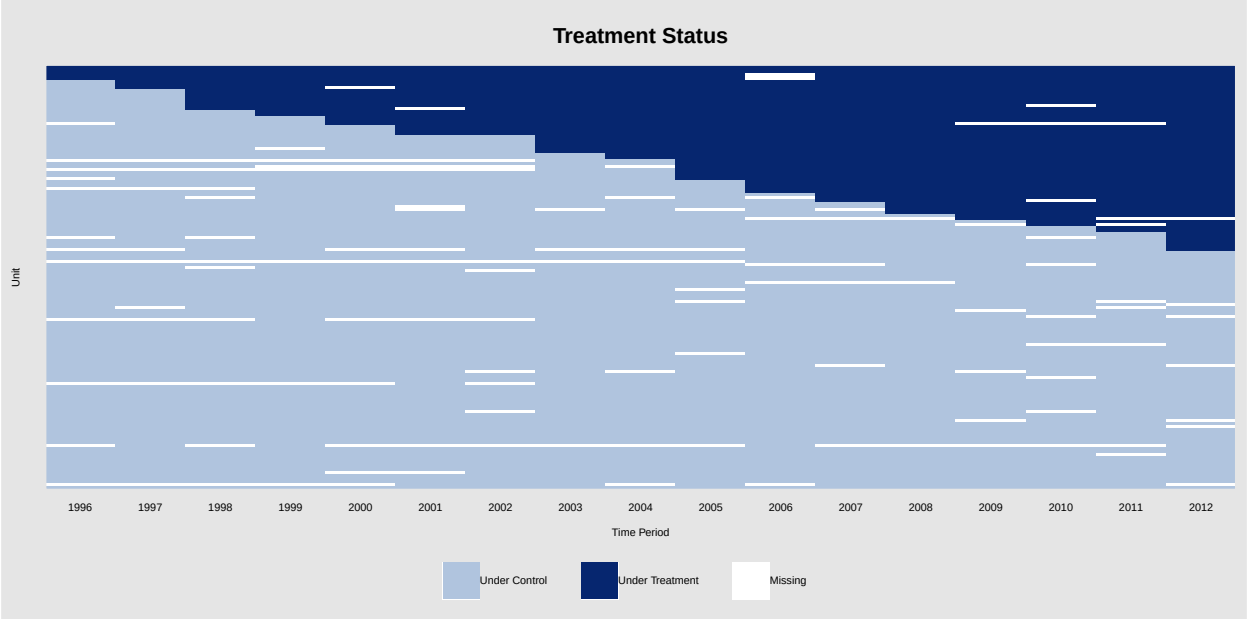
A First Look at Data

The paper investigates the effects of gender quotas on government spending priorities, using country-year level panel data from 1995-2012. One of the main findings of this paper is that "quota adoption is associated with increased health spending in adopting countries relative to nonadopting countries in the same period" (p924).

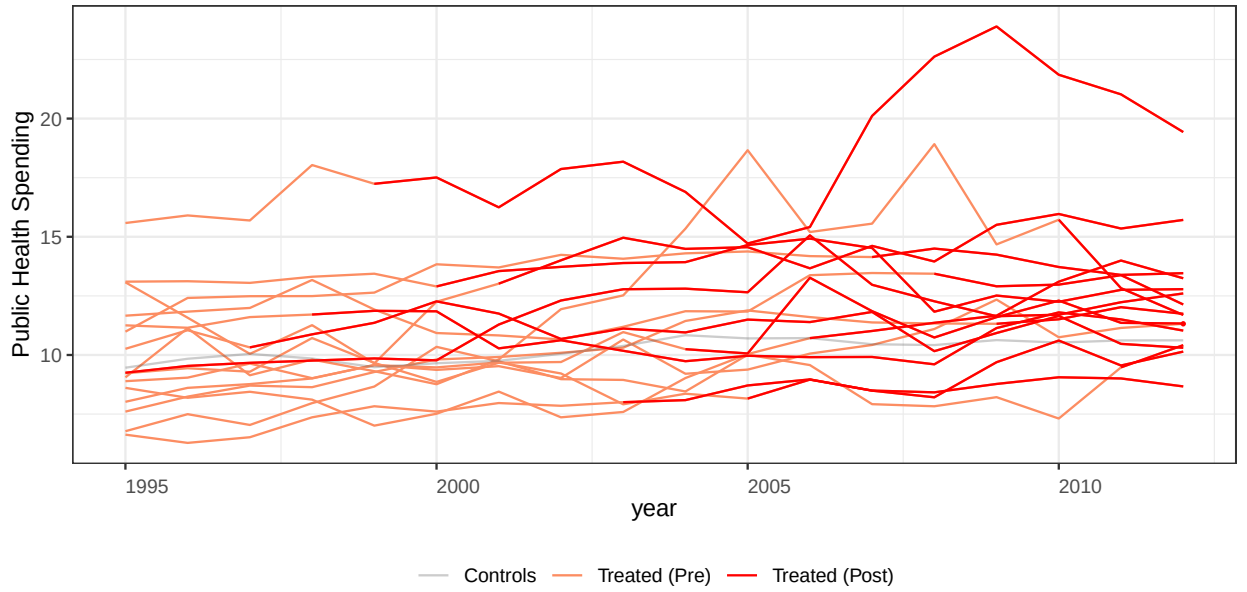
Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the country level.

Replication Summary	
Unit of analysis	Country × year
Treatment	Gender quota
Outcome	Public health spending
Treatment type	Staggered
Outcome type	Continuous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a country). The missing data problem is not severe.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each country. The trajectories of the control units are depicted in gray. For the ever-treated units, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this case) and apply the TWFE (with country and year fixed effects), Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DID multiple estimator (DIDM) and imputation estimator (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

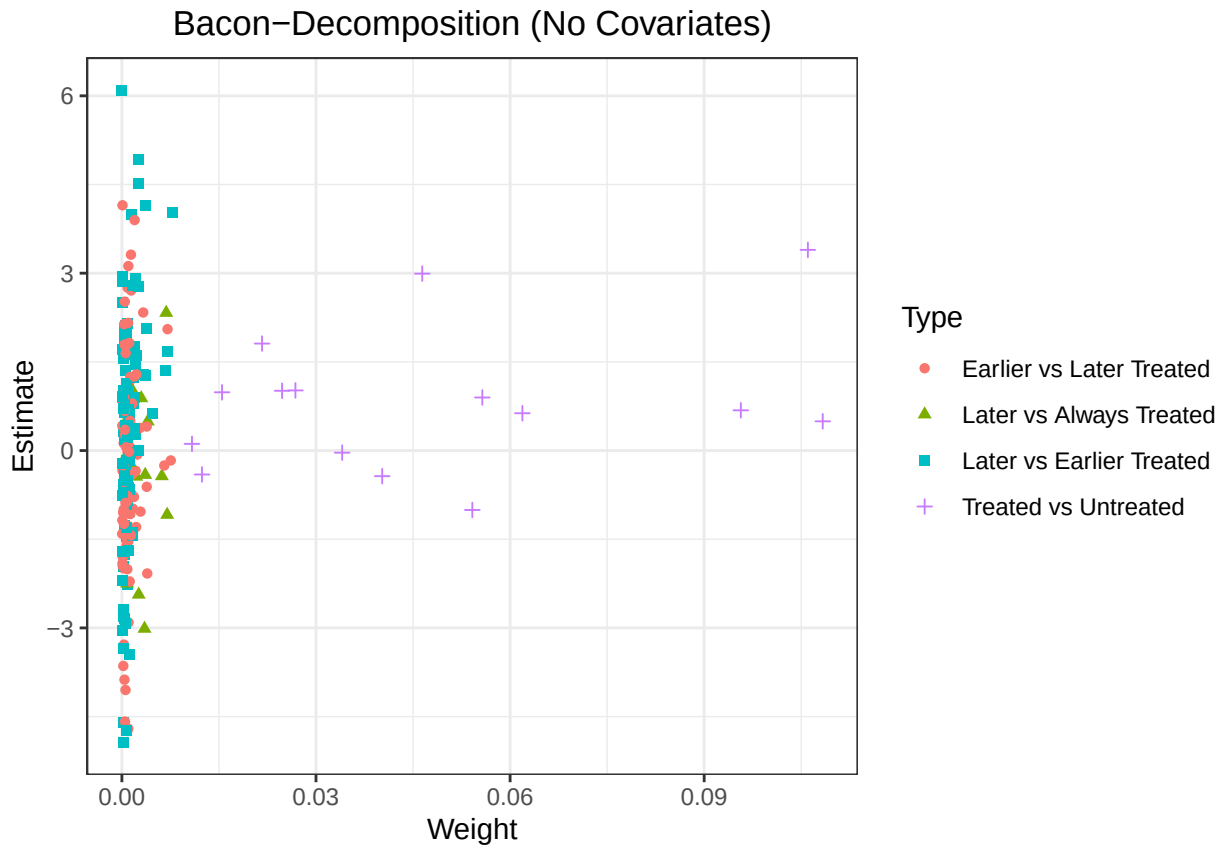
Original Results

```
sol <- feols(percenthealth~adopt_quotalag+demlag+dem_cumlag+
             autoclag+leftlag+left_cumlag+lnODAlag+
             lnOillag+lnTradelag+lnFDIlag+time_trend+time_trend_2|country+year,
             data = df,cluster = "country")
summary(sol)
```

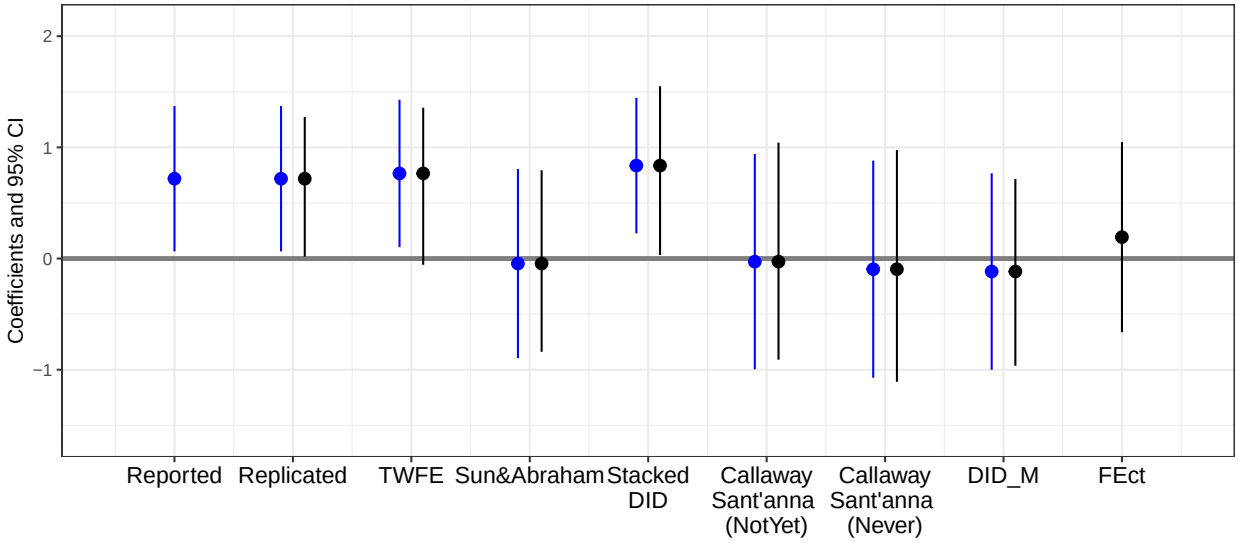
```
## OLS estimation, Dep. Var.: percenthealth
## Observations: 2,227
## Fixed-effects: country: 139,  year: 17
## Standard-errors: Clustered (country)
##                 Estimate Std. Error   t value Pr(>|t|)
## adopt_quotalag  0.718196   0.333559  2.153129 0.033047 *
## demlag          0.346997   0.437014  0.794018 0.428548
## dem_cumlag      0.082651   0.044922  1.839902 0.067931 .
## autoclag       -0.377825   0.360192 -1.048956 0.296032
## leftlag         0.036662   0.307365  0.119279 0.905228
## left_cumlag    -0.022371   0.052916 -0.422761 0.673127
## lnODAlag        0.064913   0.048115  1.349140 0.179502
## lnOillag       -0.118146   0.060436 -1.954911 0.052615 .
## lnTradelag      0.084469   0.225735  0.374197 0.708832
## lnFDIlag        0.024076   0.077208  0.311831 0.755640
## ... 2 variables were removed because of collinearity (time_trend and time_trend_2)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.05217     Adj. R2: 0.754292
##                 Within R2: 0.027646
```

Goodman-Bacon Decomposition

In the Goodman-Bacon decomposition, the comparison between ever-treated cohorts and the control cohort contributes the most to the estimated ATT using TWFE.



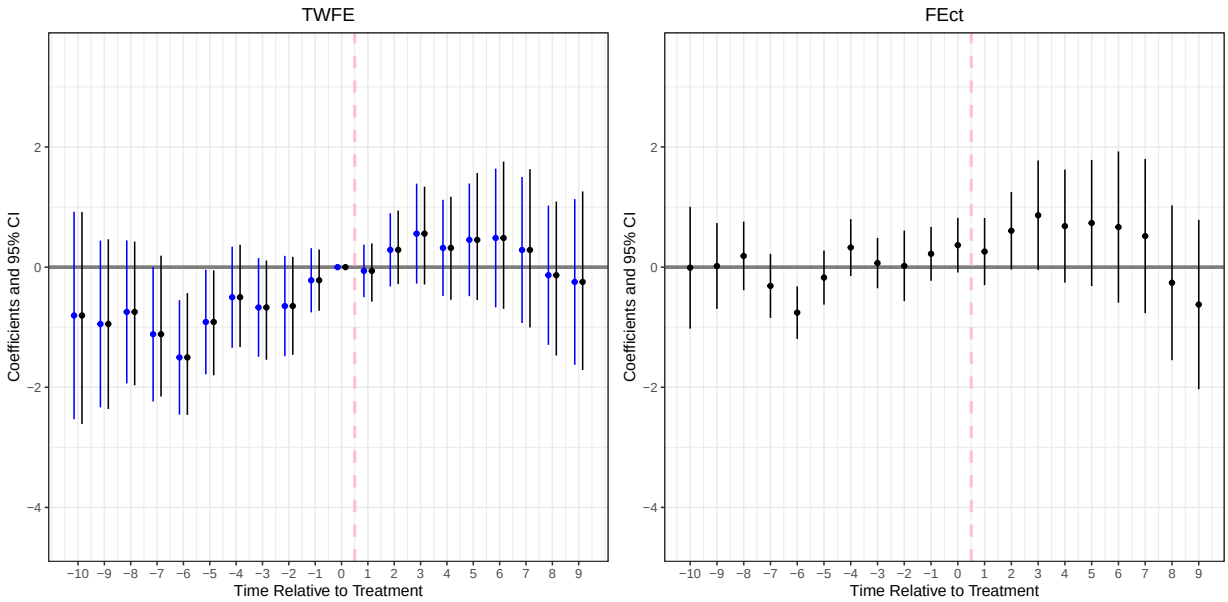

```
## [1] "baseline finished"
##   |                                                                      |
```

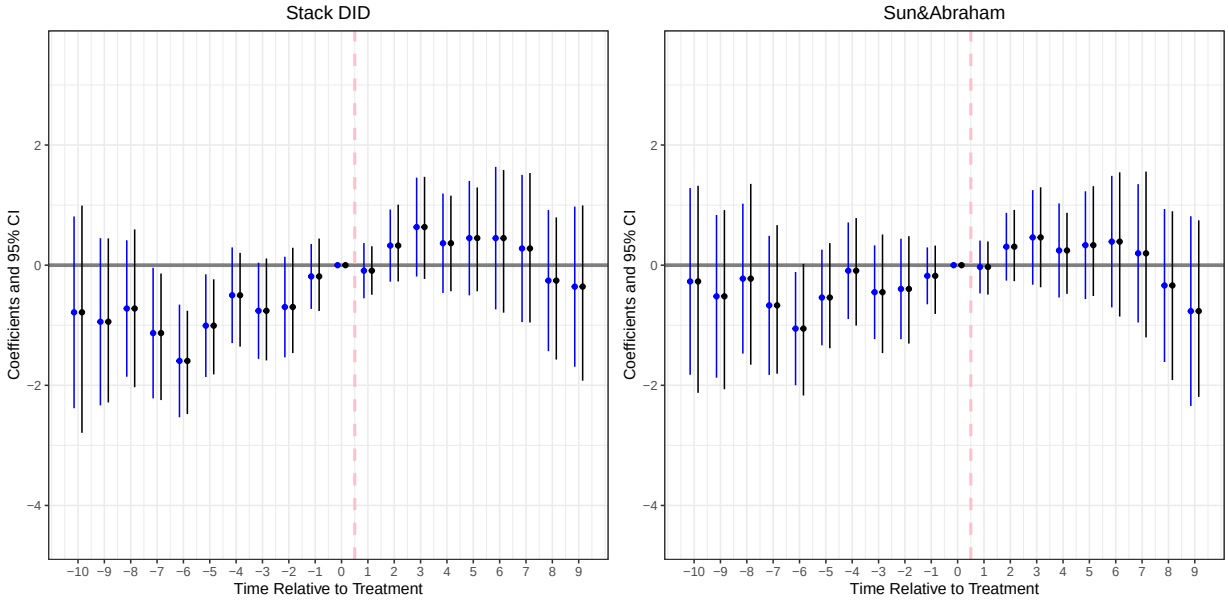


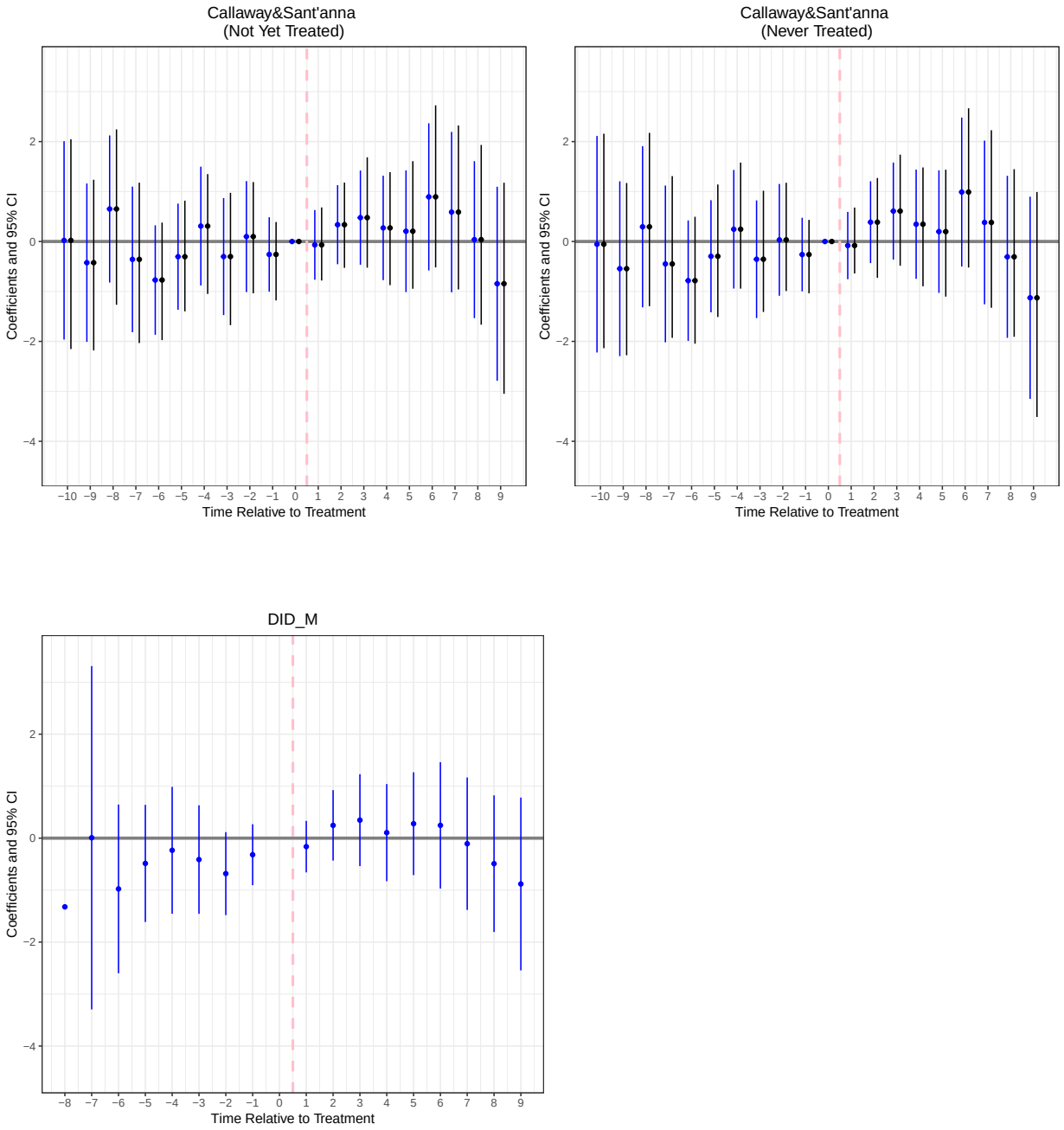
The estimated ATT is significantly positive for both TWFE and Stacked DID estimators. However, the IW estimator, CS estimator, DIDM estimator, and FEct yield ATT estimates that are statistically insignificant and close to zero. The discrepancy in results may be attributed to the fact that the TWFE and Stacked DID employ a comparable weighting scheme assigned to the "treated vs untreated" comparisons.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, FEct, CS, IW, DIDM, and Stacked DID.



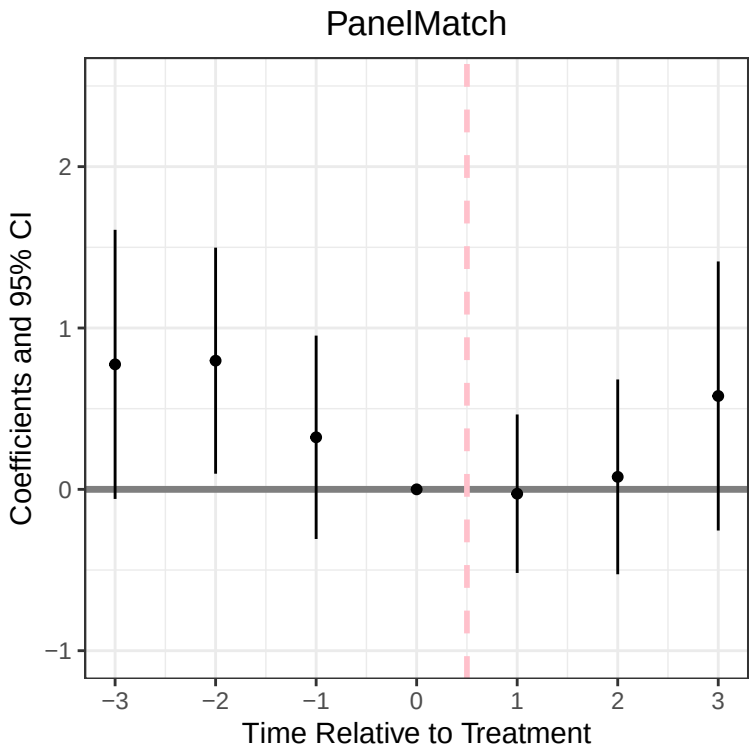




While the estimated DTE for these estimators display a similar pattern in the post-treatment periods (initial increase followed by a gradual decrease), there are some discernible differences in the estimated DTE during the pre-treatment periods. Specifically, both versions of CSDID do not have any significant DTE estimates in the pre-treatment period, whereas all other estimators have at least one.

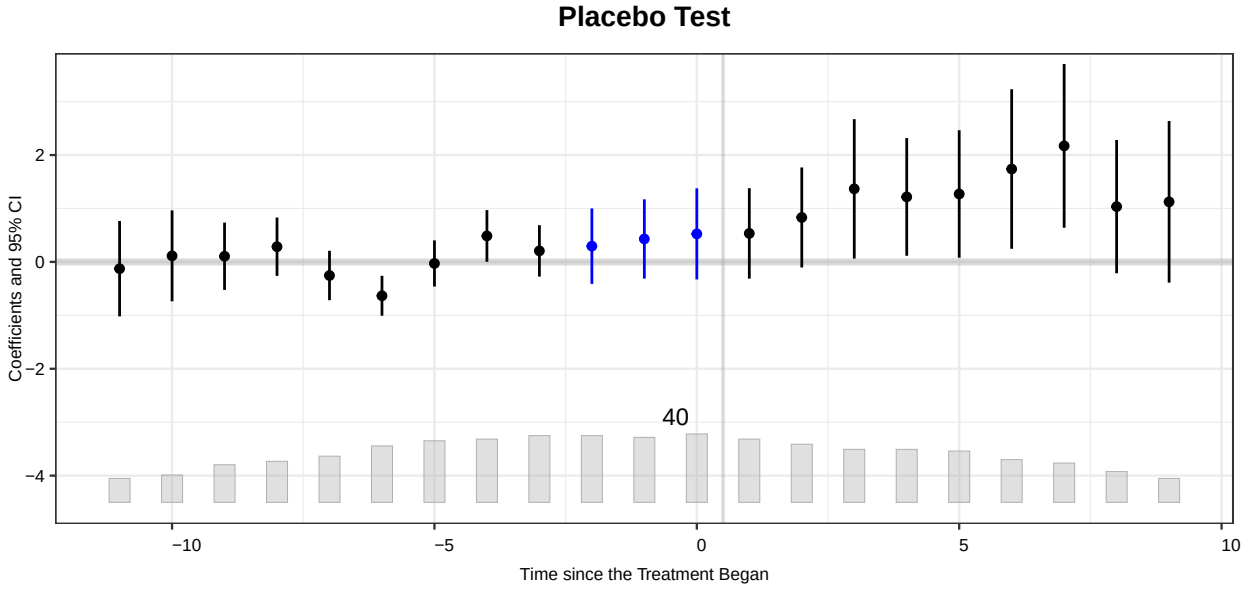
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 3 and lag = 4). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



Test Statistics

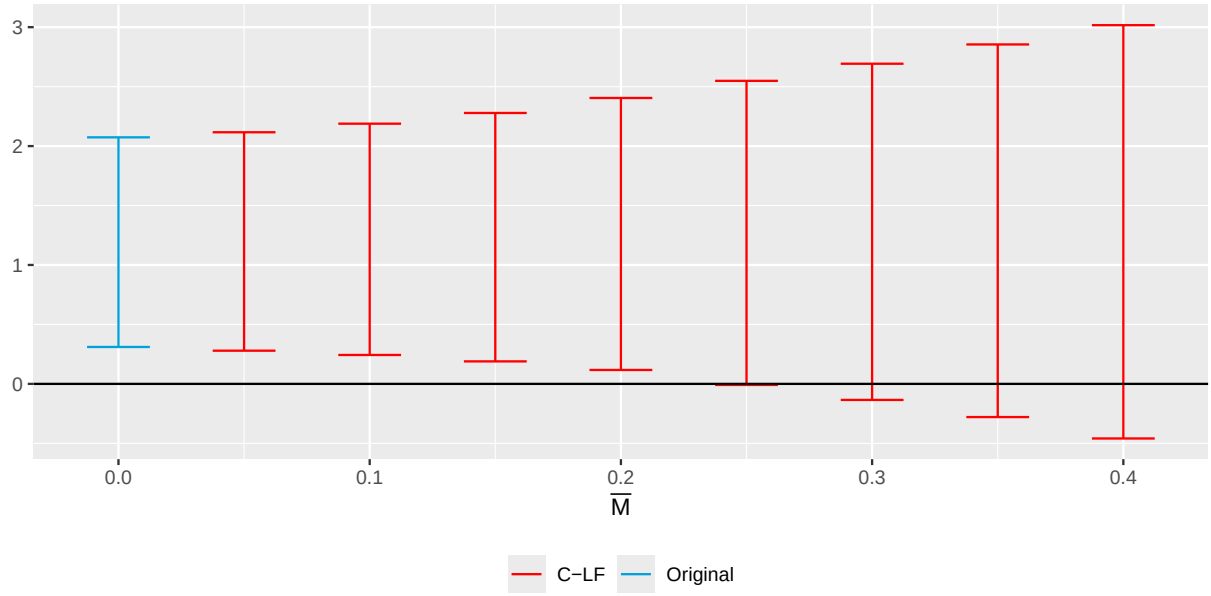
```
##                                   p-value
## F test                              0.086
## Equivalence test (default)          0.511
## Equivalence test (threshold=ATT)    0.994
## Placebo test                        0.205
## Carryover effect test                  NA
```

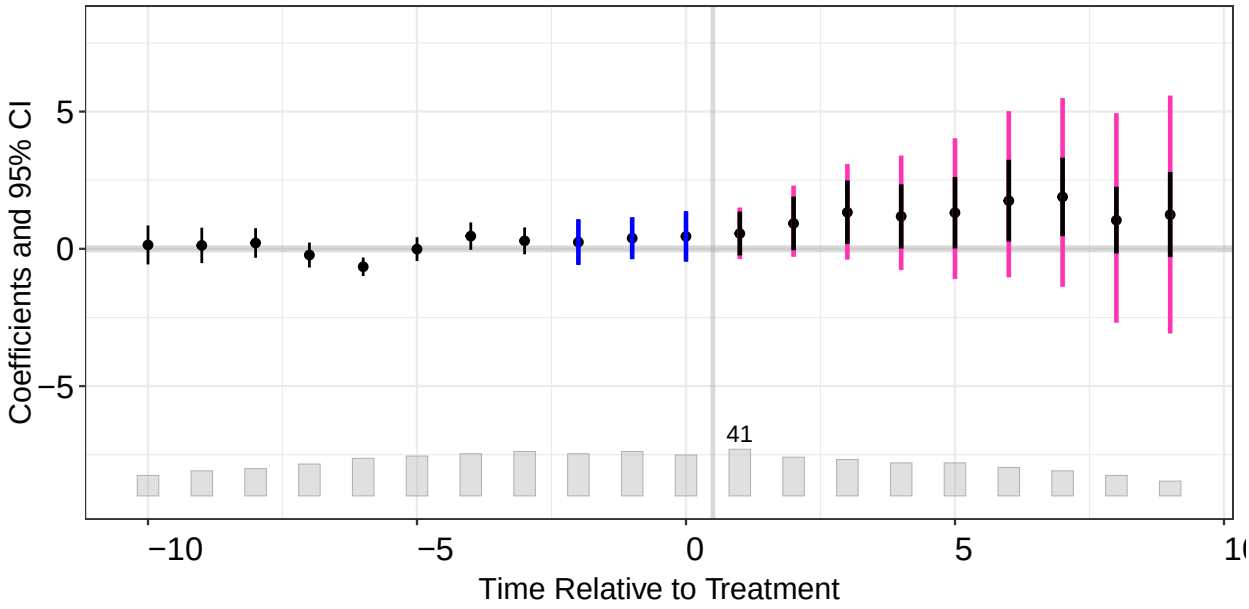
We find some evidence of violations of the parallel trends (PT) assumption in the F test. The equivalence test fails to reject the null hypothesis that the residuals in the pre-treatment periods exceed the estimated ATT, possibly due to limited statistical power.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator turns out to be significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.25$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.25 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

The study is not robust to HTE-robust estimators, and there are indications of parallel trends violations based on the estimated pretrend. However, the study does not reject either the F test or the placebo test. HTE-robust estimators produce noticeably different estimates for the average effect. After removing the placebo periods, the estimated ATT based on the imputation estimator is statistically significant at the 5% level, but appears sensitive to mild PT violations with a breakdown threshold of $M = 0.25$.

Cox and Dincecco (2021)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

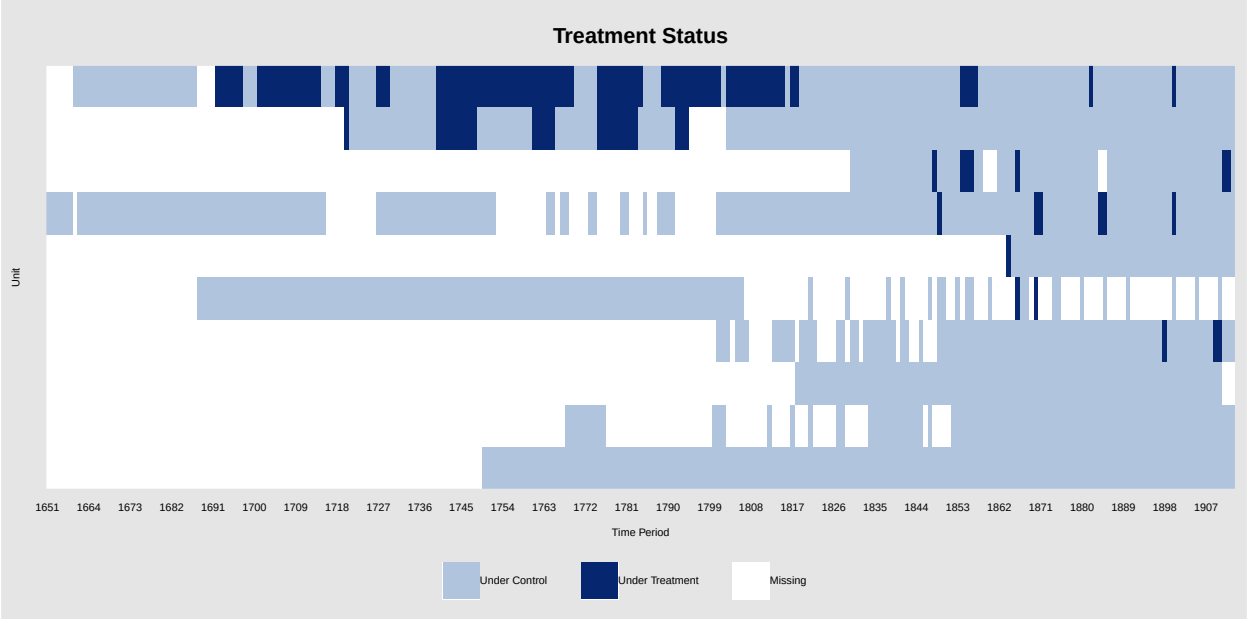
A First Look at Data

The paper investigates the effects of credible budgets on larger wartime expenditures, using European country-year level panel data between 1650 and 1913. One of the main findings of this paper is that "having a budgetary regime was associated with a 39% increase in per capita wartime spending" (p857).

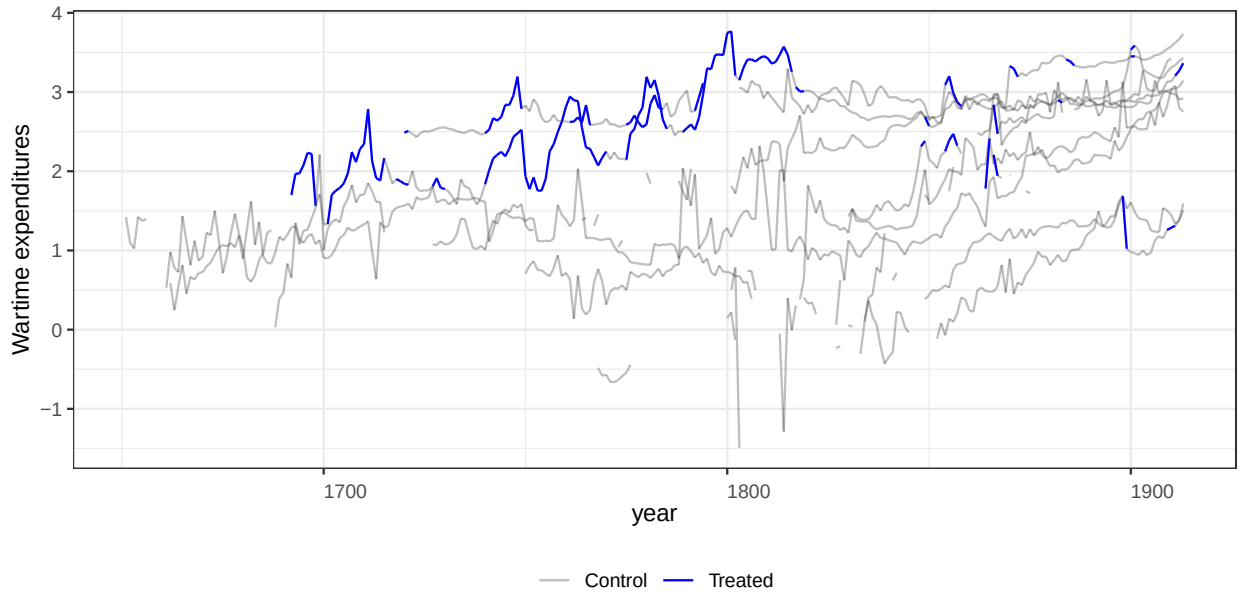
Model. We focus on **Model 1 of Table 1** in the paper, which includes a lagged dependent variable (LDV). The authors use a two-way fixed effects (TWFE) model and report robust standard errors (SEs) clustered at the country level.

Replication Summary	
Unit of analysis	Country × year
Treatment	Credible budget
Outcome	Wartime expenditures
Treatment type	General
Outcome type	Continuous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a country). There are treatment reversals and data missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each country. The observations under treated status are marked in blue.



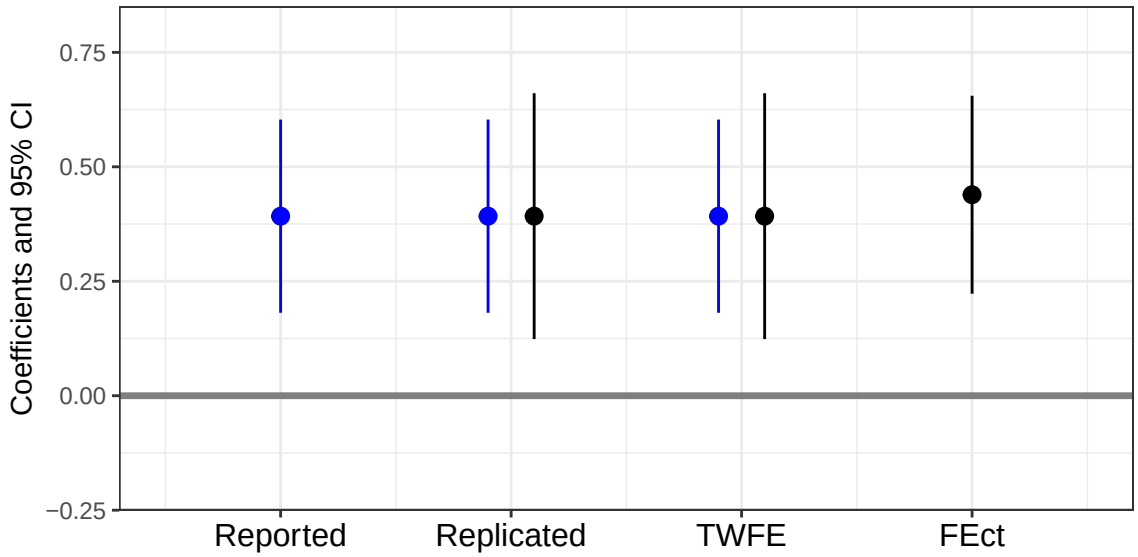
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and imputation estimator (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-jackknifed CIs, respectively.

Original Results

```r
sol <- feols(logexppercap~interactionBE +budget+extwar+linwar+lurbanrate|countrycode+year,
             data = df,cluster = "countrycode")
summary(sol)
```

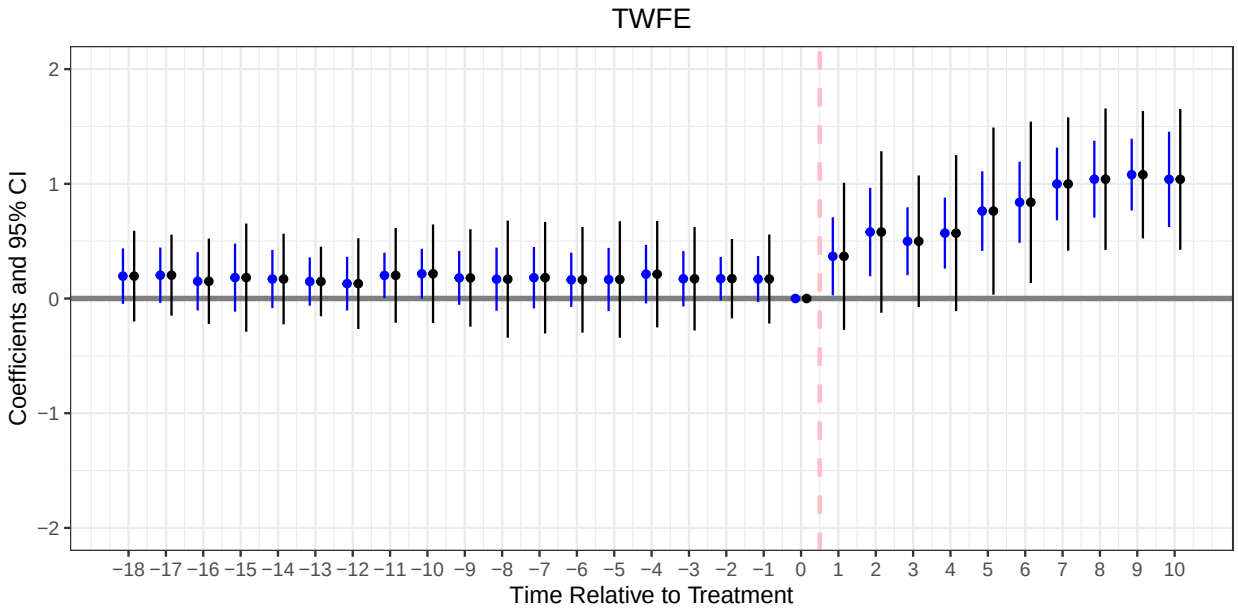
```
## OLS estimation, Dep. Var.: logexppercap
## Observations: 1,361
## Fixed-effects: countrycode: 10,  year: 259
## Standard-errors: Clustered (countrycode)
##                Estimate Std. Error   t value  Pr(>|t|)
## interactionBE  0.392174   0.107644  3.643251 0.0053736 **
## budget         0.309072   0.127834  2.417765 0.0387524 *
## extwar        -0.120769   0.112041 -1.077899 0.3091203
## linwar         0.170655   0.077939  2.189607 0.0562863 .
## lurbanrate     0.339564   1.531121  0.221775 0.8294396
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.317646     Adj. R2: 0.857698
##                  Within R2: 0.157561
```

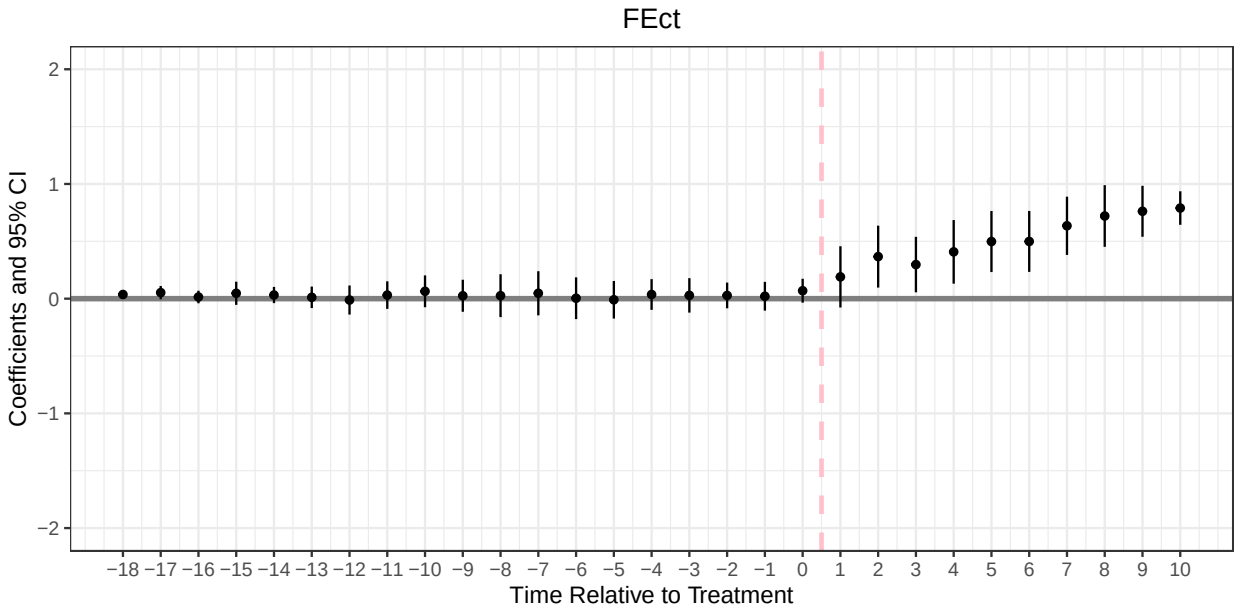


The TWFE estimate loses statistical significance when cluster-jackknife SEs are being used. The FEct estimate is significantly positive and has a smaller confidence interval than the TWFE estimate under cluster-robust SEs.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct, with cluster-robust CIs and cluster-jackknifed CIs.

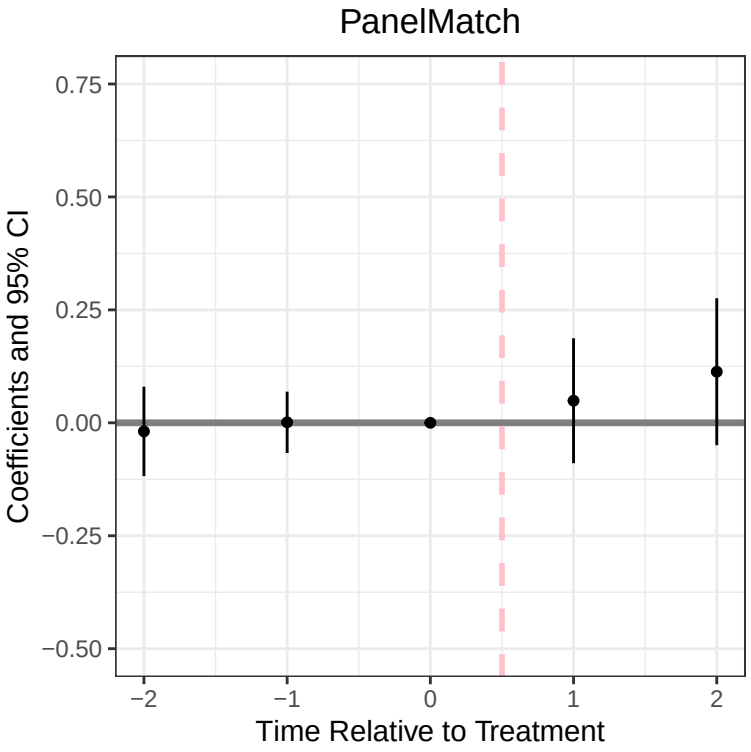




All estimated DTEs exhibit an upward trend in the post-treatment periods.

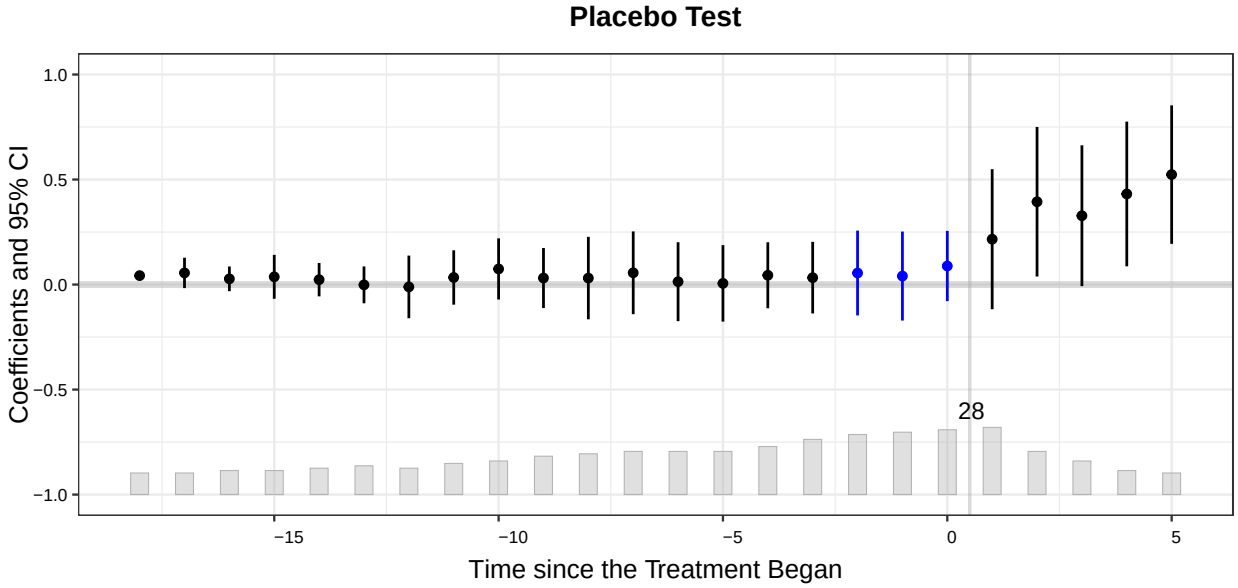
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 2 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.

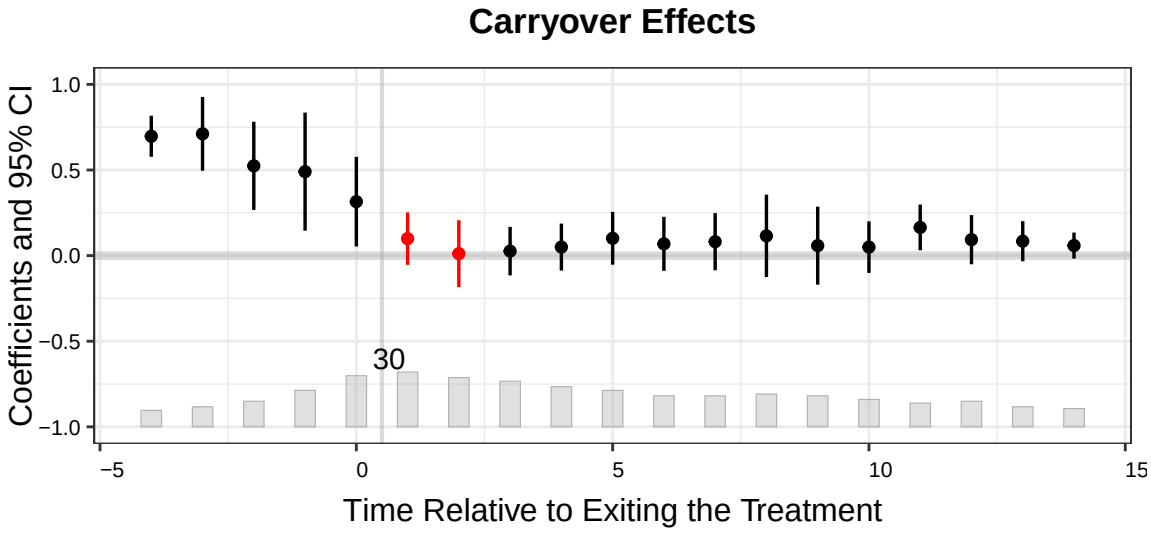


Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.



Carryover Test



Test Statistics

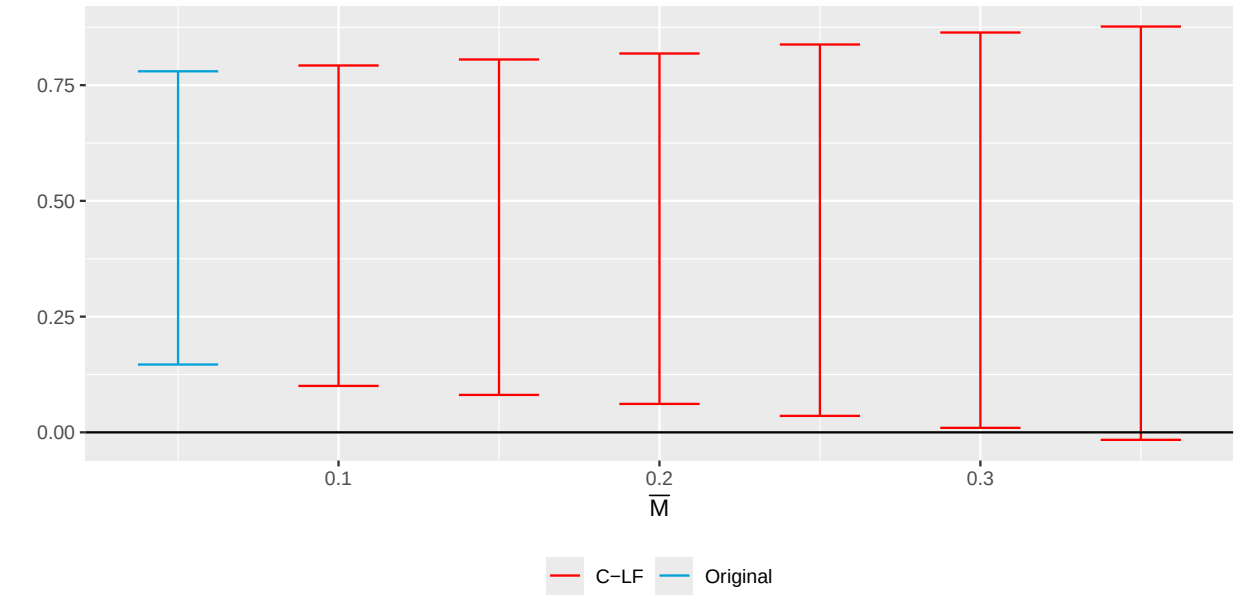
```
##                                  p-value
## F test                             0.116
## Equivalence test (default)         0.104
## Equivalence test (threshold=ATT)   0.000
## Placebo test                       0.529
## Carryover effect test              0.510
```

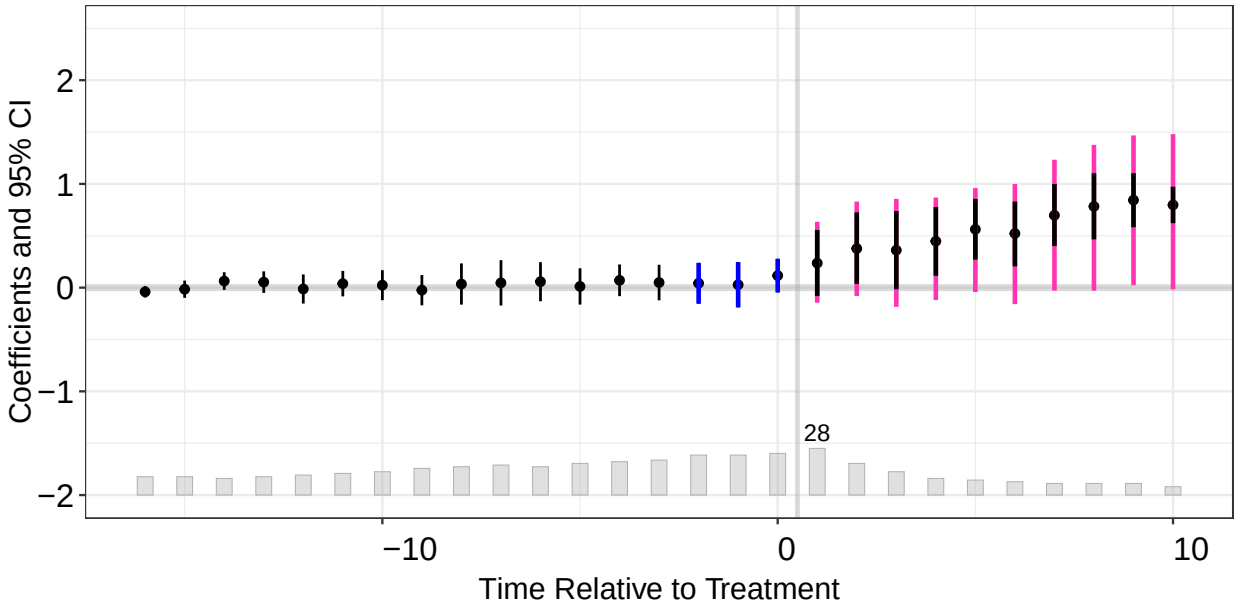
We find little evidence for potential violations of no-carryover-effect assumption. The F-test and placebo test cannot reject the null of no pre-trend or no anticipation effect, and the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.3$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.3 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model seems to be HTE-robust, but the uncertainties are likely to be under-estimated due to the fact that the number of units is very small (cluster-jackknife produces much larger SE estimates). We find no evidence for potential violations of the parallel trend assumption and find evidence that a violation would be small relative to the ATT if it exists. The missing data problem is severe in this study, but not surprising given the research context. The estimated ATT appears to be sensitive to mild PT violations in sensitivity analysis with breakdown threshold $M = 0.3$.

Dahlstrom and Holmgren (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

A First Look at Data

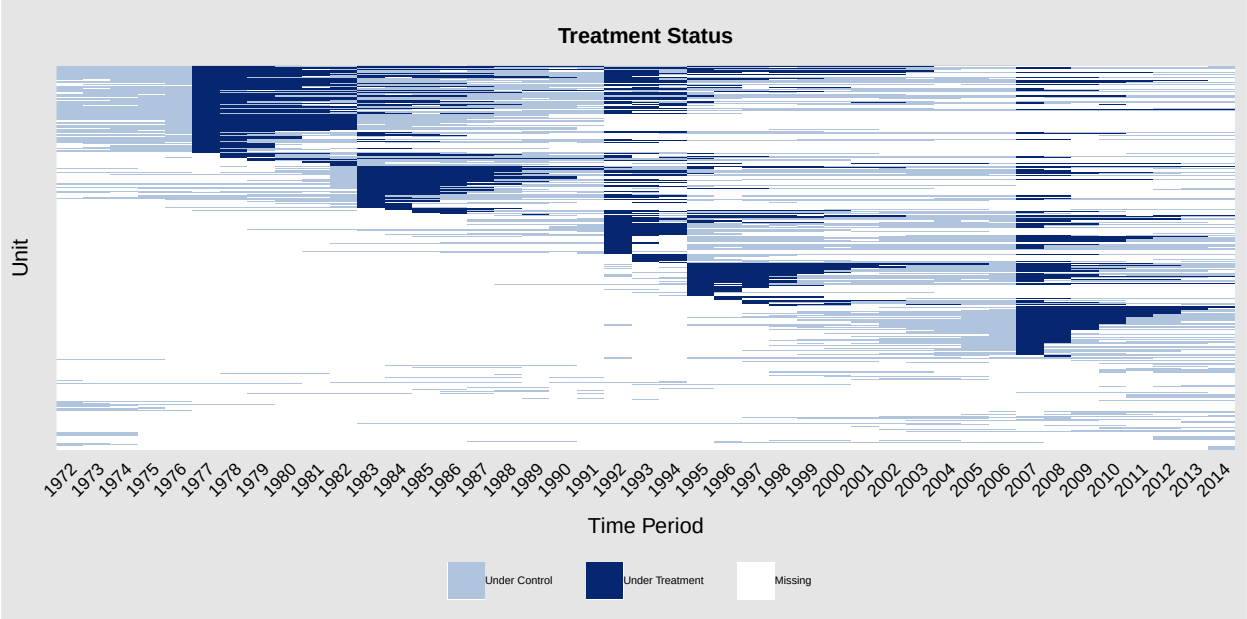
The paper investigates the effects of ideological conflict on budget growth rate, using Swedish agency-fiscal year level panel data between 1972 and 2014. One of the main findings of this paper is "when the ministers that control the bureaucracy appointments represent an ideological bloc different from the legislators who control the bureaucracy resources, we estimate a growth contraction from about

Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the agency level.

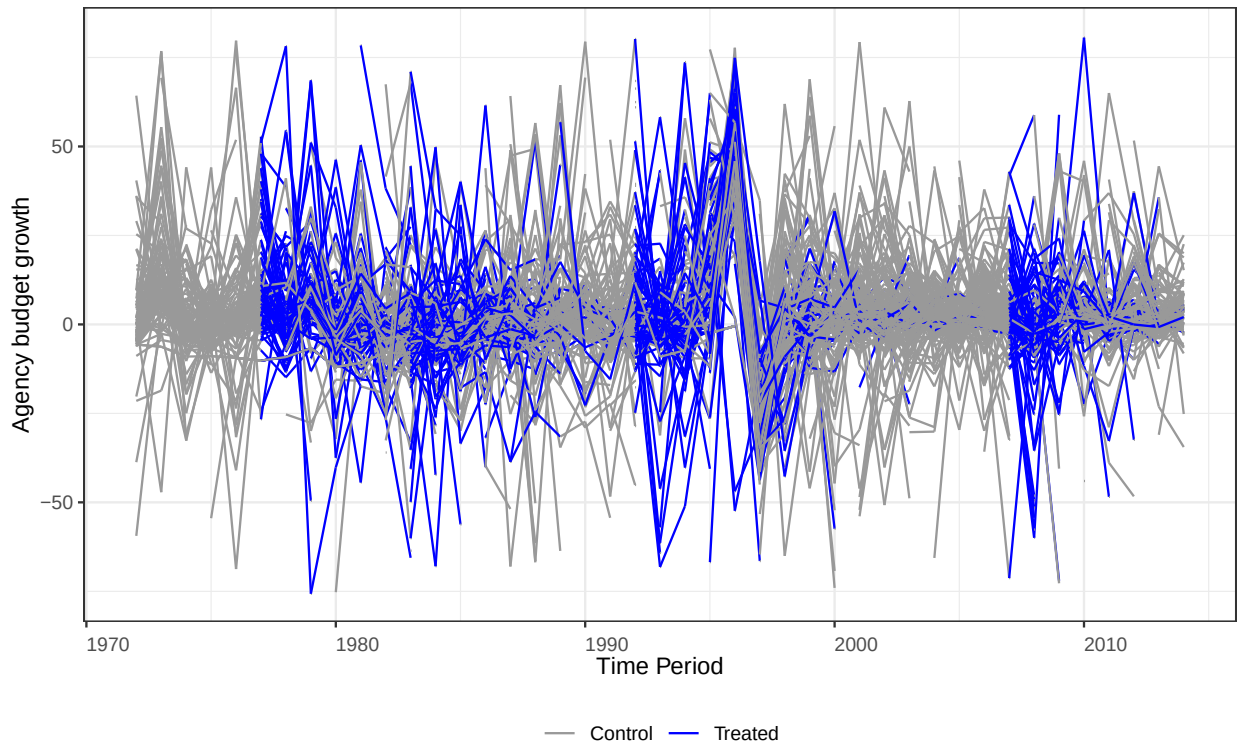
Table 1: Replication Summary

Unit of analysis	Agency × year
Treatment	Ideological conflict
Outcome	Agency budget growth
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (an agency). The treatment has reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

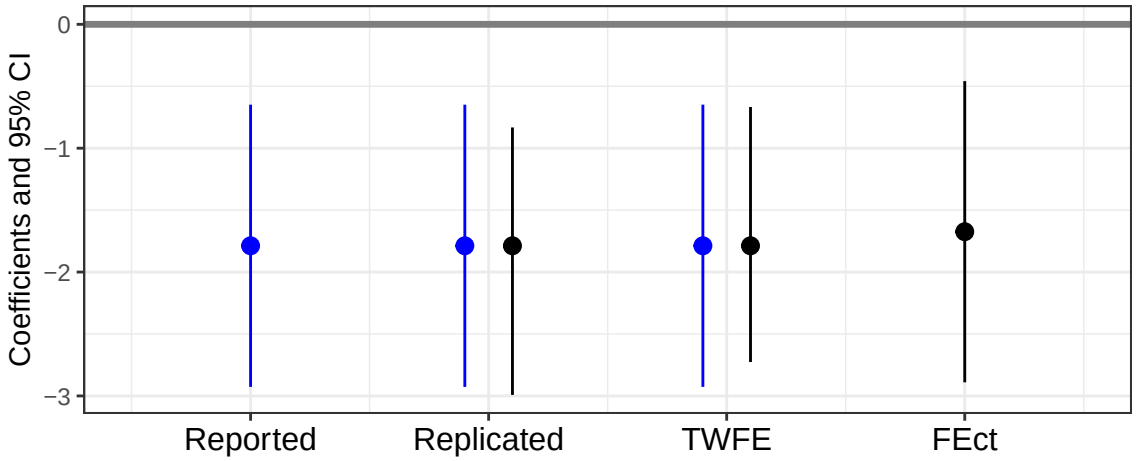
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and imputation estimator (FEct). The point estimates and

their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(bgrowth~conflict+academic+political+private+public+council+
               board+creator+lni_budget_control|agency_id+year,
             data = df,cluster = "agency_id")
summary(sol)
```

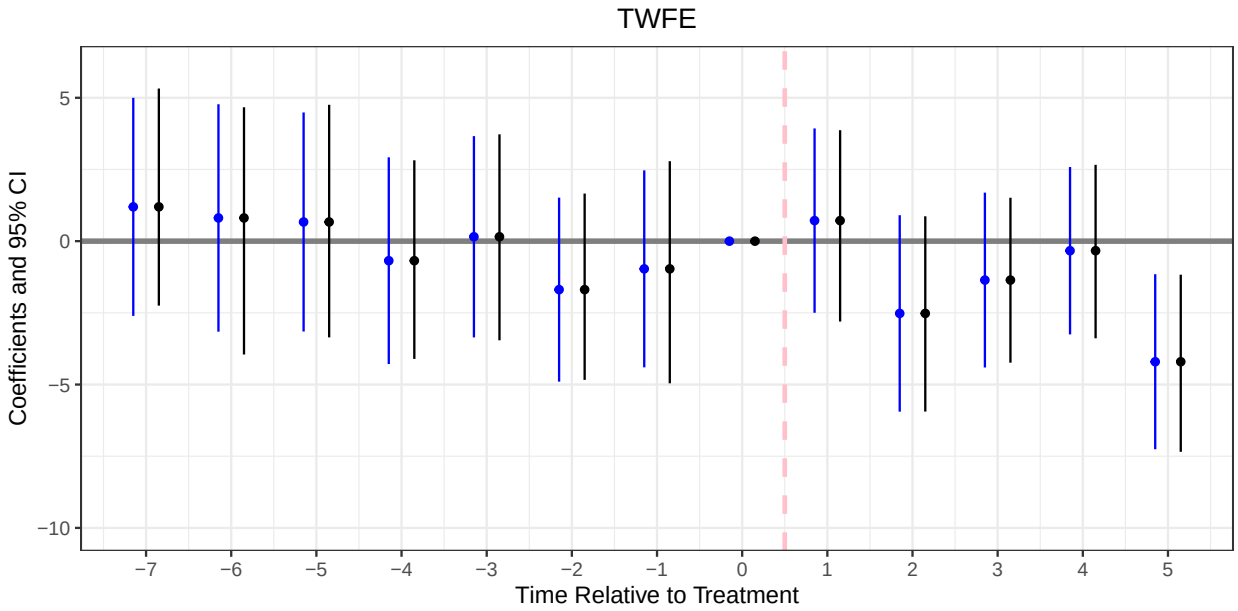
```
## OLS estimation, Dep. Var.: bgrowth
## Observations: 5,316
## Fixed-effects: agency_id: 371,  year: 43
## Standard-errors: Clustered (agency_id)
##                      Estimate Std. Error   t value  Pr(>|t|)
## conflict           -1.787876   0.581251 -3.075913 0.0022547 **
## academic           -1.222382   0.939515 -1.301078 0.1940414
## political          -0.706021   0.697041 -1.012883 0.3117779
## private             0.659686   0.971911  0.678752 0.4977196
## public              0.832722   0.636232  1.308834 0.1914030
## council             0.381913   0.868201  0.439890 0.6602733
## board              -2.052907   0.882296 -2.326778 0.0205163 *
## creator            -0.120529   0.493695 -0.244137 0.8072598
## lni_budget_control -0.791153   0.536348 -1.475073 0.1410432
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 14.2     Adj. R2: 0.272928
##               Within R2: 0.007011
```

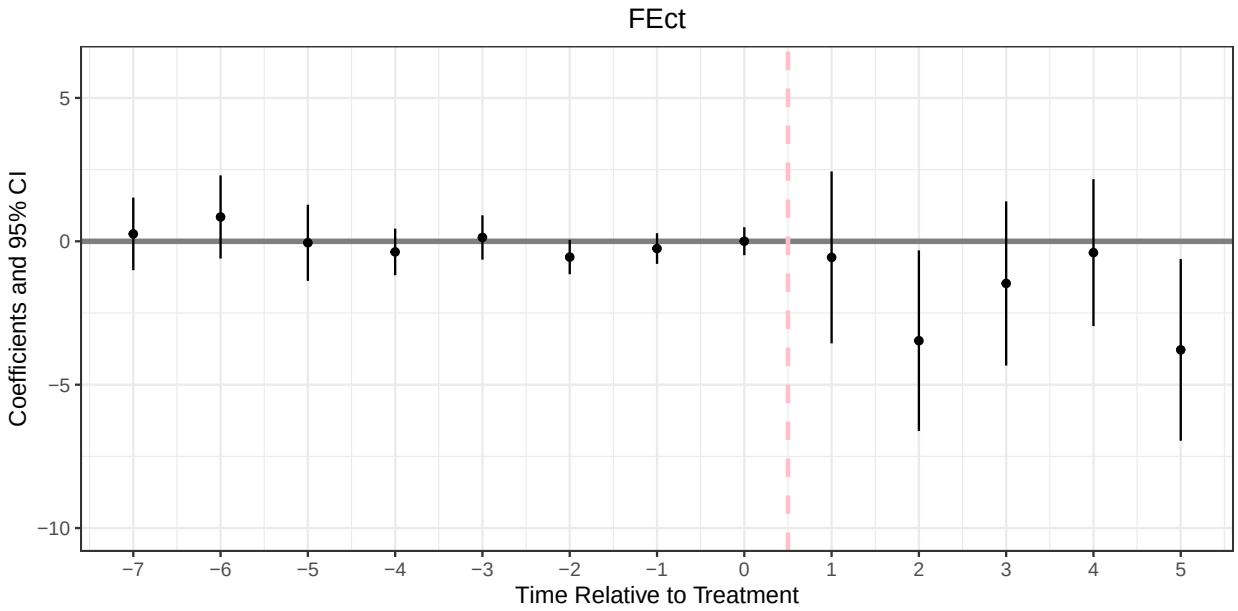


The TWFE and FEct estimator are consistent with each other.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct.

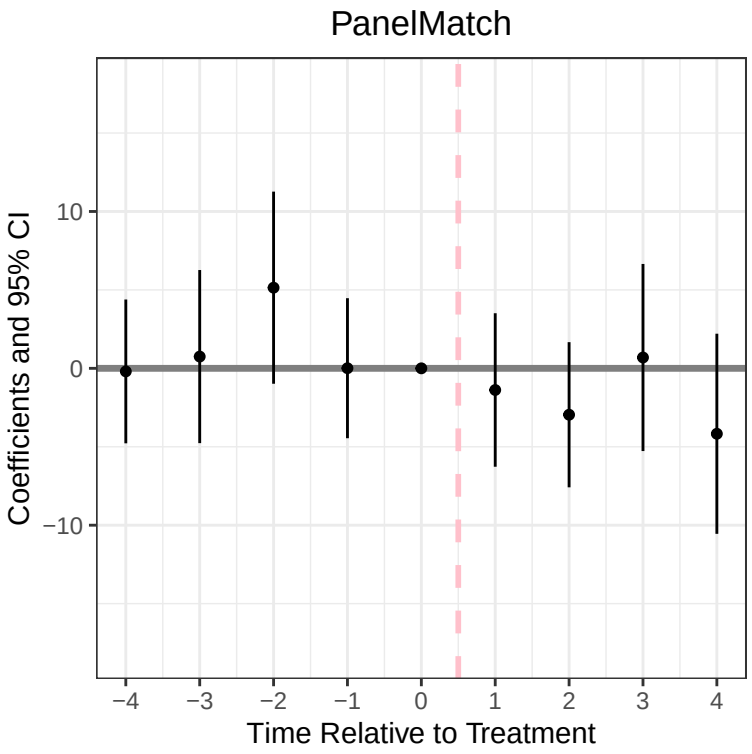




The TWFE and FEct estimator are consistent with each other. The estimated DTEs are all positive in the plot.

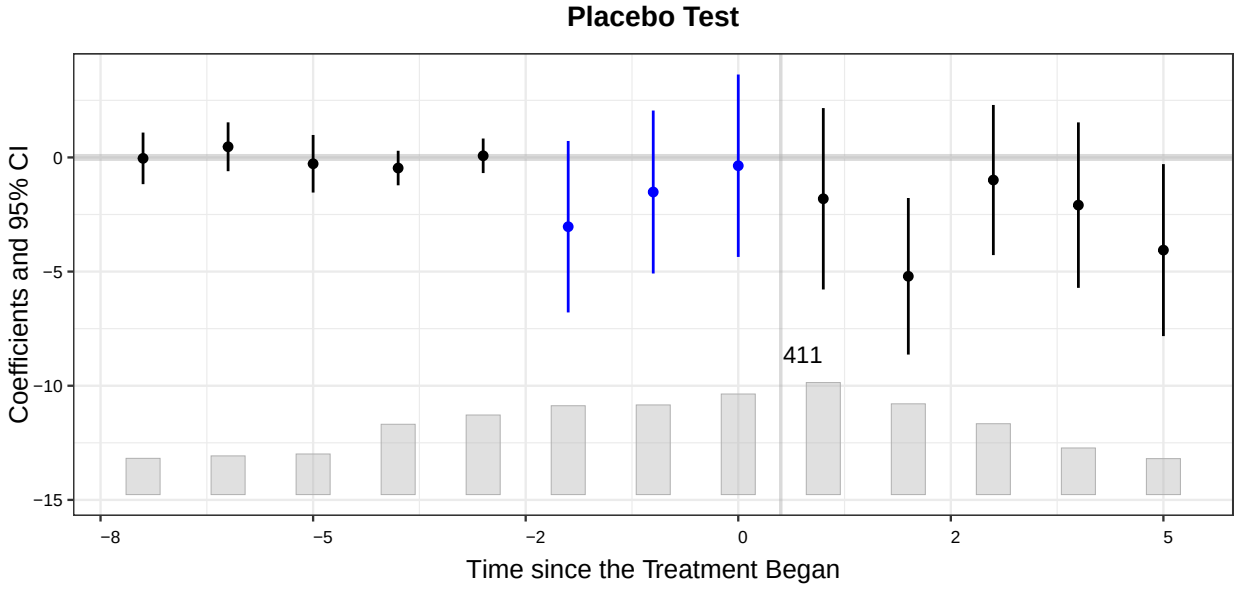
DTE of a Balanced Subsample

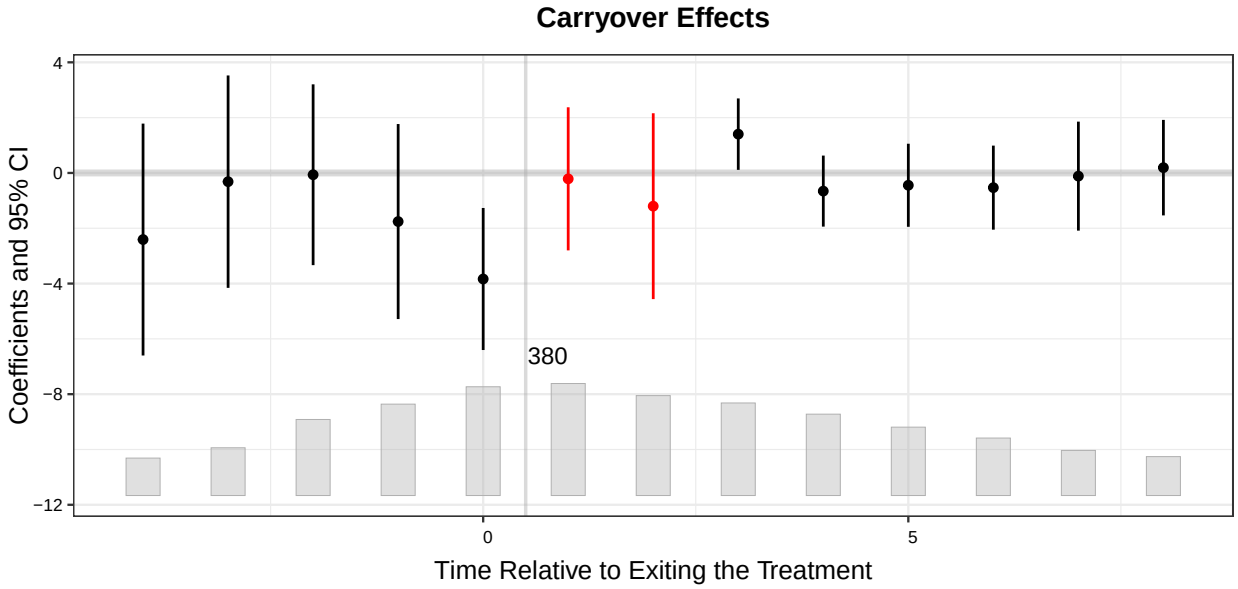
We also estimate the DTE using PanelMatch (lead = 4 and lag = 5). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





Test Statistics

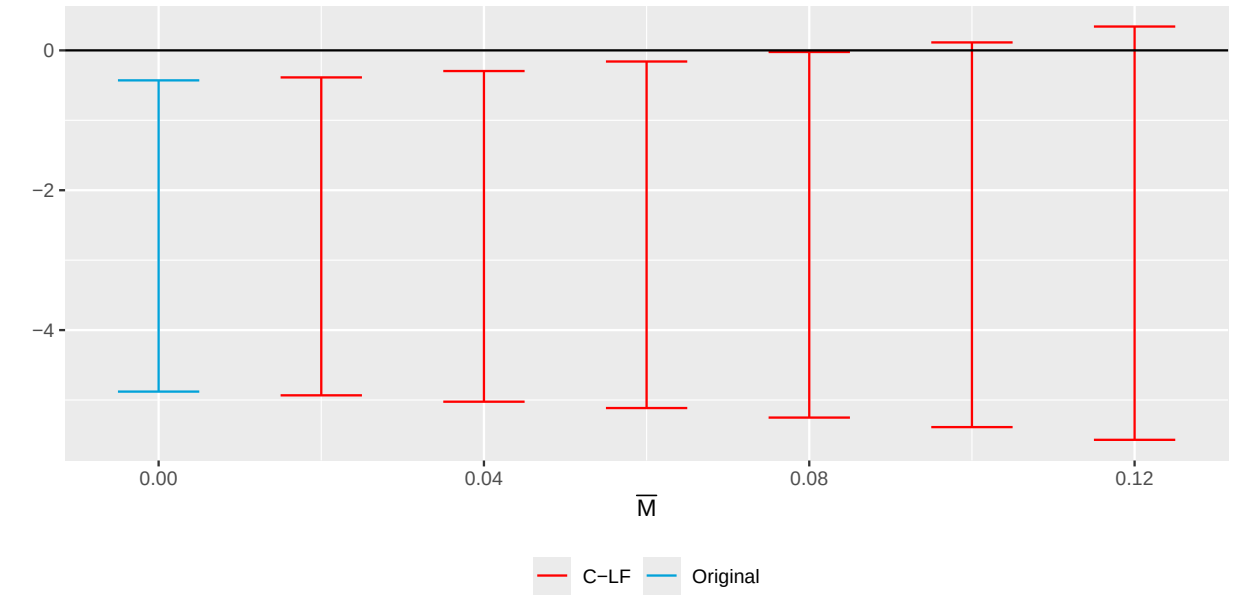
```
##                                  p-value
## F test                             0.264
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.132
## Placebo test                       0.233
## Carryover effect test              0.587
```

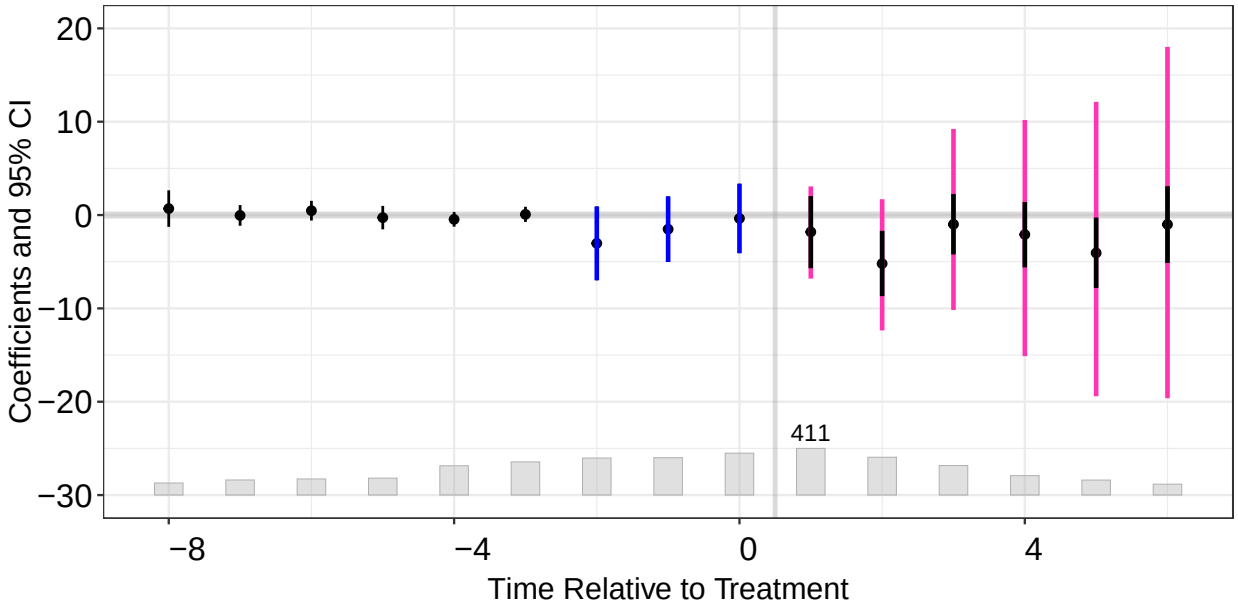
We find little evidence of violations of the parallel trend (PT) assumption or the no-carryover-effect assumption.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.08$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.08 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust when using HTE-robust estimators. We find little evidence of violations of the PT assumption. However, due to limited power, we cannot reject the null that the ATT is different from zero under mild PT violations benchmarked against the pre-treatment placebo estimates.

Dipoppa, Grossman, and Zonszein (2023)

Contents

A First Look at Data 1

Point Estimates 2

Dynamic Treatment Effects 3

Diagnostic Tests 4

Sensitivity Analysis 5

Summary 6

A First Look at Data

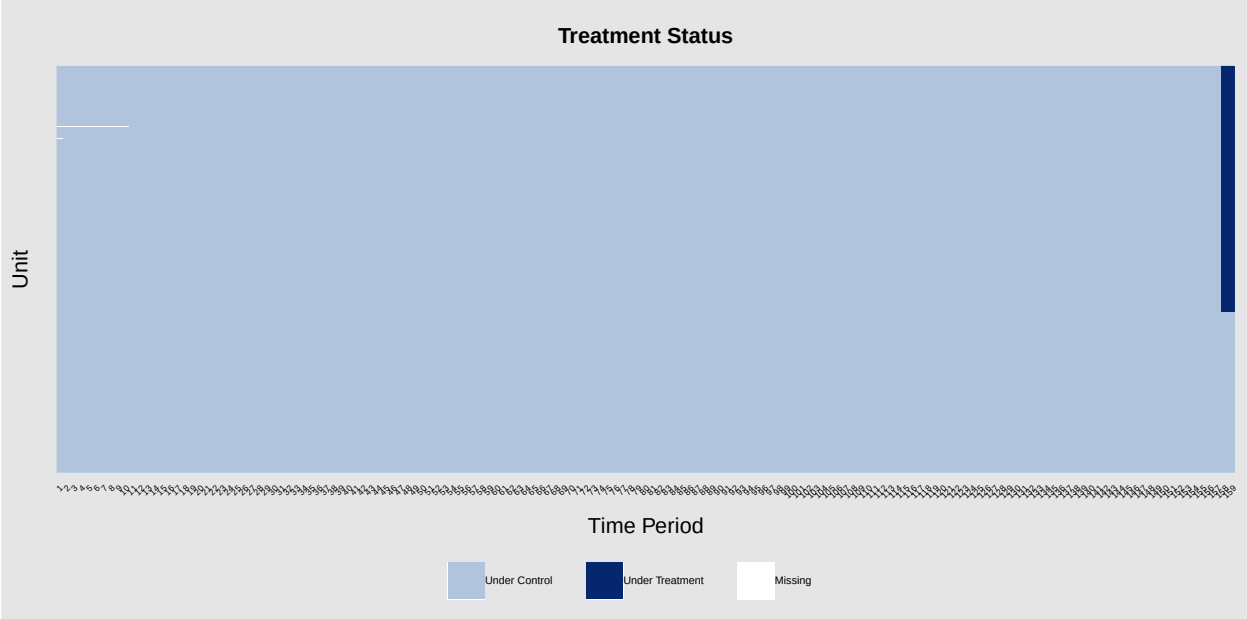
The paper investigates the effects of COVID-19 economic shocks on Asian hate crimes, using Italian municipality-month level panel data between 2007 and 2020. One of the main findings of this paper is that Asian hate crimes "significantly increase in municipalities with high expected unemployment." (p398)

Model. We focus on **Model 2, Panel B of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the municipality level.

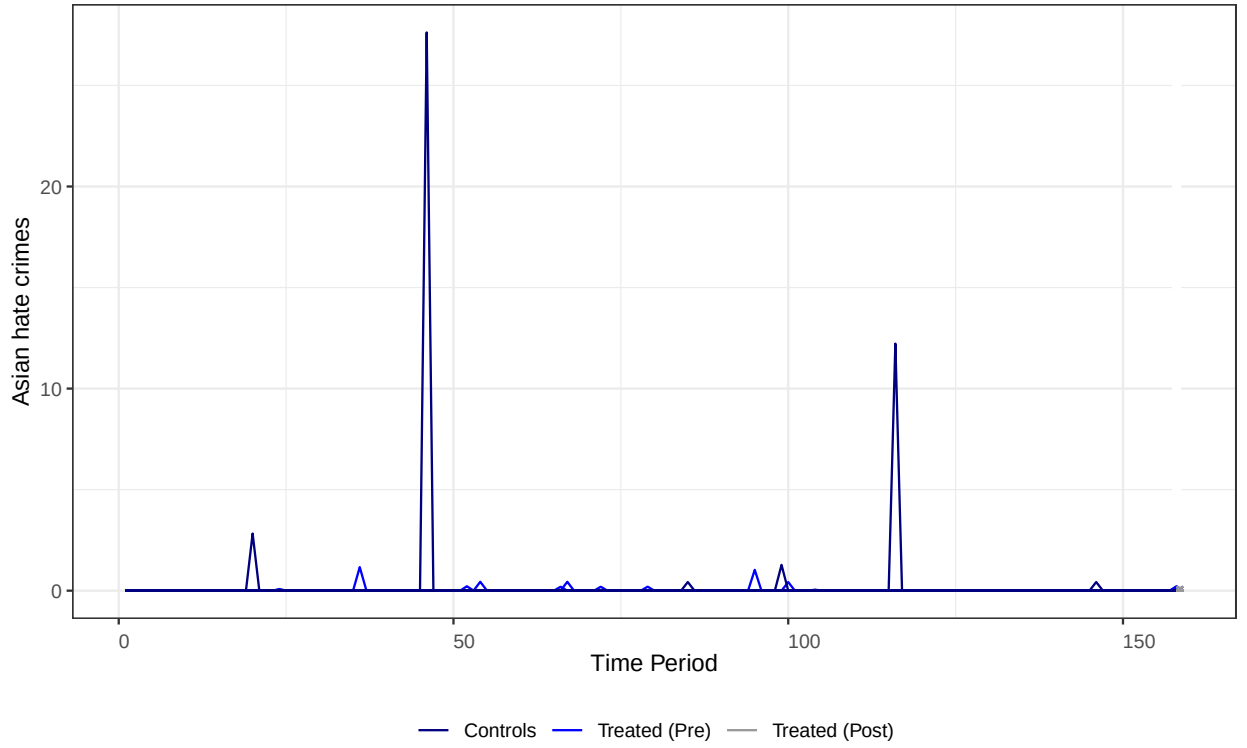
Table 1: Replication Summary

Unit of analysis	Municipality × month
Treatment	Above the median share of workers in economic sectors affected by COVID-19
Outcome	Asian hate crimes per 10,000 residents
Treatment type	Classic
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a municipality). The treatment has no reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

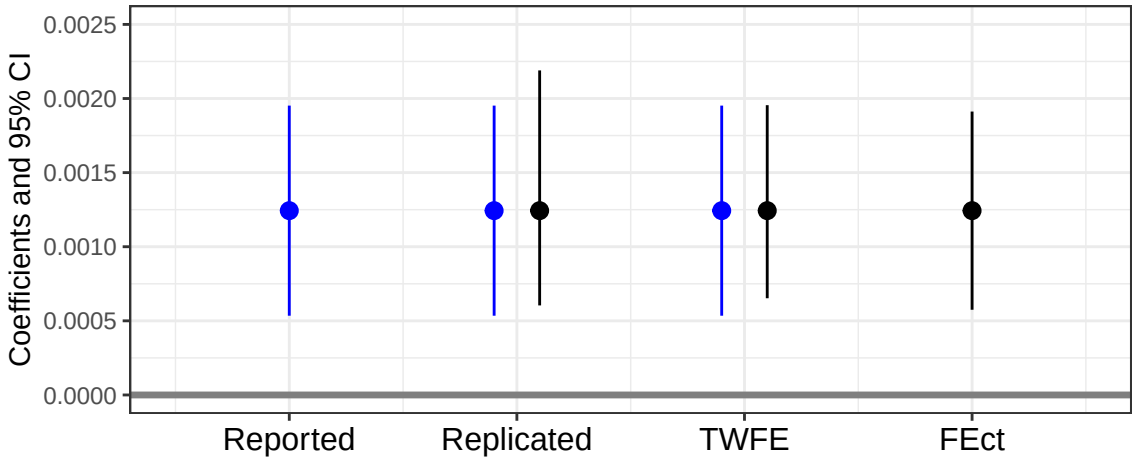
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and imputation estimator (FEct). The point estimates and

their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

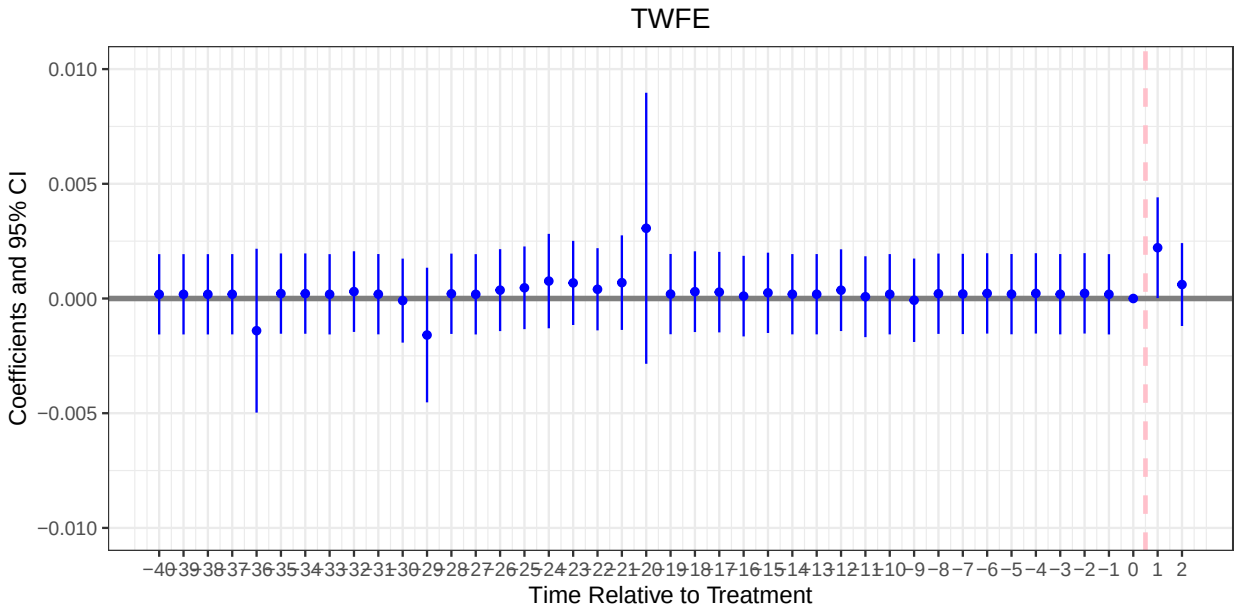
```
sol <- feols(hc_pc_asians~covid_econ|panelvar+month,
             data = df,cluster = "panelvar")
summary(sol)
```

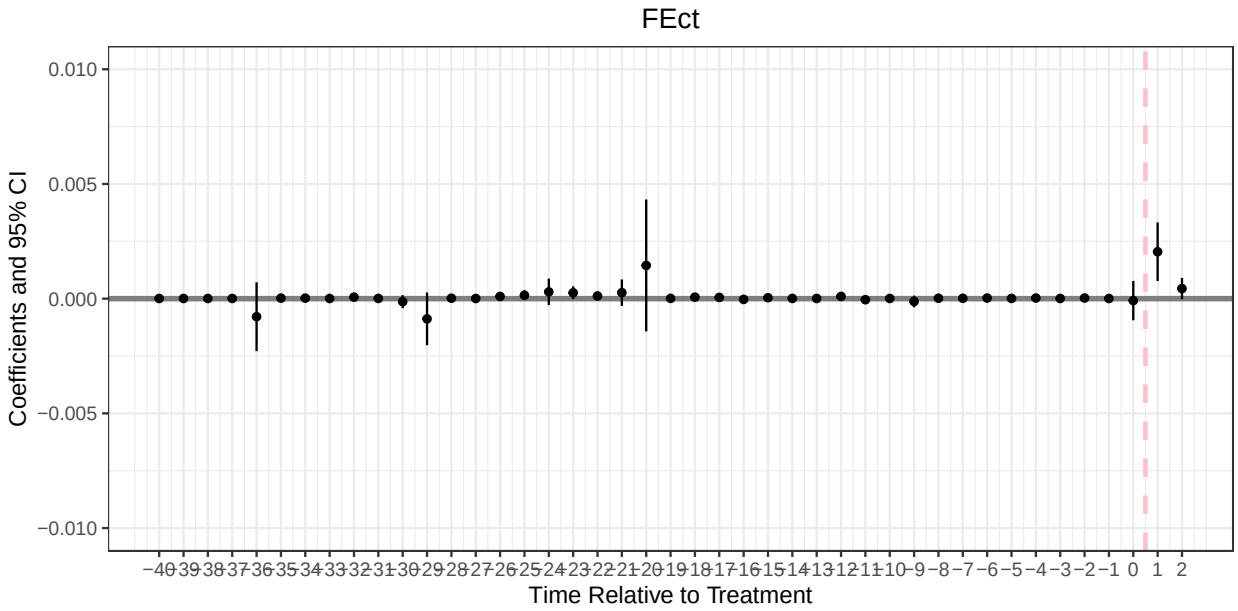
```
## OLS estimation, Dep. Var.: hc_pc_asians
## Observations: 1,259,477
## Fixed-effects: panelvar: 7,922,  month: 159
## Standard-errors: Clustered (panelvar)
##             Estimate Std. Error t value   Pr(>|t|)
## covid_econ 0.001243   0.000362 3.43859 0.00058775 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.033541     Adj. R2: 9.03e-5
##                  Within R2: 4.268e-6
```



Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct.

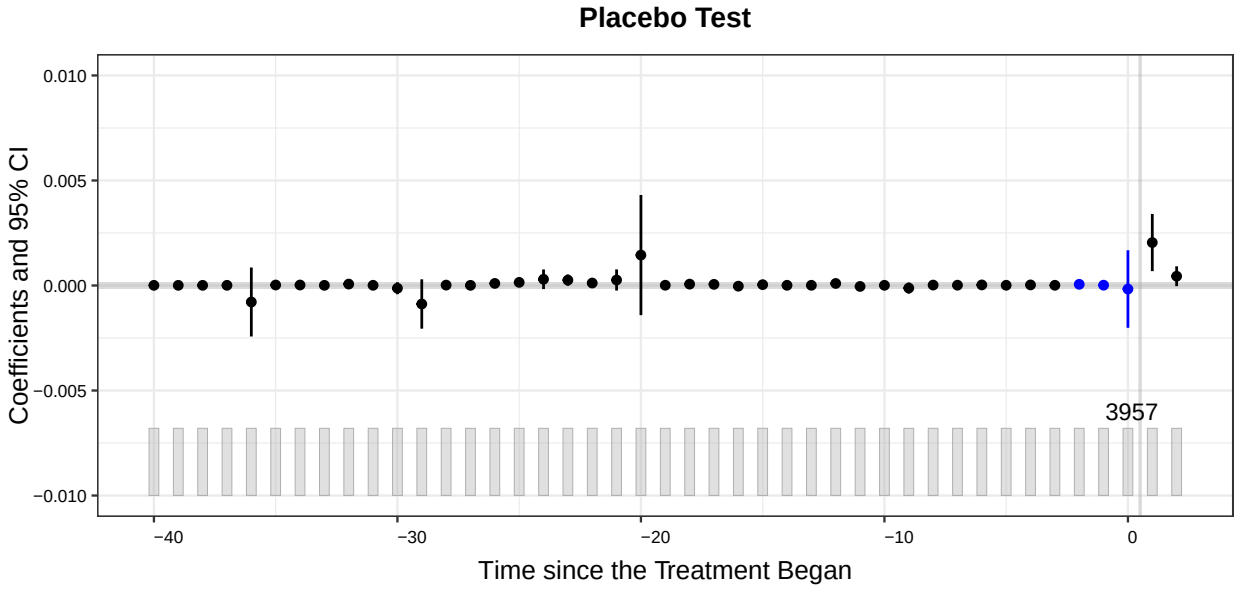




The TWFE and FEct estimator are consistent with each other.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



Test Statistics

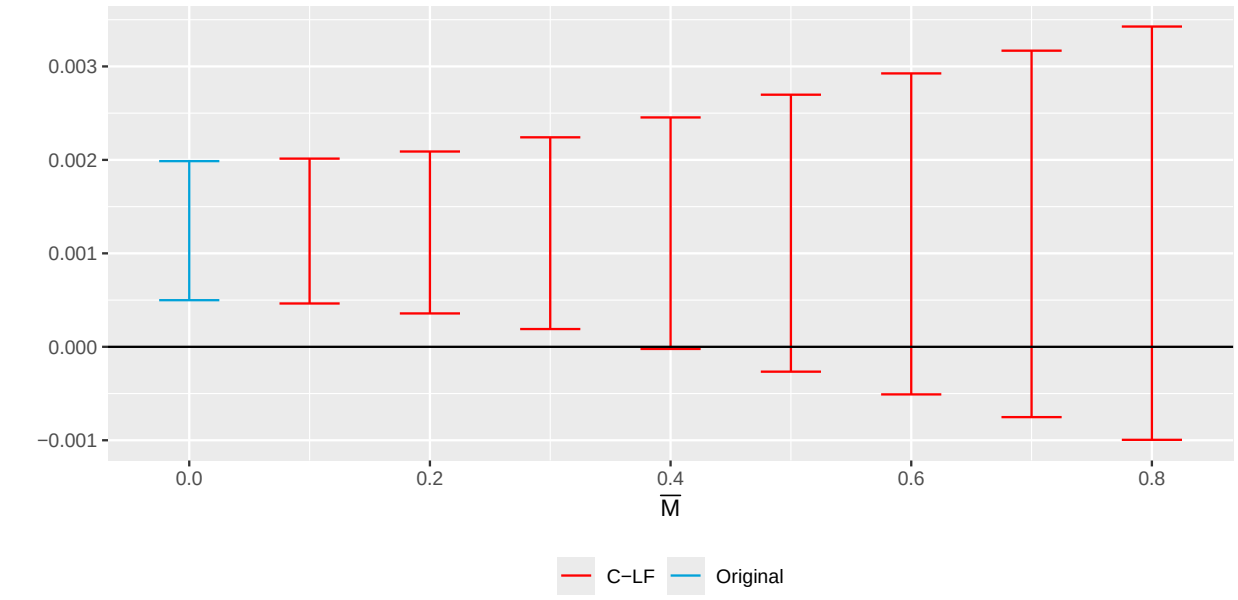
```
##                                  p-value
## F test                           0.30700
## Equivalence test (default)       0.00500
## Equivalence test (threshold=ATT) 0.00399
## Placebo test                     0.91800
## Carryover effect test                 NA
```

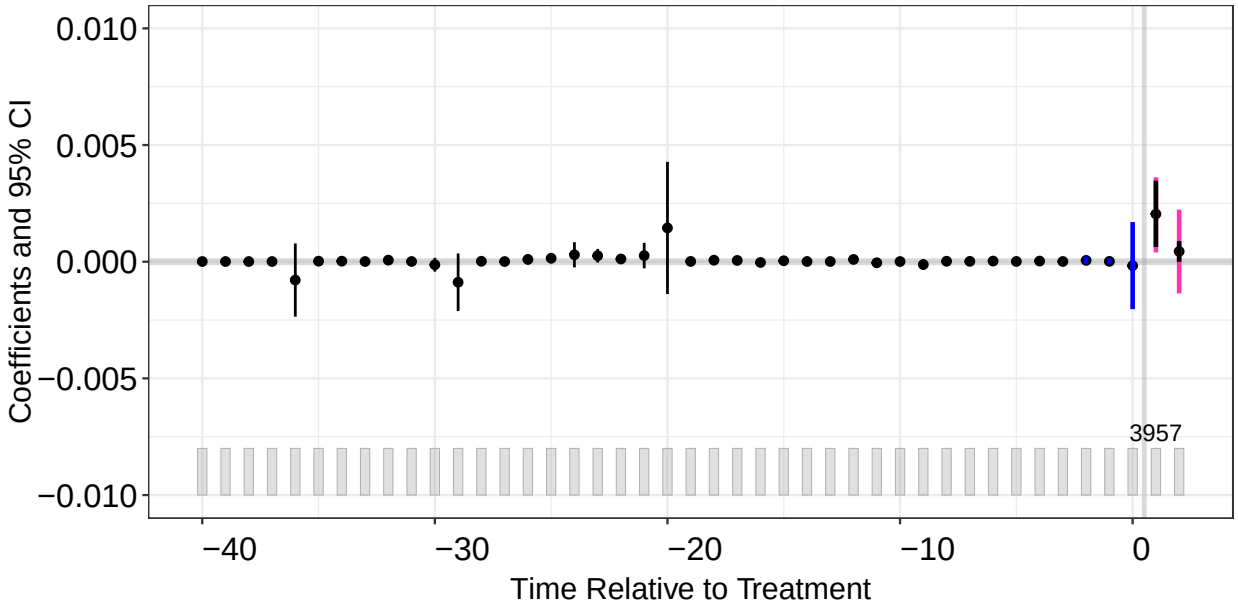
We find little evidence of violations of the parallel trend (PT) assumption. The equivalence test also rejects the null hypothesis that the residuals in the pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.4$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.4 times the maximum deviation observed between consecutive periods in the placebo periods.

We further evaluate the robustness of the estimated DTE under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, accounting for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model remain statistically significantly at the 5% level when the imputation estimator is applied. We find no evidence of potential PT violations or anticipation effects, the estimated ATT appears to be relatively robust to mild PT violations in sensitivity analysis with breakdown threshold $M = 0.4$.

Distelhorst and Locke (2018)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

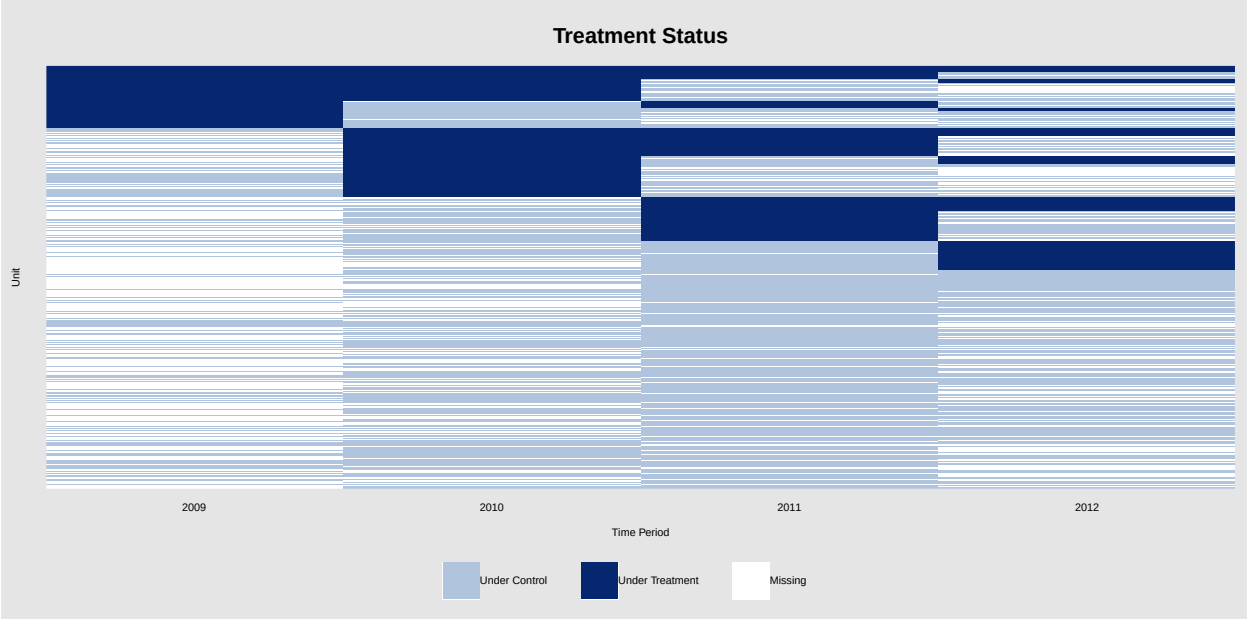
A First Look at Data

The paper investigates the effects of compliance on order value, using factory-year level panel data from 36 countries during 2009-2012. One of the main findings of this paper is that "achieving compliance is associated with a 4% [1%, 7%] average increase in annual purchasing" (p695).

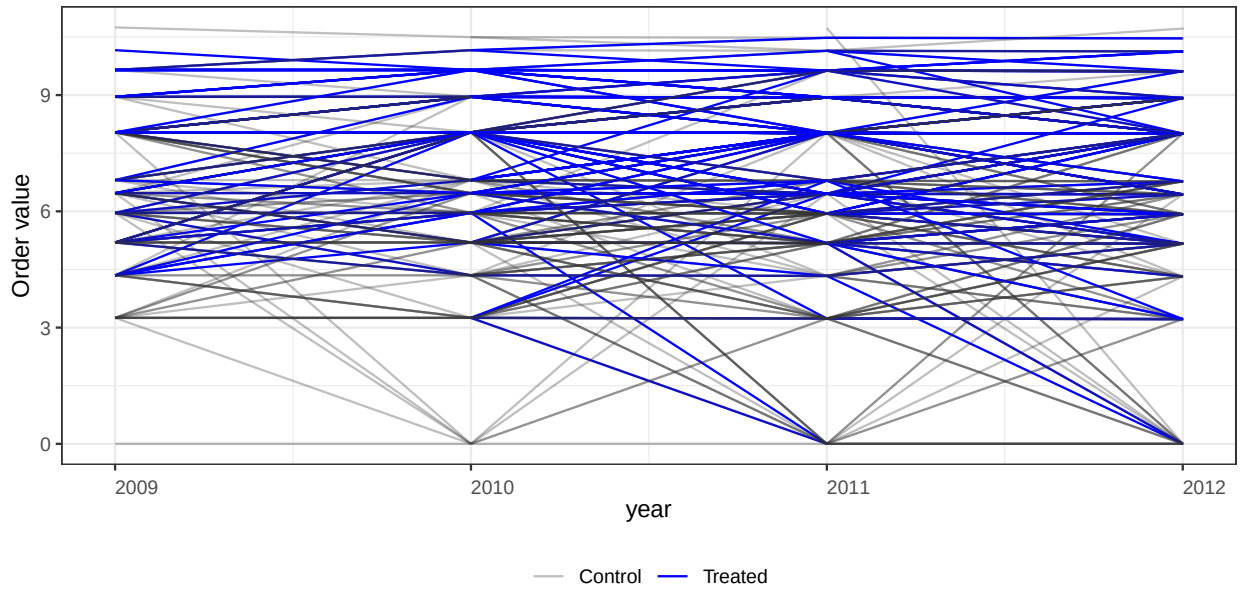
Model. We focus on **Model 1 of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the factory level.

Replication Summary	
Unit of analysis	Factory × year
Treatment	Compliance
Outcome	Order value
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a factory). We see that around half of units are treated at various time points and there are treatment reversals. There are some missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each factory. The observations under treated status are marked in blue.



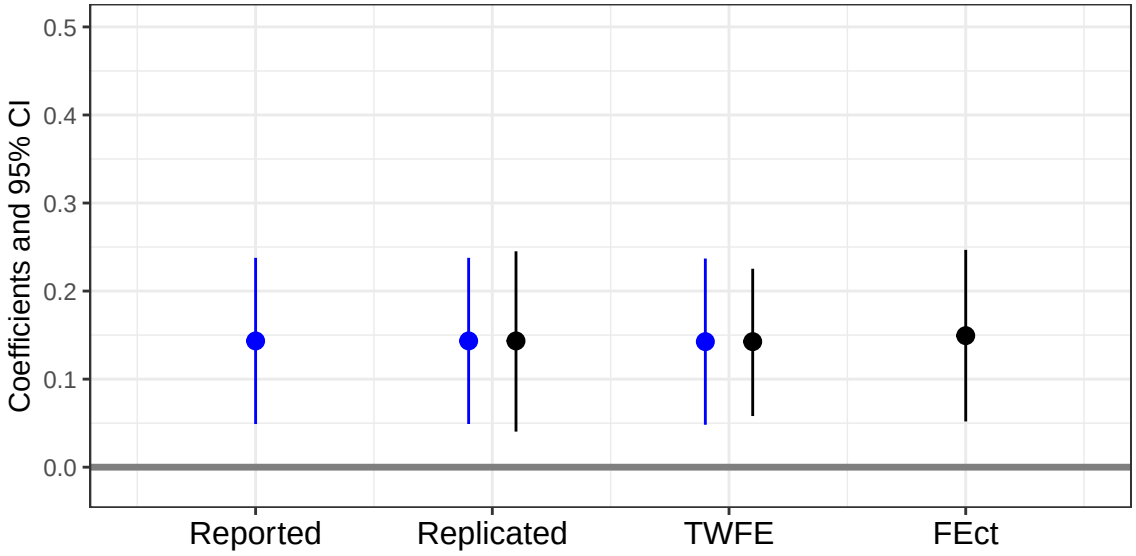
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and imputation estimator (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

```r
sol <- feols(lestspend_dall~compyr|fcode+year,
             data = df,cluster = "fcode")
summary(sol)
```

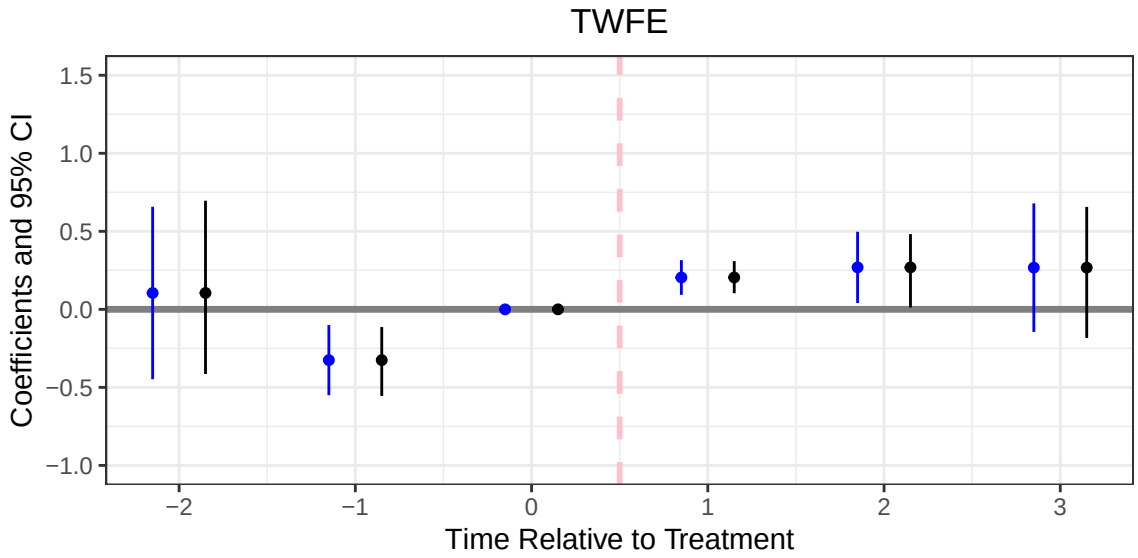
```
## OLS estimation, Dep. Var.: lestspend_dall
## Observations: 6,915
## Fixed-effects: fcode: 2,447,  year: 4
## Standard-errors: Clustered (fcode)
##         Estimate Std. Error t value  Pr(>|t|)
## compyr 0.143383   0.048141 2.97837 0.0029264 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.05863     Adj. R2: 0.61391
##                  Within R2: 0.001831
```

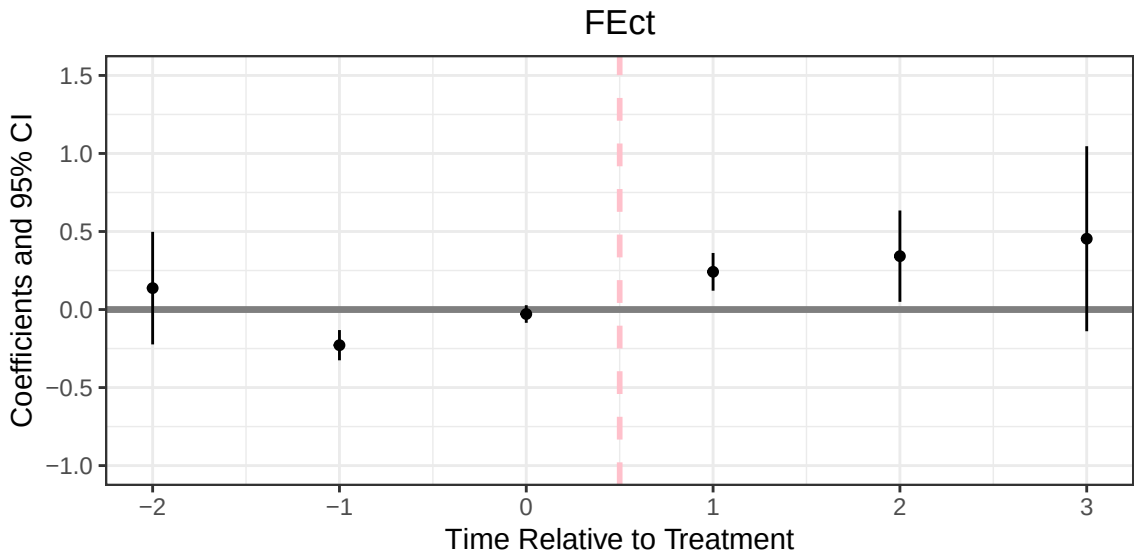


TWFE and FEct give similar results. The estimated ATT are all positive and statistically significant.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

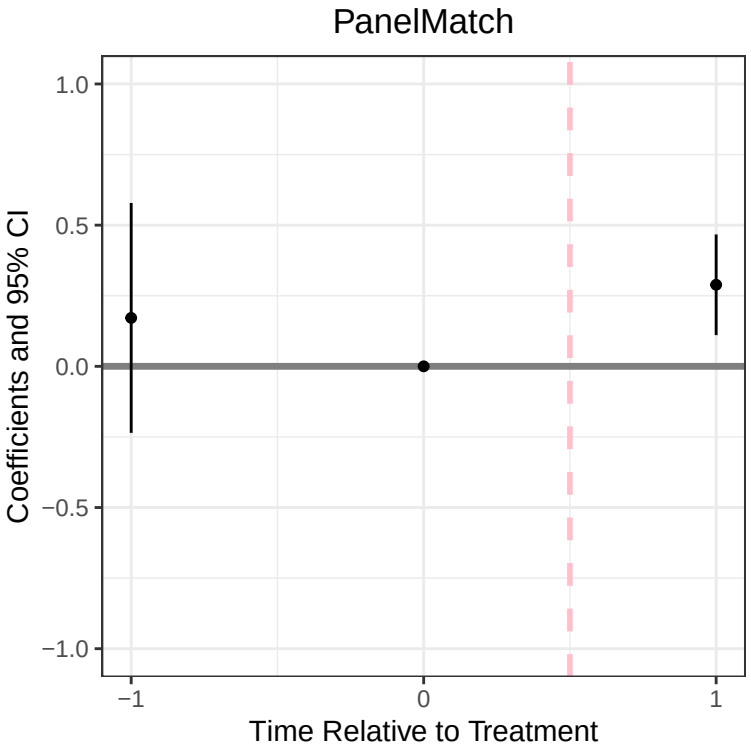




The estimated DTEs given by TWFE and FEct are similar. They exhibit upward trends on post-treatment periods in the plot.

DTE of a Balanced Subsample

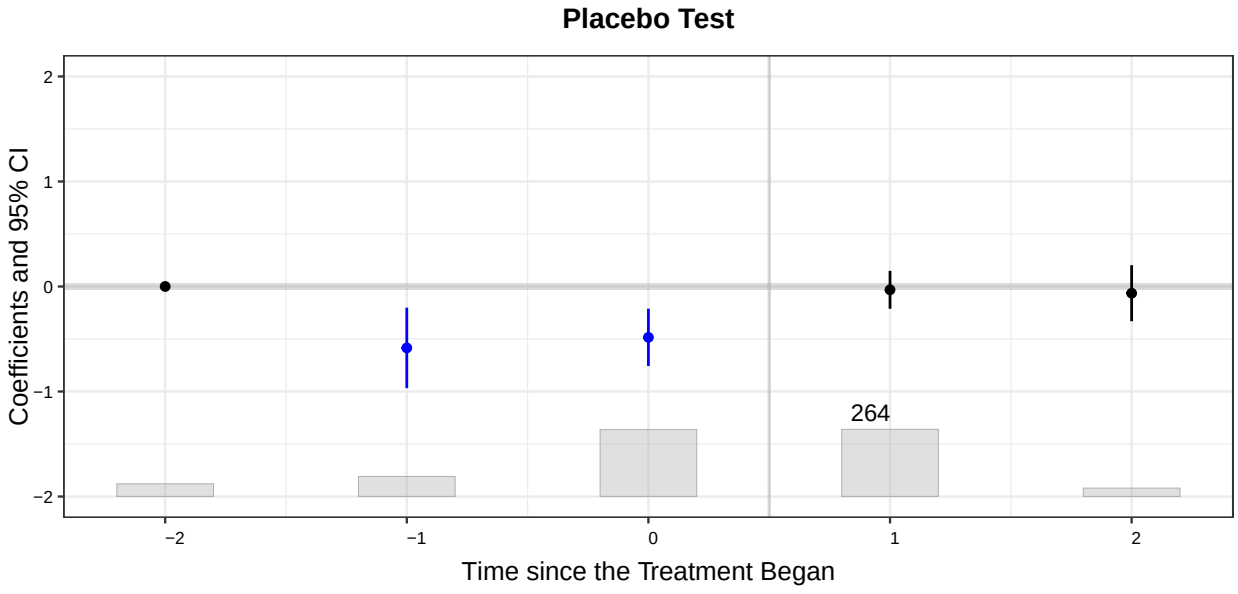
We also estimate the DTE using PanelMatch (lead = 1 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.

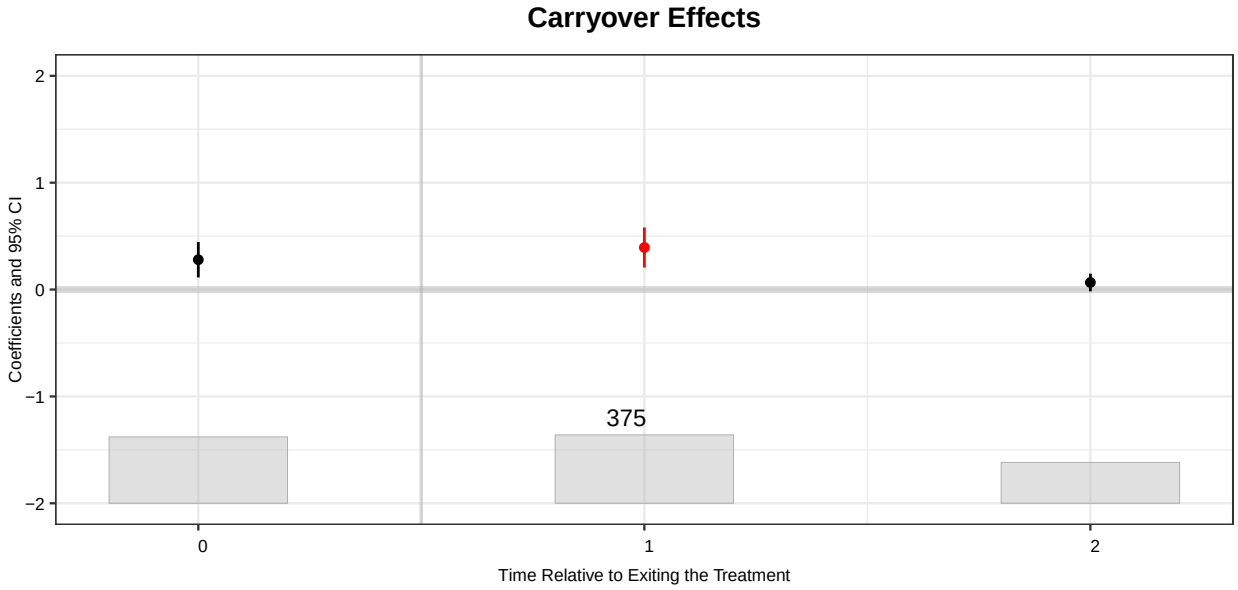


PanelMatch and FEct estimates are broadly consistent with each other on the post-treatment period, though they are different on the pre-treatment periods.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





Test Statistics

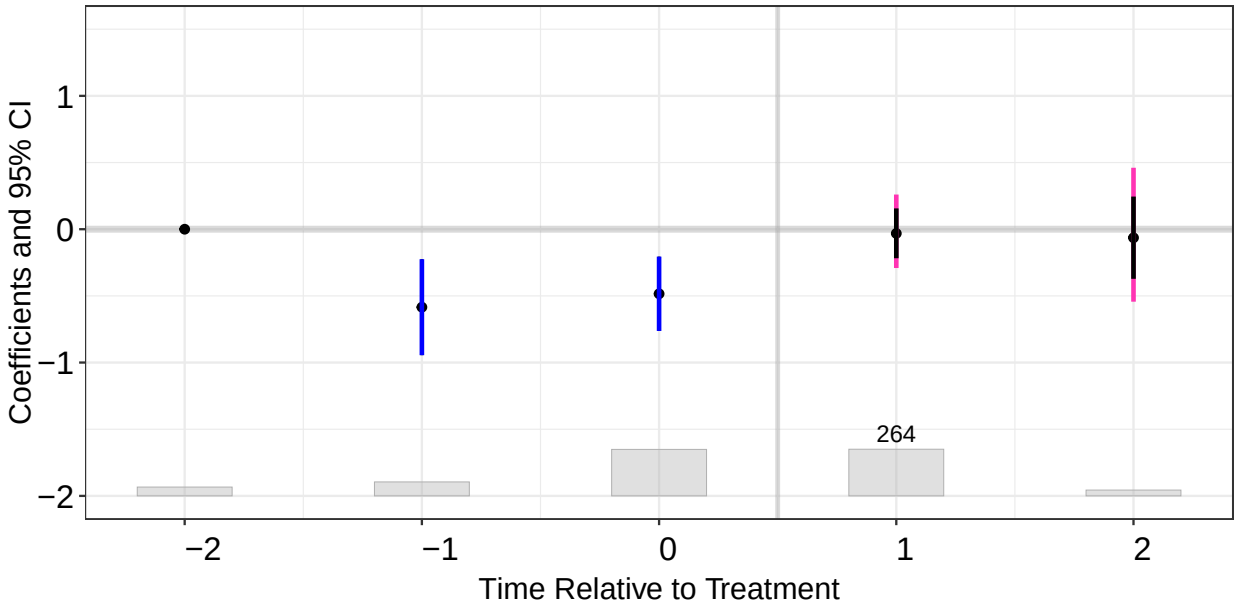
```
## Cannot use full pre-treatment periods in the F test. The first period is removed.
```

```
##                                   p-value
## F test                                  0
## Equivalence test (default)              0
## Equivalence test (threshold=ATT)        0
## Placebo test                            0
## Carryover effect test                   0
```

We find evidence for violations of the parallel trend (PT) assumption using F-test and the placebo test rejects the null. The rejection in the carryover test also shows violations of the no-carryover-effect assumption.

Sensitivity Analysis

After holding out two pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Both TWFE and FEct yield similar positive and statistically significant point estimates. A carryover effect is found in the FEct carryover effect test. Visually, both TWFE and HTE-robust estimators show some evidence of potential violations of the PT assumption. The estimated ATT is not robust to the removal of placebo periods and potential violations of the PT assumption benchmarked against the pre-treatment placebo estimates.

Eckhouse (2022)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 6
Sensitivity Analysis 6
Summary 7

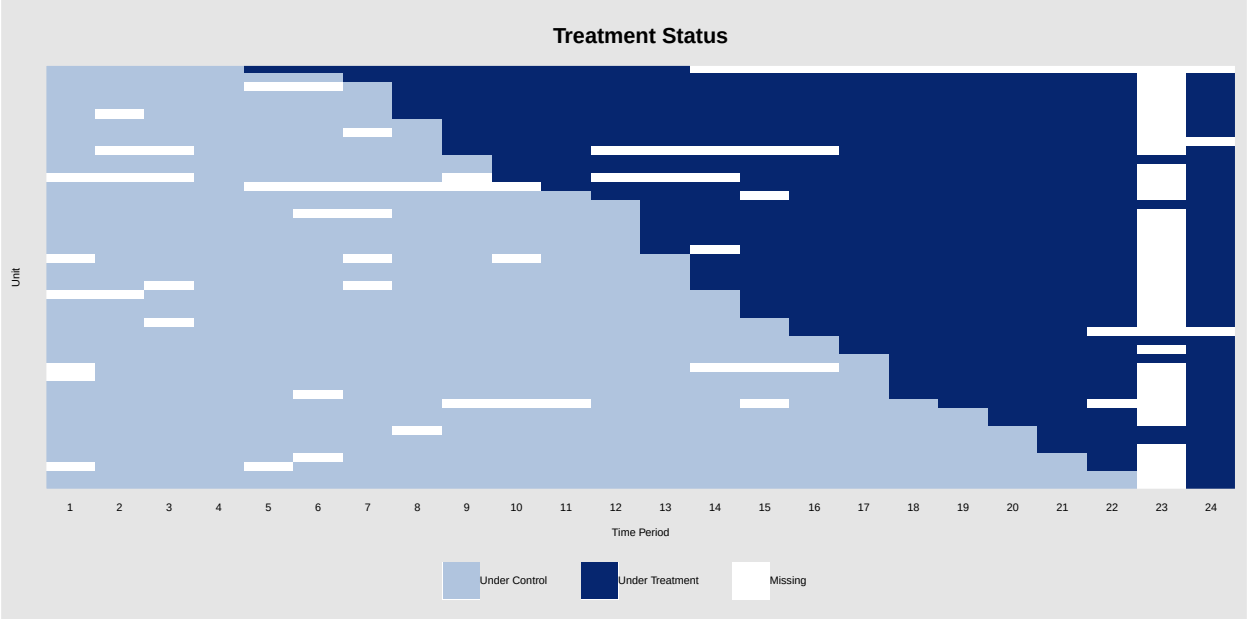
A First Look at Data

The paper investigates the effects of metrics management on policing, using US city-year level panel data from between 1990 and 2013. One of the main findings of this paper is that "adopting CompStat is associated with … an increase in 1.3 percentage points in the share of Part 2 arrests" (p715).

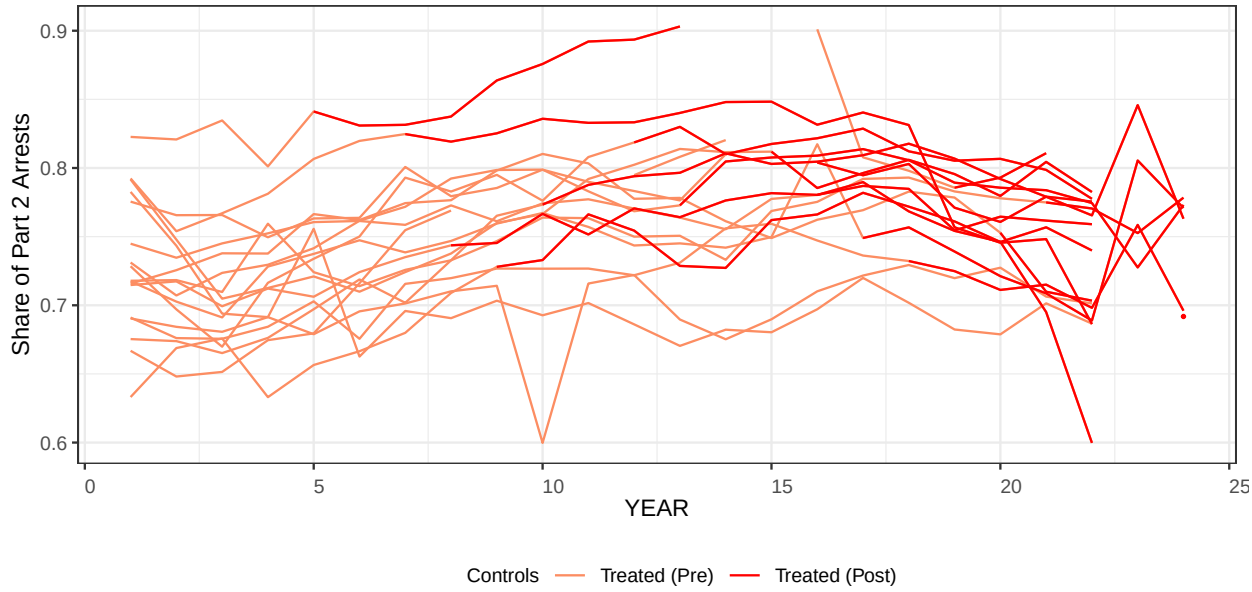
Model. We focus on **Model 1 of Table 4** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors (not clustering).

Replication Summary	
Unit of analysis	City × year
Treatment	Metrics management
Outcome	Share of Part 2 Arrests
Treatment type	Staggered
Outcome type	Continuous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a city).



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units (there is none in this data) and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DIDM estimator, and imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent

confidence intervals (CIs) based on cluster-robust SEs (for the "reported" ATT, we report the robust SEs) and cluster-bootstrapped CIs, respectively. Note that the point estimates are no longer statistically significant at the 5% level since in the original paper the authors do not use cluster-robust SEs.

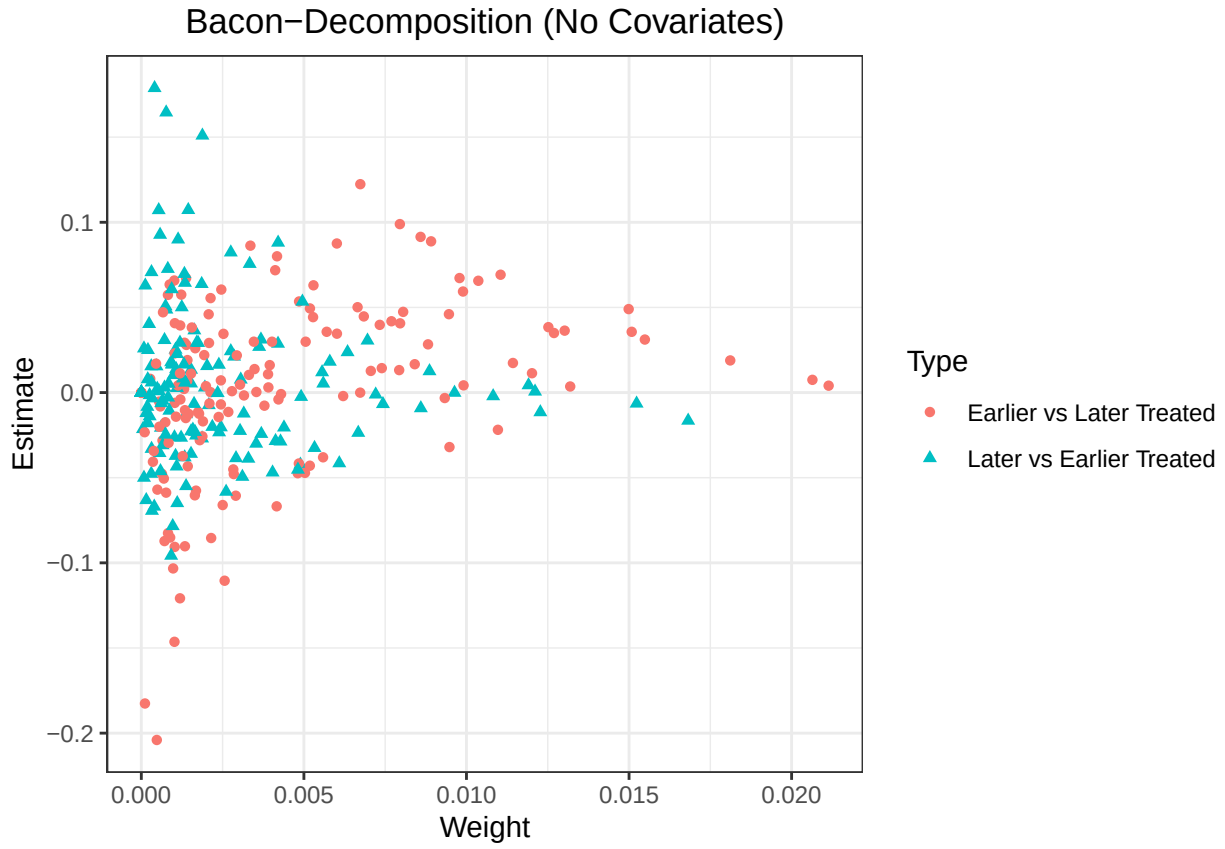
Original Results

```
sol <- feols(SHAREPT2~HASCOMPSTAT|AGENCY+YEAR,data = df,vcov = "hetero")
summary(sol)
```

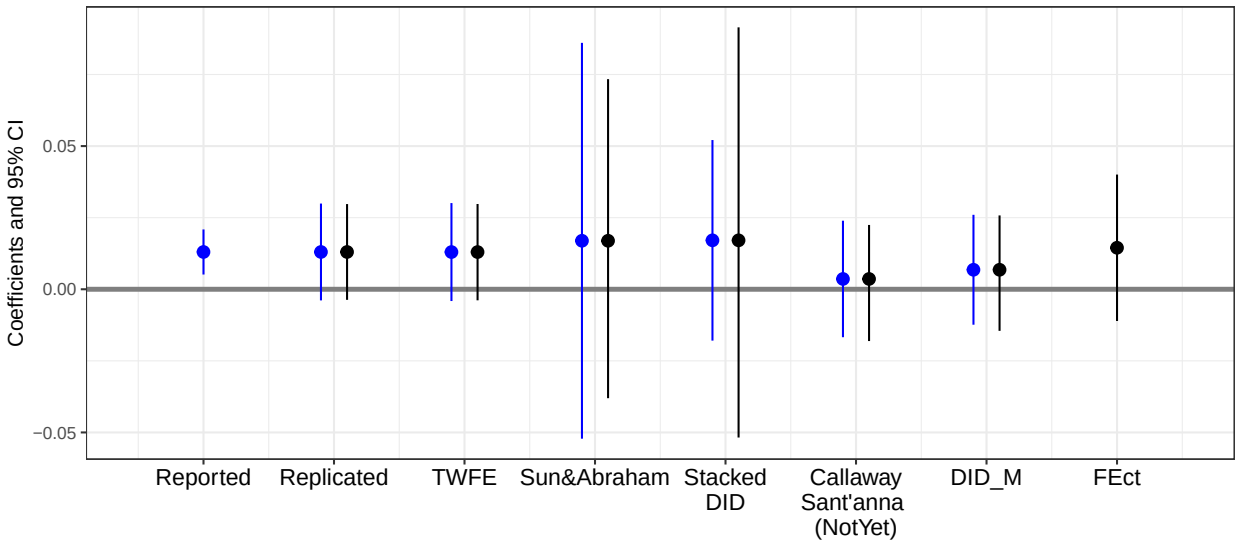
```
## OLS estimation, Dep. Var.: SHAREPT2
## Observations: 1,023
## Fixed-effects: AGENCY: 47,  YEAR: 24
## Standard-errors: Heteroskedasticity-robust
##              Estimate Std. Error t value  Pr(>|t|)
## HASCOMPSTAT  0.01302   0.004016 3.24197 0.0012283 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.037664     Adj. R2: 0.644087
##                  Within R2: 0.008948
```

Goodman-Bacon Decomposition

As there is no never-treated nor always-treated units, we only have the Earlier vs Later Treated and Later vs Earlier Treated comparisons in the Goodman-Bacon decomposition.



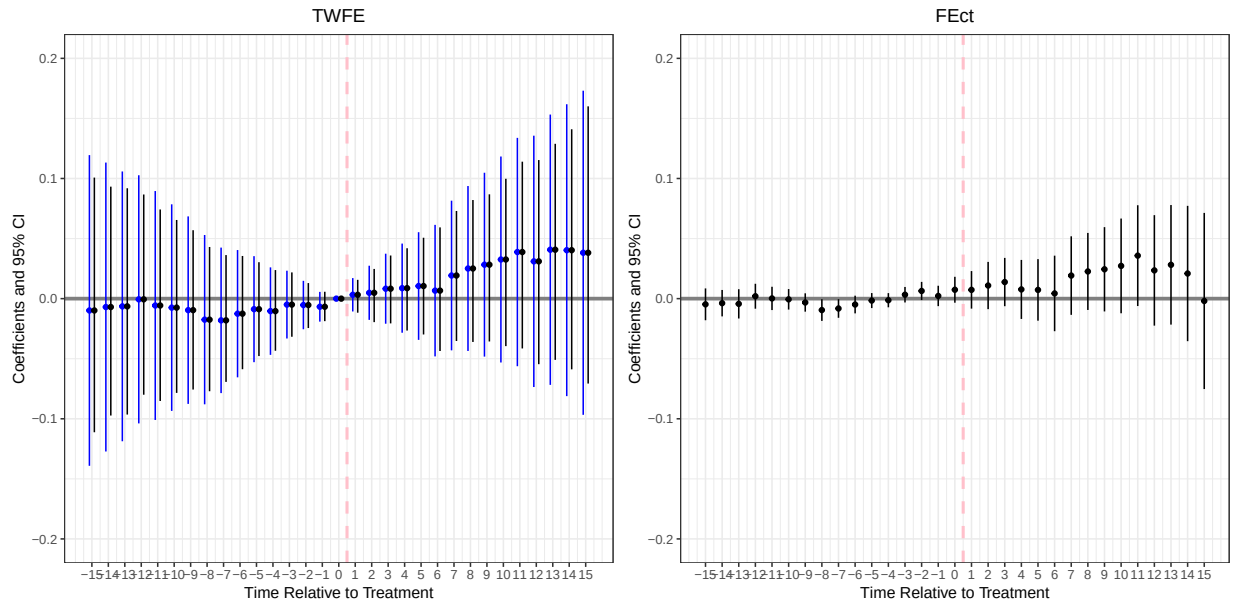

```
## [1] "baseline finished"
##   |                                                                      |
```

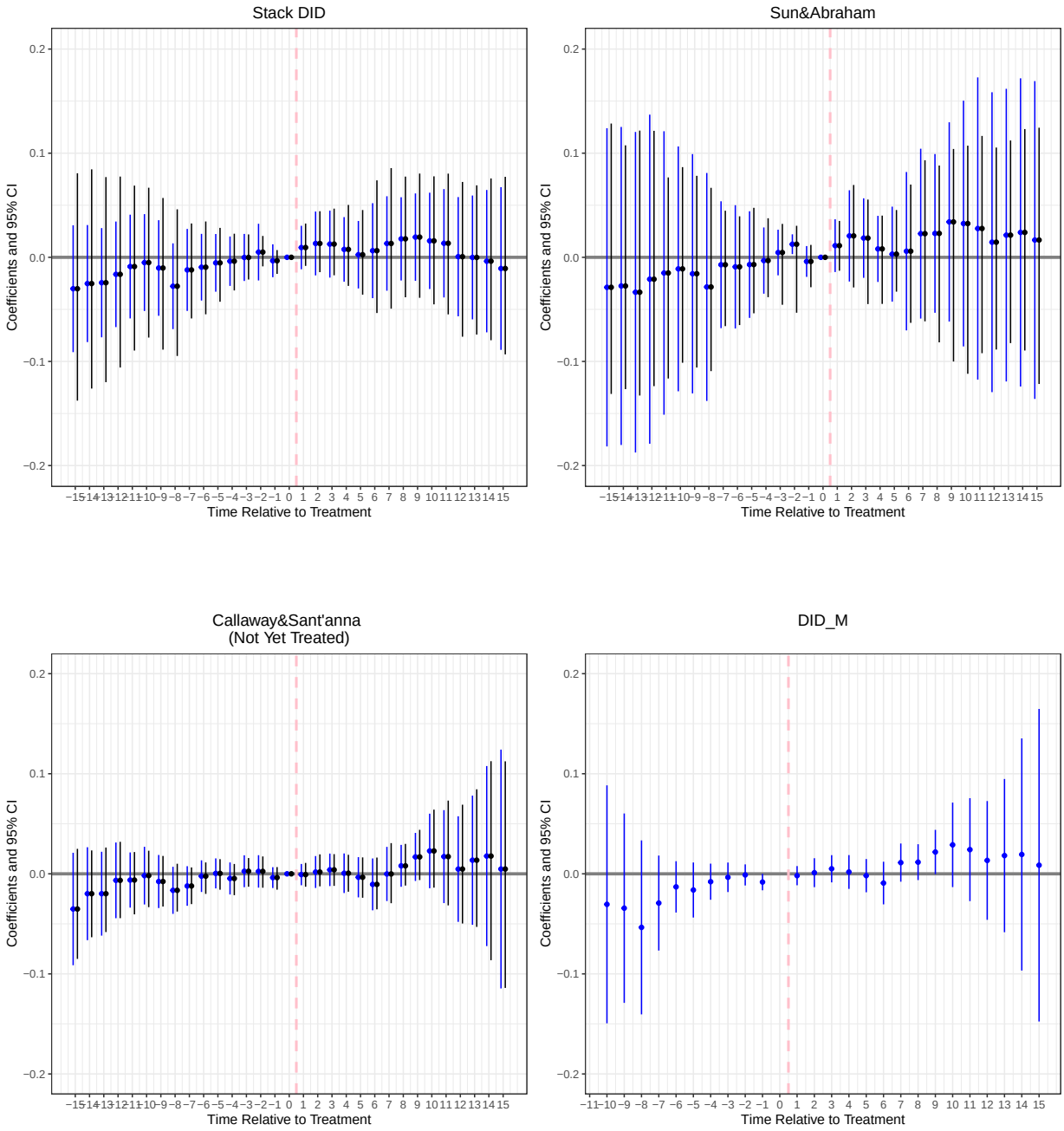


The statistical significance of the TWFE estimate diminishes when using cluster-robust SEs or cluster-bootstrap SEs. All HTE-robust estimators yield estimated ATT that are positive but statistically insignificant.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DIDM, and FEct. As TWFE, IW and Stacked DID estimators need to use the never-treated units as the control or reference group, we drop the observations in the last two periods so the units treated at the 22nd periods can serve as the never-treated units. The results are shown below.





The estimated DTEs given by TWFE and HTE-robust estimators have similar shapes in post-treatment periods. For the TWFE, Stacked DID and IW estimators, the CIs are wide during the earlier pre-treatment periods as well as the later post-treatment periods. This can be attributed to the fact that these three estimators utilize the units treated in the 22nd period, which have a relatively small sample size, as the reference group.

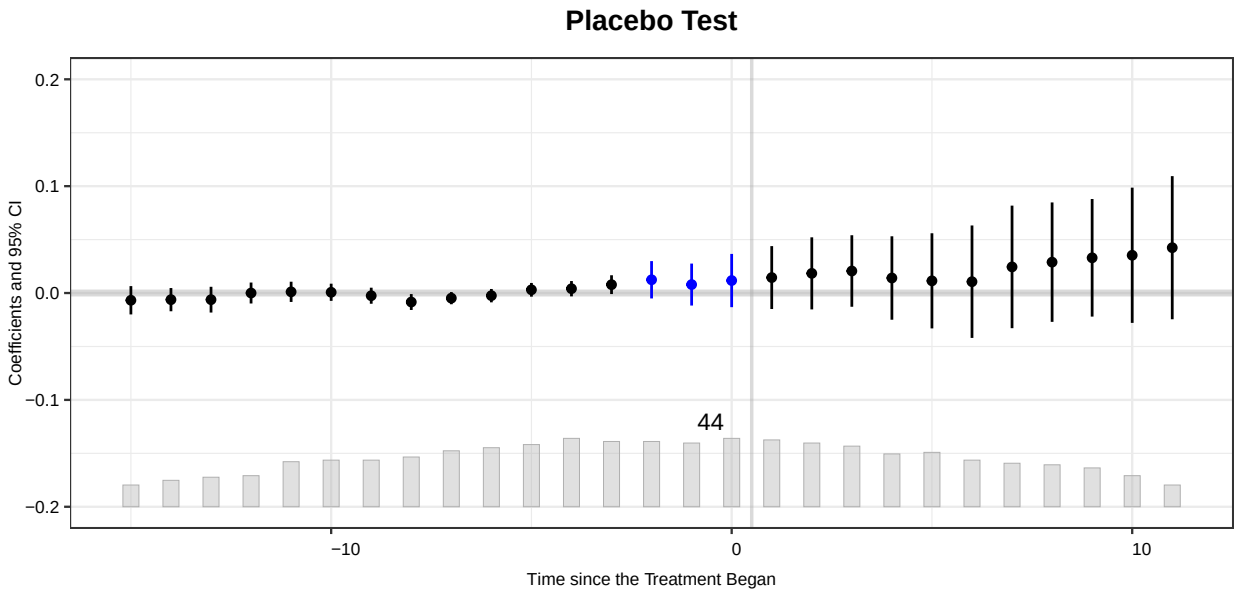
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 4 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate

time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



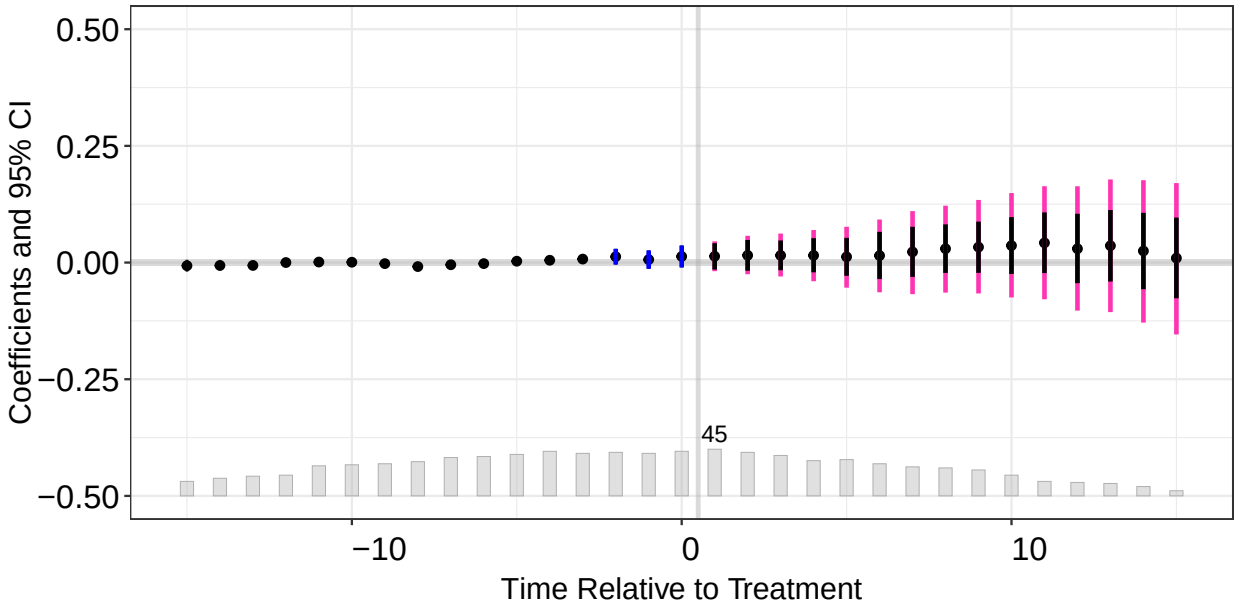
Test Statistics

```
##                                  p-value
## F test                             0.395
## Equivalence test (default)         0.284
## Equivalence test (threshold=ATT)   0.140
## Placebo test                       0.284
## Carryover effect test                 NA
```

We find little evidence of violations of the parallel trend (PT) assumption using the F-test and the placebo test. However, the equivalence test fails to reject the null hypothesis that the residuals in pre-treatment periods exceed the estimated ATT, possibly due to limited power.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

We do not find strong evidence for violations of the PT using statistical tests but the estimates are no longer statistically significant once we use cluster-robust SEs or cluster-bootstrap SEs. We also cannot reject the null that the ATT is different from zero under mild PT violations benchmarked against the pre-treatment placebo estimates.

Esberg and Siegel (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 6
Sensitivity Analysis 7
Summary 8

A First Look at Data

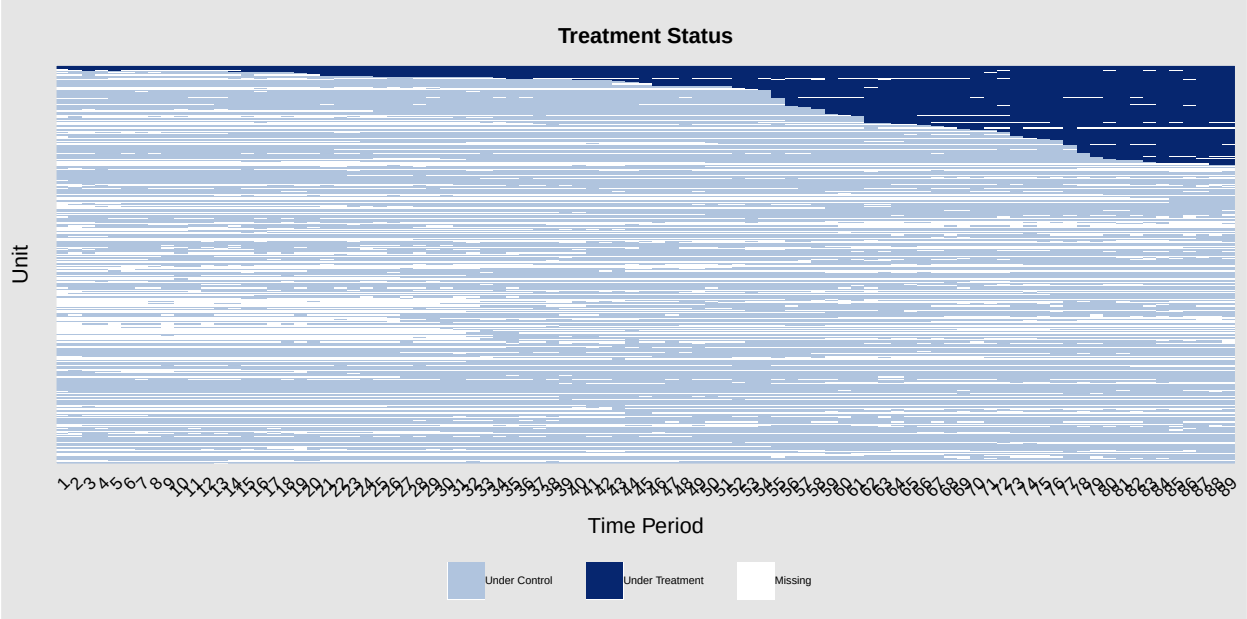
The paper investigates the effects of exile on activists' public discourse, using Venezuela activists' tweet account-month level panel data between 2013 and 2020. One of the main findings of this paper is that exile " significantly increases Twitter discussions of foreign policy solutions to Venezuela crisis and harsh criticisms of the regime." (p398)

Model. We focus on **Model 1 of Figure 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the account level.

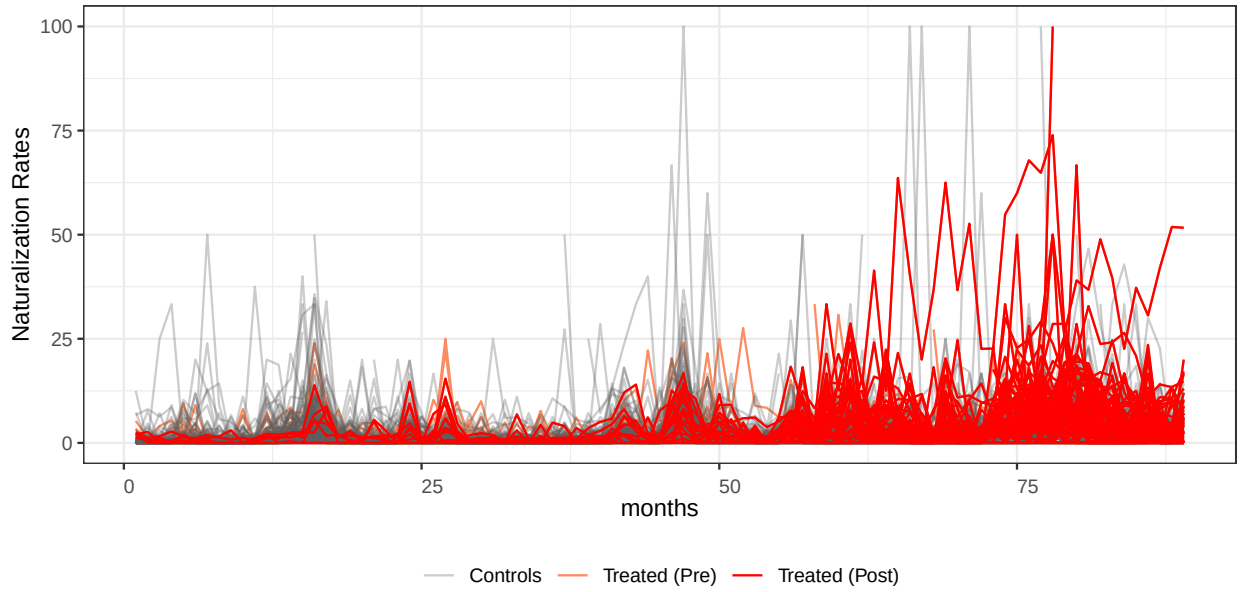
Table 1: Replication Summary

Unit of analysis	Twitter account × month
Treatment	Exile
Outcome	Percentage of posts about foreign policy
Treatment type	Staggered
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a month) and each row represents a unit.



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units (there is none in this data) and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DIDM estimator, and imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent

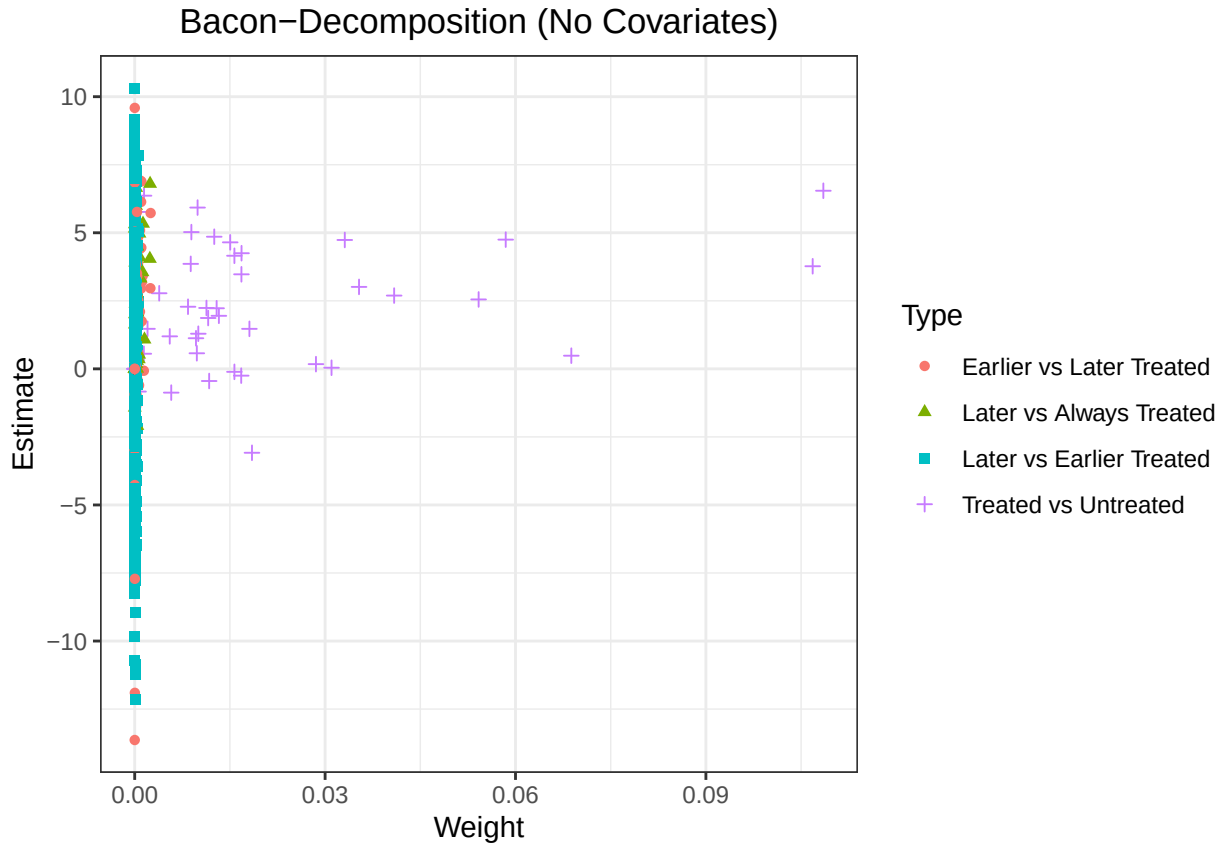
confidence intervals (CIs) based on cluster-robust SEs (for the "reported" ATT, we report the robust SEs) and cluster-bootstrapped CIs, respectively. Note that the point estimates are no longer statistically significant at the 5% level since in the original paper the authors do not use cluster-robust SEs.

Original Finding

```
sol <- feols(perc_foreign_policy~tweeted_exile + log_tweets |actorid+months,
             data = df,cluster = "actorid")
summary(sol)
```

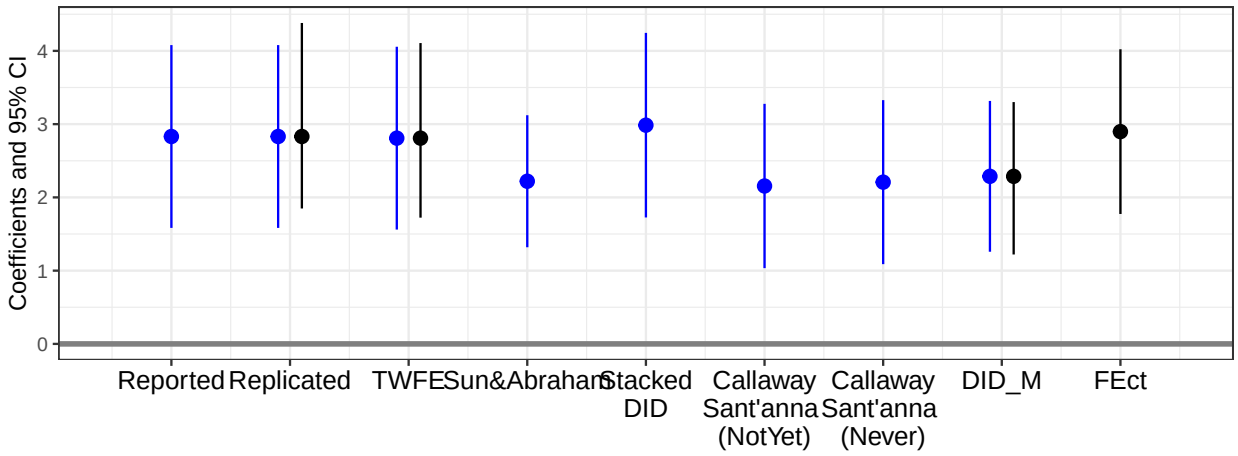
```
## OLS estimation, Dep. Var.: perc_foreign_policy
## Observations: 22,491
## Fixed-effects: actorid: 355,  months: 89
## Standard-errors: Clustered (actorid)
##                Estimate Std. Error t value   Pr(>|t|)
## tweeted_exile 2.831351   0.636670 4.44713 1.1666e-05 ***
## log_tweets    0.125293   0.046303 2.70593 7.1411e-03 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 3.75237     Adj. R2: 0.258946
##                  Within R2: 0.023422
```

Goodman-Bacon Decomposition




```
##   |                                                                      |
```

```
##   |                                                                      |
```

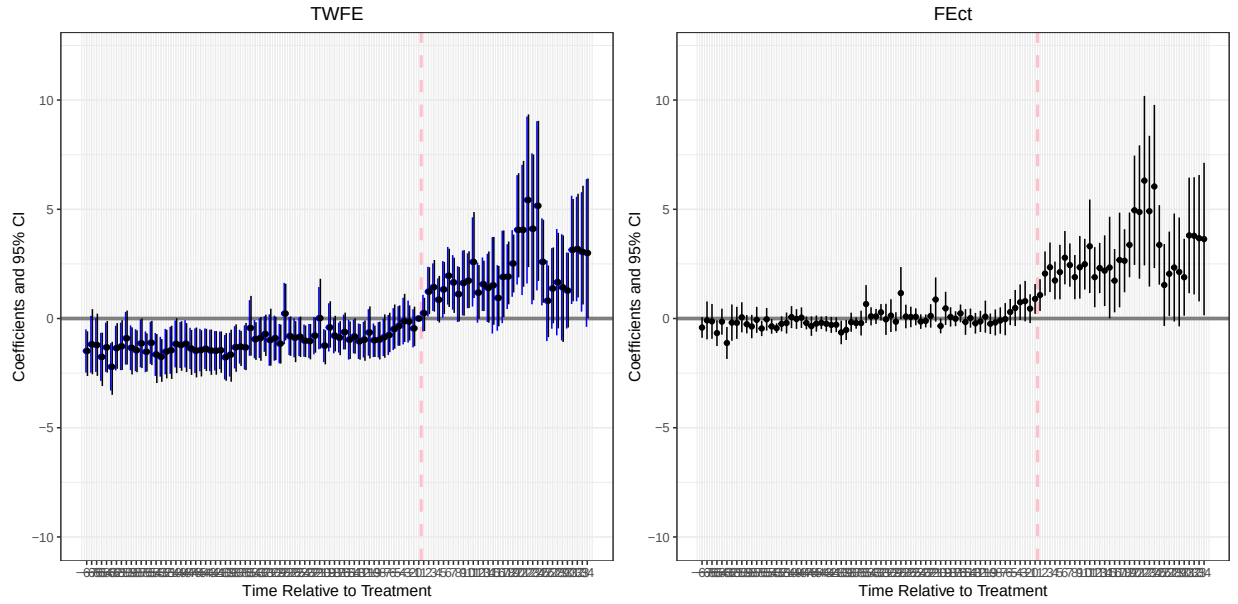


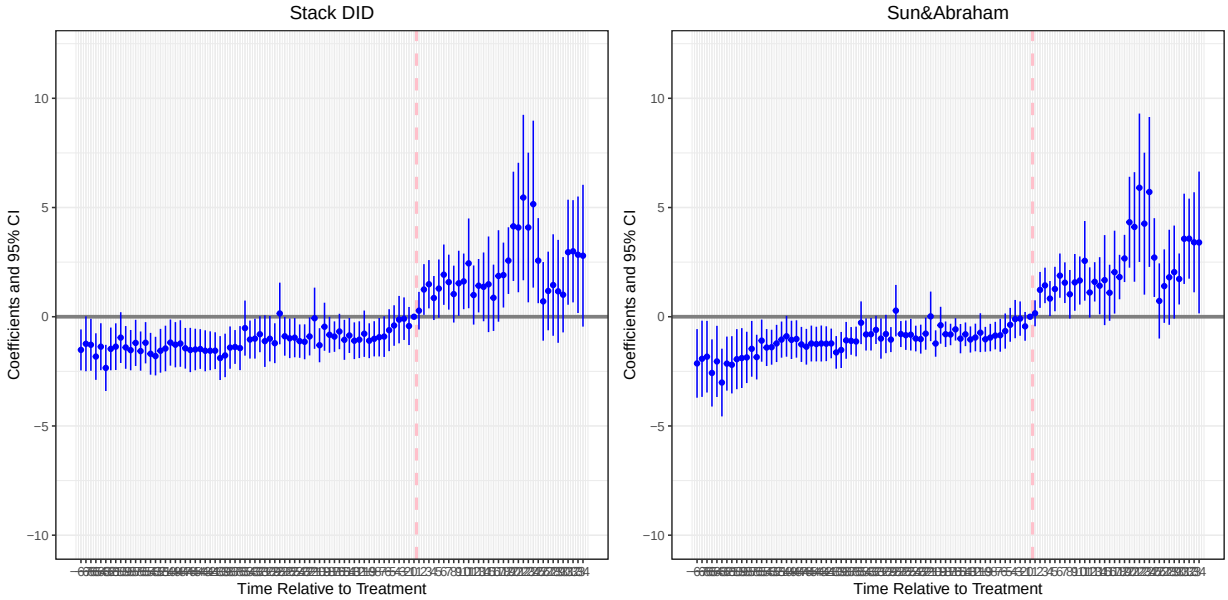

```
##   |                                                                      |
##   |                                                                      |
```

Dynamic Treatment Effects

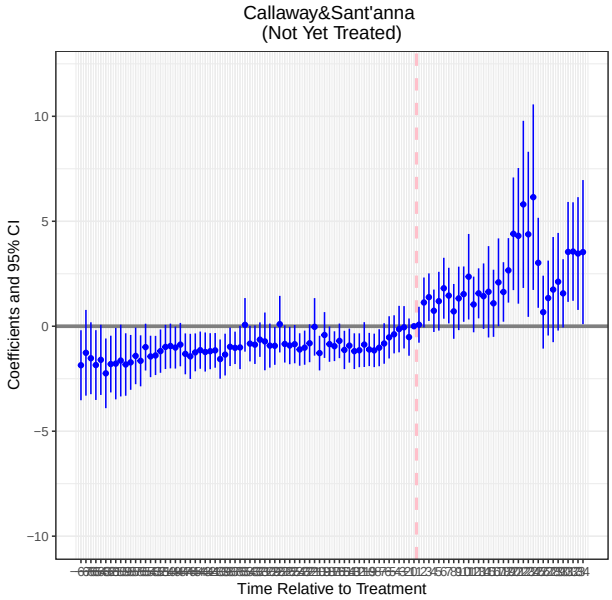
We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DIDM, and FEct. The results are shown below.

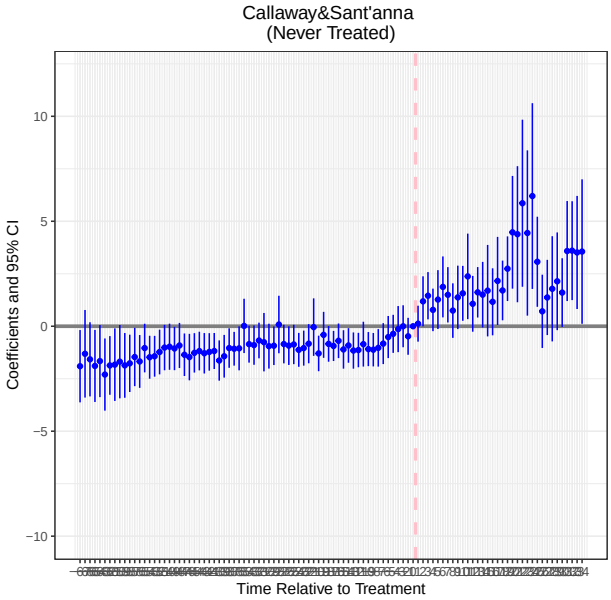
```
##   |                                                                      |
```

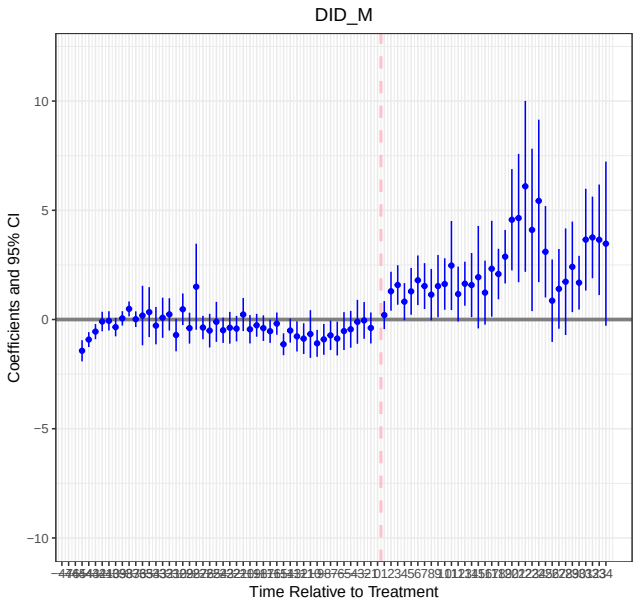



Stack DID
Coefficients and 95% CI
10
5
0
−5
−10
Time Relative to Treatment

Sun&Abraham
Coefficients and 95% CI
10
5
0
−5
−10
Time Relative to Treatment


Callaway&Sant'anna
(Not Yet Treated)
Coefficients and 95% CI
10
5
0
−5
−10
Time Relative to Treatment


Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
10
5
0
−5
−10
Time Relative to Treatment

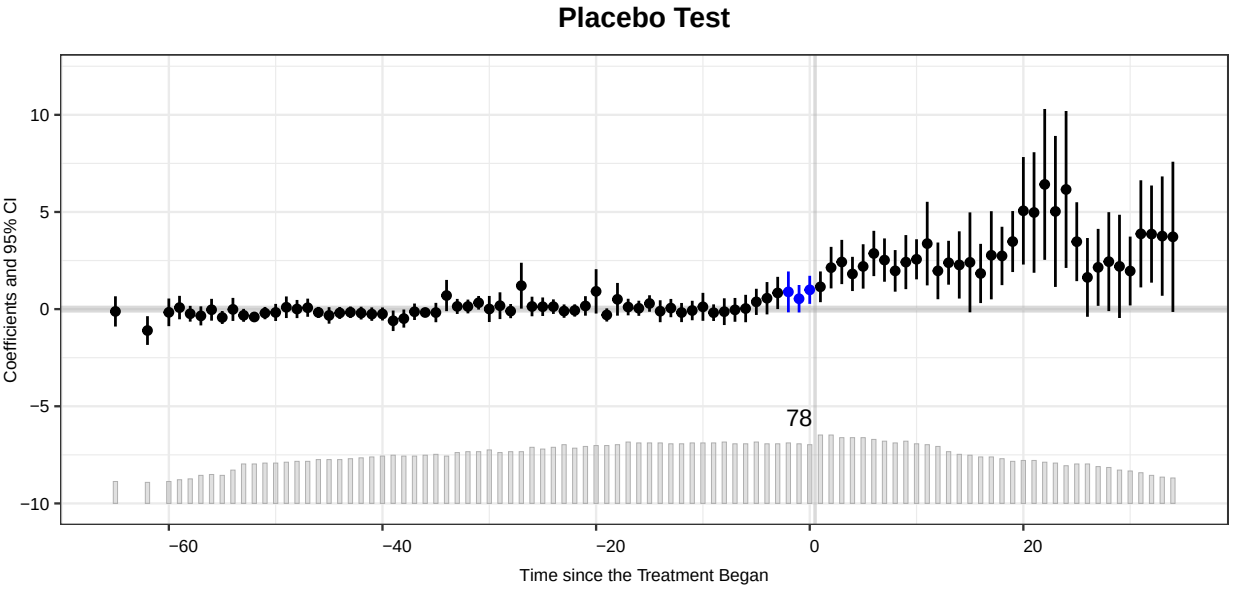


All HTE-robust estimators and TWFE yield similar estimated DTE. The estimated DTE are mostly positive and show two humps on post-treatment periods. On pre-treatment periods, there exists some signs of pre-trends.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.

```
##   |                                                                      |
```



Test Statistics

```
##                                    p-value
## F test                               0.010
## Equivalence test (default)           0.468
## Equivalence test (threshold=ATT)     0.002
## Placebo test                         0.016
## Carryover effect test                   NA
```

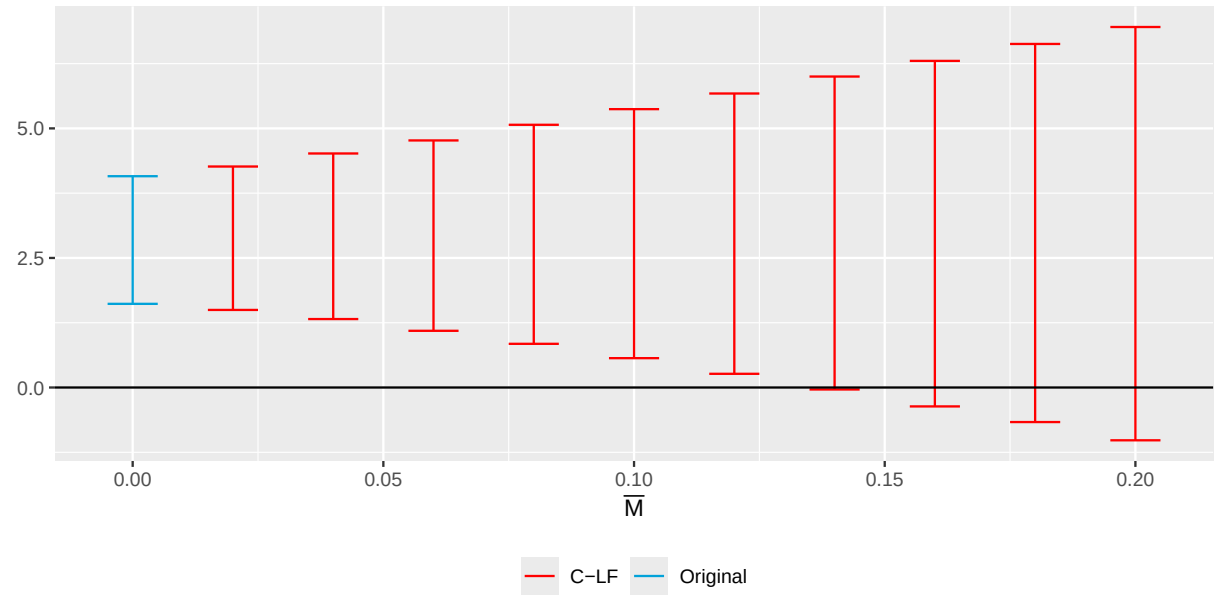
The rejection in the F-test and placebo test casts some doubt on the parallel trend (PT) assumption but the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT, suggesting potential confounding likely has very limited impact on the causal estimates.

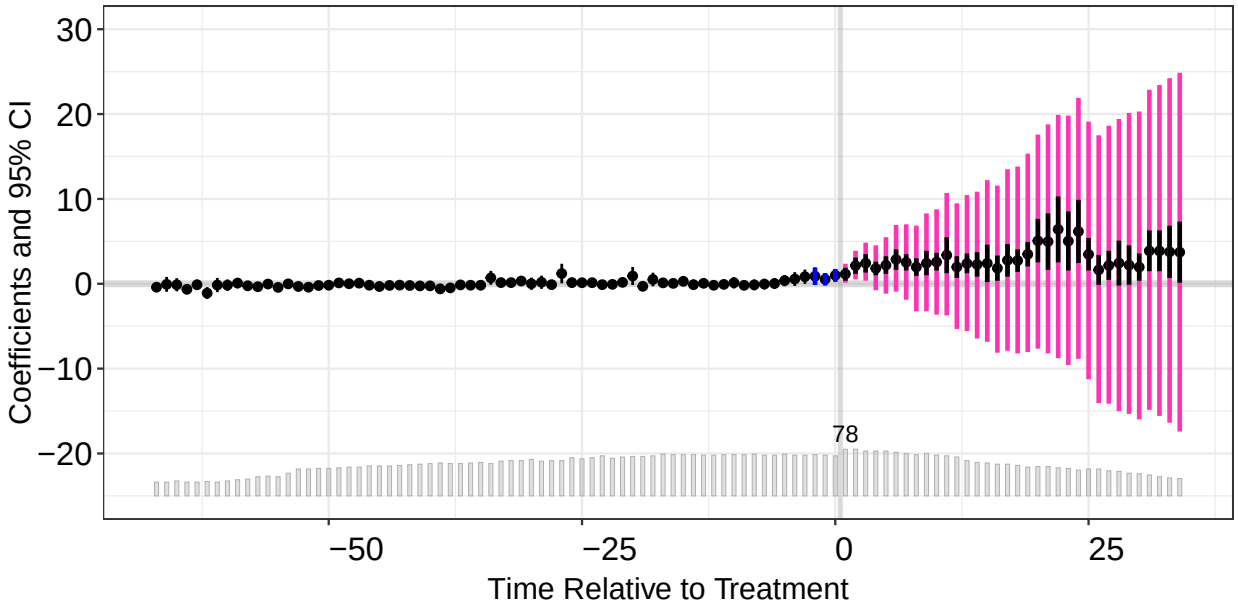
Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.14$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 0.14 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

```
##   |                                                                      |
```





Summary

Overall, the main result of the chosen model seems to be robust to HTE-robust estimators. There is some evidence (placebo test) for potential violations of the PT, but they seem inconsequential. The estimated ATT is sensitive to mild potential violations of the PT assumption benchmarked against the pre-treatment placebo estimates.

Fouirnaies (2018)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 6
Summary 7

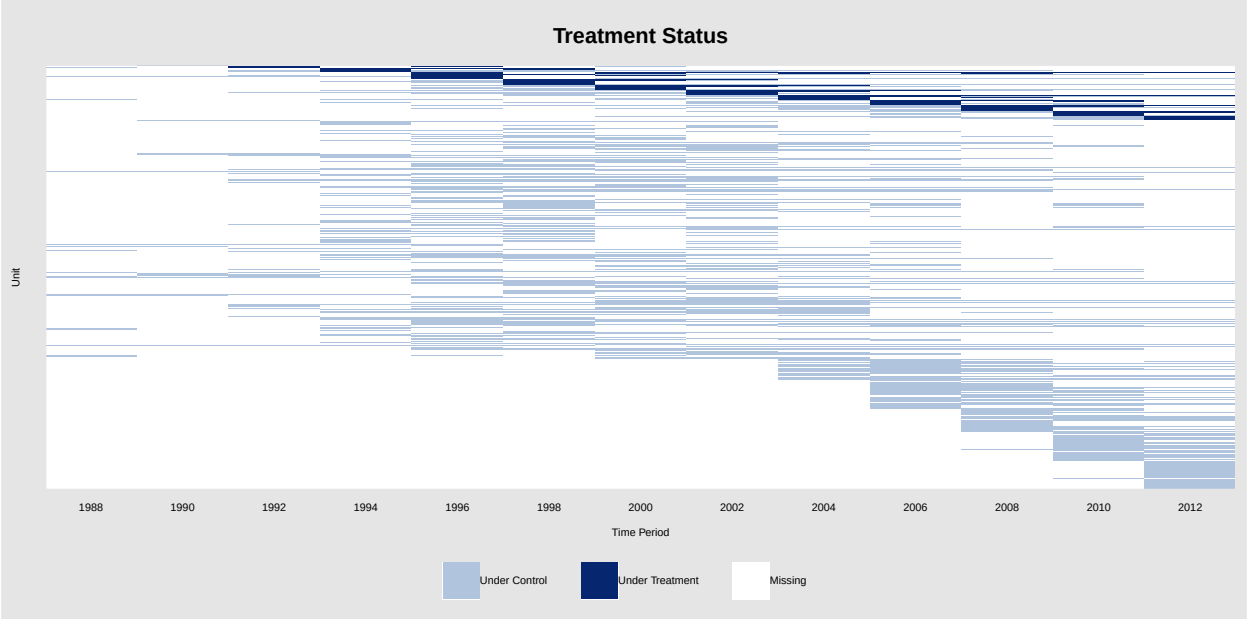
A First Look at Data

The paper investigates the effects of committee and party leader positions on industry contributions, using US legislator-year level panel data from 1988-2012. One of the main findings of this paper is that "when legislators advance to either a committee or party leader position, they experience a significant boost in corporate campaign contributions relative to other legislators in the chamber" (p179).

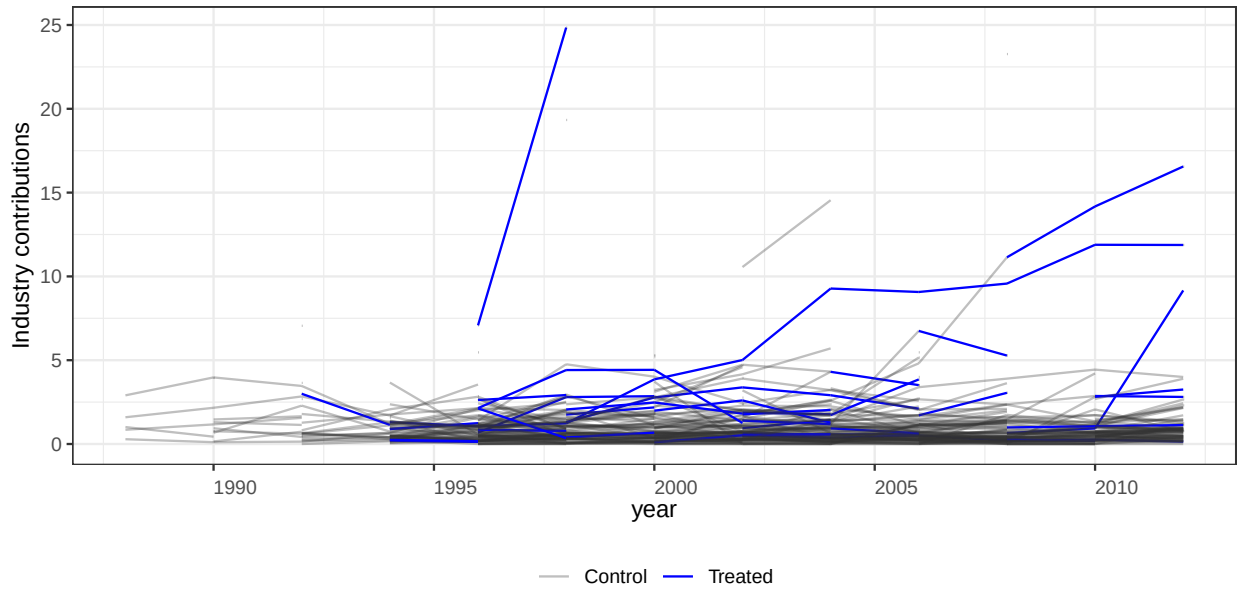
Model. We focus on **Model 1 of Table 1** in the paper. The authors use legislator (unit) and State's Chamber × Year fixed effects and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	Legislator × Year
Treatment	Leader position
Outcome	Industry contributions
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Higher-level Unit*Time dummy

View treatment status First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a legislator). There are a large number of missing values in the data. The treatment has reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each legislator. The observations under treated status are marked in blue.



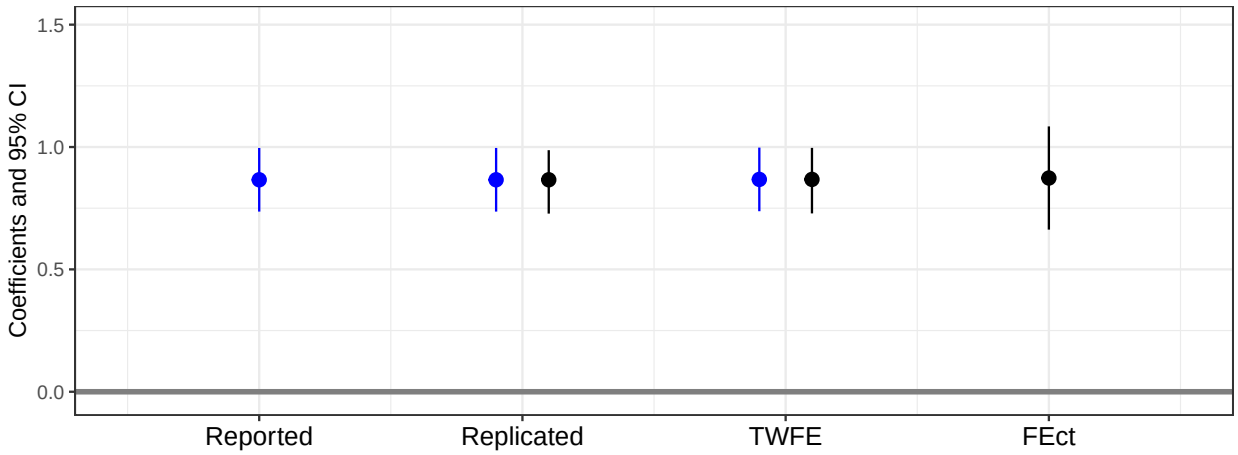
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE (using State's Chamber × Year and unit fixed effects) and imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

```r
sol <- feols(pct_amountindustry_1~Leader + Chair + IncMaj|CandId + StateChamberYear,
             data = df,cluster = "CandId")
summary(sol)
```

```
## OLS estimation, Dep. Var.: pct_amountindustry_1
## Observations: 45,639
## Fixed-effects: CandId: 16,404,  StateChamberYear: 911
## Standard-errors: Clustered (CandId)
##        Estimate Std. Error  t value   Pr(>|t|)
## Leader 0.866165   0.066336 13.05733  < 2.2e-16 ***
## Chair  0.169930   0.024401  6.96405 3.4316e-12 ***
## IncMaj 0.125683   0.022134  5.67834 1.3831e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.817671     Adj. R2: 0.72438
##                  Within R2: 0.036021
```



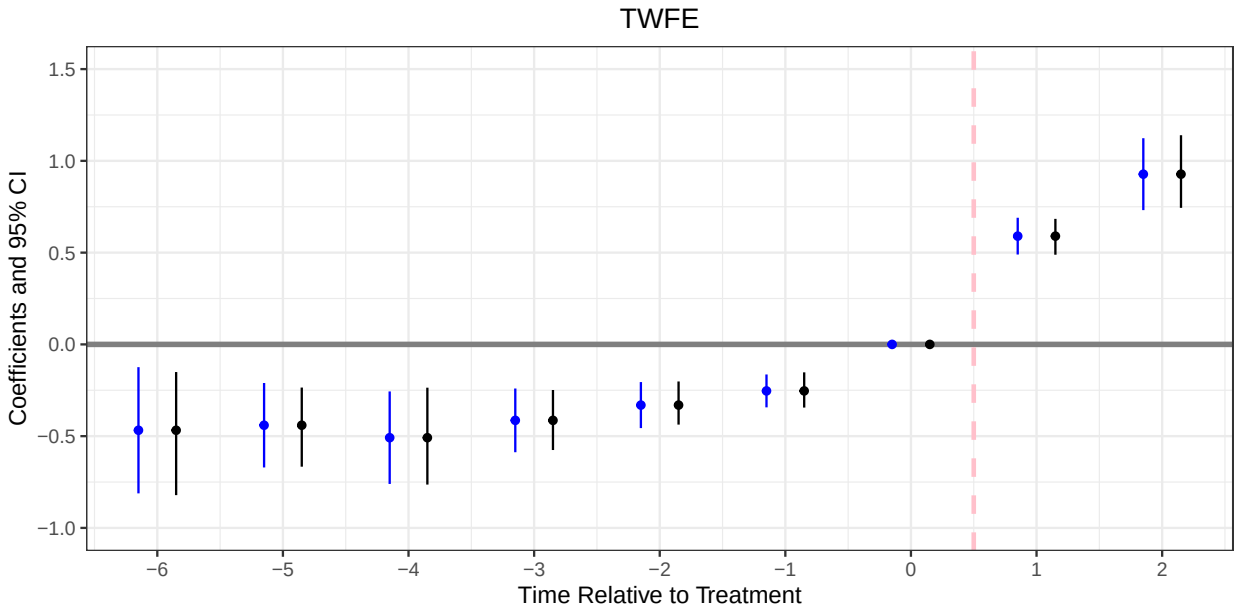
The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

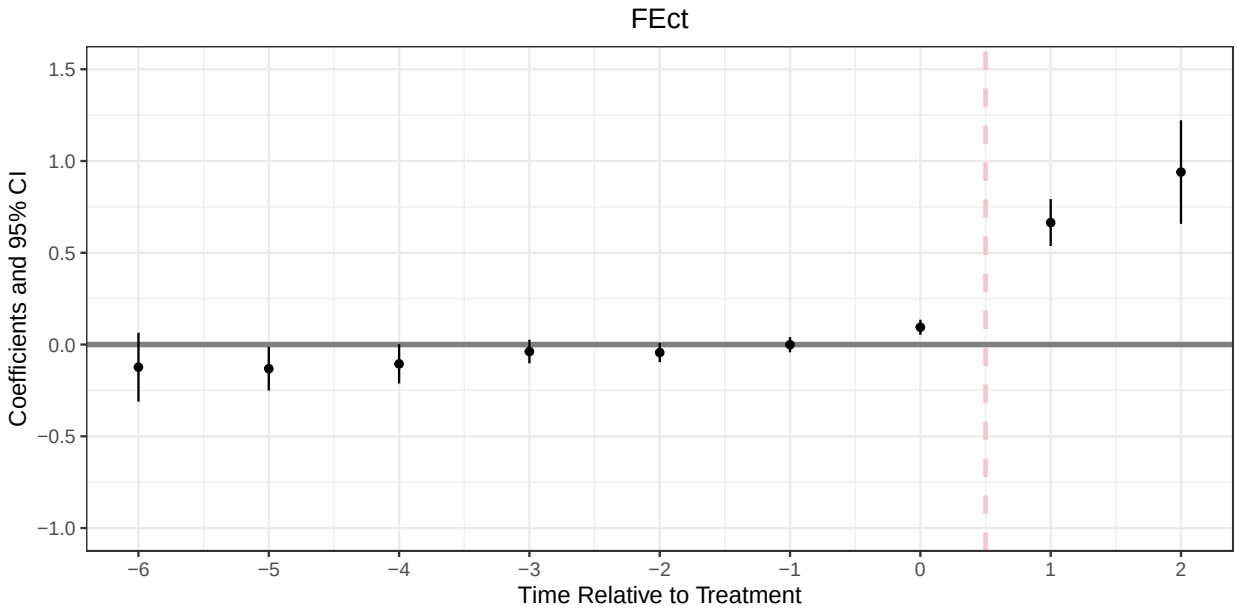
Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

```
##   |                                                                      |
```

```
##   |                                                                      |
```



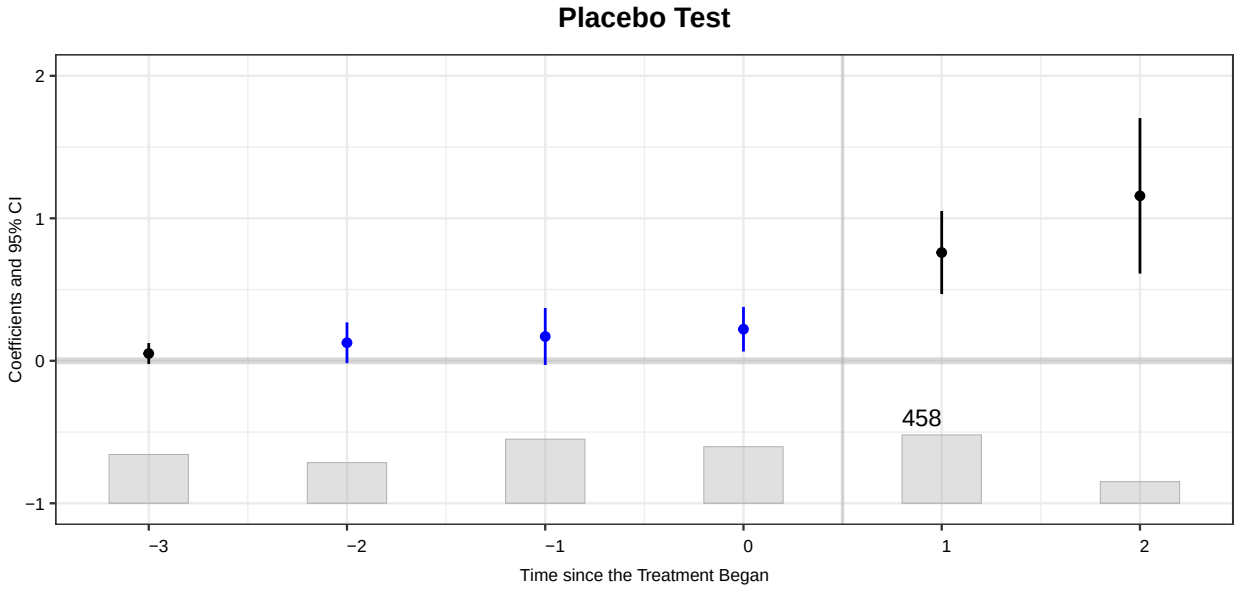


The TWFE and FEct estimator yield similar results. The estimated DTEs are positive on all two post-treatment periods in the plot. The estimated DTEs using TWFE exhibits some pre-trend. For FEct, the pre-trend appears to be weaker.

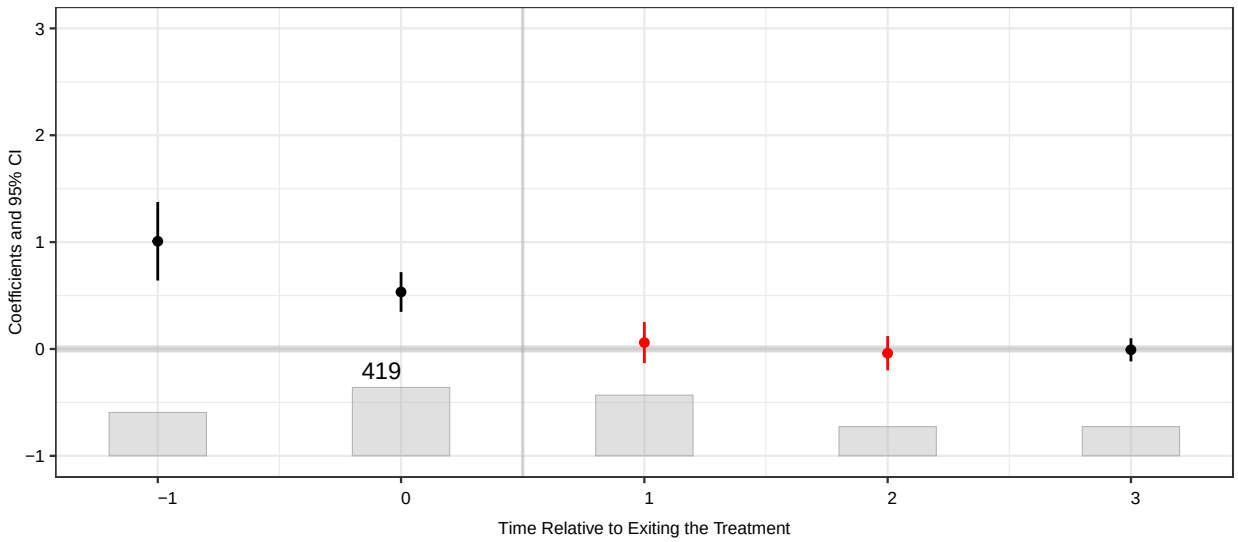
Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.

```
##   |                                                                      |
```




```
##   |                                                                      |
##   |                                                                      |
```



Test Statistics

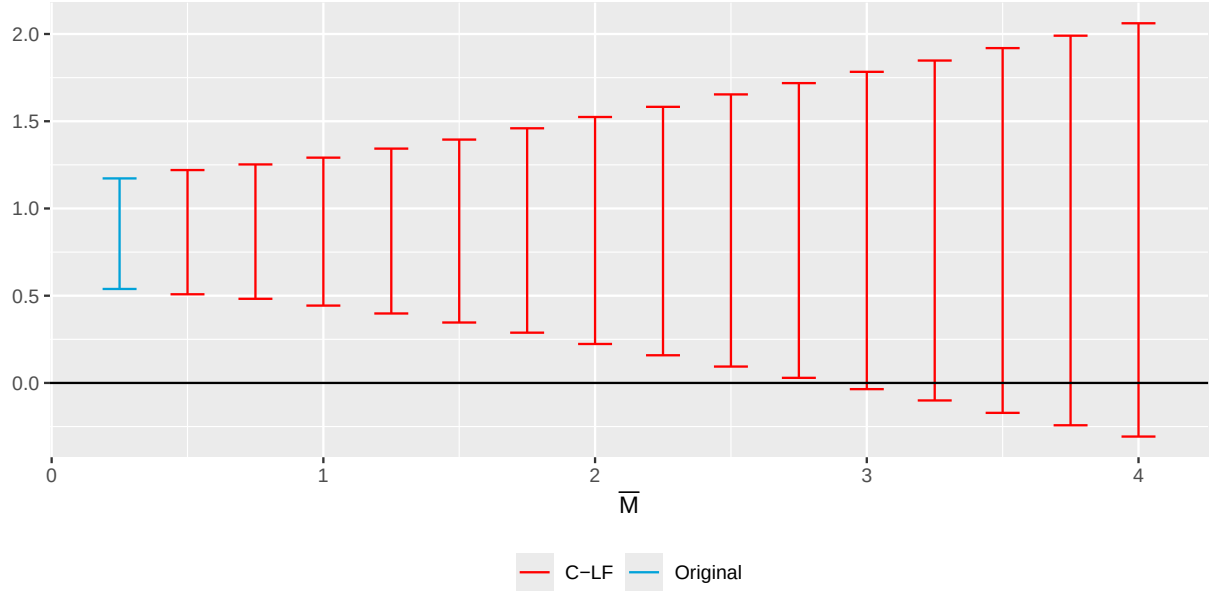
```
##                                  p-value
## F test                             0.000
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.000
## Placebo test                       0.008
## Carryover effect test              0.722
```

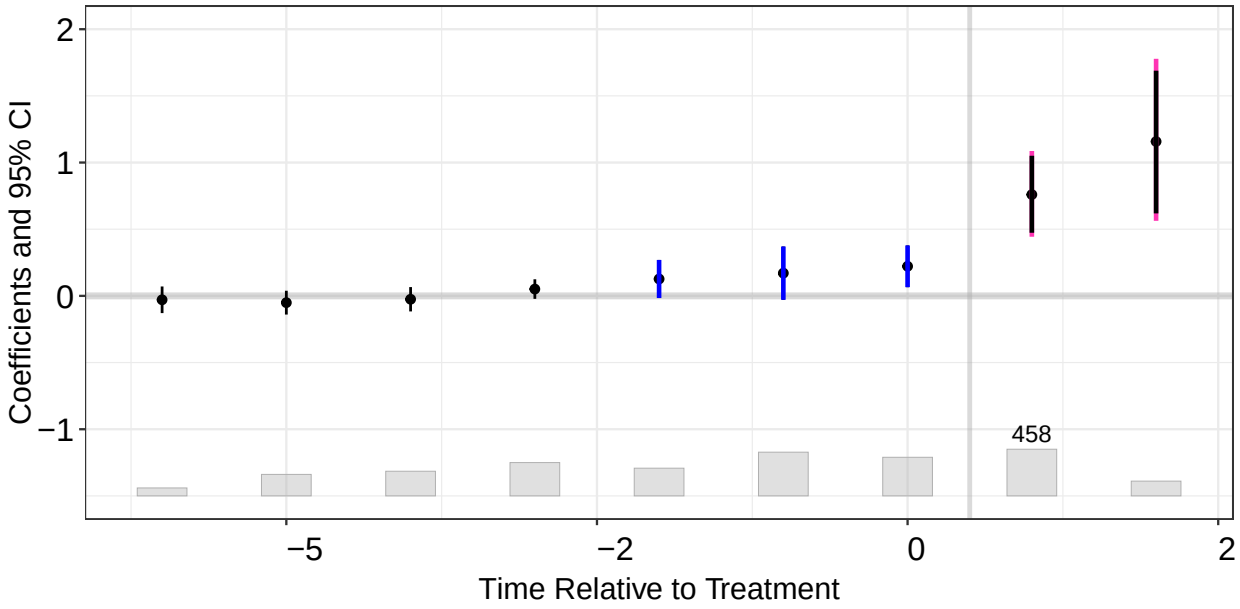
We find some evidence for violations of the parallel trend (PT) assumption. However, the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT, suggesting that the violation is inconsequential for the ATT estimate. There is no evidence for the violation to the no-carryover-effect assumption.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 2.8$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 2.8 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model seems to be robust to the HTE-robust estimator, FEct. We find some evidence of violations of the PT assumption using the F test and placebo tests; however, these violations seem inconsequential for credible inference of the ATT estimate, as shown by the equivalence test and sensitivity analysis, with the breakdown threshold $M = 2.8$.

Fouirnaies and Hall (2018)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 6
Summary 7

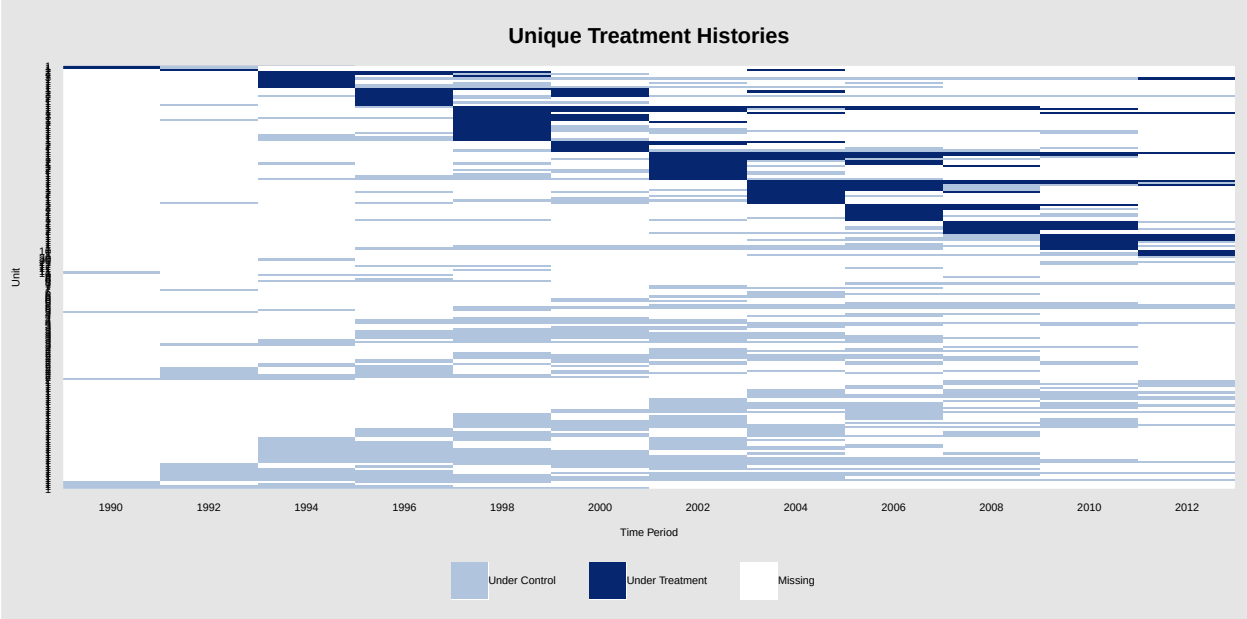
A First Look at Data

The paper investigates the effects of committee membership on contributions from interested donors, using candidate-industry-year level panel data from US state legislative committees during 1988-2014. One of the main findings of this paper is that "Joining the committee appears to cause roughly a 27% increase in the amount of contributions from interest groups whose business interests correspond to the committee's policy jurisdiction" (p138).

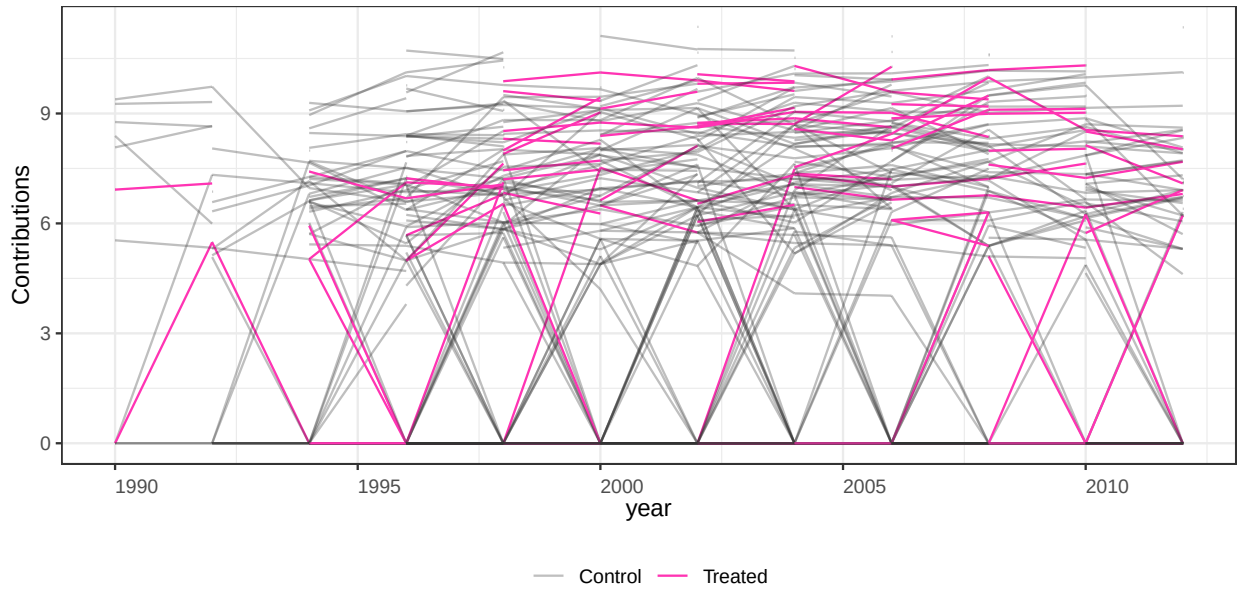
Model. We focus on **Model 2 of Table 2** in the paper. The authors use candidate-industry (unit) and candidate× year fixed effects and report robust standard errors (SEs) clustered at the unit level.

Replication Summary	
Unit of analysis	Candidate-industry × year
Treatment	Committee membership
Outcome	Contributions from interested donors
Treatment type	General
Outcome type	continous
Fixed Effects	Unit+Higher-level Unit*Time dummy

View treatment status First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (candidate-industry pair). There are a large number of missing values. The treatment has reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each candidate-industry pair. The observations under treated status are marked in blue.



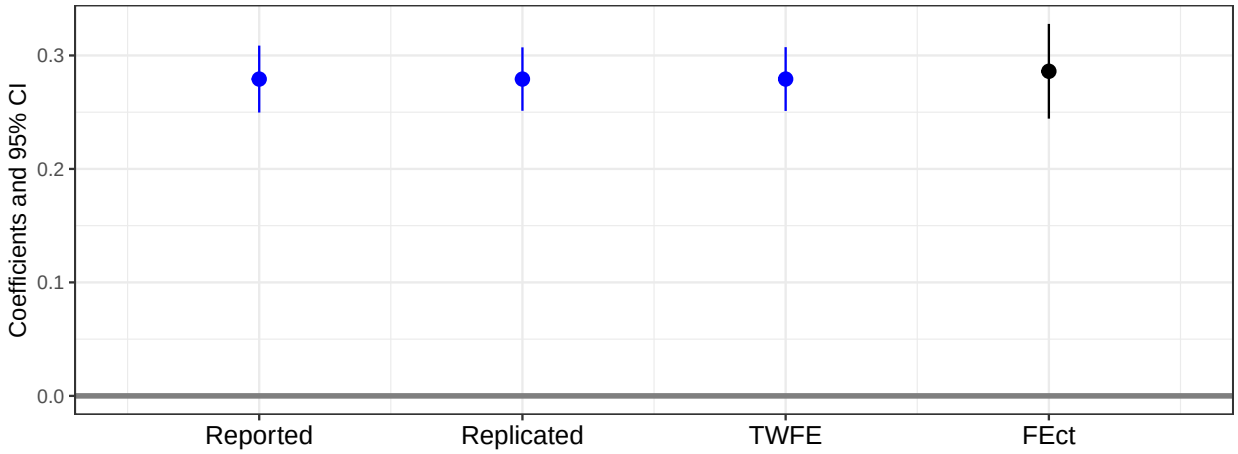
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE (using candidate-industry and candidate× year fixed effects) and imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs (for the "reported'' ATT, we report the robust SEs) and cluster-bootstrapped CIs, respectively.

Original Finding

```r
sol <- feols(log_amount~cmt|cand_industry+cand_year,
             data = df,vcov = "hetero")
summary(sol)
```

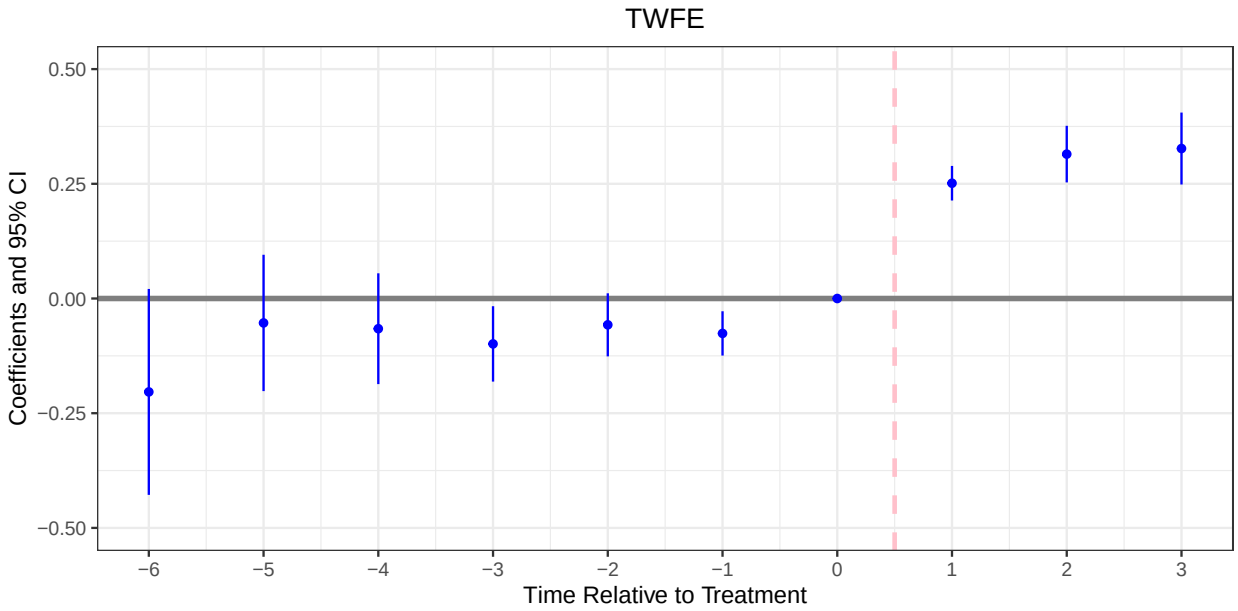
```
## OLS estimation, Dep. Var.: log_amount
## Observations: 443,490
## Fixed-effects: cand_industry: 161,820,  cand_year: 44,349
## Standard-errors: Heteroskedasticity-robust
##     Estimate Std. Error t value  Pr(>|t|)
## cmt 0.279182   0.015051  18.549 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.30443     Adj. R2: 0.776132
##                  Within R2: 0.001693
```

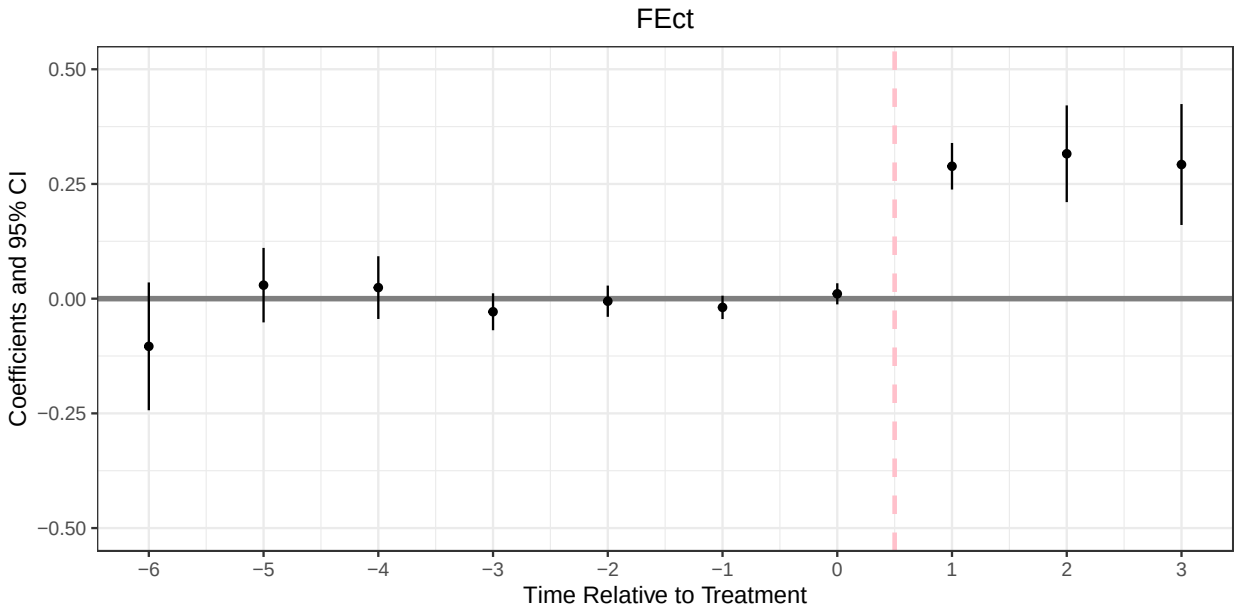


The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

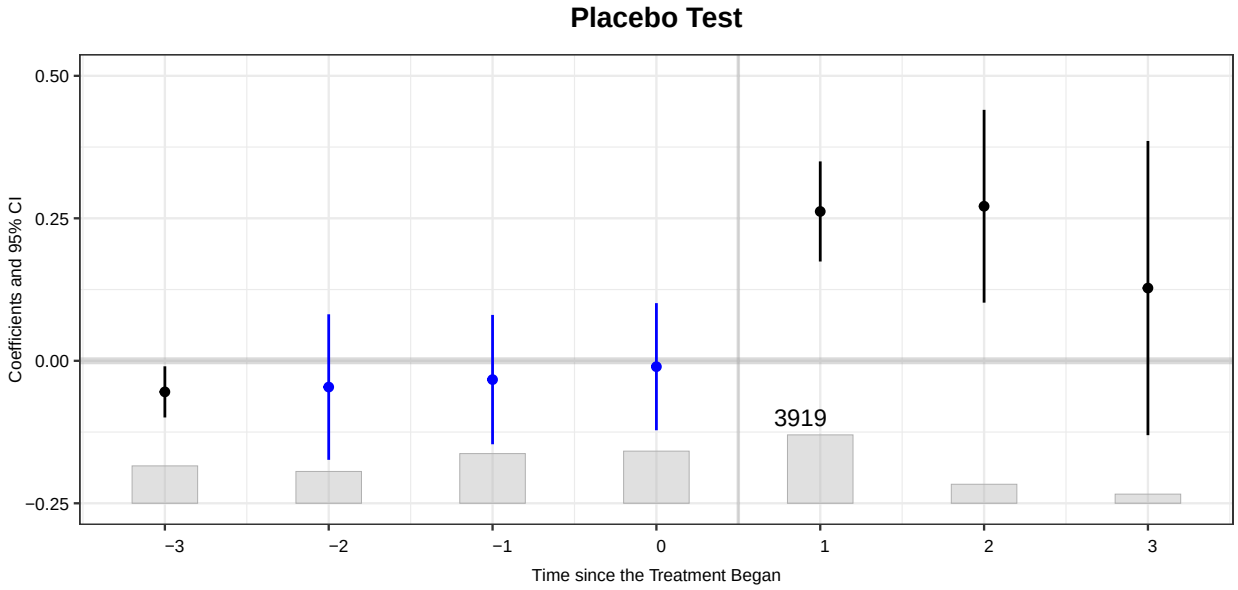




TWFE and FEct estimates are broadly consistent with each other. The estimated DTE are all positive on post-treatment periods and the pre-trends seem to be weak.

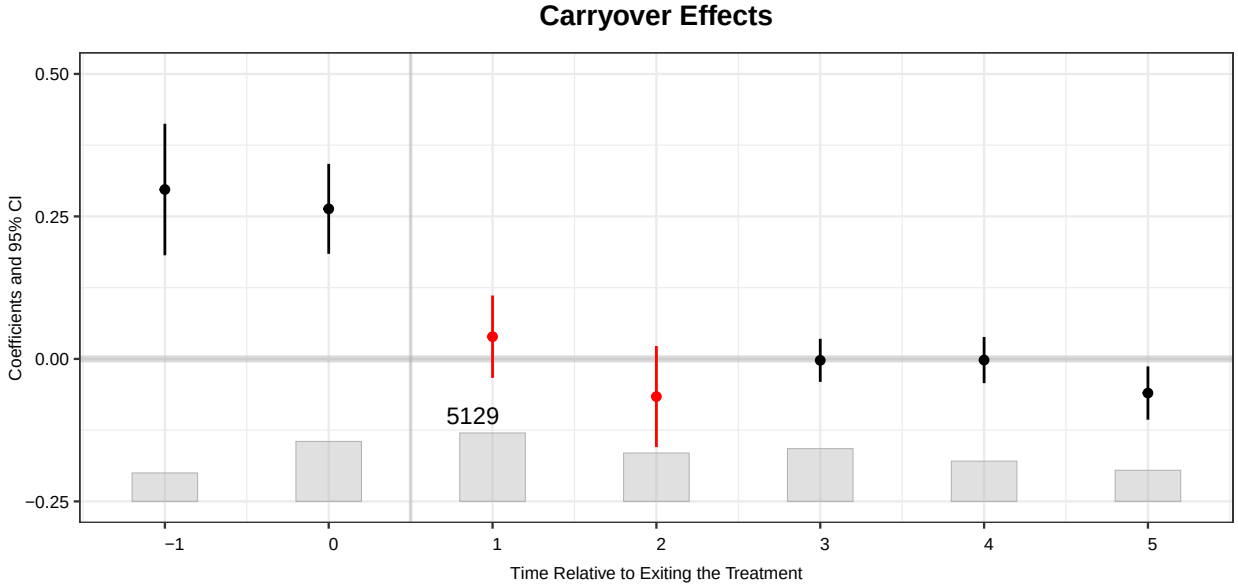
Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.



Carryover Test

```
## Problem getting FE, maximum iterations reached (1st order loop).  |
```



Test Statistics

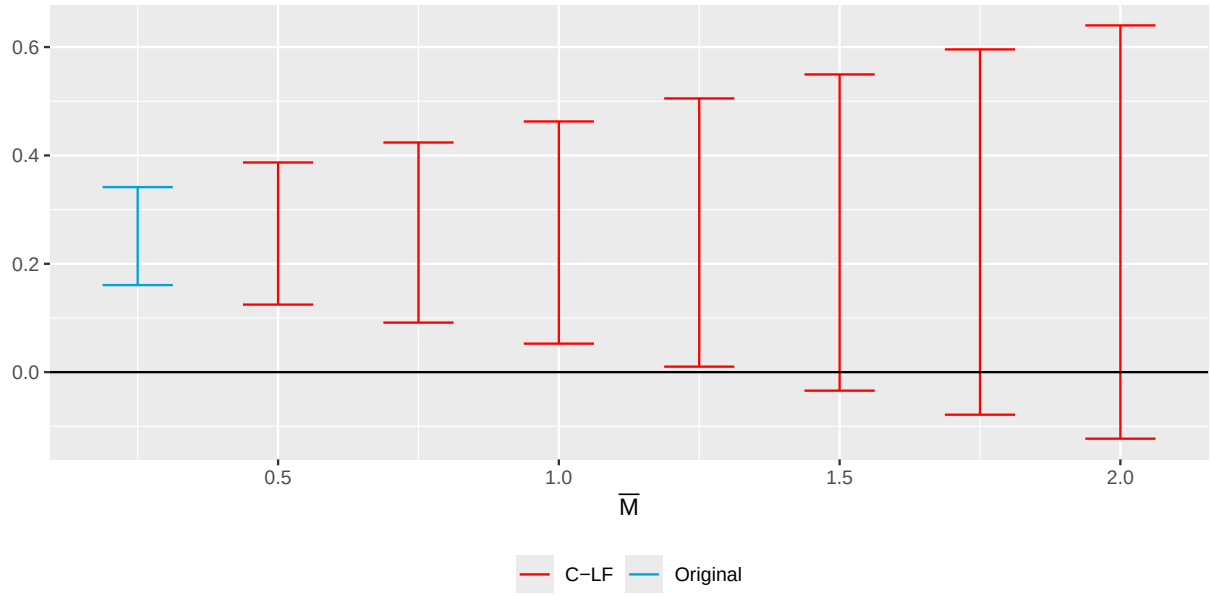
```
##                                  p-value
## F test                             0.209
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.000
## Placebo test                       0.534
## Carryover effect test              0.885
```

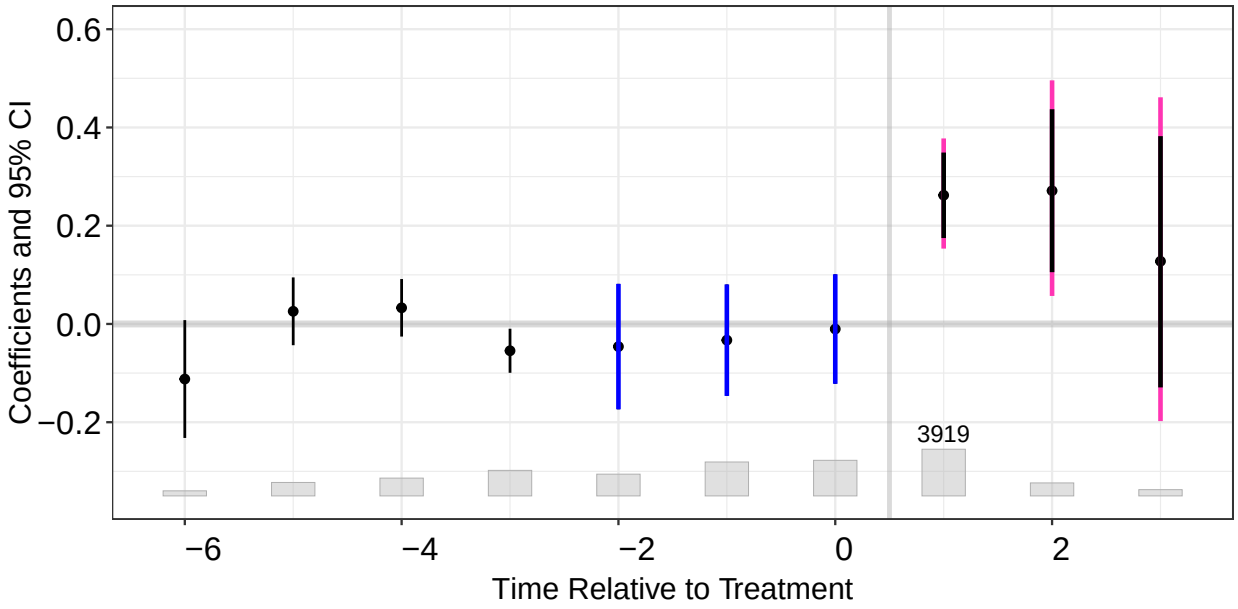
We find little evidence for violations to the parallel trend (PT) assumption, the no-anticipation assumption, and the no-carryover-effect assumption. The equivalence test also rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 1.25$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 1.25 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust to FEct, an HTE-robust estimator for the ATT. We do not find strong evidence for violations of the PT and no-carryover-effect assumption. The estimated ATT is also robust to mild violations of the PT assumption with the breakdown threshold $M = 1.25$.

Fouirnaies and Hall (2022)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 6
Summary 7

A First Look at Data

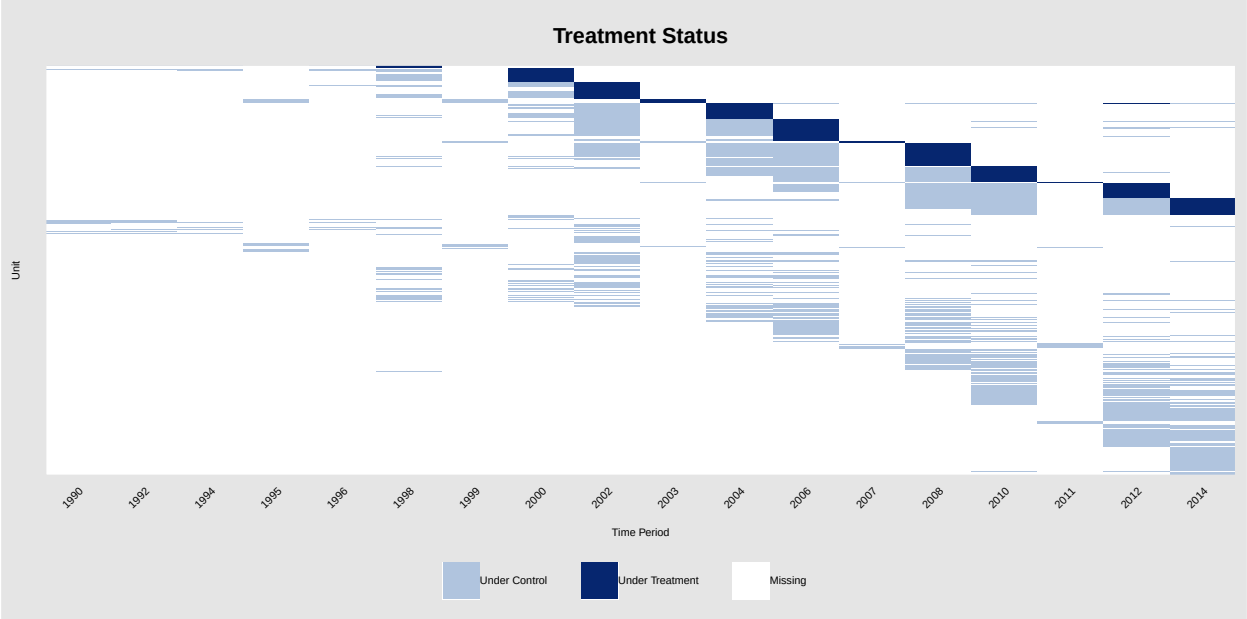
The paper investigates the effects of term limits on the perfomance of legislators, using US legislator-year level panel data between 1990 and 2014. One of the main findings of this paper is that "a noticeable on-average decrease in productivity for legislators when they can no longer seek reelection" (p669, Table 2).

Model. We focus on **Model 1 of Table 2** in the paper. The authors use a unit (legislator) and Chamber $\times$ Term fixed effects model and report robust standard errors clustered at the unit level.

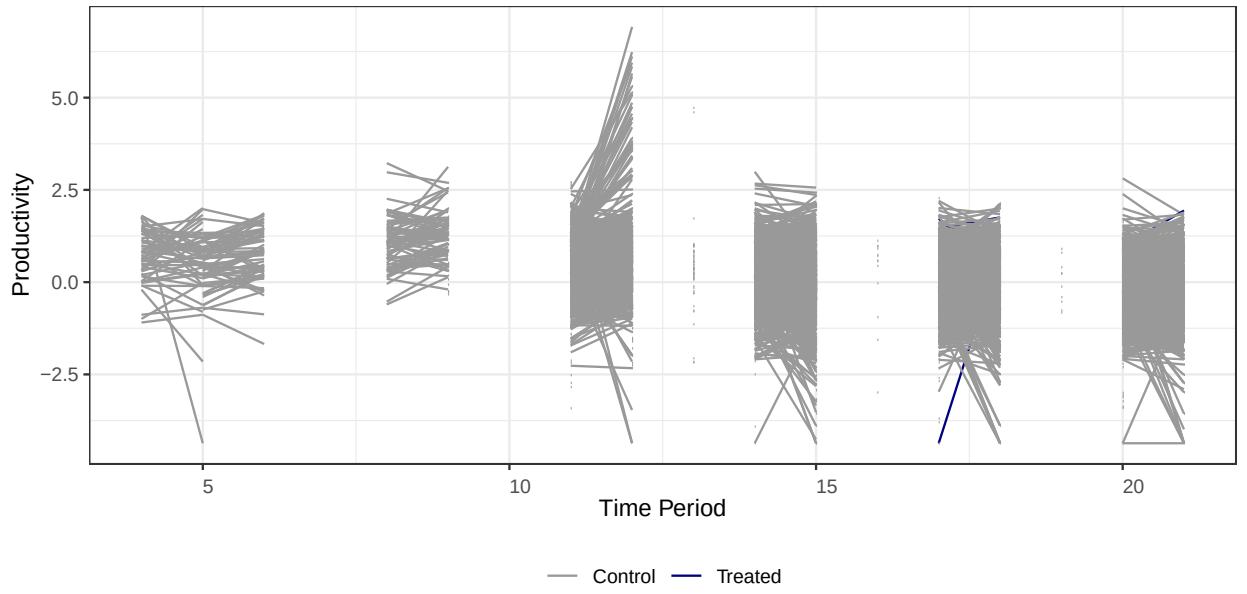
Table 1: Replication Summary

Unit of analysis	Legislator $\times$ year
Treatment	Term limit
Outcome	Productivity
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Higher-level Unit*Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a legislator). There are treatment reversals and a large amount of missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each candidate-industry pair. The observations under treated status are marked in blue.



Point Estimates

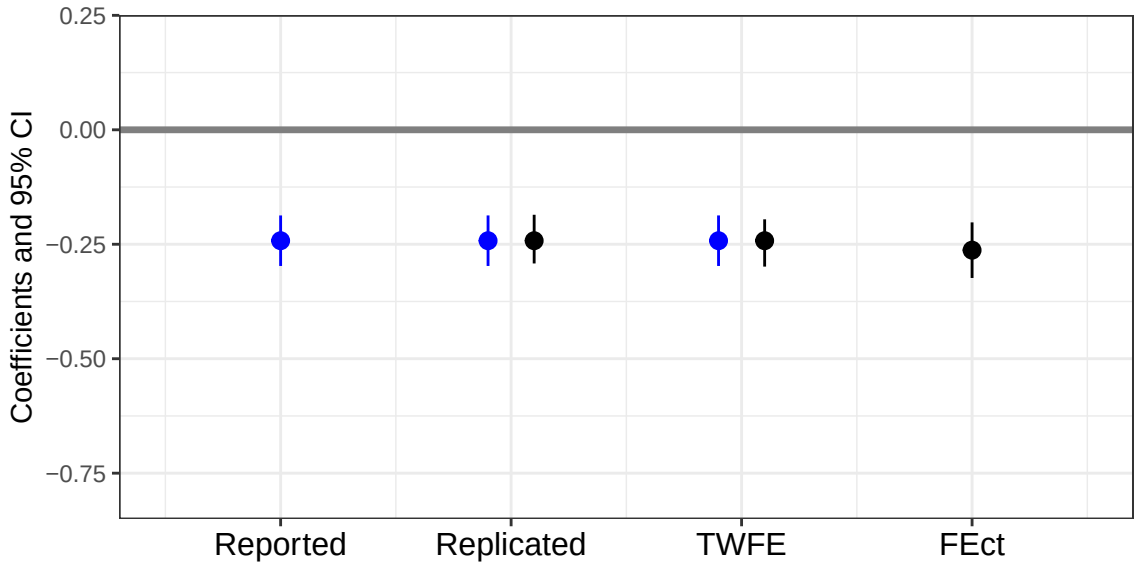
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE (using legislator and Chamber × Term fixed effects) and imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```r
sol <- feols(pc1~termlimited|CandId+chamber_term,
             data = df,cluster = "CandId")
summary(sol)
```

```
## OLS estimation, Dep. Var.: pc1
## Observations: 11,109
## Fixed-effects: CandId: 4,642,  chamber_term: 130
## Standard-errors: Clustered (CandId)
##              Estimate Std. Error  t value  Pr(>|t|)
## termlimited -0.242228   0.028139 -8.60839 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.450454     Adj. R2: 0.644293
##                  Within R2: 0.014726
```

```
##   |                                                      |
```



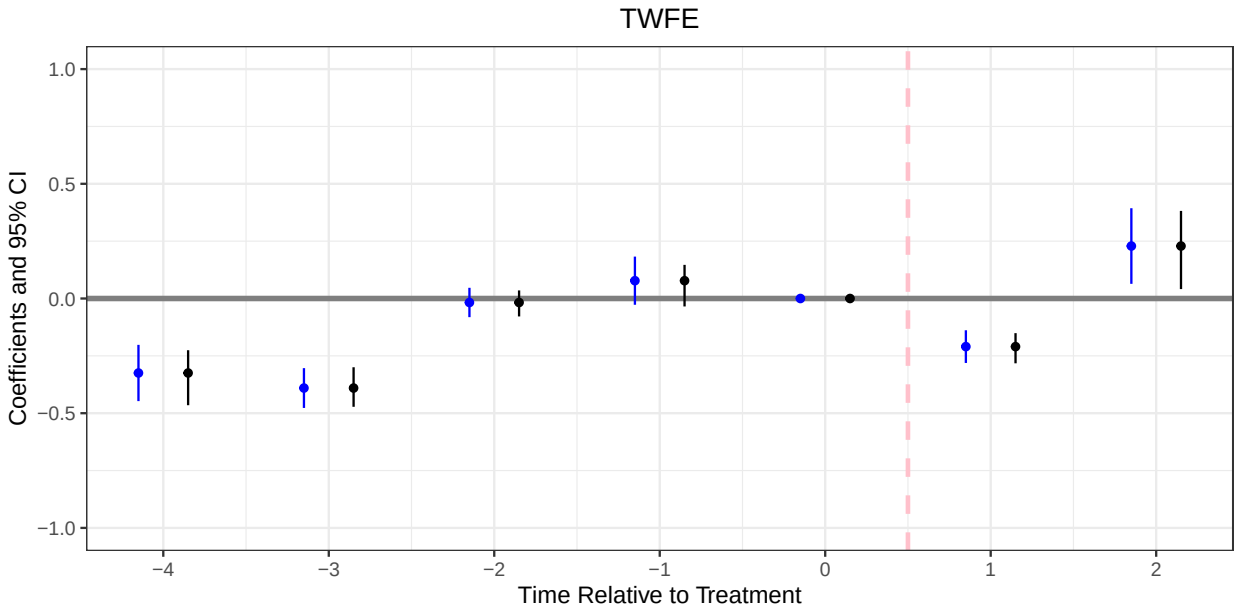

```
##   |                                                      |
```

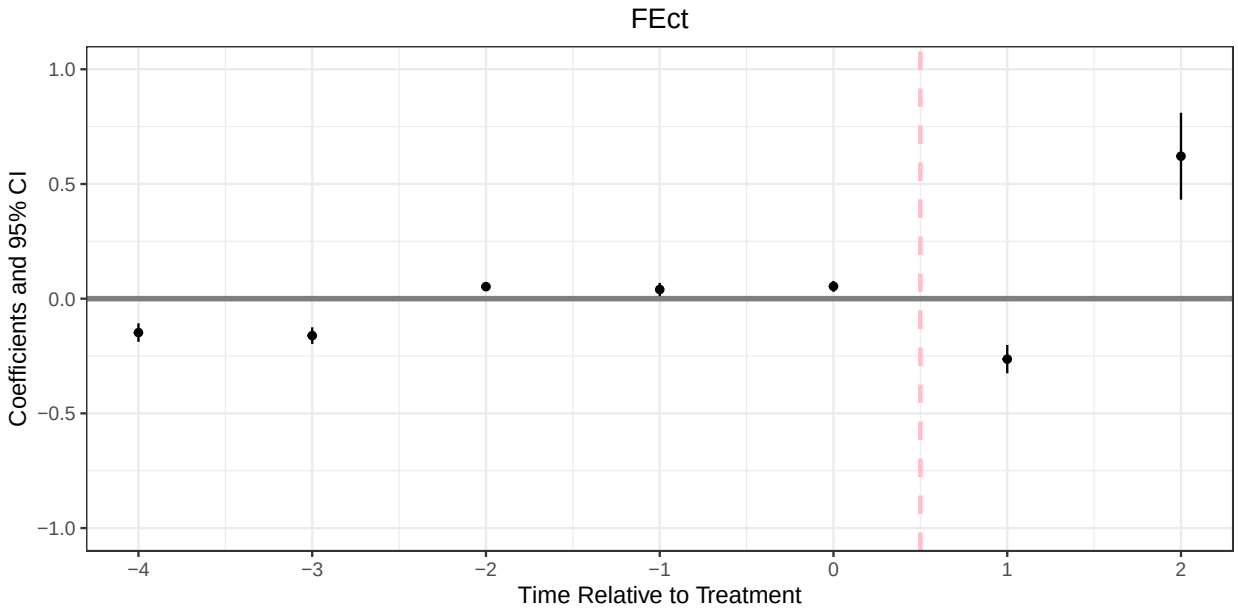
```
##   |                                                      |
```

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

```
##   |                                                      |
```





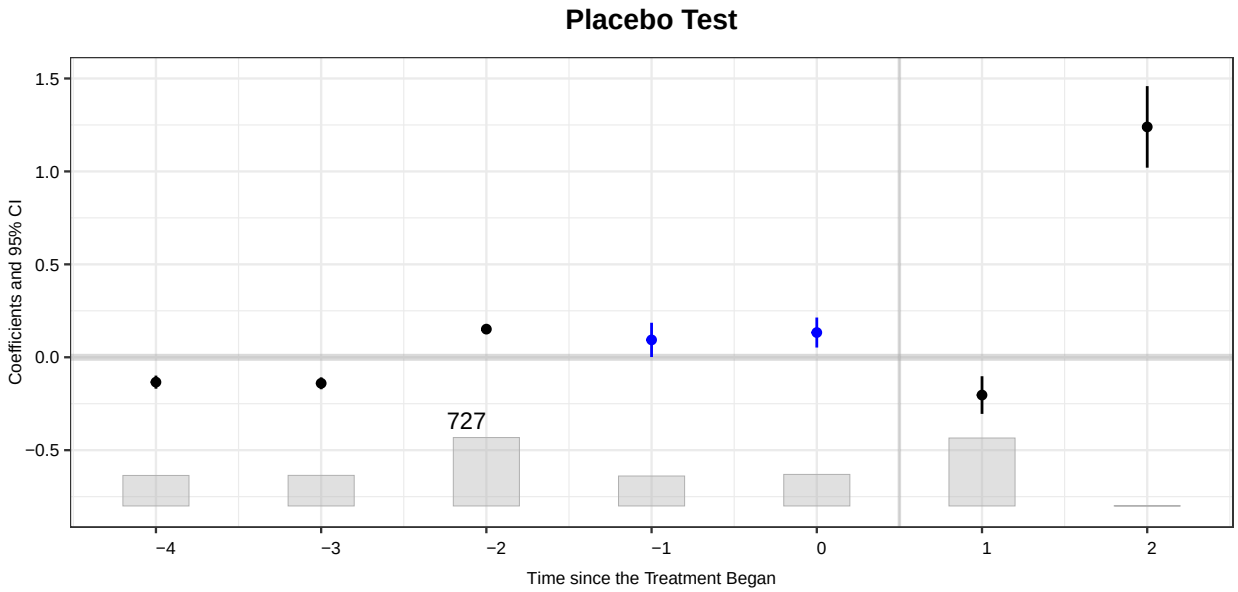
TWFE and FEct estimates are broadly consistent with each other. The estimated DTEs are negative in the first post-treatment period and positive in the second post-treatment period. The estimated DTEs using TWFE exhibits some pre-trend. For FEct, the pre-trend appears to be weaker.

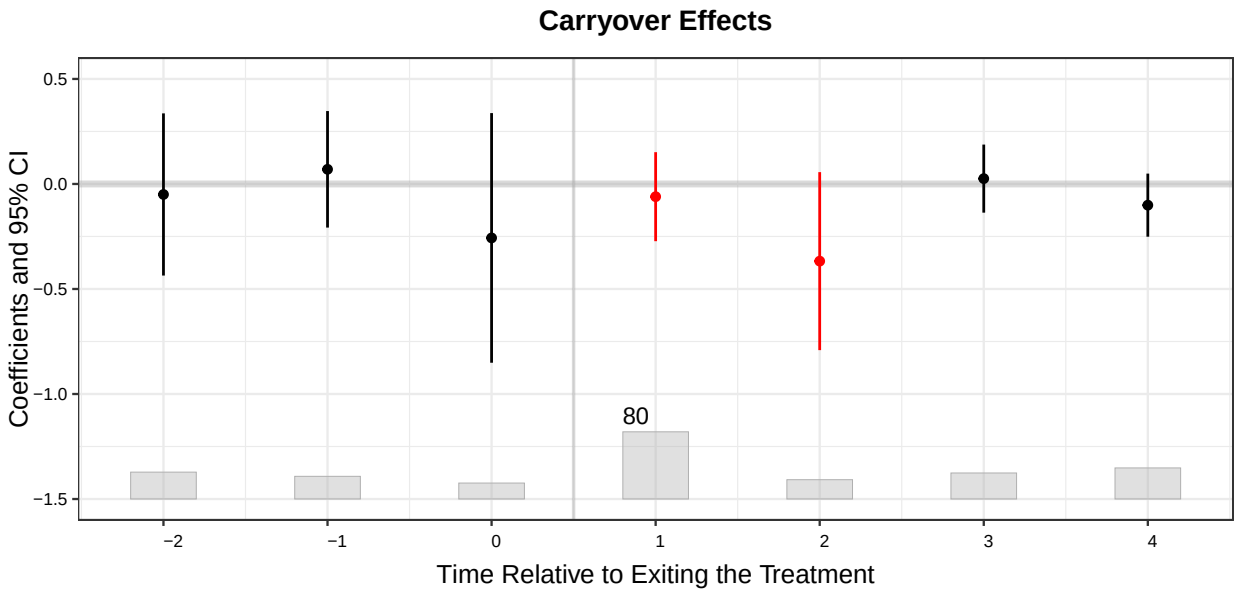
Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.

```
##   |                                                                      |
```

```
##   |                                                                      |
##   |                                                                      |
```





Test Statistics

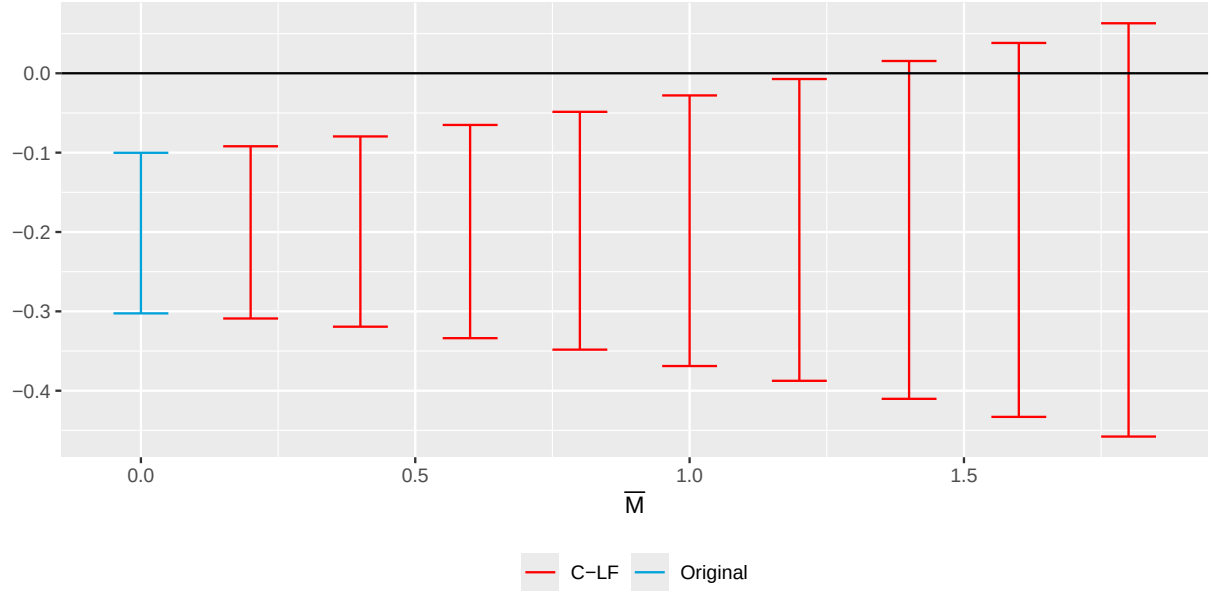
	p-value
F test	0.000
Equivalence test (default)	0.000
Equivalence test (threshold=ATT)	0.000
Placebo test	0.001
Carryover effect test	0.237

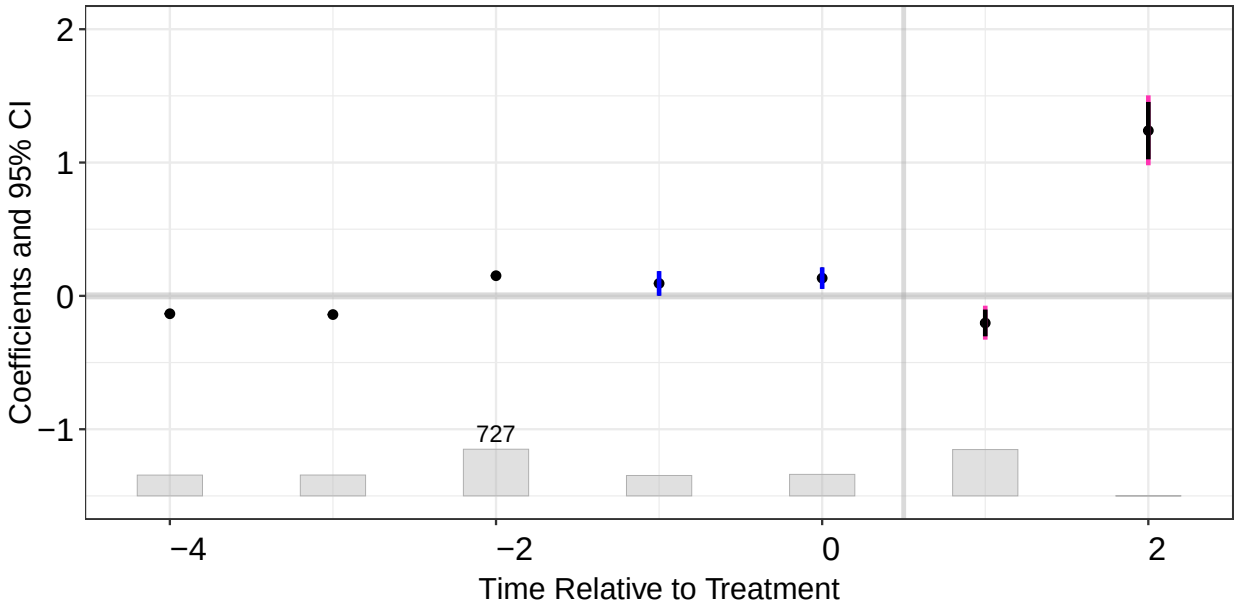
The F-test and the placebo test indicate violations of the parallel trend (PT) assumption and no-anticipation effect assumption. However, the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT, suggesting the violations may be inconsequential. We find little evidence of violation of the no-carryover-effect assumption.

Sensitivity Analysis

After holding out two pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the two placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 1.2$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 1.2 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust to the imputation estimator. We find some evidence for violations of the PTA, though they seem inconsequential for credible inference of the average effect, with the breakdown threshold $M = 1.2$.

Fresh (2018)

Contents

Summary 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Summary 6

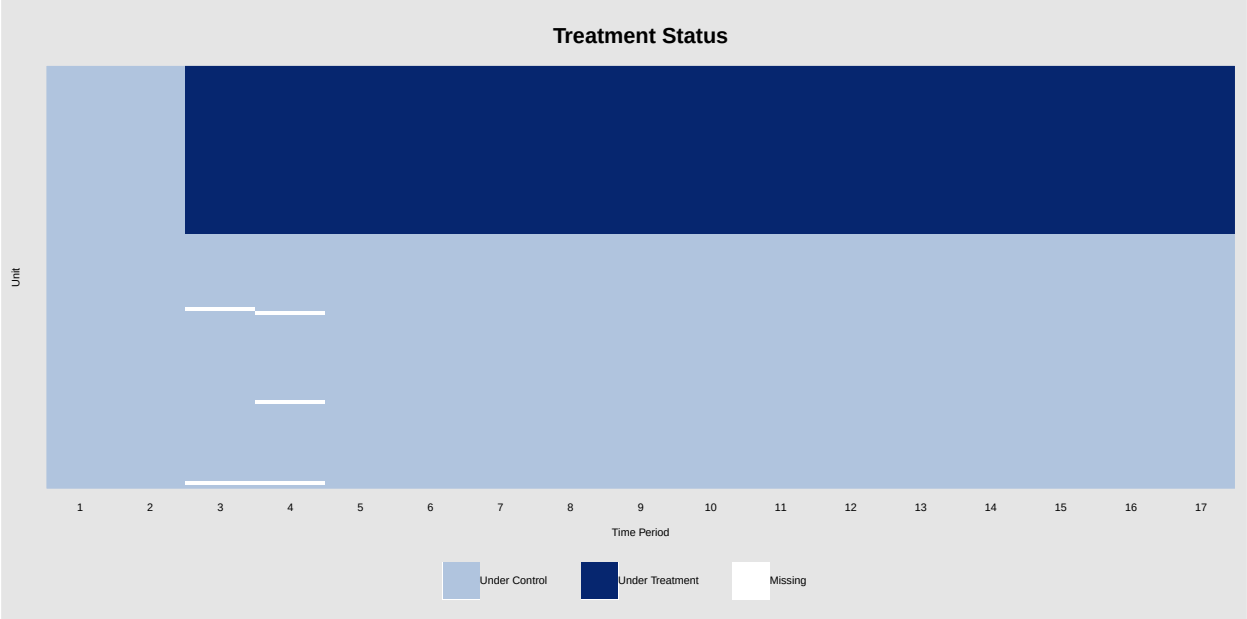
Summary

The paper investigates the effects of Section 5 of the 1965 Voting Rights Act (VRA) on black voter registration, using county-year level panel data from North Carolina, between 1958 and 1993. One of the main findings of this paper is that "Black voter registration rates increased (statistically significantly) in response to Section 5 coverage" (p715).

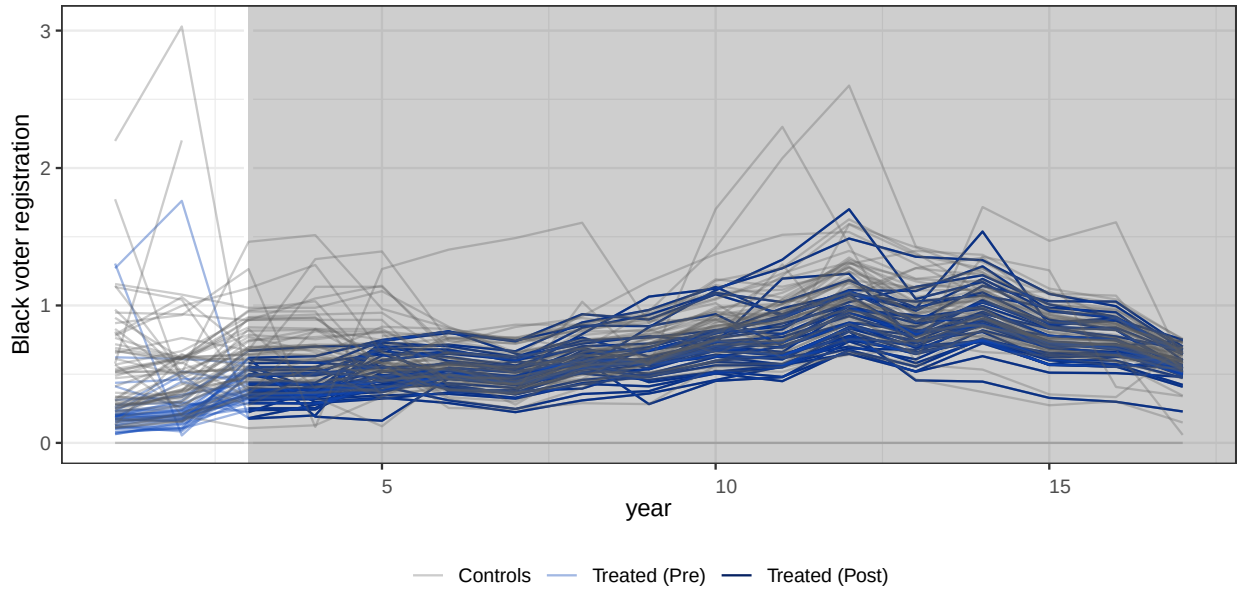
Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	County × year
Treatment	1965 Voting Rights Act
Outcome	Black voter registration
Treatment type	Classic
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a county).



Plotting the outcome variable. We plot the trajectory of the outcome variable for each county. The control units are represented in gray. We highlight the observations of treated units under untreated and treated status using light blue and deep blue, respectively.



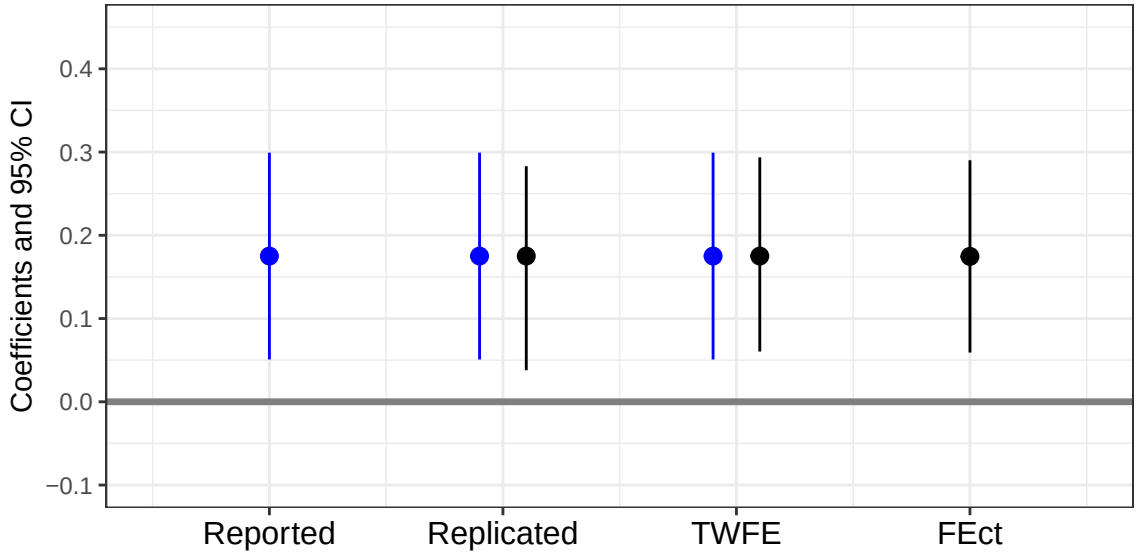
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this data) and apply two estimators: TWFE and imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(p_vreg_bnw~post_x_covered|county_id+year,data = df,cluster = "county_id")
summary(sol)
```

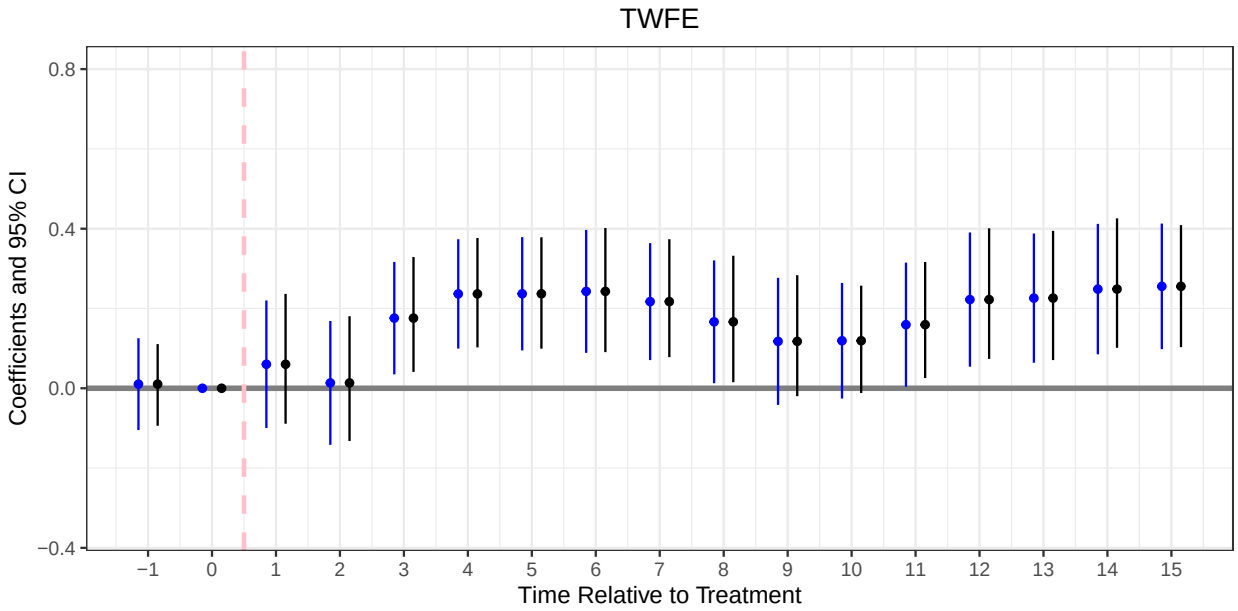
```
## OLS estimation, Dep. Var.: p_vreg_bnw
## Observations: 1,695
## Fixed-effects: county_id: 100,  year: 17
## Standard-errors: Clustered (county_id)
##                 Estimate Std. Error t value  Pr(>|t|)
## post_x_covered 0.175045   0.063368 2.76237 0.0068415 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.185574     Adj. R2: 0.60345
##                   Within R2: 0.021742
```

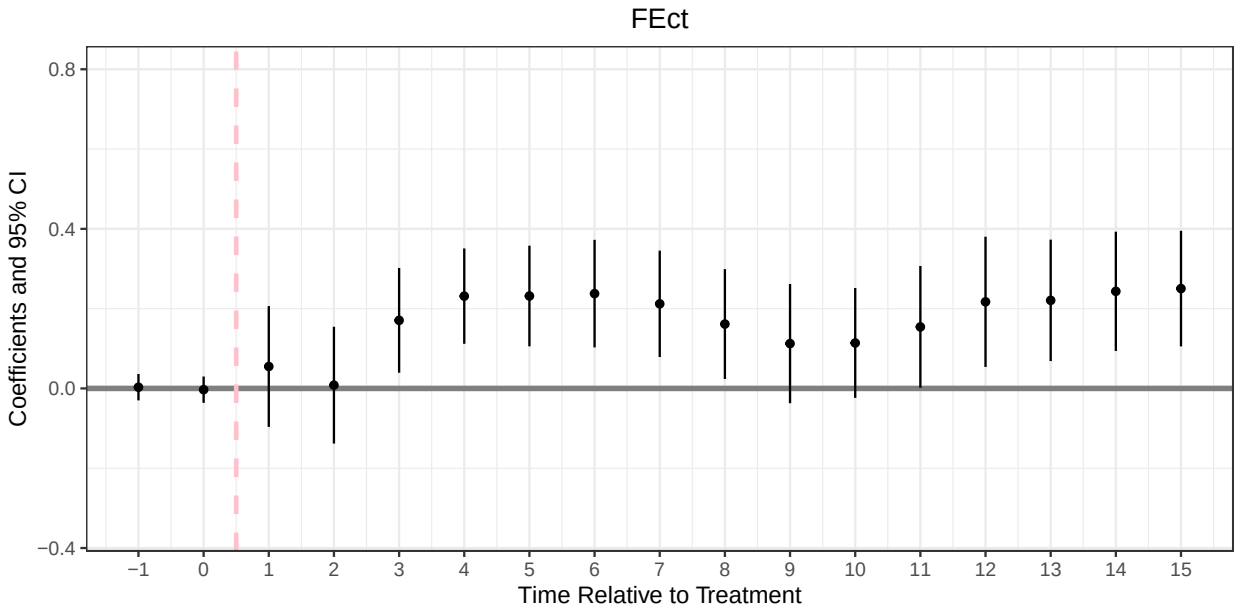


The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

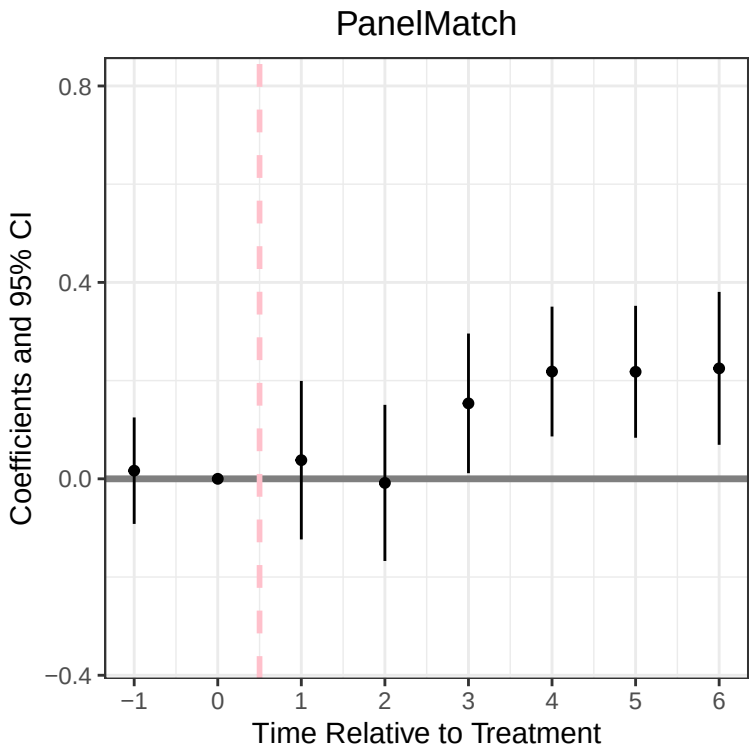




TWFE and FEct estimates are broadly consistent with each other. The estimated DTEs are positive on post-treatment periods. The number of pre-treatment periods is small, therefore, it is hard to tell if there exists some pre-trends.

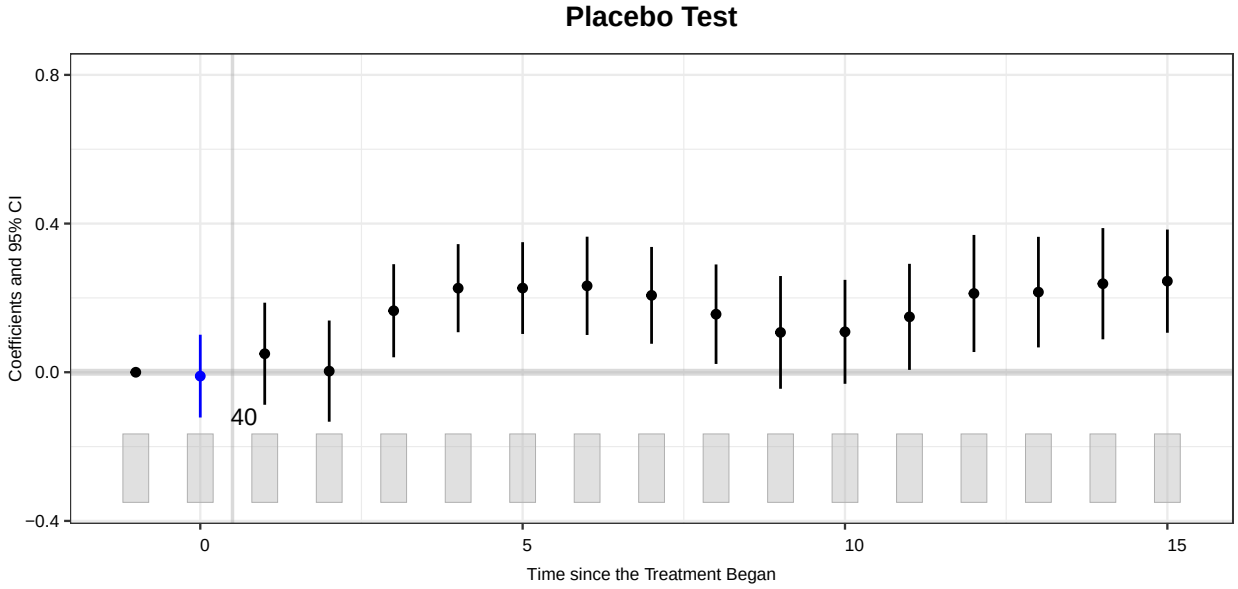
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 6 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



Test Statistics

```
##                                   p-value
## F test                              0.854
## Equivalence test (default)          0.000
## Equivalence test (threshold=ATT)    0.000
## Placebo test                        0.855
## Carryover effect test                  NA
```

Due to the small number of pre-treatment periods, we find little evidence for potential violations of the parallel trends (PT) assumption. The equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT.

Summary

Overall, the main result of the chosen model appears to be robust to the imputation estimator, FEct. We find little evidence for violations of the PT given the short pre-treatment periods. However, the limited number of pre-treatment periods restricts further investigation into the robustness of the results using placebo tests or sensitivity analyses.

Garfias (2019)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 6

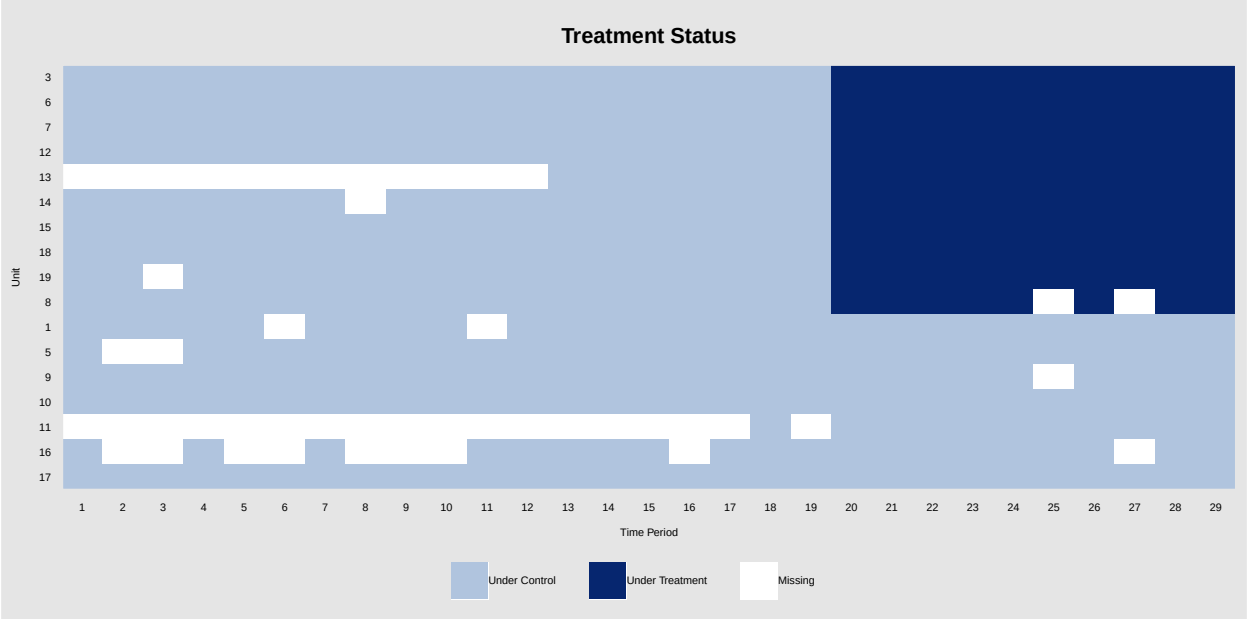
A First Look at Data

The paper investigates effects of Mining Tribunal on fiscal capacity investments, using Mexican treasury-year level panel data, between 1758 and 1786. One of the main findings of this paper is that "the Tribunal led to a substantial increase in relative civil administration expenditures in mining treasuries" (p105).

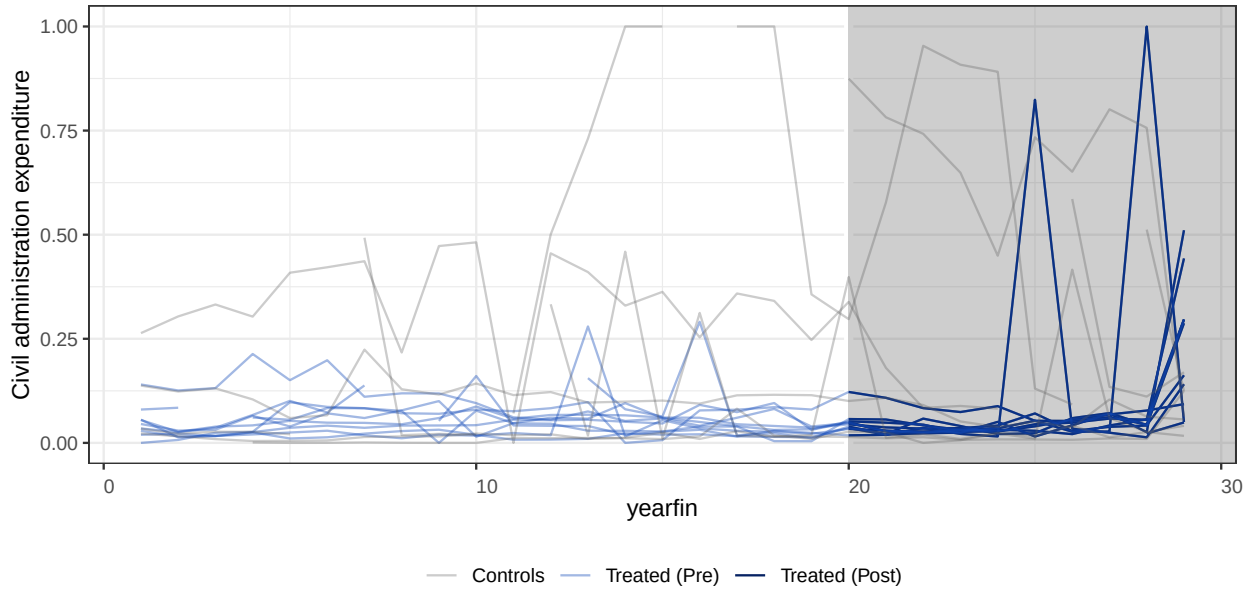
Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	Treasury × year
Treatment	Mining Tribunal
Outcome	Civil administration expenditure
Treatment type	Classic
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a treasury).



Plotting the outcome variable. We plot the trajectory of the outcome variable for each treasury. The control units are represented in gray. We highlight the observations of treated units under untreated and treated status using light blue and deep blue, respectively.



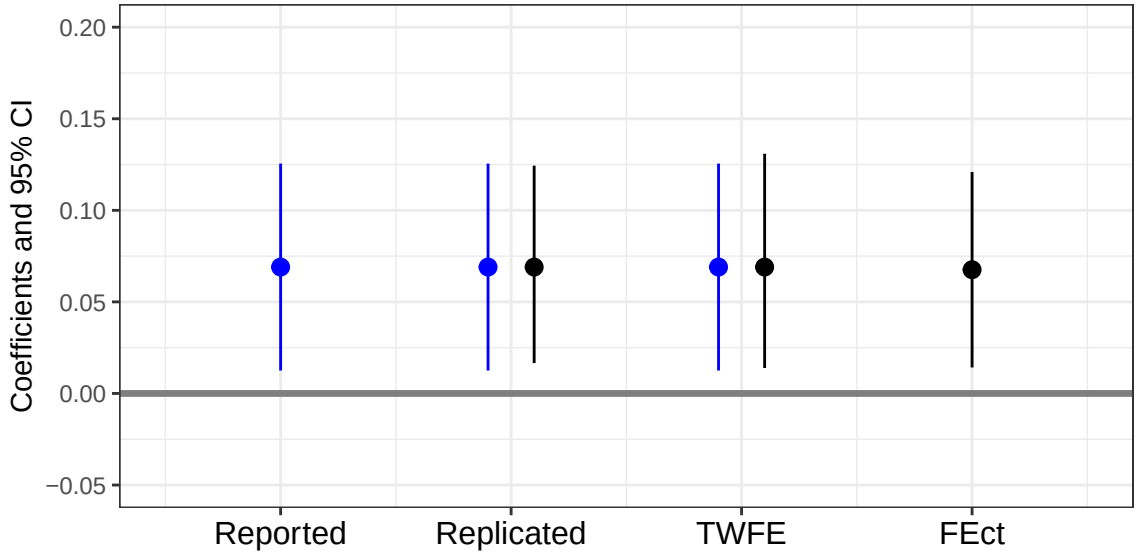
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this data) and apply two estimators: TWFE and imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(prRec~tribunalm|cajacode+yearfin,data = df,cluster = "cajacode")
summary(sol)
```

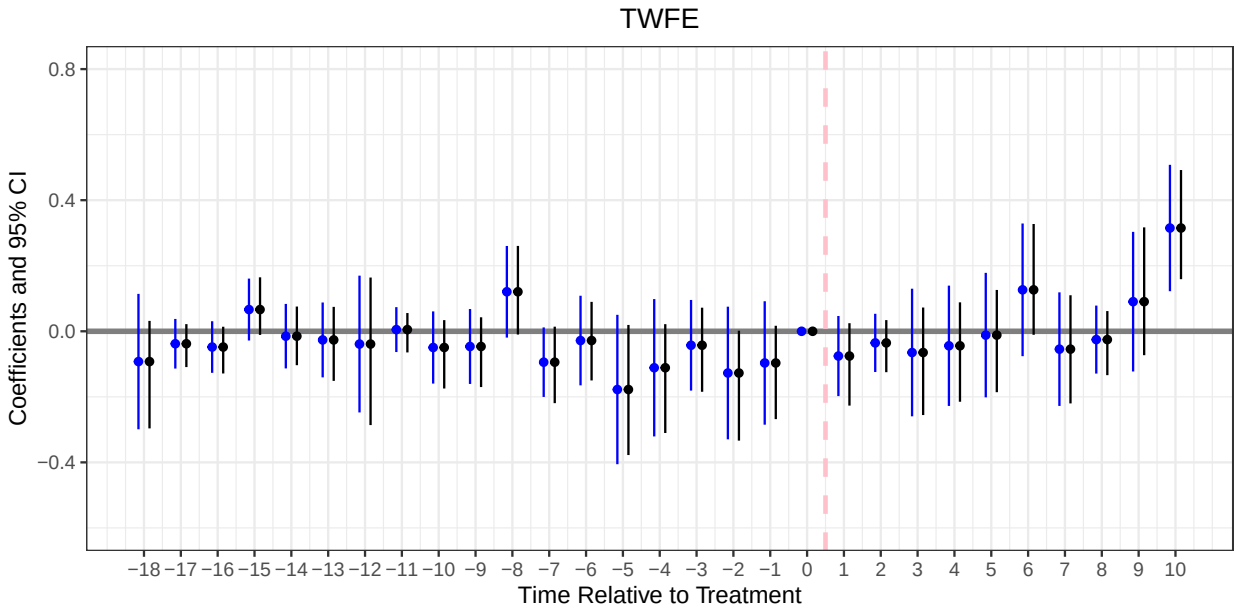
```
## OLS estimation, Dep. Var.: prRec
## Observations: 445
## Fixed-effects: cajacode: 17,  yearfin: 29
## Standard-errors: Clustered (cajacode)
##           Estimate Std. Error t value Pr(>|t|)
## tribunalm 0.069018    0.02882 2.39474 0.029223 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.121703     Adj. R2: 0.552755
##                  Within R2: 0.016603
```

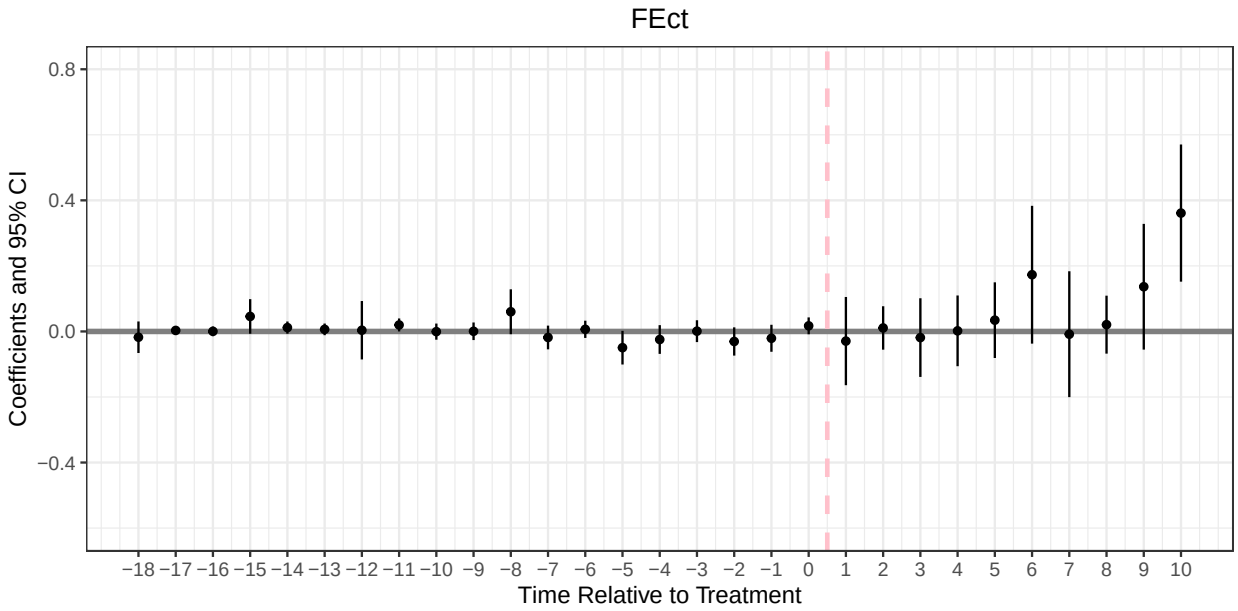


The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

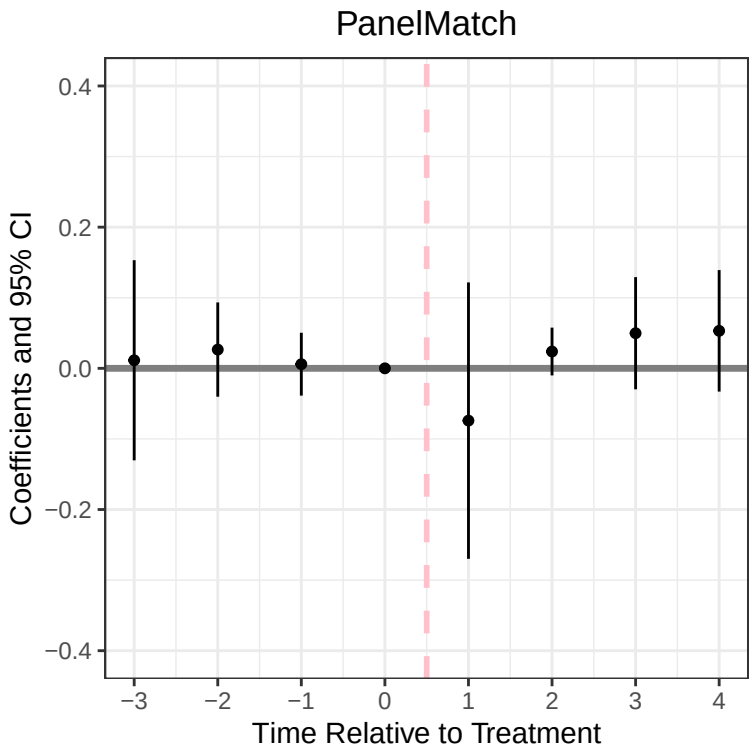




TWFE and FEct estimates are broadly consistent with each other. The estimated DTEs are close to 0 on most post-treatment periods. The estimated DTEs using TWFE exhibits a weak upward pre-trend.

DTE of a Balanced Subsample

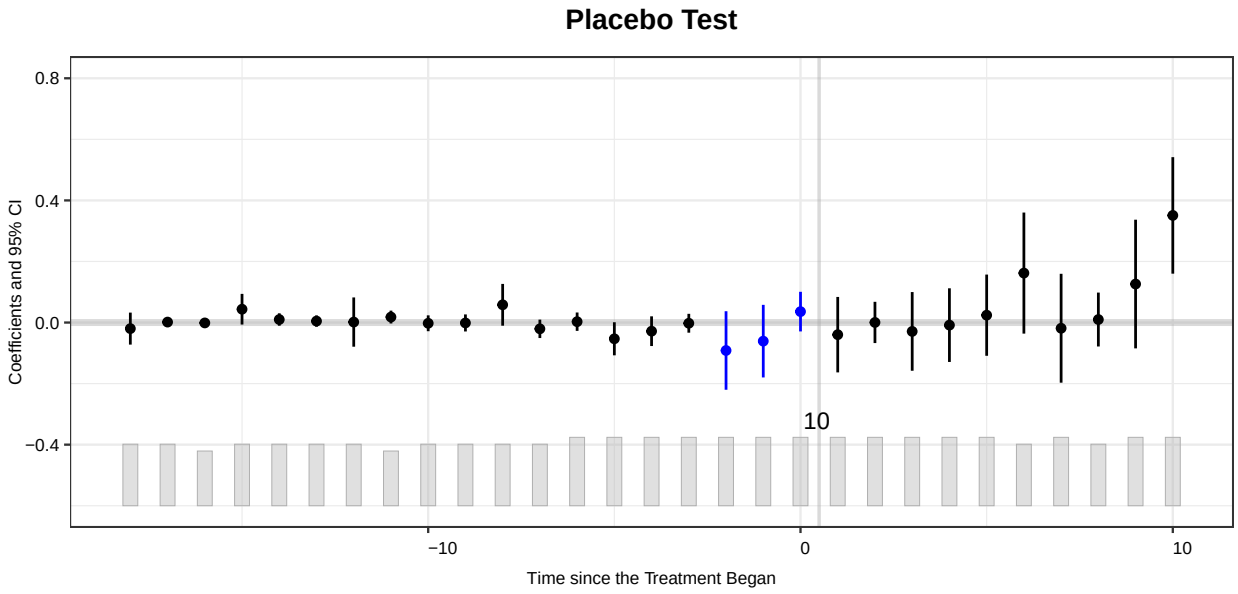
We also estimate the DTE using PanelMatch (lead = 4 and lag = 4). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



FEct and PanelMatch estimates broadly agree with each other.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.



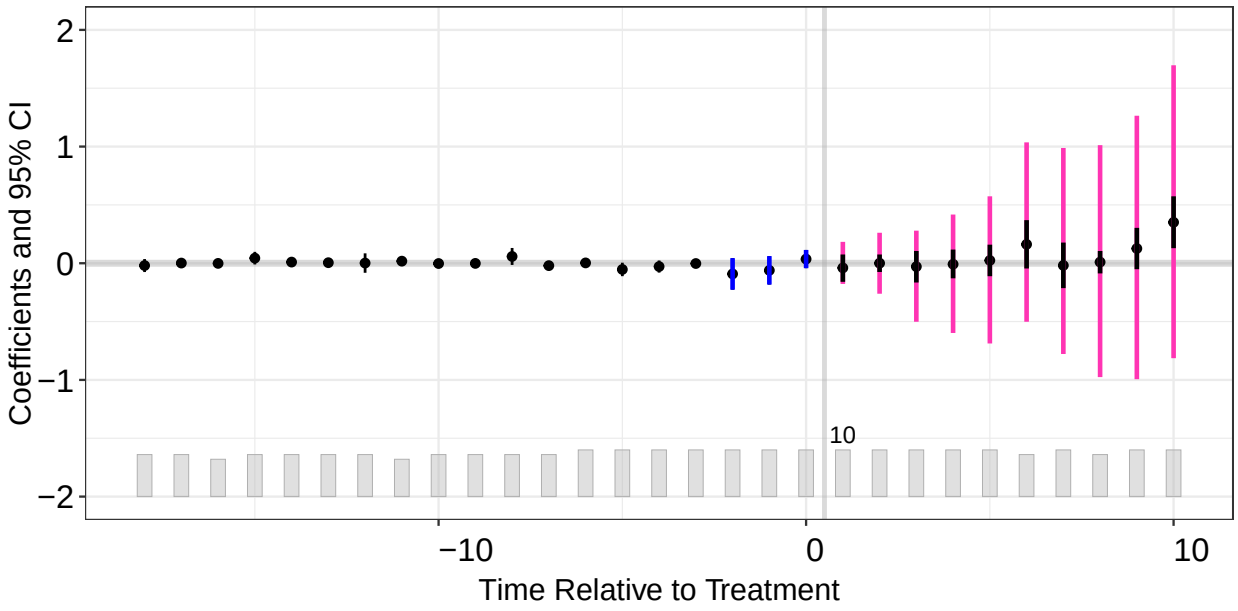
Test Statistics

```
##                                  p-value
## F test                             0.901
## Equivalence test (default)         0.680
## Equivalence test (threshold=ATT)   0.415
## Placebo test                       0.297
## Carryover effect test                 NA
```

We find little evidence for violations of the parallel trends (PT) assumption. However, the equivalence test fails to reject the null that the residuals in pre-treatment periods exceed the estimated ATT, possibly due to limited power.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is no longer significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, the main result of the chosen model seems to be robust to FEct, an HTE-robust estimator. TWFE and FEct estimators produce similar positive and statistically significant point estimates. However, the result appears to be sensitive to the removal of placebo periods and mild violations of the PT assumption benchmarked against the pre-treatment placebo estimates, possibly due to limited power.

Grumbach and Sahn (2020)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

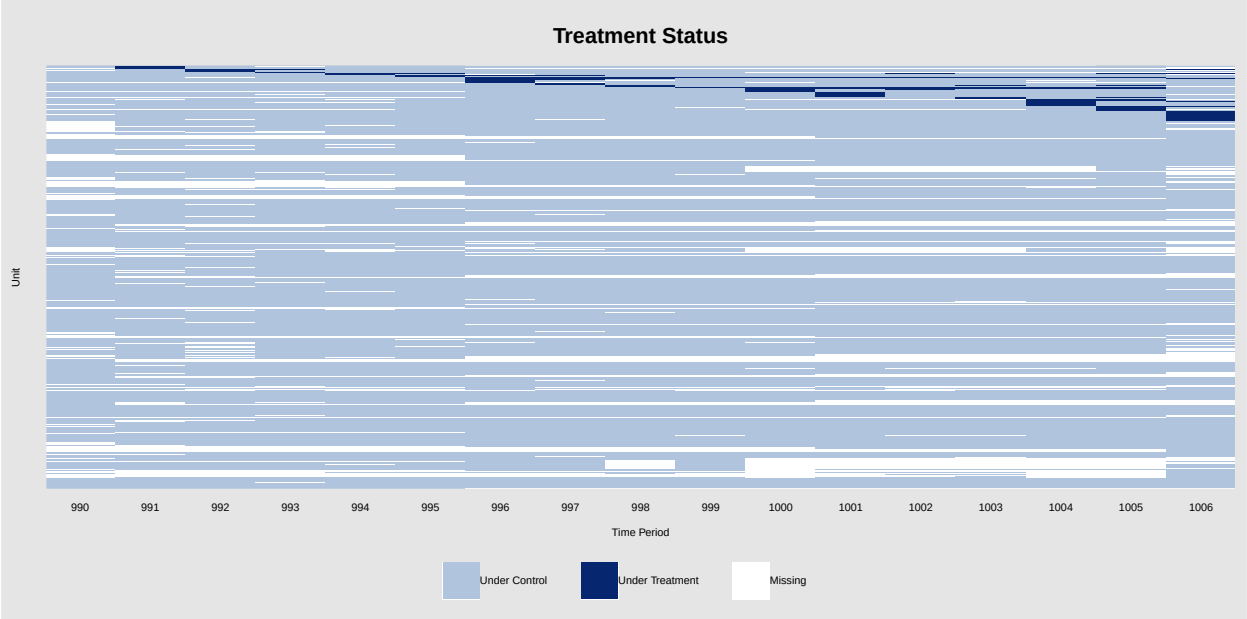
A First Look at Data

The paper investigates the effects of Asian (Black/Latino) candidates in general elections on the share of campaign contributions by Asian (Black/Latino) donors, using district-election year level panel data from US House general elections, between 1980 and 2012. One of the main findings of this paper is that "increase in coethnic contribution shares from Asian, black, or Latino donors substitutes for white contribution shares" (p10).

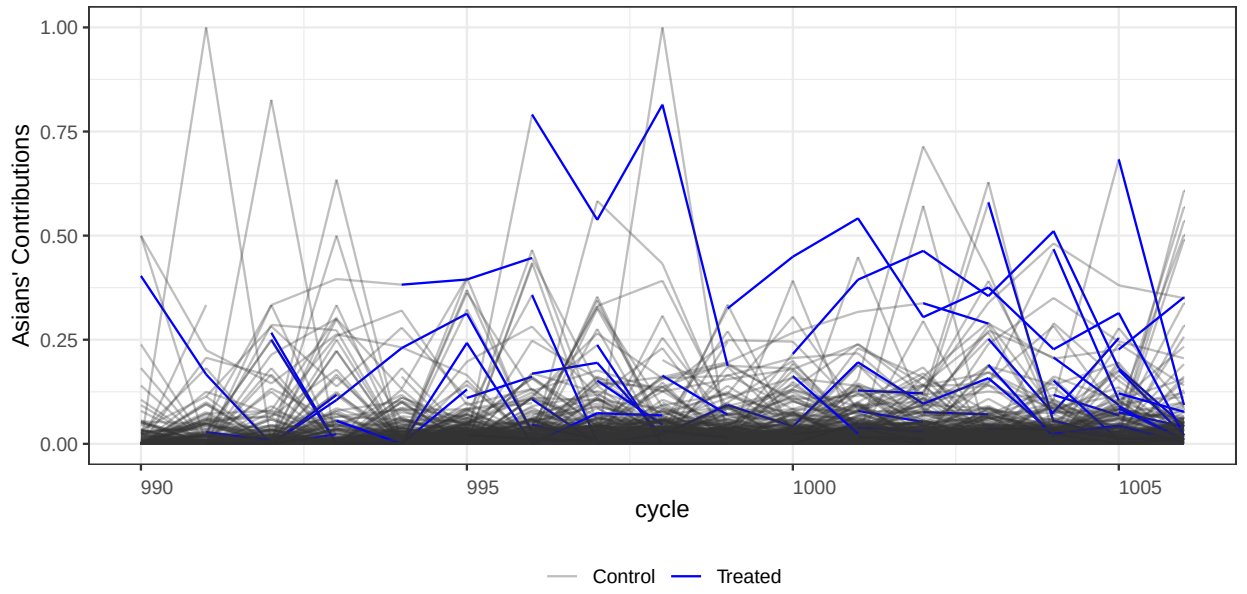
Model. For simplicity, we focus on the top left panel of **Figure 5** in the paper, the effects of Asian candidates and the corresponding specification at of Figure 5 in the paper.

Replication Summary	
Unit of analysis	District × election year
Treatment	Asian (Black/Latino) candidates in general elections
Outcome	Share of campaign contributions by Asian (Black/Latino) donors
Treatment type	General
Outcome type	continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (an election year) and each row represents a unit (a district). There are treatment reversals and some missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each district. The observations under treated status are marked in blue.



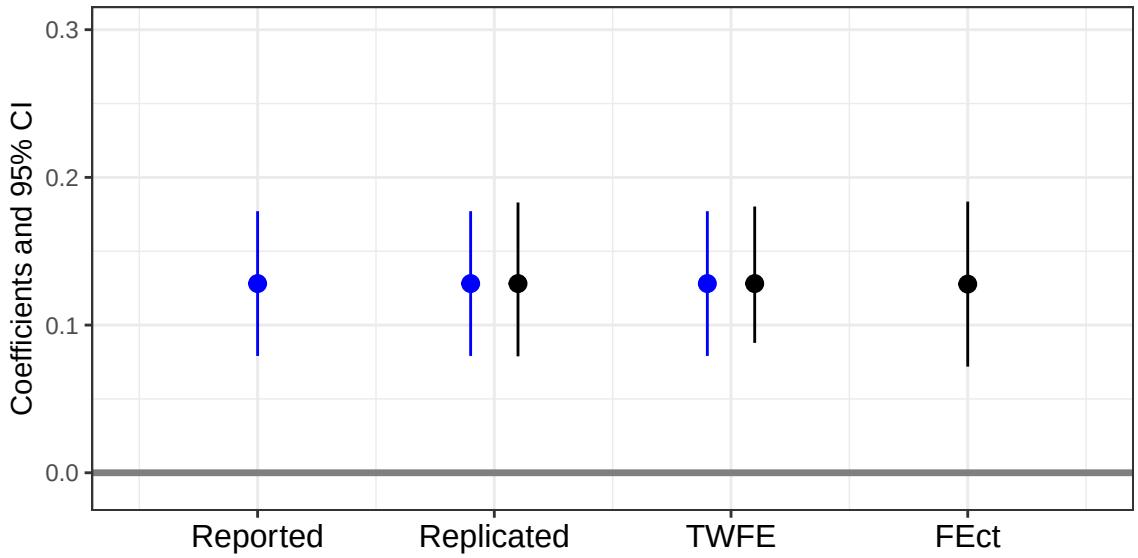
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation estimator (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```r
sol <- feols(general_sharetotal_A_all~cand_A_all+cand_H_all+cand_B_all|district_final+cycle,
             data = df,cluster = "district_final")
summary(sol)
```

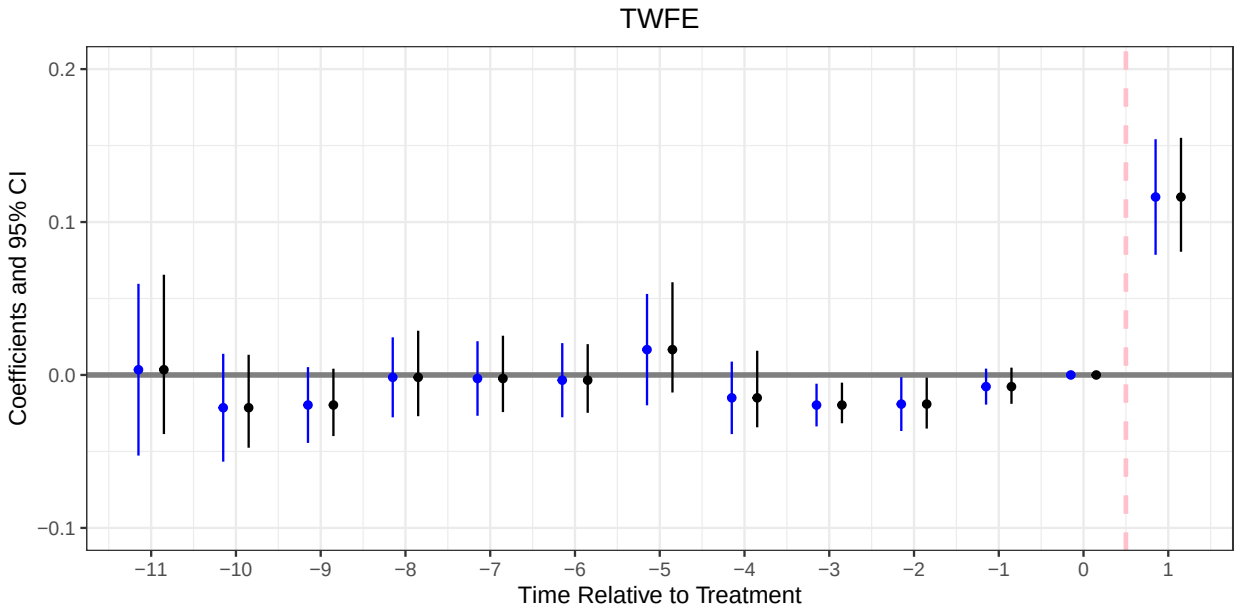
```
## OLS estimation, Dep. Var.: general_sharetotal_A_all
## Observations: 6,847
## Fixed-effects: district_final: 489,  cycle: 17
## Standard-errors: Clustered (district_final)
##             Estimate Std. Error t value   Pr(>|t|)
## cand_A_all 0.128098   0.025007 5.12256 4.3480e-07 ***
## cand_H_all 0.012181   0.005274 2.30980 2.1316e-02 *
## cand_B_all 0.010456   0.004594 2.27615 2.3269e-02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.049535     Adj. R2: 0.275898
##                  Within R2: 0.083964
```

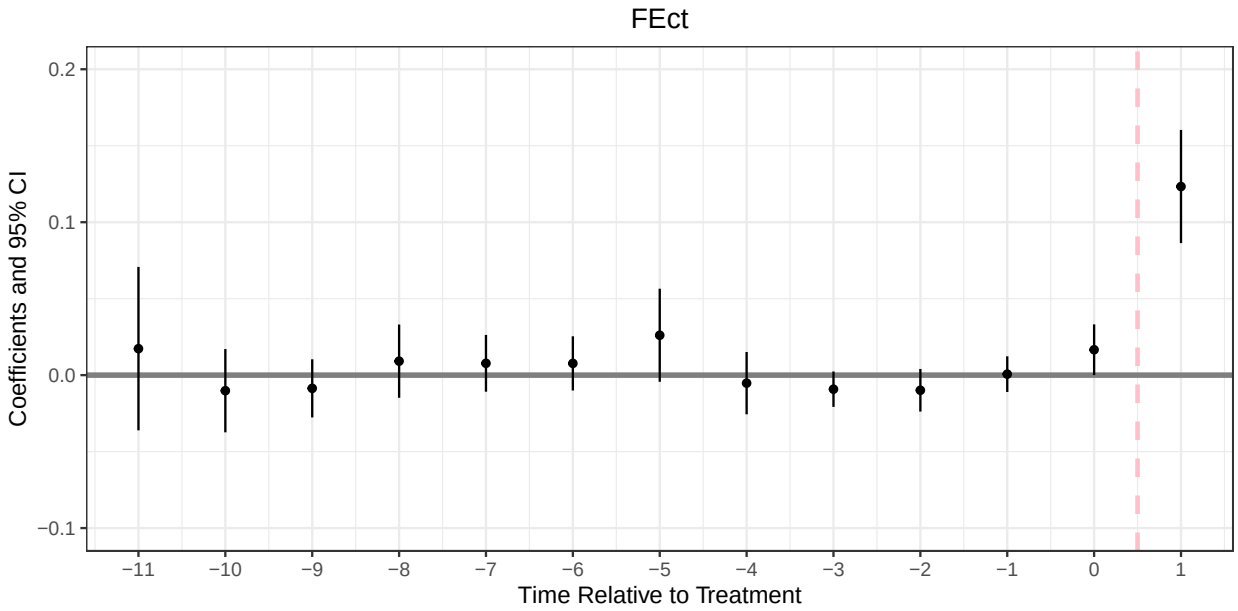


The TWFE and the imputation estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

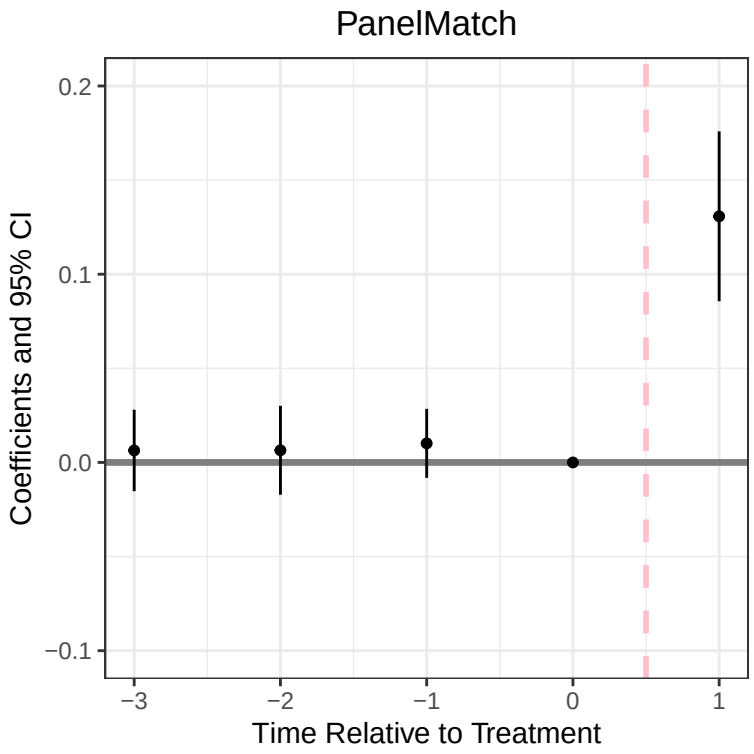




TWFE and imputation estimates agree with each other. The estimated DTEs are positive in the first post-treatment period.

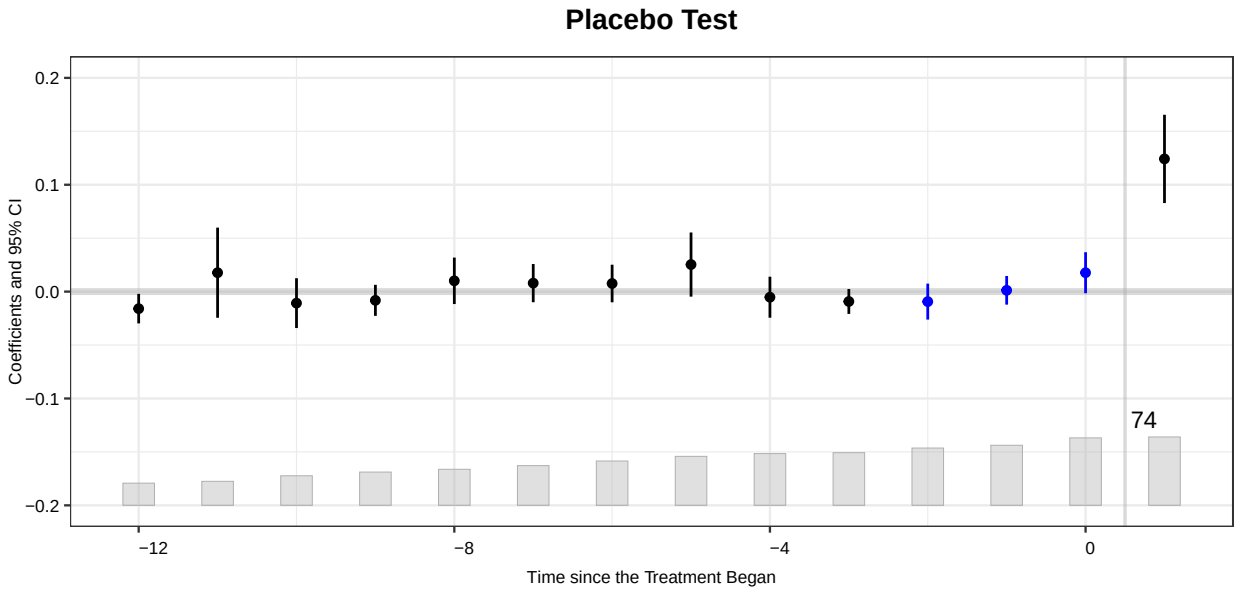
DTE of a Balanced Subsample

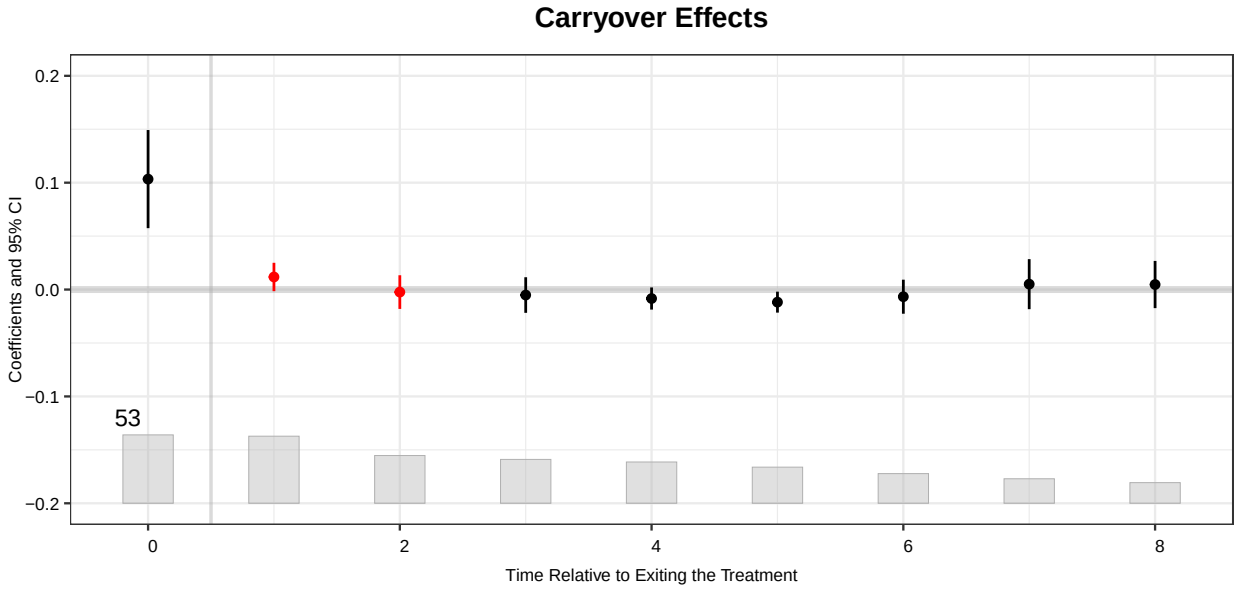
We also estimate the DTE using PanelMatch (lead = 1 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





Test Statistics

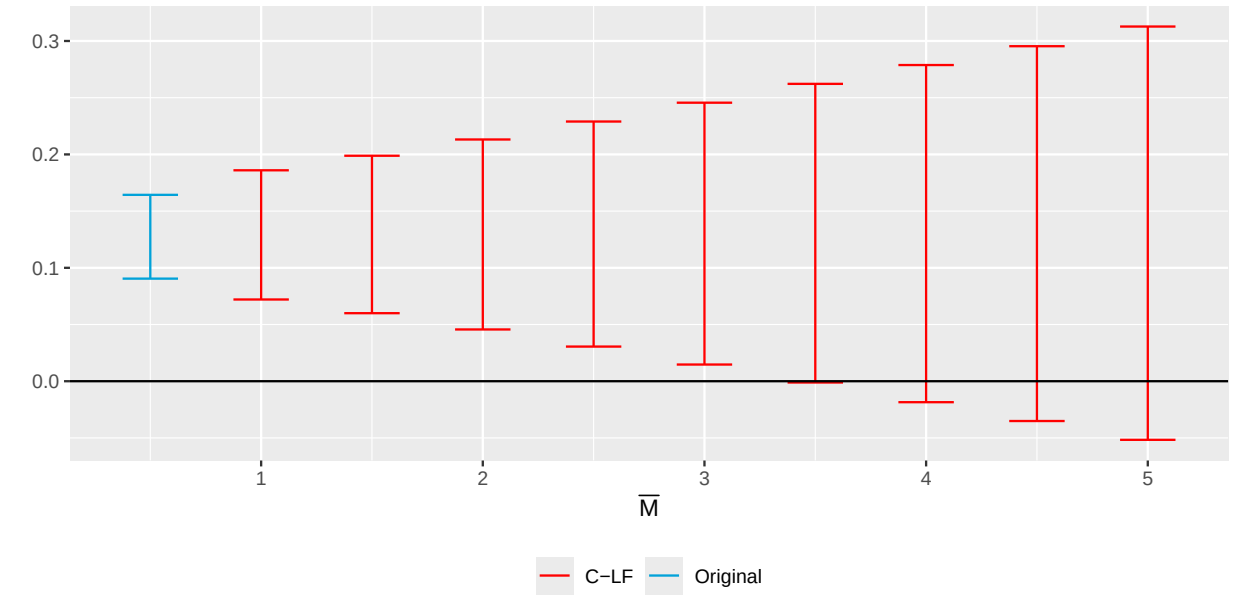
```
##                                    p-value
## F test                               0.423
## Equivalence test (default)           0.732
## Equivalence test (threshold=ATT)     0.000
## Placebo test                         0.551
## Carryover effect test                0.225
```

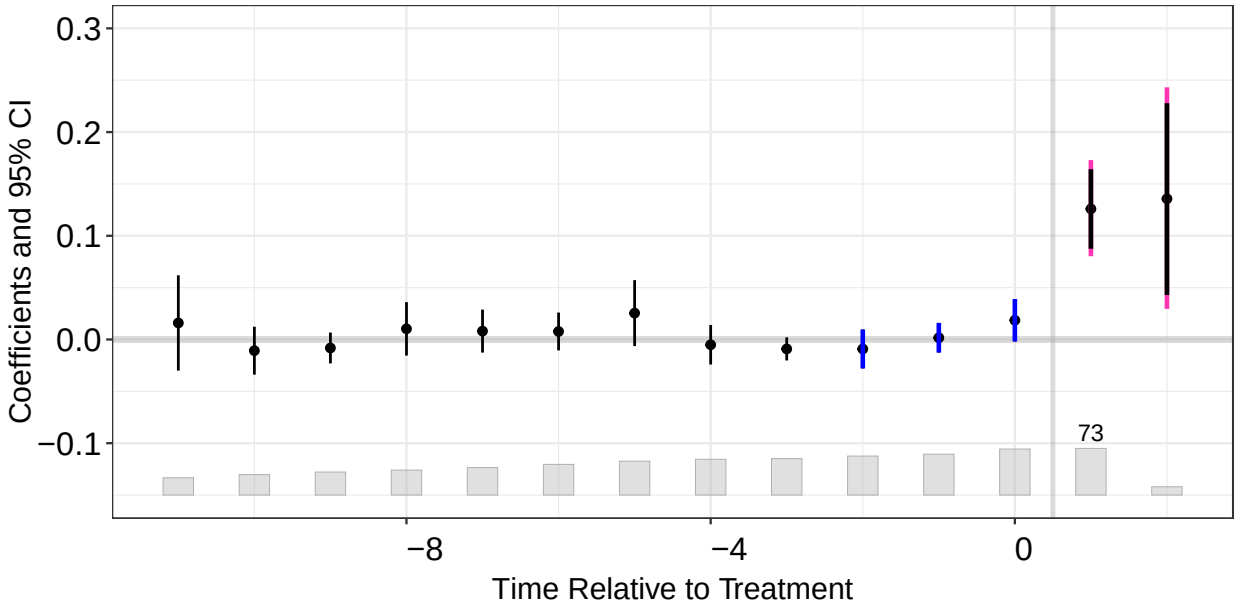
We find little evidence for violations of the parallel trends (PT) assumption and no-carryover-effect assumption. The equivalence test also reject the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the imputation estimator remains significant. We assess the robustness of this finding using the method proposed by Rambachan and Roth (2022). We use the three placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 3.5$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 3.5 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust to HTE-robust estimators. We find no evidence for violations of the PT assumption, and the equivalence test rejects. The estimated ATT appears to be robust to mild PT violations in sensitivity analysis with the breakdown threshold $M = 3.5$.

Grumbach and Hill (2022)

Contents

Summary 1
Point Estimates 2
Dynamic Treatment Effects 3
Sensitivity Analysis 6
Summary 7

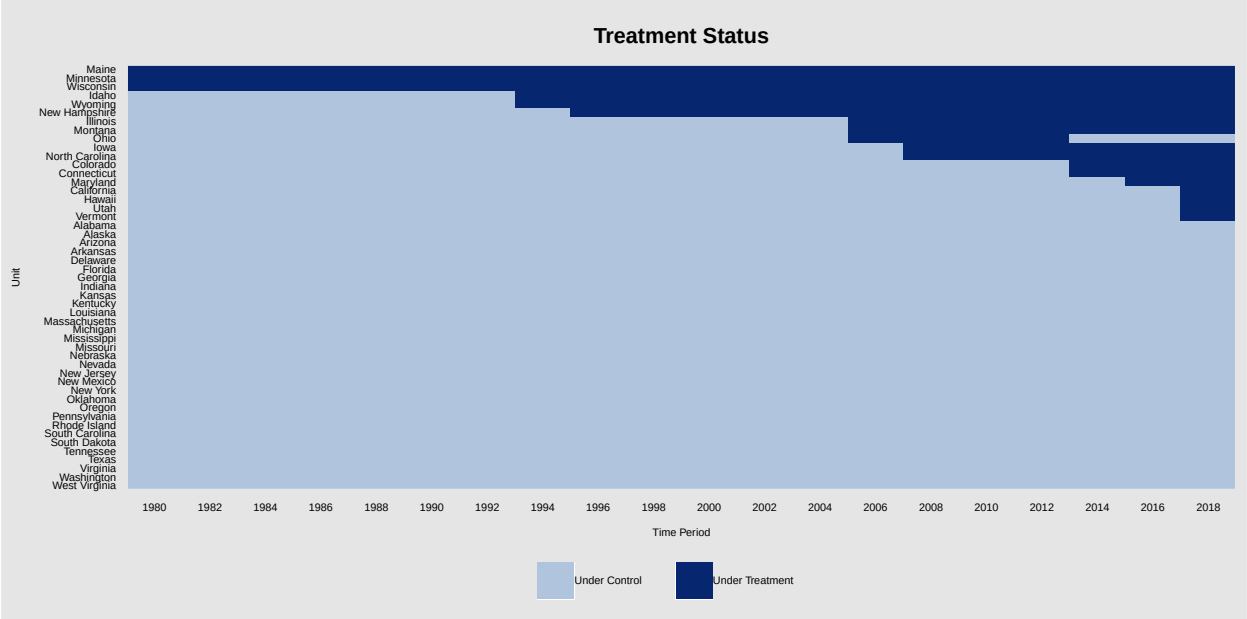
Summary

This paper investigates the effects of same day registration (SDR) laws on voter turnout, using country-year panel data from US state-year level TSCS data during 1980-2018. The paper finds that "SDR disproportionately increases turnout among individuals aged 18 to 24" (p405).

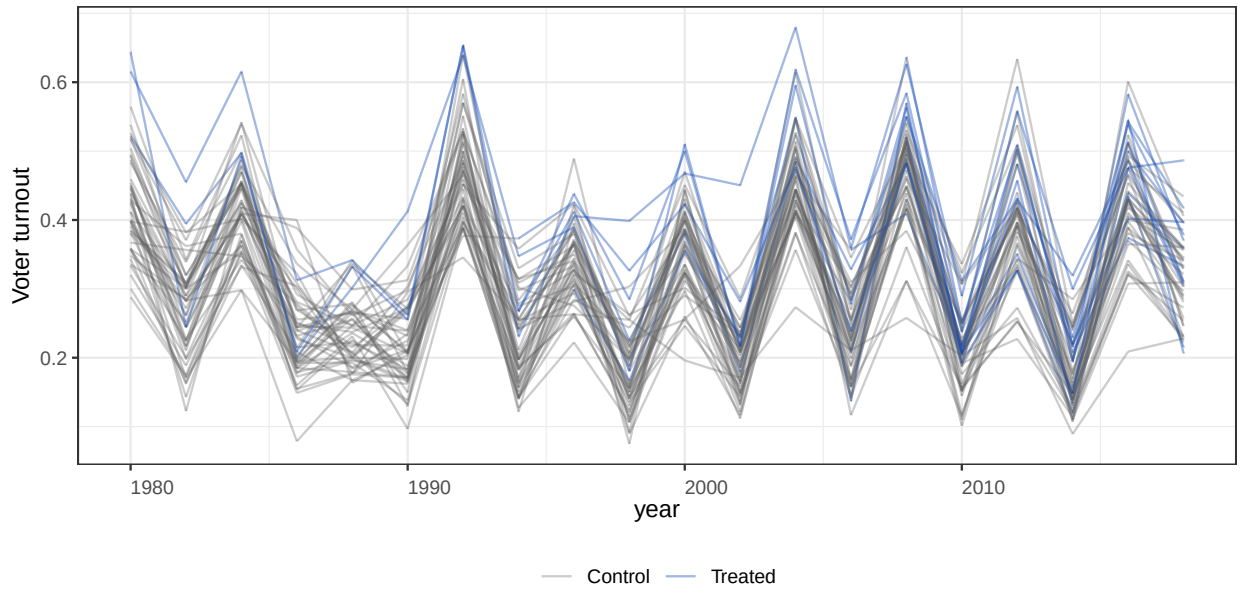
Model. We focus on state-level analysis of **Figure 3** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the state level.

Replication Summary	
Unit of analysis	State × year
Treatment	Same day registration (SDR) laws
Outcome	Voter turnout
Treatment type	General
Outcome type	Continuous
Fixed effects	Unit+Time

View treatment status First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a state. The treatment has reversals.



View the outcome We plot the trajectory of the outcome variable for each state. The observations under treated status are marked in blue.



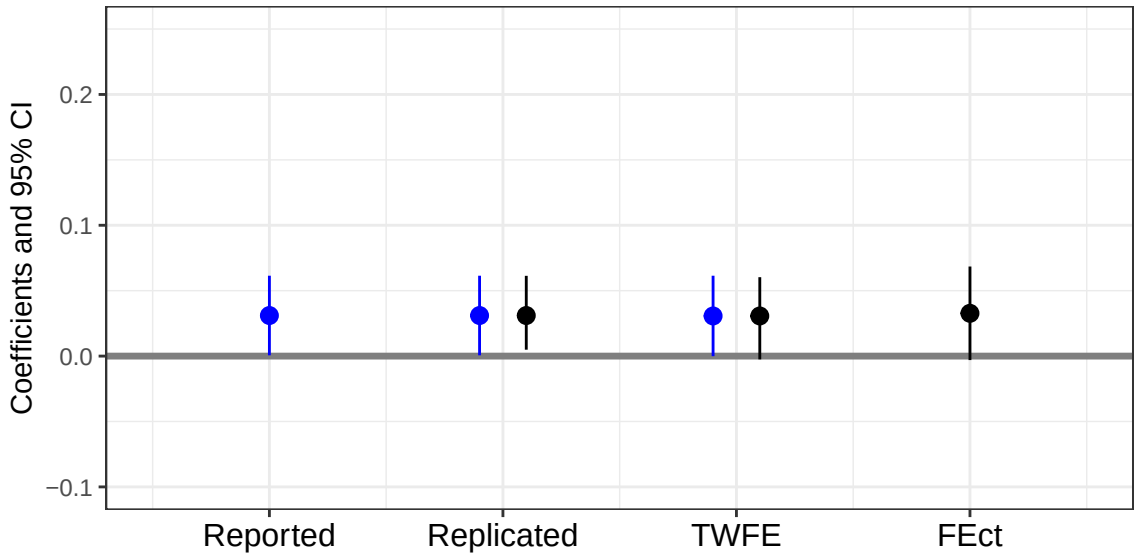
Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

```r
sol <- feols(voted~temp_var|state+year,data = df,cluster = "state")
summary(sol)
```

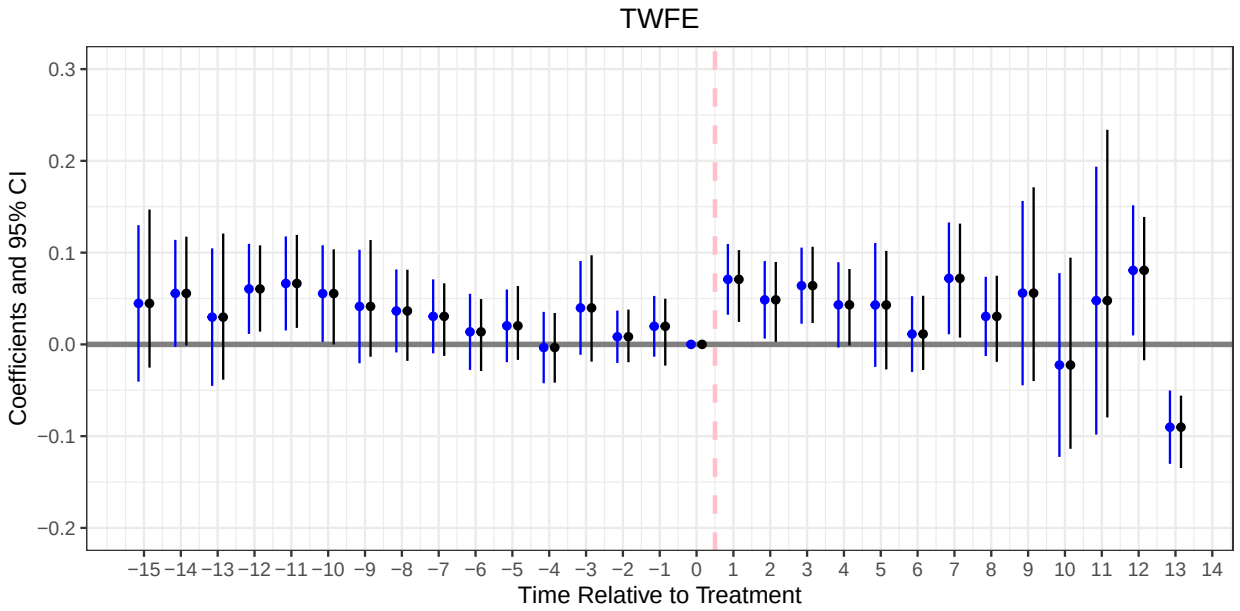
```
## OLS estimation, Dep. Var.: voted
## Observations: 980
## Fixed-effects: state: 49,  year: 20
## Standard-errors: Clustered (state)
##          Estimate Std. Error t value Pr(>|t|)
## temp_var 0.031006   0.015525 1.99724 0.051485 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.053673     Adj. R2: 0.804223
##                  Within R2: 0.012935
```

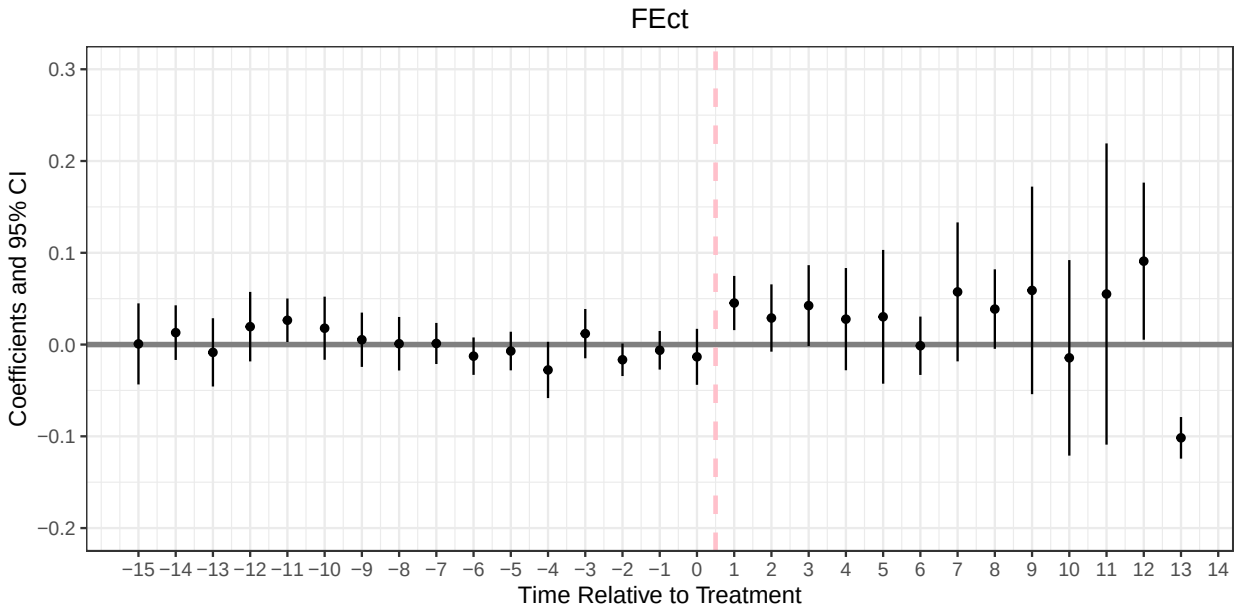


The TWFE and FEct estimator are consistent with each other. The estimated ATT are marginally significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

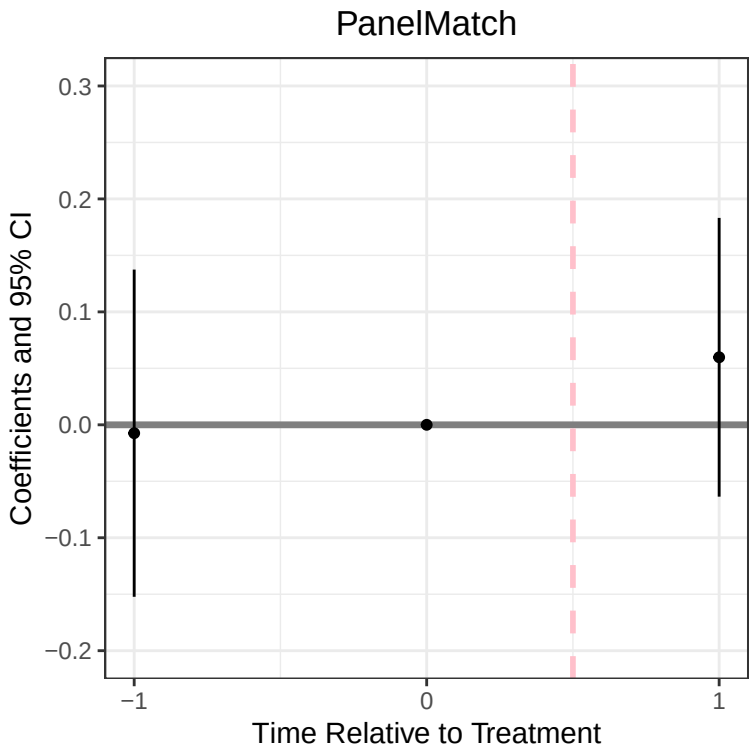




TWFE and FEct estimates are consistent with each other. The estimateds DTE are positive on most post-treatment periods. The estimated DTE using TWFE exhibits a weak downward pre-trend.

DTE of a Balanced Subsample

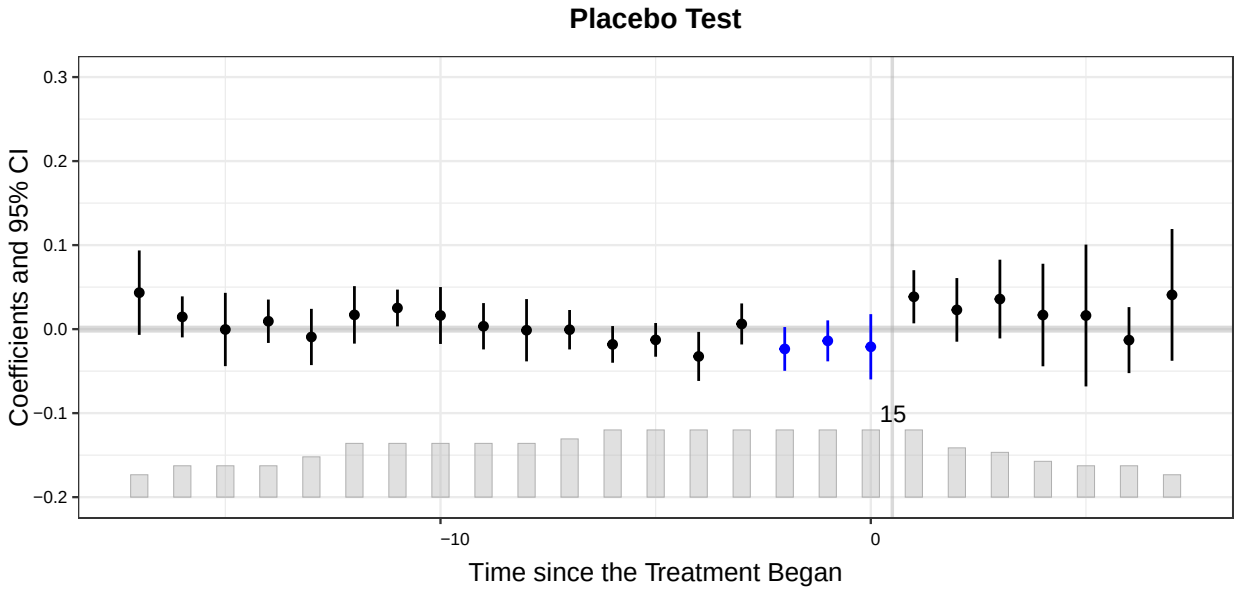
We also estimate the DTE using PanelMatch (lead = 1 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.

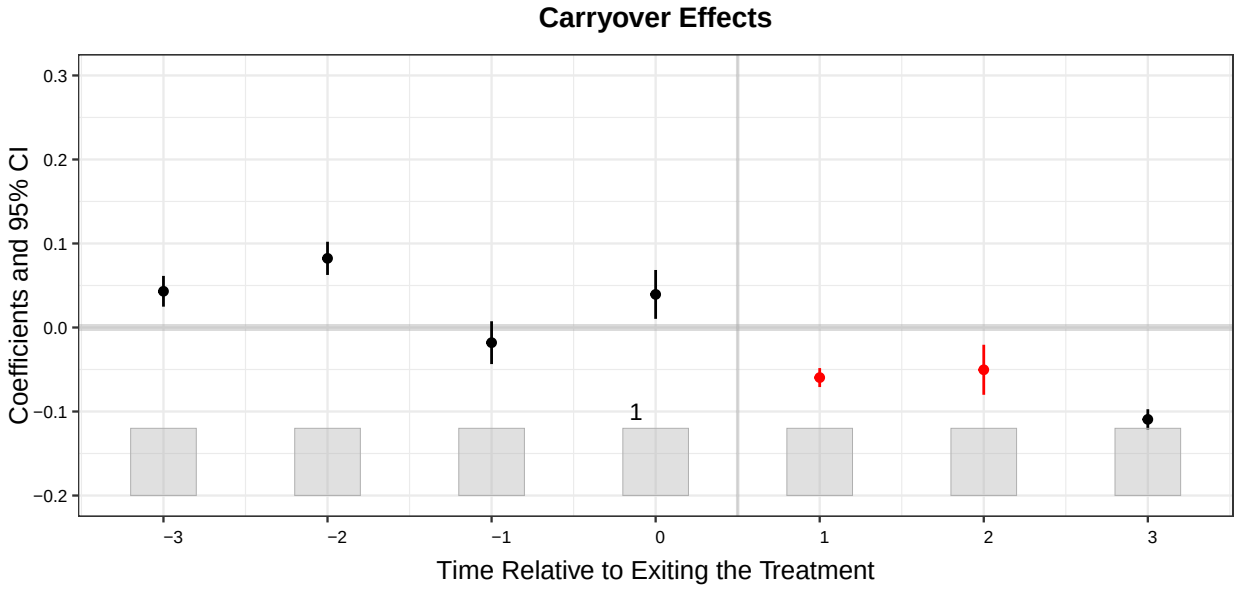


The FEct and PanelMatch estimates are broadly consistent with each other, though PanelMatch gives wider confidence intervals.

Placebo Test

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





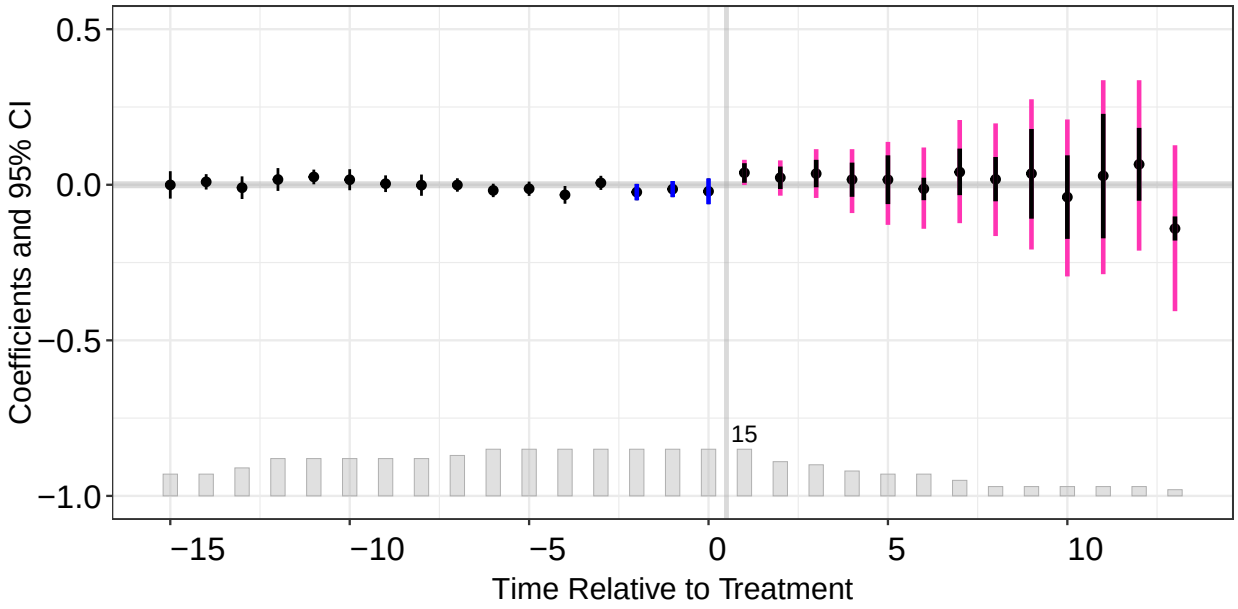
Test Results

```
##                                  p-value
## F test                             0.856
## Equivalence test (default)         0.724
## Equivalence test (threshold=ATT)   0.374
## Placebo test                       0.109
## Carryover effect test              0.000
```

We find little evidence for violations of the parallel trends (PT) assumption. However, the equivalence test fails to reject the null that the residuals in pre-treatment periods exceed the estimated ATT possibly due to limited power. The rejection in the carryover test casts some doubts on the no-carryover-effect assumption. As there is only one unit that undergoes a treatment reversal, this would have very limited impact on the conclusion.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, the main result of the chosen model appears to be robust to HTE-robust estimators. We find little evidence for violations of the PT. However, due to limited power, the result appears to be sensitive to the removal of placebo periods and potential violations of the PT assumption benchmarked against the pre-treatment placebo estimates.

Grumbach (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

A First Look at Data

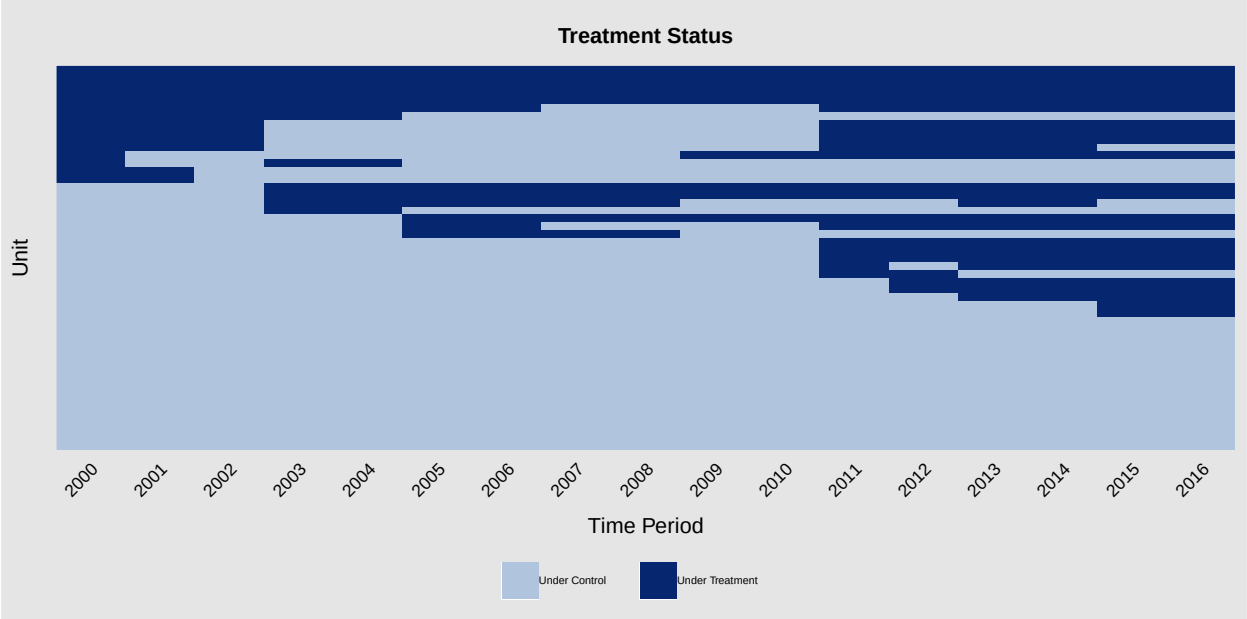
The paper investigates the effects of Republican control of state government on state democracy performance, using US state-year level panel data between 2000 and 2016. One of the main findings of this paper is "there is a large negative relationship between Republican control and democracy in the states." (p976)

Model. We focus on **Model 3 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the state level.

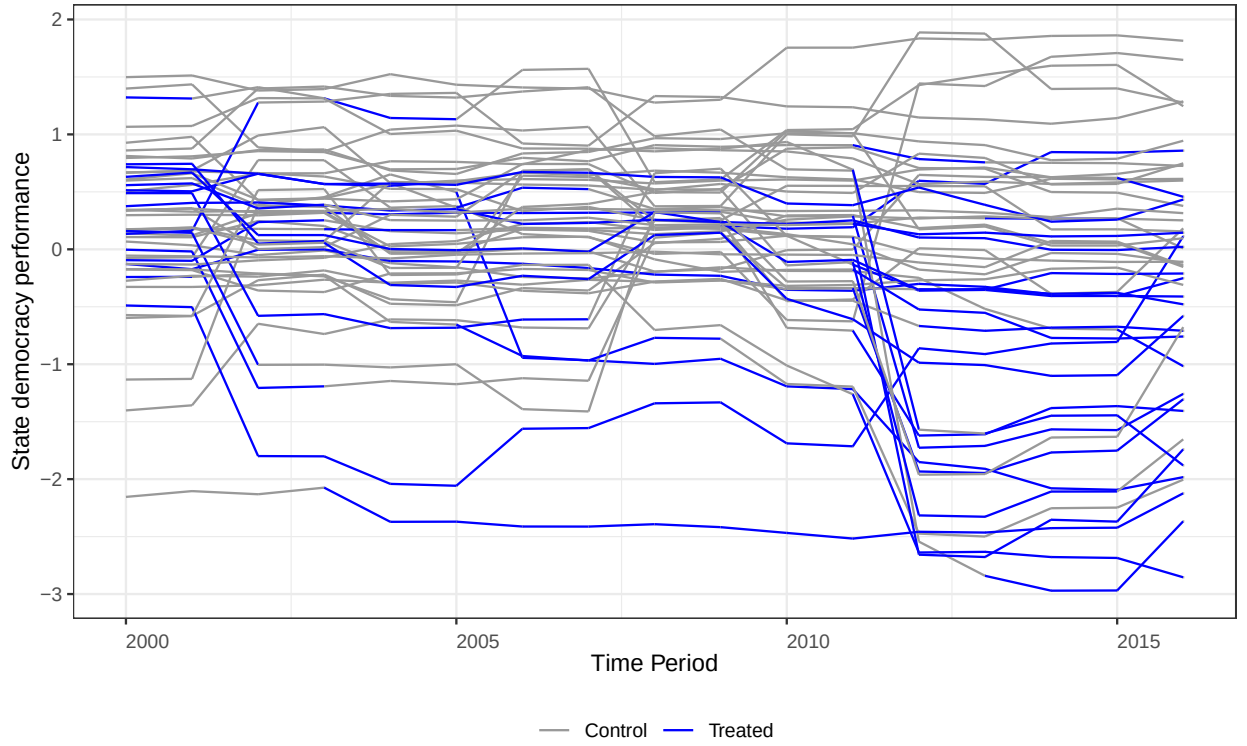
Table 1: Replication Summary

Unit of analysis	State × year
Treatment	Republican control of state government
Outcome	State democracy performance
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a state).



Plotting the outcome variable. We plot the trajectory of the outcome variable for each state. The observations under treated status are marked in blue.



Point Estimates

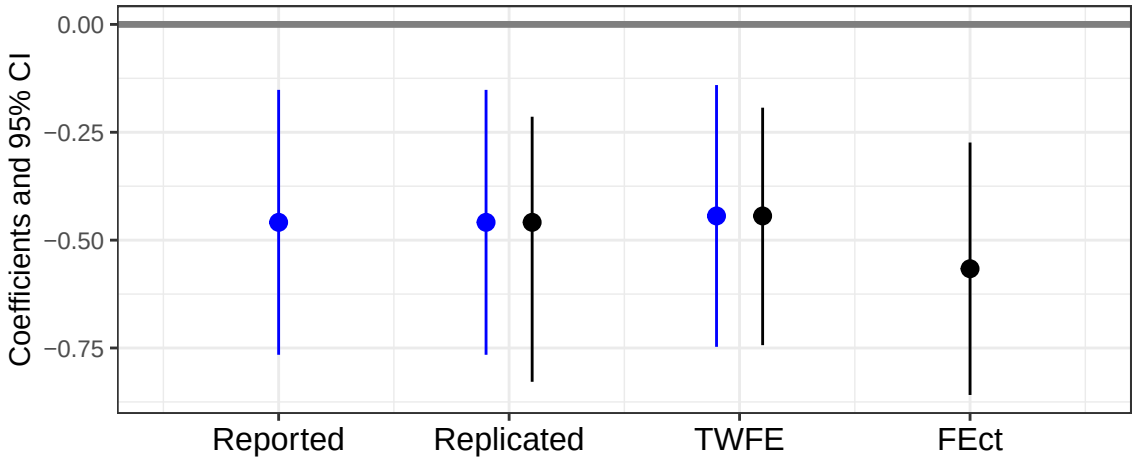
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates

and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(democracy_mcmc~Republican|state+year,
             data = df,cluster = "state")
summary(sol)
```

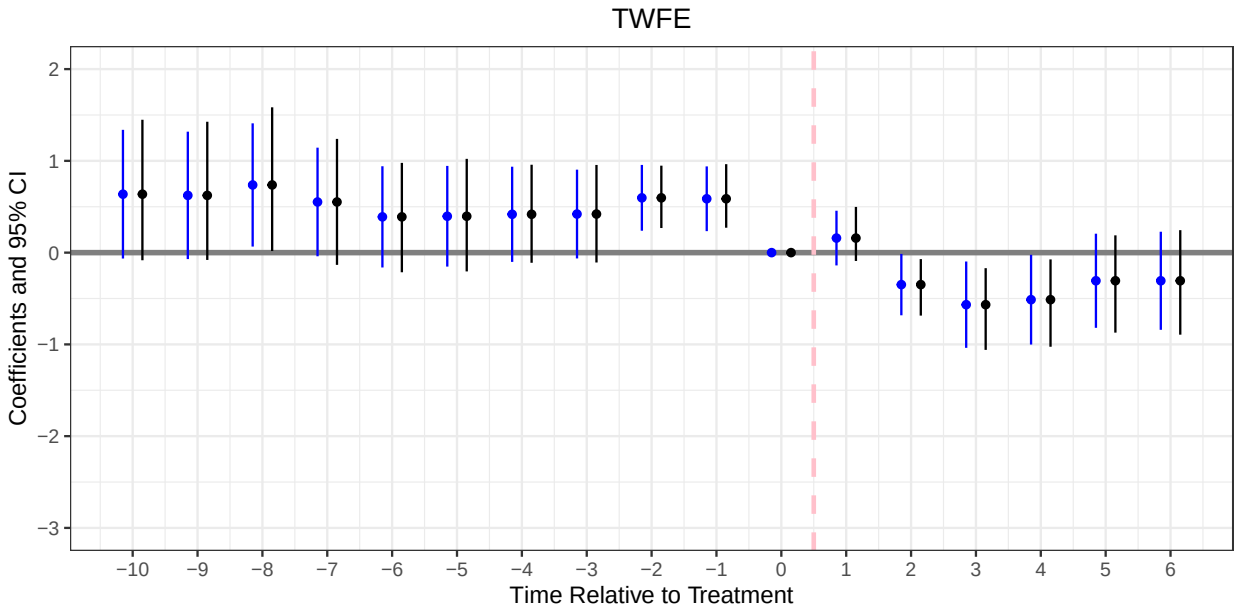
```
## OLS estimation, Dep. Var.: democracy_mcmc
## Observations: 833
## Fixed-effects: state: 49,  year: 17
## Standard-errors: Clustered (state)
##             Estimate Std. Error  t value  Pr(>|t|)
## Republican -0.458745   0.156641 -2.92863 0.0051942 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.498982     Adj. R2: 0.673875
##                  Within R2: 0.071515
```

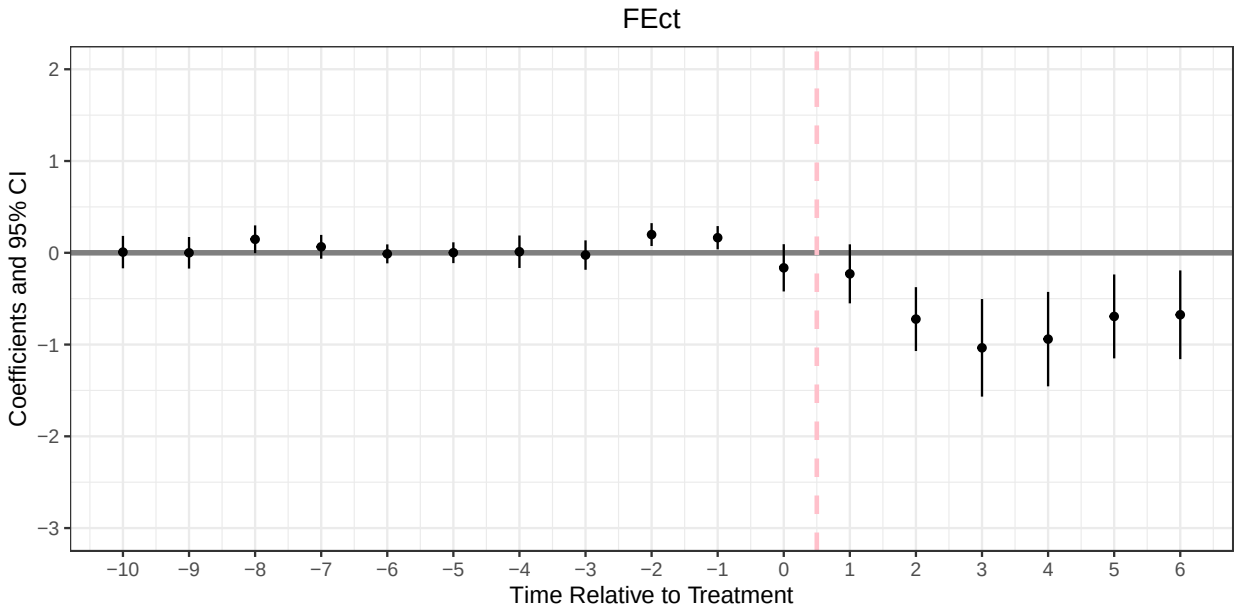


The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

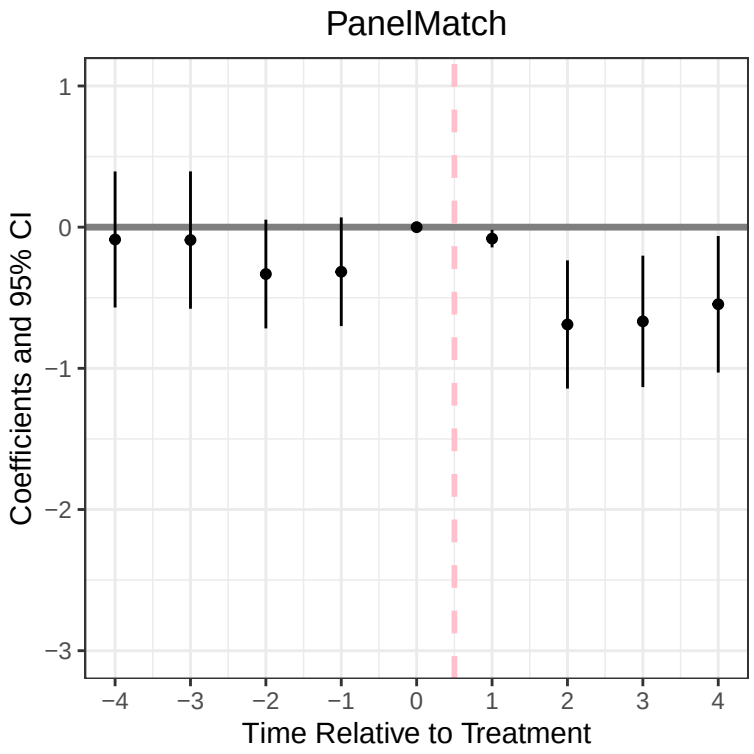
We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct.





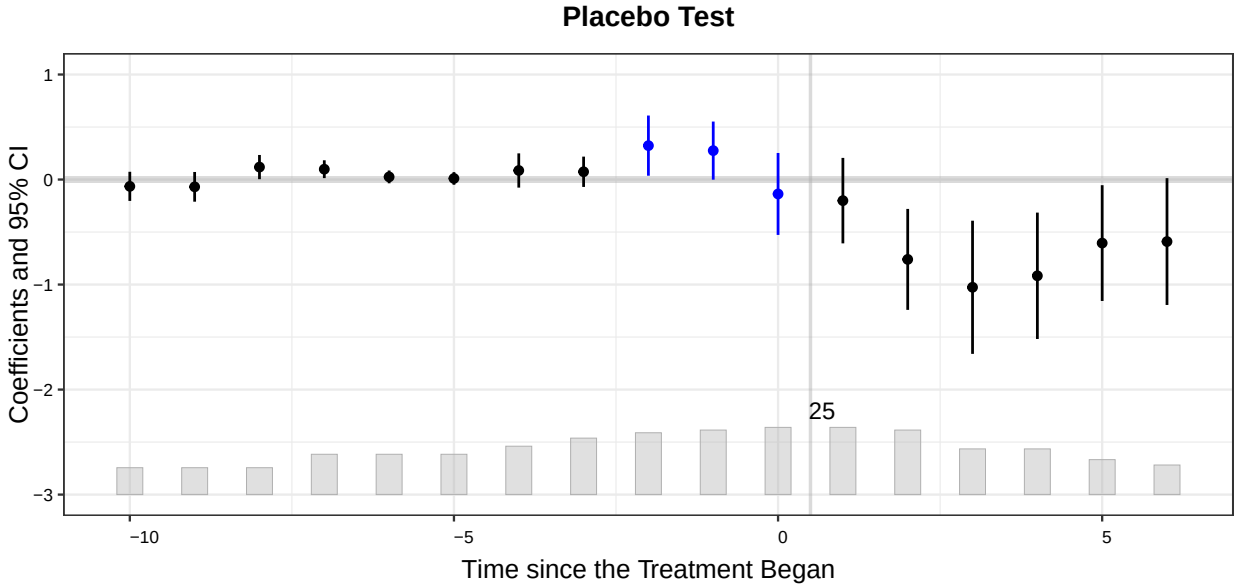
DTE of a Balanced Subsample

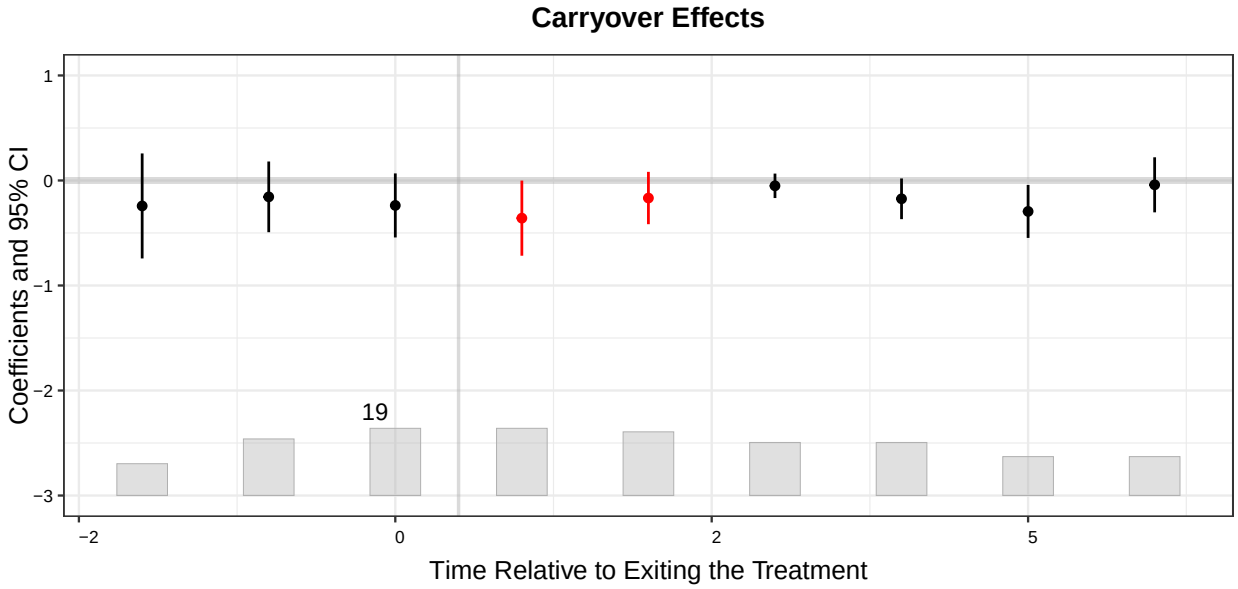
We also estimate the DTE using PanelMatch (lead = 4 and lag = 5). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





Test Statistics

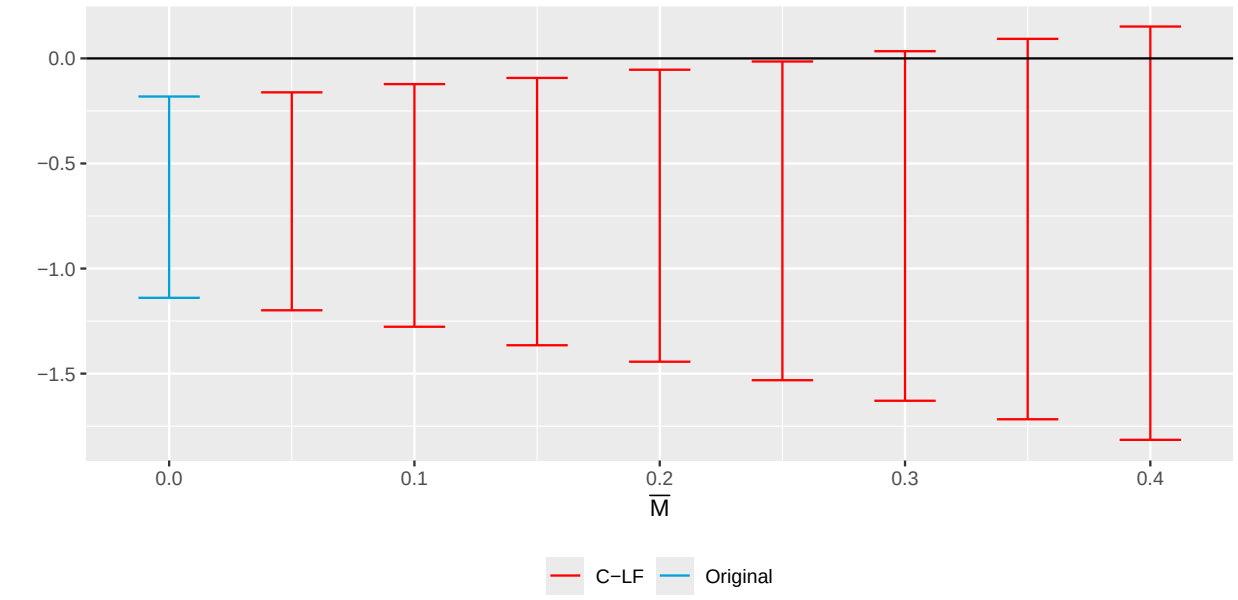
```
##                                  p-value
## F test                            0.1310
## Equivalence test (default)        0.7960
## Equivalence test (threshold=ATT)  0.0011
## Placebo test                      0.3150
## Carryover effect test             0.0640
```

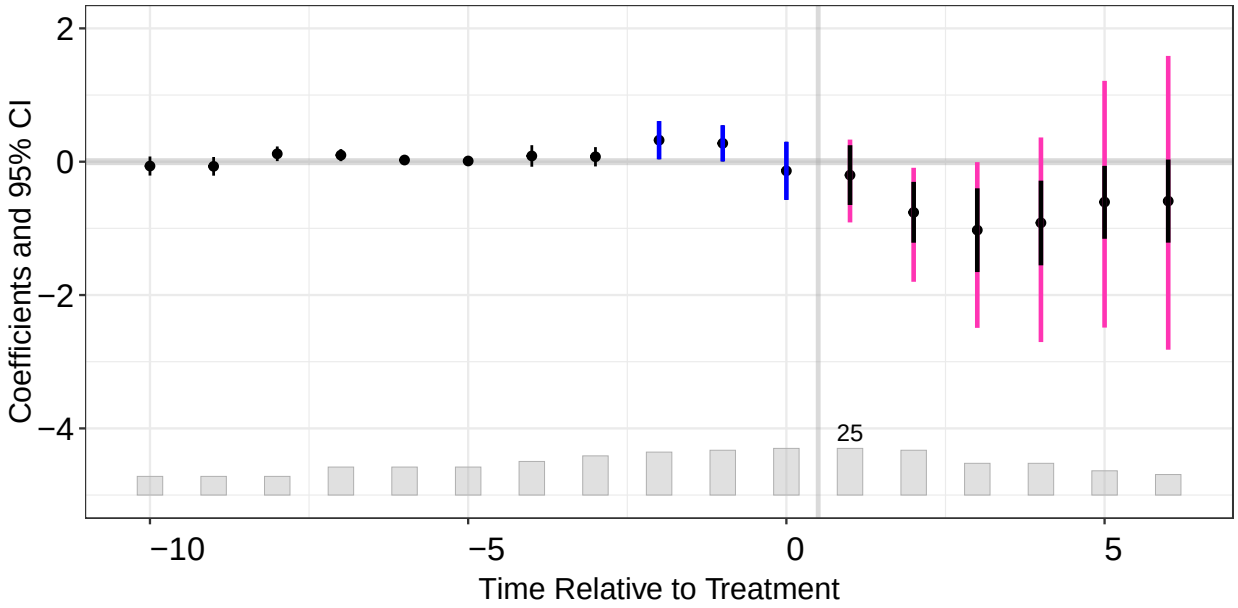
We find little evidence of violations of the parallel trend (PT) assumption or the no-carryover-effect assumption. The equivalence test also rejects the null hypothesis that the residuals in the pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.25$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 0.25 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust when using HTE-robust estimators. We find little evidence of violations of the PT assumption. However, the originally statistically significant result seems sensitive to mild violations of the PT assumption benchmarked against the estimated pretrend with the breakdown threshold $M = 0.25$.

Hainmueller and Hangartner (2019)

Contents

A First Look at Data 1

Point Estimates 2

Dynamic Treatment Effects 4

Diagnostic Tests 7

Sensitivity Analysis 8

Summary 9

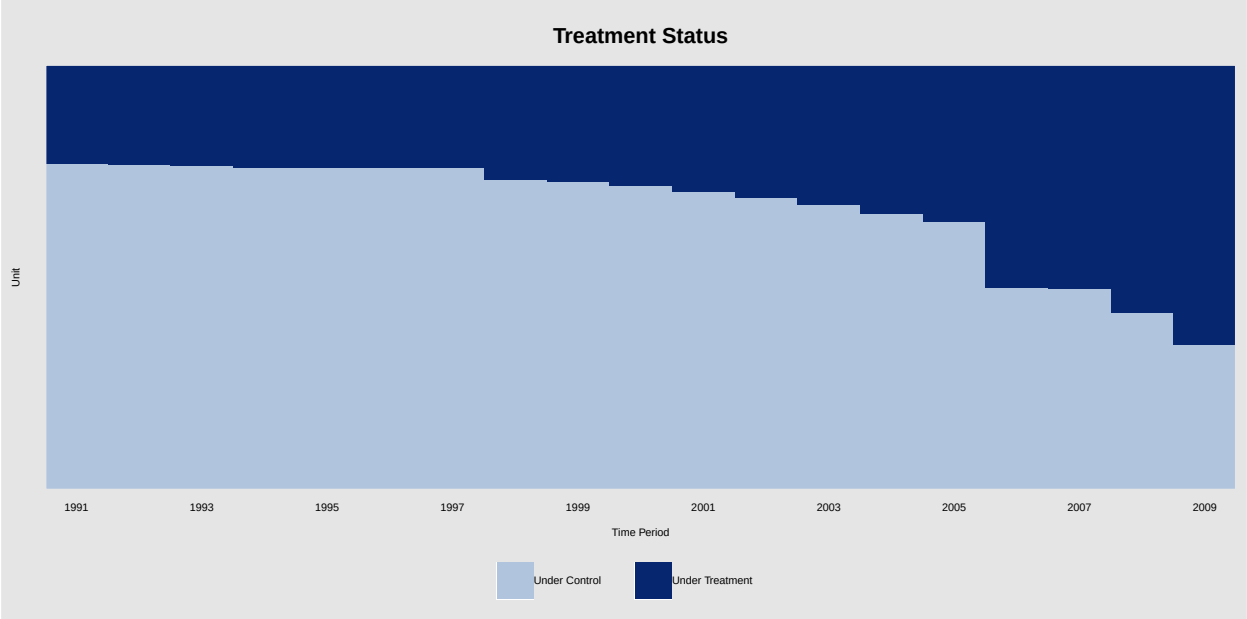
A First Look at Data

The paper investigates the effects of switching from direct to representative democracy on naturalization rates, using municipality-year level panel data from Switzerland between 1991 and 2009. One of the main findings of this paper is that "switching from direct to representative democracy increased naturalization rates by 1.22 percentage points on average" (p535).

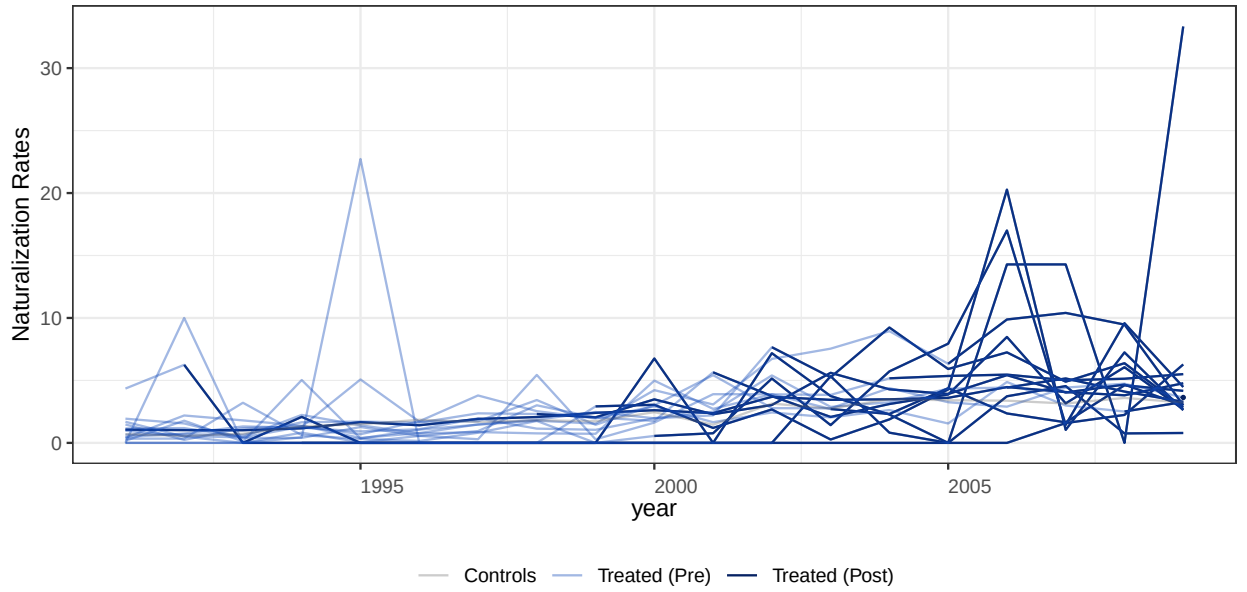
Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	Municipality × year
Treatment	Indirect democracy
Outcome	Naturalization rate
Treatment type	Staggered
Outcome type	Continous
Fixed effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a municipality). We see that two thirds of units are treated eventually, and there are no treatment reversals.



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

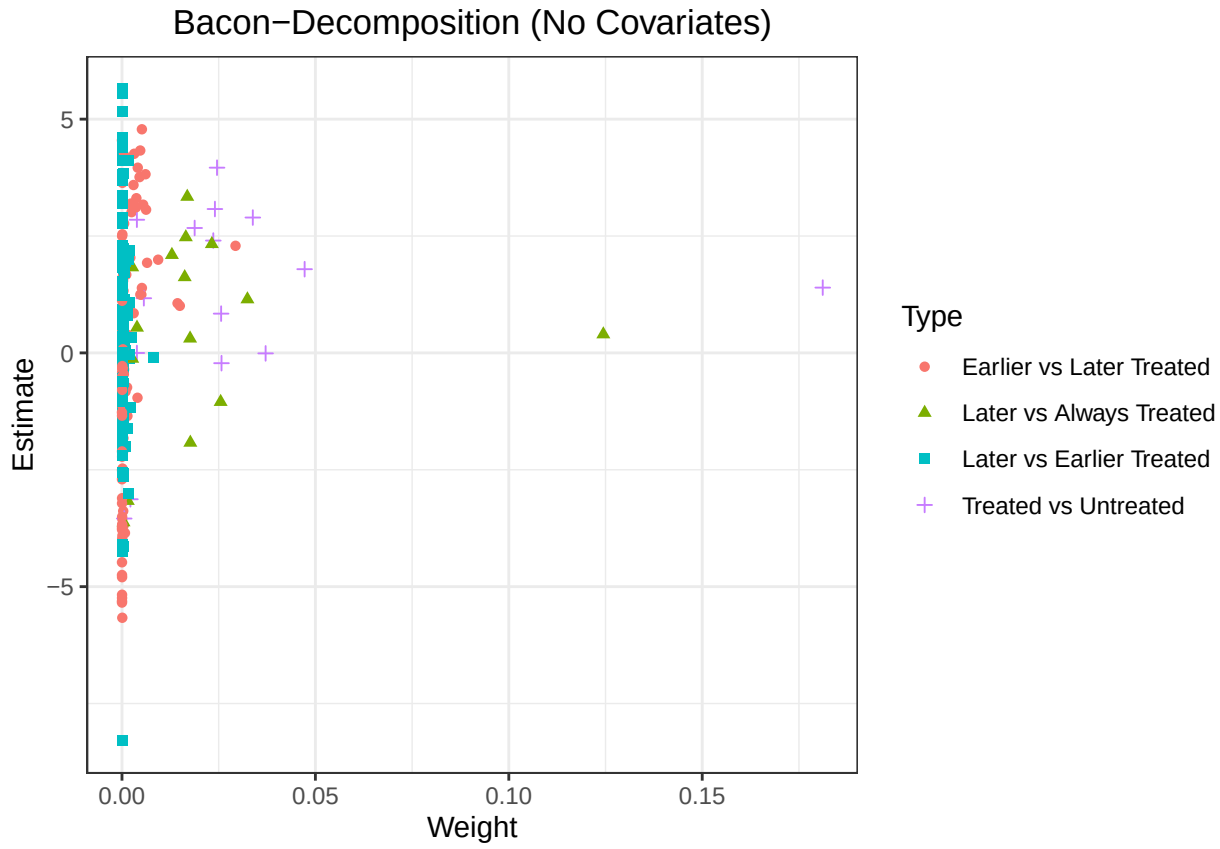
We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, and the imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

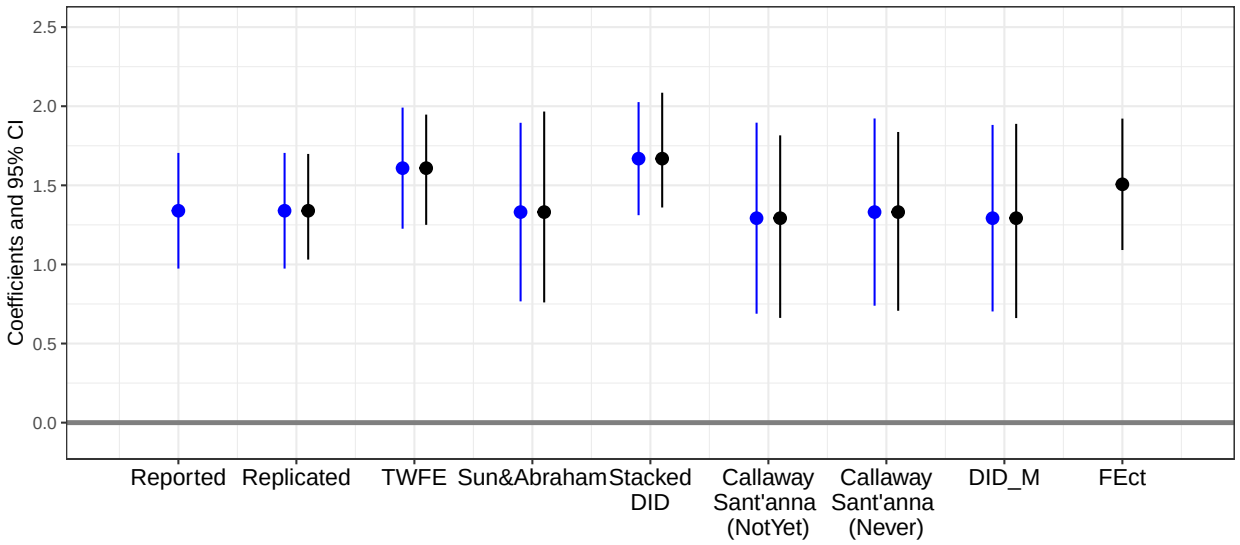
```
sol <- feols(nat_rate_ord~indirect|bfs+year,data = df,cluster = "bfs")
summary(sol)
```

```
## OLS estimation, Dep. Var.: nat_rate_ord
## Observations: 22,971
## Fixed-effects: bfs: 1,209,  year: 19
## Standard-errors: Clustered (bfs)
##          Estimate Std. Error t value   Pr(>|t|)
## indirect  1.33932   0.186525 7.18039 1.2117e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 4.09541     Adj. R2: 0.152719
##                 Within R2: 0.005173
```

Goodman-Bacon Decomposition In the Goodman-Bacon decomposition, green triangles represent the estimates of 2-by-2 DIDs that use the always-treated units as "controls," which have relatively large impact on the original estimated ATT.



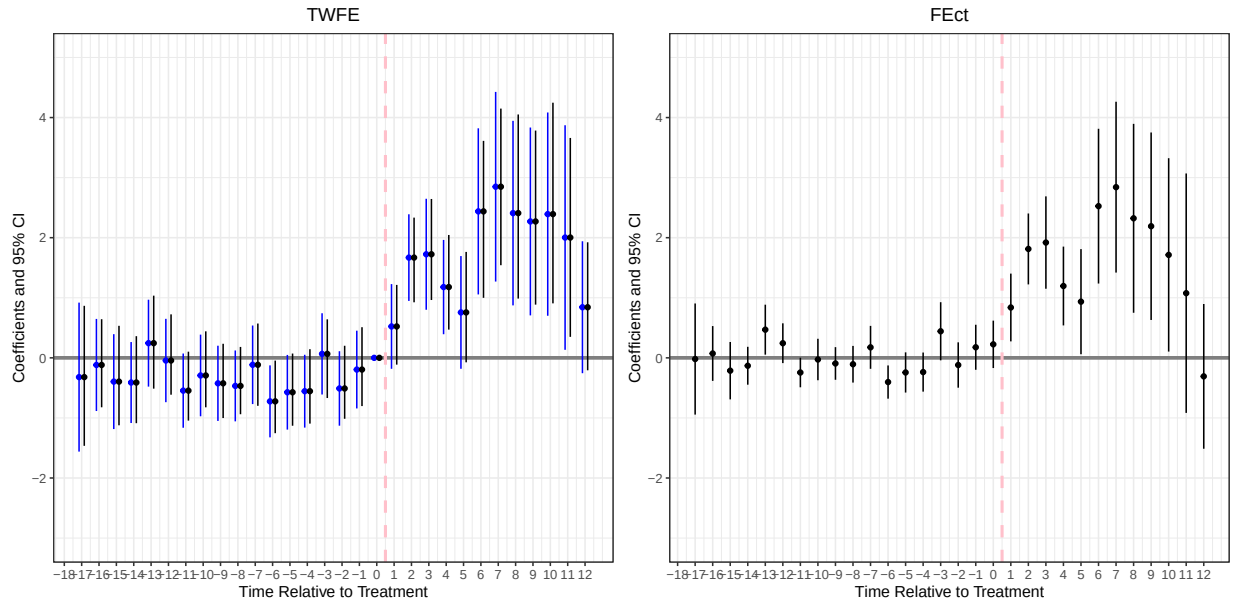

```
##   |                                                                      |
```

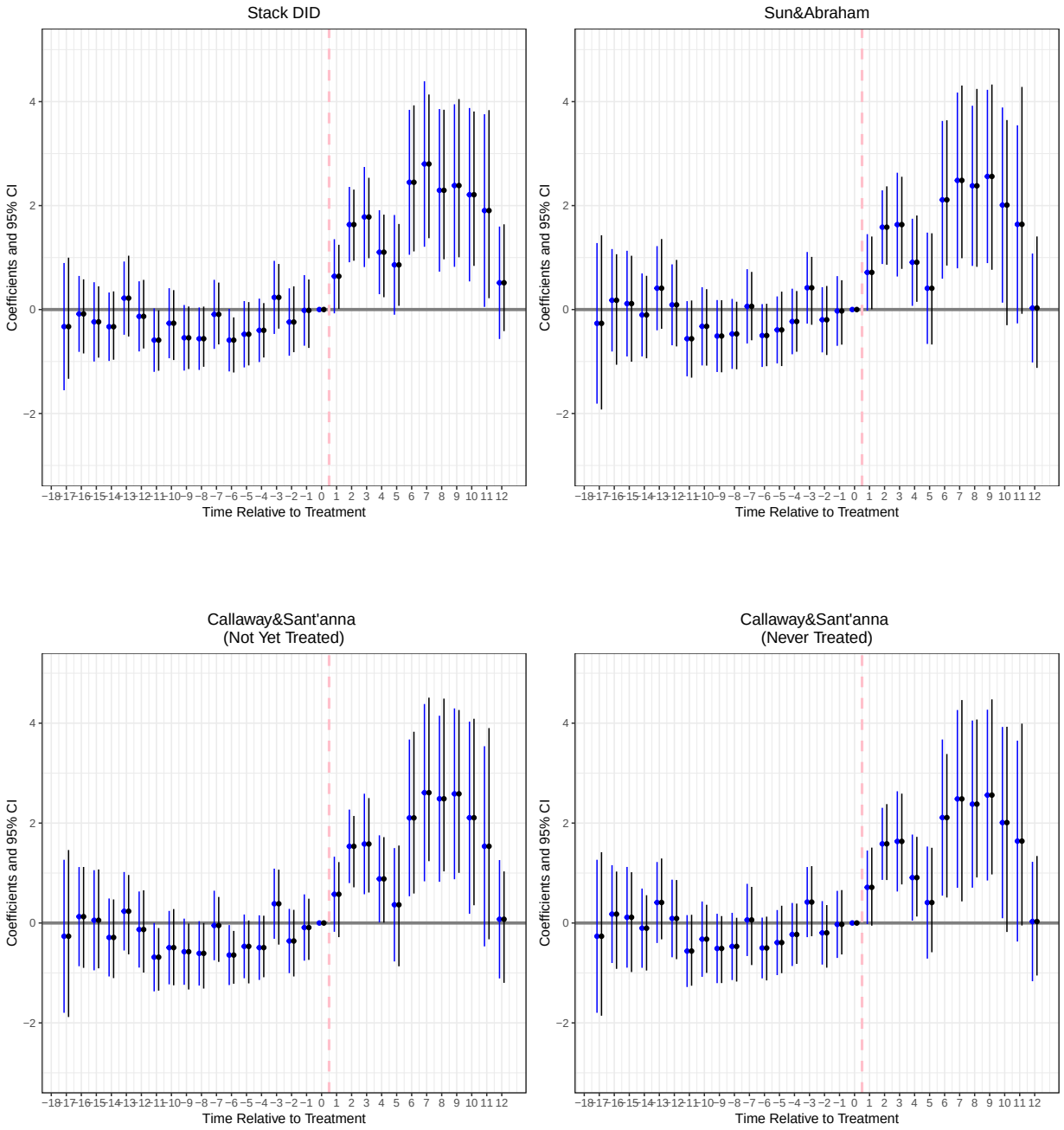


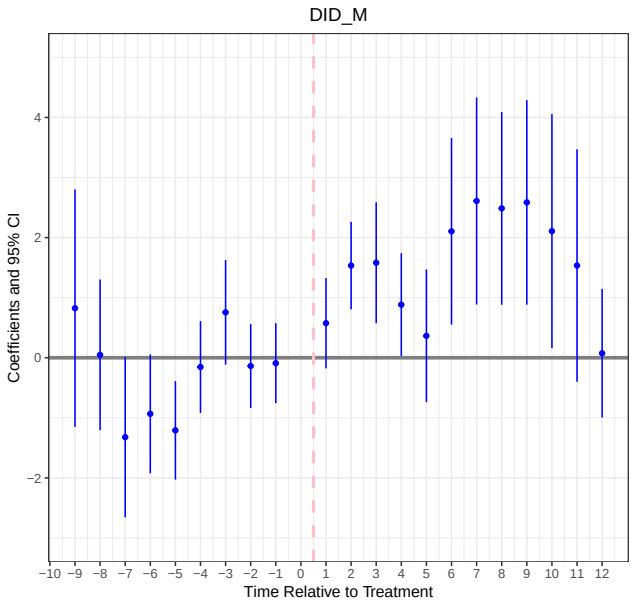
After dropping the always-treated units, the TWFE estimate becomes larger. All HTE-robut estimators and TWFE give broadly consistent results.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DID_M and FEct. The results are shown below.



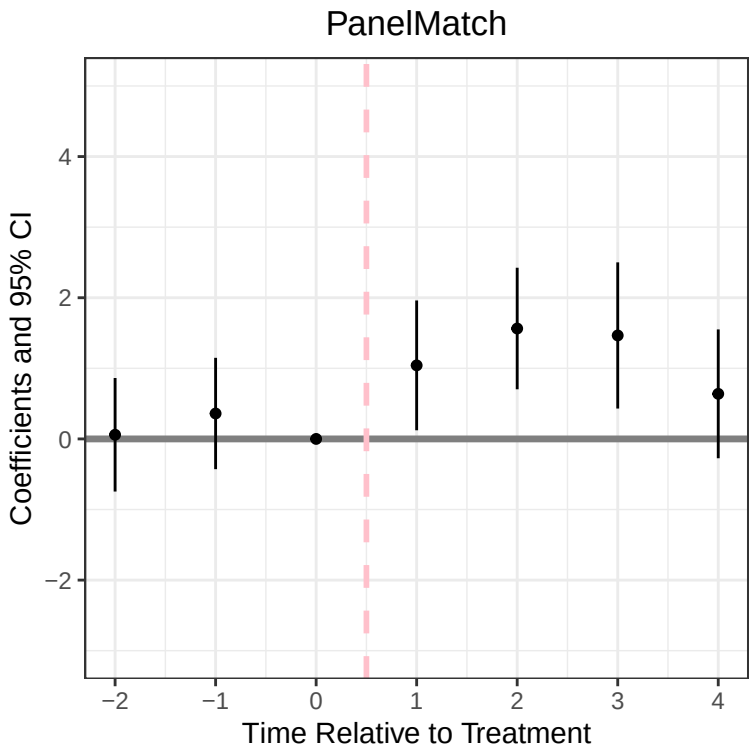

Stack DID
Sun&Abraham
Callaway&Sant'anna
(Not Yet Treated)
Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
Time Relative to Treatment



All HTE-robust estimators and TWFE yield similar estimated DTE. The estimated DTE are mostly positive and show two humps on post-treatment periods. On pre-treatment periods, there exists some signs of pre-trends.

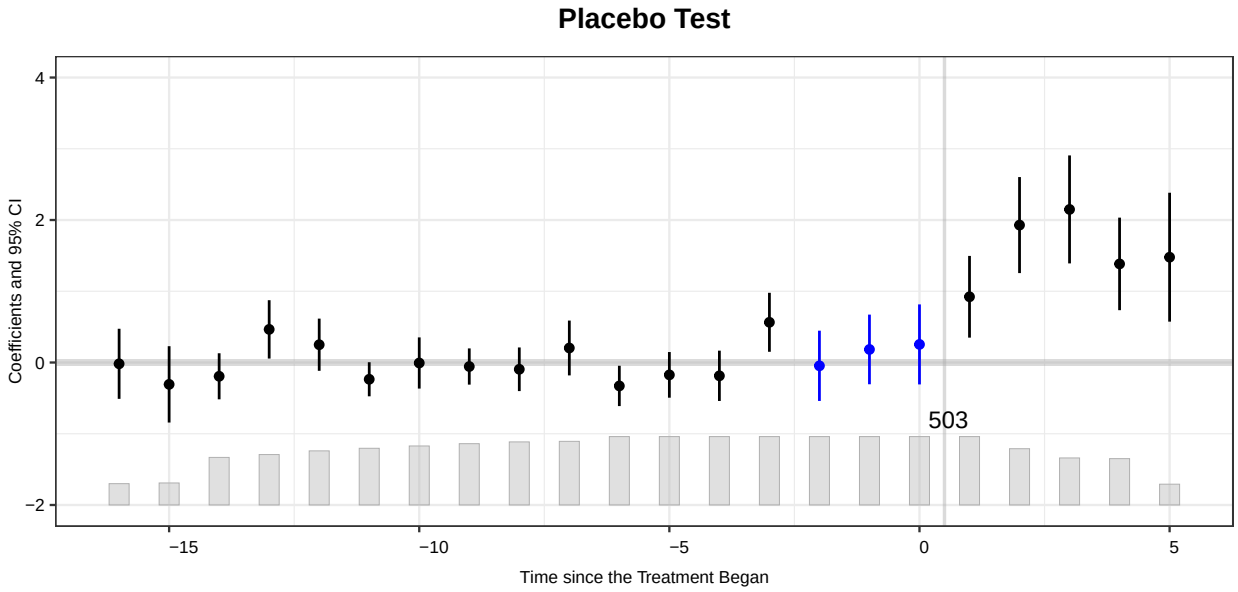
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 4 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.



Test Statistics

```
##                                 p-value
```

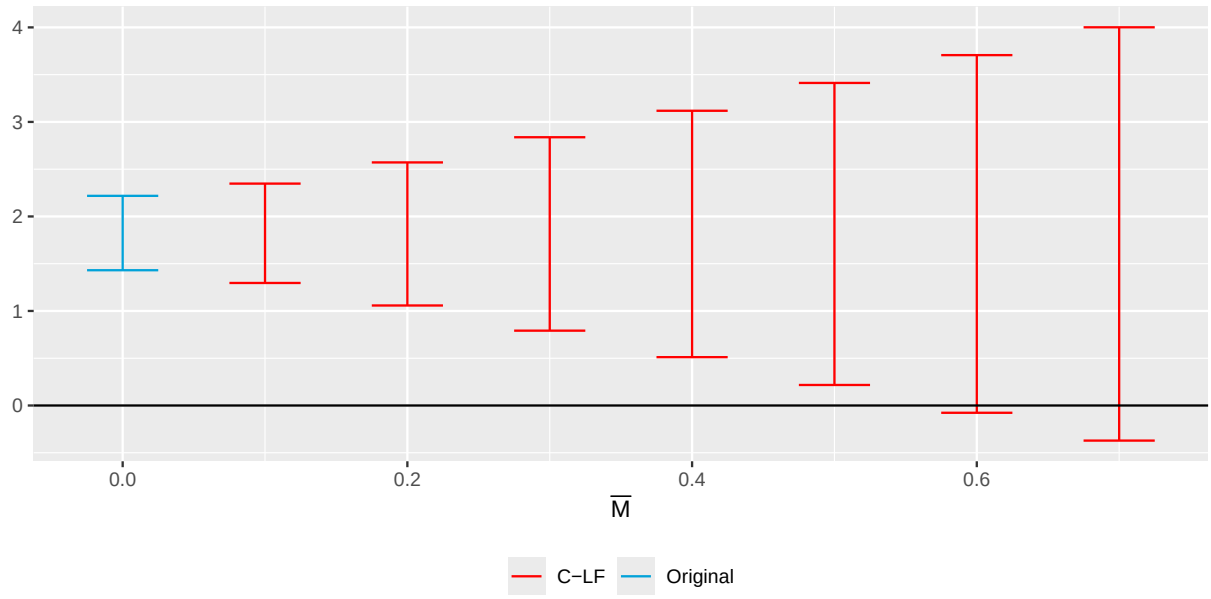
```
## F test                              0.076
## Equivalence test (default)          0.000
## Equivalence test (threshold=ATT)    0.000
## Placebo test                        0.444
## Carryover effect test                  NA
```

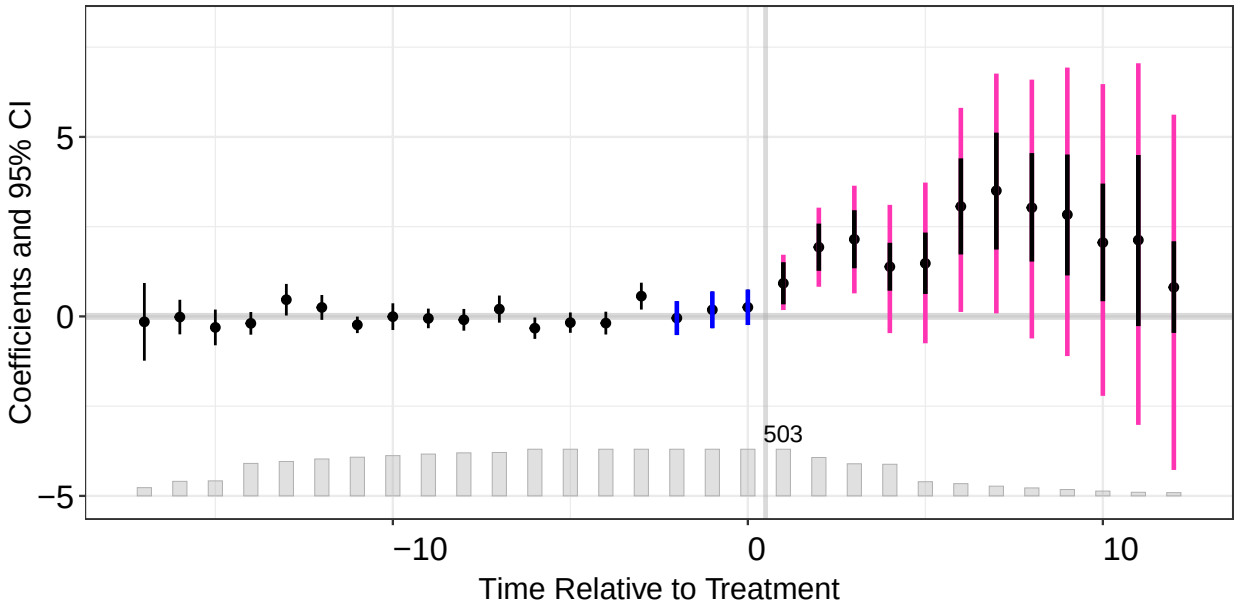
The rejection in the F-test casts some doubt on the parallel trend (PT) assumption but the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT, suggesting potential confounding likely has very limited impact on the causal estimates.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.55$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 0.55 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model seems to be robust to HTE-robust estimators. There is some evidence (F test) for potential violations of the PT but seems inconsequential. The estimated ATT appears to be robust to mild PT violations in sensitivity analysis with the breakdown threshold $M = 0.55$.

Hall and Yoder (2022)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 6
Summary 7

A First Look at Data

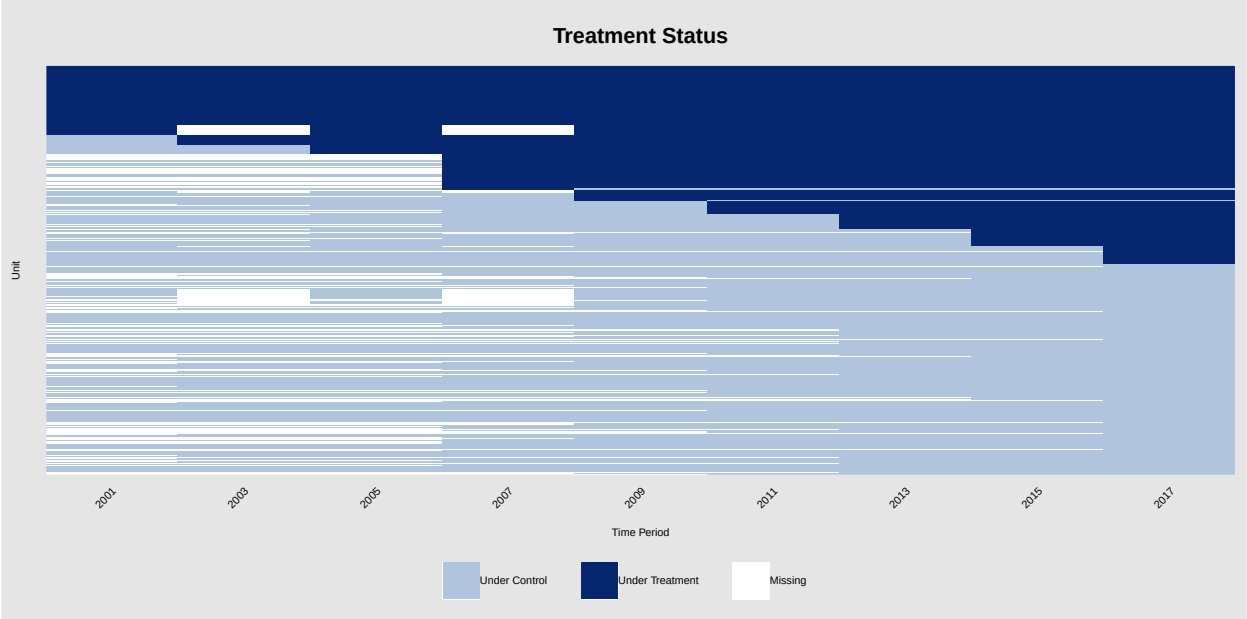
The paper investigates the effects of homeownership on voter turnout, using individual-year panel data from the state of Ohio between 2001 and 2017. One of the main findings of this paper is that "becoming a homeowner leads about a 4.9 percentage-point increase in turnout in local general elections" (p357, Table 1).

Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the unit level.

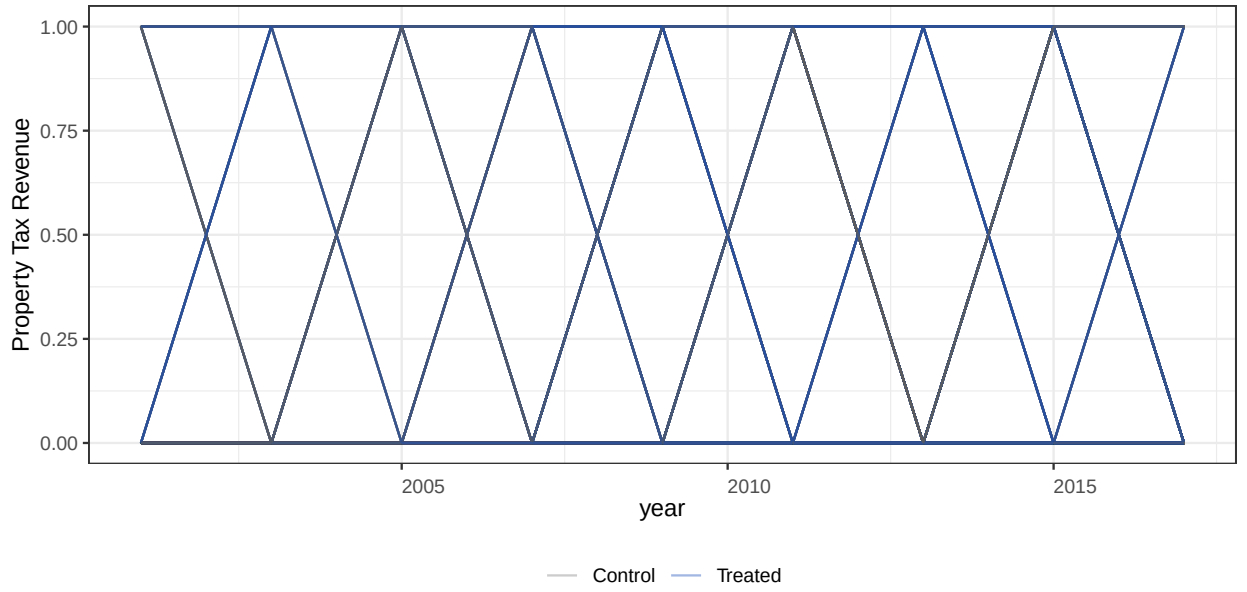
Table 1: Replication Summary

Unit of analysis	Individual × year
Treatment	Becoming a homeowner leads
Outcome	Turnout in local general elections
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (an individual). There are treatment reversals and some missingness.



View the outcome We plot the trajectory of the outcome variable for each individual. The observations under treated status are marked in blue. This plot is not very informative as the outcome is binary.



Point Estimates

We first present the regression result following the authors' original specification. As the original dataset is super large, we randomly draw a 1% sub-sample to conduct our analysis. We present the regression result using the original specification on the subsample and then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

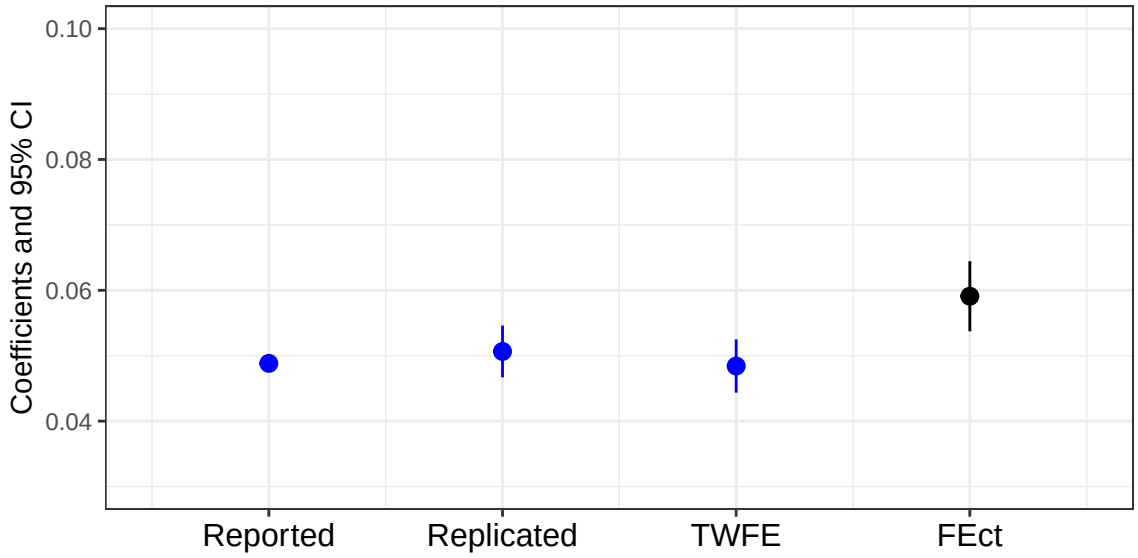
Original Results

```
#report.est.org
load("data/cleaned/hall_jop2022_org_twfe.RData")
```

```
sol <- feols(dv~homeowner|id+year,data = df,cluster = "id")
summary(sol)
```

```
## OLS estimation, Dep. Var.: dv
## Observations: 765,082
## Fixed-effects: id: 98,885,  year: 9
## Standard-errors: Clustered (id)
##           Estimate Std. Error t value  Pr(>|t|)
## homeowner 0.050652   0.002019 25.0851 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.285262     Adj. R2: 0.520782
##                  Within R2: 0.001329
```

```
##   |                                                                      |
```



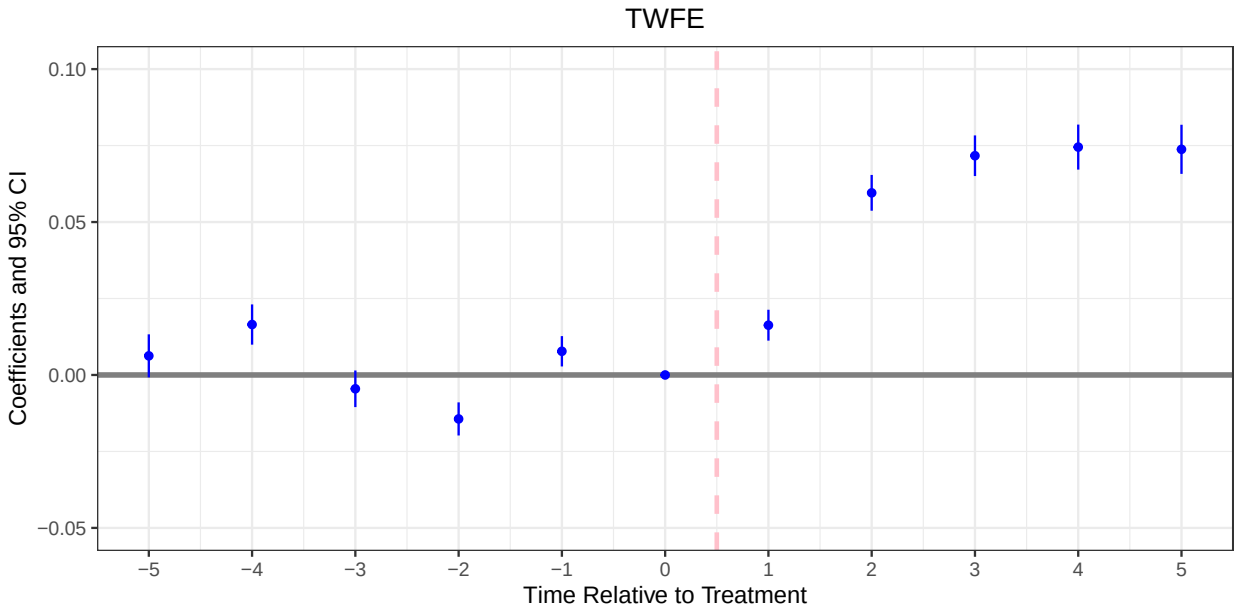
The TWFE and FEct estimator are consistent with each other. The estimated ATT are marginally significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

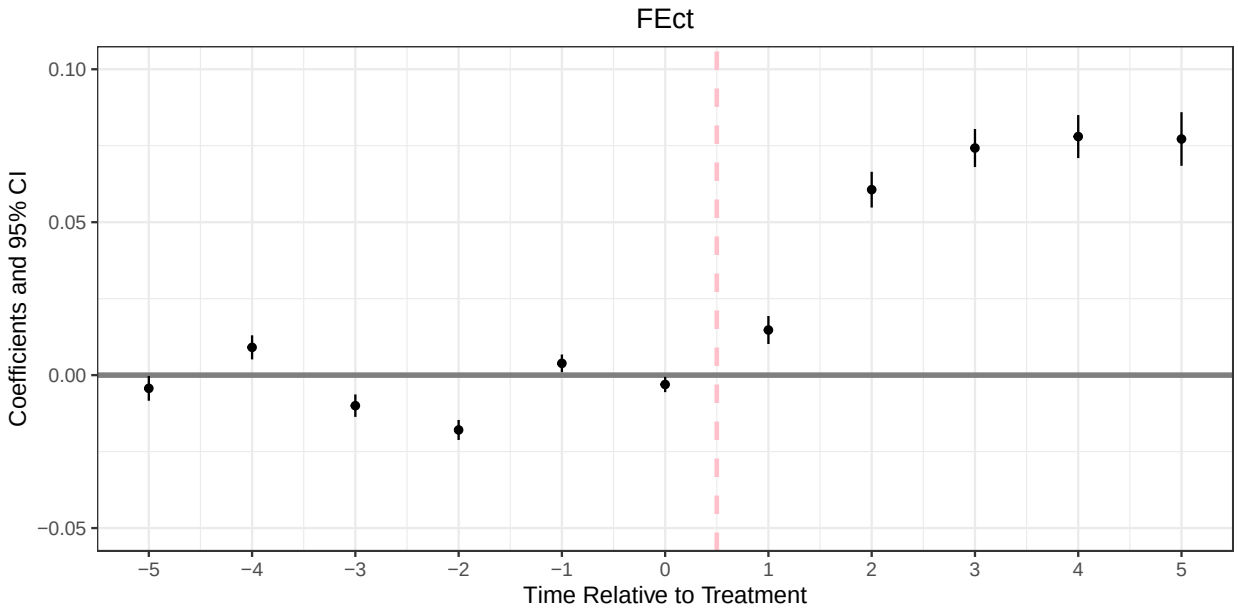
```
##   |                                                                      |
```

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.

```
##   |                                                                      |
```





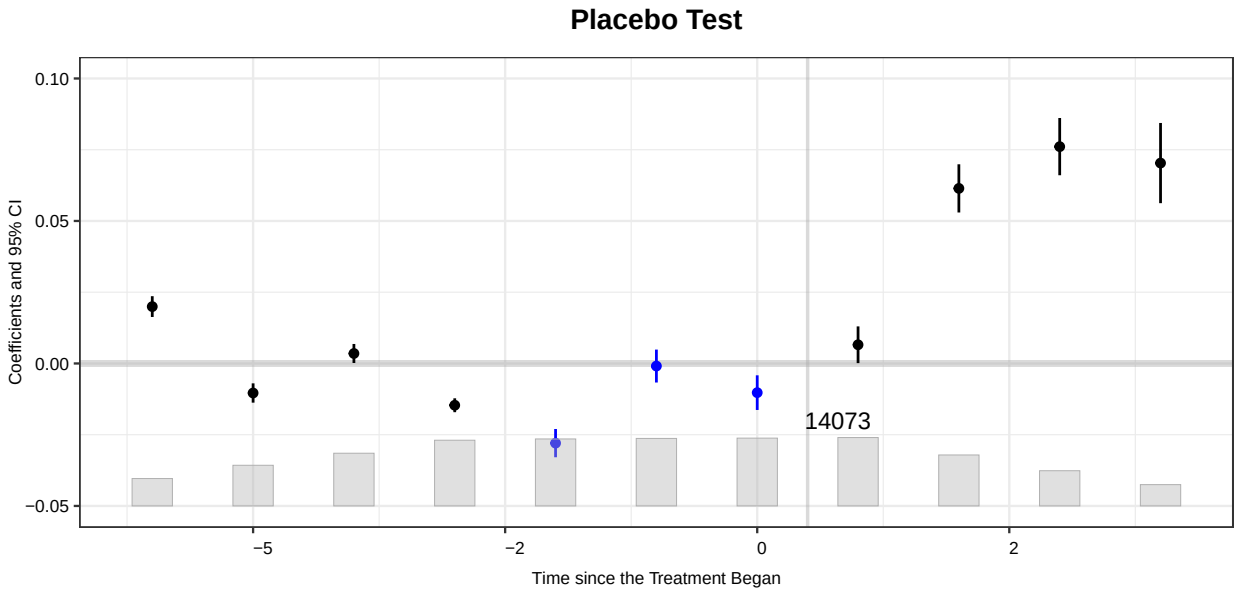
TWFE and FEct estimates are consistent with each other. The estimated DTEs are positive on all post-treatment periods. On the pre-treatment periods, the estimated DTE are different from zero but do not show a obvious clear pre-trend.

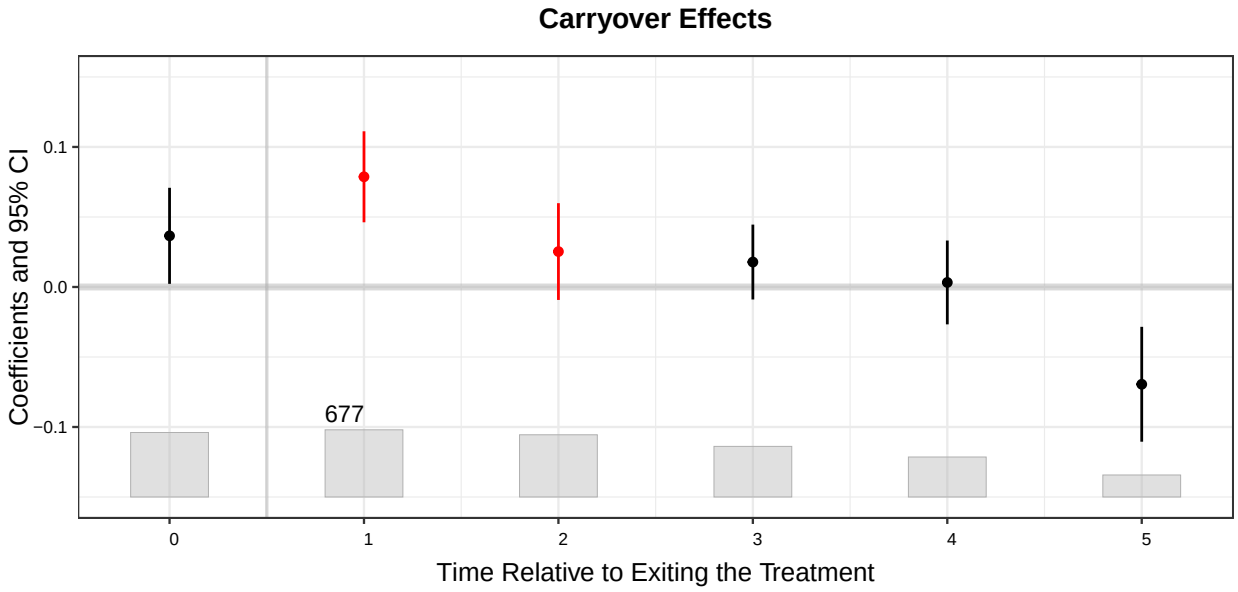
Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.

```
##   |                                                                      |
```

```
##   |                                                                      |
##   |                                                                      |
```





Test Results

	p-value
F test	0
Equivalence test (default)	0
Equivalence test (threshold=ATT)	0
Placebo test	0
Carryover effect test	0

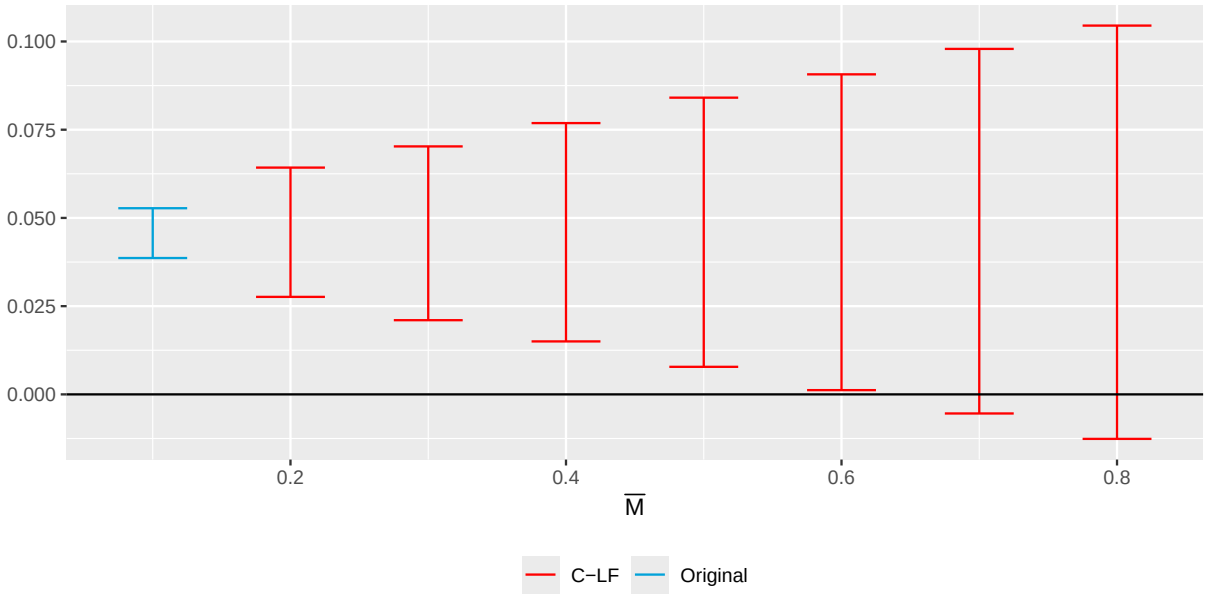
We find some evidence for potential violations to the parallel trends (PT) assumption. The DTE plot, F-test, and the placebo test, suggests at least some residual averages in the pre-treatment periods are significantly different from 0. However, the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT, suggesting that such violations likely have little impact on the causal effect estimates. Our test also shows that the no-carryover-effect assumption is unlikely to be valid.

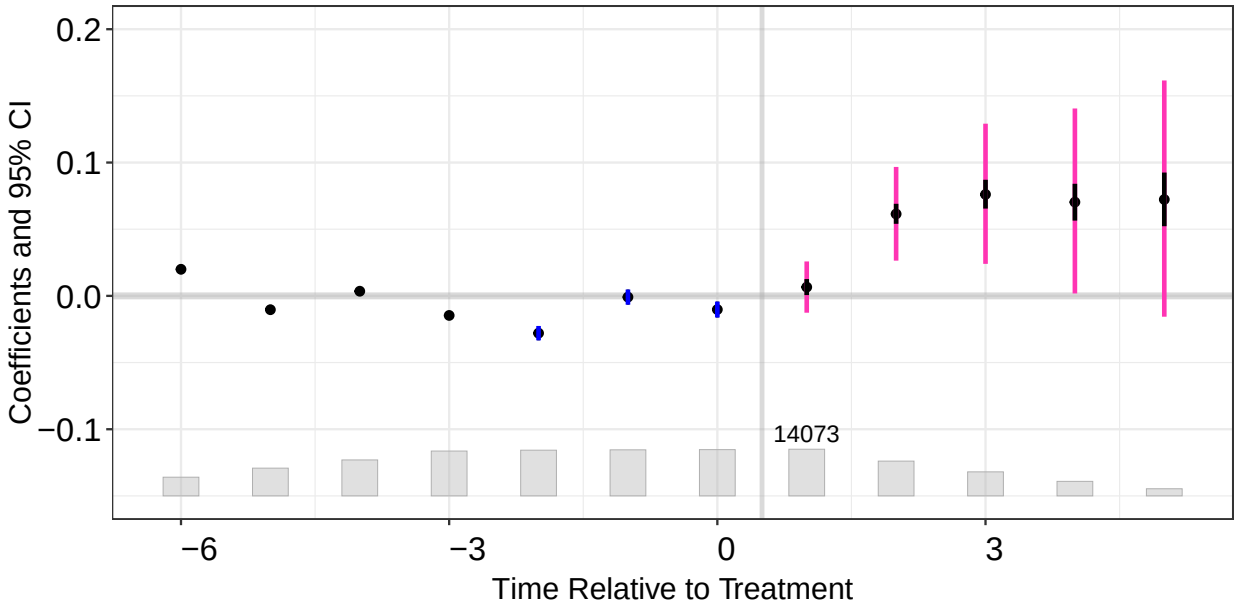
Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.6$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 0.6 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

```
##   |                                                                      |
```





Summary

Based on this 1% sub-sample, the main result of the chosen model appears to be robust to HTE-robust estimators. We find some evidence for violations of the PT, although they are unlikely to change the substantive findings of the paper. The violation of the no-carryover-effect assumption can be addressed by removing a few periods after the treatment ends. The estimated ATT appears to be robust to mild PT violations in sensitivity analysis with the breakdown threshold $M = 0.6$.

Hankinson and Magazinnik (2023)

Contents

A First Look at Data 1

Point Estimates 2

Dynamic Treatment Effects 5

Diagnostic Tests 7

Sensitivity Analysis 8

Summary 8

A First Look at Data

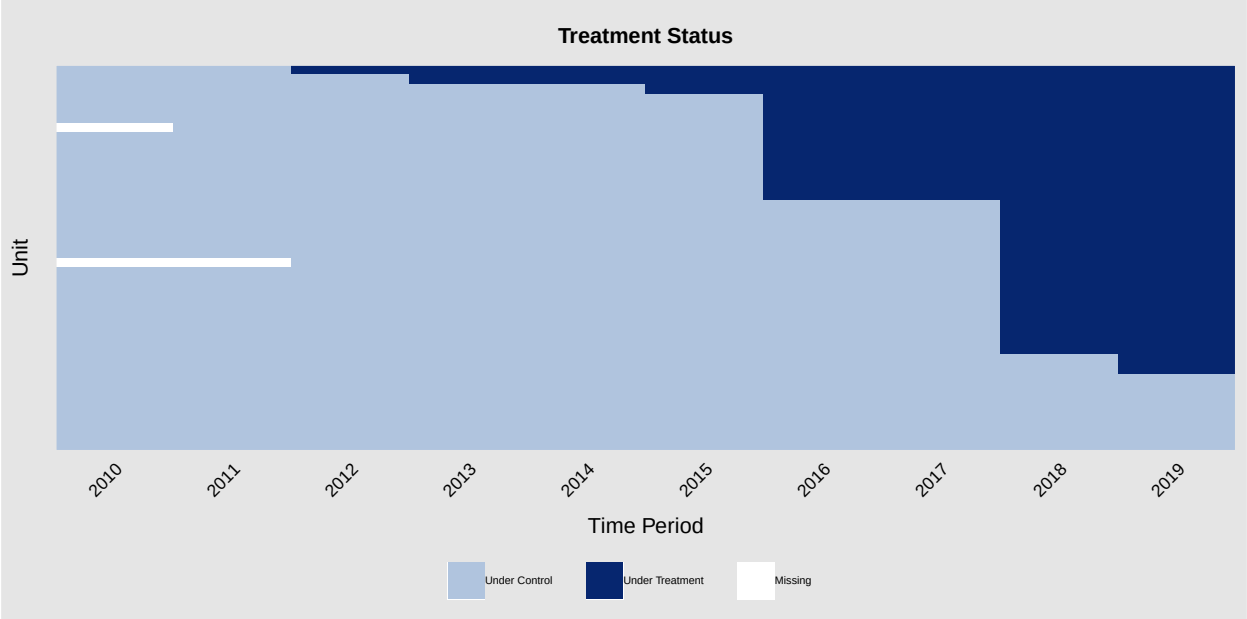
The paper investigates the effects of district elections on the permitting of multifamily housing, using Californian city-year level panel data between 2010 and 2019. One of the main findings of this paper is that "switching to districts decreases the permitting of multifamily housing units by 0.81 log points or 55%." (p1041)

Model. We focus on **Model 2 of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model with a unit-specific linear time trend and report robust standard errors clustering at the city level.

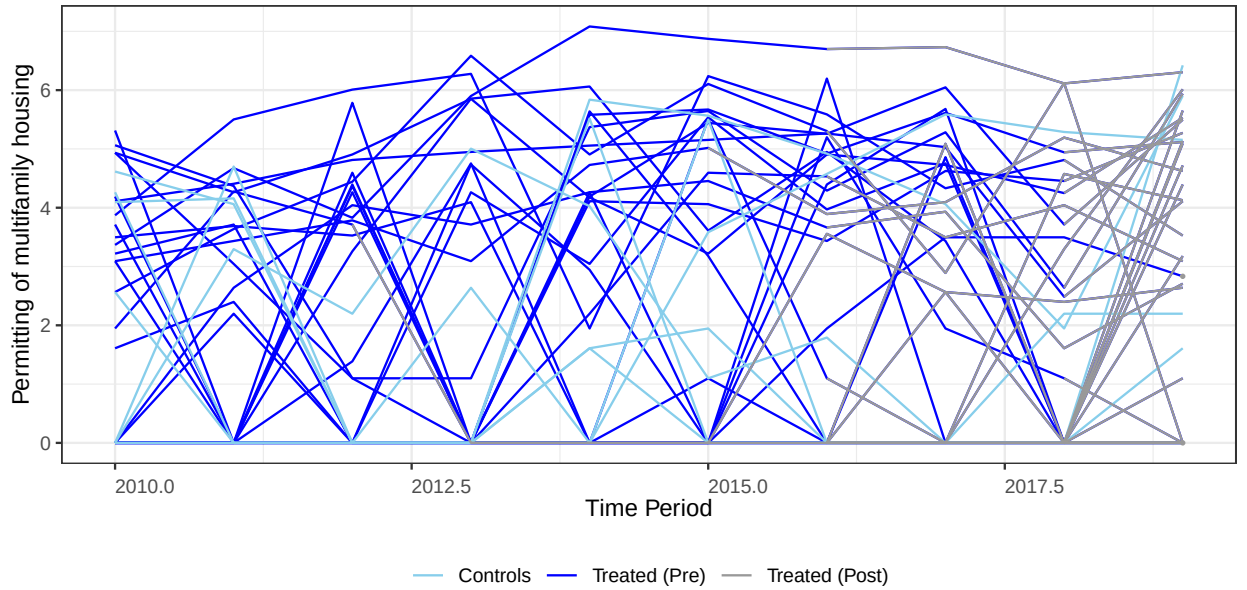
Table 1: Replication Summary

Unit of analysis	City × year
Treatment	District election
Outcome	Permitting of multifamily housing
Treatment type	Staggered
Outcome type	Continous
Fixed Effects	Unit+Time+Unit-specific linear time trend

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a city).



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

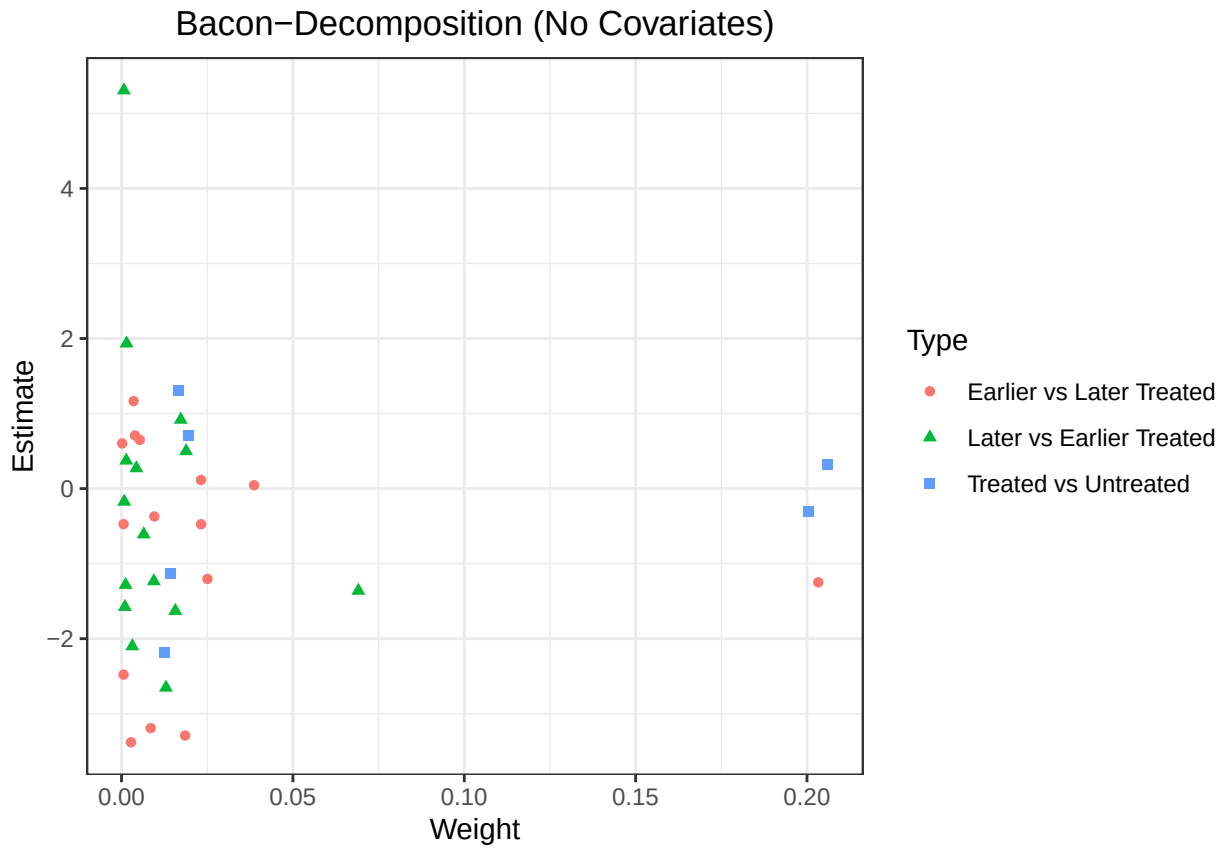
We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DIDM estimator and the imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

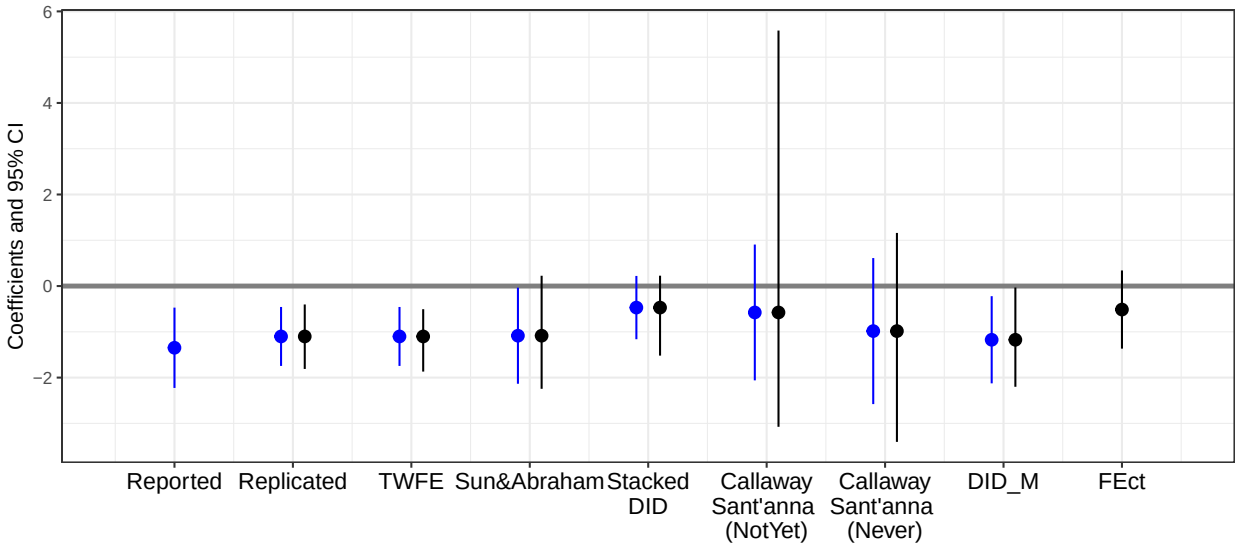
```
sol <- feols(l.multi~treat + mht+population.th+own_rate+vacancy_rate+med_value.th+
               med_income.th+prop_min_elected_past12|location+year + location[year],
             data = df,cluster = "location")
summary(sol)
```

```
## OLS estimation, Dep. Var.: l.multi
## Observations: 397
## Fixed-effects: location: 40,  year: 10
## Varying slopes: year (location): 40
## Standard-errors: Clustered (location)
##                           Estimate Std. Error   t value  Pr(>|t|)
## treat                    -1.348483   0.447591 -3.012754 0.0045301 **
## mht                       0.216117   0.537621  0.401988 0.6898878
## population.th            -0.003754   0.083098 -0.045178 0.9641962
## own_rate                  3.862714   7.672932  0.503421 0.6175001
## vacancy_rate             36.532175  21.446653  1.703397 0.0964512 .
## med_value.th             -0.010045   0.008679 -1.157322 0.2541802
## med_income.th            -0.000579   0.074666 -0.007748 0.9938573
## prop_min_elected_past12   1.908422   2.598128  0.734537 0.4670152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1.43974     Adj. R2: 0.485993
##                 Within R2: 0.044896
```

Goodman-Bacon Decomposition



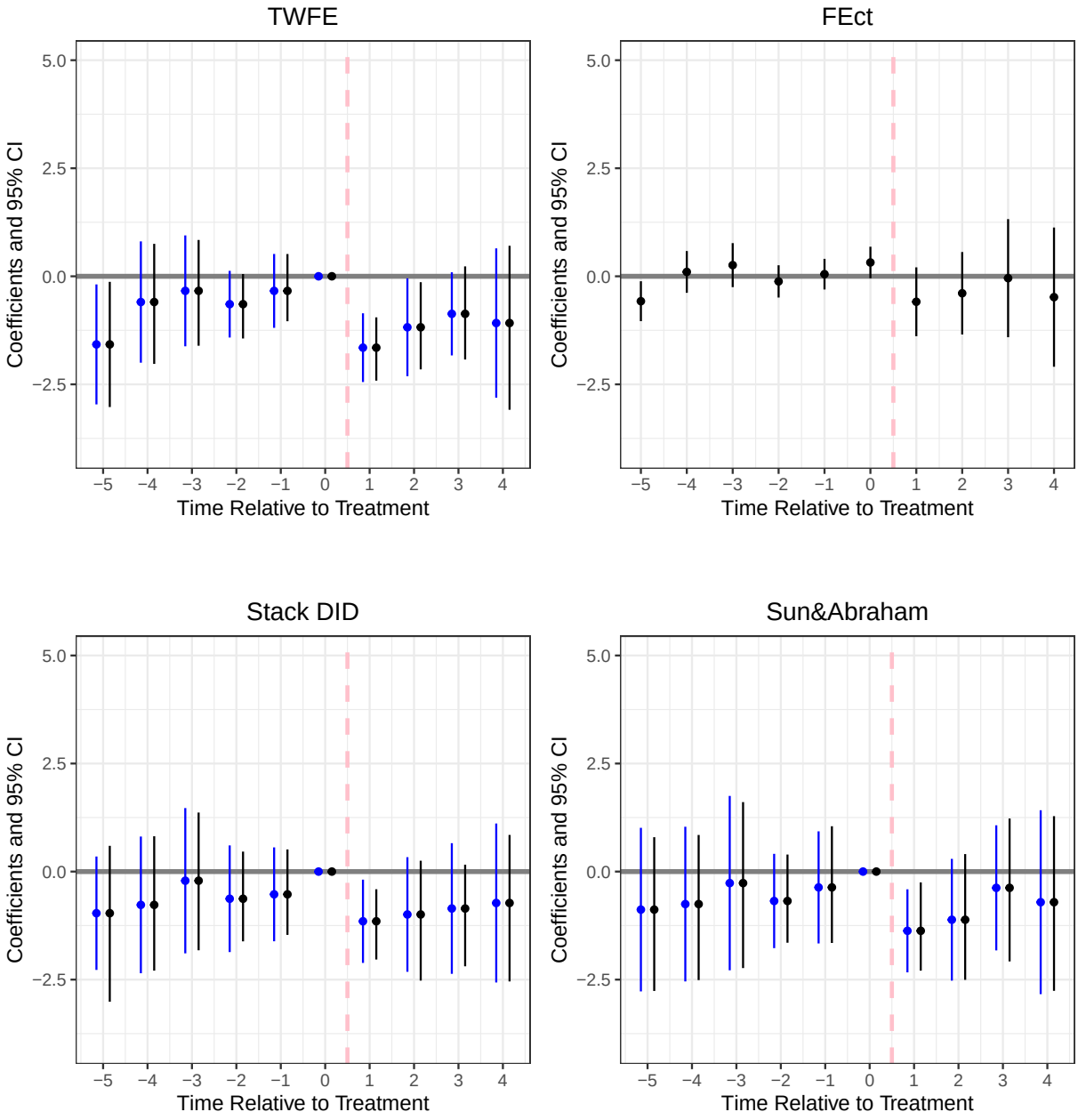

```
##   |                                                                      |
```

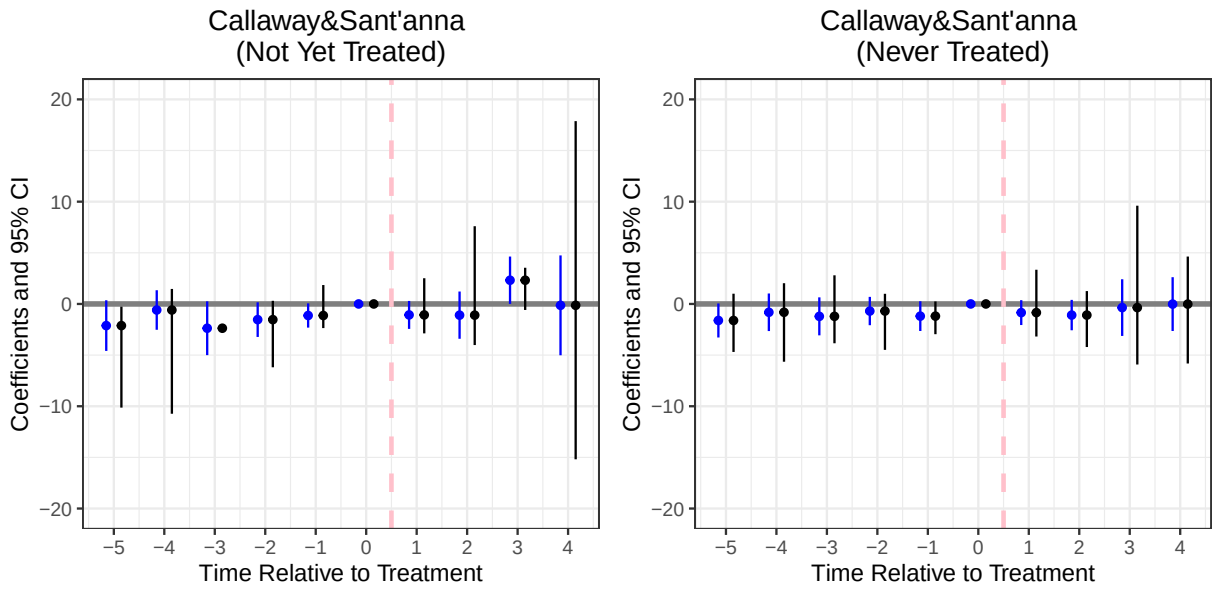


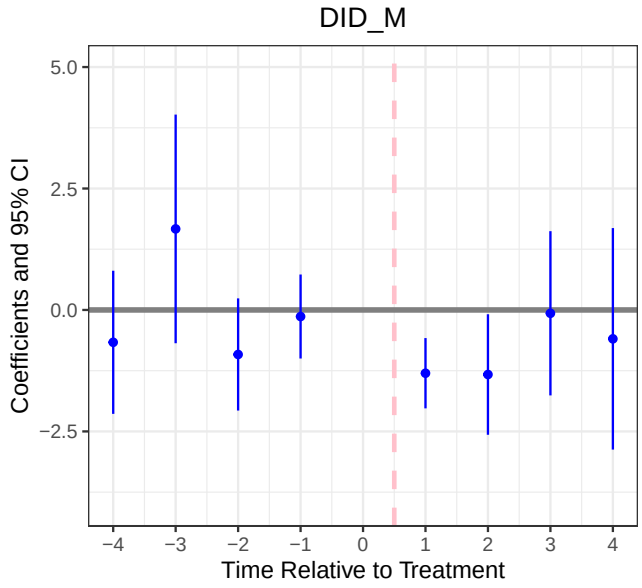
HTE-robust estimators and TWFE give different results. The point estimates are not statistically significant when using the HTE-robust estimators.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DIDM and FEct. The results are shown below.

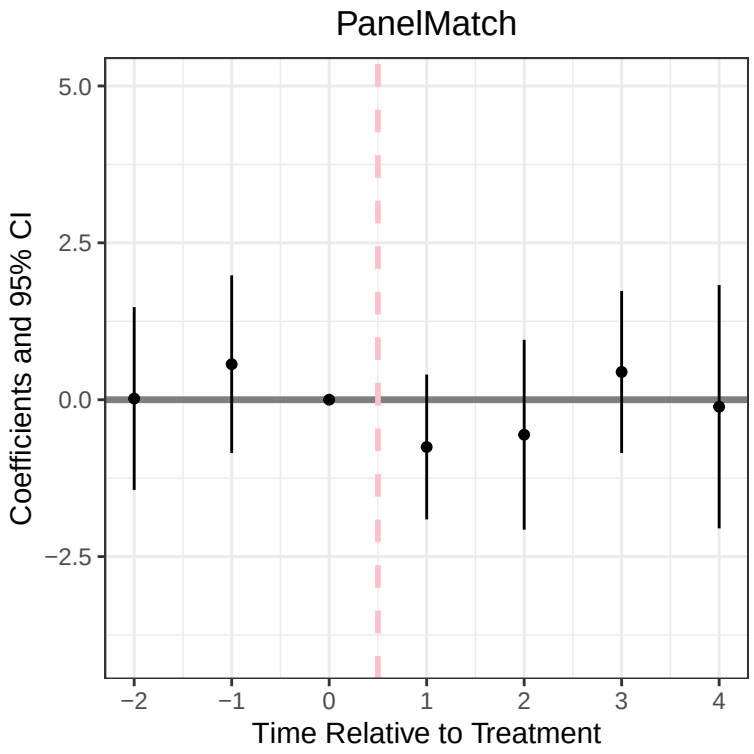






DTE of a Balanced Subsample

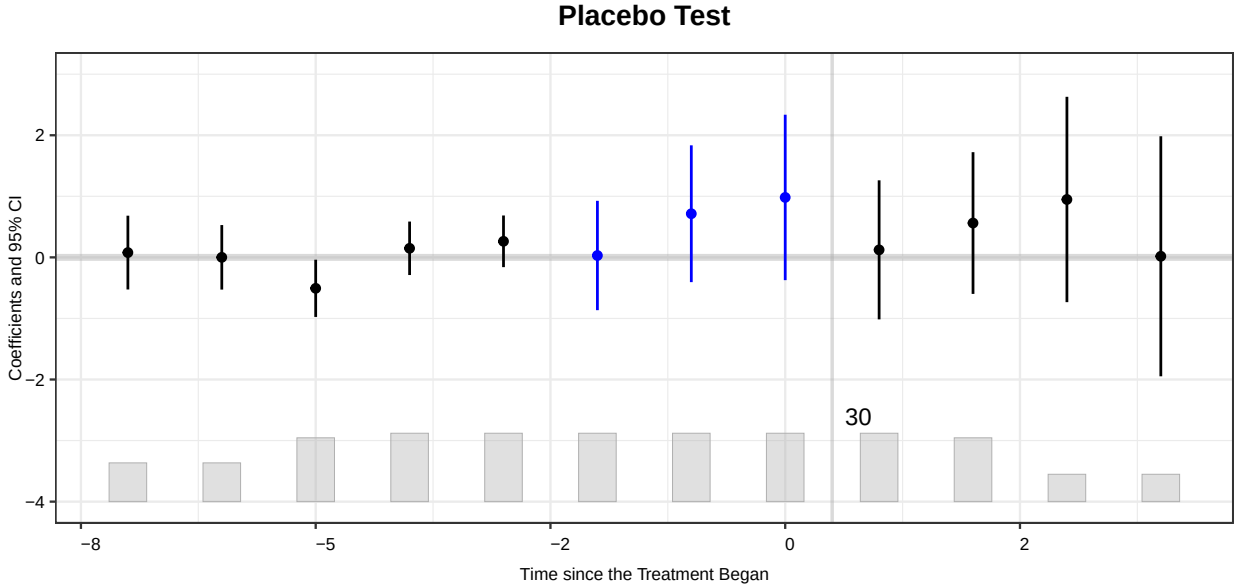
We also estimate the DTE using PanelMatch (lead = 4 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



FEct and PanelMatch estimates are broadly consistent with each other.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.



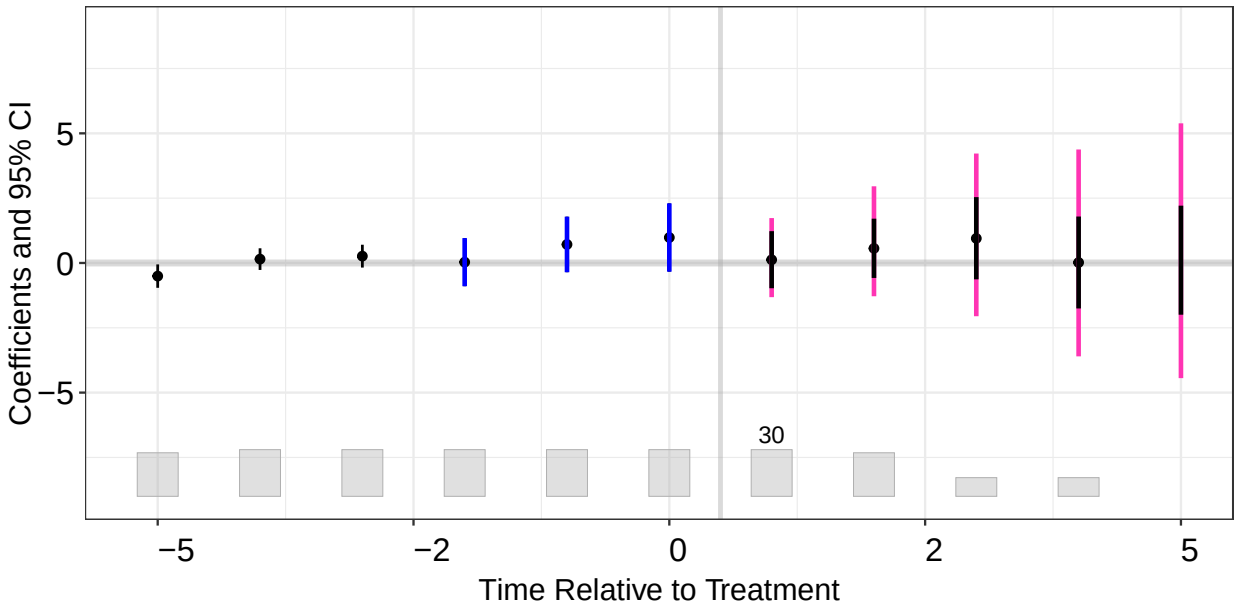
Test Statistics

```
##                                  p-value
## F test                             0.244
## Equivalence test (default)         0.385
## Equivalence test (threshold=ATT)   0.606
## Placebo test                       0.223
## Carryover effect test                 NA
```

We find little evidence for violations of the parallel trends (PT) assumption and no-carryover effects assumption.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is no longer significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, the results seem not robust to HTE and modeling choices and the study may be underpowered. We do not find strong evidence for potential violations of the PT.

Hirano et al. (2022)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 6
Sensitivity Analysis 7
Summary 8

A First Look at Data

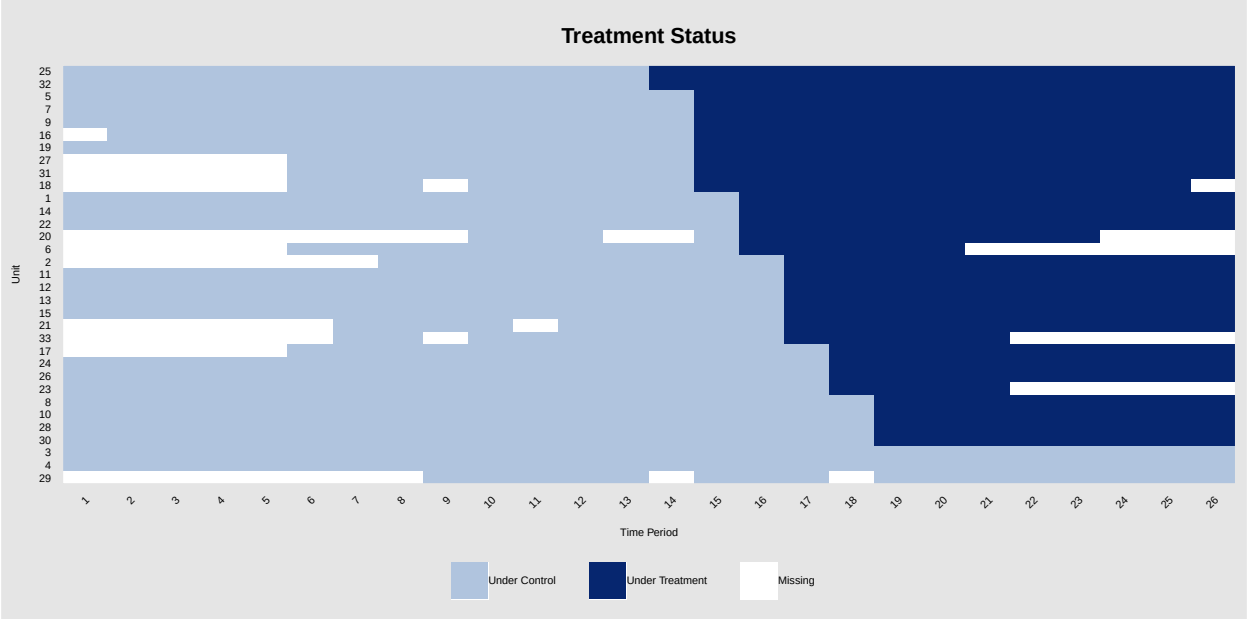
The paper investigates the effects of the direct primary on the number of campaign advertisements for candidates in general election races, using US state-year level panel data between 1880 and 1930. One of the main findings of this paper is that the "introduction of direct primaries is associated with a large and statistically significant increase in the number of candidate advertisements" (p1491).

Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the state level.

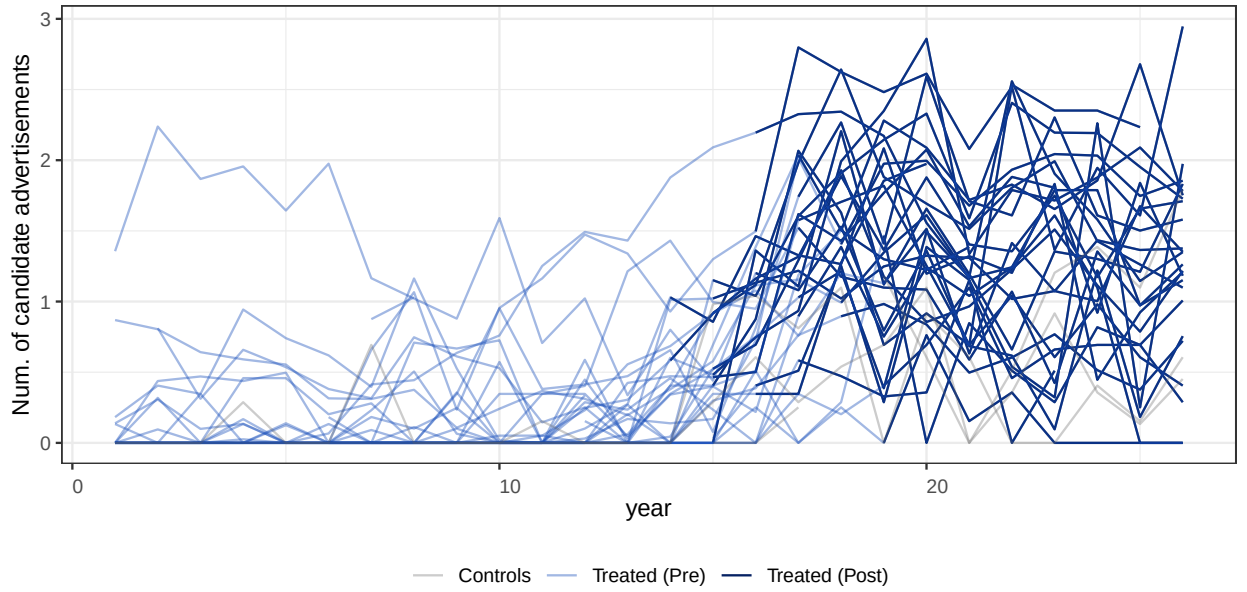
Table 1: Replication Summary

Unit of analysis	State × year
Treatment	Direct primaries
Outcome	Num. of candidate advertisements
Treatment type	Staggered
Outcome type	Continuous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a state). There are treatment reversals and some missingness.



View the outcome We plot the trajectory of the outcome variable for each state. The control units are represented in gray. We highlight the observations of treated units under untreated and treated status using light blue and deep blue, respectively.



Point Estimates

We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units (there is none in this data) and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DIDM estimator and the imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. As the number of never-treated units are not sufficient for the CS estimator, we only report the CS estimates based on not-yet-treated units.

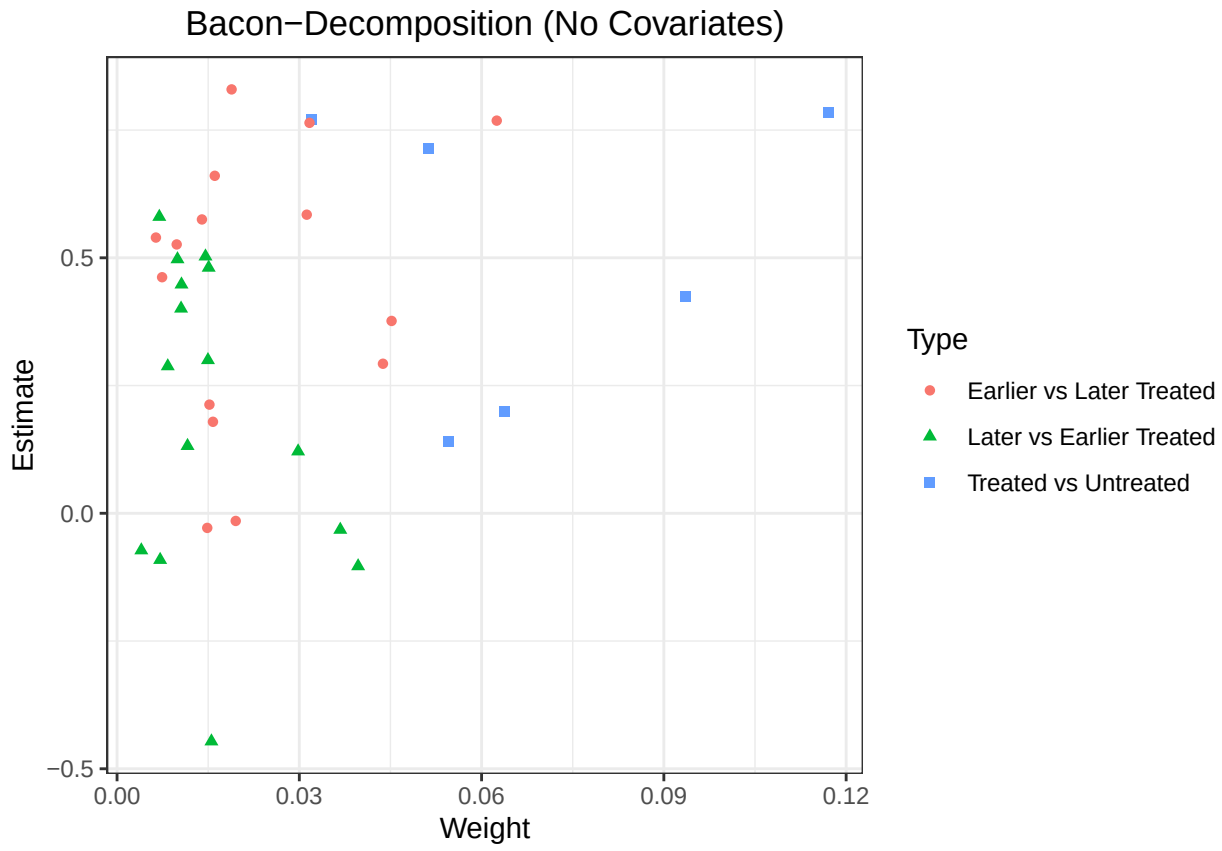
Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs (for the "reported" ATT, we report the robust SEs) and cluster-bootstrapped CIs, respectively.

Original Results

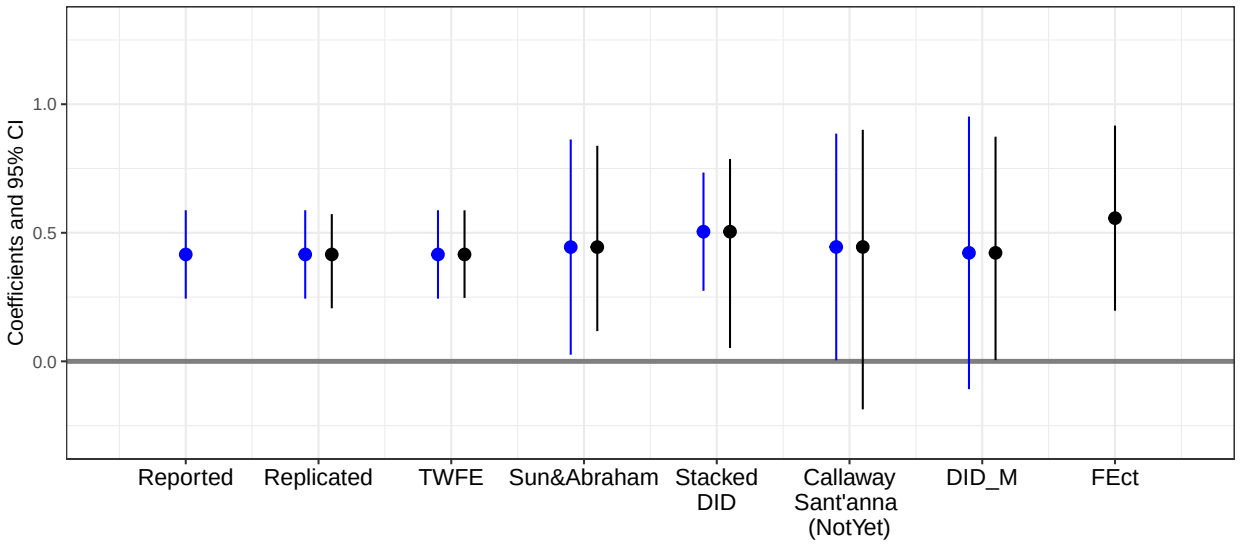
```
sol <- feols(ln_x_cand_all~prim_sw|state+year,data = df,cluster = "state")
summary(sol)
```

```
## OLS estimation, Dep. Var.: ln_x_cand_all
## Observations: 769
## Fixed-effects: state: 33,  year: 26
## Standard-errors: Clustered (state)
##         Estimate Std. Error t value   Pr(>|t|)
## prim_sw  0.41576   0.087721 4.73957 4.2298e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.41412     Adj. R2: 0.640927
##                 Within R2: 0.044162
```

Goodman-Bacon Decomposition



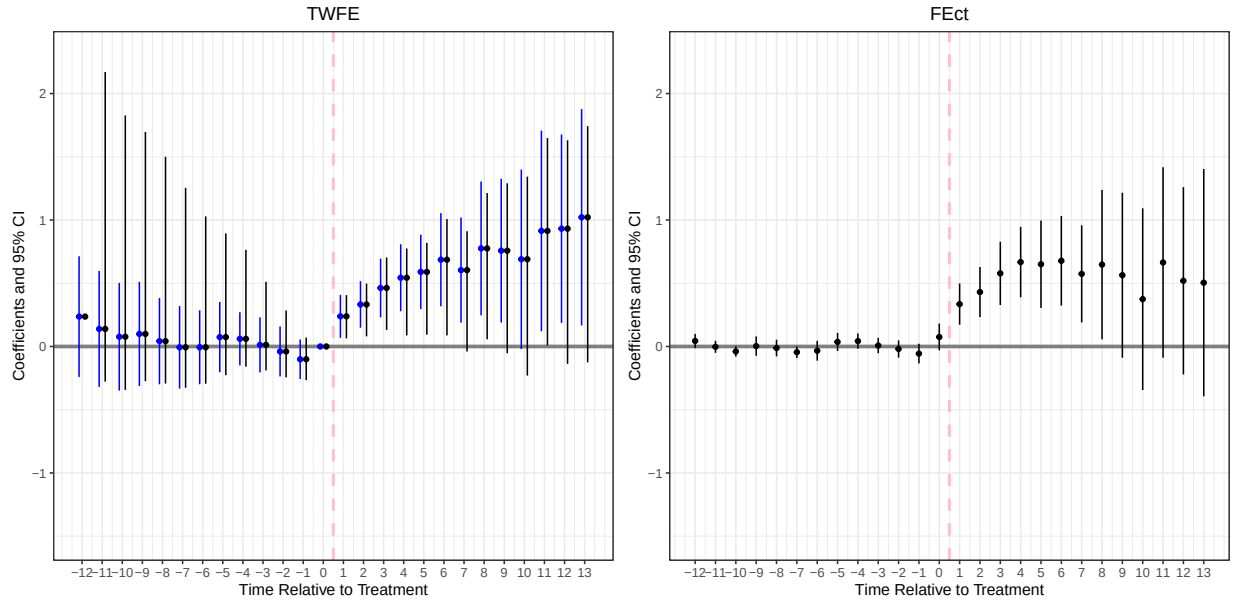

```
##   |                                                                      |
```

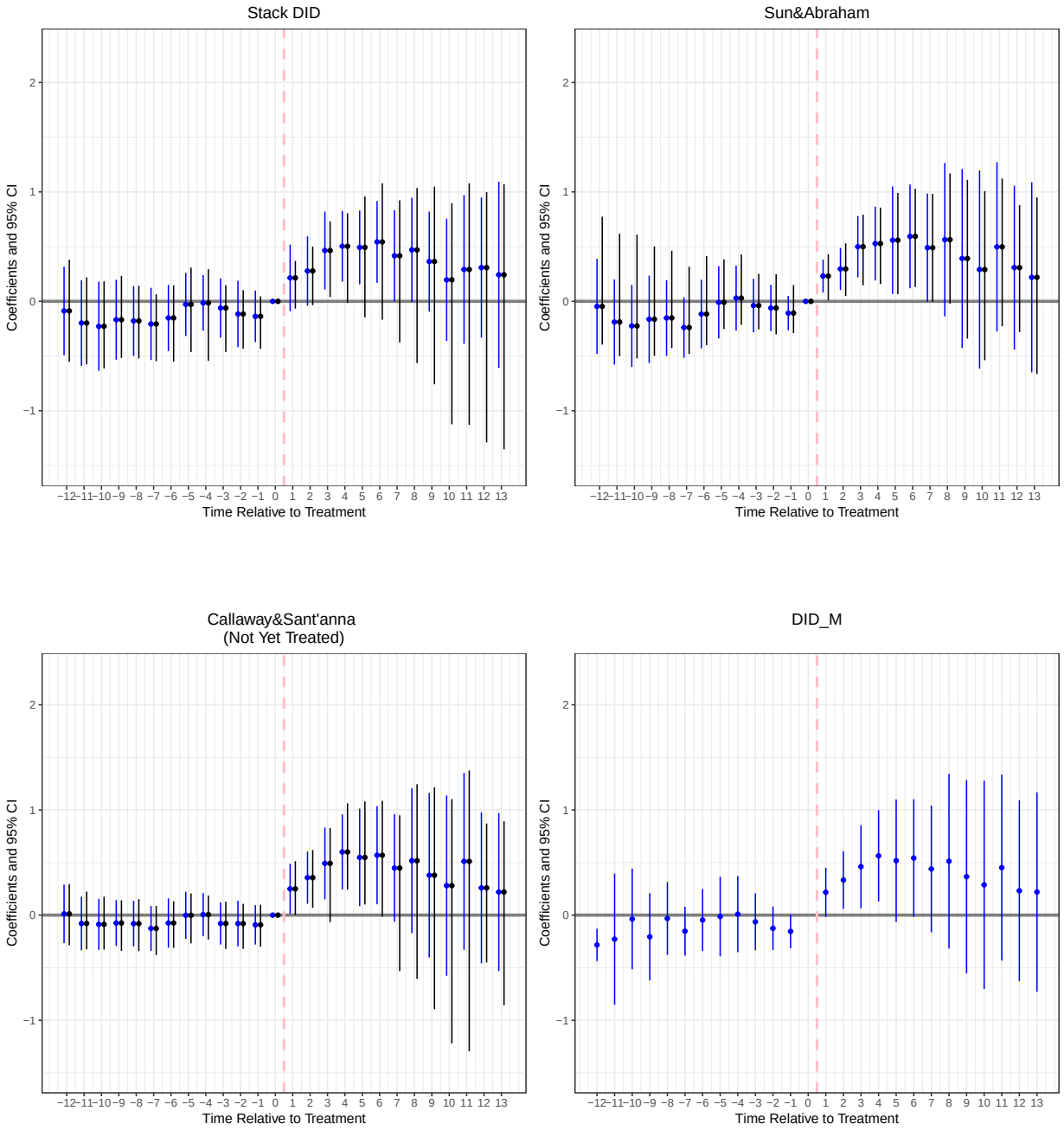


All HTE-robust estimators yield estimated ATT similar to TWFE coefficients. They are all positive and mostly statistically significant at the 5% level.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DIDM, and FEct.

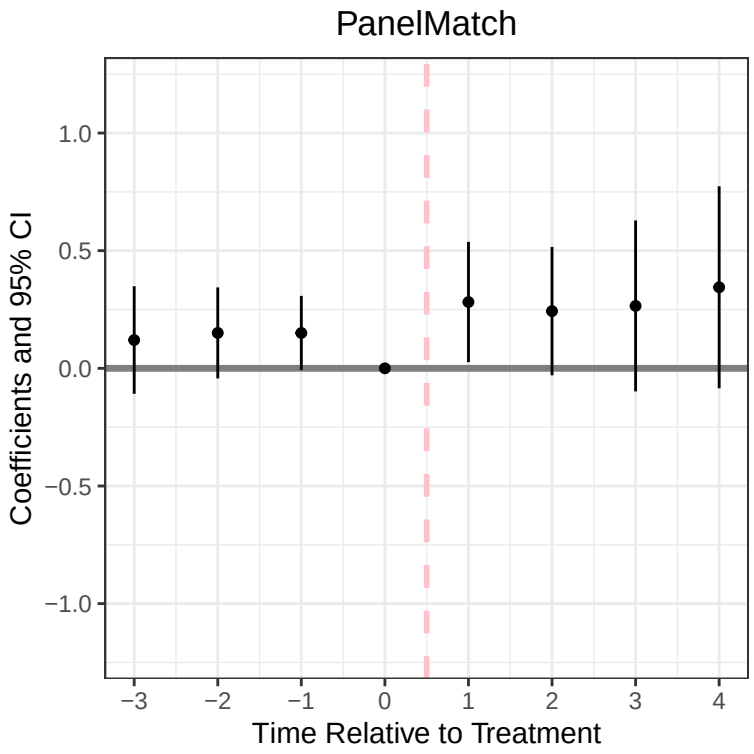




The estimated DTEs given by TWFE and HTE-robust estimators have similar shapes. There is no obvious pre-trend.

DTE of a Balanced Subsample

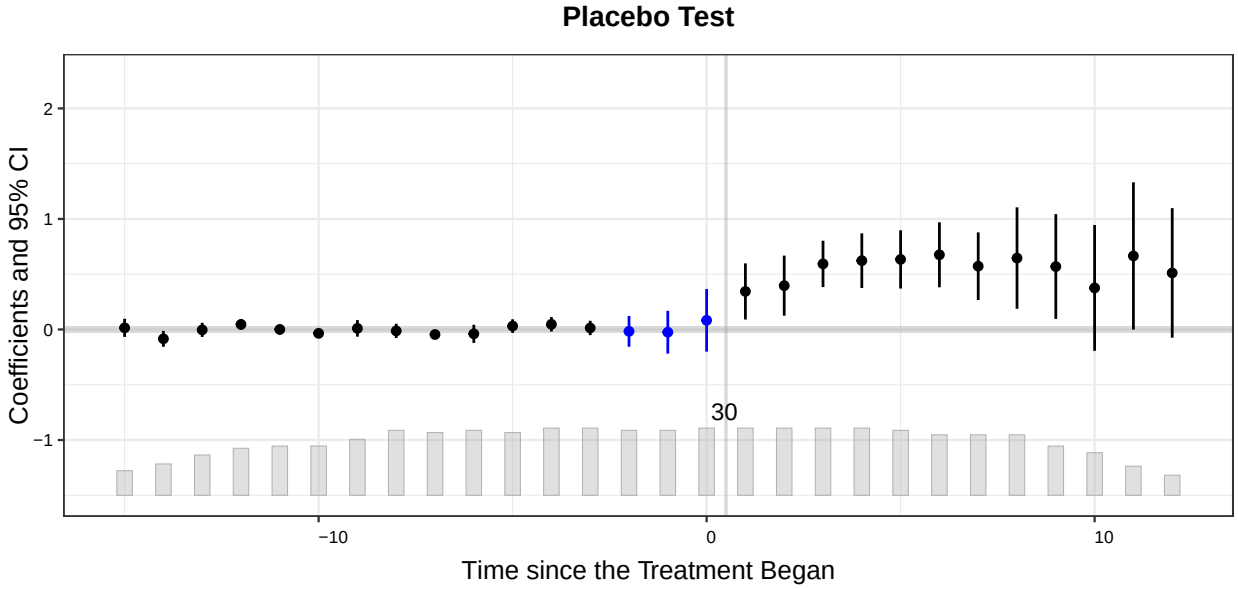
We also estimate the DTE using PanelMatch (lead = 4 and lag = 4). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test for (no) anticipation effects.

Placebo Test



Test Results

```
##                                   p-value
## F test                              0.200
## Equivalence test (default)          0.263
## Equivalence test (threshold=ATT)    0.000
## Placebo test                        0.876
## Carryover effect test                  NA
```

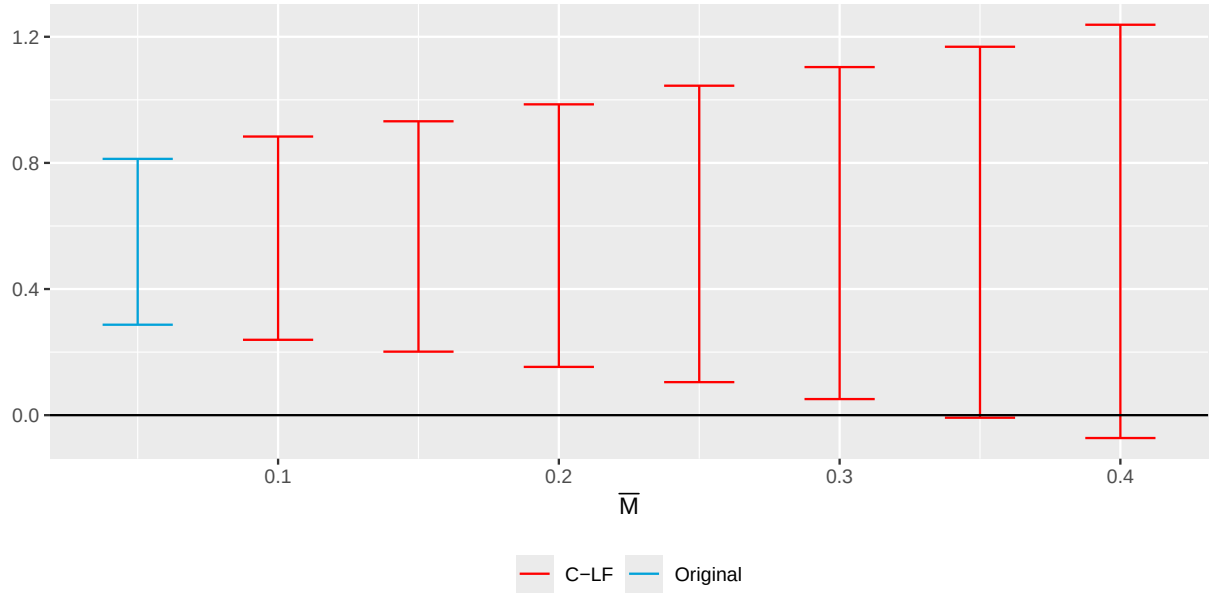
We find no evidence for violations of the parallel trends (PT) assumption. The equivalence test also rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

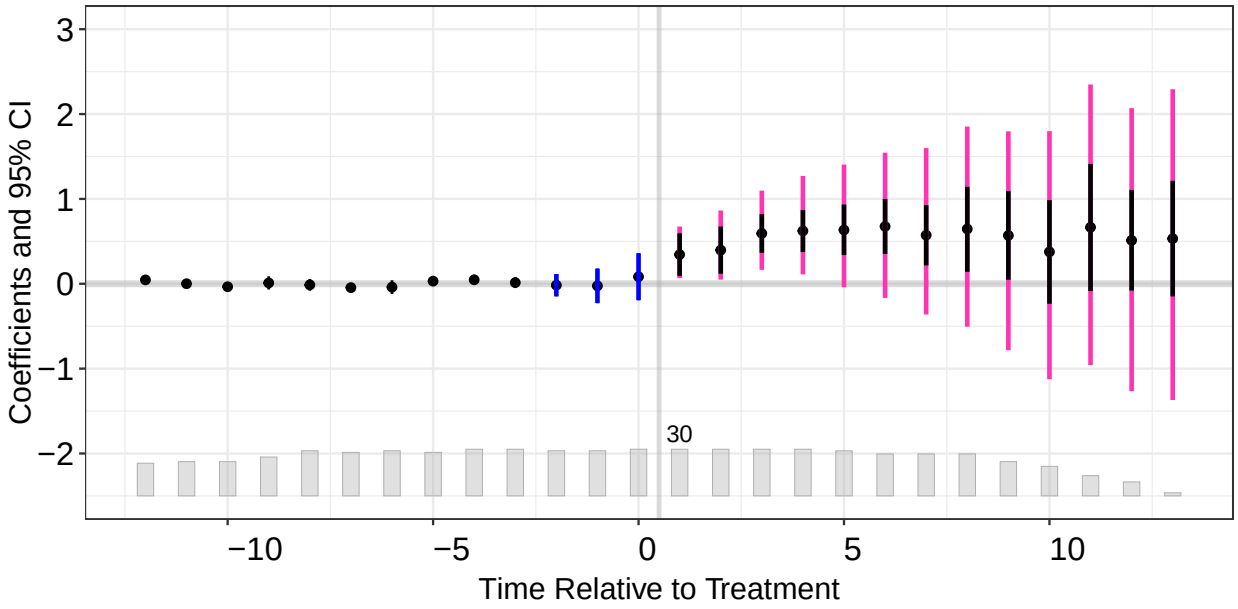
Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.35$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 0.35 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

```
##   |                                                                      |
```





Summary

Overall, the main result of the chosen model appears to be robust to HTE-robust estimators. We find little evidence of violations of the PT assumption, and the equivalence test suggests that violations would be small if they do exist. The estimated ATT appears to be relatively robust to mild PT violations in sensitivity analysis with the breakdown threshold $M = 0.35$.

Jiang (2018)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

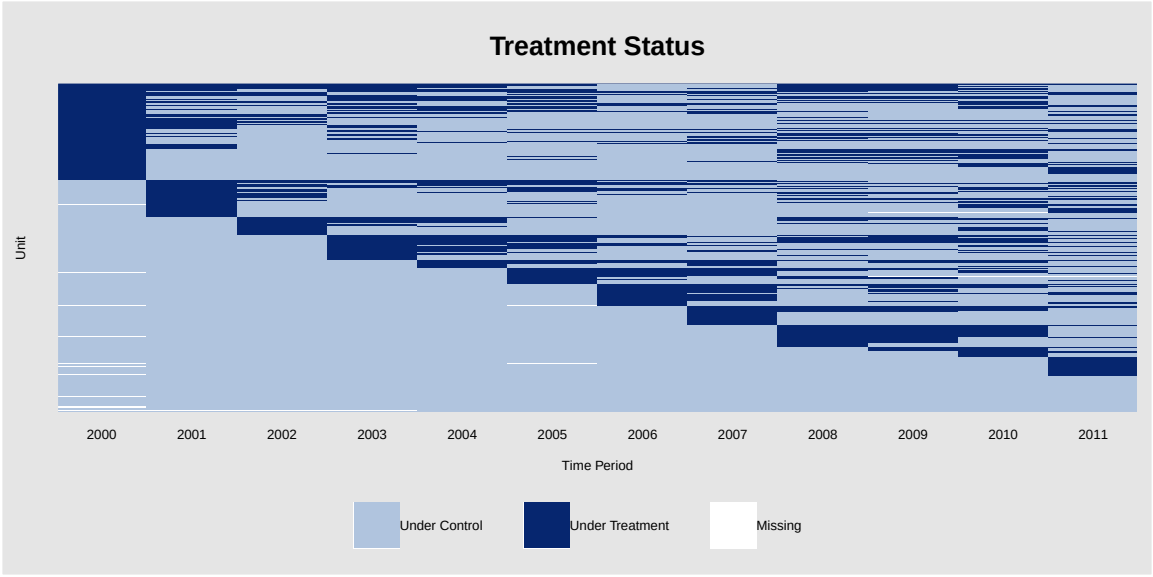
A First Look at Data

The paper investigates the effects of patronage network on economic growth, using city-year level panel data from China, between 2000 and 2011. One of the main findings of this paper is that "city leaders with informal ties to the incumbent provincial leaders deliver significantly faster economic growth than those without" (p982).

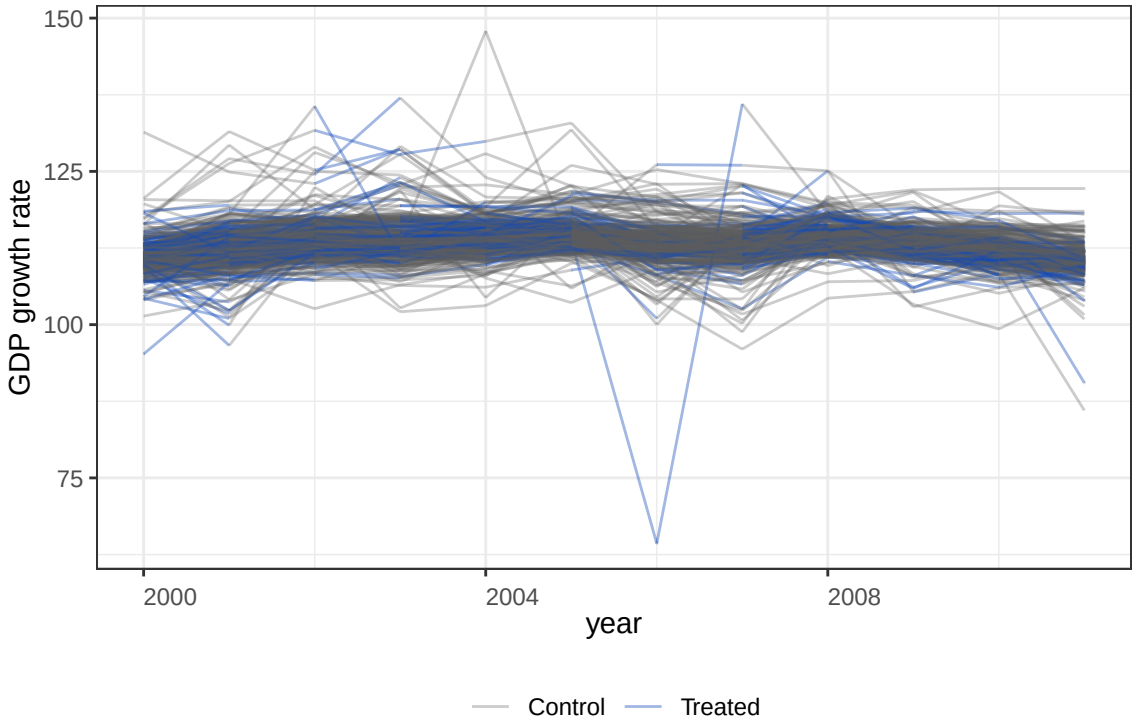
Model. We focus on **model 2 of Table 1** in the paper. We recode the treatment by flipping the value of 0 and 1 to estimate the average treatment effects on the control (ATC) with a flipped sign. The analysis below uses the recoded treatment. The treatment follows a general adoption setting with treatment reversals. The paper uses a TWFE with city and province×year fixed effects.

Replication Summary	
Unit of analysis	City × year
Treatment	Connected to provincial secretary
Outcome	GDP growth rate
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Higher-level Unit*Time dummy

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a city). There are treatment reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this data) and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis,

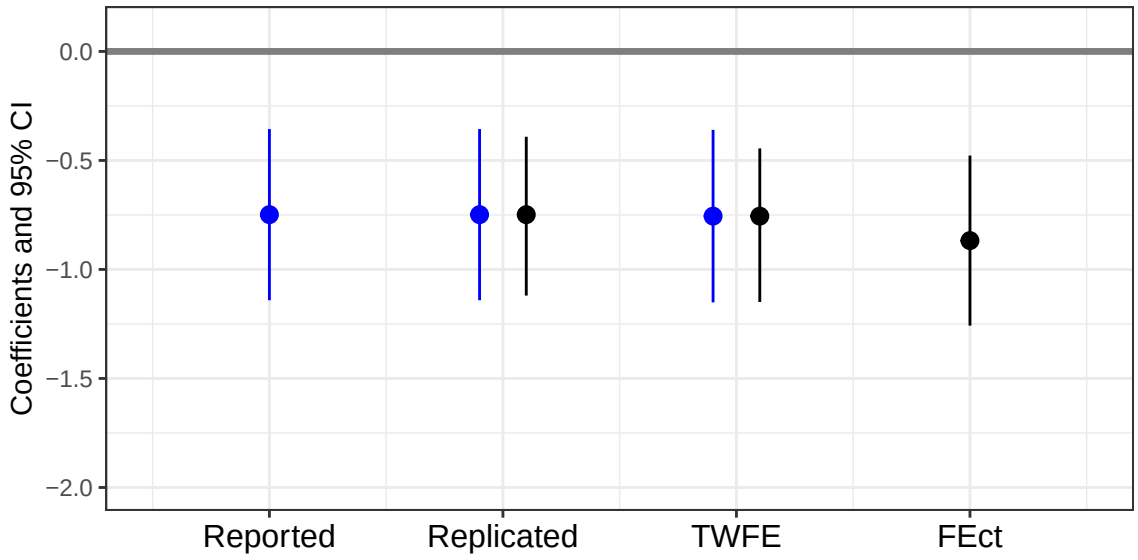
we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(dv~nd + startmsec_loggdp+startmayor_loggdp+
               startmsec_logpop+startmayor_logpop+
               startmsec_loginvest+startmayor_loginvest+
               startmsec_gdpidx+startmayor_gdpidx+dep|cityid+year + py,
             data = df,cluster = "cityid")
summary(sol)
```

```
## OLS estimation, Dep. Var.: dv
## Observations: 3,891
## Fixed-effects: cityid: 326,  year: 12,  py: 312
## Standard-errors: Clustered (cityid)
##                          Estimate  Std. Error   t value   Pr(>|t|)
## nd                    -0.748635425 0.200290917 -3.737740 0.00021933 ***
## startmsec_loggdp      -1.500542716 0.553666724 -2.710191 0.00708134 **
## startmayor_loggdp     -1.684977510 0.509247964 -3.308756 0.00104223 **
## startmsec_logpop      -0.200402499 1.094000877 -0.183183 0.85476856
## startmayor_logpop      0.613532345 0.866118965  0.708370 0.47922351
## startmsec_loginvest    0.319232126 0.308453542  1.034944 0.30146455
## startmayor_loginvest   0.408423772 0.277382230  1.472422 0.14187461
## startmsec_gdpidx       0.030579052 0.044812091  0.682384 0.49548226
## startmayor_gdpidx     -0.013686830 0.022806253 -0.600135 0.54883449
## dep                    0.000000711 0.000000284  2.506902 0.01266716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.1532     Adj. R2: 0.558873
##                Within R2: 0.045914
```

```
##   |                                                                      |
```



TWFE and FEct are broadly consistent with each other, the estimated ATC are negative and statistically significant.

```
##   |                                                                      |
```

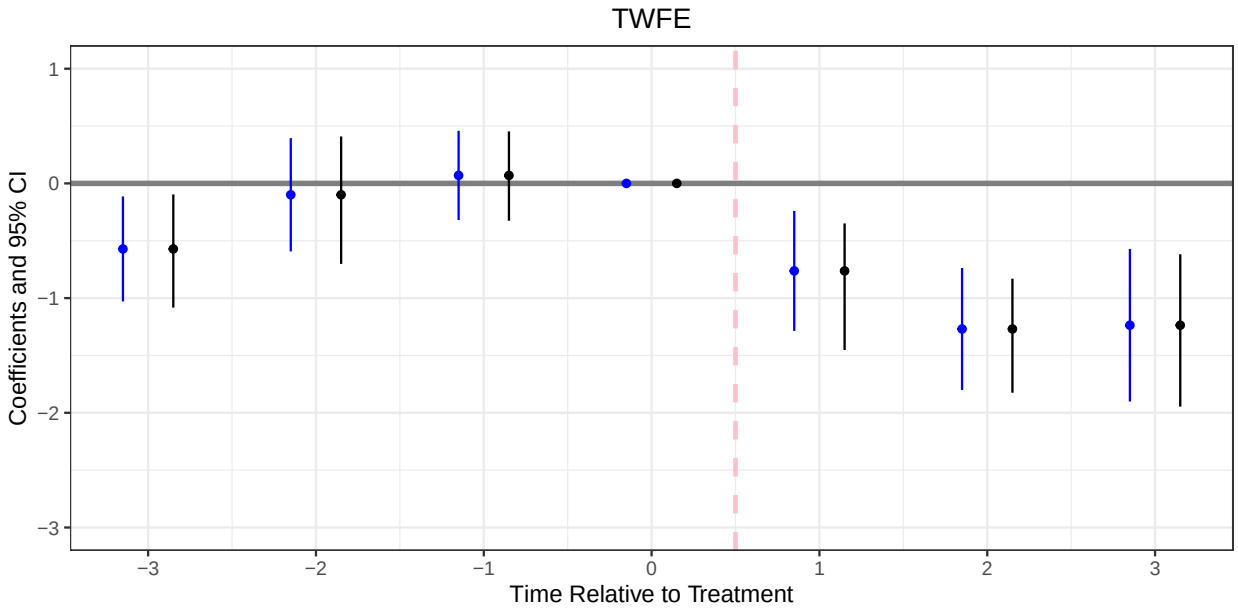
```
##   |                                                                      |
```

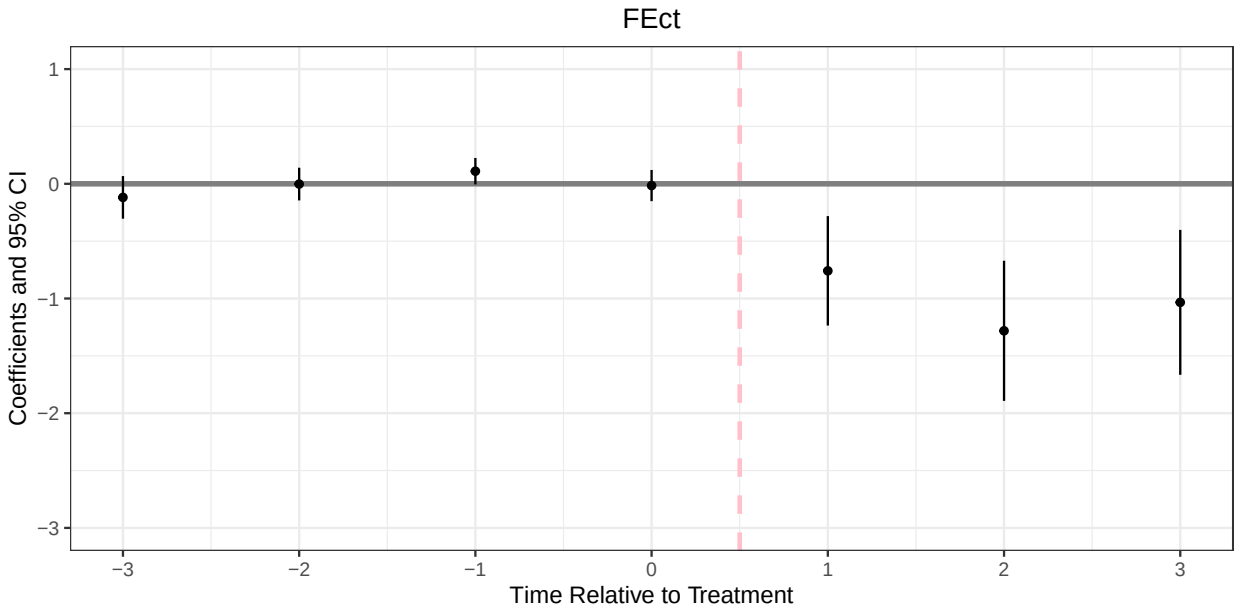
Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.

```
##   |                                                                      |
```

```
##   |                                                                      |
```



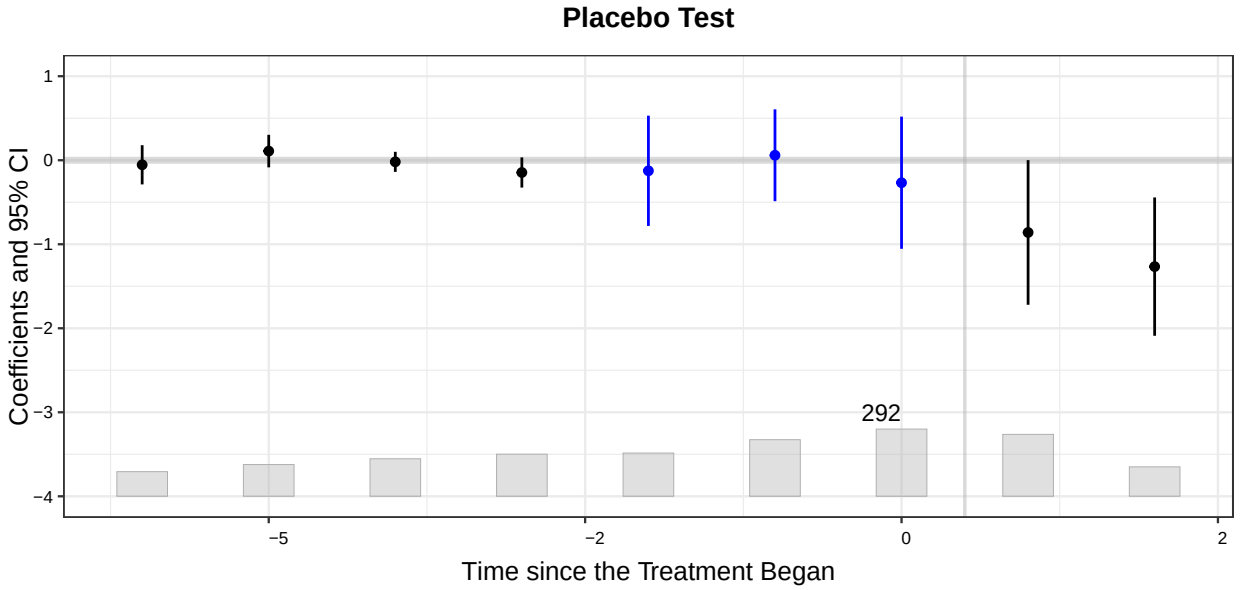


TWFE and FEct are broadly consistent with each other. The estimated DTEs are negative on all post-treatment periods. The estimated DTE using TWFE exhibits a weak upward pre-trend.

Diagnostic Tests

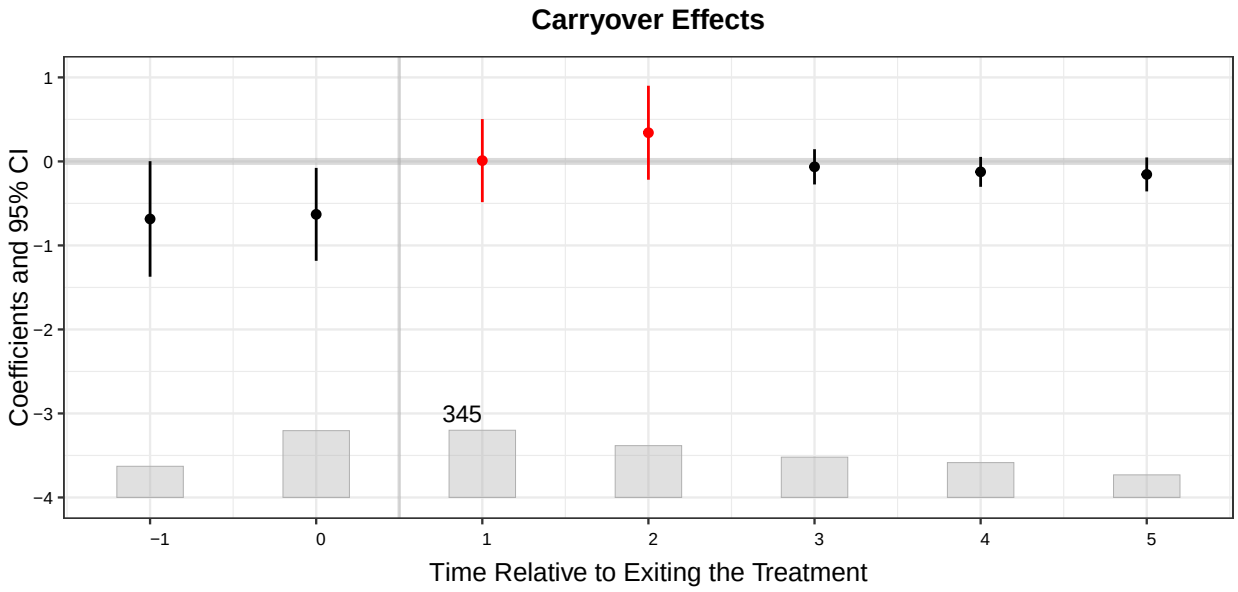
Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.

```
##   |                                                                      |
```




```
##   |                                                                      |
```

```
##   |                                                                      |
```



Test Statistics

```
##                                    p-value
## F test                               0.508
## Equivalence test (default)           0.000
## Equivalence test (threshold=ATC)     0.000
## Placebo test                         0.626
## Carryover effect test                0.504
```

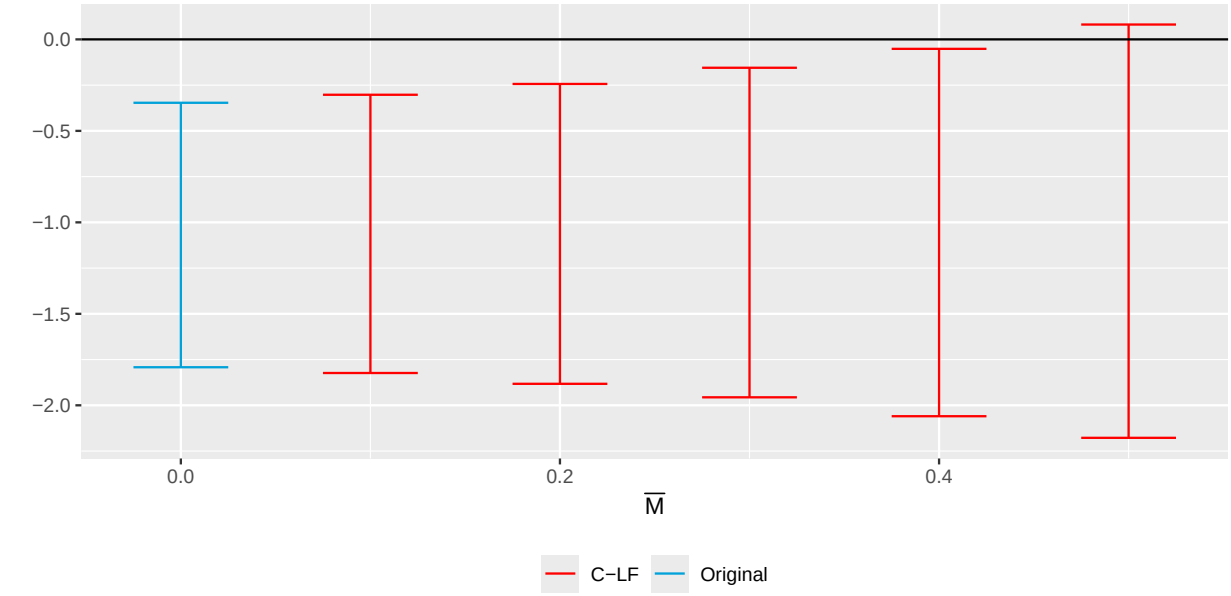
We find no evidence for violations of the parallel trends (PT) assumptions and the no-carryover-effect assumption. The equivalence test also rejects the null that the residuals in pre-treatment periods exceed the estimated ATC.

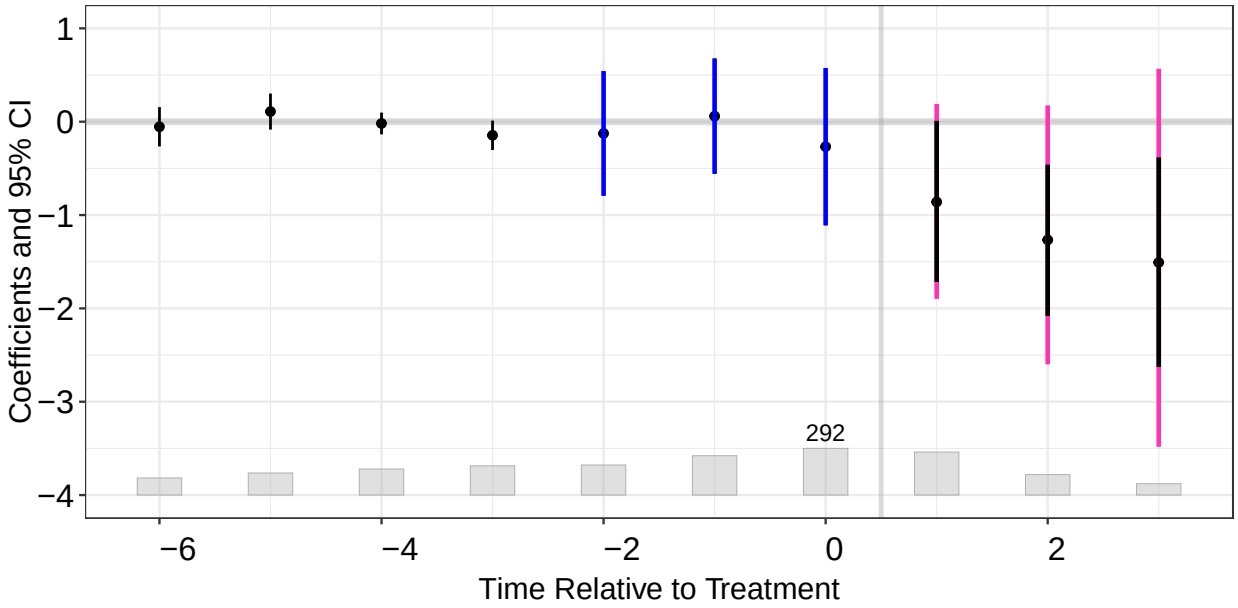
Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATC derived from the FEct estimator is significant. We assess the robustness of this significant ATC using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the PT observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.45$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 0.45 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

```
##   |                                                                      |
```





Summary

Overall, the main result of the chosen model appears to be robust HTE-robust estimators. We find no evidence for violations of the PT assumption and the no-carryover-effect assumption, and the equivalence test suggests that violations would be small if they do exist. The estimated ATC appears to be relatively robust to mild PT violations in sensitivity analysis with the breakdown threshold $M = 0.45$.

Kilborn and Vishwanath (2022)

Contents

A First Look at Data 1
Point Estimates 3
Dynamic Treatment Effects 4
Diagnostic Tests 5
Sensitivity Analysis 7
Summary 8

A First Look at Data

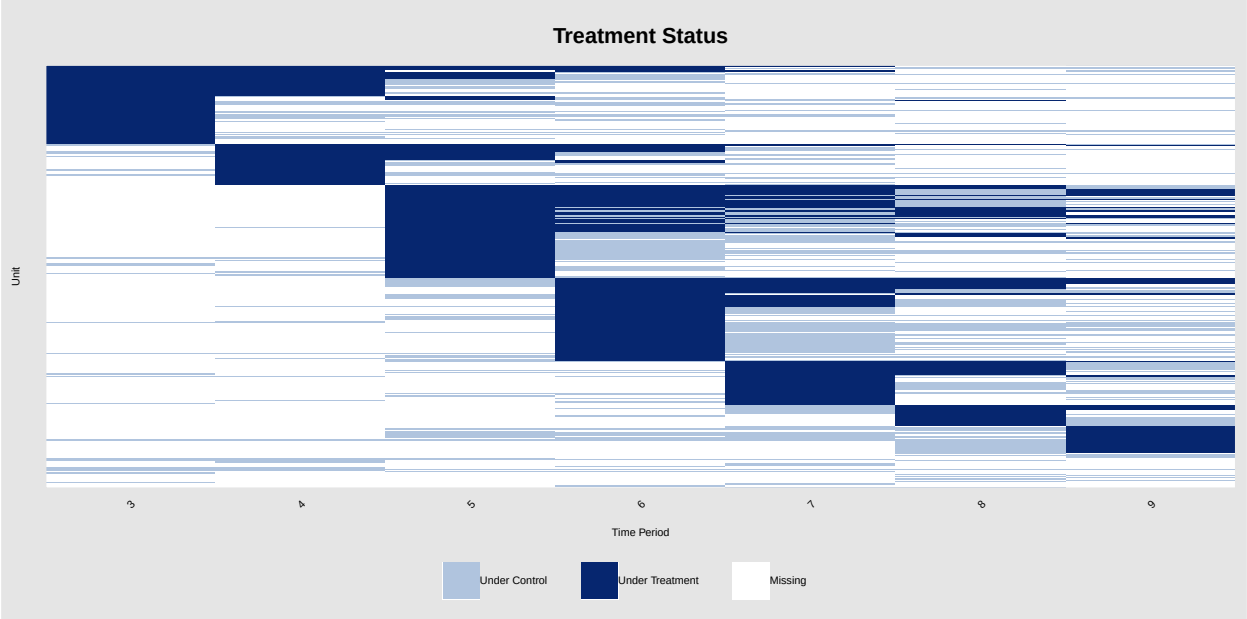
The paper investigates the effects of public campaign financing on the candidates' representation, using US politician-election year level panel data between 2004 and 2016. One of the main findings of this paper is that when "a candidate switches from private to public financing between election years, her ideological distance to the average constituent increases and vice versa" (p739, Table 3).

Model. We focus on **Model 1 of Table 3** in the paper. The authors use a two-way fixed effects (TWFE) model with an additional FE at a different level and report robust standard errors using error propagation method.

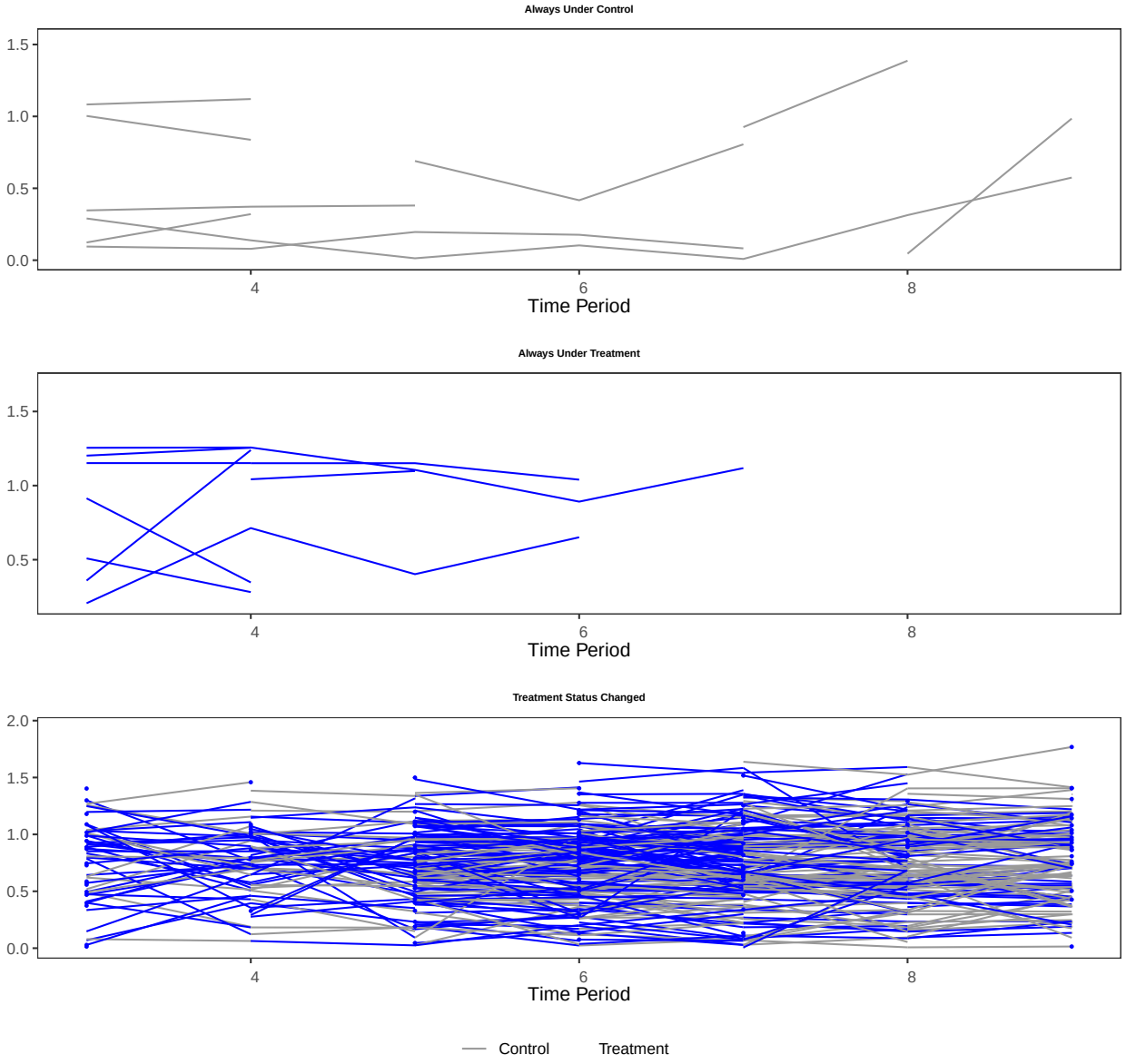
Table 1: Replication Summary

Unit of analysis	Politicians × election year
Treatment	Public campaign finance
Outcome	Ideological distance
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time+Higher-level Unit

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a politician). There are treatment reversals and some missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each politician. The observations under treated status are marked in blue.



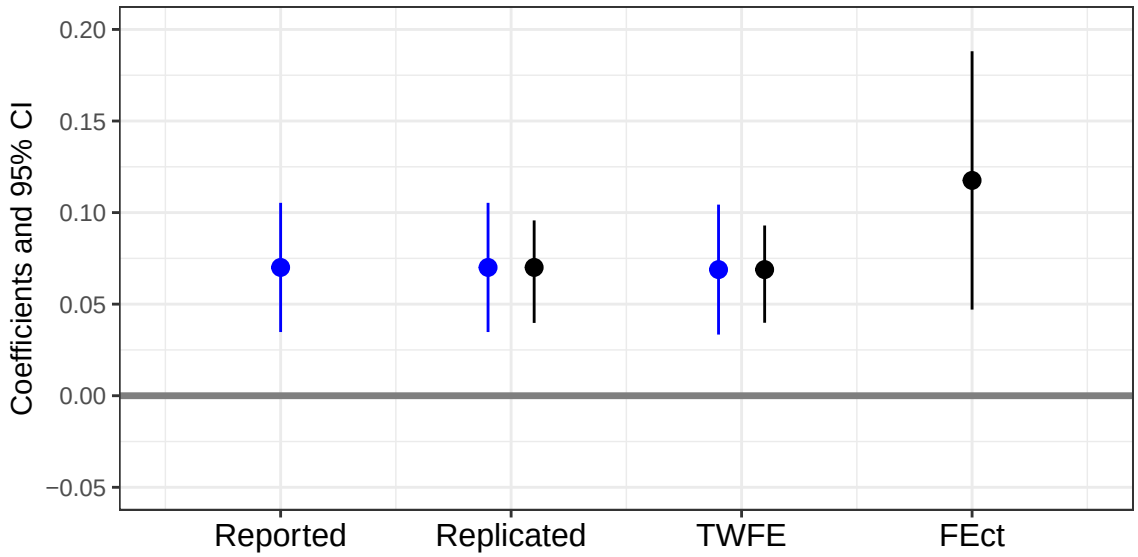
Point Estimates

We first present the regression result following the authors' original specification (using cluster-robust SEs). We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(Distance_CFDyn~CleanYear|id + year + UniqueDistrict_CensusGroup,
             data = df,cluster = "id")
summary(sol)
```

```
## OLS estimation, Dep. Var.: Distance_CFDyn
## Observations: 1,062
## Fixed-effects: id: 347,  year: 7,  UniqueDistrict_CensusGroup: 453
## Standard-errors: Clustered (id)
##           Estimate Std. Error t value   Pr(>|t|)
## CleanYear 0.070049   0.017997 3.89232 0.00011912 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.106124     Adj. R2: 0.603181
##                  Within R2: 0.042905
```

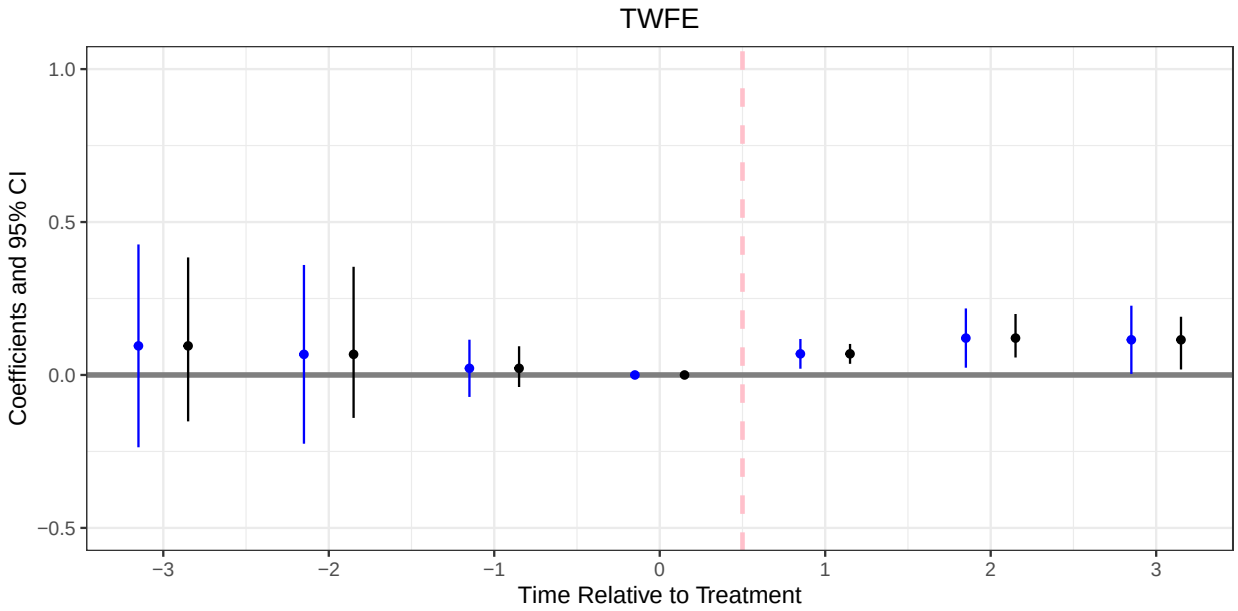


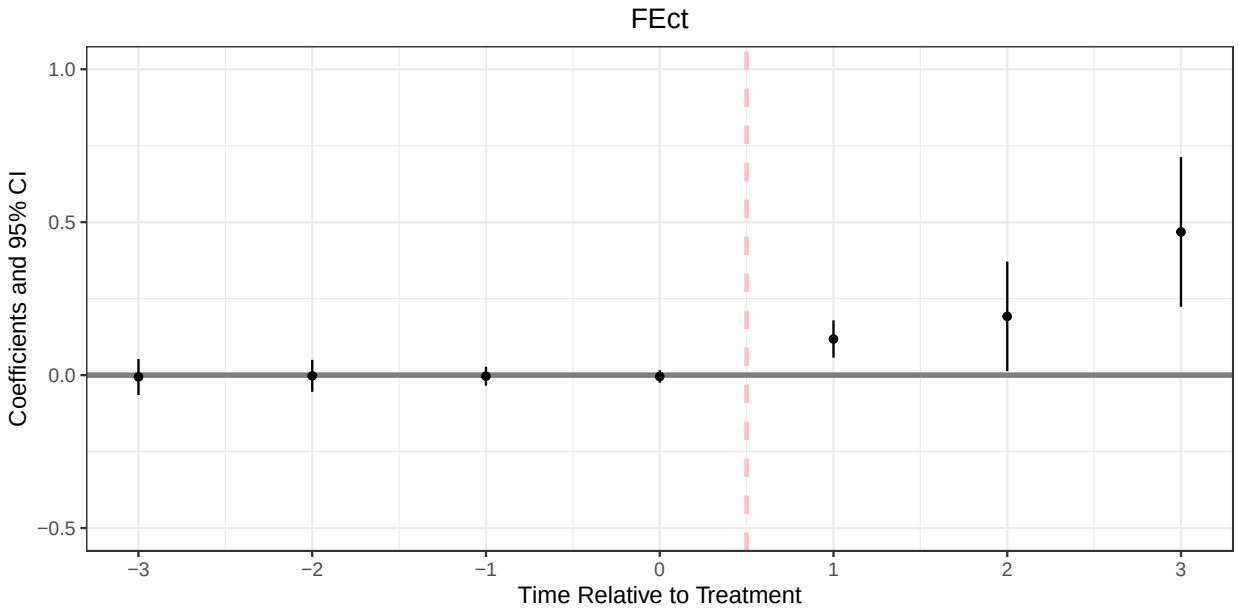
Both TWFE and FEct estimates are positive and statistically significant, while the latter is larger in magnitude and has wider CIs.

The TWFE and FEct estimates are robust to LDV. The FEct estimate is no longer significant under ULT. Note that FEct with ULT requires a large number of untreated observations for each treated unit, so the result should be interpreted with caution.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.

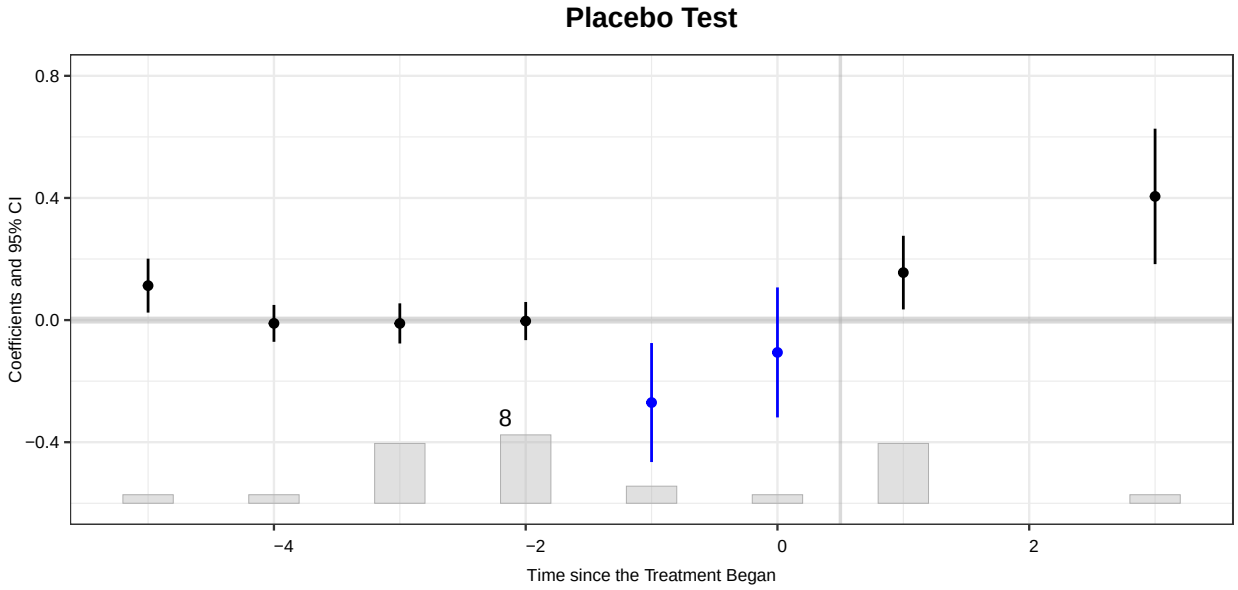


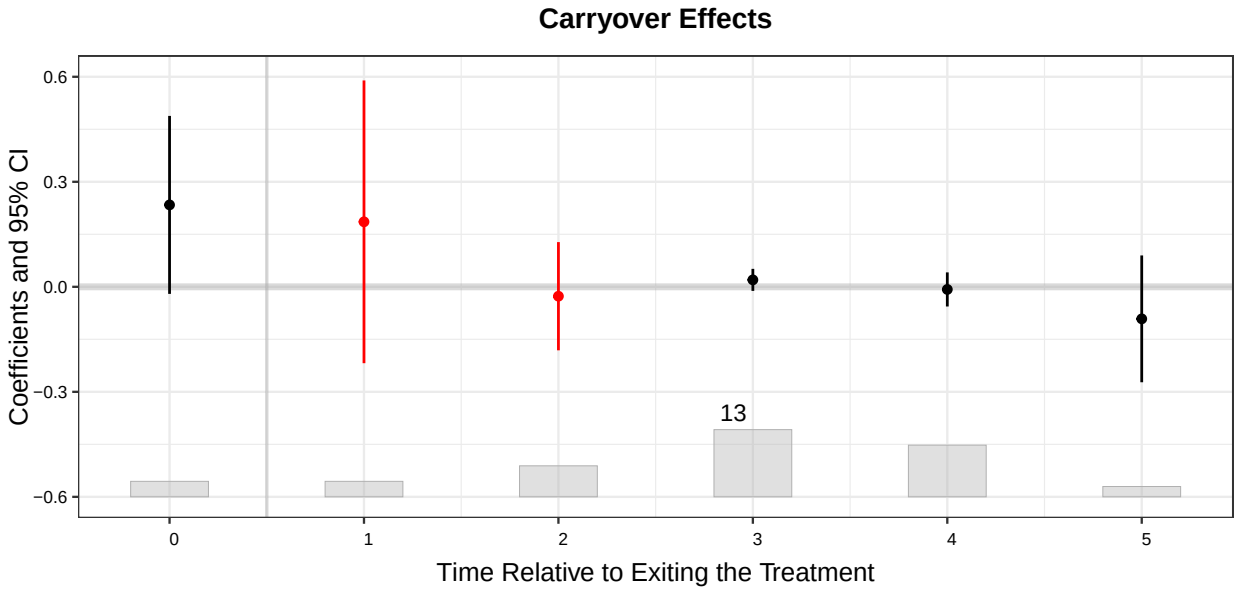


The estimated DTE are all positive during post-treatment periods.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test for (no) anticipation effects, and a test for (no) carryover effects.





Test Statistics

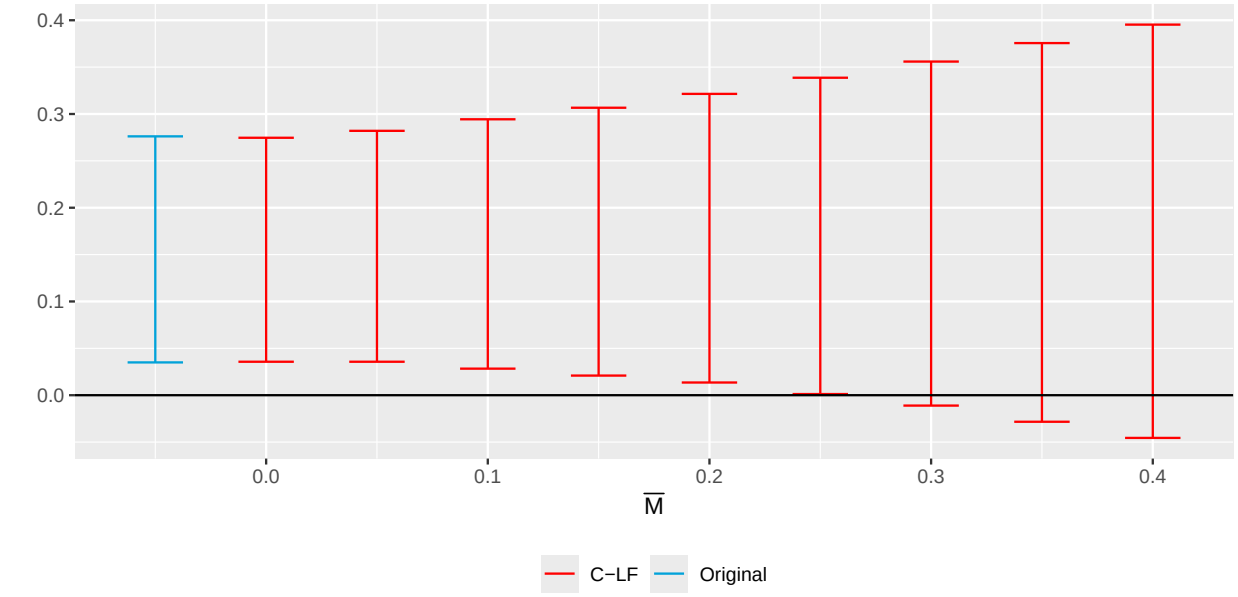
```
##                                  p-value
## F test                             0.794
## Equivalence test (default)         0.027
## Equivalence test (threshold=ATT)   0.000
## Placebo test                       0.000
## Carryover effect test              0.650
```

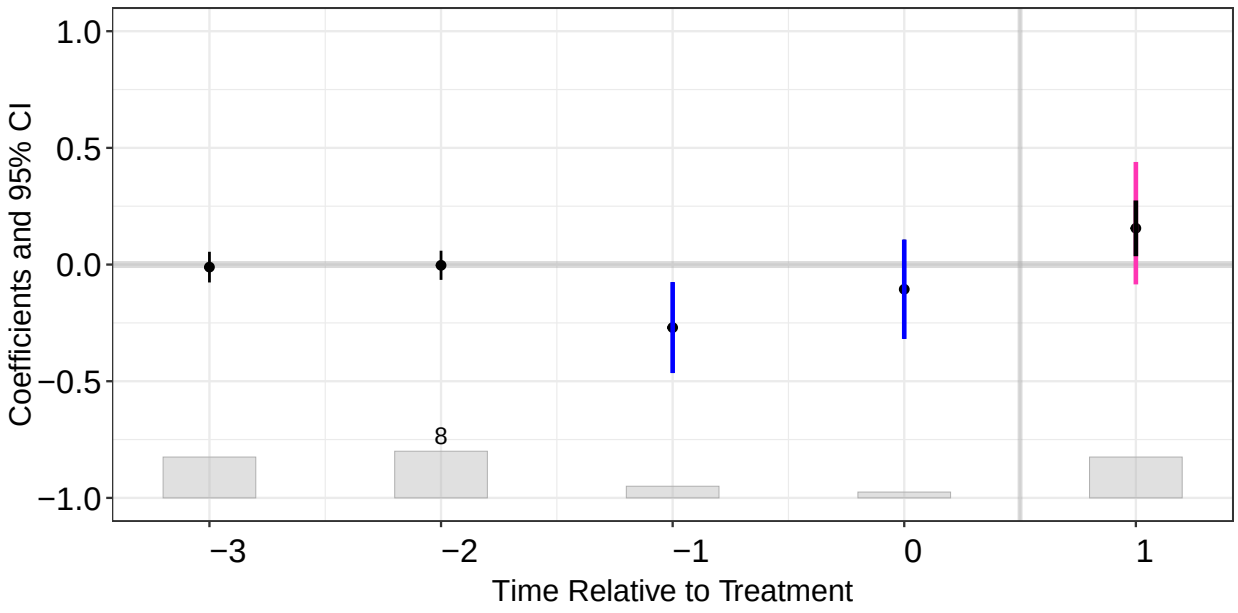
We find some evidence for potential violations of the no-anticipation effects assumption using the placebo test. However, the F test cannot reject the null, and the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT. We also find little evidence for violations of the no-carryover-effect assumption.

Sensitivity Analysis

After excluding three pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the three placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.25$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.25 times the maximum deviation observed between consecutive periods in the placebo periods.

We further evaluate the robustness of the estimated DTE under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, accounting for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust to FEct, an HTE-robust estimator, although the point estimates from FEct and TWFE diff substantially. We find some evidence for potential violations of the PT. The estimated ATT appears to be relatively sensitive to mild PT violations with the breakdown threshold $M = 0.25$.

Kogan (2021)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic tests 4
Sensitivity Analysis 5
Summary 5

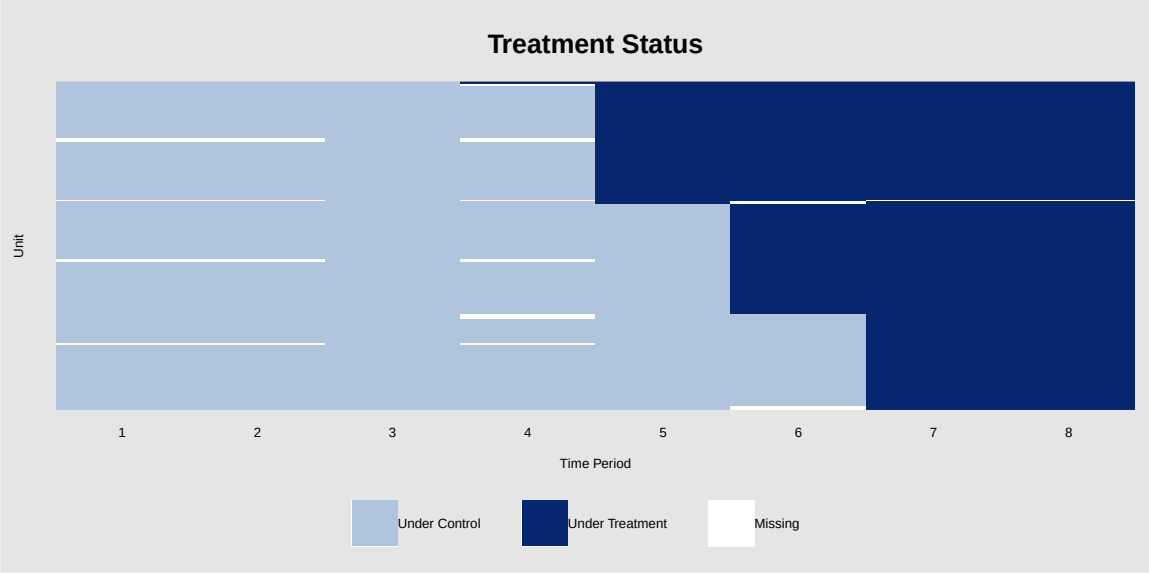
A First Look at Data

The paper investigates the effects of American Food Stamp Program (FSP) on Democratic Presidential vote share, using county-year level panel data from 1952 and 1980. One of the main findings of this paper is that "FSP introduction increased the share of the two-party vote won by Democratic presidential candidates by between 0.9 and 1.6 percentage points" (p65).

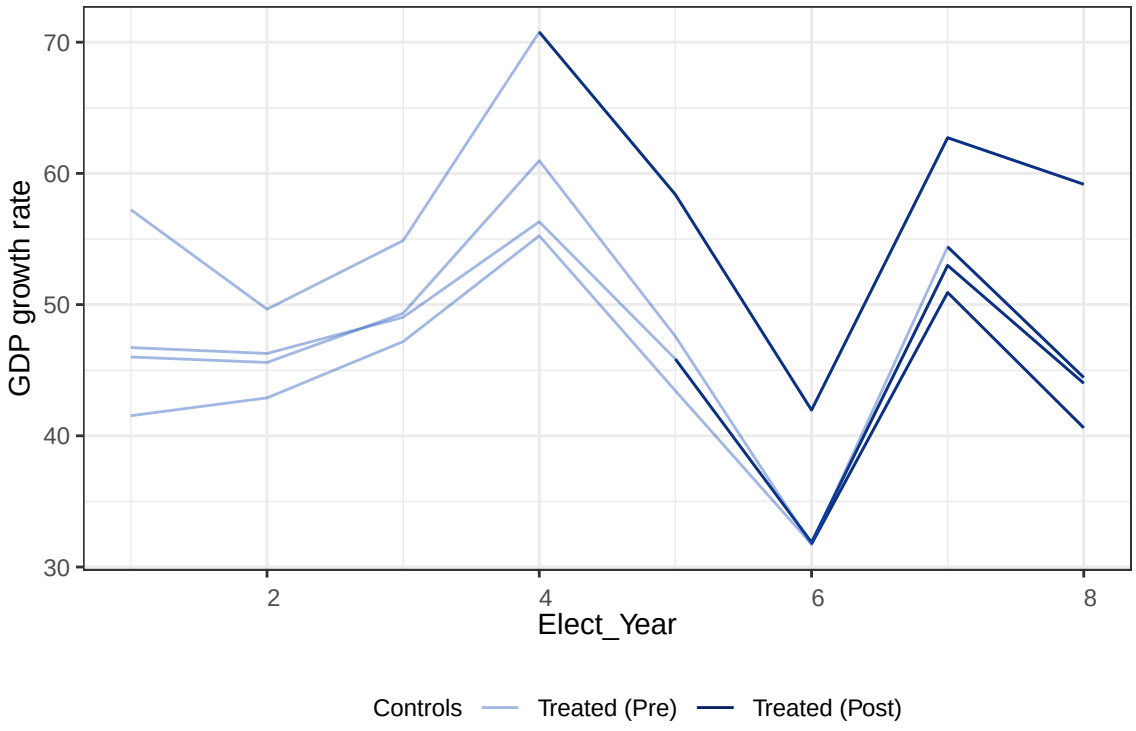
Model. We focus on **Model 1 of Table 3** in the paper. The author uses a unit (county), higher-level unit*time dummy fixed effect (state × year), and unit specific linear time trend (ULT) (County × linear trend), and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	County × election year
Treatment	FSP
Outcome	Democratic Presidential vote share
Treatment type	Staggered
Outcome type	Continuous
Fixed Effects	Unit+Higher-level Unit*Time dummy+Unit specific linear time trend

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a county).



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this data) and apply two estimators: TWFE (with state × year FE and

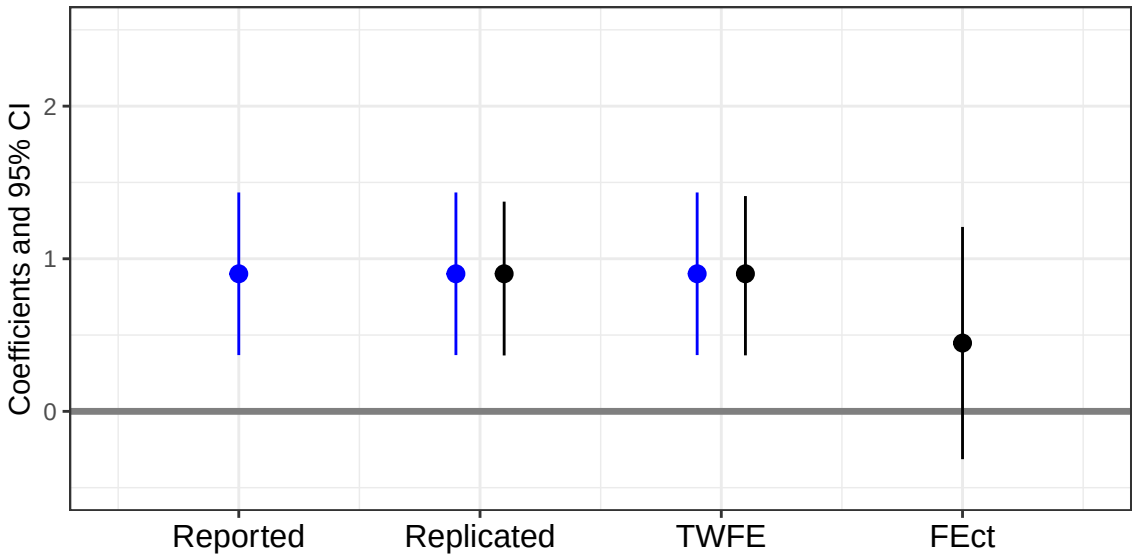
County × linear trend) and the imputation method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(dem_share~rollout |Elect_Year+
               FIPS_State + fips + fips[Elect_Year] + fez,data = df,cluster = "fips")
summary(sol)
```

```
## OLS estimation, Dep. Var.: dem_share
## Observations: 23,610
## Fixed-effects: Elect_Year: 8,  FIPS_State: 48,  fips: 3,005,  fez: 377
## Varying slopes: Elect_Year (fips): 3,005
## Standard-errors: Clustered (fips)
##         Estimate Std. Error t value   Pr(>|t|)
## rollout 0.901587   0.271844 3.31657 0.00092215 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 4.15944     Adj. R2: 0.895989
##                 Within R2: 0.001172
```

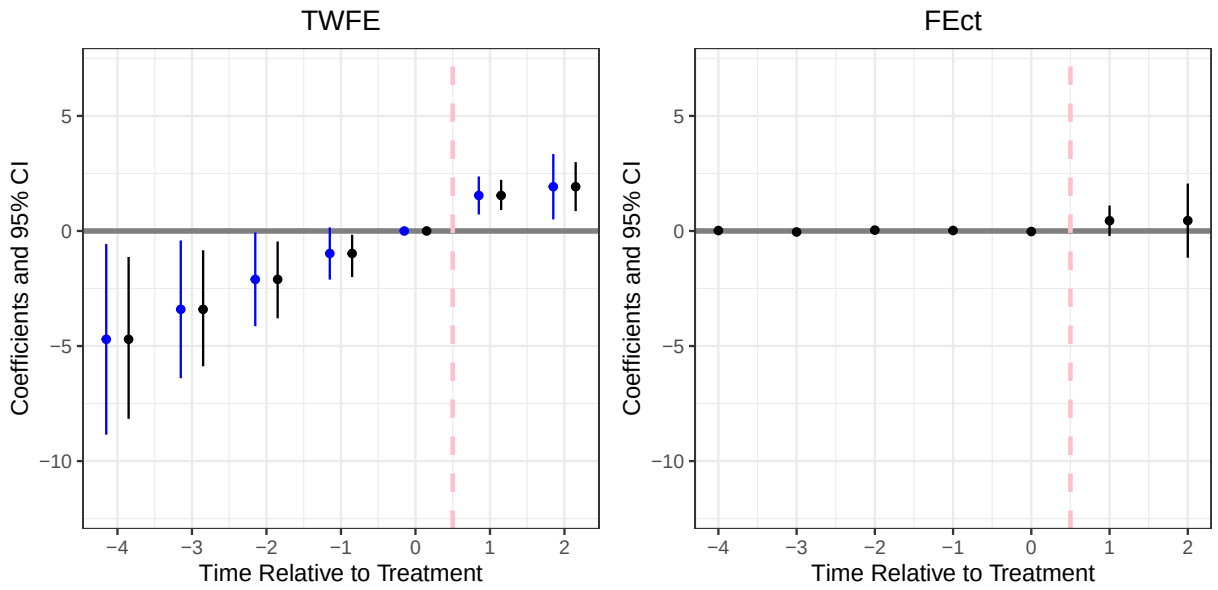
```
##   |                                                                      |
```



Both TWFE and FEct estimates are positive, while the latter is not statistically significant and has wider CIs.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The observations in the last two periods are dropped automatically because there is no untreated unit. The results are shown below.

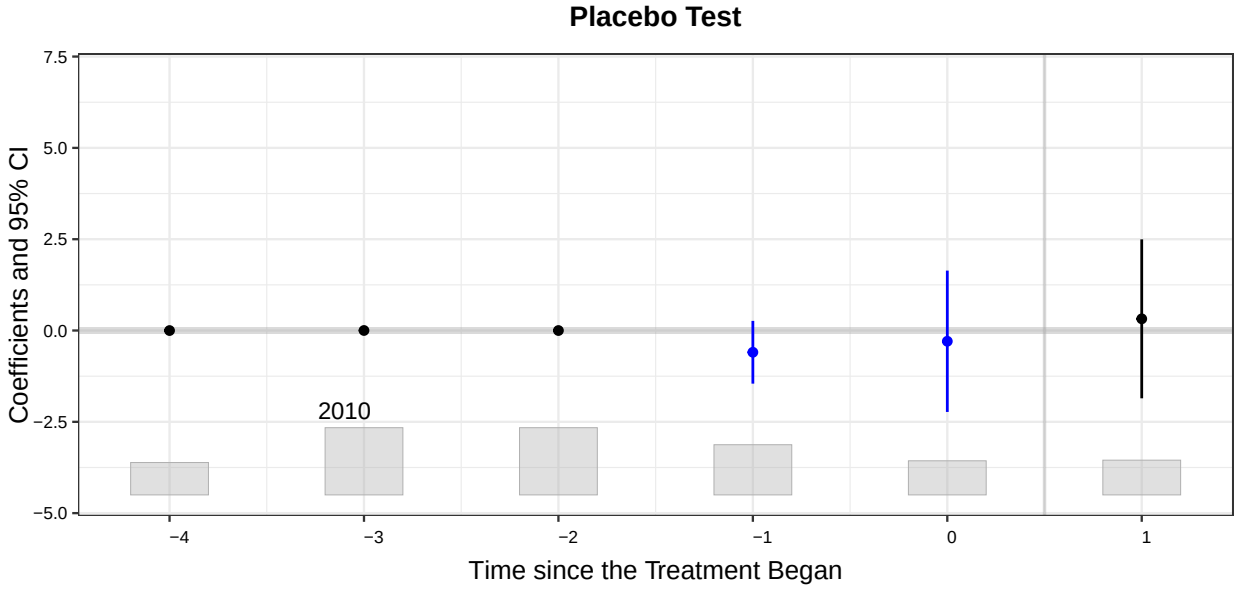


The DTE estimated using TWFE are significant in most pre-treatment periods and positive and significant during post-treatment periods. The DTEs estimated using FEct are insigificant in all periods.

Diagnostic tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.

```
##   |                                                                      |
```



Test Statistics

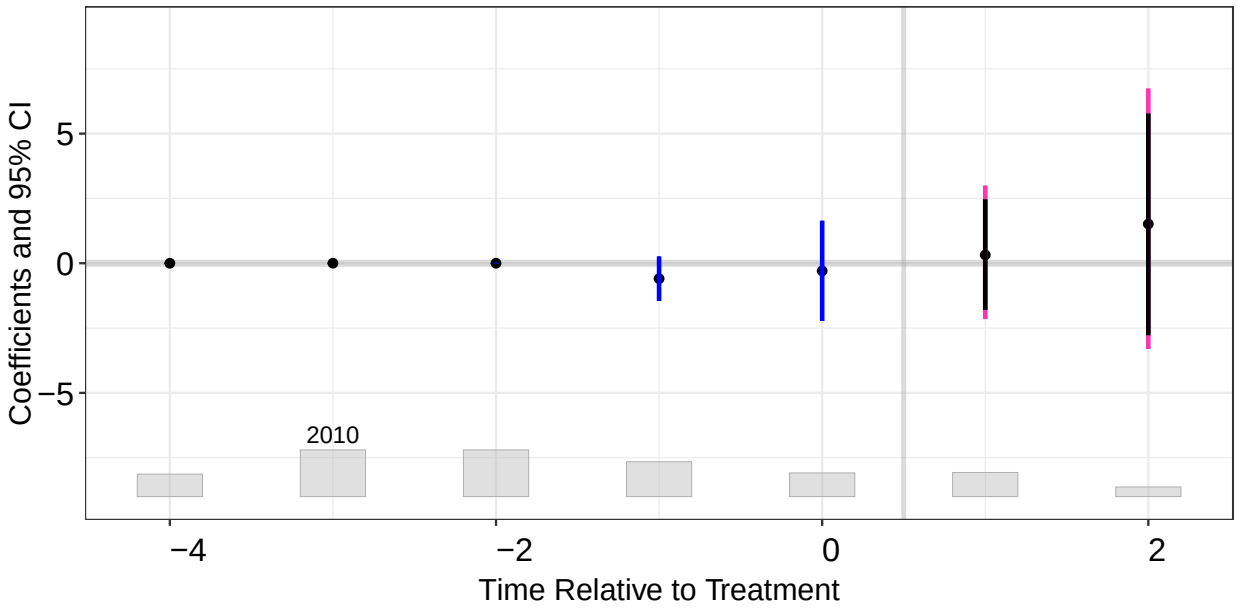
```
## Cannot use full pre-treatment periods in the F test. The first period is removed.
```

```
##                                  p-value
## F test                             0.851
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.000
## Placebo test                       0.442
## Carryover effect test                 NA
```

We find no evidence for violations of the parallel trends (PT) assumptions or the no-carryover effects assumption. The equivalence test also rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out two pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, we find little evidence for violations of the PT or no-carryover effects assumption. However, the main finding of the chosen model does not seem to be robust to the imputation estimator (FEct), an HTE-robust estimator.

Kroeger and Silfa (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 5
Summary 6

A First Look at Data

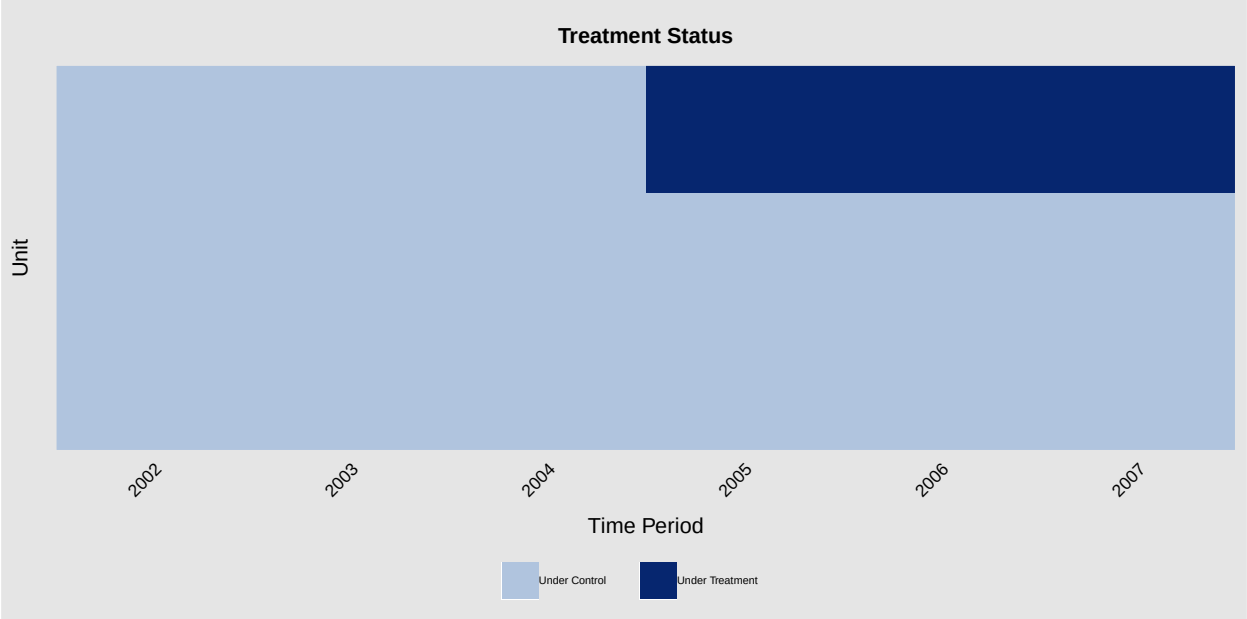
The paper investigates the effects of exempted from price restrictions on short selling on firms' lobbying, using US firm-year level panel data between 2002 and 2007. One of the main findings of this paper is that " Being treated is associated with a change in the probability of lobbying approximately 2.8 percentage points higher than those in the control group." (p1142)

Model. We focus on **Model 1 of Figure 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the firm level.

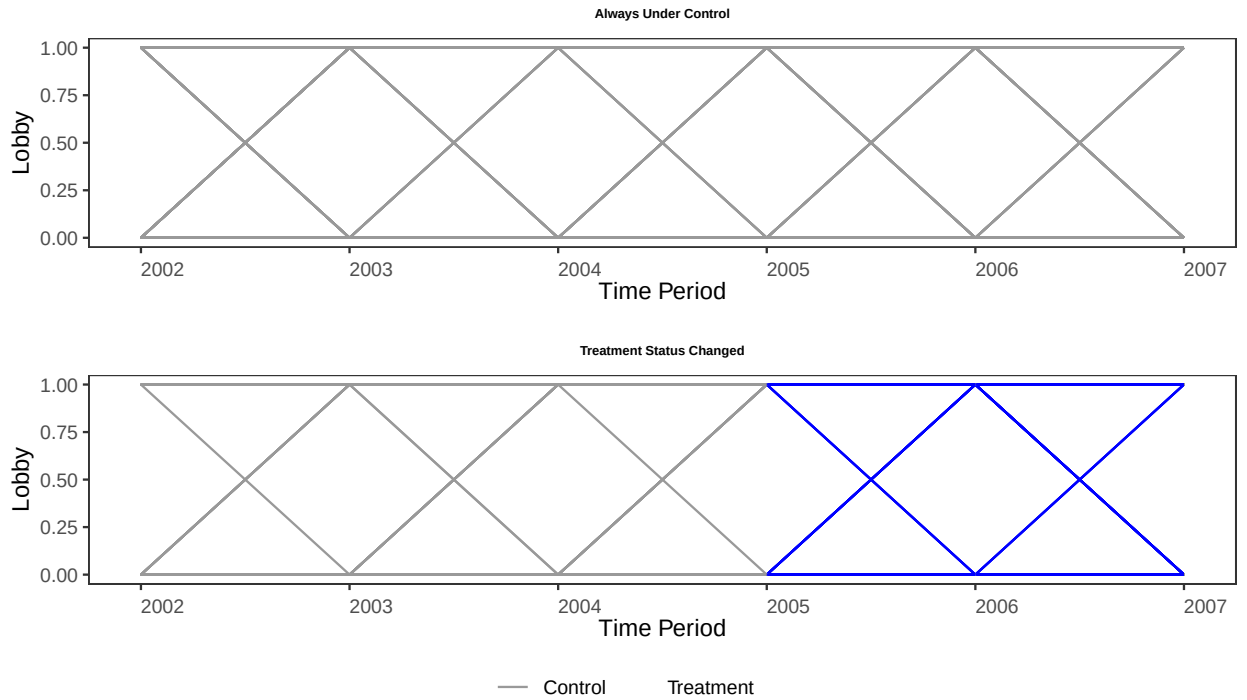
Table 1: Replication Summary

Unit of analysis	Firm × year
Treatment	Exempted from price restrictions on short selling
Outcome	Lobby
Treatment type	Classic
Outcome type	Binary
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a firm). The treatment has no reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

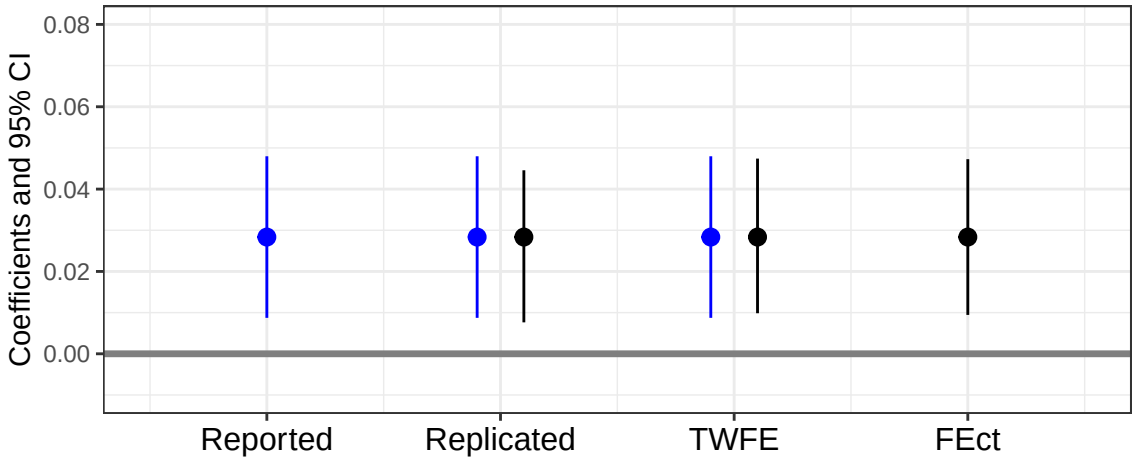
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates

and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(lobby~ treated|ticker+year,
             data = df,cluster = "ticker")
summary(sol)
```

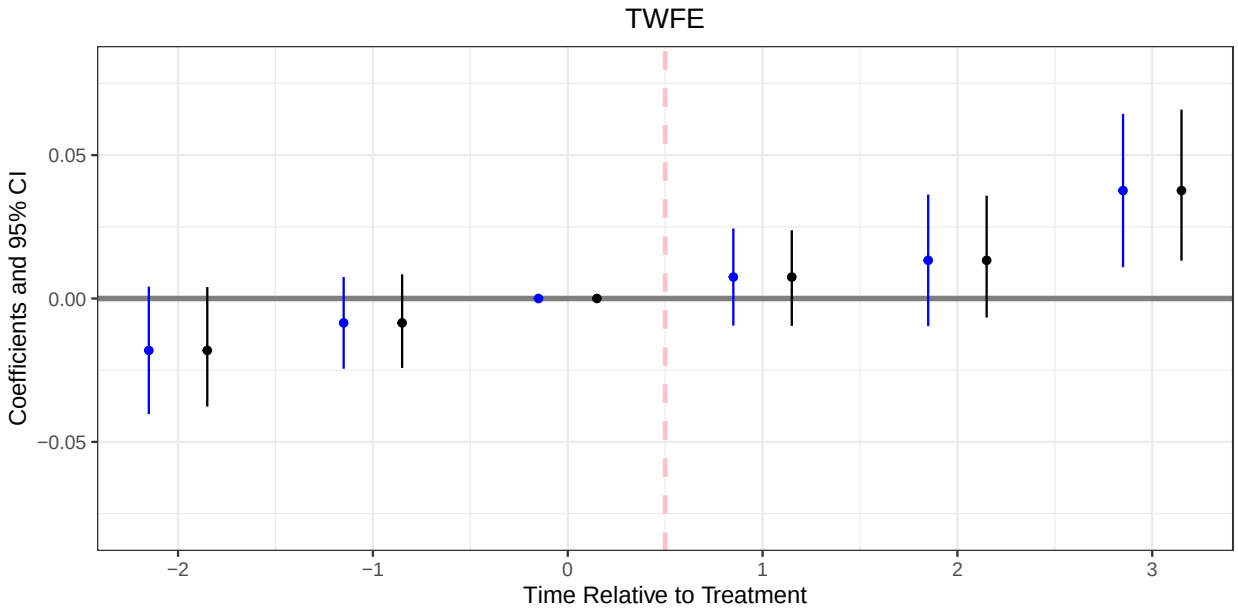
```
## OLS estimation, Dep. Var.: lobby
## Observations: 17,010
## Fixed-effects: ticker: 2,835,  year: 6
## Standard-errors: Clustered (ticker)
##         Estimate Std. Error t value  Pr(>|t|)
## treated  0.02835   0.010011  2.8319 0.0046599 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.191143     Adj. R2: 0.791777
##                  Within R2: 0.001219
```

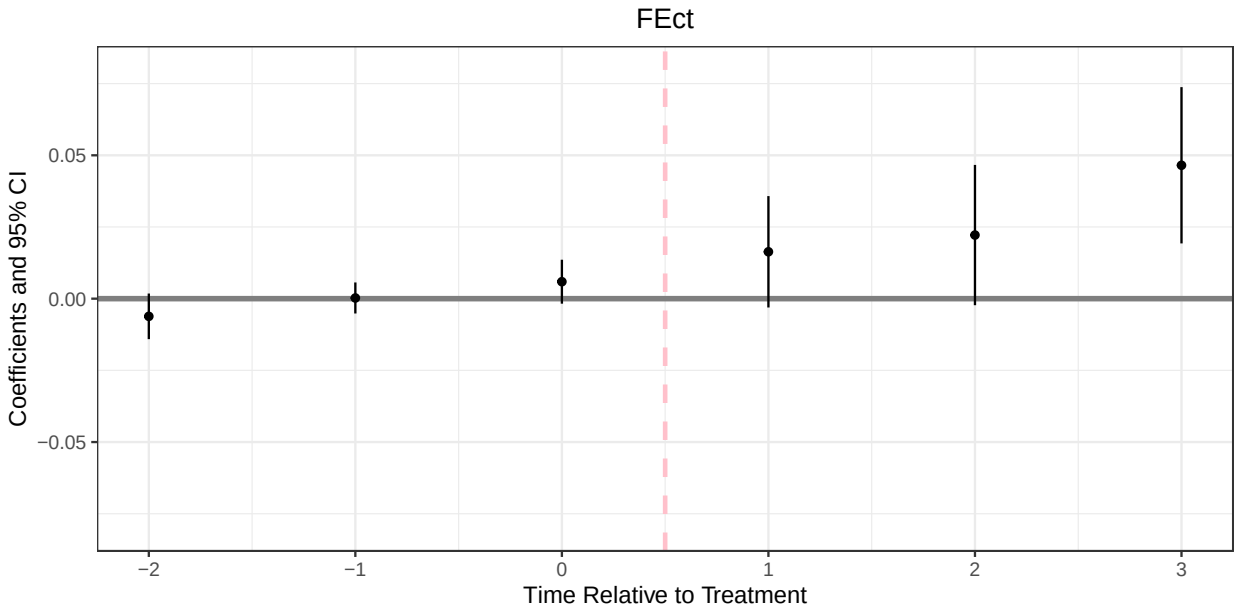


The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct.

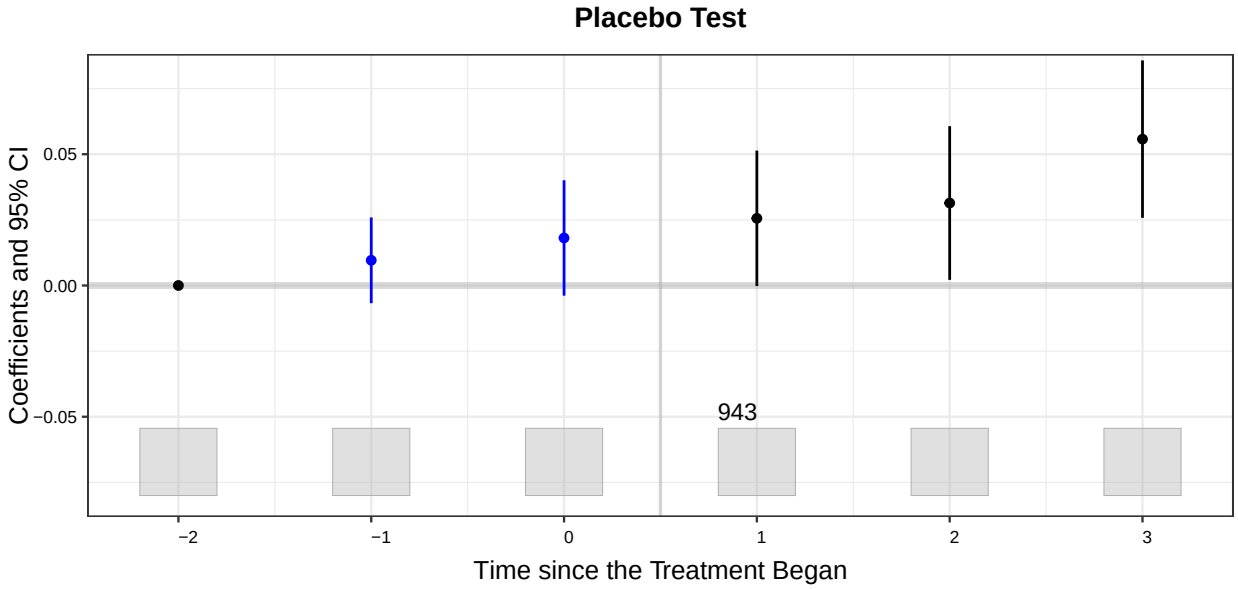




The TWFE and FEct estimator are consistent with each other. The estimated DTEs are all positive in the plot.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.



Test Statistics

```
## Cannot use full pre-treatment periods in the F test. The first period is removed.
```

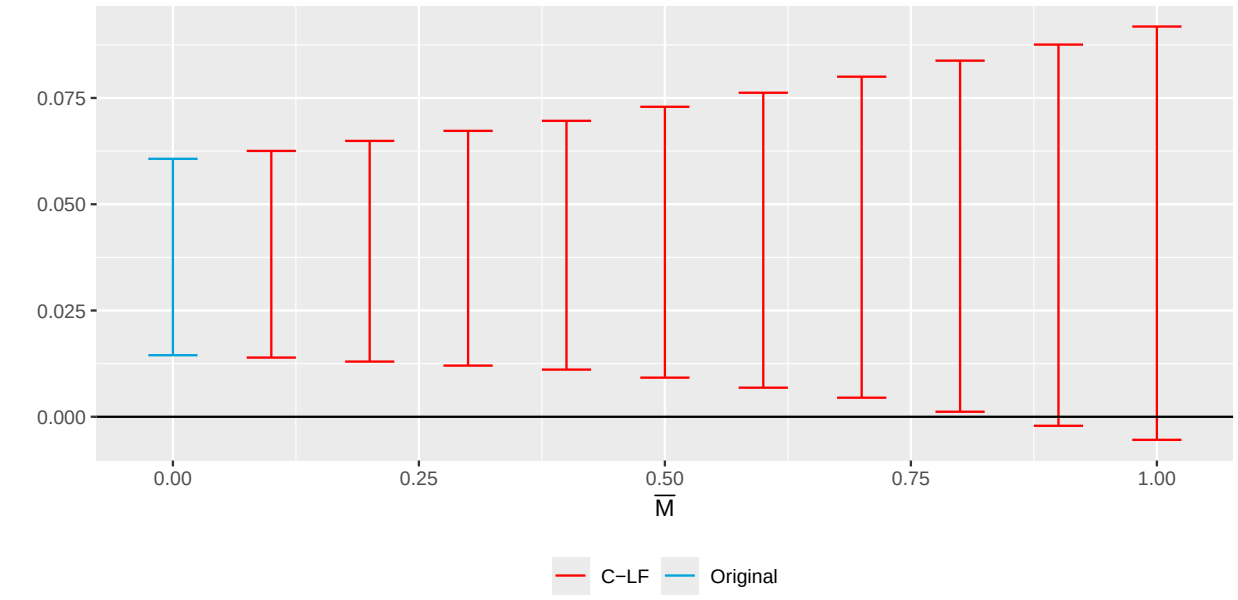
```
##                                  p-value
## F test                          2.72e-01
## Equivalence test (default)      0.00e+00
## Equivalence test (threshold=ATT) 4.78e-09
## Placebo test                    1.28e-01
## Carryover effect test                NA
```

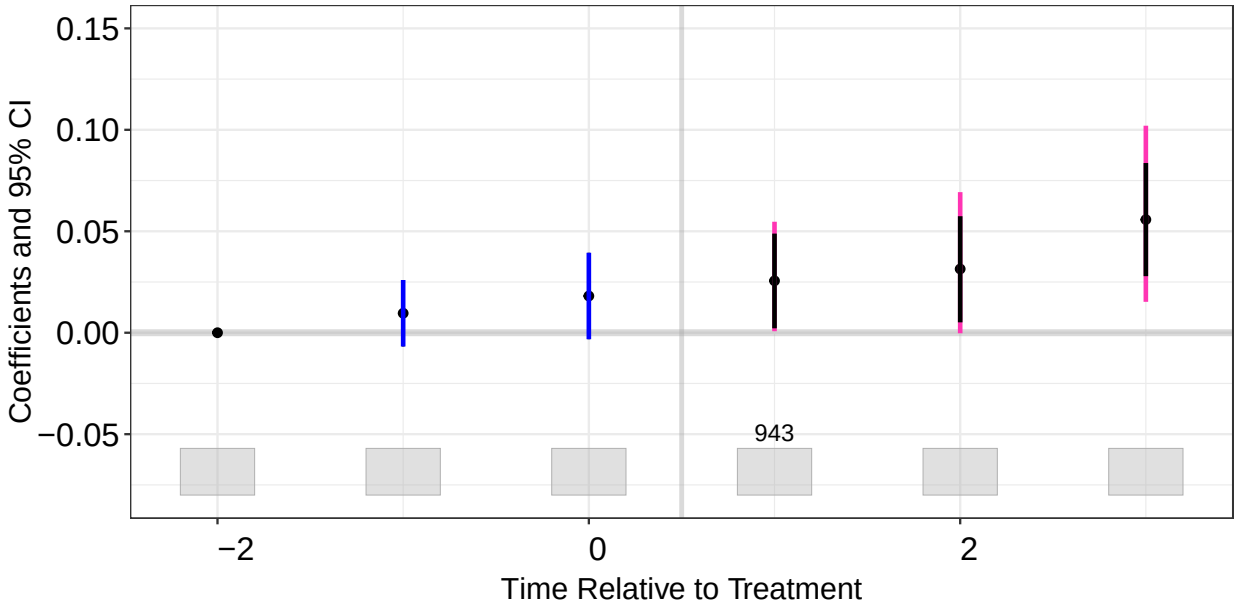
We find little evidence of violations of the parallel trend (PT) assumption. The equivalence test also reject the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After excluding two pre-treatment periods, the estimated ATT derived from the FEct estimator remains significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). Specifically, we utilize the two placebo periods to establish the baseline deviation from the PT assumption observed during the pre-treatment periods. The breaking value for a significant effect is determined to be $\bar{M} \approx 0.8$, allowing us to reject the null hypothesis of no treatment effects, provided that PT deviations in the post-treatment periods are no greater than 0.8 times the maximum deviation observed between consecutive periods in the placebo periods.

We further evaluate the robustness of the estimated DTE under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, accounting for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust when using HTE-robust estimators. We find little evidence of violations of the PT assumption. The estimated ATT appears to be robust to mild PT violations in sensitivity analysis with the breakdown threshold $M = 0.8$.

Kuipers and Sahn (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 7
Sensitivity Analysis 7
Summary 8

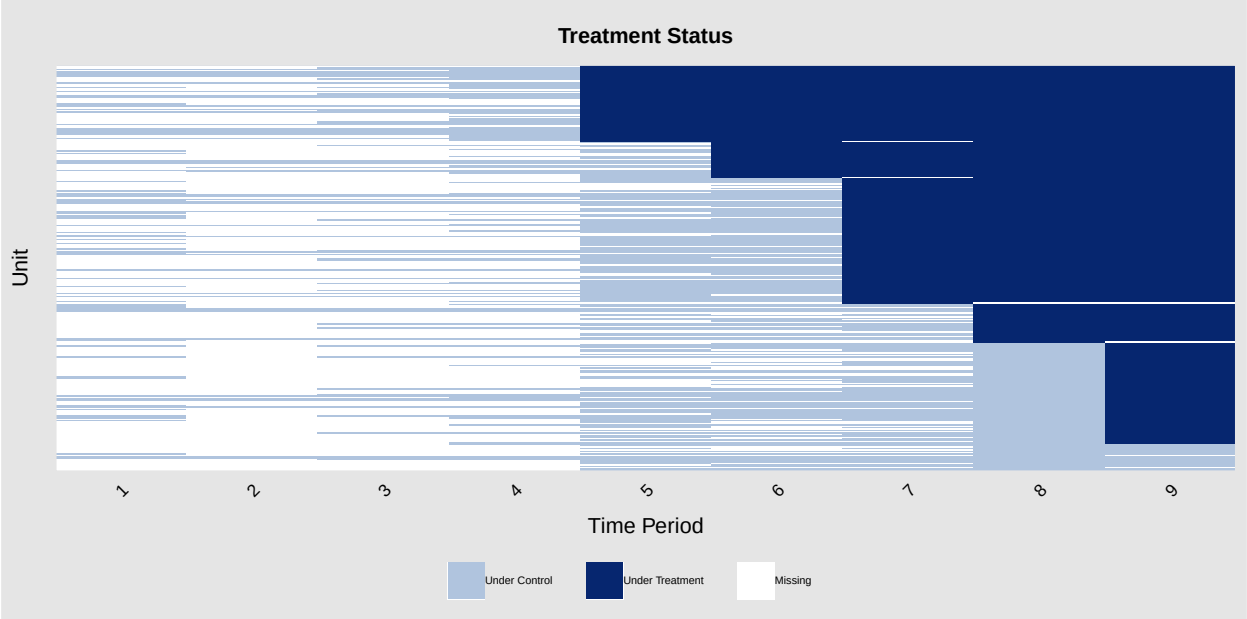
A First Look at Data

The paper investigates the effects of civil service reforms on share of municipal jobs held by different ethnic groups, using US municipality-year level panel data between 1850 and 1940. One of the main findings of this paper is that " foreign-born whites see their representation in blue-collar work significantly increase following the introduction of civil service reforms." (p209)

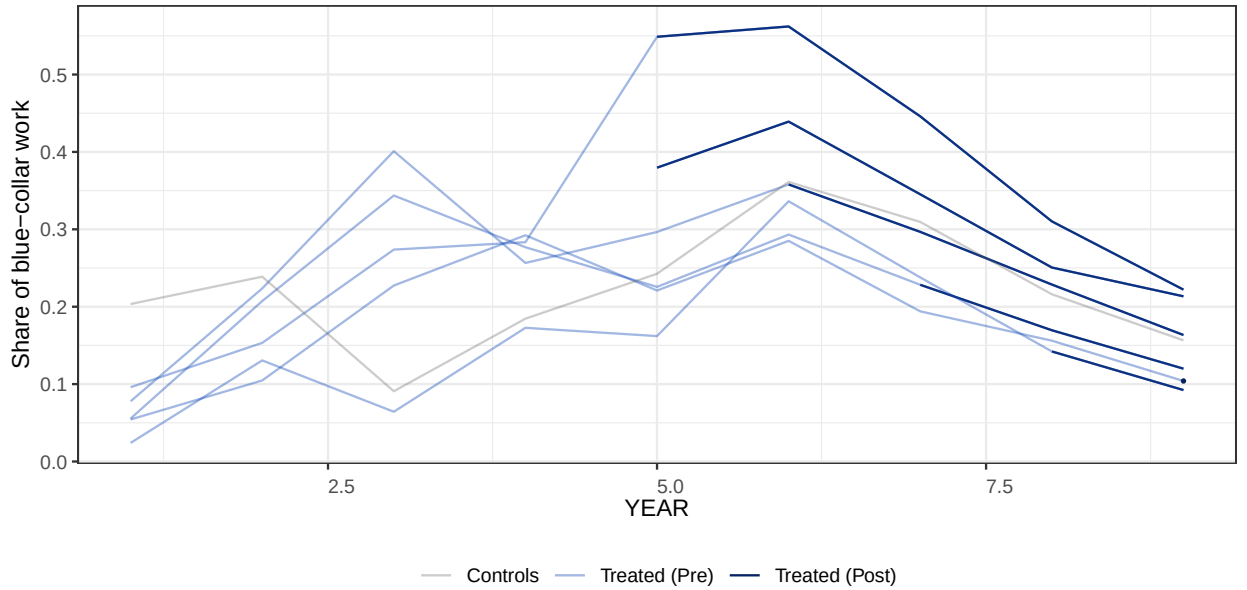
Model. We focus on **Model on White Foreign Born of Figure 3** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the city level.

Replication Summary	
Unit of analysis	Municipality × year
Treatment	Civil service reform
Outcome	Share of blue-collar work held by foreign-born whites
Treatment type	Staggered
Outcome type	Continous
Fixed effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a municipality). We see that a large majority of units are treated eventually, and there are no treatment reversals.



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

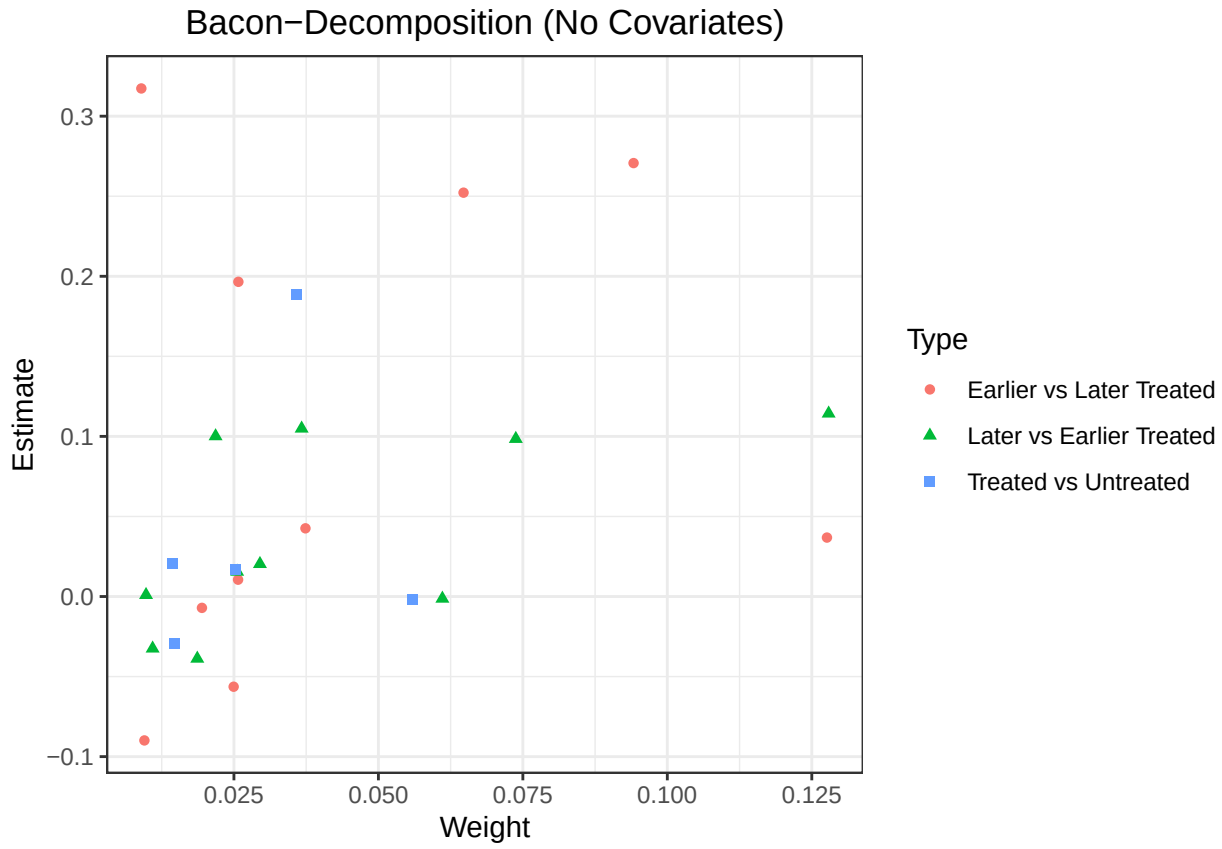
We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DIDM estimator and the imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

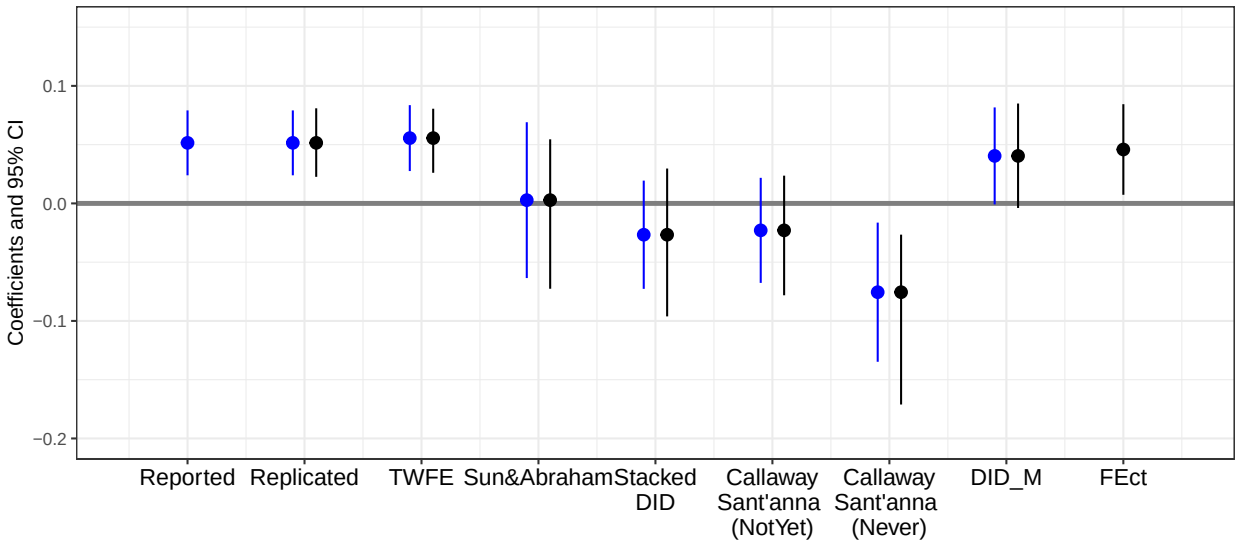
```
sol <- feols(govt_prop ~ treat + pop_prop|city+YEAR,
             data = df,cluster = "city")
summary(sol)
```

```
## OLS estimation, Dep. Var.: govt_prop
## Observations: 1,604
## Fixed-effects: city: 294,  YEAR: 9
## Standard-errors: Clustered (city)
##          Estimate Std. Error  t value  Pr(>|t|)
## treat    0.051525   0.014092  3.65633 0.0003032 ***
## pop_prop 1.250796   0.094563 13.22710 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.124832     Adj. R2: 0.594435
##                  Within R2: 0.168943
```

Goodman-Bacon Decomposition



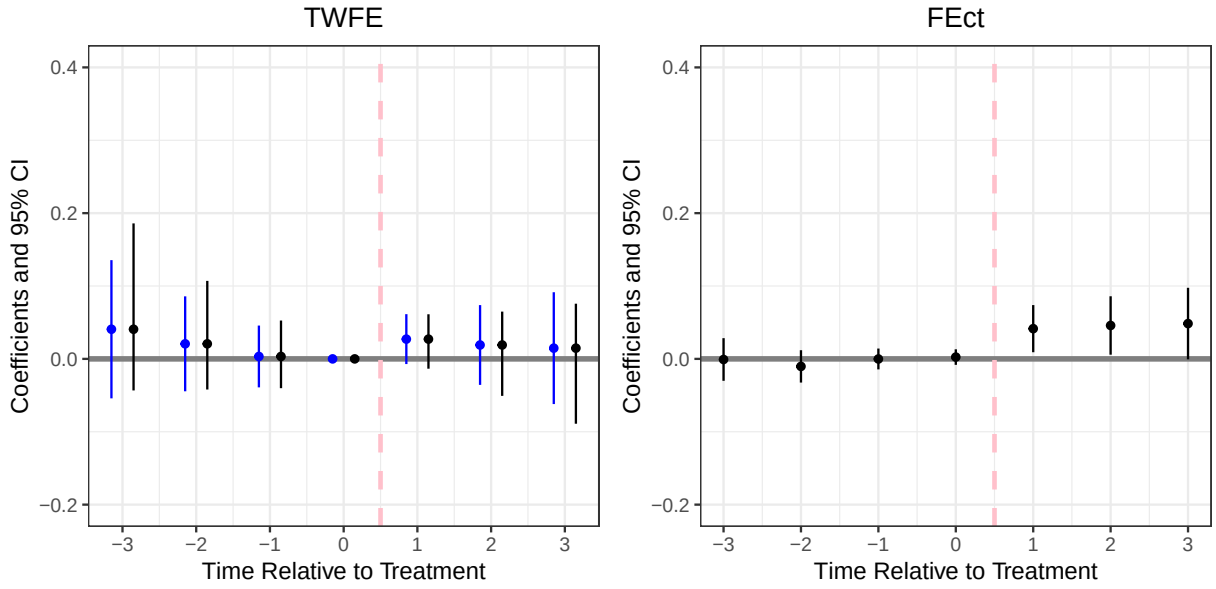

```
##   |                                                                      |
```

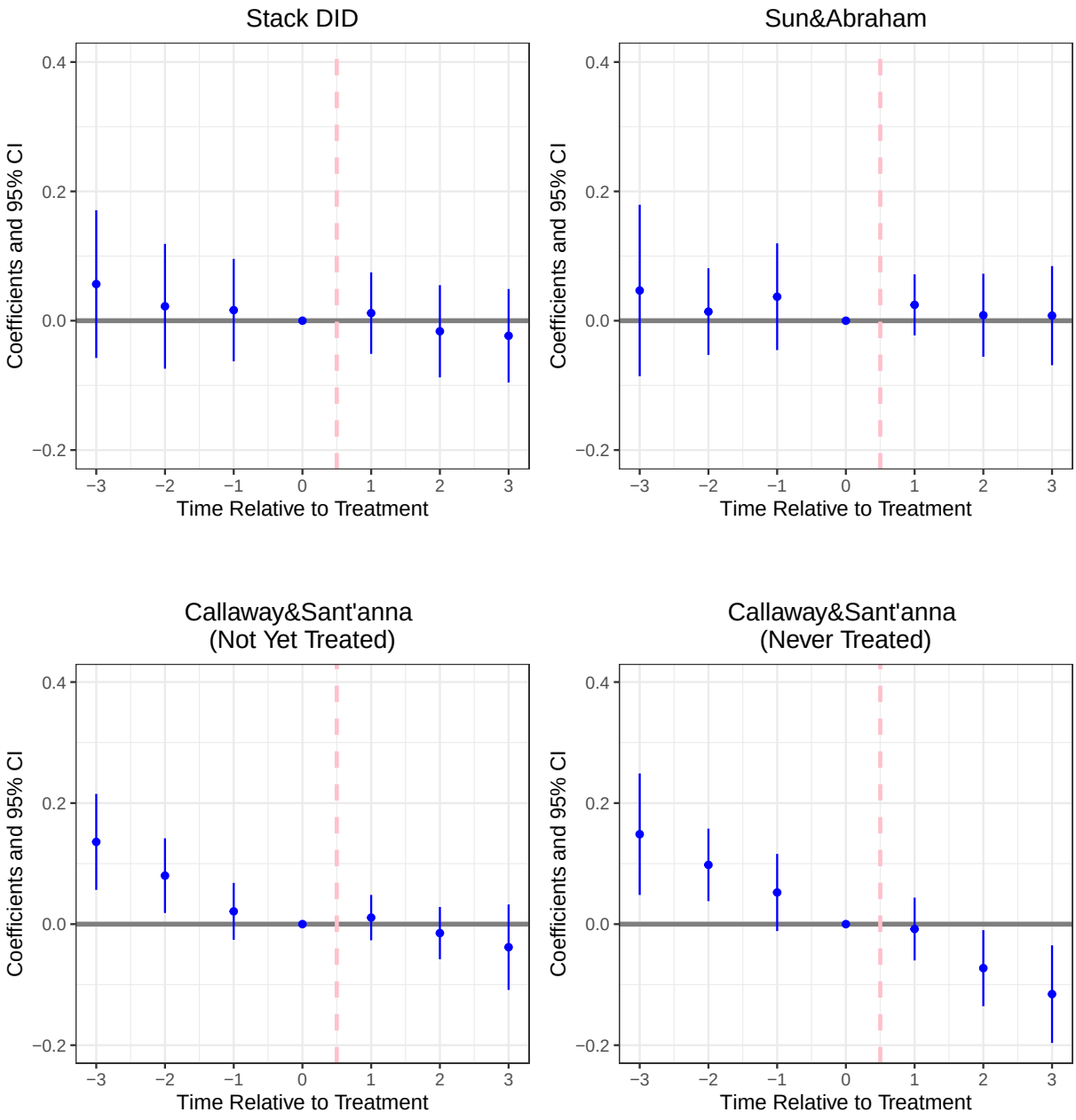


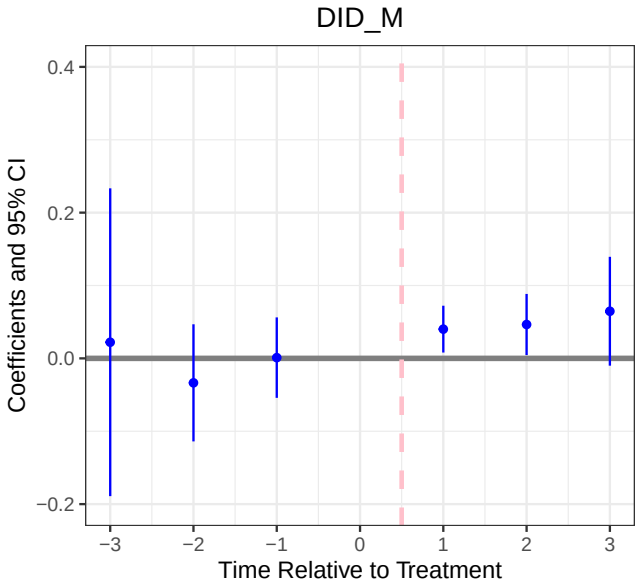
Various HTE-robust estimators yield qualitatively different results: some align with TWFE, while others do not.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DIDM and FEct. The results are shown below.



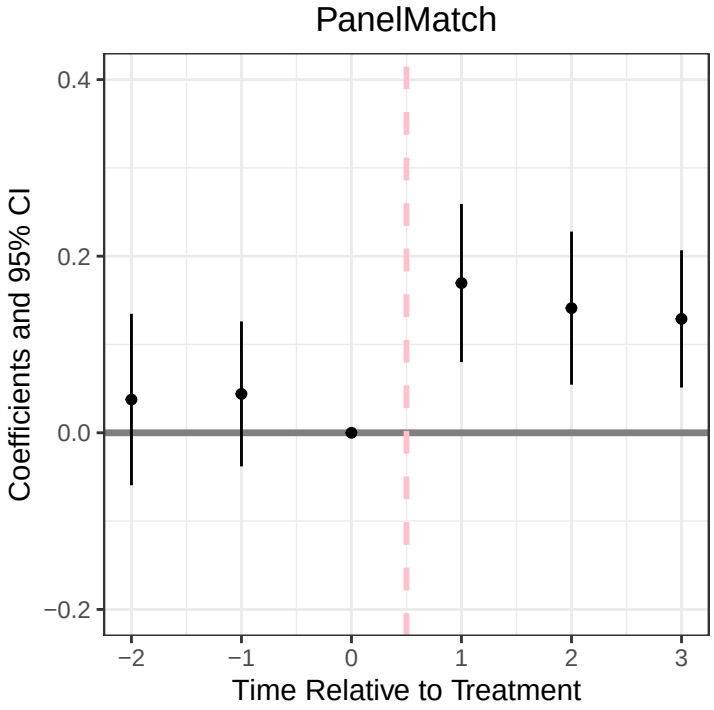

Stack DID
Sun&Abraham
Callaway&Sant'anna
(Not Yet Treated)
Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
Time Relative to Treatment
0.4
0.2
0.0
−0.2
−3
−2
−1
0
1
2
3



Various HTE-robust estimators yield very different results. Some show clear violation to the parallel trend (PT) assumptions while others don't.

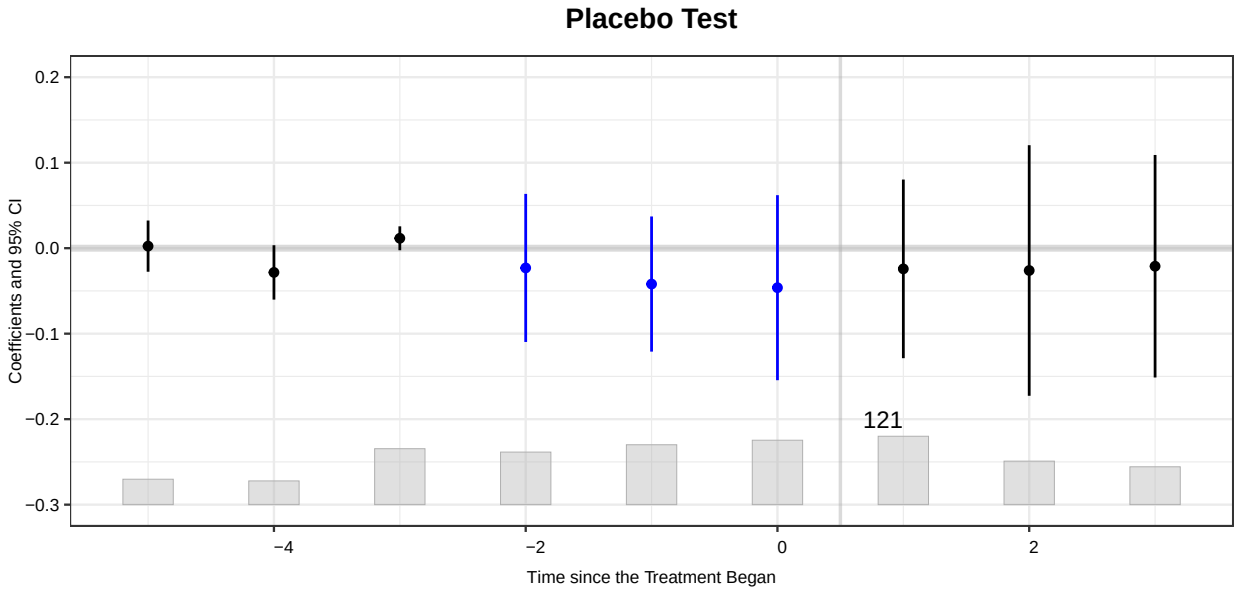
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 3 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.



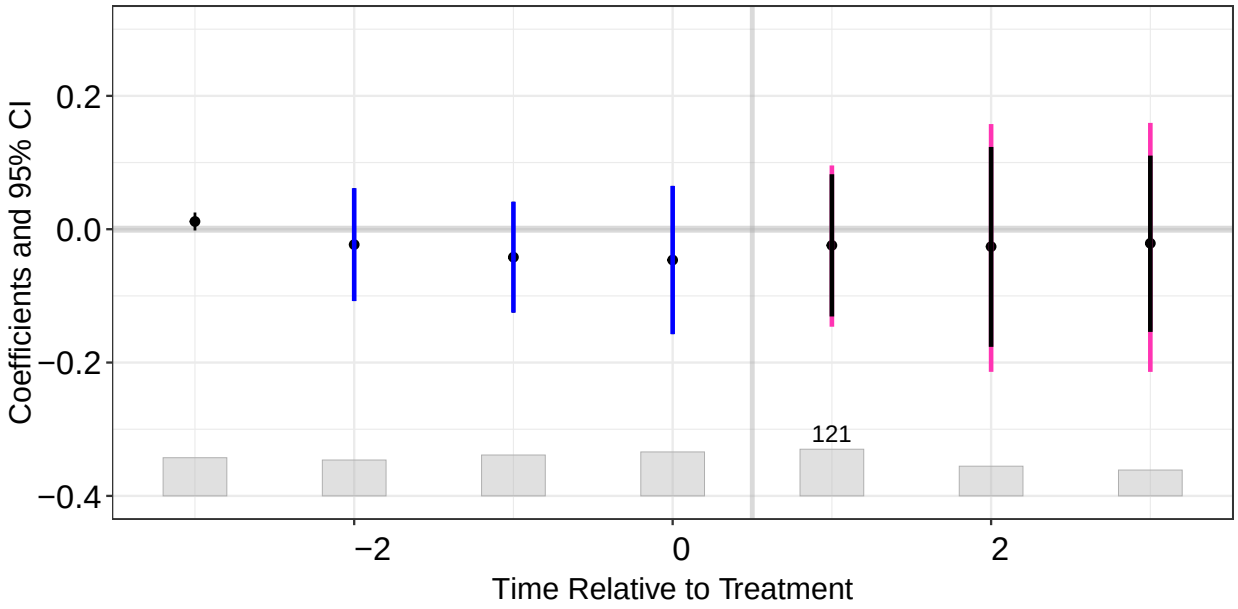
Test Statistics

```
##                                  p-value
## F test                             0.912
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.001
## Placebo test                       0.379
## Carryover effect test                 NA
```

Based on the results given by FEct, we find little evidence of violations of the PT assumption and anticipation effects. The equivalence test also rejects the null hypothesis that the residuals in the pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is no longer significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Various HTE-robust estimators yield qualitatively different results: some align with TWFE, while others do not. This discrepancy is likely due to the PT assumption failing across certain cohorts and time periods, leading to differences in CATT estimates. Moreover, differences in the choice of reference points among these estimators likely contribute to the discrepancies (see Section A.4.5 in the SM for more details). Neither the ATT nor DTE estimates appear robust to potential PT violations in the sensitivity analysis.

Latura and Weeks (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 7
Sensitivity Analysis 7
Summary 8

A First Look at Data

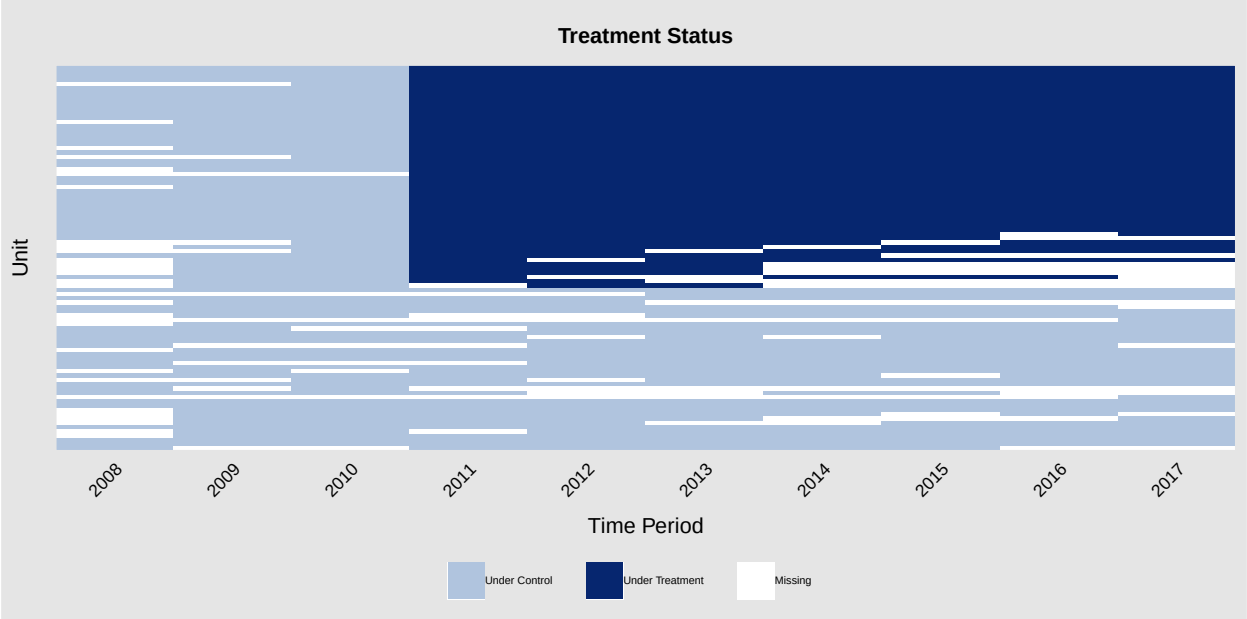
The paper investigates the effects of quota law on company attention to gender equality, using Italian company-year level panel data between 2007 and 2017. One of the main findings of this paper is that the quota law " is associated with a 0.03% increase in attention to gender equality." (p614)

Model. We focus on **Model 1 of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the account level.

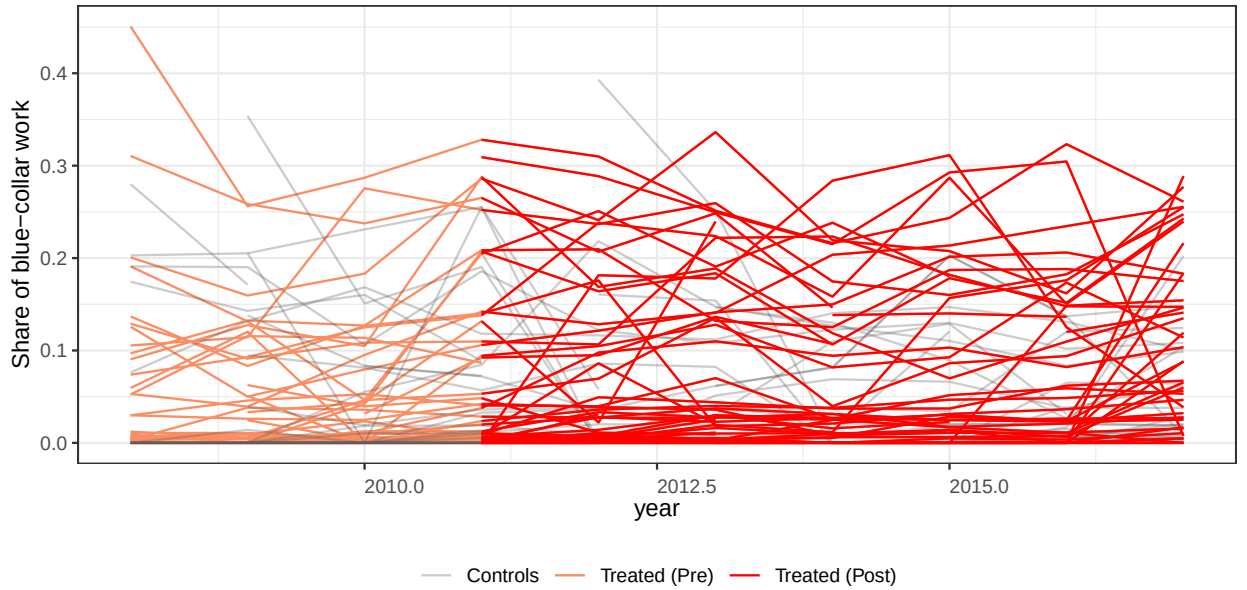
Table 1: Replication Summary

Unit of analysis	Company × year
Treatment	Quota law
Outcome	Company attention to gender equality
Treatment type	Classic
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a municipality). The treatment has no reversals.



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

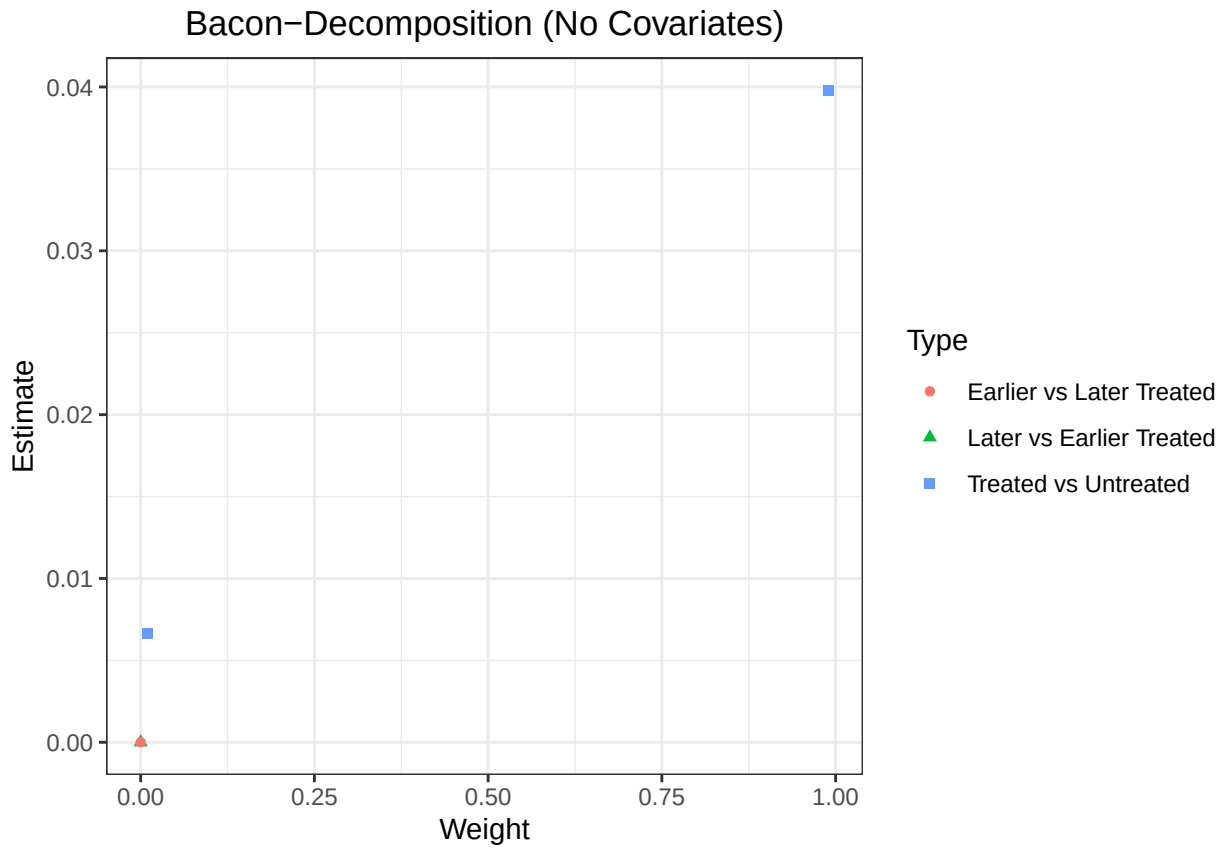
We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units and apply the TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DID_M estimator and the imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

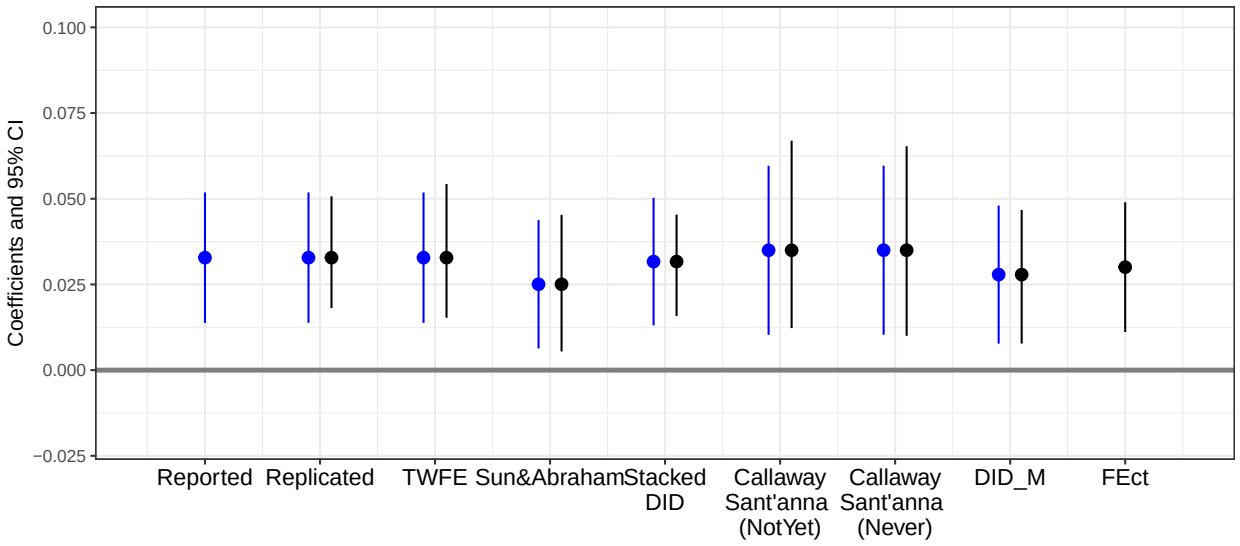
```
sol <- feols(prop_attn ~quota + sustain + pct_rev_change|companyid+year,
             data = df,cluster = "companyid")
summary(sol)
```

```
## OLS estimation, Dep. Var.: prop_attn
## Observations: 761
## Fixed-effects: companyid: 90,  year: 10
## Standard-errors: Clustered (companyid)
##                  Estimate Std. Error   t value   Pr(>|t|)
## quota            0.032810   0.009700  3.382428 1.0695e-03 **
## sustain          0.121966   0.014095  8.653392 1.9584e-13 ***
## pct_rev_change -0.000011   0.000022 -0.475337 6.3571e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.03736     Adj. R2: 0.770144
##                  Within R2: 0.365086
```

Goodman-Bacon Decomposition



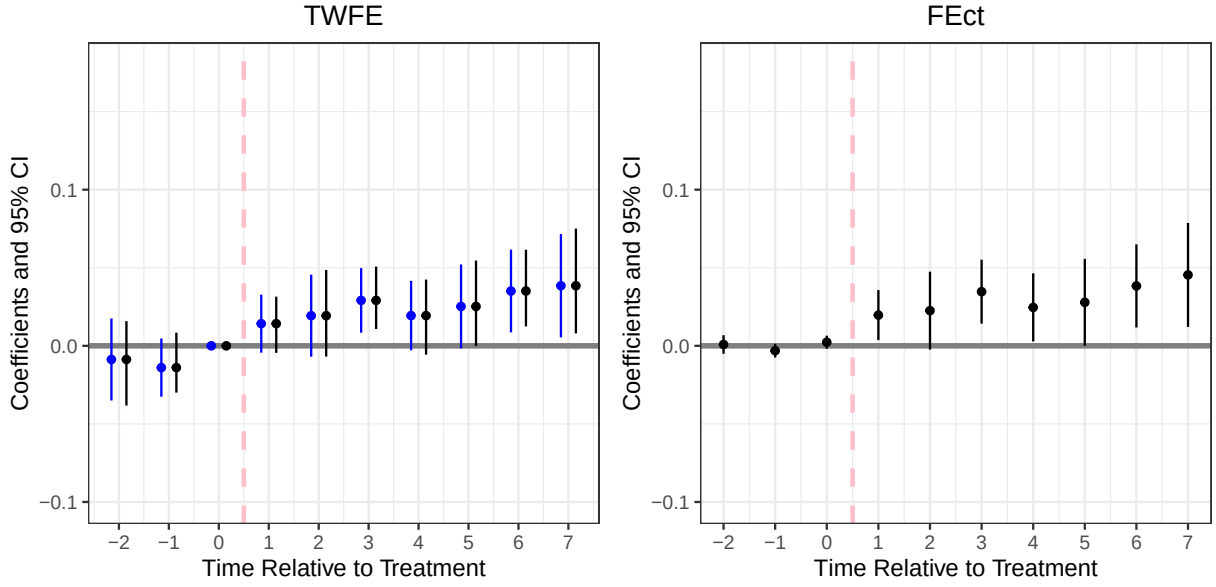

```
## [1] "baseline finished"
##   |                                                                      |
```

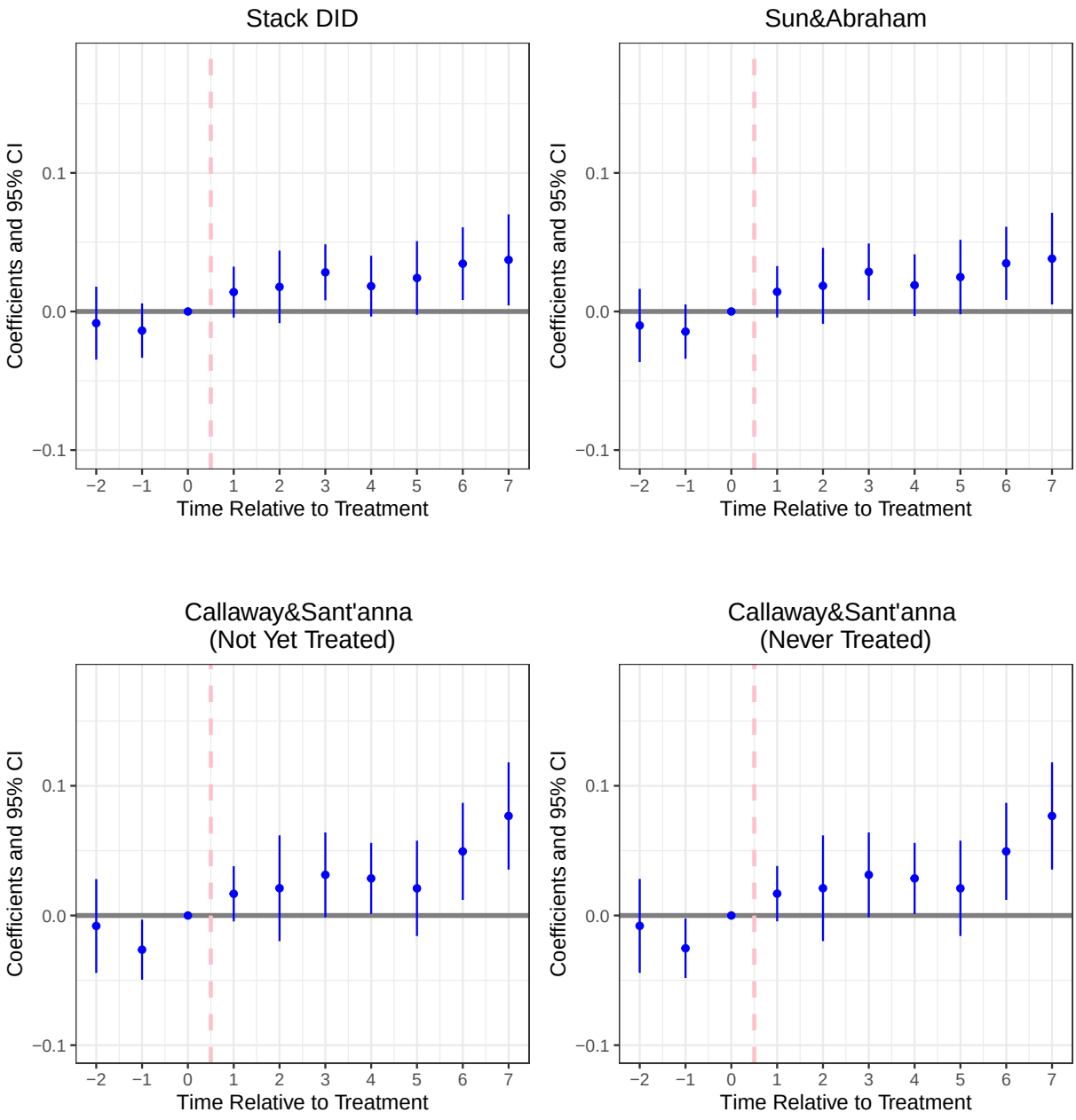


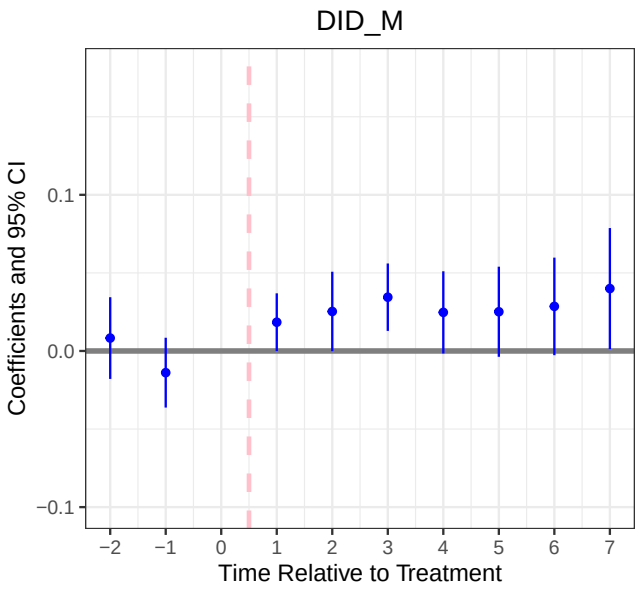
All HTE-robut estimators and TWFE give broadly consistent results.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DIDM and FEct. The results are shown below.



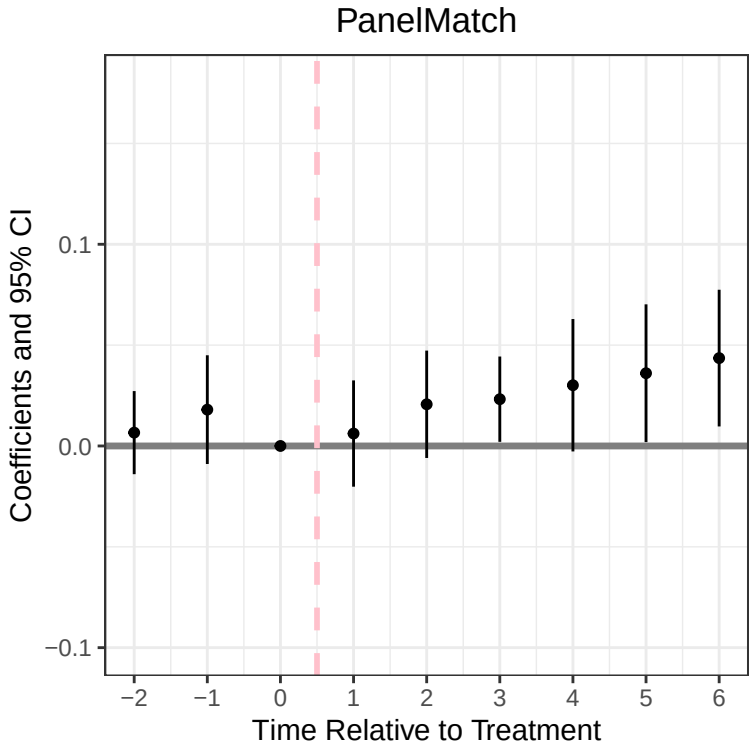

Stack DID
Sun&Abraham
Callaway&Sant'anna
(Not Yet Treated)
Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
0.1
0.0
−0.1
−2
−1
0
1
2
3
4
5
6
7
Time Relative to Treatment



All HTE-robust estimators and TWFE yield similar estimated DTE.

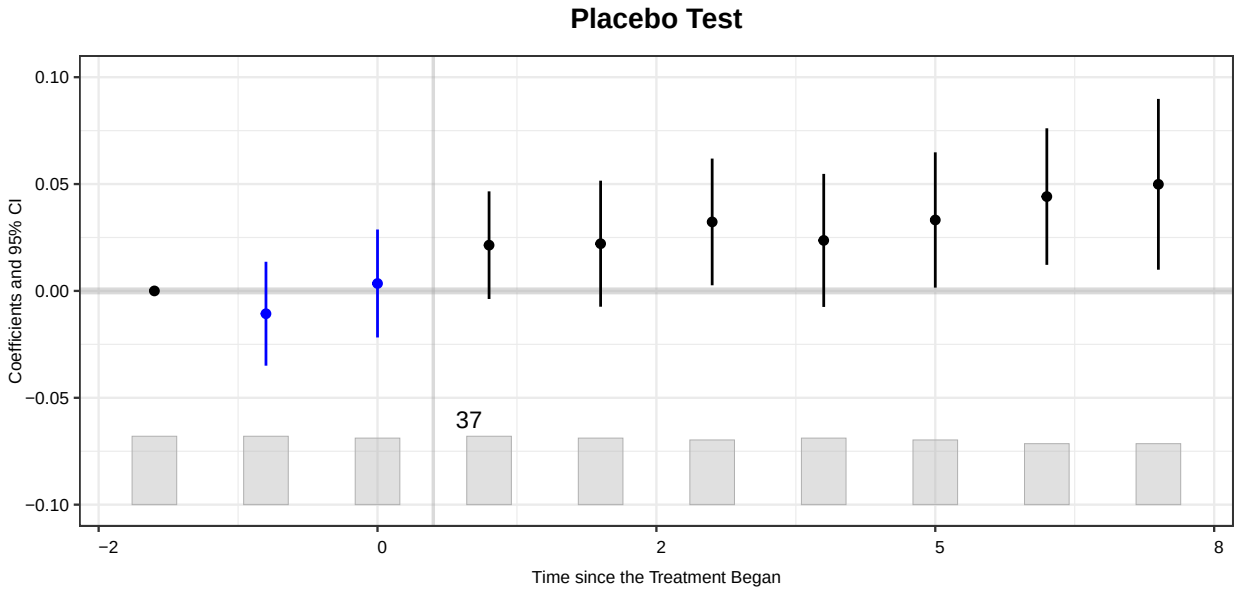
ATT for a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 6 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.



Test Statistics

```
## Cannot use full pre-treatment periods in the F test. The first period is removed.
```

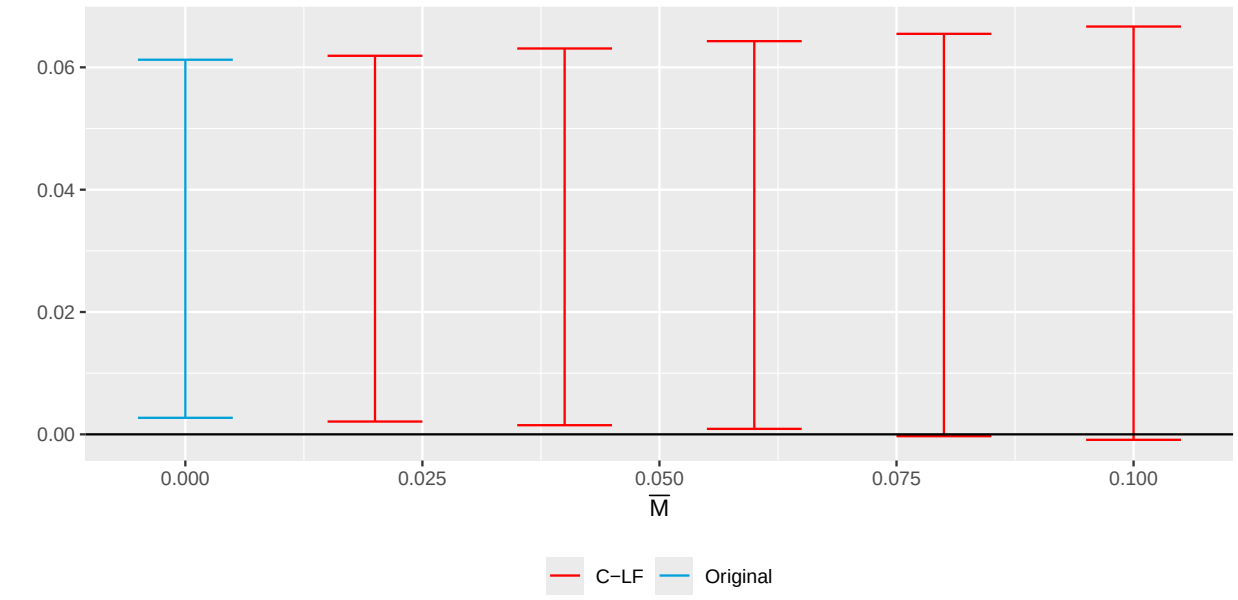
```
##                                 p-value
## F test                            0.345
## Equivalence test (default)        0.000
## Equivalence test (threshold=ATT)  0.000
## Placebo test                      0.744
## Carryover effect test                NA
```

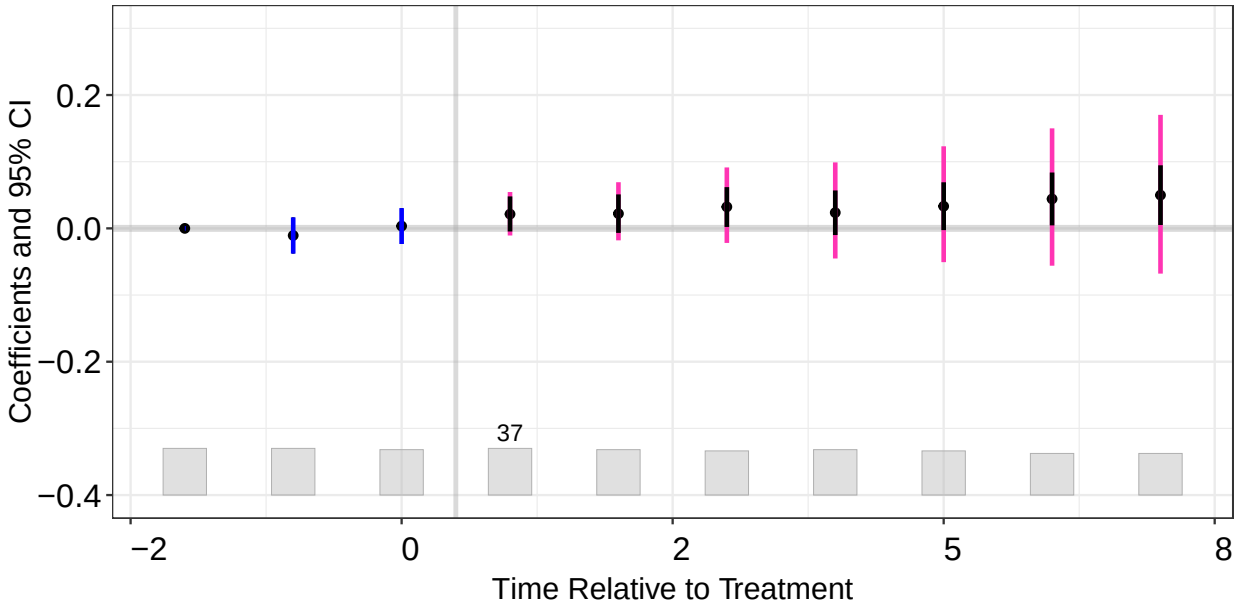
We find little evidence of violations of the parallel trend (PT) assumption.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.08$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 0.08 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model seems to be robust to applying HTE-robust estimators. However, due to limited power, the estimated ATT appears to be sensitive to mild PT violations with the breakdown threshold $M = 0.08$.

Liao (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

A First Look at Data

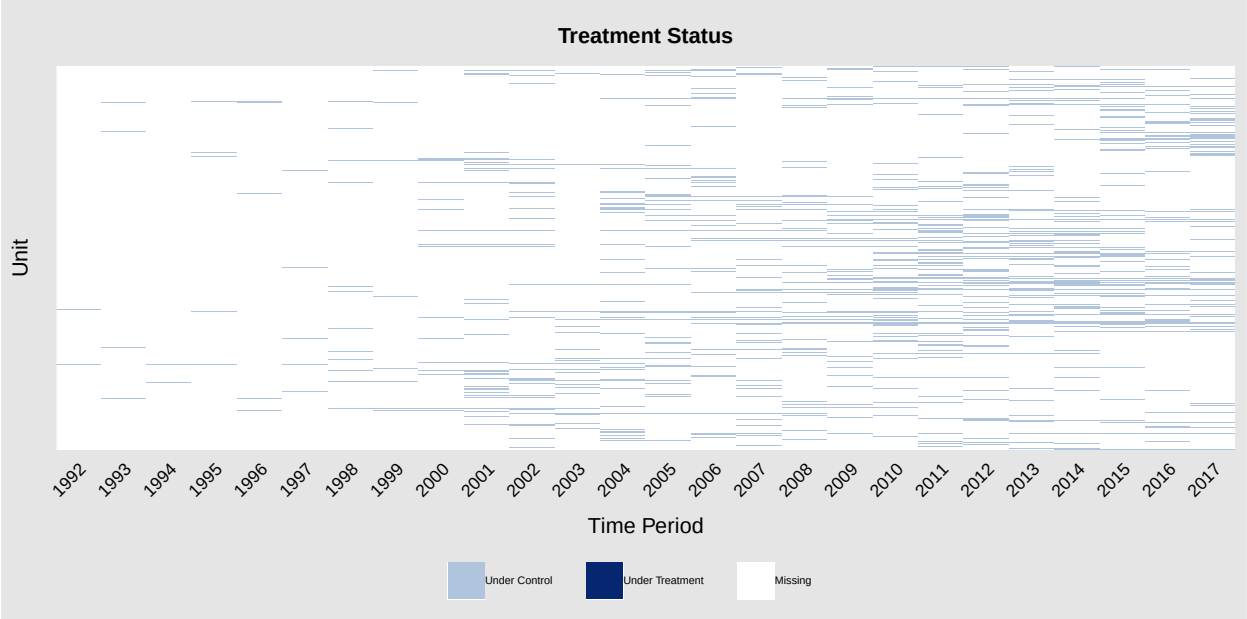
The paper investigates the effects of lobbying on H-1B denial rate, using US firm-year level panel data between 1992 and 2017. One of the main findings of this paper is that immigration lobbying " reduced firms' H-1B denial rate by around 0.045." (p1425)

Model. We focus on **Model 1 of Figure 7** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the firm level.

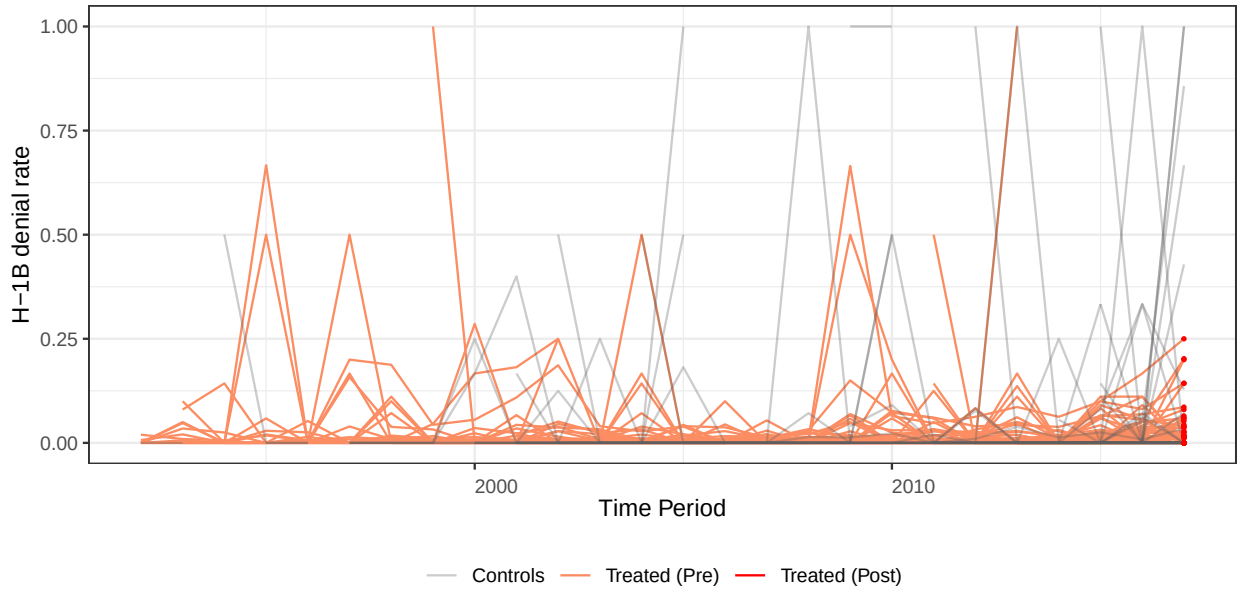
Table 1: Replication Summary

Unit of analysis	Firm × year
Treatment	Immigration lobbying
Outcome	H-1B denial rate
Treatment type	Classic
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a firm). The treatment has reversals.



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

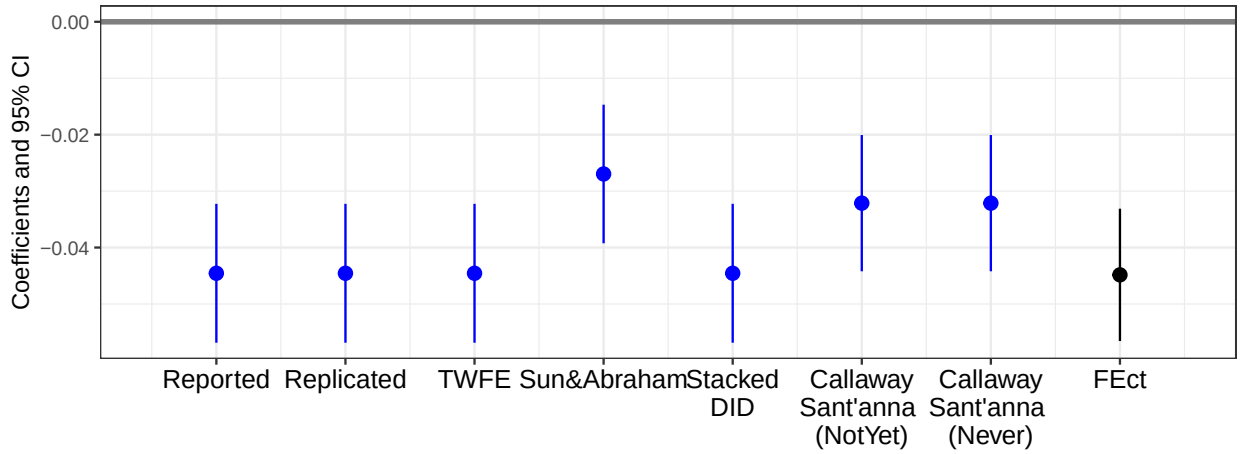
We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, the DID_M estimator and FEct to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(h1b_deny_rate ~ treatment|bvd_id+year,
             data = df,cluster = "bvd_id")
summary(sol)
```

```
## OLS estimation, Dep. Var.: h1b_deny_rate
## Observations: 981,096
## Fixed-effects: bvd_id: 384,462,  year: 26
## Standard-errors: Clustered (bvd_id)
##             Estimate Std. Error  t value   Pr(>|t|)
## treatment -0.044558   0.006277 -7.09803 1.2677e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.13718     Adj. R2: 0.437928
##                 Within R2: 7.315e-6
```

```
##   |                                                                      |
```

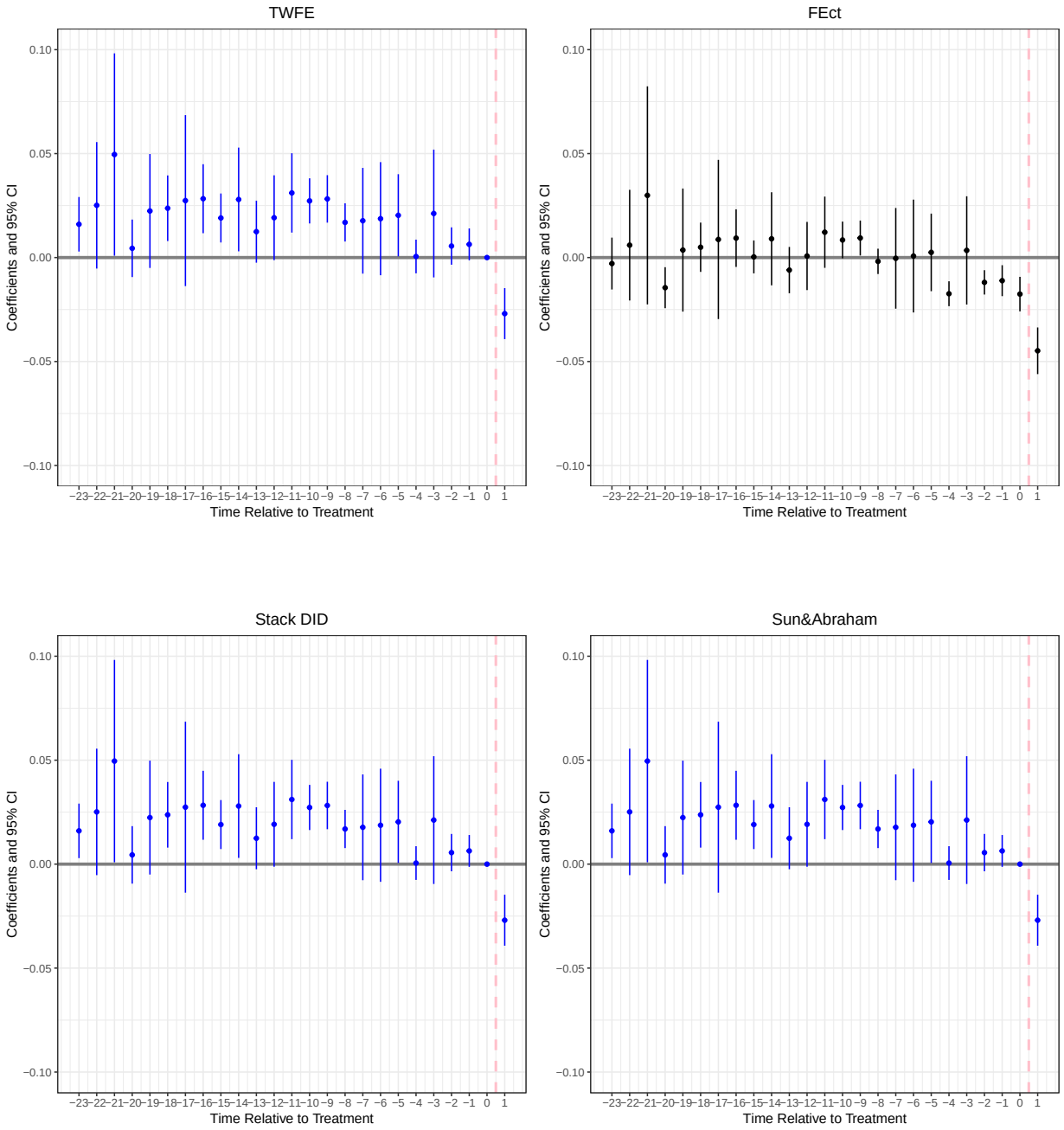


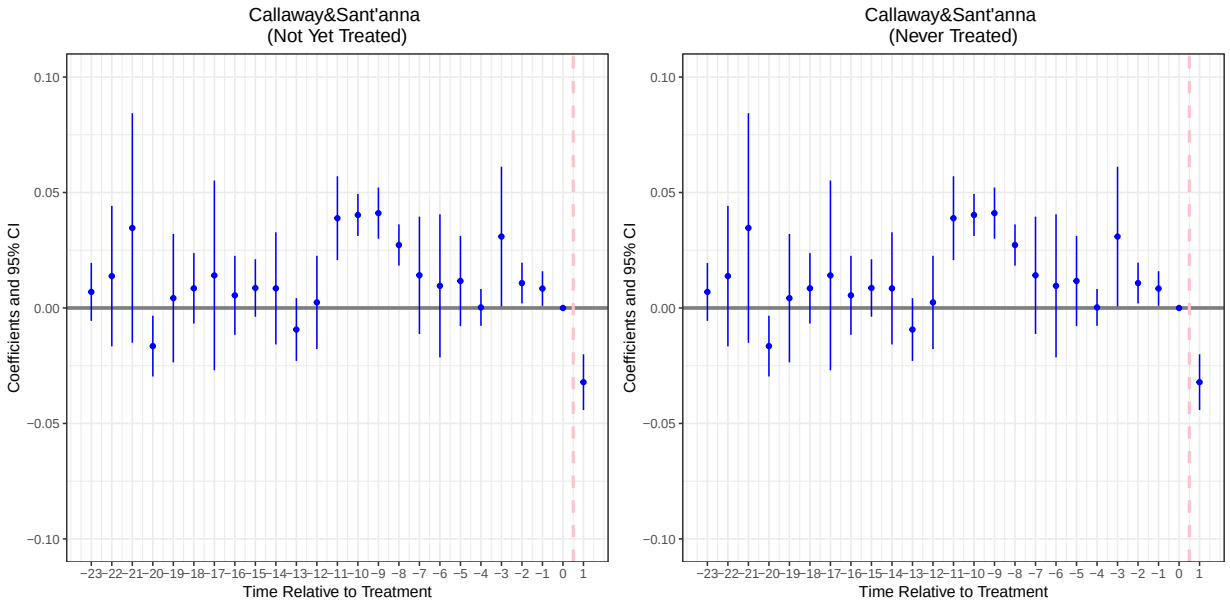
All HTE-robut estimators and TWFE give broadly consistent results.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use five estimators, TWFE, IW, CS, Stacked DID, and FEct. The results are shown below.

```
##   |                                                                      |
```


TWFE
FEct
Stack DID
Sun&Abraham
Coefficients and 95% CI
Time Relative to Treatment
0.10
0.05
0.00
−0.05
−0.10
−23 −22 −21 −20 −19 −18 −17 −16 −15 −14 −13 −12 −11 −10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1

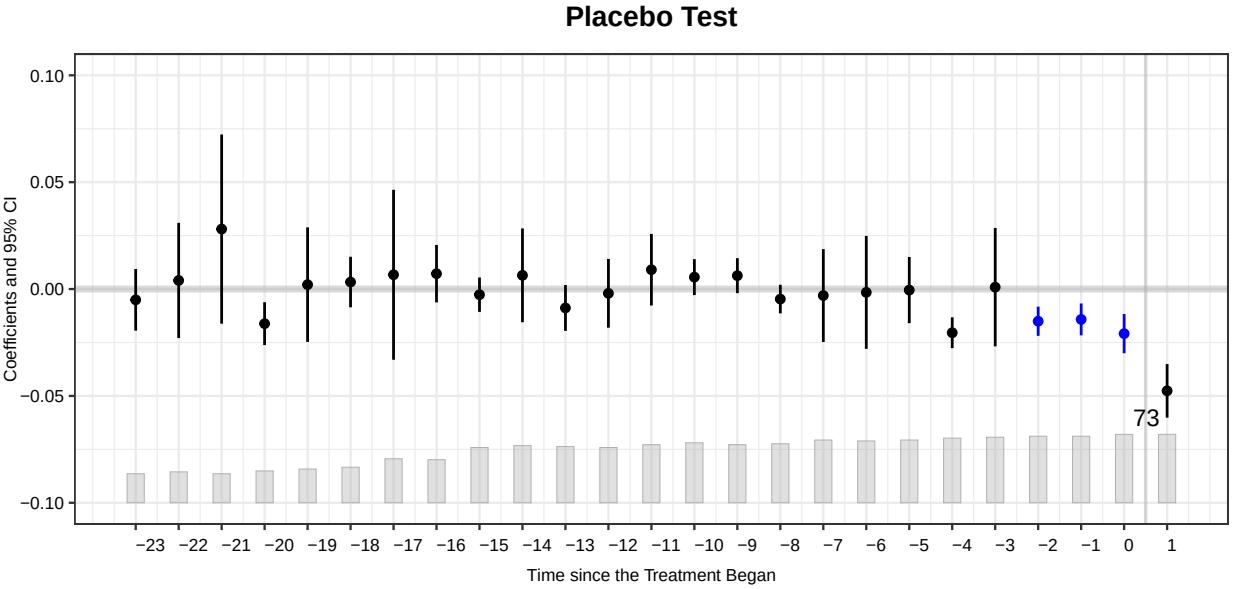


All HTE-robut estimators and TWFE give broadly consistent DTE.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.

```
##   |                                                                      |
```



Test Statistics

```
##                                    p-value
## F test                               0.000
## Equivalence test (default)           0.003
## Equivalence test (threshold=ATT)     0.032
## Placebo test                         0.000
## Carryover effect test                   NA
```

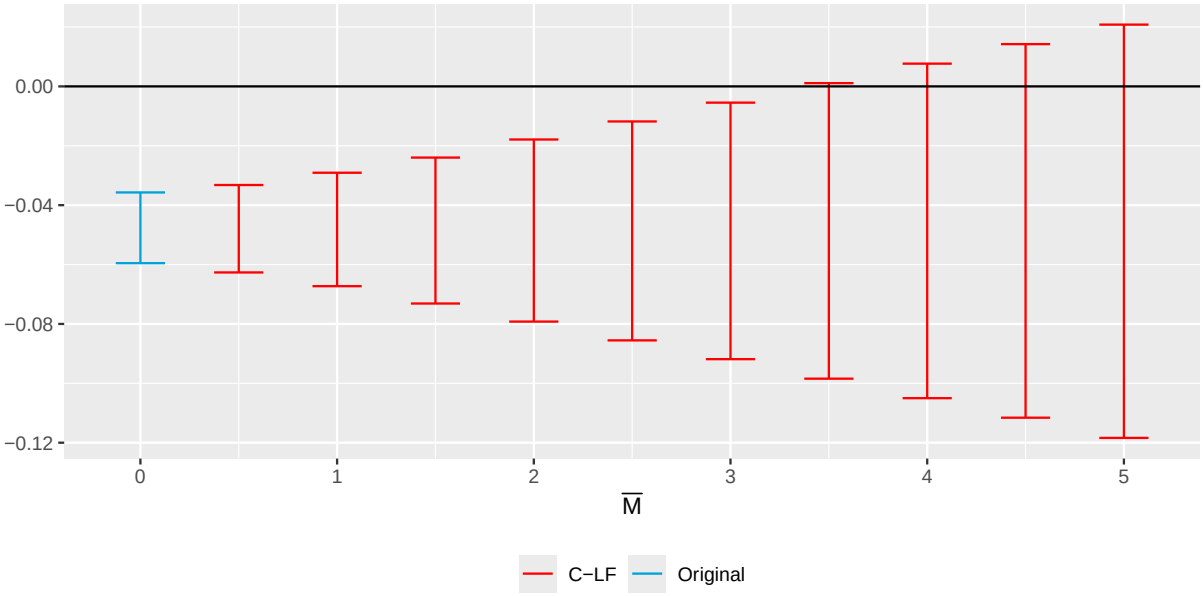
The rejection in the F-test and the placebo test casts some doubt on the parallel trend (PT) assumption but the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT, suggesting potential confounding likely has very limited impact on the causal estimates.

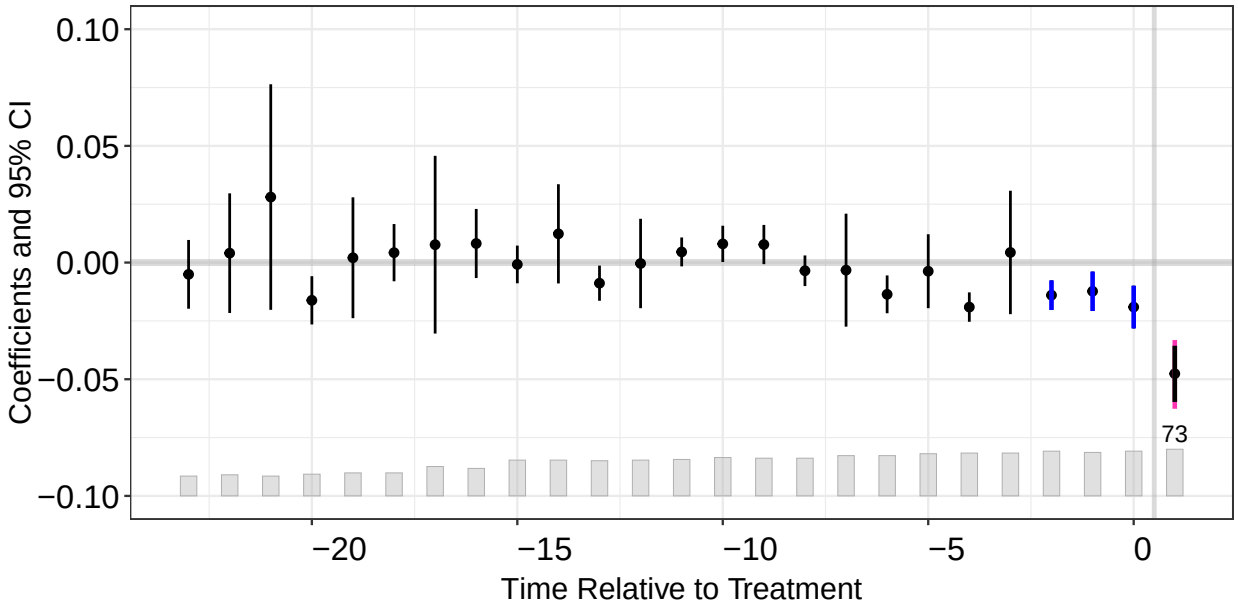
Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 3.5$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 3.5 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

```
##   |                                                                      |
```





Summary

Overall, the main result of the chosen model seems to be robust to HTE-robust estimators. There is some evidence for potential violations of the PT (F test) , but they seem inconsequential. The estimated ATT appears to be robust to mild PT violations with the breakdown threshold $M = 3.5$.

Magaloni, Franco-Vivanco, and Melo (2020)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

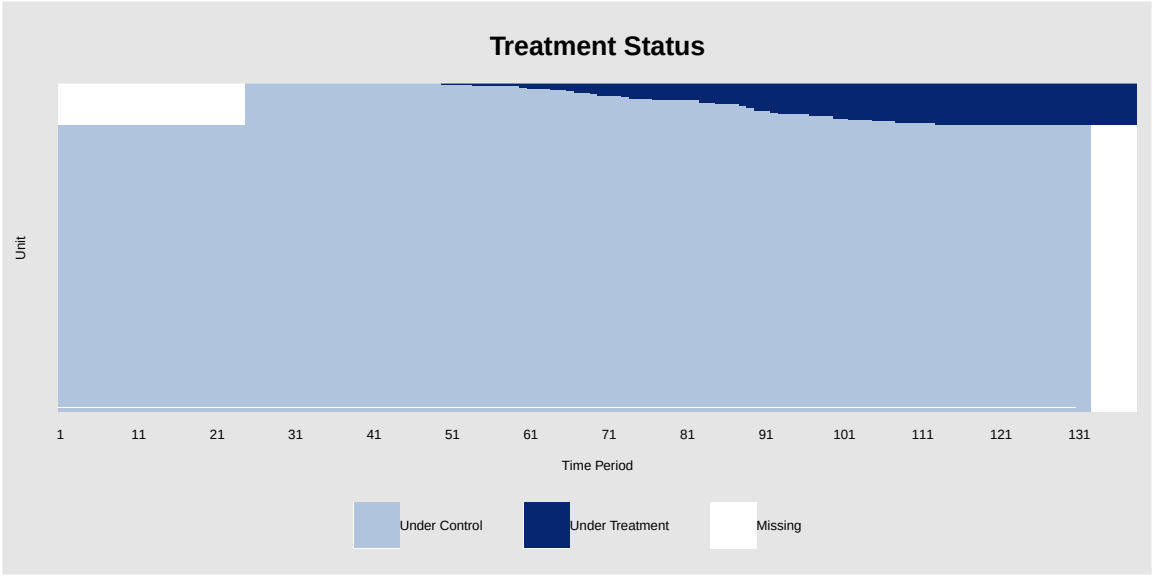
A First Look at Data

The paper investigates the effects of Rio de Janeiro's "Pacifying Police Units" (UPPs) program on police violence, using favela-month level panel data from Rio de Janeiro during 2008-2014. One of the main findings of this paper is that "the UPP had a substantial effect reducing killings by the police" (p568).

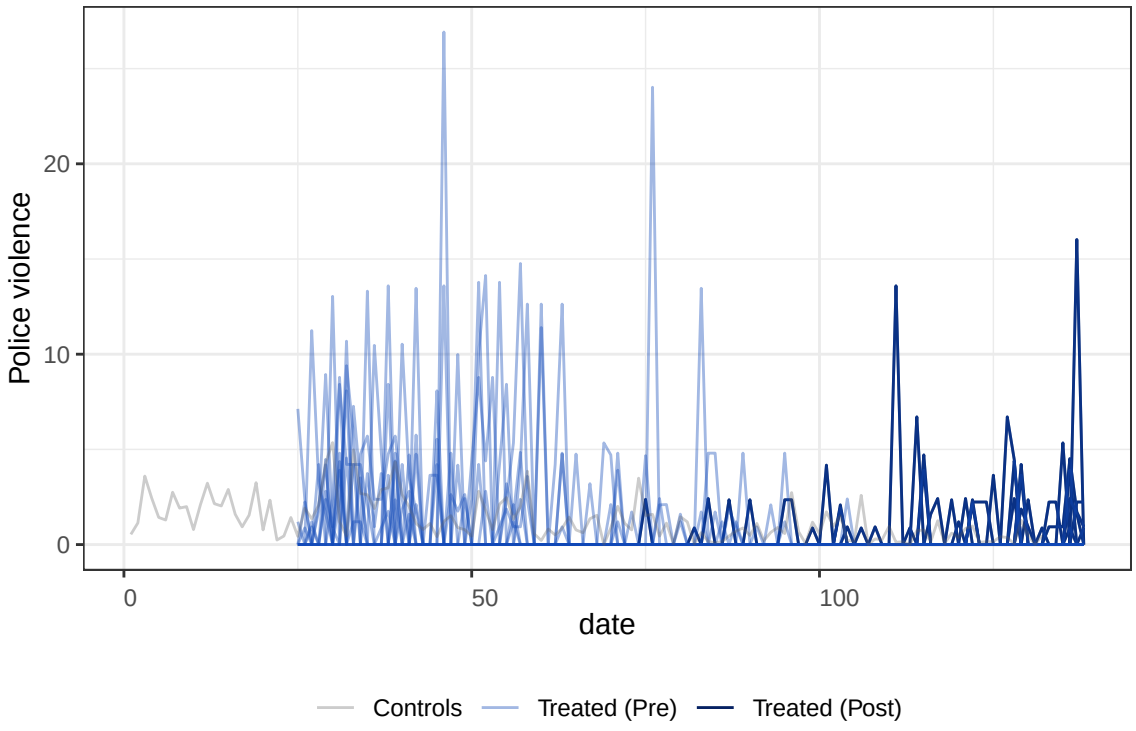
Model. We focus on **Model 3 of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model with unit-specific time trend (ULT) and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	Favela × year
Treatment	UPP
Outcome	City Police violence
Treatment type	Staggered
Outcome type	Continous
Fixed Effects	Unit+Time+Unit specific time trend

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a complexo).



Plotting the outcome variable. We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this data) and apply two estimators: TWFE (with ULT) and the inputation

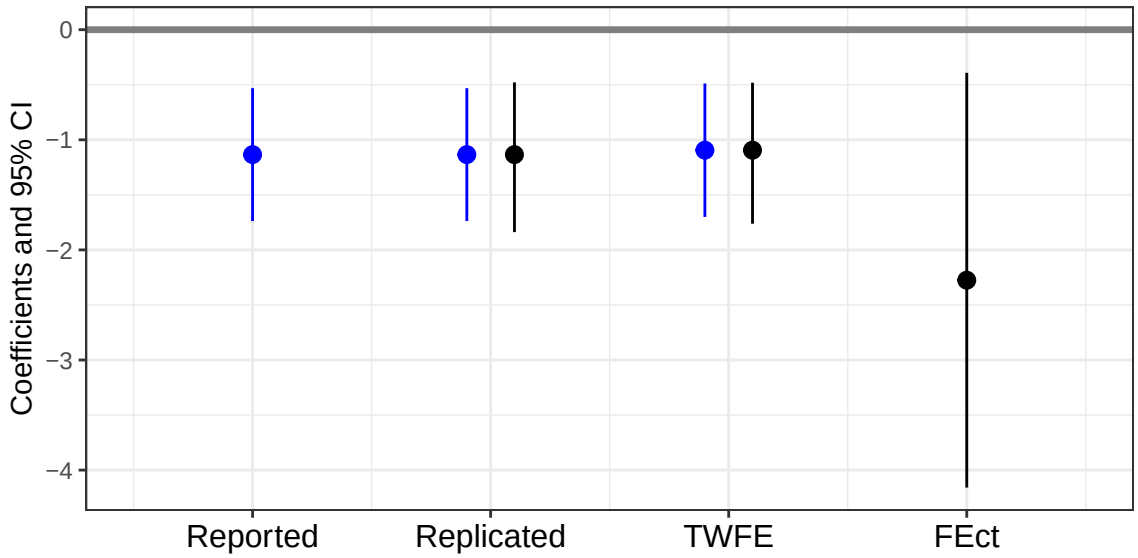
method (FEct). The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(autos_rate2~upp_post+bope|complexo_id+date+complexo_id[date],
             data = df,cluster = "complexo_id")
summary(sol)
```

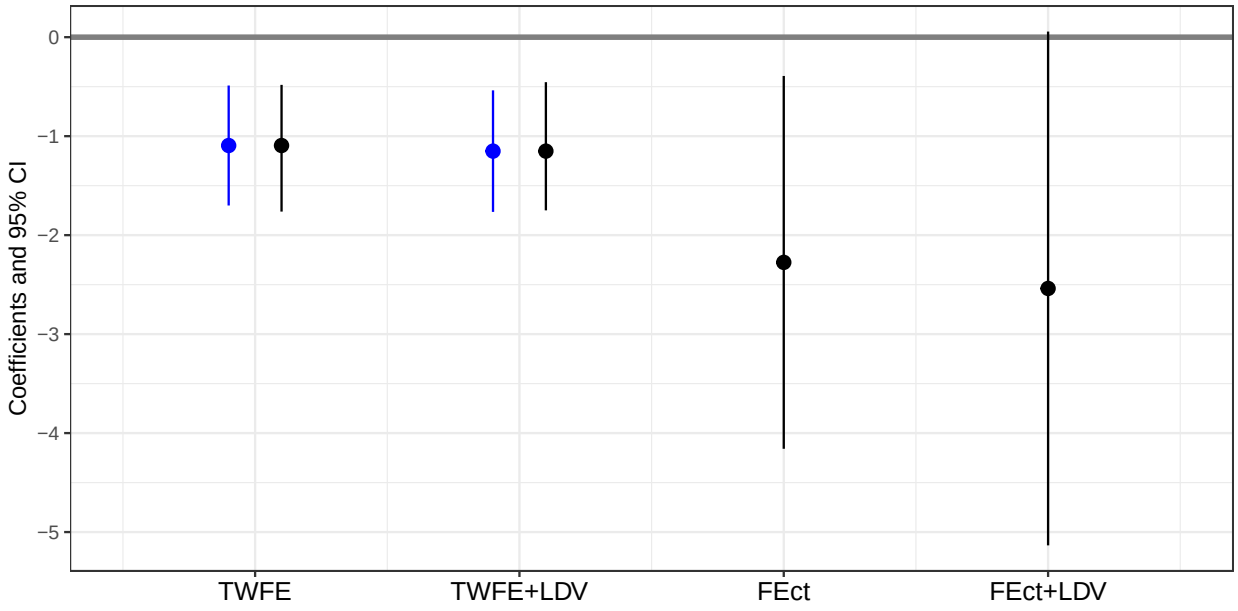
```
## OLS estimation, Dep. Var.: autos_rate2
## Observations: 36,956
## Fixed-effects: complexo_id: 286,  date: 138
## Varying slopes: date (complexo_id): 286
## Standard-errors: Clustered (complexo_id)
##           Estimate Std. Error  t value   Pr(>|t|)
## upp_post -1.134475   0.307845 -3.68521 0.00027346 ***
## bope     -0.523970   0.310239 -1.68892 0.09232797 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.8     Adj. R2: 0.027152
##              Within R2: 7.248e-5
```

```
##   |                                                                      |
```



Both TWFE and FEct estimates are negative and statistically significant, while the latter is larger in magnitude and has wider CIs.

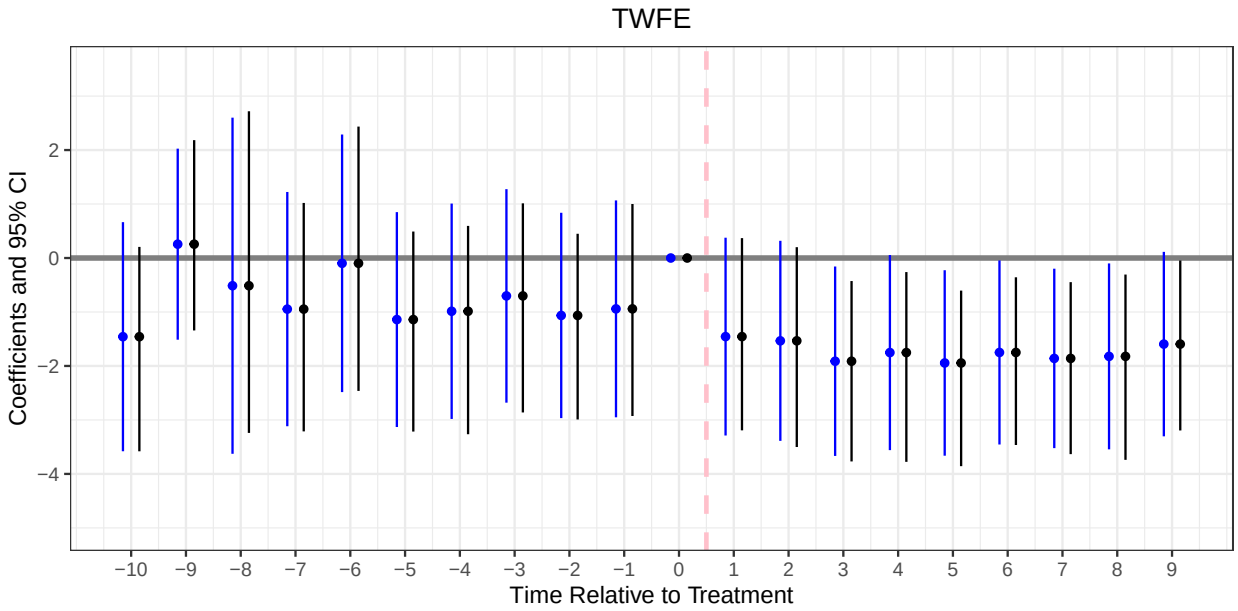
```
##   |                                                                      |
```

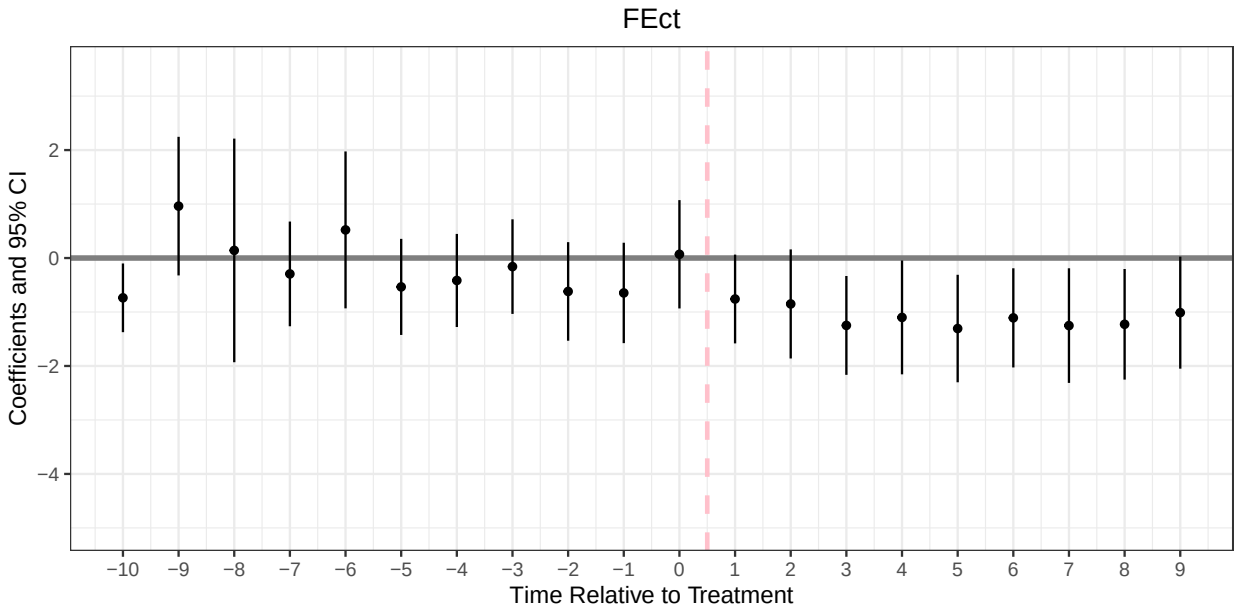


Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.

```
##   |                                                                      |
```



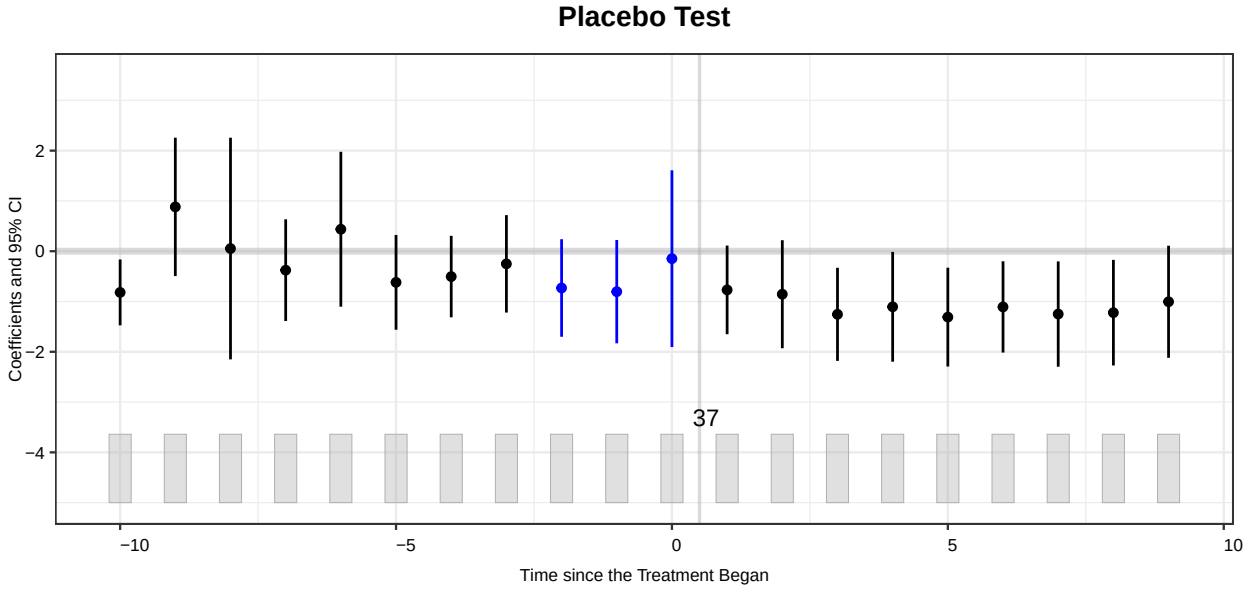


The estimated DTE given by TWFE and FEct have similar shapes and demonstrates negative values during post-treatment periods. The former have larger magnitudes.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.

```
##   |                                                                      |
```



Test Statistics

```
##                                    p-value
## F test                               0.591
## Equivalence test (default)           0.000
## Equivalence test (threshold=ATT)     0.022
## Placebo test                         0.186
## Carryover effect test                   NA
```

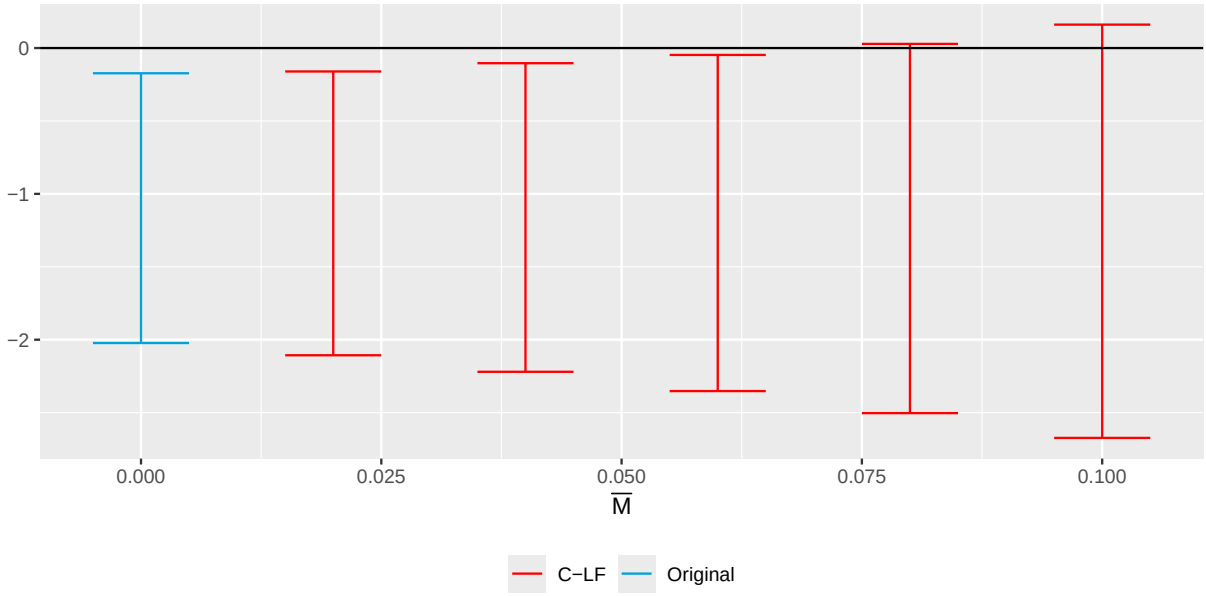
We find no evidence for violations of the parallel trends (PT) assumption using the F-test and the placebo test. The equivalence test also rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

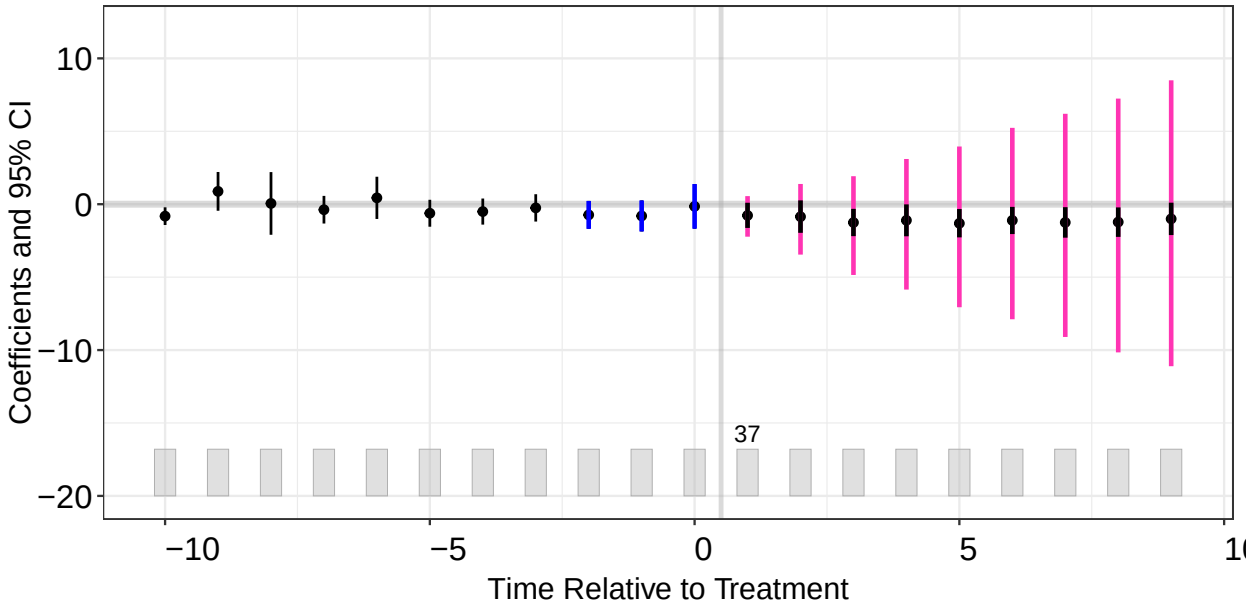
Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.07$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 0.07 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

```
##   |                                                                      |
```





Summary

Overall, the main result of the chosen model appears to be robust to FEct, an HTE-robust estimator, although the point estimate from FEct is much larger in magnitude. We find little evidence for potential violations of PT. However, due to limited power, the estimated ATT appears to be sensitive to mild PT violations with the breakdown threshold $M = 0.07$.

Marsh (2023)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 6
Summary 7

A First Look at Data

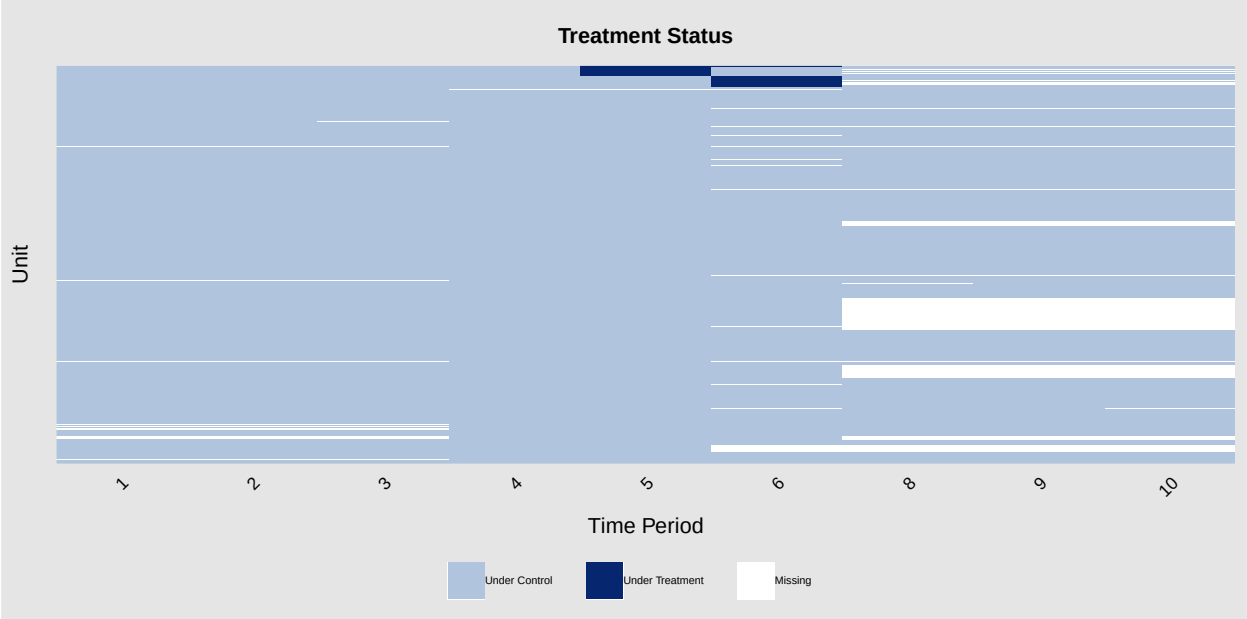
The paper investigates the effects of Black church arson attack on voter turnout, using US county-election year panel data between 1980 and 2016 One of the main findings of this paper is that experiencing Black church arson attack caused "a 3.7 percentage-point decrease in turnout for counties." (p1044)

Model. We focus on **Model 1 of Figure 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the county level.

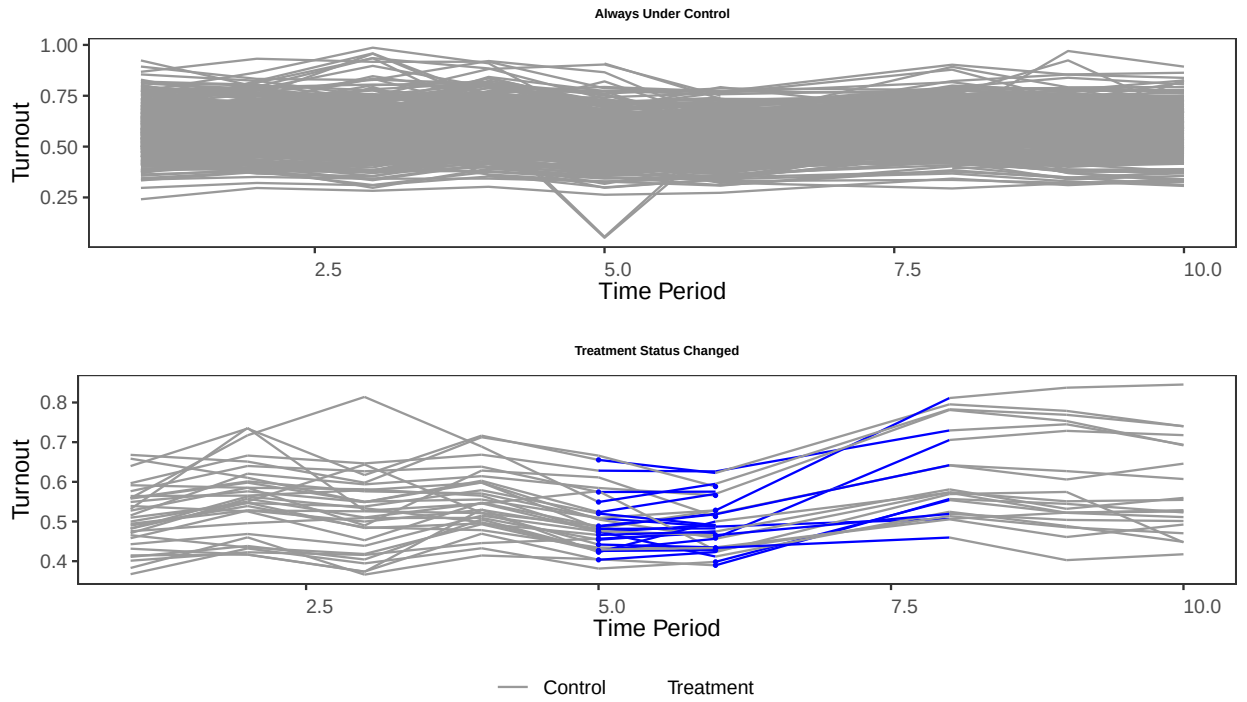
Table 1: Replication Summary

Unit of analysis	County × election year
Treatment	Black church arson attack
Outcome	Voter turnout
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (an election year) and each row represents a unit (a county). The treatment has reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

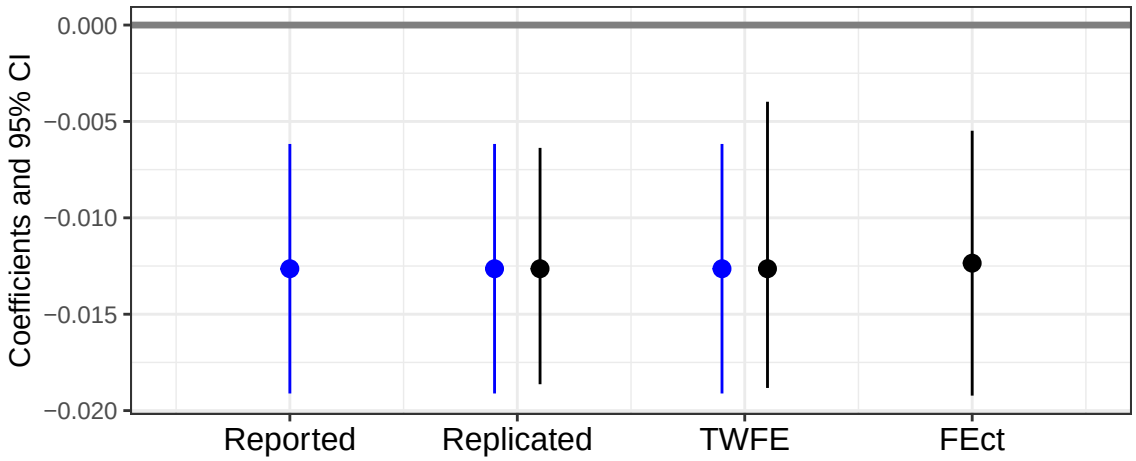
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates

and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(turnout ~ tr_ar +perc_bl+pop_tot+income|fips+year,
             data = df,cluster = "fips")
summary(sol)
```

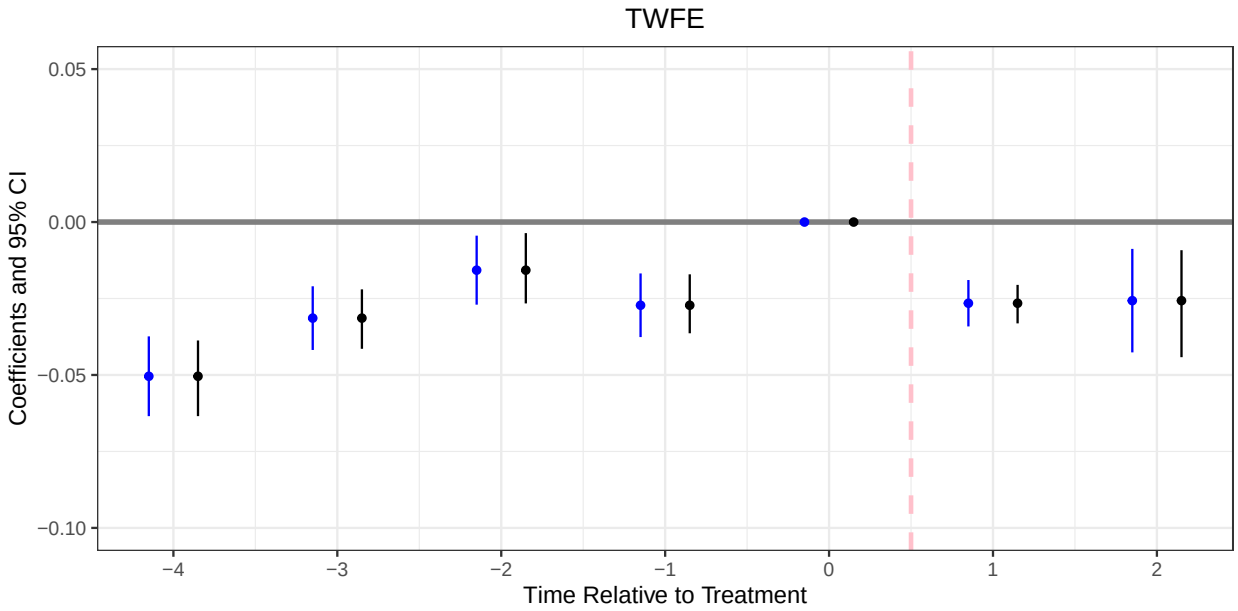
```
## OLS estimation, Dep. Var.: turnout
## Observations: 23,952
## Fixed-effects: fips: 2,889,  year: 9
## Standard-errors: Clustered (fips)
##              Estimate   Std. Error  t value   Pr(>|t|)
## tr_ar   -0.0126404814 0.0033019512 -3.82819 1.3183e-04 ***
## perc_bl -0.1092172971 0.0625119218 -1.74714 8.0719e-02 .
## pop_tot  0.0000000735 0.0000000242  3.04281 2.3649e-03 **
## income   0.0000014319 0.0000002265  6.32117 2.9972e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.045496     Adj. R2: 0.770435
##                  Within R2: 0.015326
```

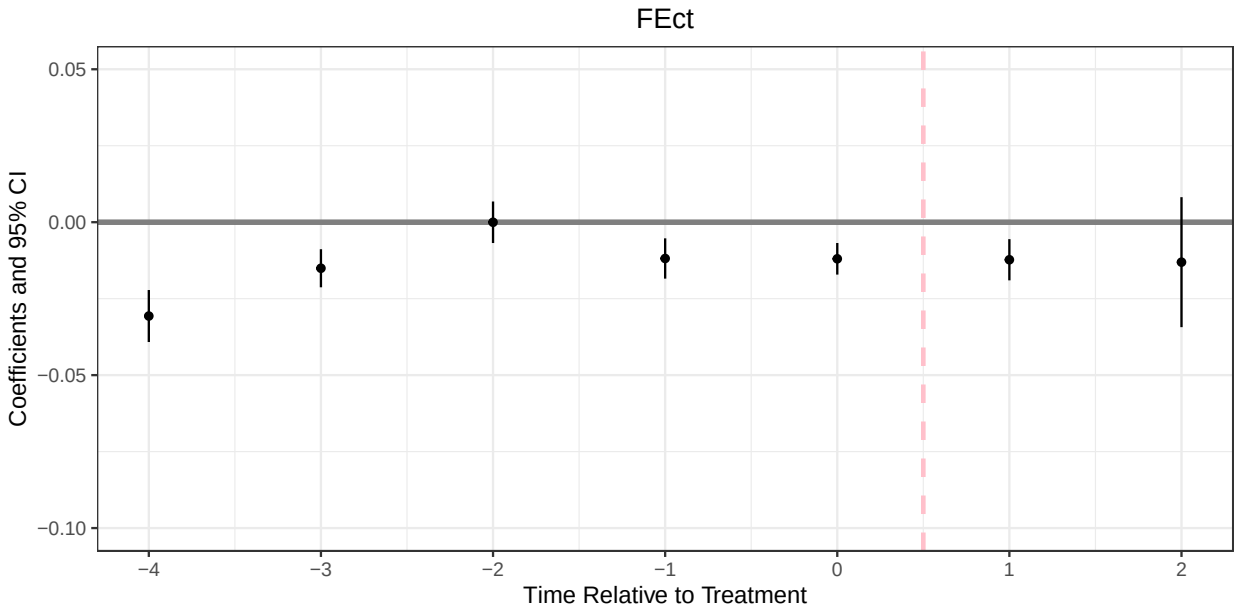


The TWFE and FEct estimator are consistent with each other. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct.

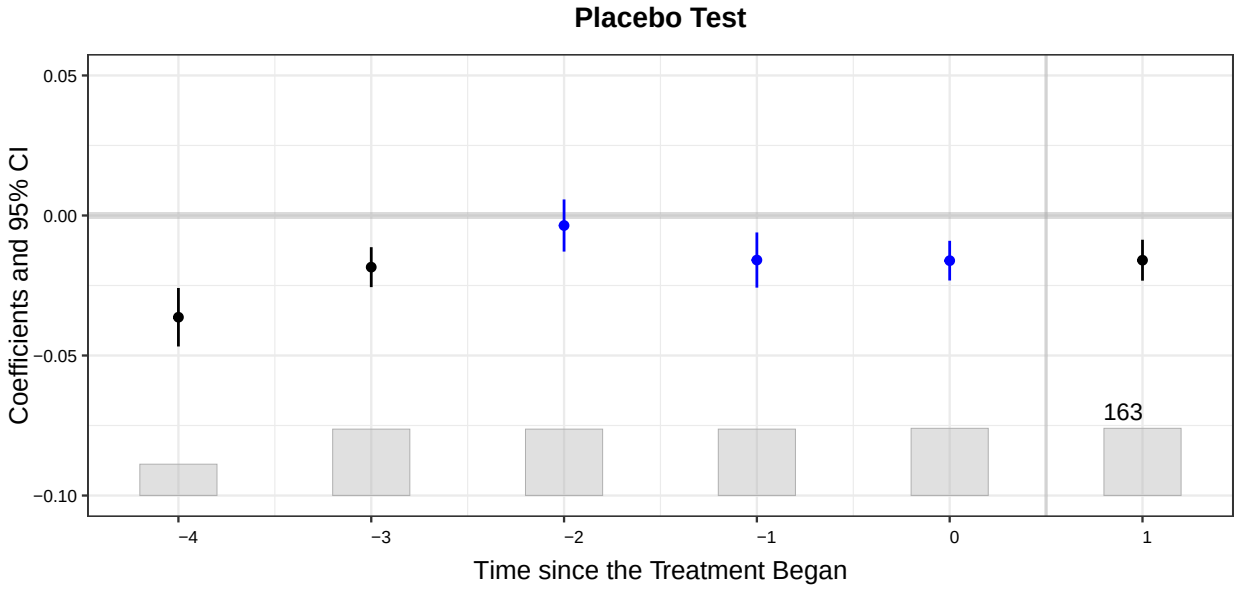


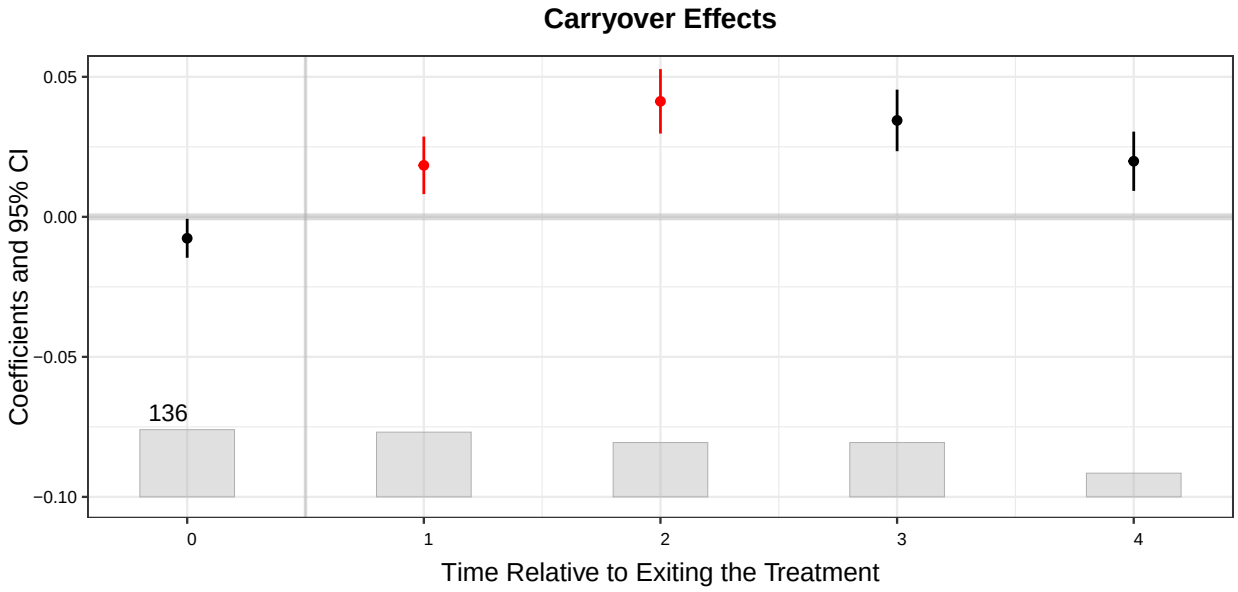


The TWFE and FEct estimator are consistent with each other.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.





Test Statistics

```
## Cannot use full pre-treatment periods in the F test. The first period is removed.
```

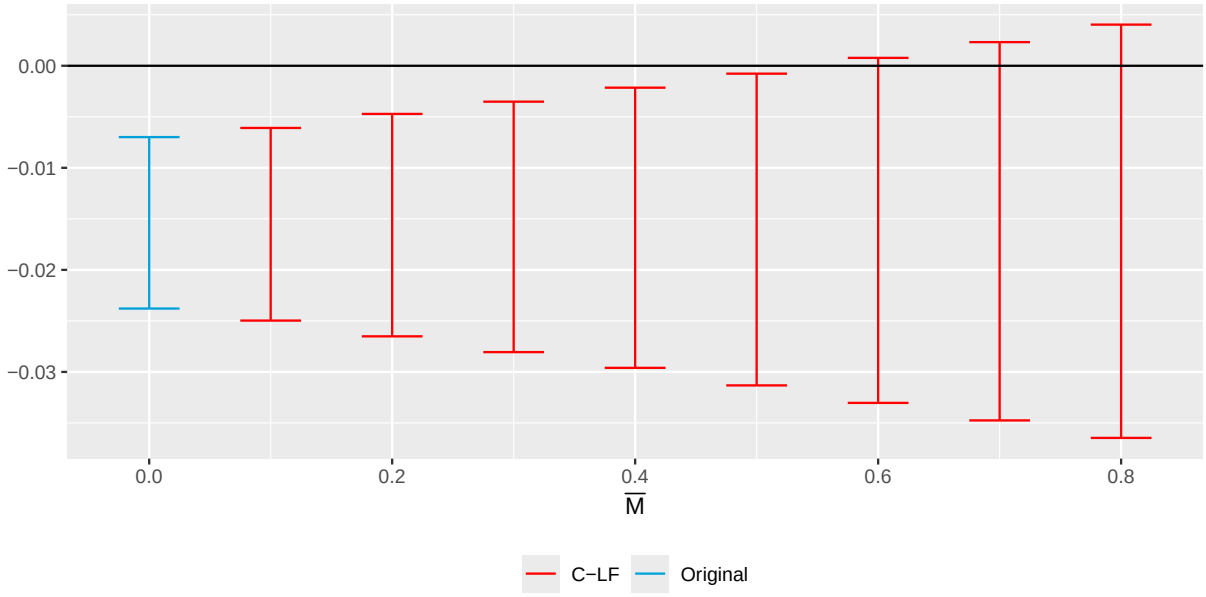
```
##                                  p-value
## F test                             0.000
## Equivalence test (default)         0.221
## Equivalence test (threshold=ATT)   0.803
## Placebo test                       0.002
## Carryover effect test              0.000
```

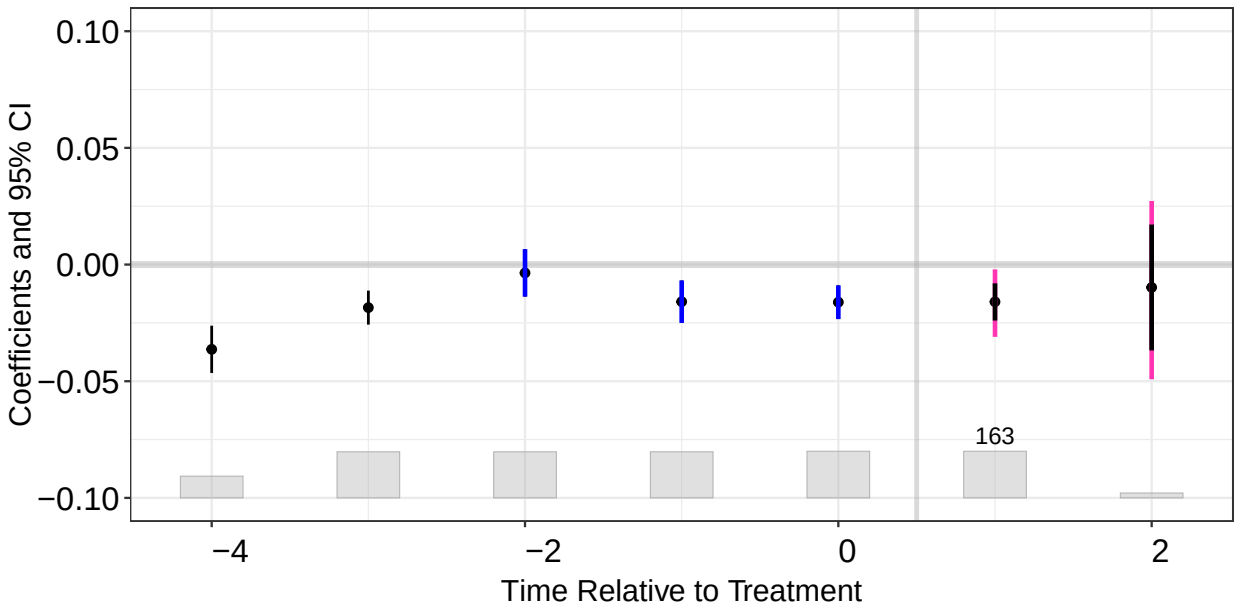
We find some evidence of a the parallel trends violation: the F test and placebo test reject, while neither of the equivalence tests reject. We also find some evidence of carryover effects.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.55$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 0.55 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust when using HTE-robust estimators. We find fairly strong evidence of PT violations. The estimated ATT appears to be robust to mild PT violations with the breakdown threshold $M = 0.55$.

Paglayan (2022)

Contents

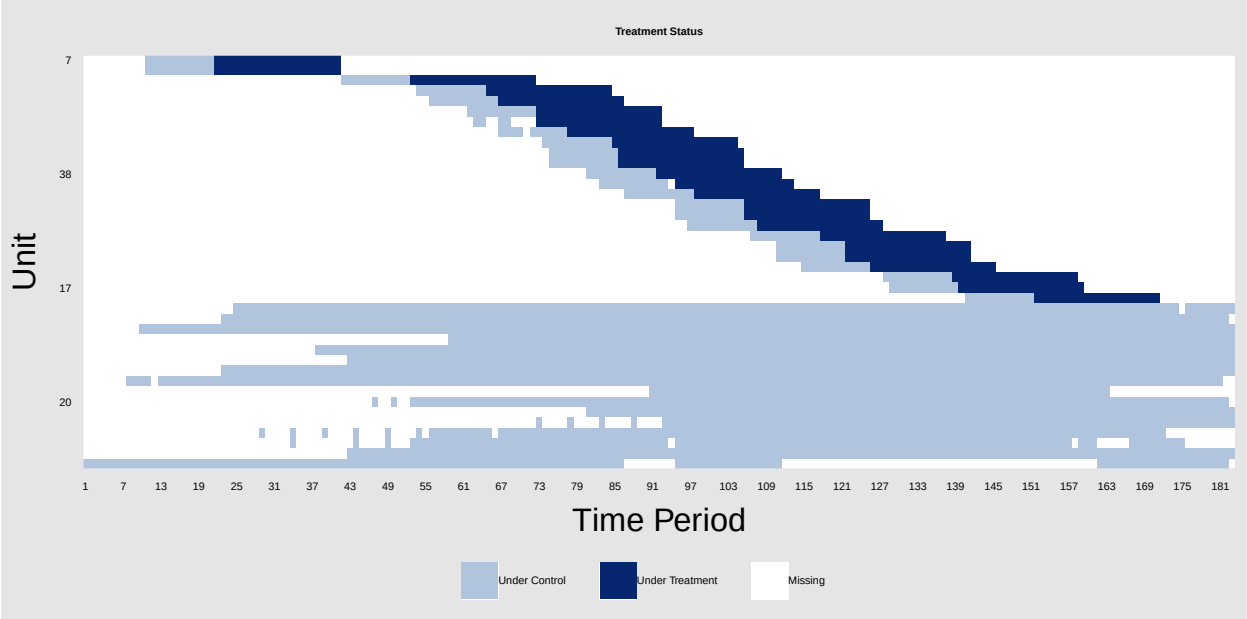
Summary 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 7
Sensitivity Analysis 8
Summary 9

Summary

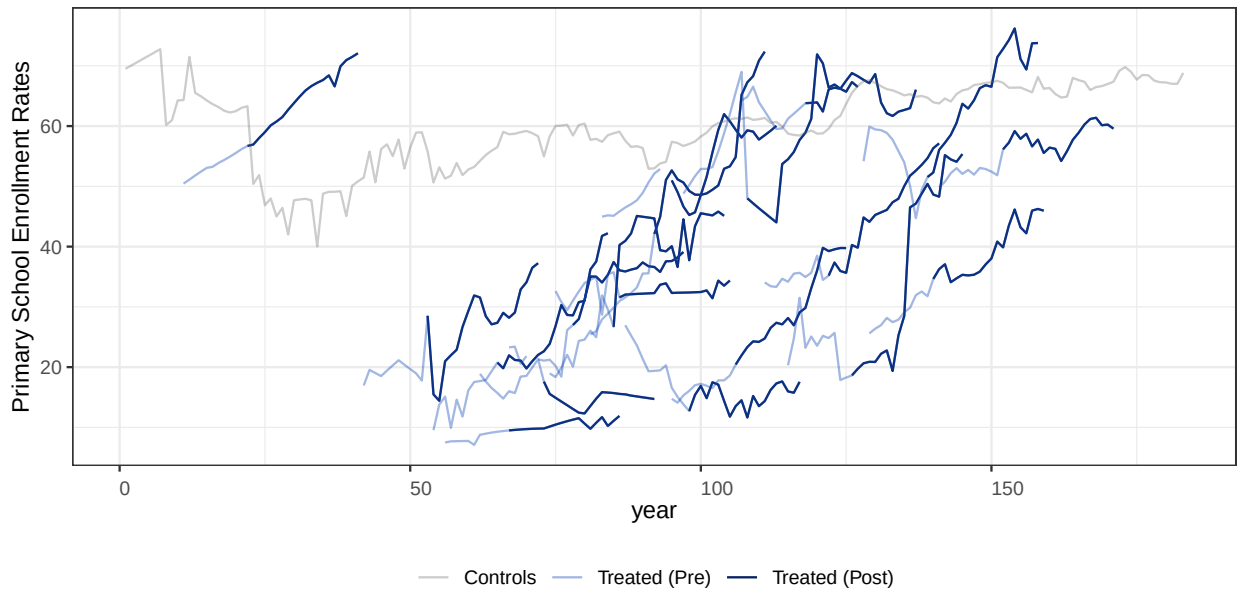
This paper investigates the effects of civil wars on Primary School Enrollment Rates (SERs), using country-year panel data from 23 Latin American and European countries during 1830-2015. The paper finds that "war-afflicted countries saw a gradual and sustained increase in primary SERs after the war that exceeded the contemporaneous increase in countries that did not experience civil war" (p1250). We replicate Figure 2 of the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the Country level. Throughout the analysis, blue and black bars represent CIs based on cluster SEs and cluster-bootstrapped CIs, respectively.

Replication Summary	
Unit of analysis	Country × year
Treatment	Civil Wars
Outcome	Primary School Enrollment Rates
Treatment type	Staggered
Outcome type	Continous
Fixed effects	Unit+Time

View treatment status First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a country). We see that over a half of countries are treated eventually and there is no treatment reversals.



View the outcome We plot the trajectory of the average outcome for each cohort. The trajectory of the control cohort is depicted in gray. For the ever-treated cohorts, we mark their pre-treatment periods in light blue and highlight their post-treatment periods in deep blue.



Point Estimates

We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units (there is none in this data) and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna) estimator, DIDM and the imputation estimator (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below.

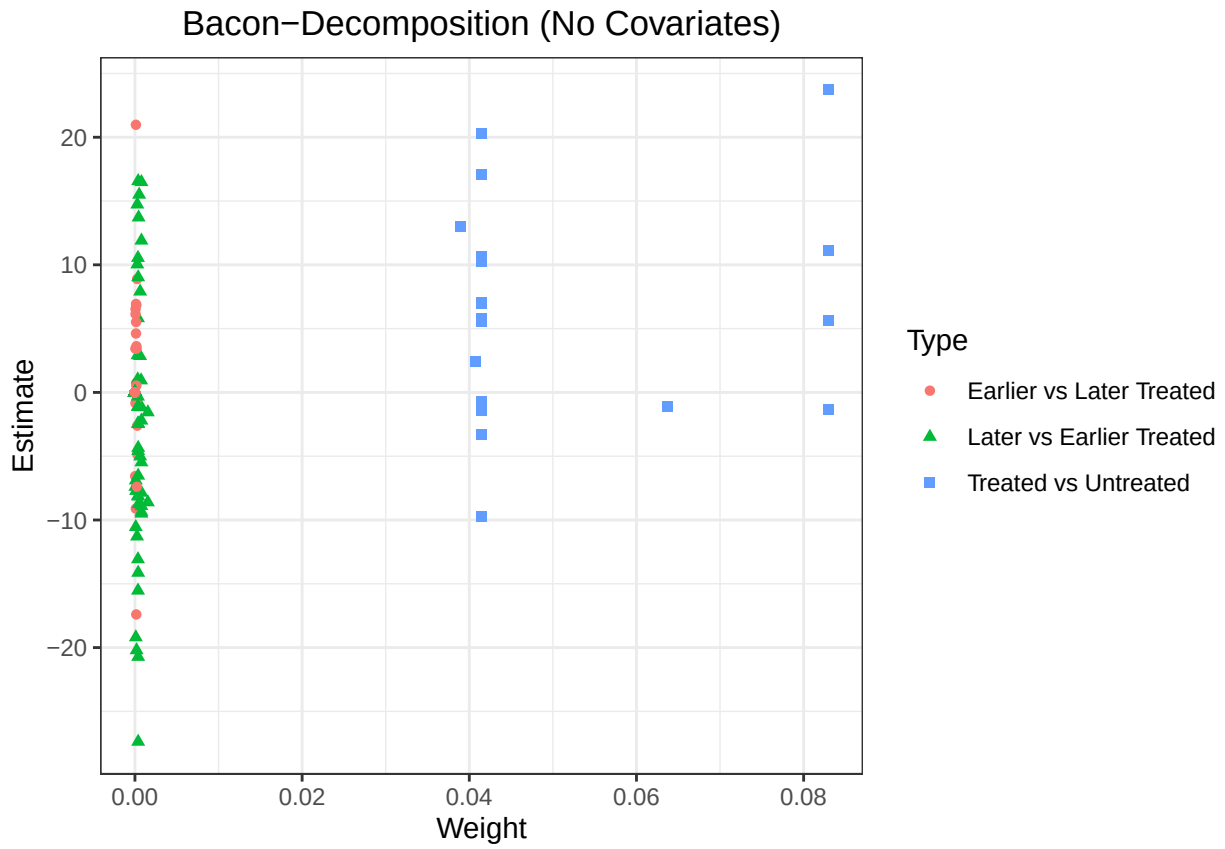
Original Results

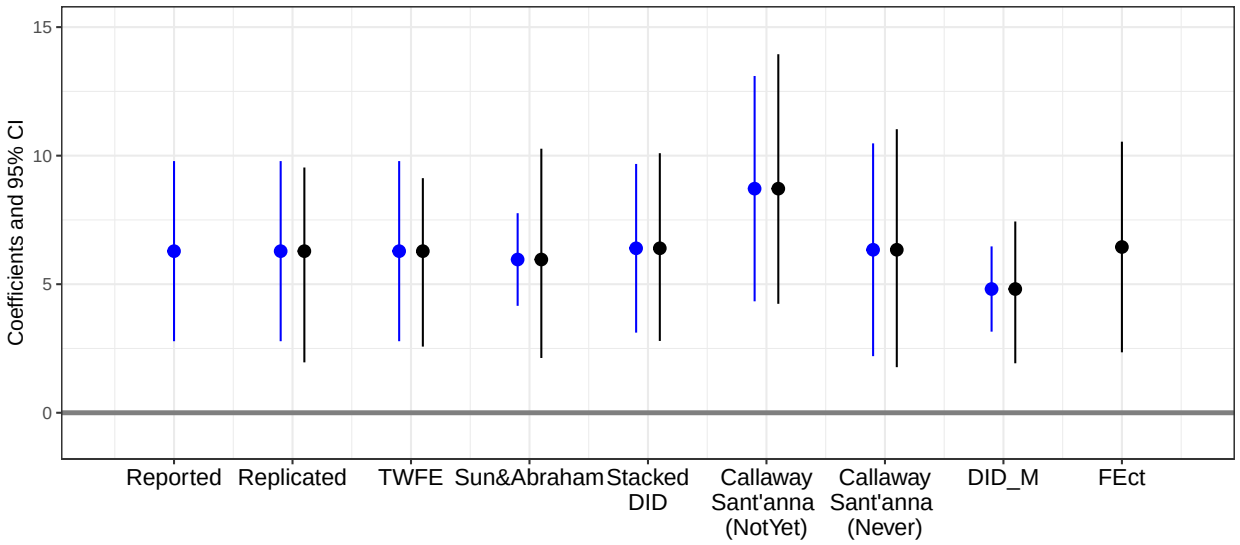
```
sol <- feols(primratio~treatment|Countryid+year,data = df,cluster = "Countryid")
summary(sol)
```

```
## OLS estimation, Dep. Var.: primratio
## Observations: 2,882
## Fixed-effects: Countryid: 40,  year: 183
## Standard-errors: Clustered (Countryid)
##           Estimate Std. Error t value  Pr(>|t|)
## treatment  6.28577    1.78714 3.51722 0.0011241 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 10.0     Adj. R2: 0.733145
##              Within R2: 0.02072
```

Goodman-Bacon Decomposition

In the Goodman-Bacon decomposition, we can see that the comparison "treated vs untreated" has a dominating impact on the original point estimate.

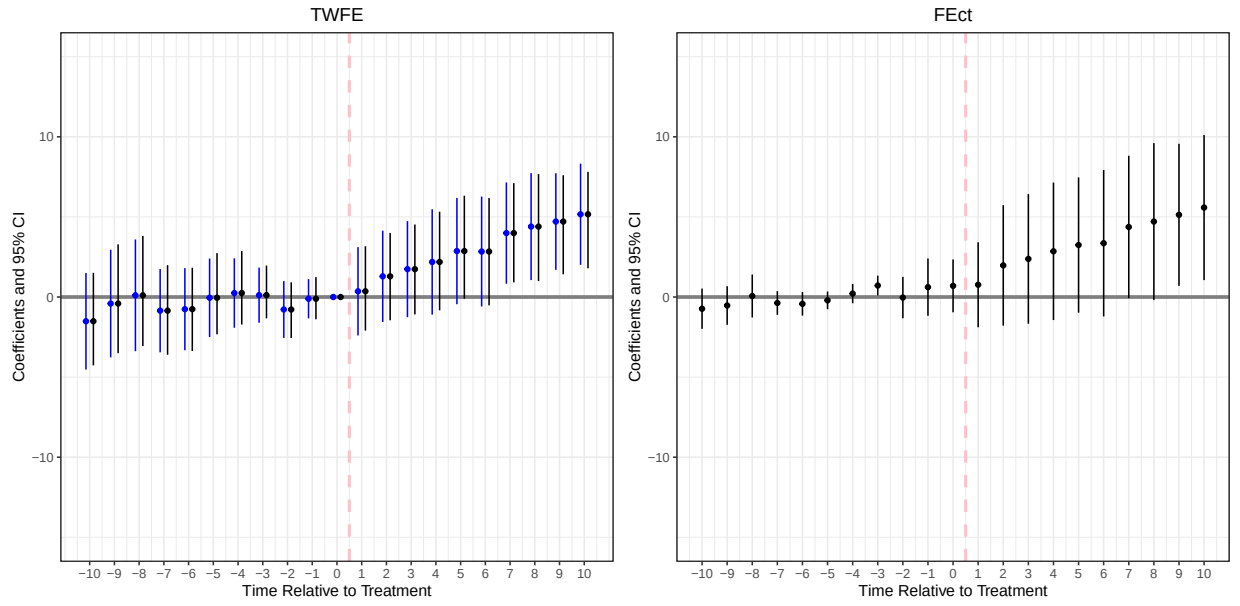


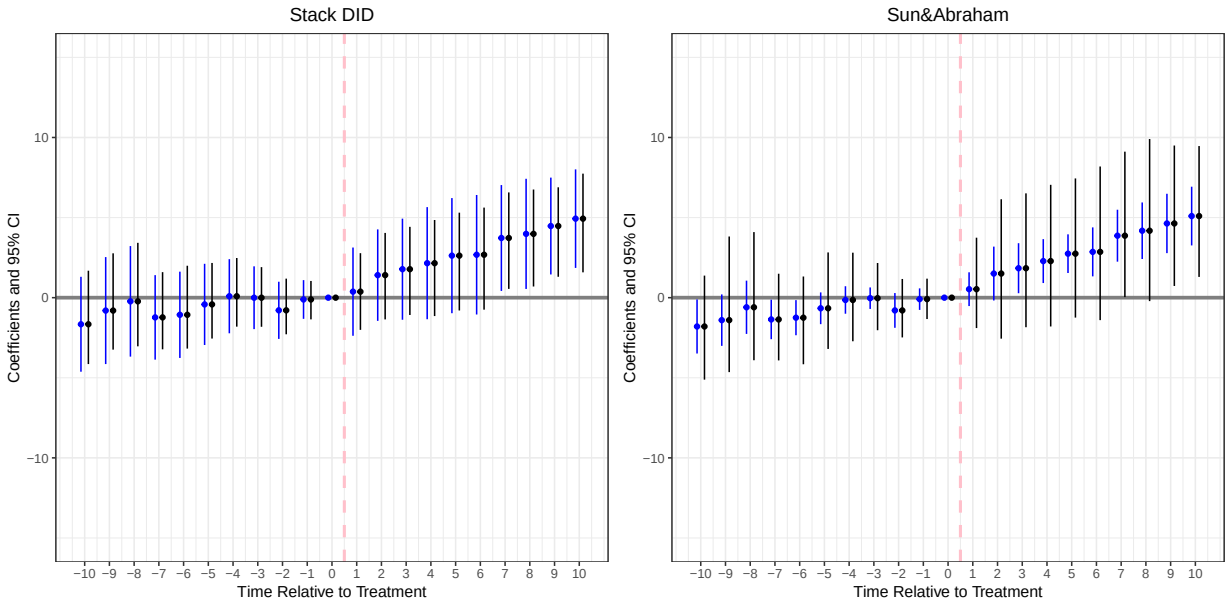


All HTE-robust estimators and TWFE are broadly consistent.

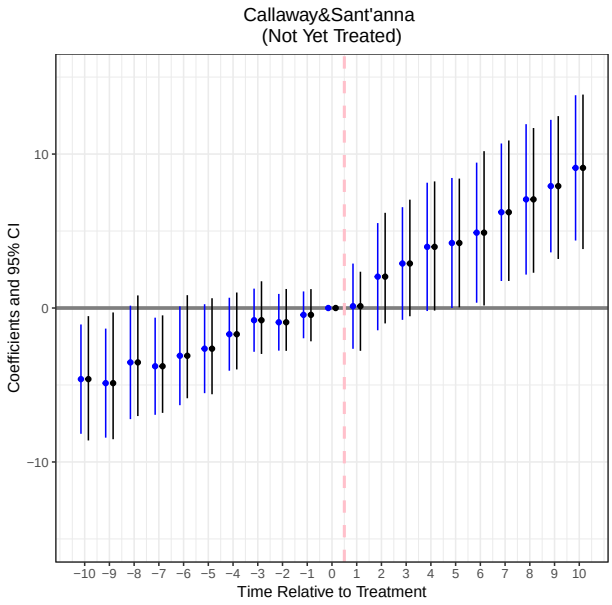
Dynamic Treatment Effects

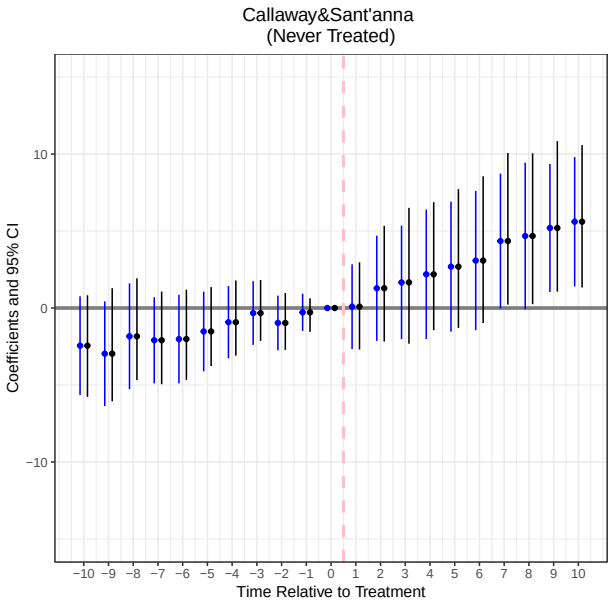
We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use six estimators, TWFE, IW, CS, Stacked DID, DIDM and FEct.

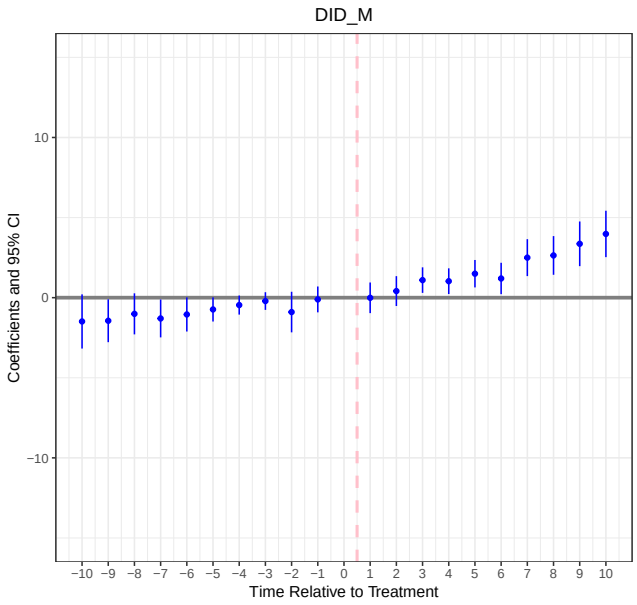



Stack DID
Coefficients and 95% CI
10
0
−10
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10
Time Relative to Treatment

Sun&Abraham
Coefficients and 95% CI
10
0
−10
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10
Time Relative to Treatment


Callaway&Sant'anna
(Not Yet Treated)
Coefficients and 95% CI
10
0
−10
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10
Time Relative to Treatment

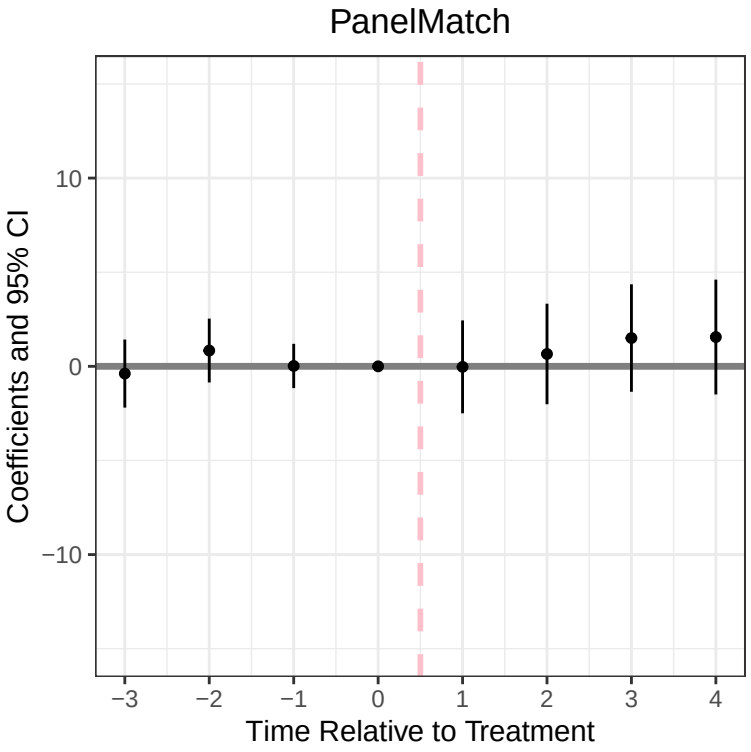

Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
10
0
−10
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10
Time Relative to Treatment



All HTE-robust estimators and TWFE are broadly consistent. The estimated DTEs show obvious upward trends during post-treatment periods but are also occasionally statistically significant in the earliest pre-treatment periods when they are largest in magnitude.

DTE of a Balanced Subsample

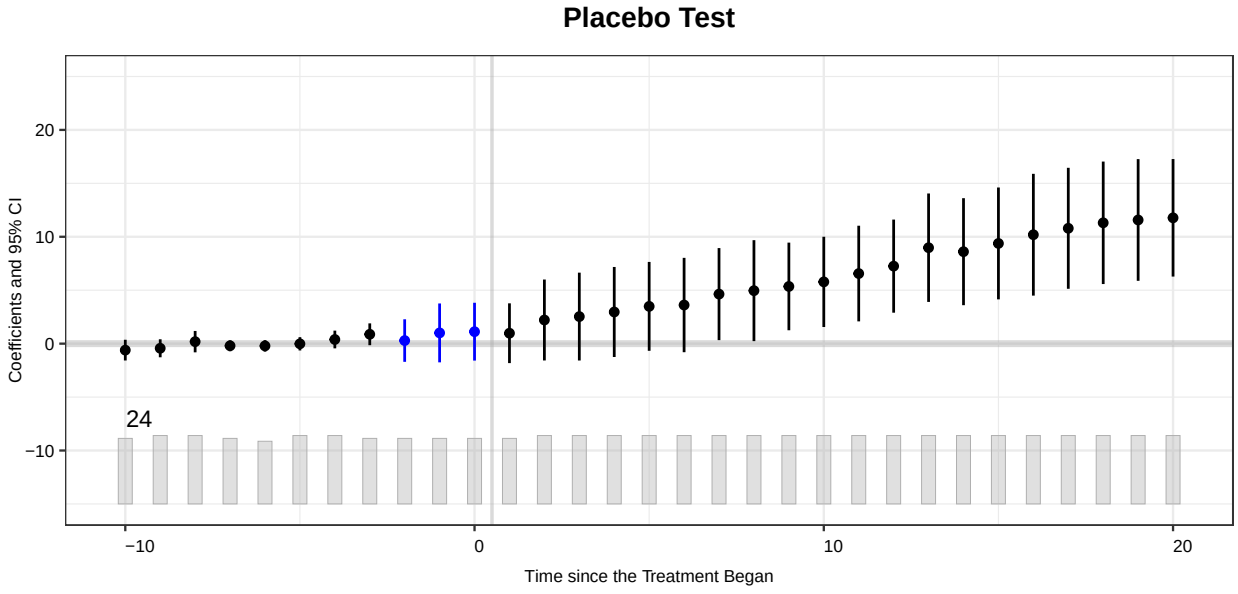
We also estimate the DTE using PanelMatch (lead = 4 and lag = 4). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.

Placebo Test



Test Results

```
##                                    p-value
## F test                               0.639
## Equivalence test (default)           0.000
## Equivalence test (threshold=ATT)     0.000
## Placebo test                         0.499
## Carryover effect test                   NA
```

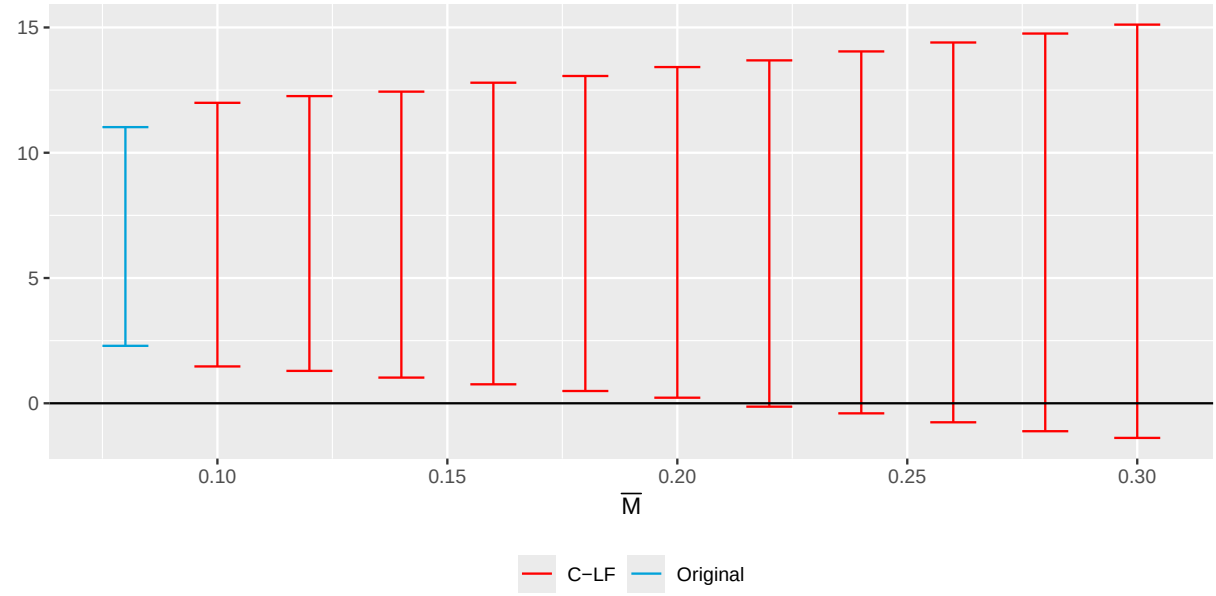
We find no evidence for violations of the parallel trends (PT) assumption using the F test and the placebo test. The equivalence test also rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

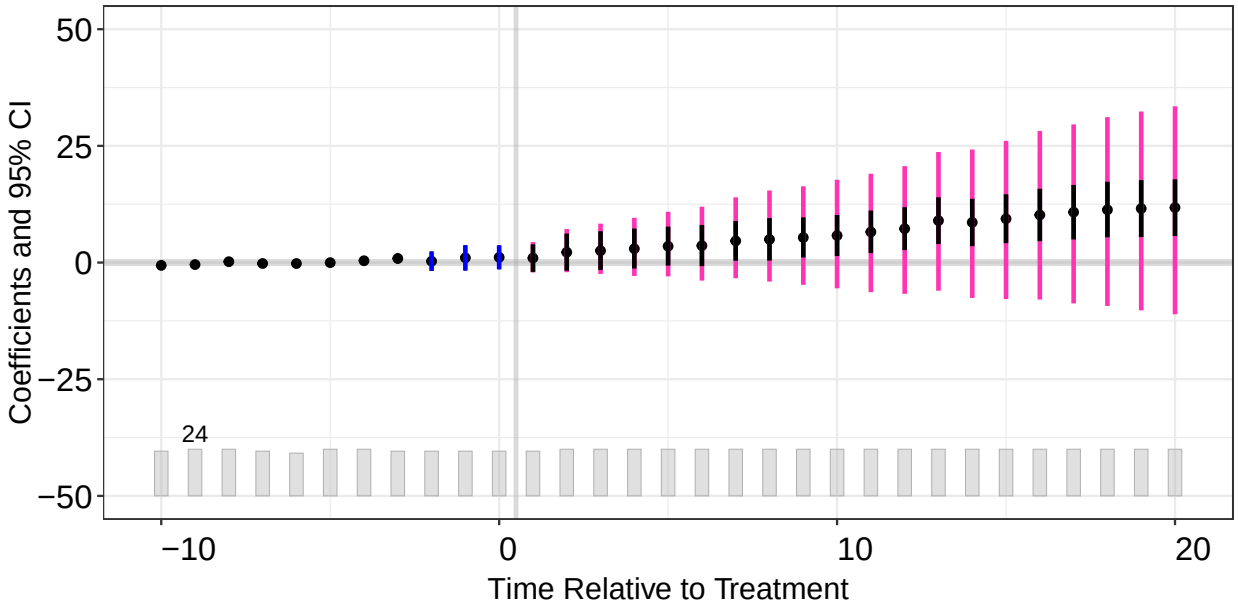
Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.2$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 0.2 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.

```
##   |                                                                      |
```





Summary

Overall, the main result of the chosen model seems to be robust to HTE-robust estimators. However, we find some evidence for violations of the PT assumption, which is likely driven by the distinctive trends of the never-treated units. The estimated ATT is sensitive to mild PT violations with the breakdown threshold $M = 0.2$.

Payson (2020a)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

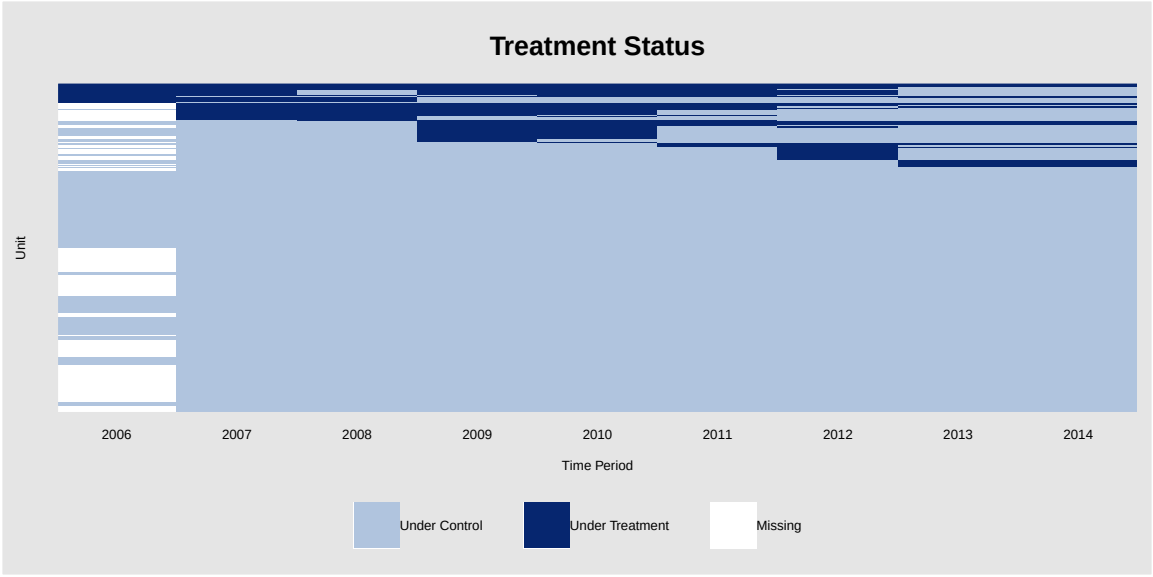
A First Look at Data

The paper investigates the effects of partisan mismatch on city lobbying, using US city-year level panel data, between 2006 and 2014. One of the main findings of this paper is that "When an election leads to a partisan mismatch between a city's residents and the party of their state representative, the probability of lobbying increases between 4% and 7%" (p682).

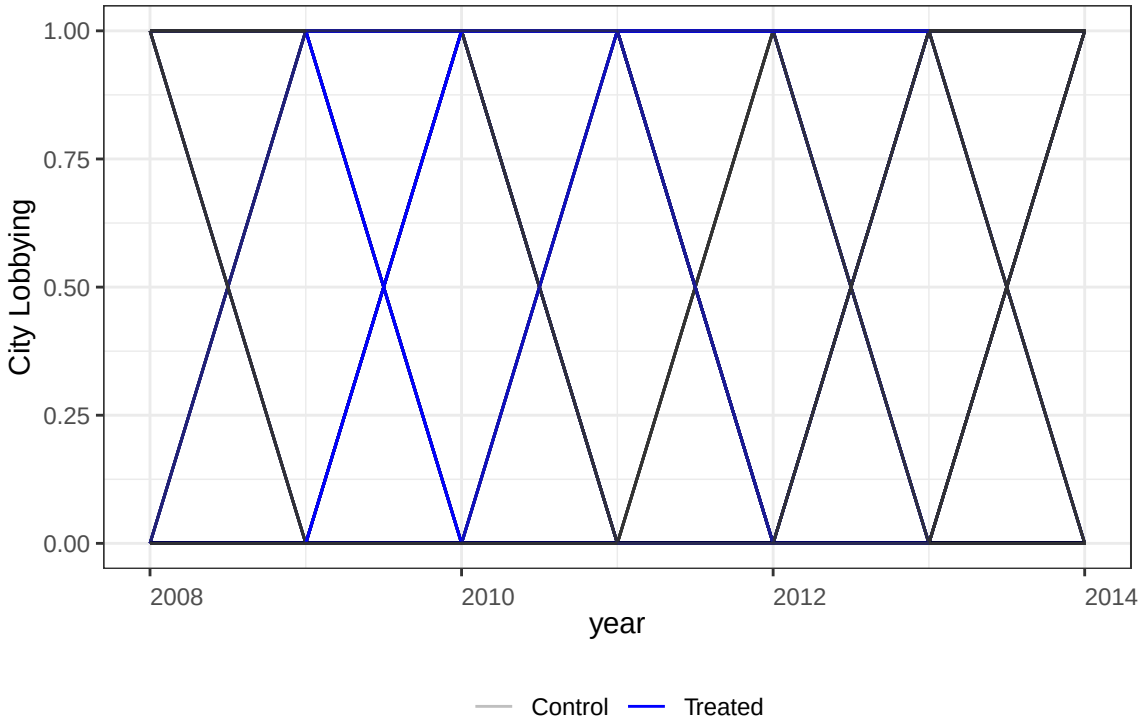
Model. We focus on **Model 1 of Table 1** in the paper. The author uses a two-way fixed effects (TWFE) model and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	City × year
Treatment	Partisan mismatch
Outcome	City lobbying
Treatment type	General
Outcome type	Binary
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a city). There are treatment reversals and some missingness.



Plotting the outcome variable. We plot the outcome variable, city lobbying, for each city. Treated observations are in blue.



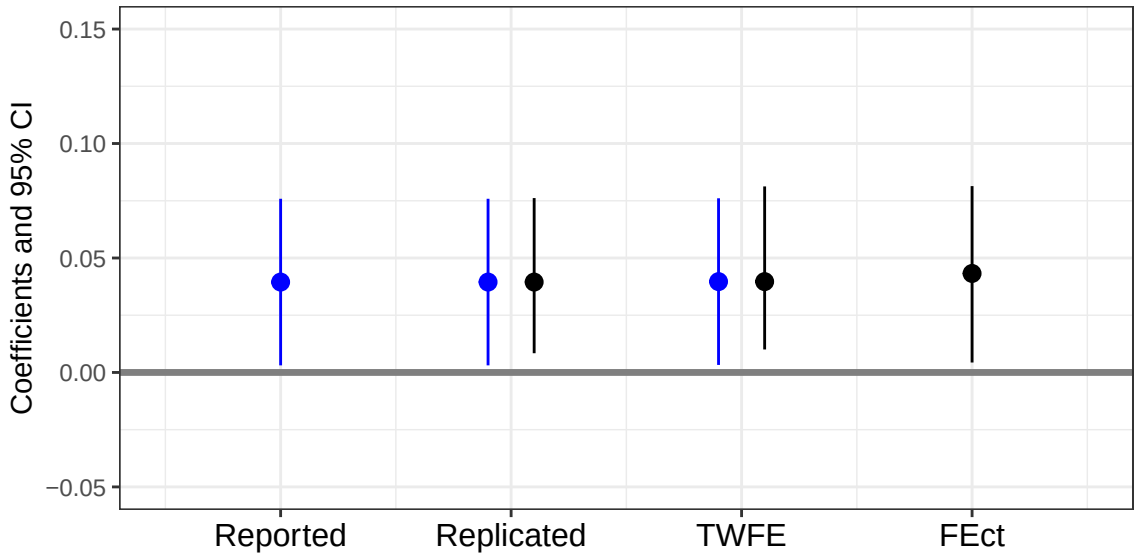
Point Estimates

We first replicate the authors' using the original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(lobby~ mismatch + log.pop| fips + year, data = df, cluster = "fips")
summary(sol)
```

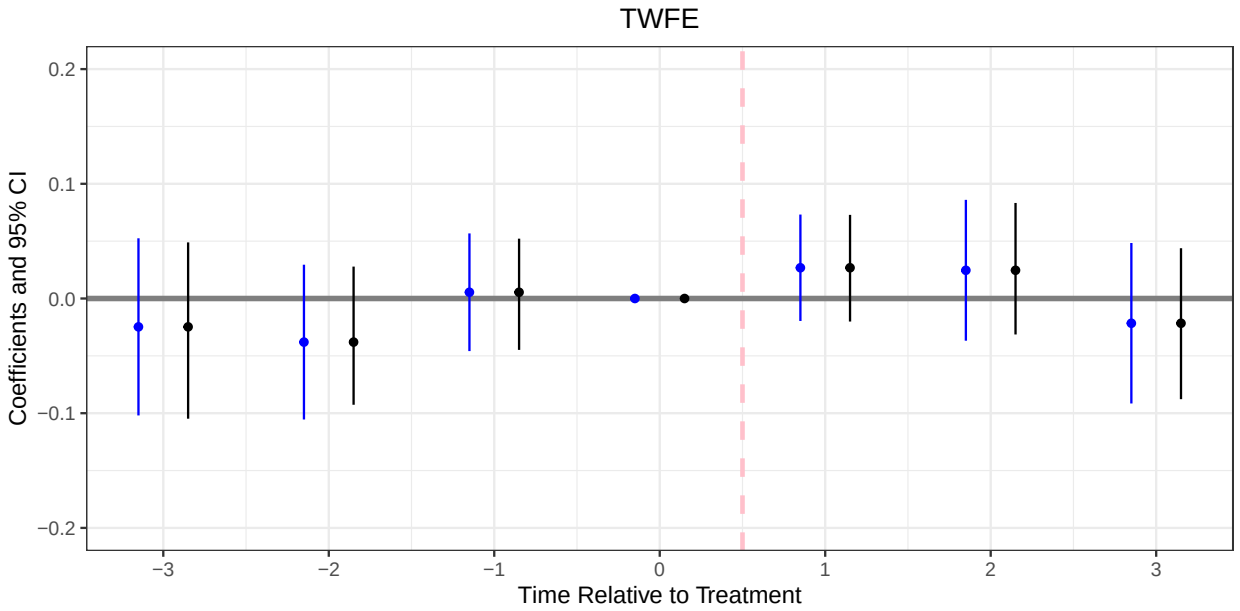
```
## OLS estimation, Dep. Var.: lobby
## Observations: 6,307
## Fixed-effects: fips: 738,  year: 9
## Standard-errors: Clustered (fips)
##          Estimate Std. Error t value Pr(>|t|)
## mismatch 0.039483   0.018562 2.12706 0.033747 *
## log.pop  0.241633   0.120521 2.00490 0.045339 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.245625     Adj. R2: 0.723702
##                  Within R2: 0.002931
```

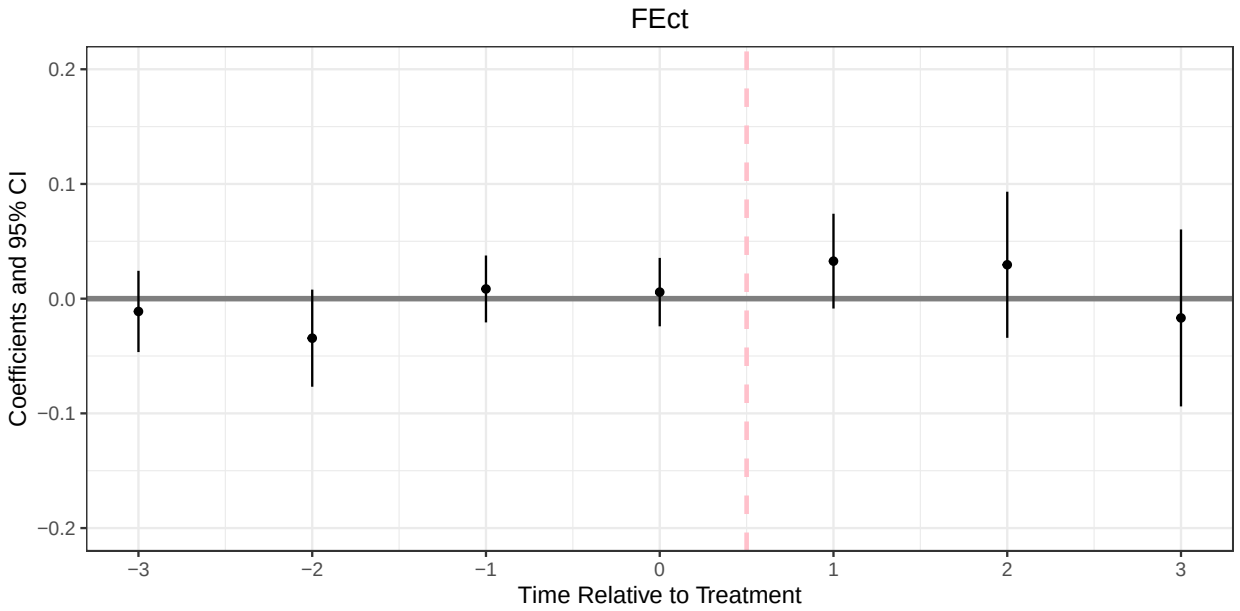


The TWFE and FEct estimates are consistent, but both are only marginally significant.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct.

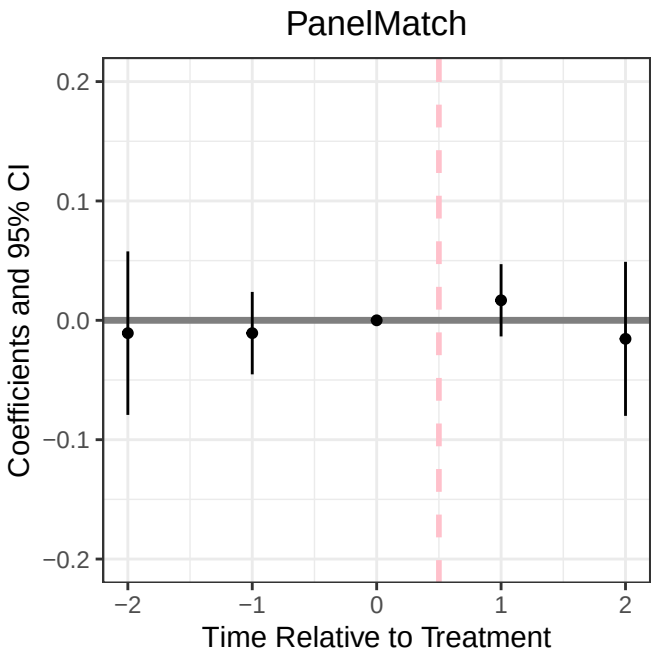




The estimated DTEs using TWFE and FEct exhibit similar shapes, initially showing positive values during the first two post-treatment periods, but transitioning to negative values in the third period.

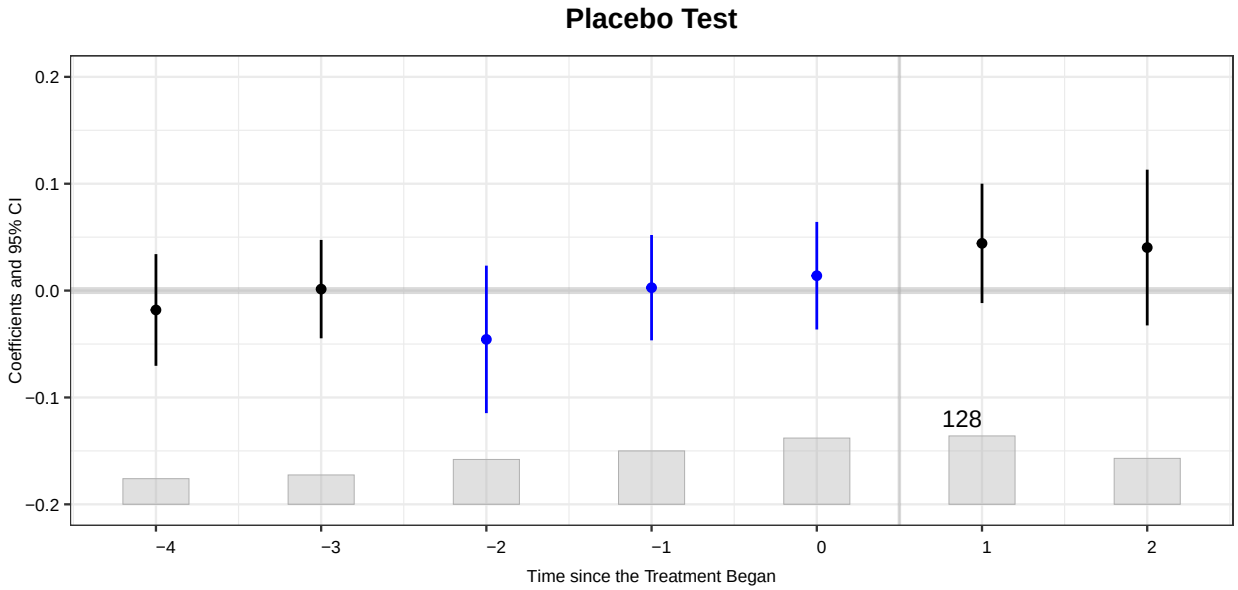
DTE of a Balanced Subsample

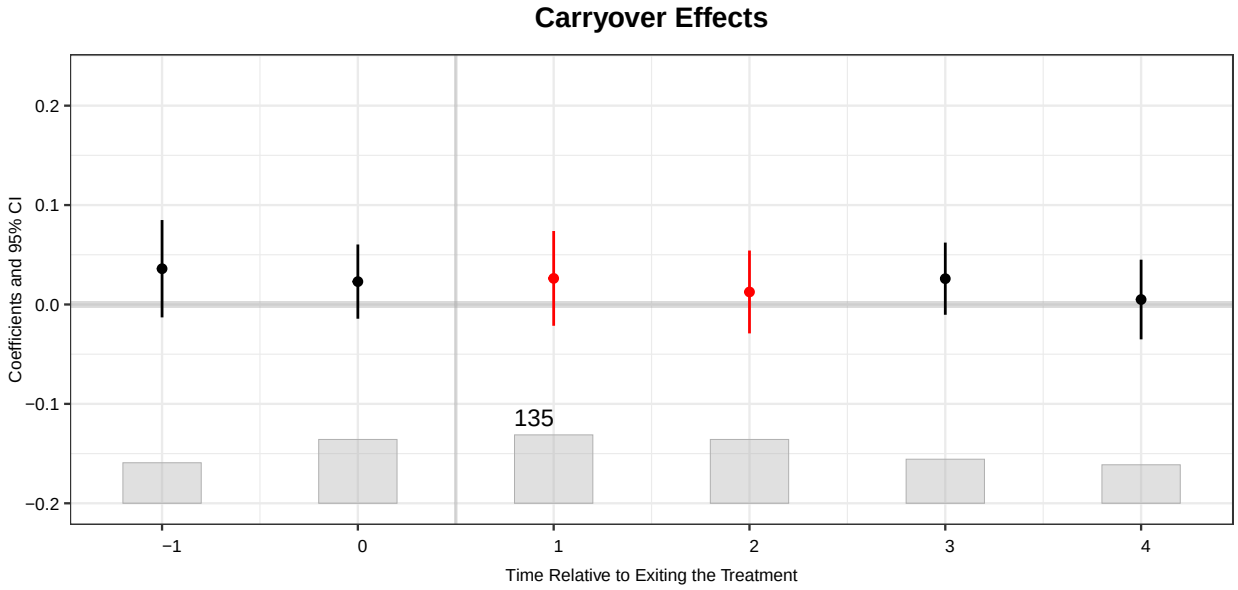
We also estimate the DTE using PanelMatch (lead = 2 and lag = 3). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.





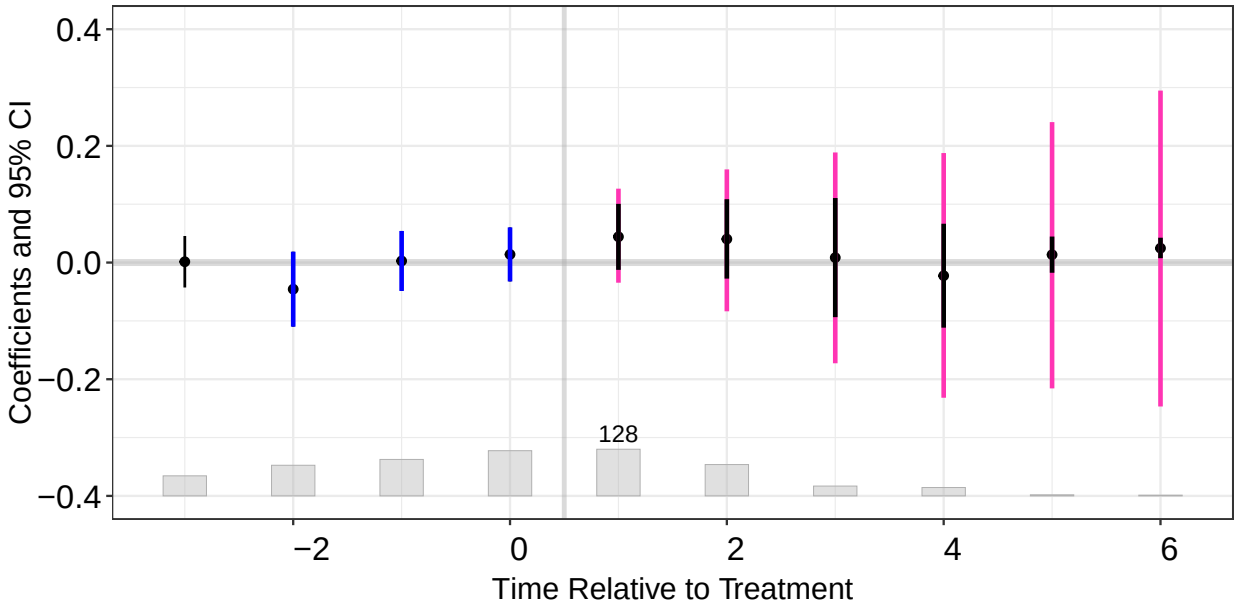
Test Statistics

```
##                                    p-value
## F test                               0.324
## Equivalence test (default)           0.013
## Equivalence test (threshold=ATT)     0.342
## Placebo test                         0.798
## Carryover effect test                0.338
```

We find little evidence of violations of the parallel trend (PT) assumption or the no-carryover-effect assumption. However, the equivalence test fails to reject the null hypothesis that the residuals in pre-treatment periods exceed the estimated ATT, possibly due to limited power.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, the main result of the chosen model appears to be robust to HTE-robust estimators. We find little evidence for PT violations. However, due to limited power, the estimated ATT is no longer statistically significant after we hold out three periods for the placebo test, which is the basis of the sensitivity analysis.

Payson (2020b)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

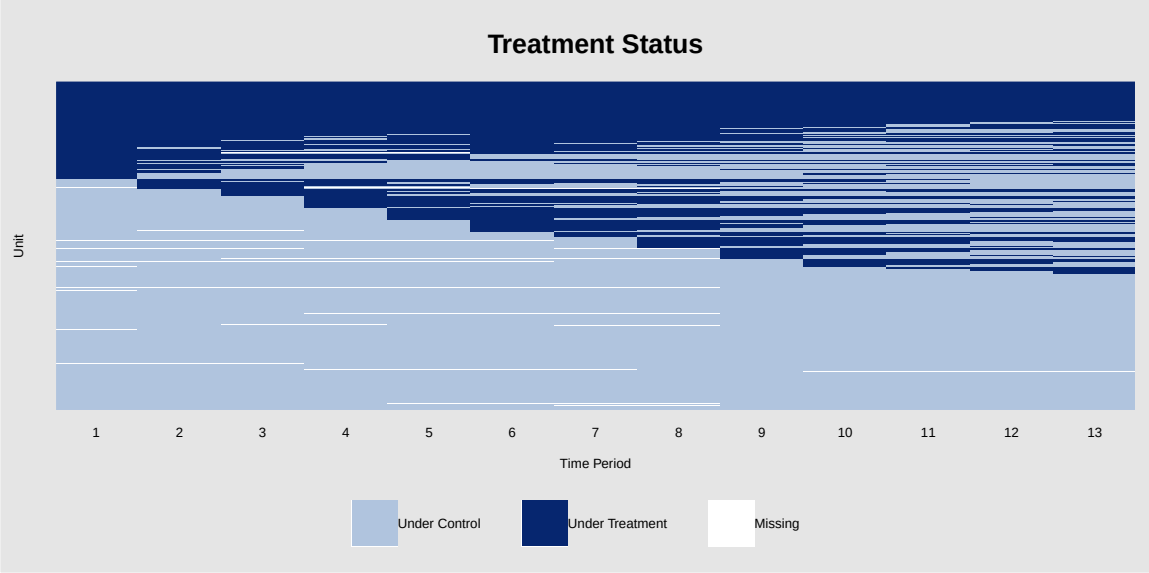
A First Look at Data

The paper investigates the effects of city lobbying on state transfers, using California's city-year level panel data, between 2002 and 2015. One of the main findings of this paper is that "the decision to lobby has a positive effect on subsequent city revenue per capita received from the state" (p409).

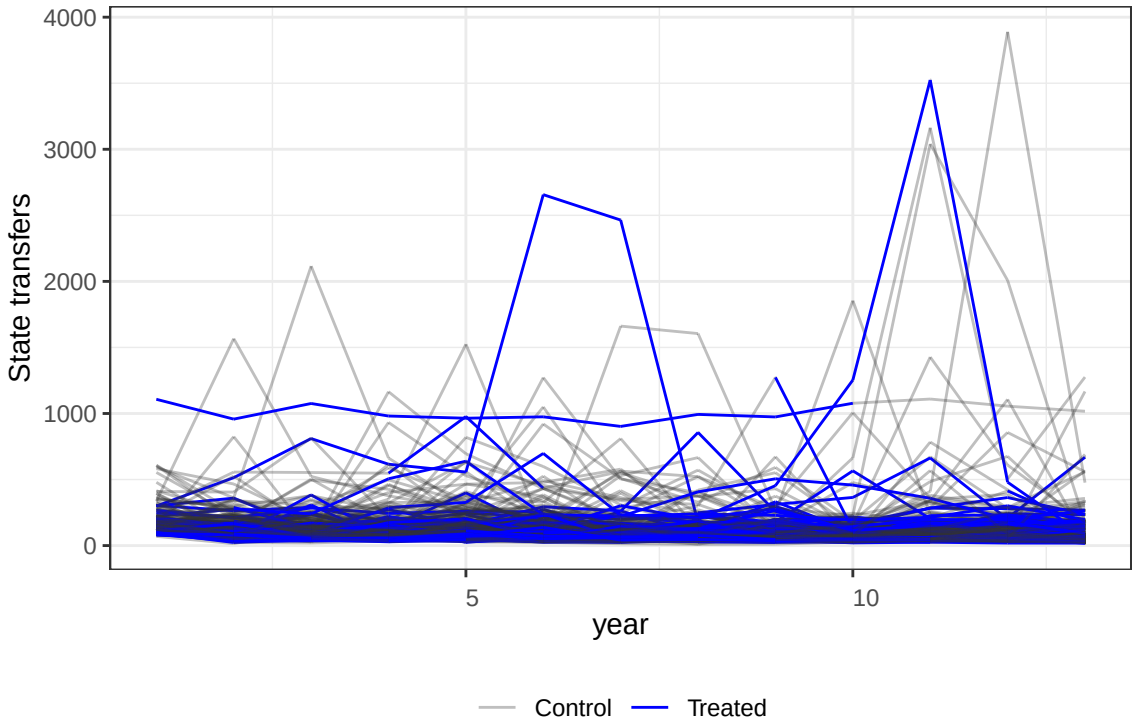
Model. We focus on **Model 2 of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the city level.

Replication Summary	
Unit of analysis	City × Year
Treatment	City lobbying
Outcome	State transfers
Treatment type	General
Outcome type	Continuous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represenst a unit (a city). We see that around half of the units are ever-treated and there are treatment reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable, state transfers, for each city. Treated observations are in blue.



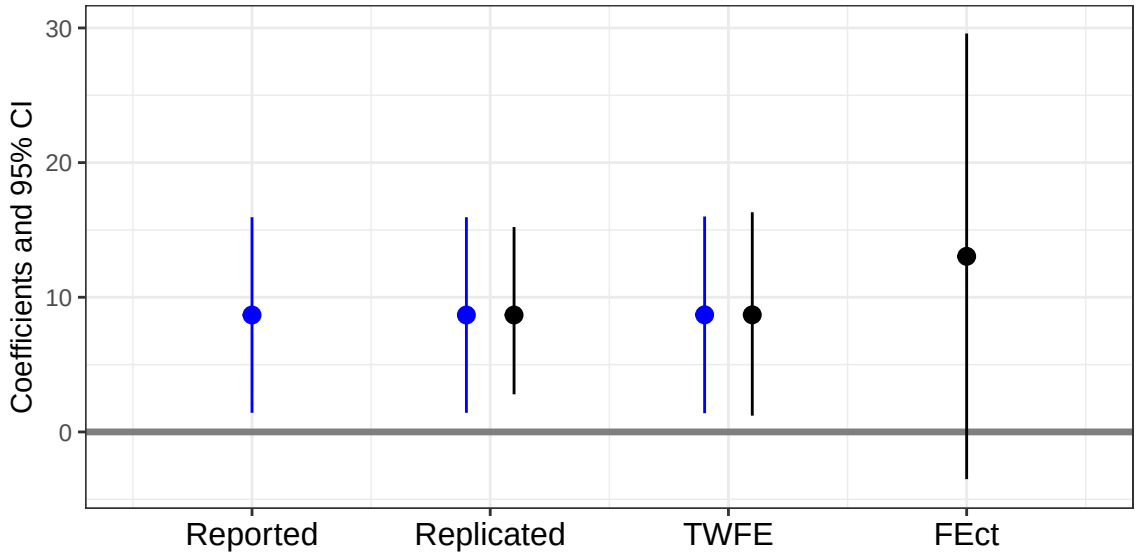
Point Estimates

We first replicate the authors' using the original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(state.pp1~ lobby + log.pop+ log.taxes| city + year, data = df, cluster = "city")
summary(sol)
```

```
## OLS estimation, Dep. Var.: state.pp1
## Observations: 5,982
## Fixed-effects: city: 467,  year: 13
## Standard-errors: Clustered (city)
##            Estimate Std. Error  t value Pr(>|t|)
## lobby       8.68196    3.70546  2.34301 0.019548 *
## log.pop   -78.12786   42.12754 -1.85456 0.064291 .
## log.taxes -19.38591   14.93624 -1.29791 0.194960
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 120.6     Adj. R2: 0.338389
##               Within R2: 0.002057
```

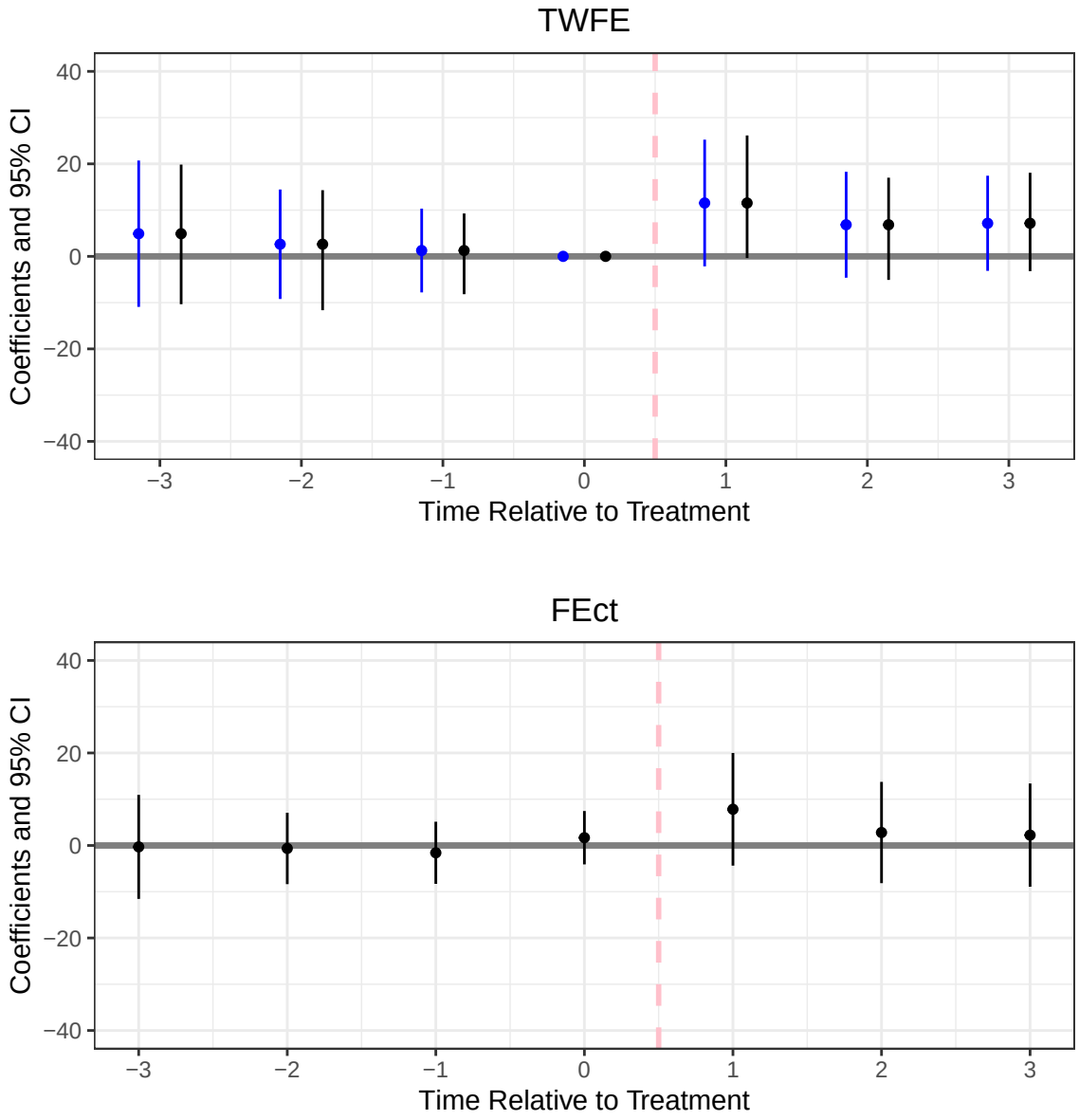


We also test the robustness of the finding by adding Unit-specific linear time trends (ULT) and lagged dependent variables (LDV) to both models. The results are shown in the figure below.

It appears the results are robust to HTE but potentially underpowered.

Dynamic Treatment Effects

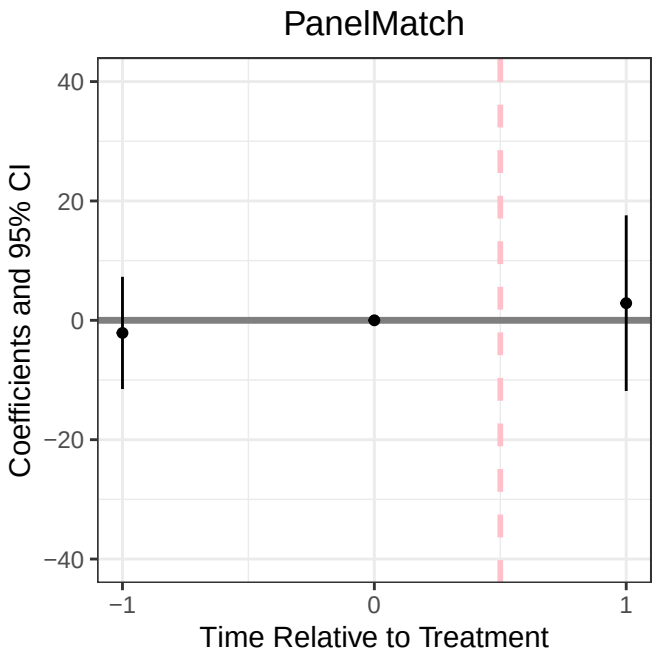
We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.



The estimated DTEs using TWFE and FEct exhibit similar shapes, showing positive values during the three post-treatment periods in the plot.

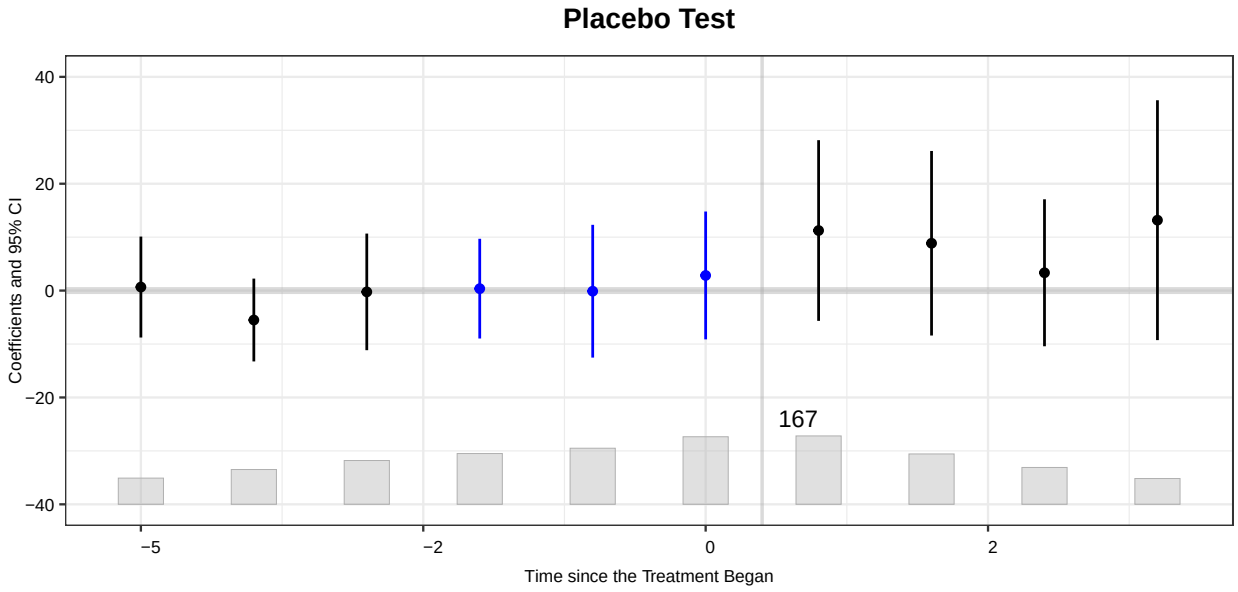
DTE of a Balanced Subsample

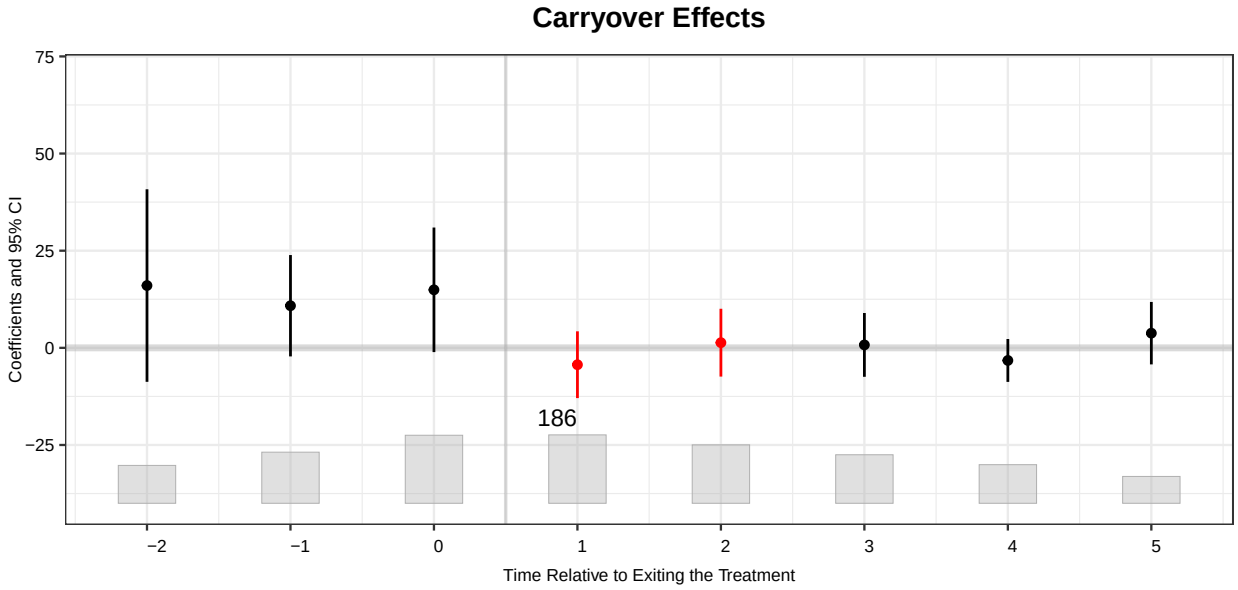
We also estimate the DTE using PanelMatch (lead = 1 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.





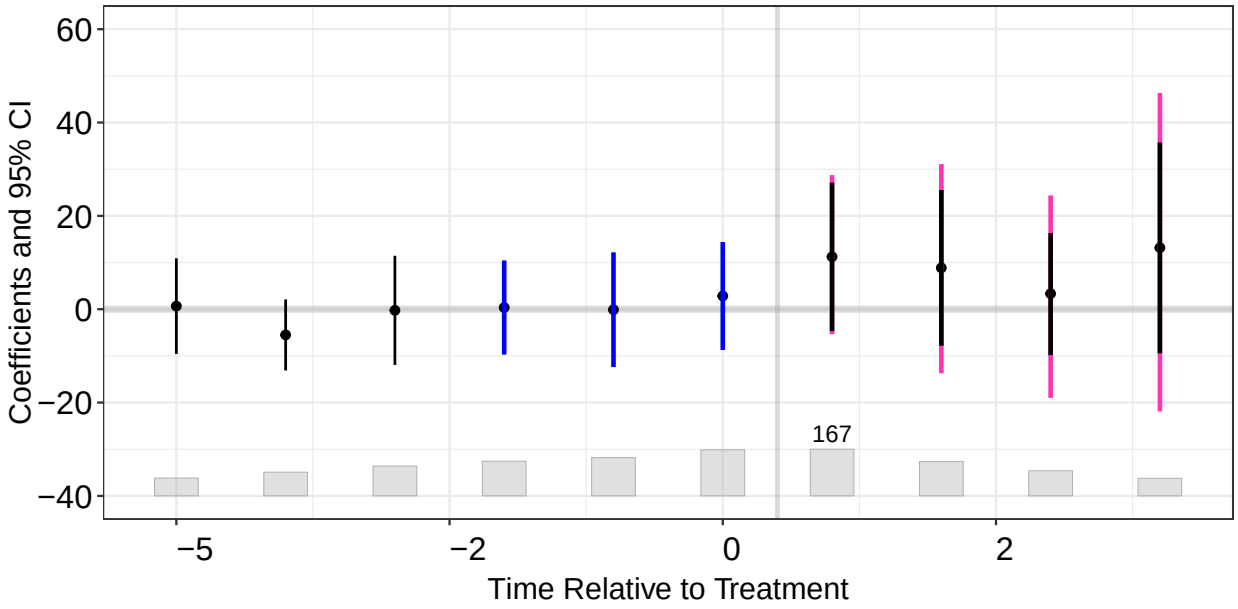
Test Statistics

```
##                                    p-value
## F test                               0.920
## Equivalence test (default)           0.000
## Equivalence test (threshold=ATT)     0.034
## Placebo test                         0.808
## Carryover effect test                0.658
```

We find little evidence for violations of the parallel trends (PT) assumption and no-carryover-effect assumption. The equivalence test also reject the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, the results seem relatively robust to HTE and modeling choices, although the study is underpowered after HTE-robust estimators are used. We do not find strong evidence for potential PT violations.

Pierskalla and Sacks (2021)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 5
Sensitivity Analysis 6
Summary 7

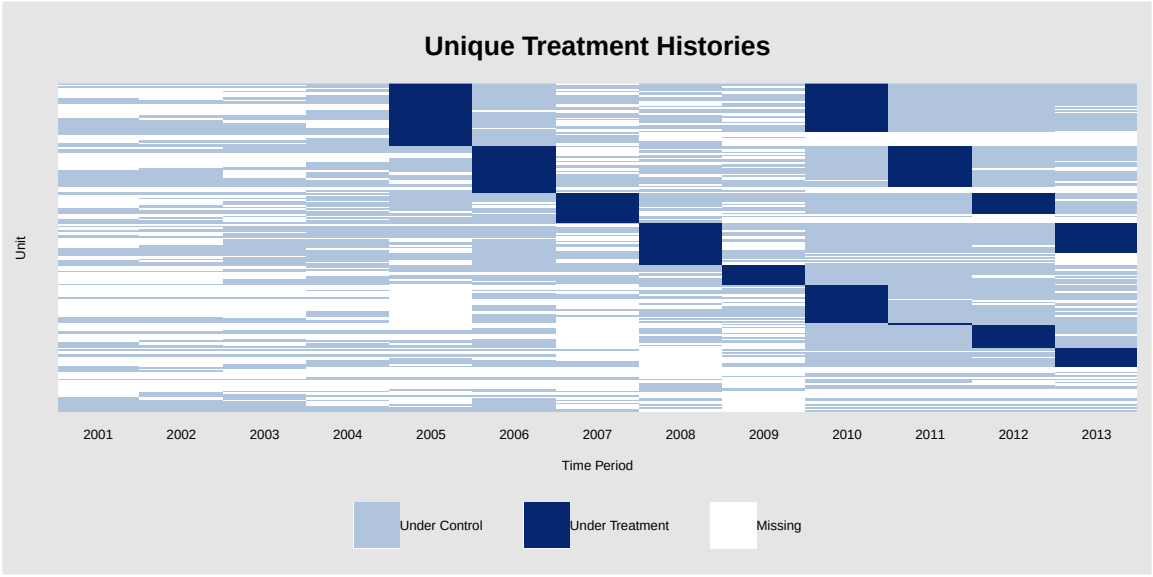
A First Look at Data

The paper investigates the effects of elections on per capita expenditures, using Indonesian district-year level panel data, between 2001 and 2012. One of the main findings of this paper is that "a statistically significant drop in overall expenditures in election years … is driven by a reduction in capital expenditures" (p518).

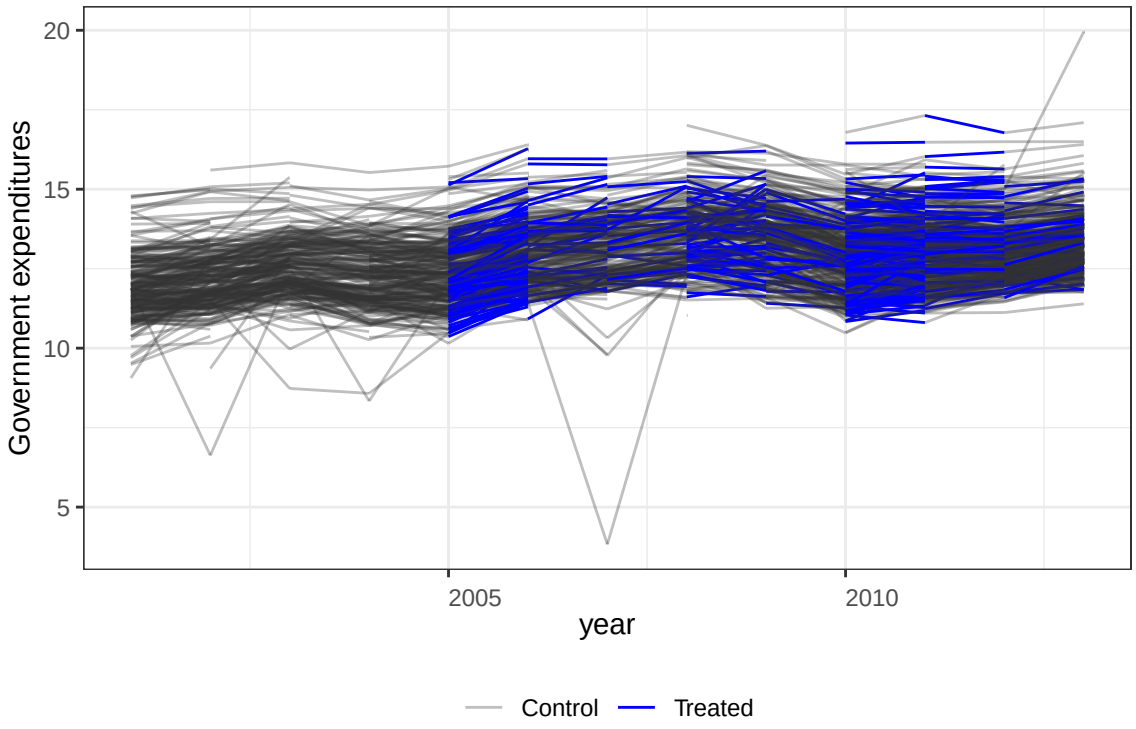
Model. We focus on **Model 2 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the district level.

Replication Summary	
Unit of analysis	District × year
Treatment	Election
Outcome	Government expenditures
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a district). There are treatment reversals and some missingness.



Plotting the outcome variable. We plot the trajectories of the outcome variable, Government expenditures, for each district. Ever-treated units are in blue.



Point Estimates

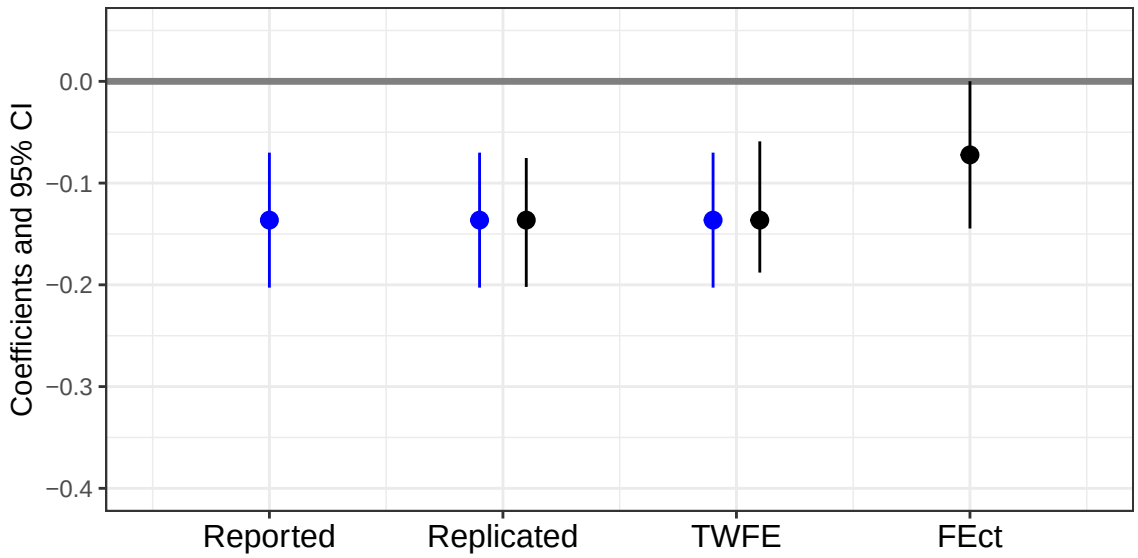
We first replicate the authors' using the original specification. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis,

we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(lcapital_exp_pc~election_year+ election_year_lead1+
               election_year_l+elected_leader_l+incumbency+enp_all+
               golkar_share_all+pdip_share_all+services_provision+
               rev_natural_pc_l+gini_l+rev_total_pc2_l+lpop_l+poverty_pc_l+lgdppc_l|kode_neil+year,
             data = df,cluster = "kode_neil")
summary(sol)
```

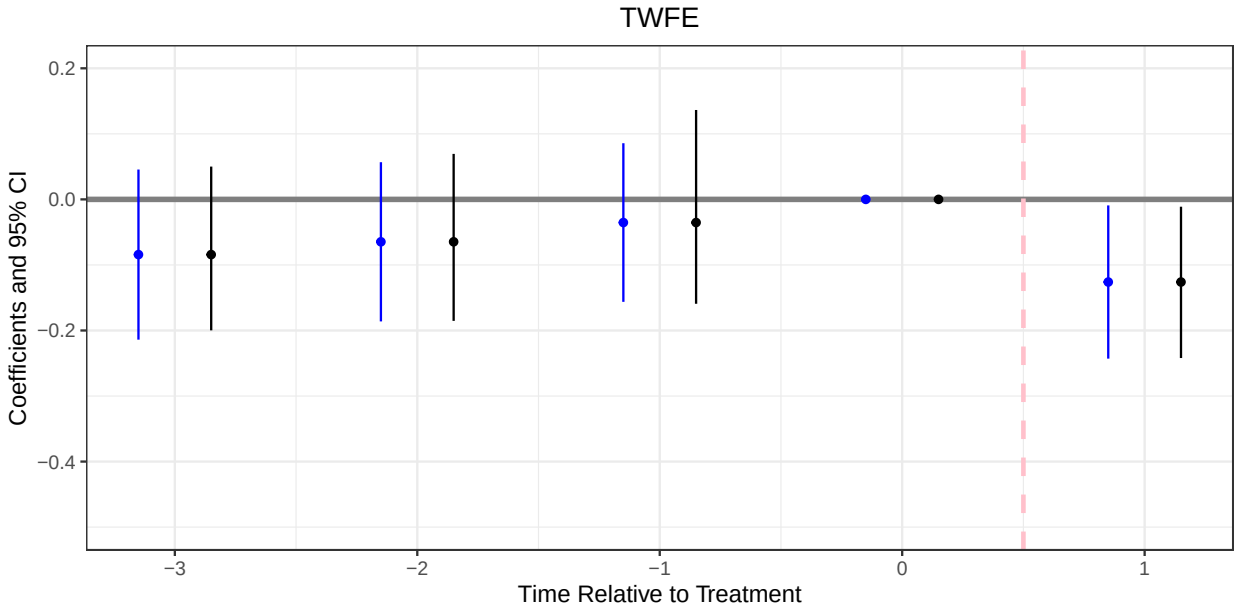
```
## OLS estimation, Dep. Var.: lcapital_exp_pc
## Observations: 2,524
## Fixed-effects: kode_neil: 455,  year: 9
## Standard-errors: Clustered (kode_neil)
##                          Estimate   Std. Error   t value   Pr(>|t|)
## election_year       -1.364003e-01 3.384890e-02 -4.029683 0.00006548 ***
## election_year_lead1 -7.306826e-03 2.704802e-02 -0.270143 0.78717324
## election_year_l      1.611981e-02 2.899168e-02  0.556015 0.57847449
## elected_leader_l    -5.927882e-02 5.046107e-02 -1.174743 0.24071313
## incumbency           2.018822e-02 3.043486e-02  0.663325 0.50745879
## enp_all              8.508674e-03 1.735848e-02  0.490174 0.62424761
## golkar_share_all    -3.369140e-01 1.719618e-01 -1.959238 0.05069653 .
## pdip_share_all       1.426236e-03 2.404505e-01  0.005932 0.99526997
## services_provision  -2.504004e-02 1.124148e-02 -2.227468 0.02640545 *
## rev_natural_pc_l     7.982000e-08 5.081000e-08  1.570822 0.11692056
## gini_l              -2.128116e-03 2.642332e-03 -0.805393 0.42101445
## rev_total_pc2_l      6.920000e-09 8.640000e-09  0.801445 0.42329347
## lpop_l              -3.461378e-01 1.881608e-01 -1.839585 0.06648192 .
## poverty_pc_l         4.479169e-02 1.446521e-01  0.309651 0.75696826
## lgdppc_l             9.434559e-02 1.312579e-01  0.718780 0.47264614
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.35957     Adj. R2: 0.870228
##                 Within R2: 0.02391
```

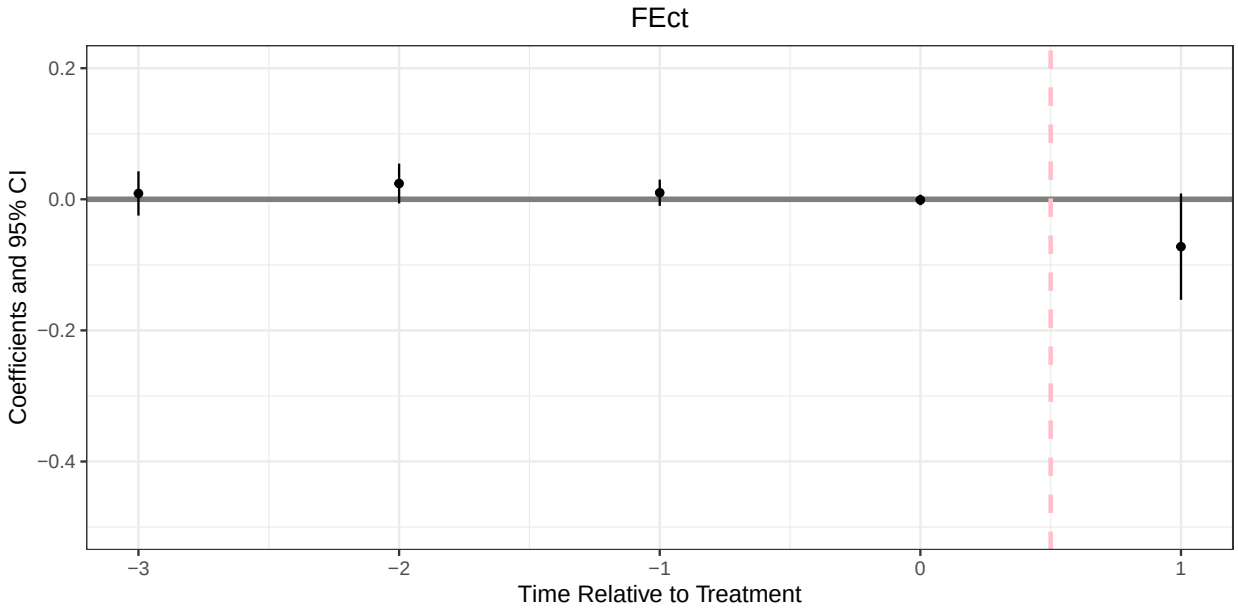


The estimate from FEct is smaller than that from TWFE and seems to be underpowered.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.



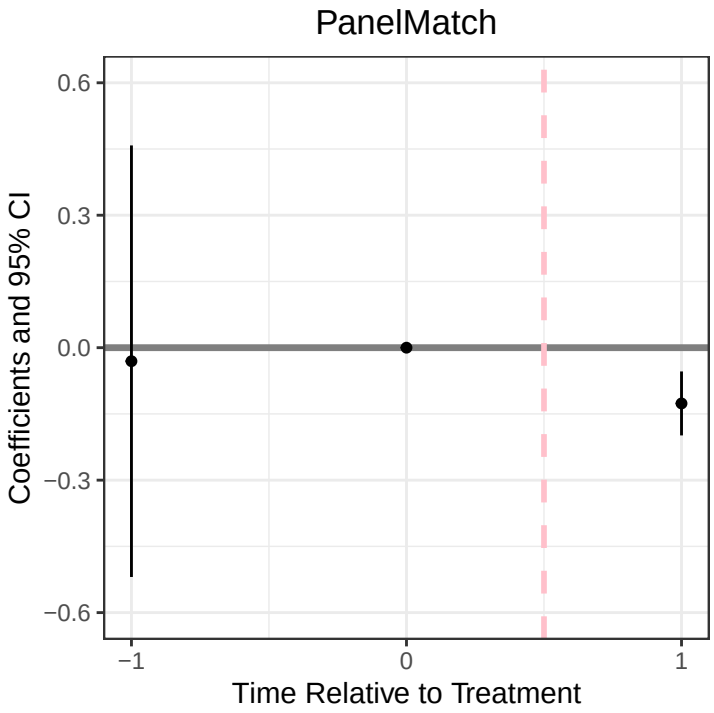


The two sets of estimates are broadly consistent.

DTE of a Balanced Subsample

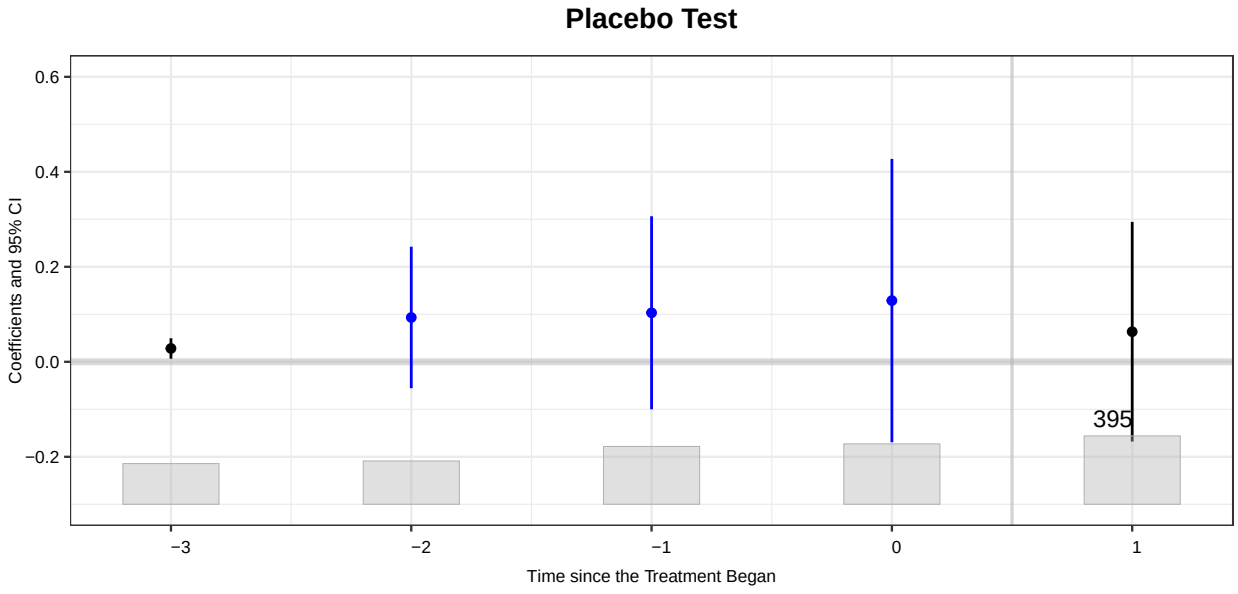
We also estimate the DTE using PanelMatch (lead = 1 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding

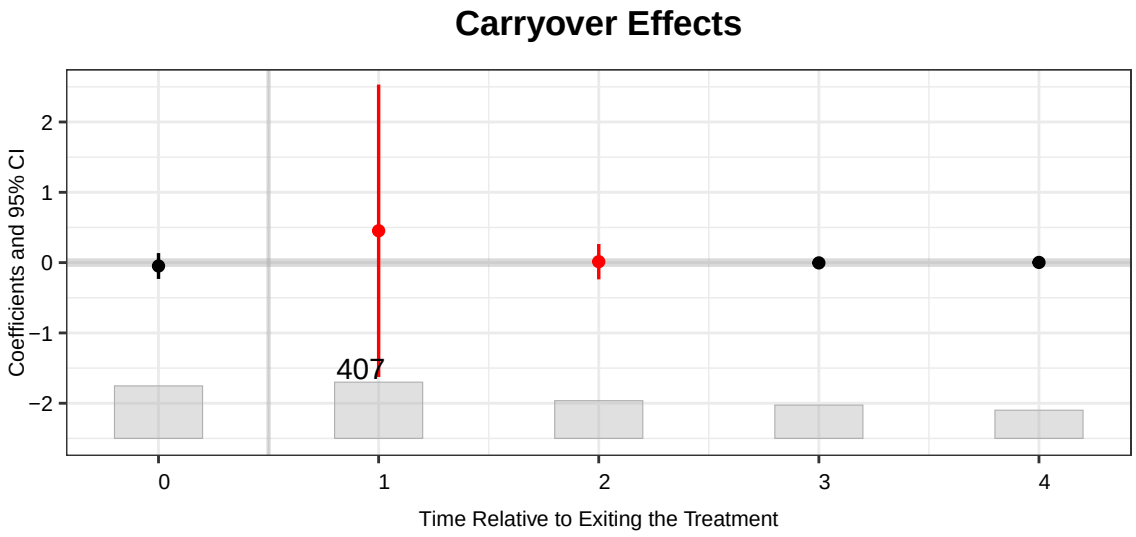
the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.





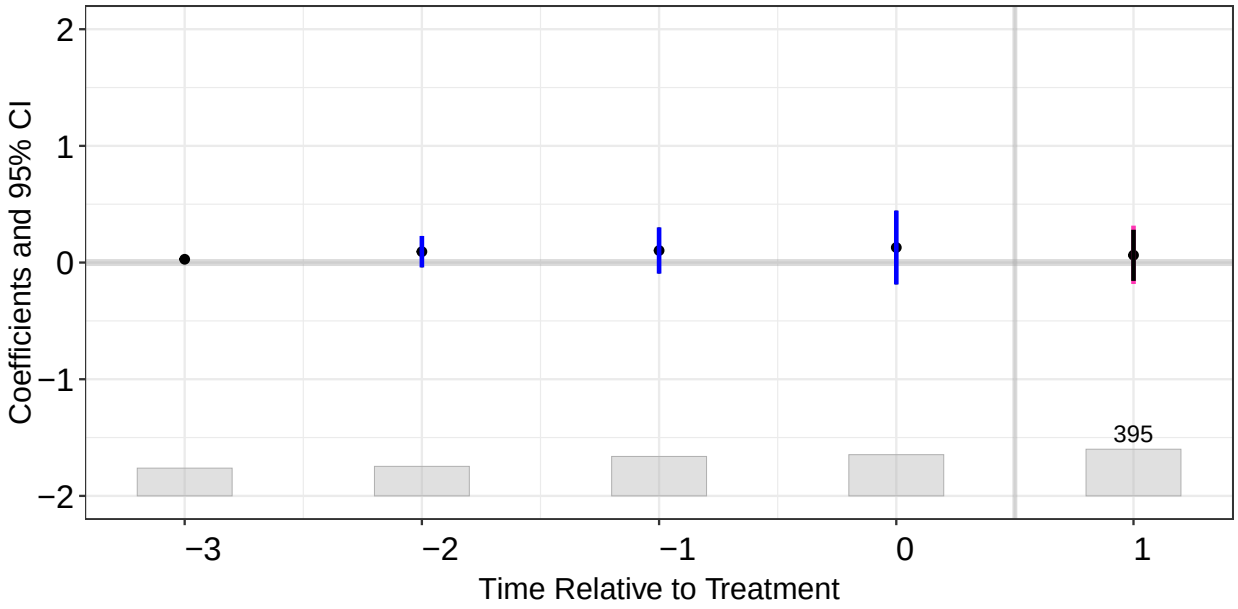
Test Statistics

```
##                                  p-value
## F test                             0.006
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.001
## Placebo test                       0.289
## Carryover effect test              0.685
```

We find some evidence of violation of the parallel trend (PT) assumption in the F-test, but the equivalence test rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

The imputation estimator (FEct) produces a point estimate similar to it of the TWFE model; however, this estimate is no longer statistically significant. There is no strong evidence of potential PT violations, anticipation effects or carryover effects. However, due to limited power, we cannot reject the null hypothesis that the ATT differs from zero under mild PT violations benchmarked against the pre-treatment placebo estimates.

Ravanilla et al. (2022)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Summary 4

A First Look at Data

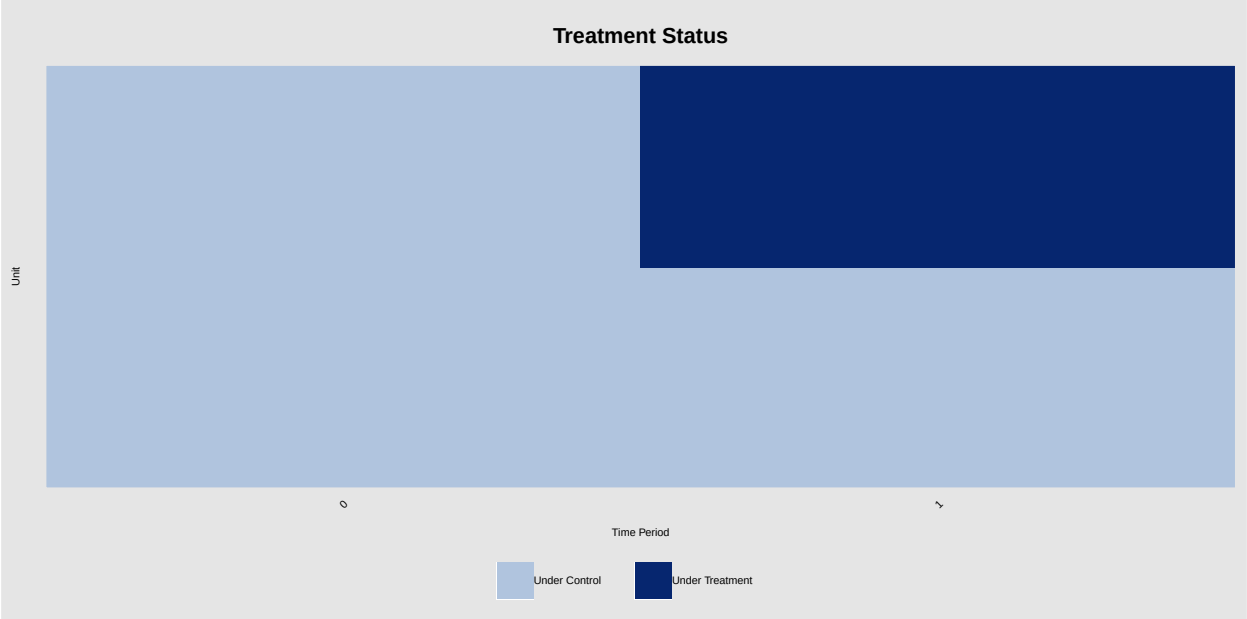
The paper investigates the effects of mayors from independent or minority parties on the execution of President Rodrigo Duterte's War on Drugs, using Filipino municipality-year level panel data before and after Duterte's first day of office. One of the main findings of this paper is that "in municipalities with an outsider mayor, there are about 40% more drug-related crimes reported during the postperiod than those with insider mayoralties" (p1048-1049, Table 3).

Model. We focus on **Model 1 of Table 3** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the municipality level.

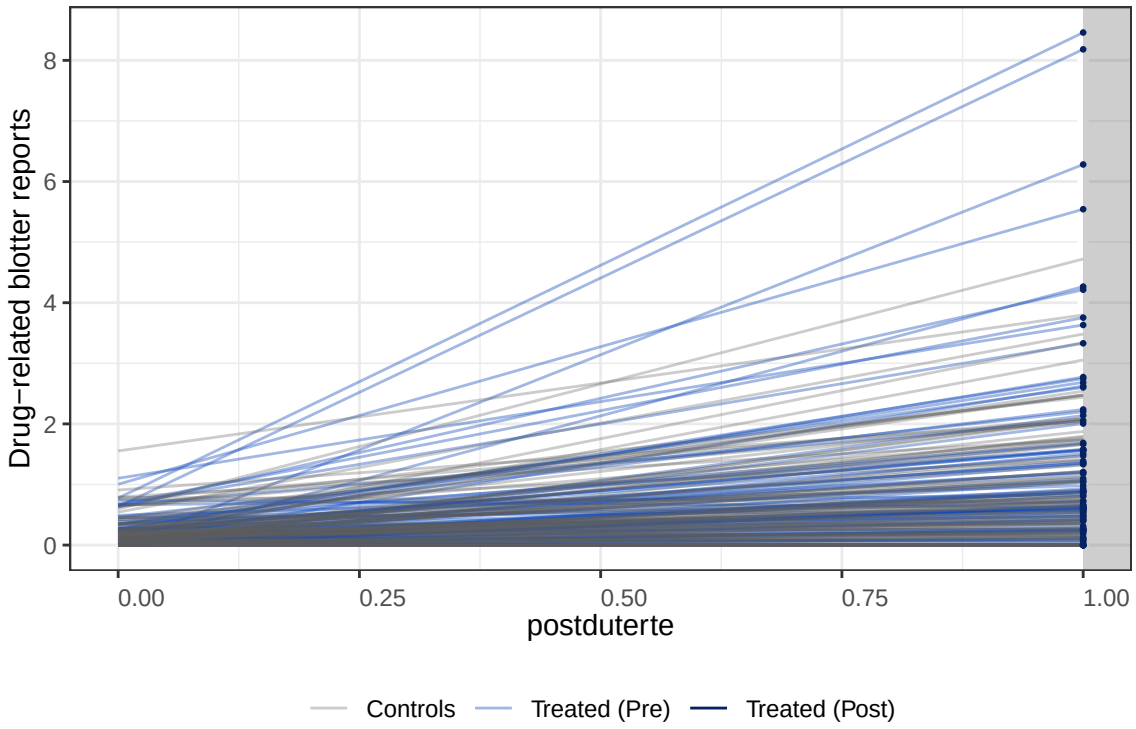
Table 1: Replication Summary

Unit of analysis	Municipality × period
Treatment	Outsider mayors after Duterte's first day in office
Outcome	Drug-related blotter reports
Treatment type	Classic
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period and each row represents a unit (a municipality).



Plotting the outcome variable We plot the trajectory of the outcome variable for each municipality. The observations under treated status are marked in blue.



Point Estimates

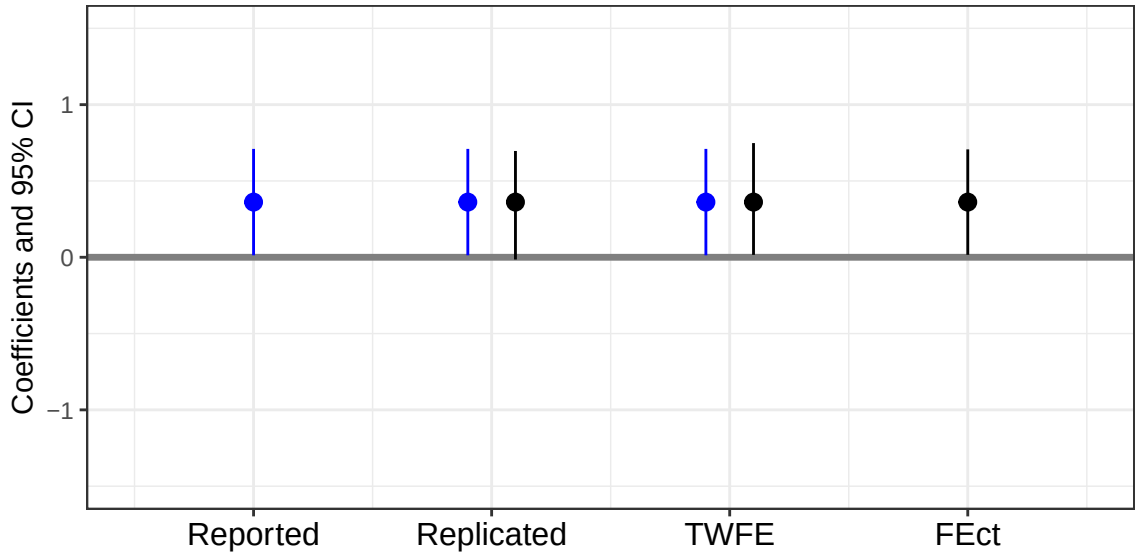
We first present the regression result following the authors' original specification. We then drop the always-treated units (there is none in this data) and apply two estimators: TWFE and the imputation method

(FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Results

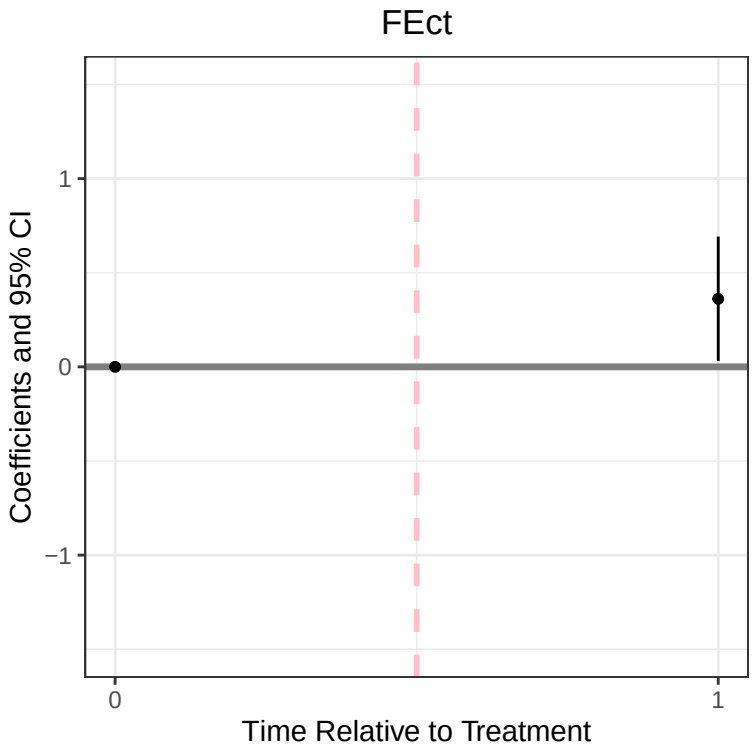
```
sol <- feols(drug_related_pcap~interaction|municipality+postduterte,data = df,cluster = "municipality")
summary(sol)
```

```
## OLS estimation, Dep. Var.: drug_related_pcap
## Observations: 378
## Fixed-effects: municipality: 189,  postduterte: 2
## Standard-errors: Clustered (municipality)
##             Estimate Std. Error t value Pr(>|t|)
## interaction 0.361468   0.177847 2.03247 0.043513 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.596392     Adj. R2: 0.364709
##                  Within R2: 0.022414
```



As this data poses a classic difference-in-differences pattern. The TWFE and FEct estimator give the same result. The estimated ATT are statistically significant when cluster-robust SEs or cluster-bootstrap SEs are being used.

Dynamic Treatment Effects



Summary

Overall, the main result of the chosen model seems to be robust to bootstrapped SEs. Because there is one pre-treatment period, it is difficult for us to evaluate whether the parallel trends (PT) assumption is plausible.

Sanford (2023)

Contents

A First Look at Data 1
Point Estimates 3
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 6
Summary 7

A First Look at Data

The paper investigates the effects of democratization on deforestation, using forested Cell-year level panel data between 1982 and 2016. One of the main findings of this paper is that "forested areas have about 1 percentage point greater forest cover loss during democratic years compared to the nondemocratic years." (p755, Table 1)

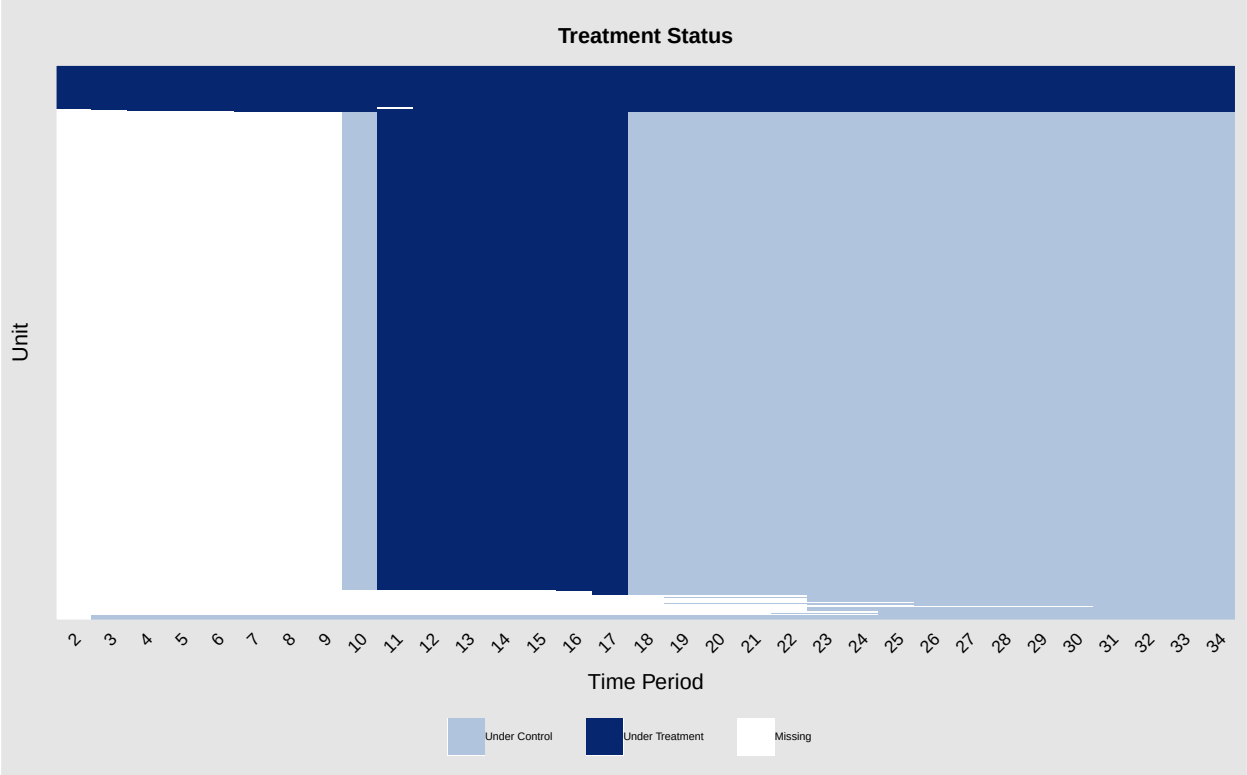
Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the agency level.

Table 1: Replication Summary

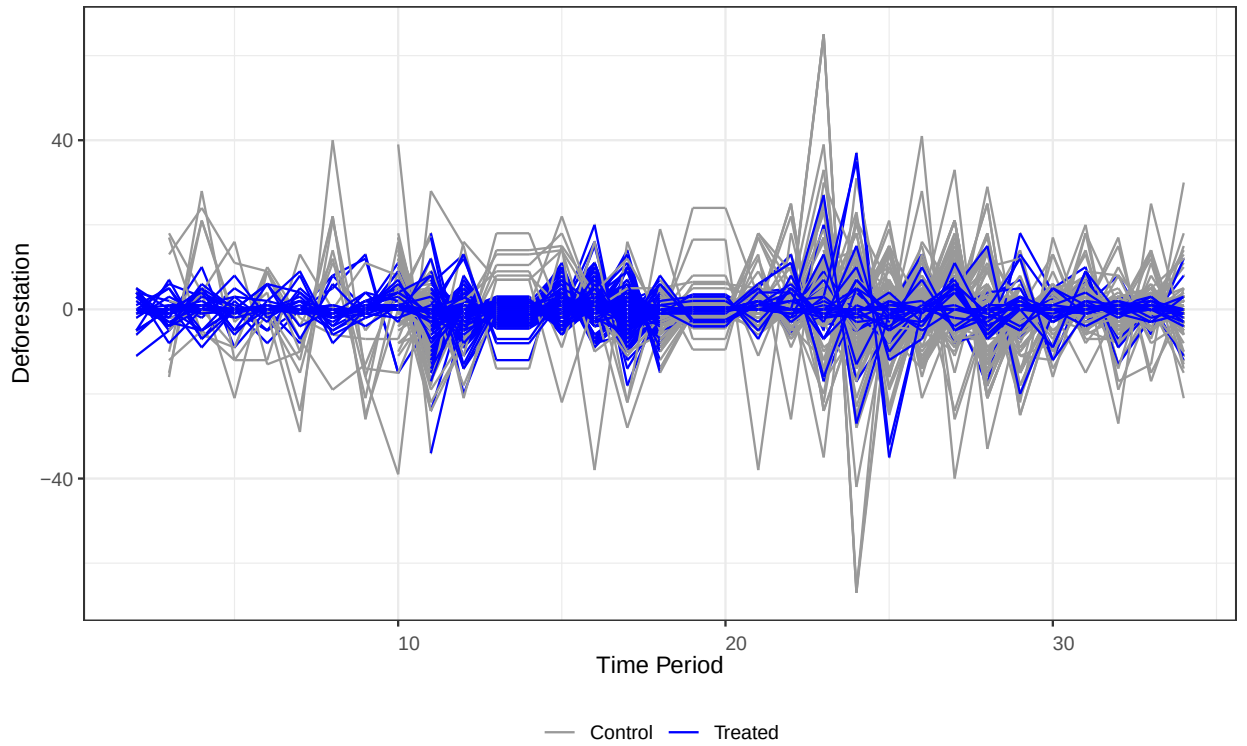
Unit of analysis	Forested Cell × year
Treatment	Democratization
Outcome	Deforestation
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (an agency). The treatment has reversals.

#View the outcome



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



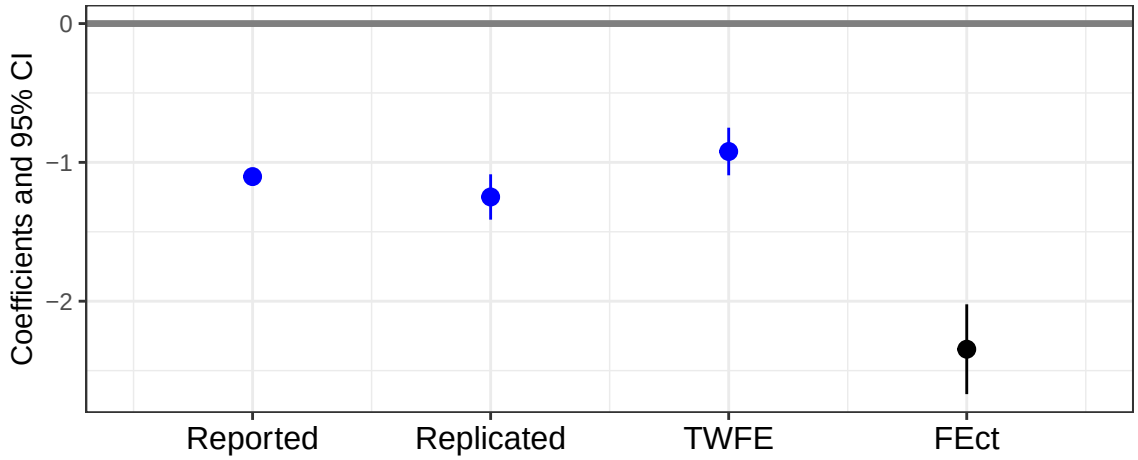
Point Estimates

We first present the regression result following the authors' original specification. As the original dataset is super large, we randomly draw a 0.2% sub-sample to conduct our analysis. We drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

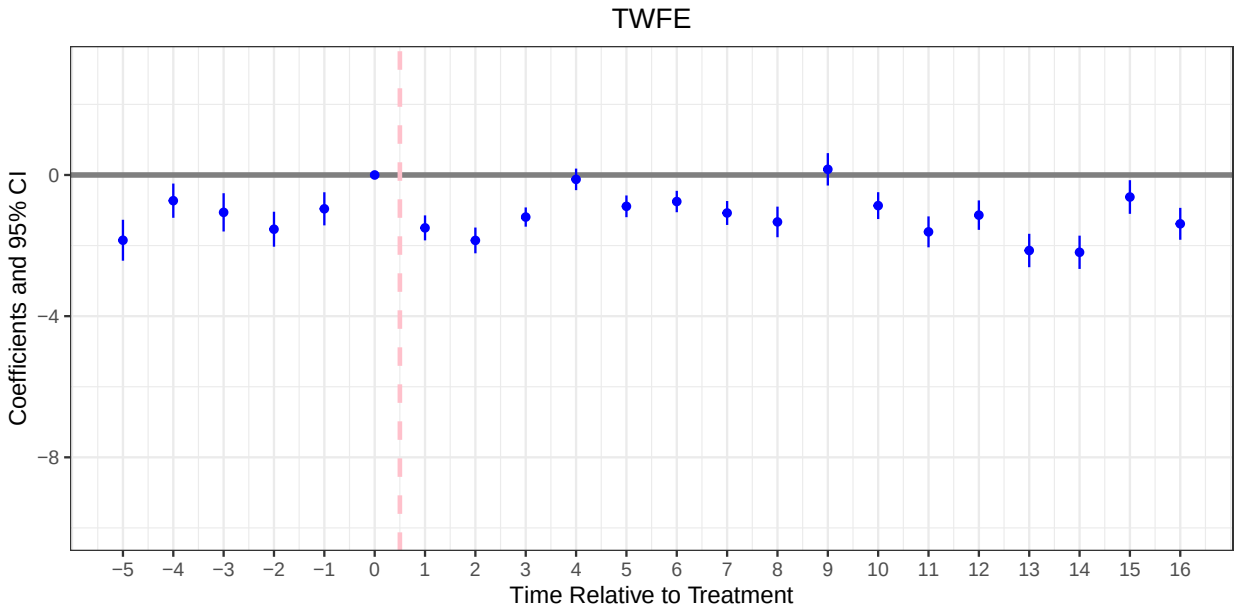
```
sol <- feols(forest.diff~democracy_BX+forest.l+PCGDP.l+PCGDP.change.l+Pop.growth.l|FID+year,
             data = df,cluster = "FID")
summary(sol)
```

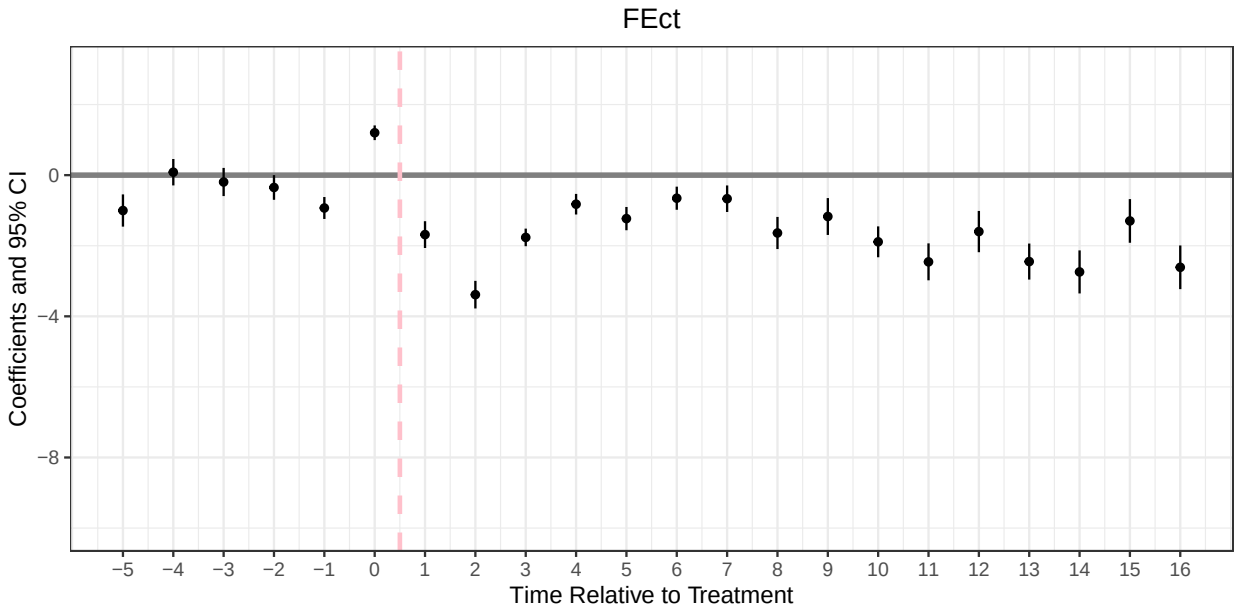
```
## OLS estimation, Dep. Var.: forest.diff
## Observations: 136,743,524
## Fixed-effects: FID: 4,588,157,  year: 33
## Standard-errors: Clustered (FID)
##                  Estimate Std. Error    t value  Pr(>|t|)
## democracy_BX   -1.102446   0.003720  -296.3635 < 2.2e-16 ***
## forest.l       -0.770400   0.000183 -4218.2187 < 2.2e-16 ***
## PCGDP.l         0.079487   0.000383   207.6031 < 2.2e-16 ***
## PCGDP.change.l -7.859223   0.190876   -41.1744 < 2.2e-16 ***
## Pop.growth.l   -0.129144   0.002188   -59.0237 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 7.25483     Adj. R2: 0.37714
##                  Within R2: 0.387479
```



Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct.

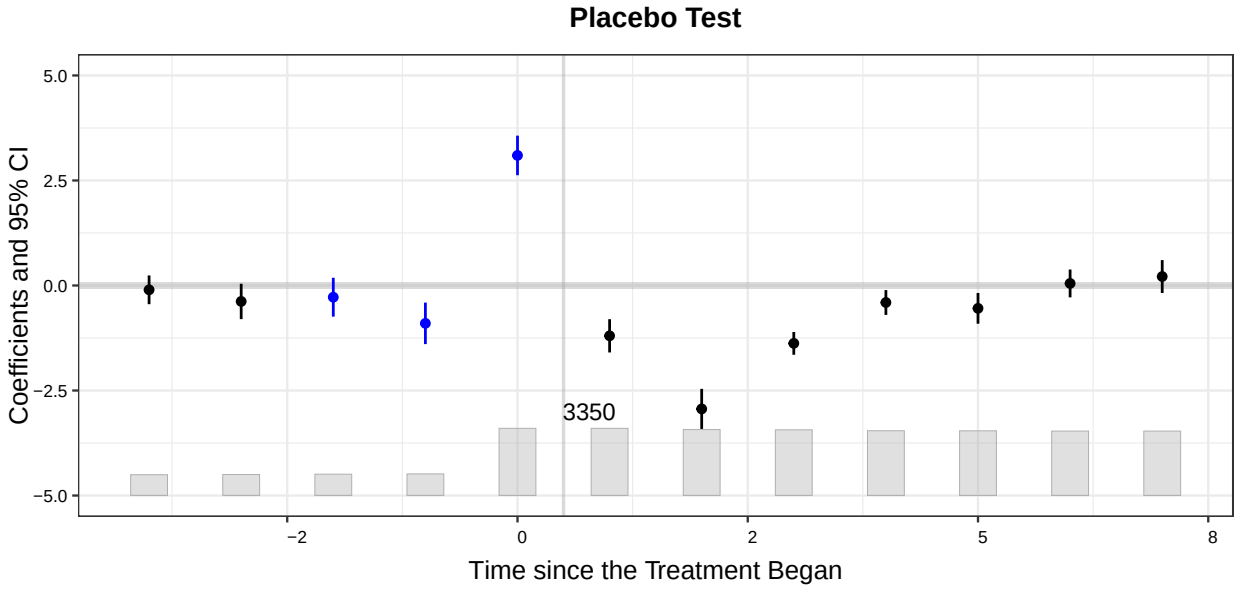


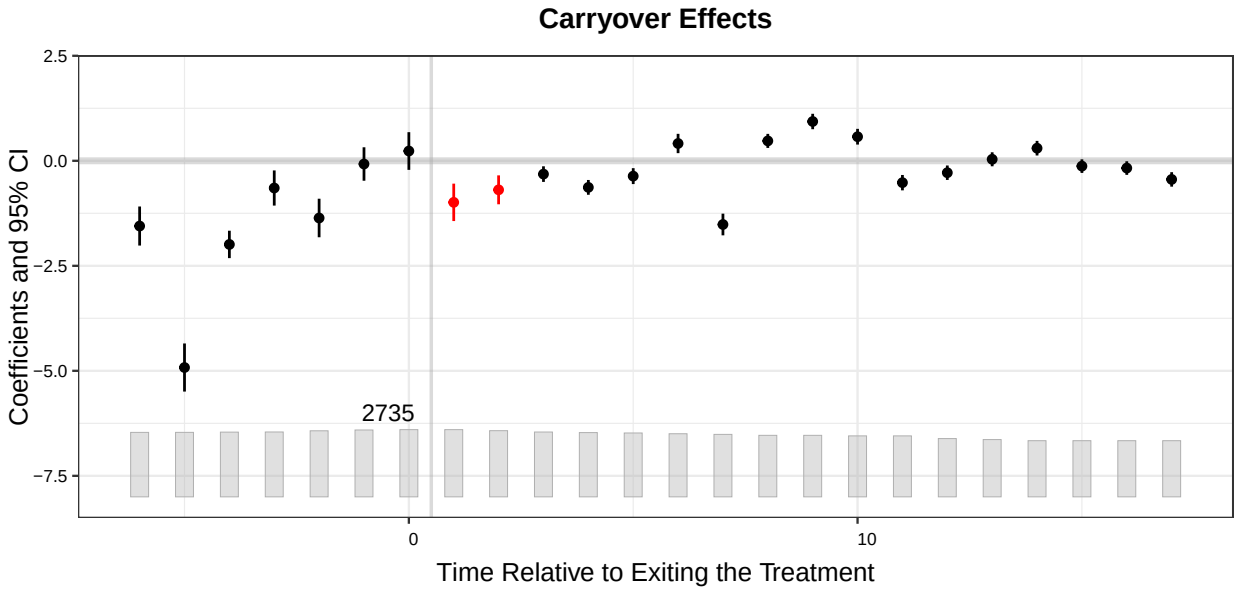


TWFE and FEct yield different results, and both show sort of pre-trend.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.





Test Statistics

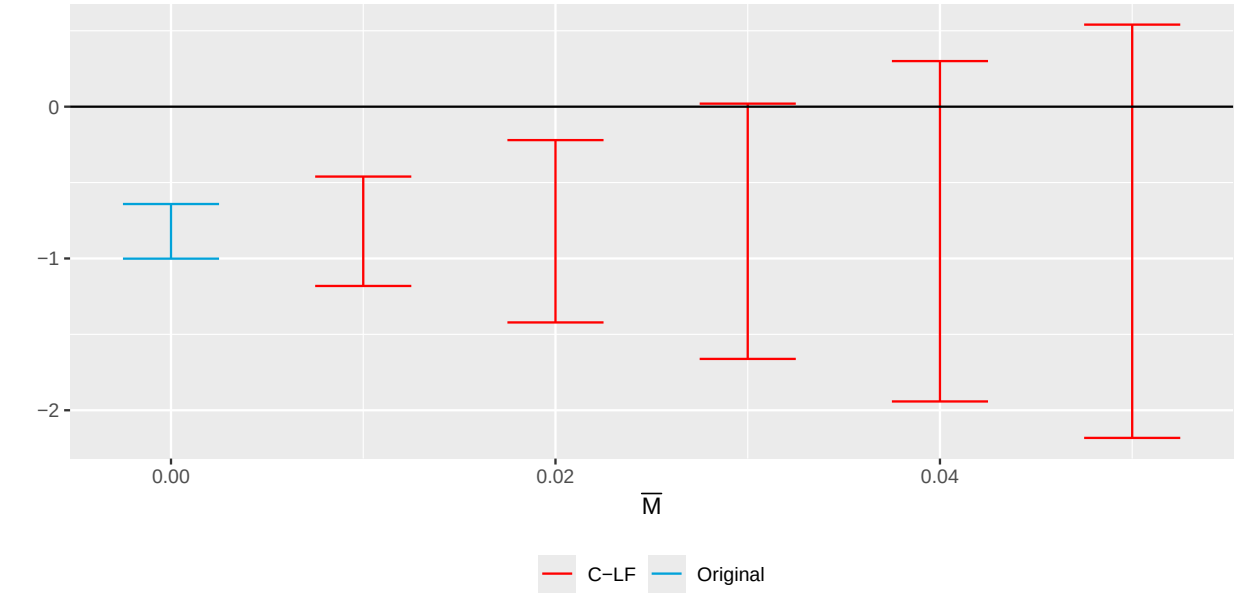
```
##                                    p-value
## F test                                   0
## Equivalence test (default)               0
## Equivalence test (threshold=ATT)         0
## Placebo test                             0
## Carryover effect test                    0
```

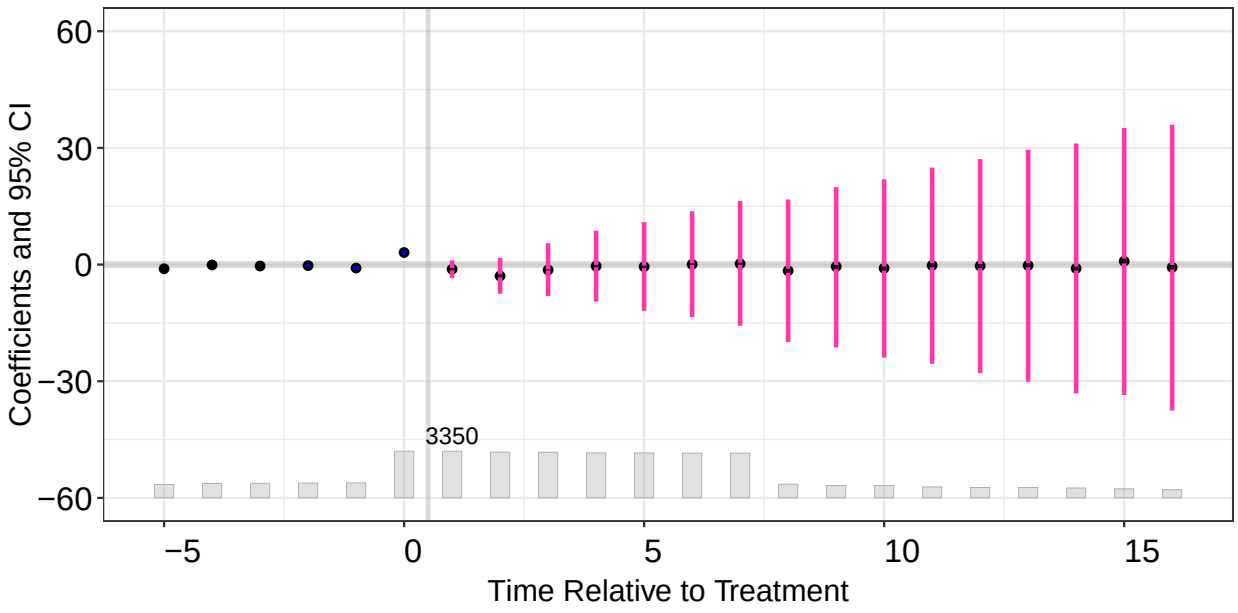
Both of the F-test and the placebo test indicate both violations of the parallel trend assumption (PT). However, the equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT, suggesting the violations may be inconsequential. We find some evidence of violation of the no-carryover-effect assumption.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.03$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 0.03 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result of the chosen model appears to be robust to FEct, an HTE-robust estimator, while the estimated ATT are sort of different. We find some evidence for violations of the PT, though they may be inconsequential for credible inference of the average effect. The estimated ATT is is sensitive to mild PT violations with the breakdown threshold $M = 0.03$.

Schafer et al. (2022)

Contents

A First Look at Data 1

Point Estimates 3

Dynamic Treatment Effects 3

Diagnostic Tests 4

Summary 6

A First Look at Data

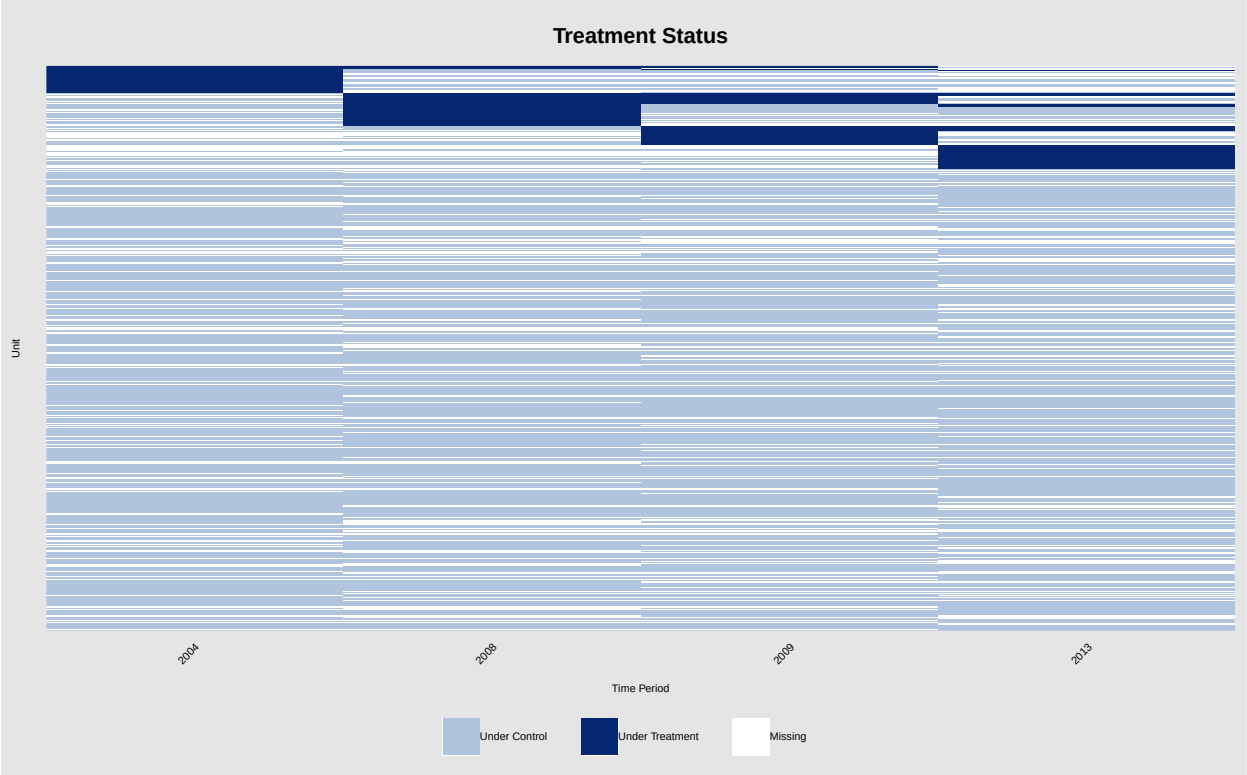
The paper investigates the effects of income shocks on voter turnout, using Italian individual-election year level panel data between 2004 and 2013. One of the main findings of this paper is that "voter turnout decreases by 3 percentage points (p.p.) when households stop earning any taxable income" (p746).

Model. We focus on **Model 3, column 2, of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model with an addition of the age fixed effect and report robust standard errors clustered at the individual and household levels.

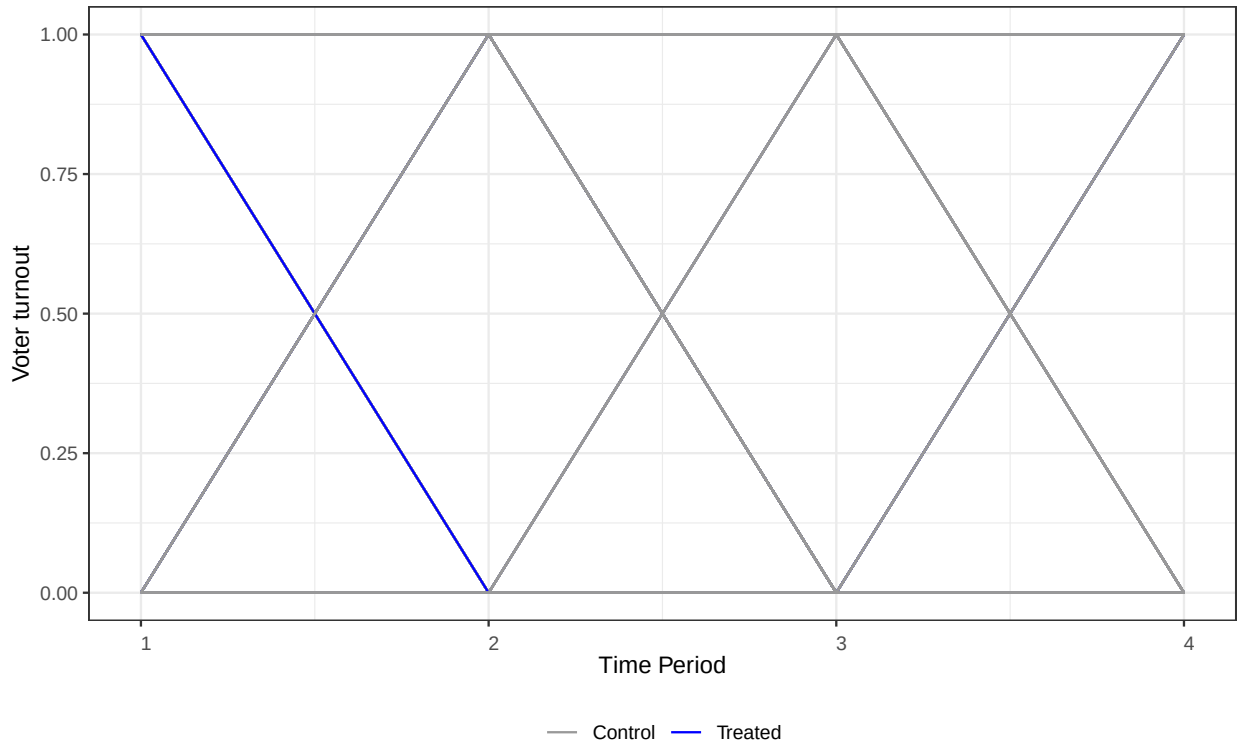
Table 1: Replication Summary

Unit of analysis	Individual × election year
Treatment	Negative household income shock
Outcome	Voter turnout
Treatment type	General
Outcome type	Binary
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (an election year) and each row represents a unit (an individual). We see that a small number of units are treated at various time points and there are treatment reversals. There are some missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each municipality. The observations under treated status are marked in blue.



Point Estimates

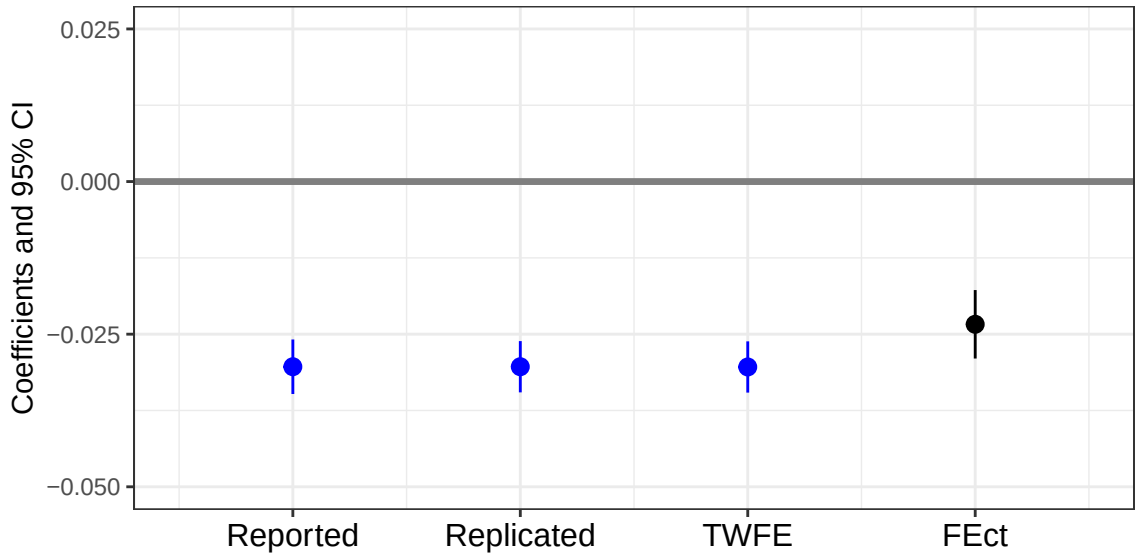
We first replicate the authors' using the original speicification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(dv~treatment_alt|ID_voter+year+age,data = df,cluster = c("ID_voter", "ID_hh"))
summary(sol)
```

```
## OLS estimation, Dep. Var.: dv
## Observations: 1,163,307
## Fixed-effects: ID_voter: 381,256,  year: 4,  age: 92
## Standard-errors: Clustered (ID_voter & ID_hh)
##                Estimate Std. Error  t value  Pr(>|t|)
## treatment_alt -0.030332   0.002277 -13.3185 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.235246     Adj. R2: 0.432743
##                  Within R2: 3.781e-4
```

```
##   |                                                      |
```

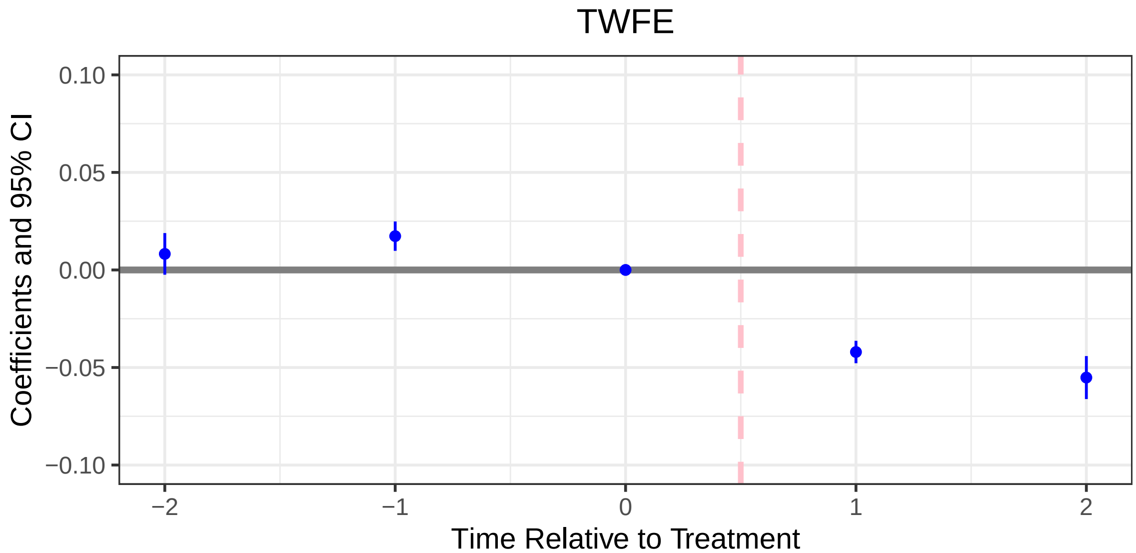


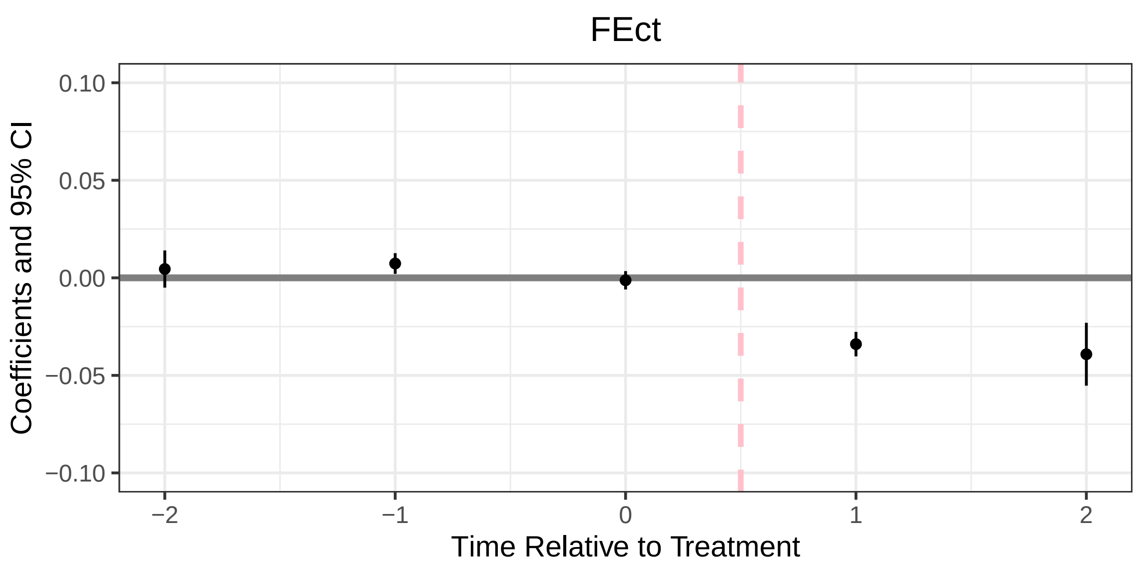
The TWFE and FEct estimates are consistent with each other. The estimated ATTs are negative and statistically significant.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTE) and and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.

| |





TWFE and FEct are consistent, and the DTE estimates are all negative during post-treatment periods.

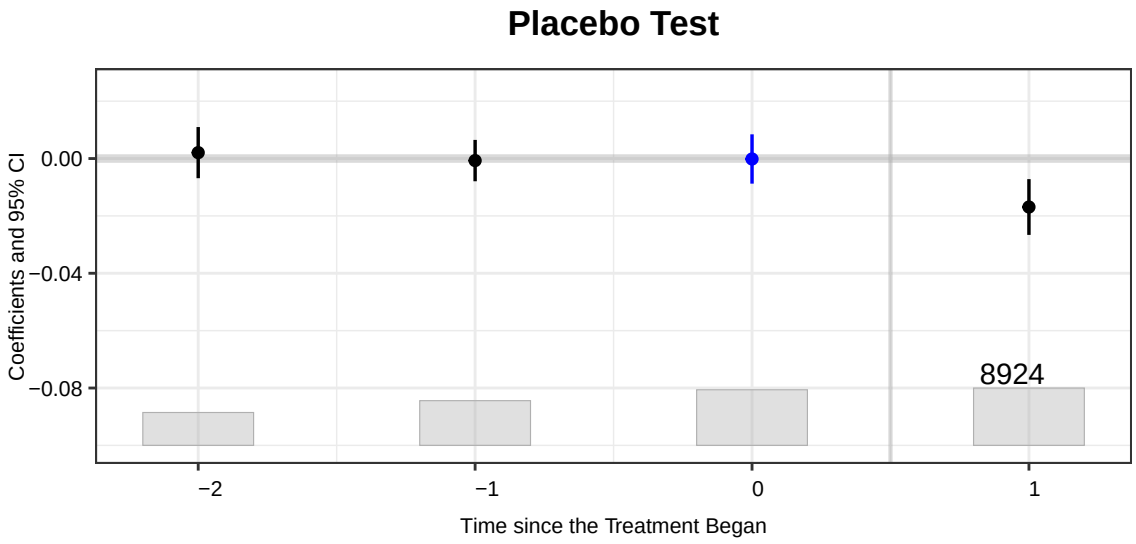
Diagnostic Tests

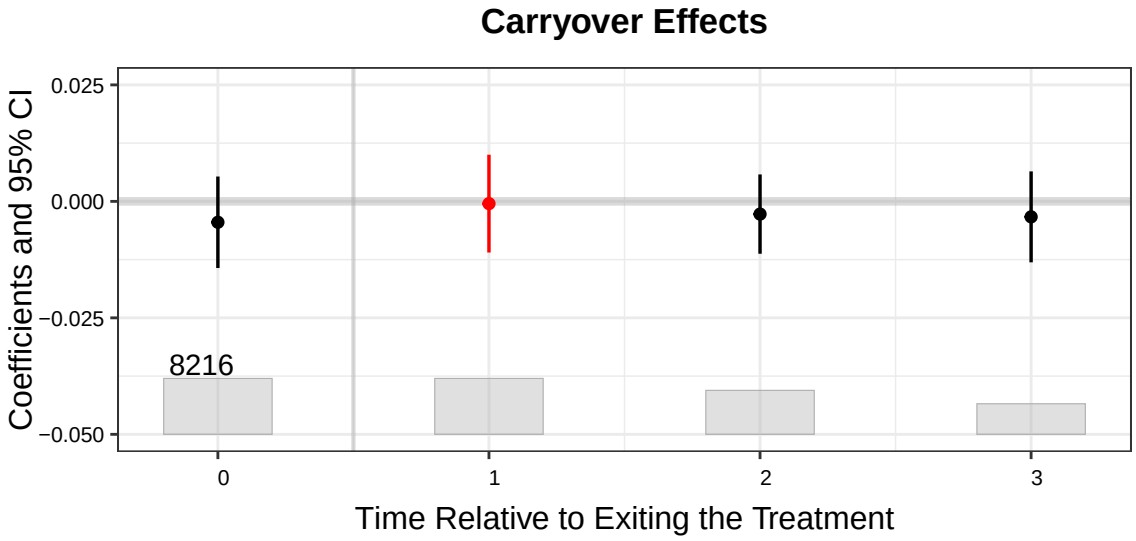
Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.

| |

| |

| |





Test Statistics

```
## Cannot use full pre-treatment periods in the F test. The first period is removed.
##
## Cannot use full pre-treatment periods in the F test. The first period is removed.
##
## Cannot use full pre-treatment periods in the F test. The first period is removed.
```

```
##                                  p-value
## F test                             0.017
## Equivalence test (default)         0.000
## Equivalence test (threshold=ATT)   0.000
## Placebo test                       0.972
## Carryover effect test              0.928
```

We find some evidence of violations of the parallel trends (PT) assumption using the F-test but not in the placebo test. The equivalence test rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

Summary

Overall, the main result of the chosen model seems robust to HTE-robust estimators. There is some evidence for PT violations (F test), but they seem inconsequential for the credibility of the ATT estimates. We cannot conduct a sensitivity analysis due to too few pre-treatment periods.

Schubiger (2021)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Summary 4

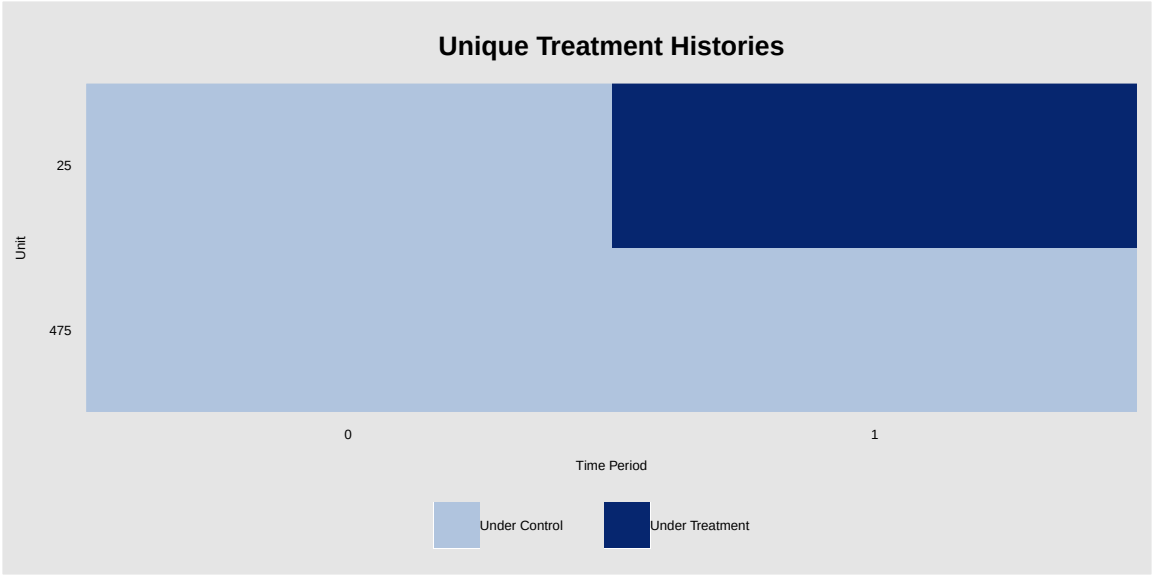
A First Look at Data

The paper investigates the effects of state violence on counterinsurgent mobilization, using Peruvian centros poblados-period panel data. One of the main findings of this paper is that "State violence has a positive and substantial effect on subsequent counterinsurgent mobilization in all specifications" (p1395).

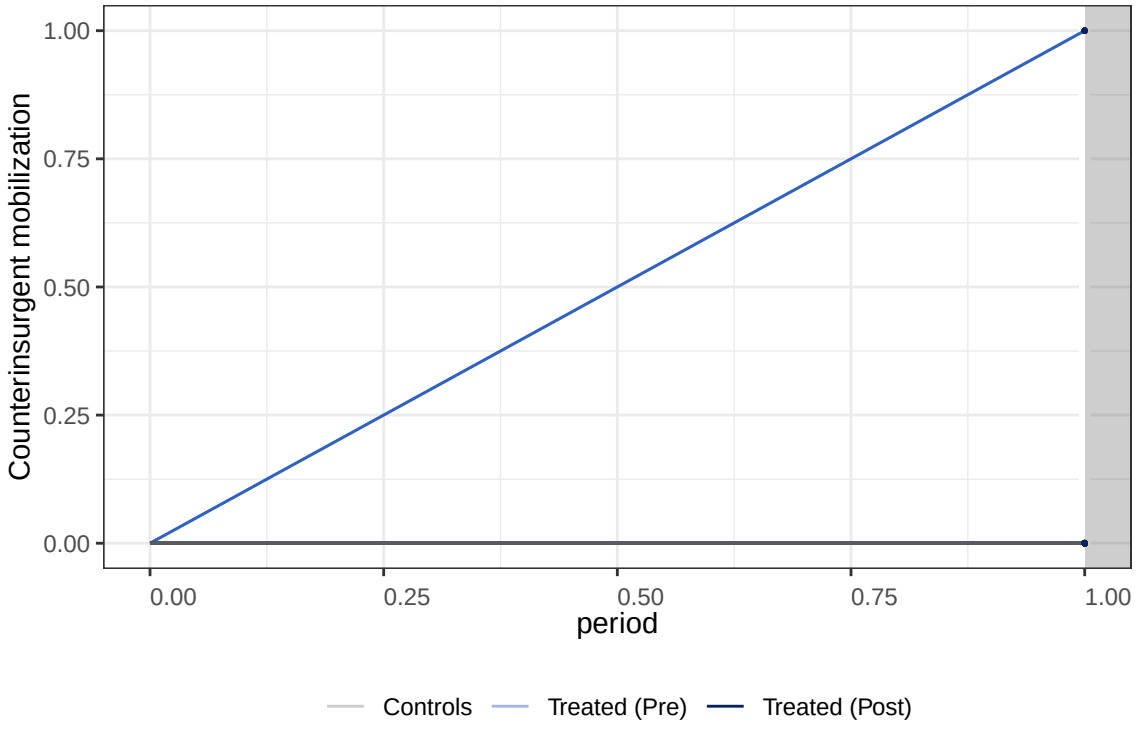
Model. We focus on **Model 1 of Table 3** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	Centros poblados × period
Treatment	State violence
Outcome	Counterinsurgent mobilization
Treatment type	Classic
Outcome type	Binary
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period and each row represents a unit. This is a classic 2 × 2 DID.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

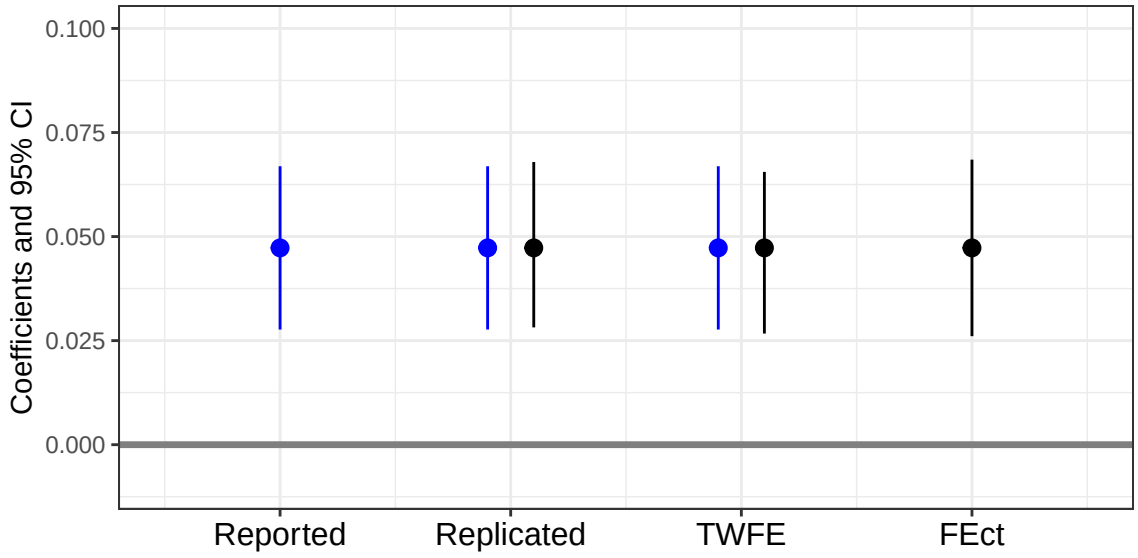
We first replicate the authors' using the original specification. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis,

we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

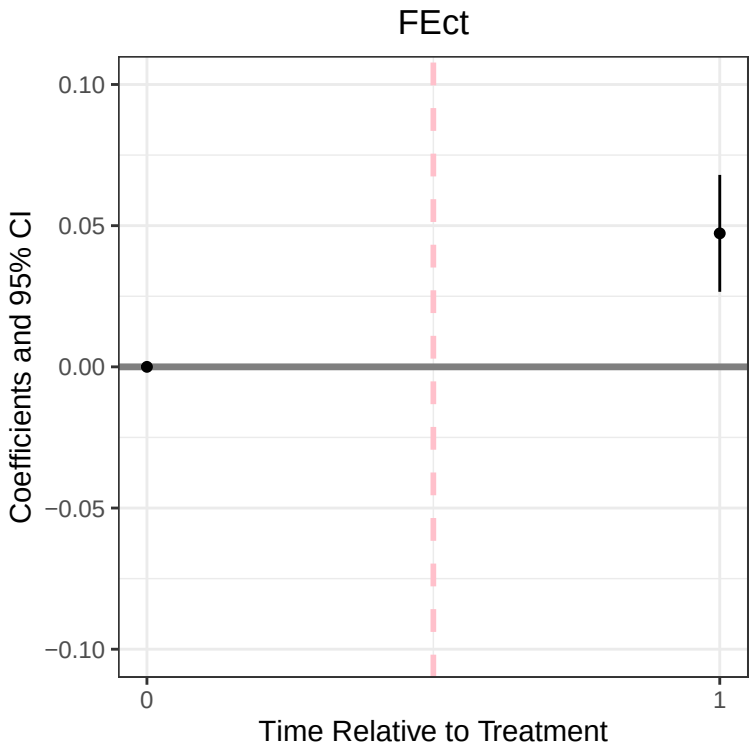
Original Finding

```
sol <- feols(autodefensa~treatment|id+period,data = df,cluster = "id")
summary(sol)
```

```
## OLS estimation, Dep. Var.: autodefensa
## Observations: 23,916
## Fixed-effects: id: 11,958,  period: 2
## Standard-errors: Clustered (id)
##           Estimate Std. Error t value   Pr(>|t|)
## treatment 0.047282   0.010009 4.72377 2.3415e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.031254     Adj. R2: 0.024553
##                  Within R2: 0.023911
```



Dynamic Treatment Effects



Summary

Not surprisingly, FEct and TWFE produce identical estimates for this paper with 2 by 2 classic DID setting. Because there is one pre-treatment period, it is difficult for us to evaluate whether the parallel trends assumption is plausible.

Schuit and Rogowski (2017)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 3
Diagnostic Tests 4
Sensitivity Analysis 6
Summary 7

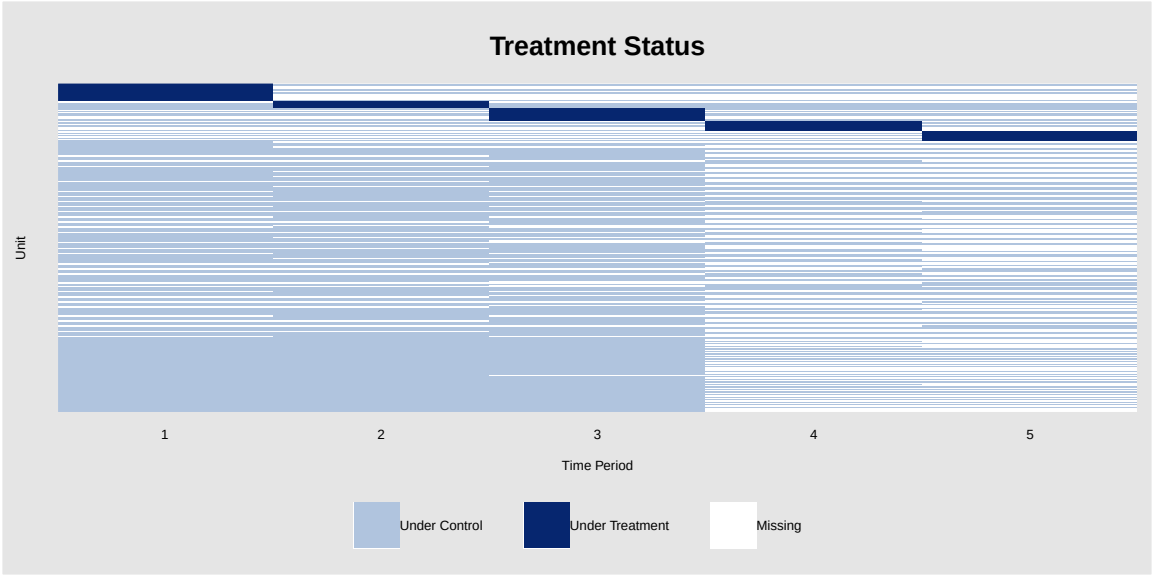
A First Look at Data

The paper investigates the effects of Voting Rights Act (VRA) on Black political representation using US legislator-congress level panel data from the 86th to the 105th congress. One of the main findings of this paper is that "members of Congress who represented jurisdictions subject to the pre-clearance requirement were substantially more supportive of civil rights-related legislation than legislators who did not represent covered jurisdictions" (p513).

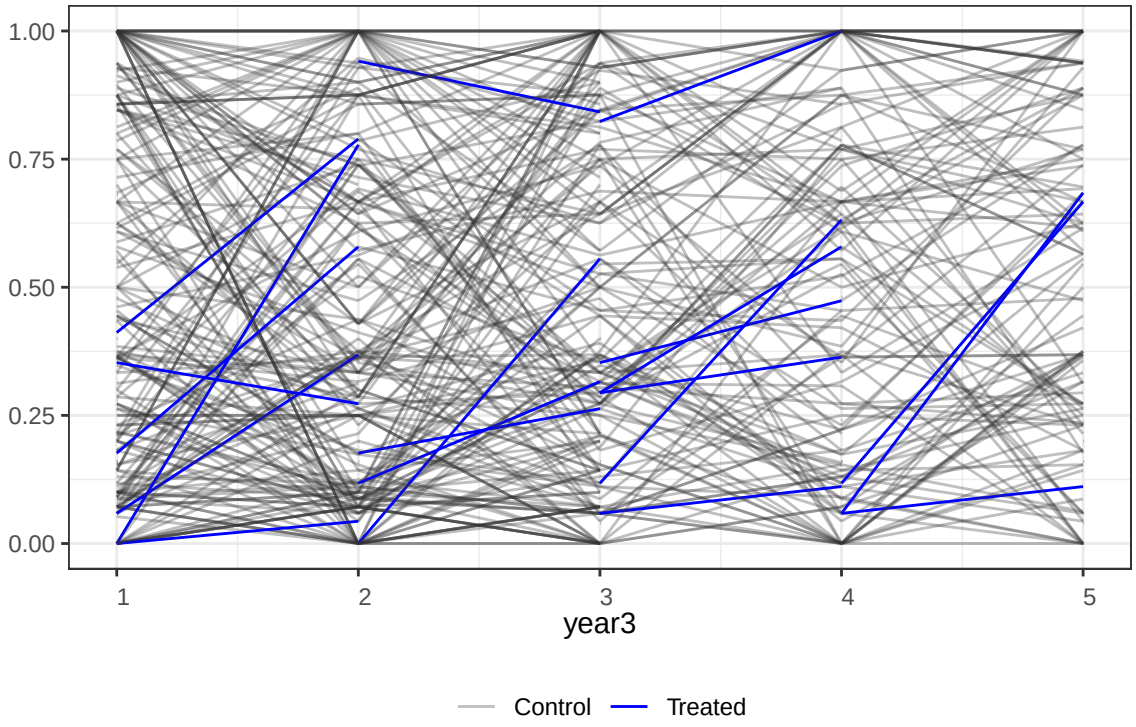
Model. Authors present an analysis using district-year level panel data of the US South. They compare legislators' voting records from Texas districts with the states of the Deep South (Alabama, Georgia, Louisiana, Mississippi, and South Carolina) using a two-way fixed effects (TWFE) model with district and congress (time) fixed effects. They find that "requiring preclearance for Texas jurisdictions was associated with a substantively large increase in support for civil rights legislation among legislators from that state relative to legislators from the Deep South." We replicate this analysis and adjust the treatment variable by interchanging the values of 0 and 1, thereby enabling us to derive estimates for the average treatment effects on the control (ATC) with a flipped sign.

Replication Summary	
Unit of analysis	District × year
Treatment	VRA coverage
Outcome	Civil rights support score
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a congress indicator) and each row represents a unit (a district). We see that a small number of units are treated at various time points and there are treatment reversals. There is some missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

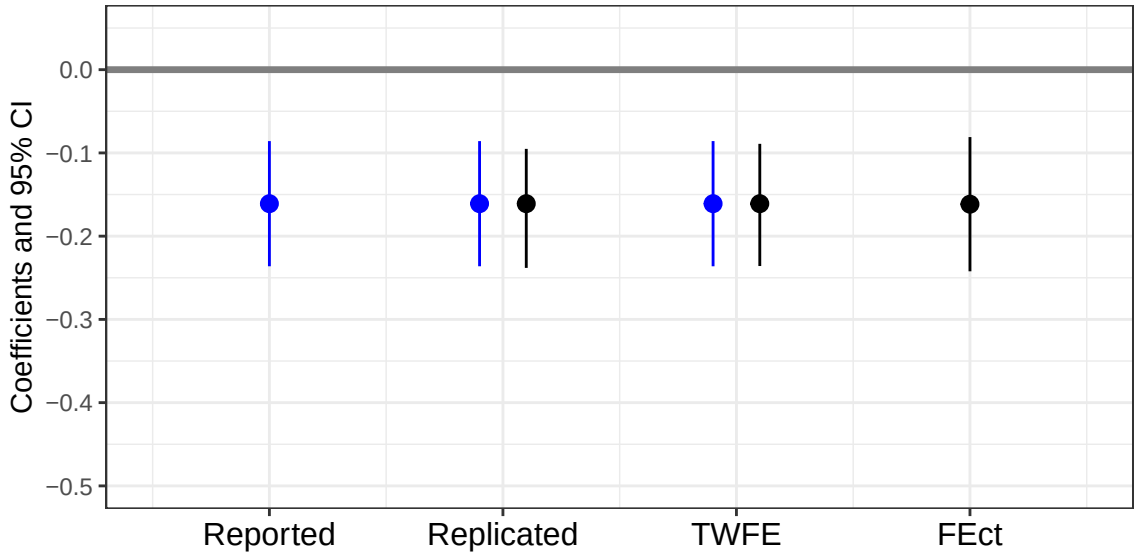
We first replicate the authors' using the original specification—we are able to successfully replicate the point estimate. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs)

are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(correct~ treatment+competitive+dpres+
               party+black_legis+percentblack| redistrict_decade + year3,
             data = df, cluster = "redistrict_decade")
summary(sol)
```

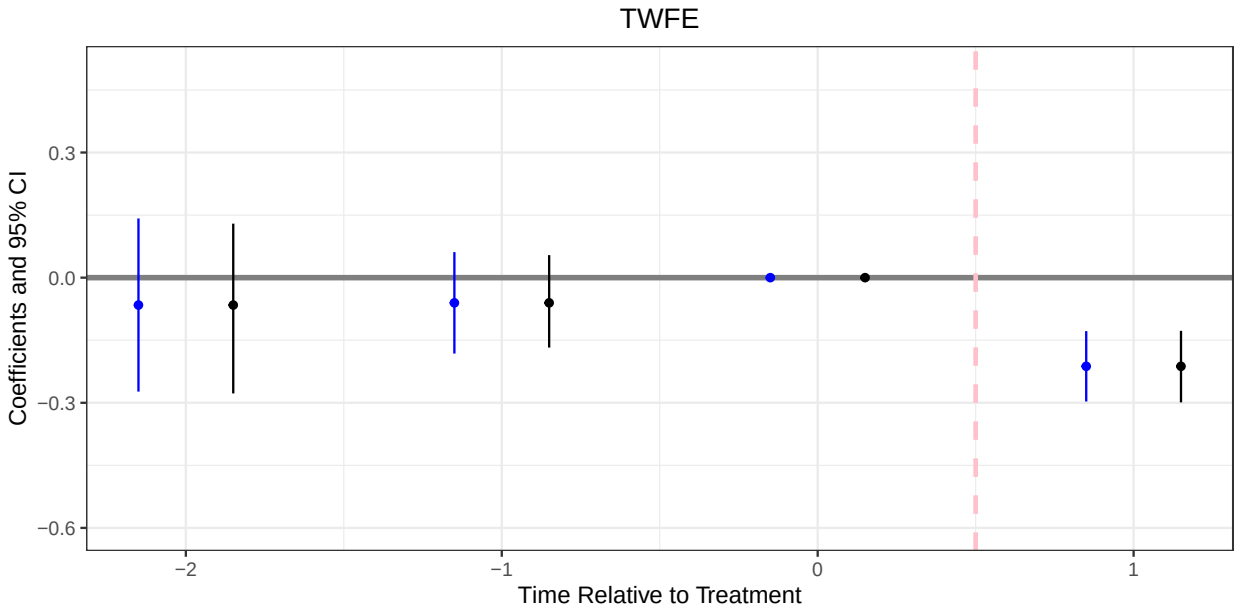
```
## OLS estimation, Dep. Var.: correct
## Observations: 902
## Fixed-effects: redistrict_decade: 261,  year3: 5
## Standard-errors: Clustered (redistrict_decade)
##                Estimate Std. Error   t value   Pr(>|t|)
## treatment    -0.160981   0.038354 -4.197275 3.7118e-05 ***
## competitive   0.085340   0.018915  4.511750 9.7442e-06 ***
## dpres         0.008271   0.010778  0.767371 4.4356e-01
## party        -0.003851   0.000538 -7.163181 8.1044e-12 ***
## black_legis   0.404061   0.095349  4.237692 3.1388e-05 ***
## percentblack  0.204240   0.323709  0.630935 5.2864e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.175361     Adj. R2: 0.624712
##                  Within R2: 0.269714
```

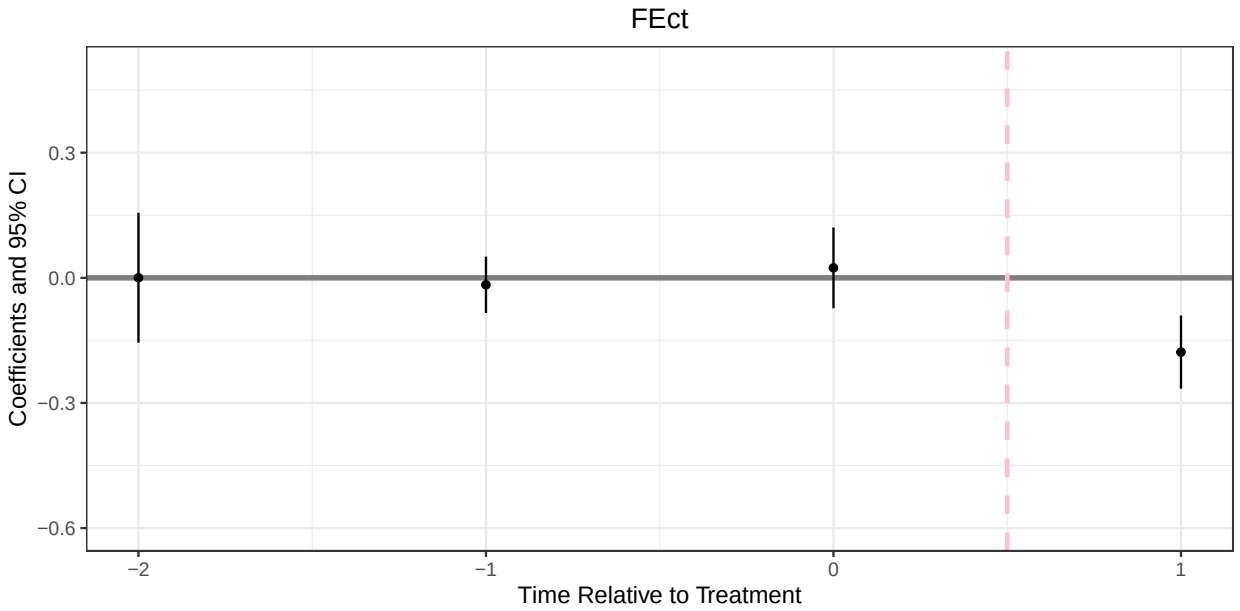


The TWFE and FEct estimates are consistent with each other and statistically significant.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTE) and obtaining corresponding event study plots. We use two estimators, TWFE and FEct. The results are shown below.

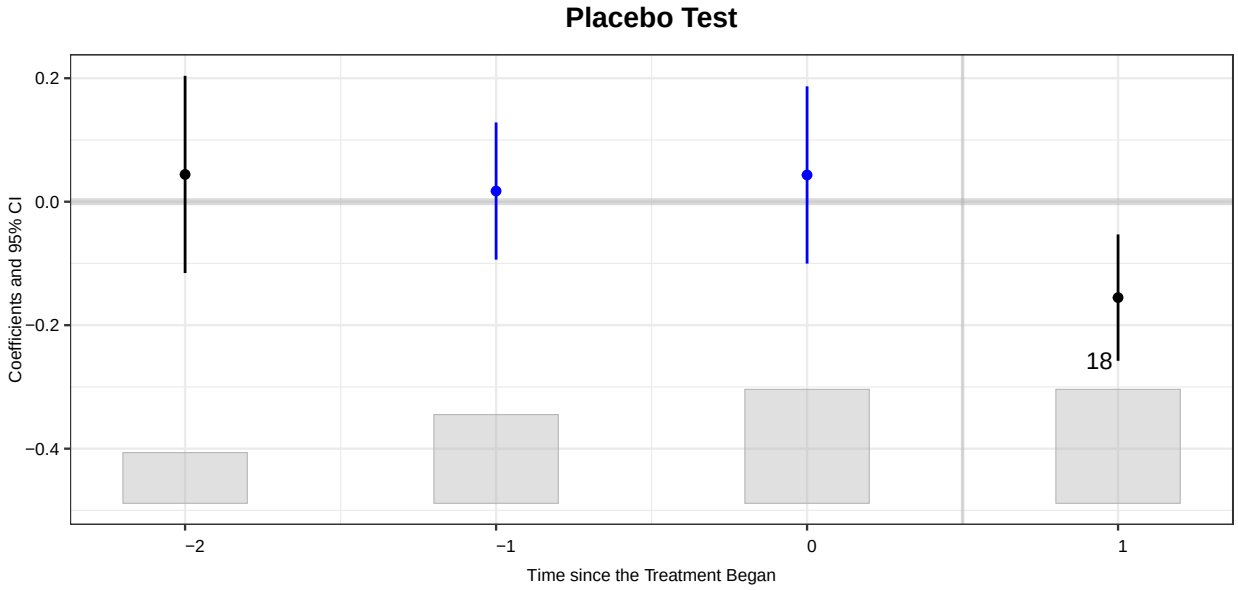


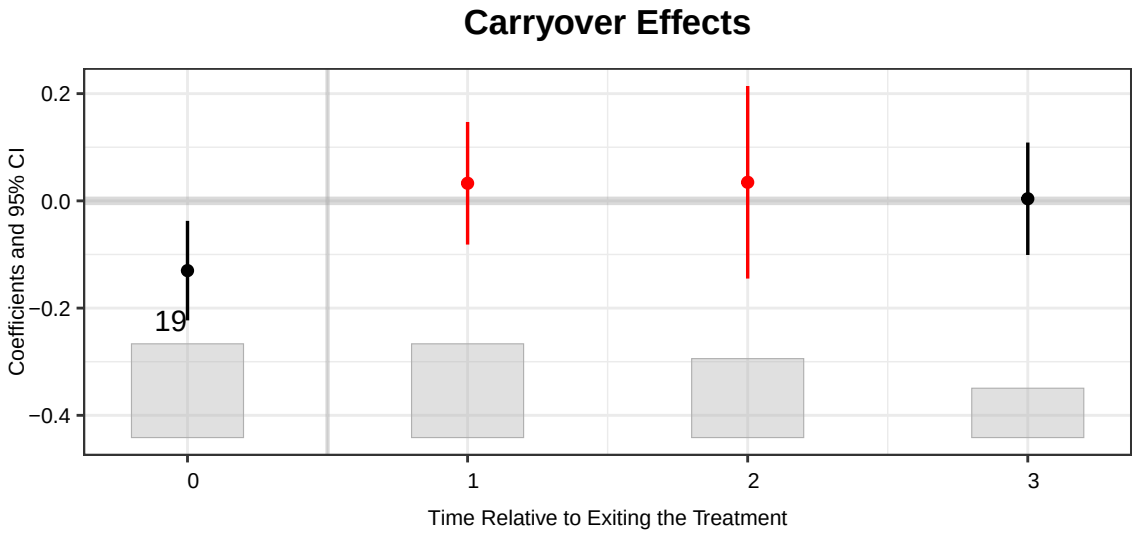


The two sets of estimates are consistent with each other, and the estimated DTEs are negative in the post-treatment period.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.





Test Statistics

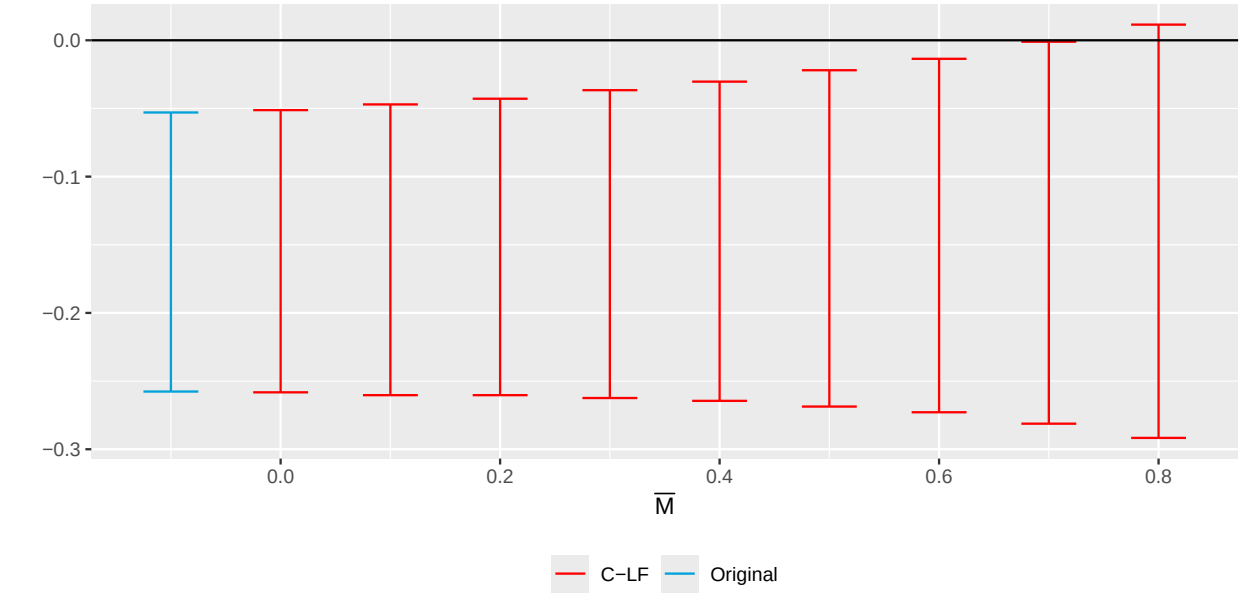
```
##                                  p-value
## F test                             0.942
## Equivalence test (default)         0.170
## Equivalence test (threshold=ATT)   0.021
## Placebo test                       0.567
## Carryover effect test              0.533
```

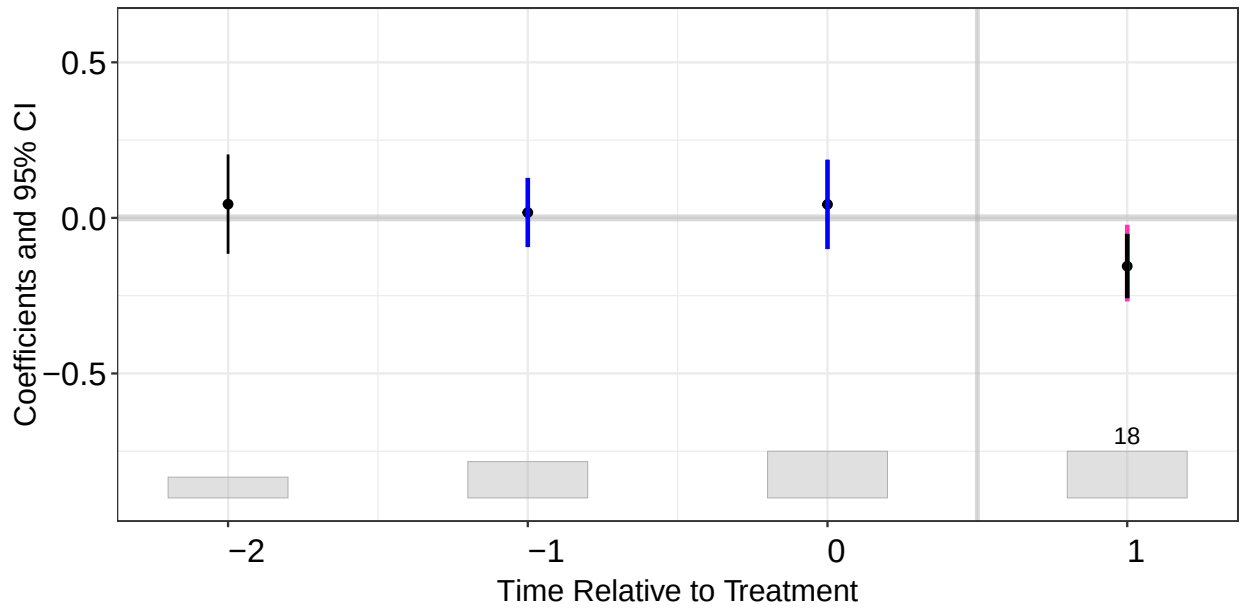
We do not find evidence for parallel trends (PT) assumption violations or carryover effects. The equivalence test also rejects the null that the residuals in pre-treatment periods exceed the estimated ATT.

Sensitivity Analysis

After holding out two pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the two placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 0.6$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for PT deviations in post-treatment periods up to 0.6 times the maximum deviation observed between consecutive periods in the placebo periods.

We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.7$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.7 times the maximum observed deviation in the placebo periods, are depicted in pink.





Summary

Overall, the main result seems robust to HTE and modeling choices. We do not find strong evidence for potential PT violations. The result also appears to be robust to mild violations of the PT with the breakdown threshold $M = 0.7$.

Skorge (2023)

Contents

A First Look at Data 1
Point Estimates 2
Diagnostic Tests 3
Summary 5

A First Look at Data

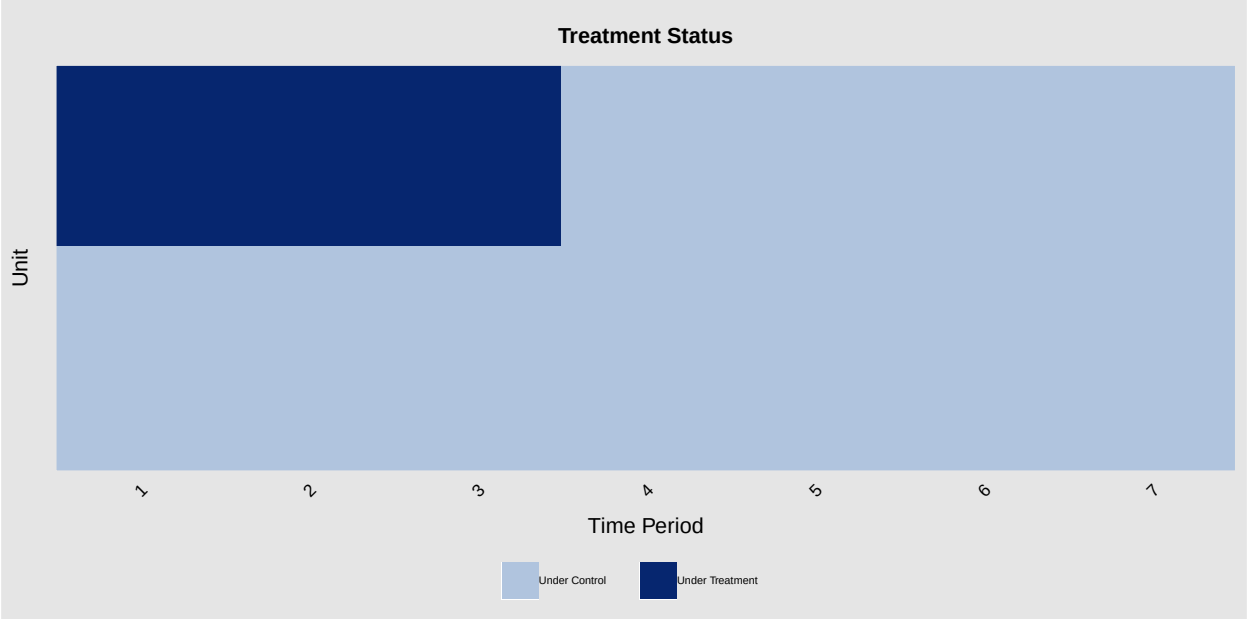
The paper investigates the effects of proportional representation (PR) reform on the fraction of female voters, using Norway municipality-election year level panel data between 1910 and 1928. One of the main findings of this paper is that the PR reform"increased the fraction female voters by an average of 0.096 points for the reformed municipalities." (p398)

Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustering at the municipality level.

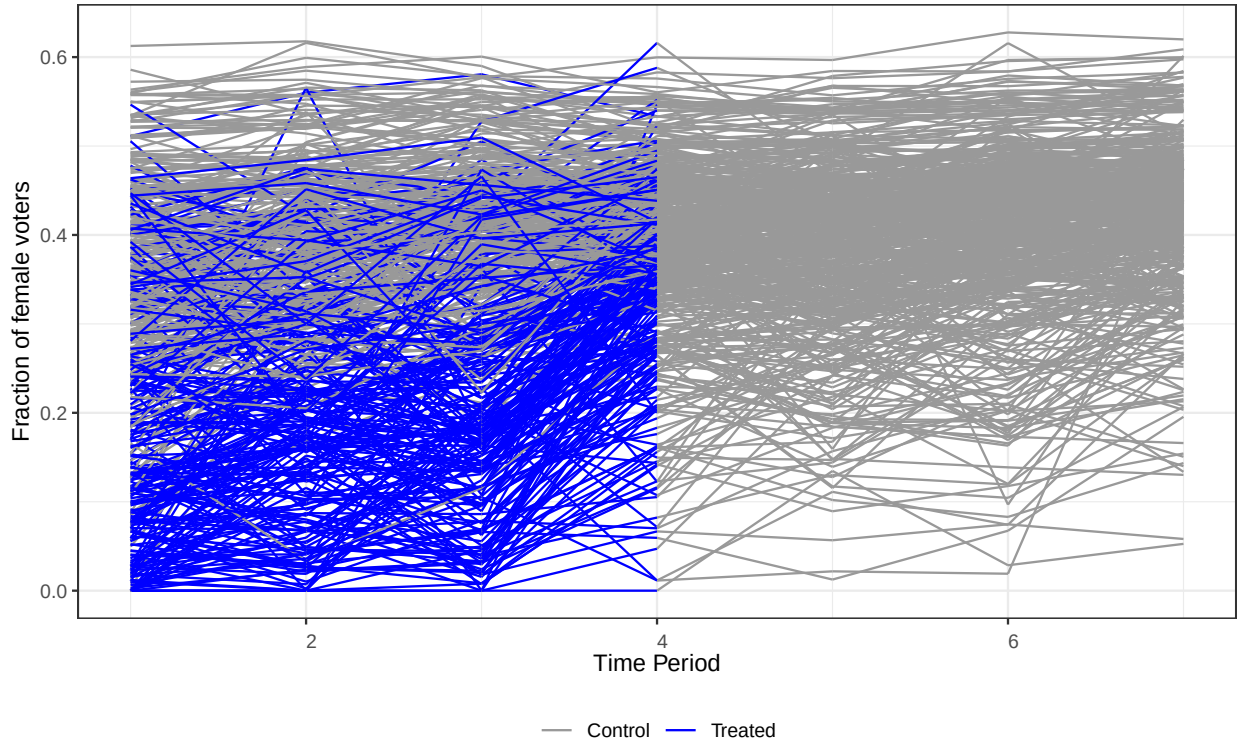
Table 1: Replication Summary

Unit of analysis	Municipality × election year
Treatment	PR reform
Outcome	Fraction of female voters
Treatment type	Staggered
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a municipality). The treatment has no reversals.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

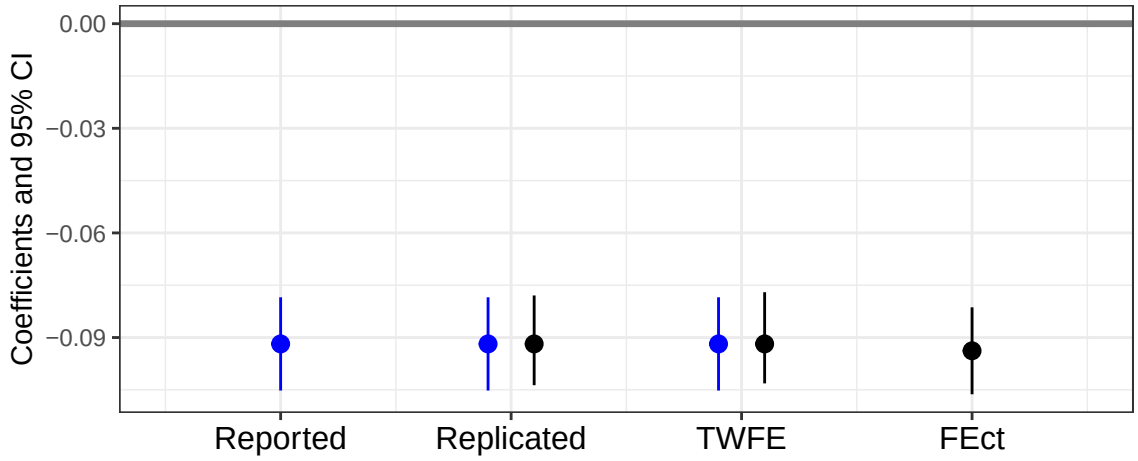
We first present the regression result following the authors' original specification. We then drop the always-treated units and apply two estimators: TWFE and the imputation method (FEct). The point estimates

and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

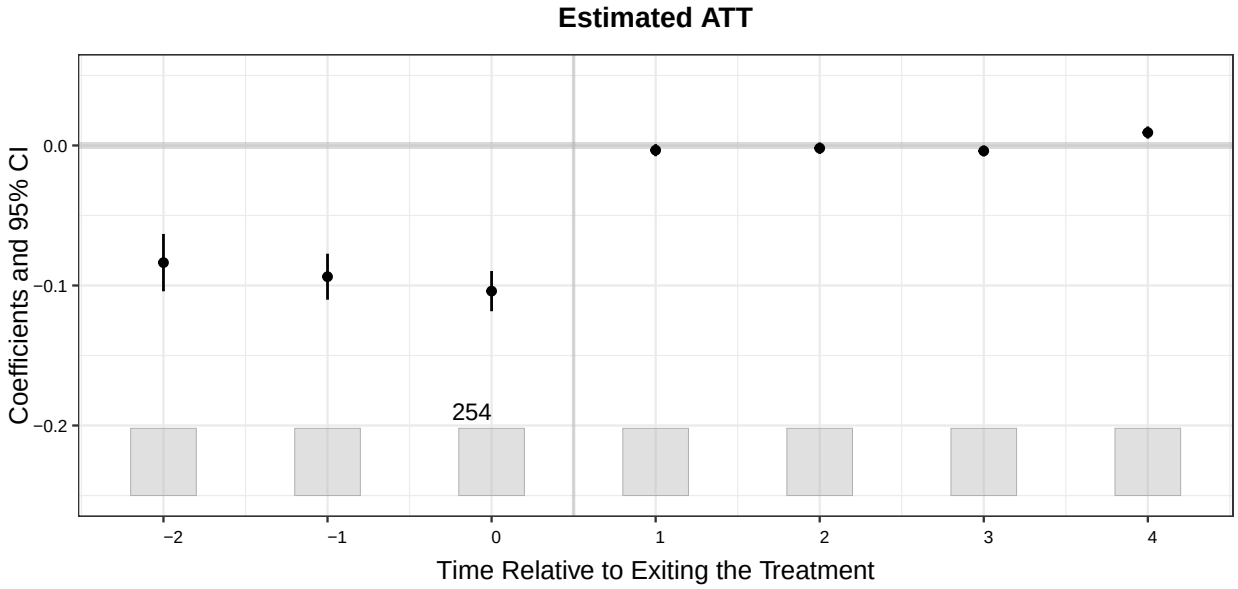
```
sol <- feols(relTurnoutW ~ neg_treatment+log.elVotersT+relElVotersW+log.repsT+log.pop+log.popDens+relPo
             data = df,cluster = "muni.")
summary(sol)
```

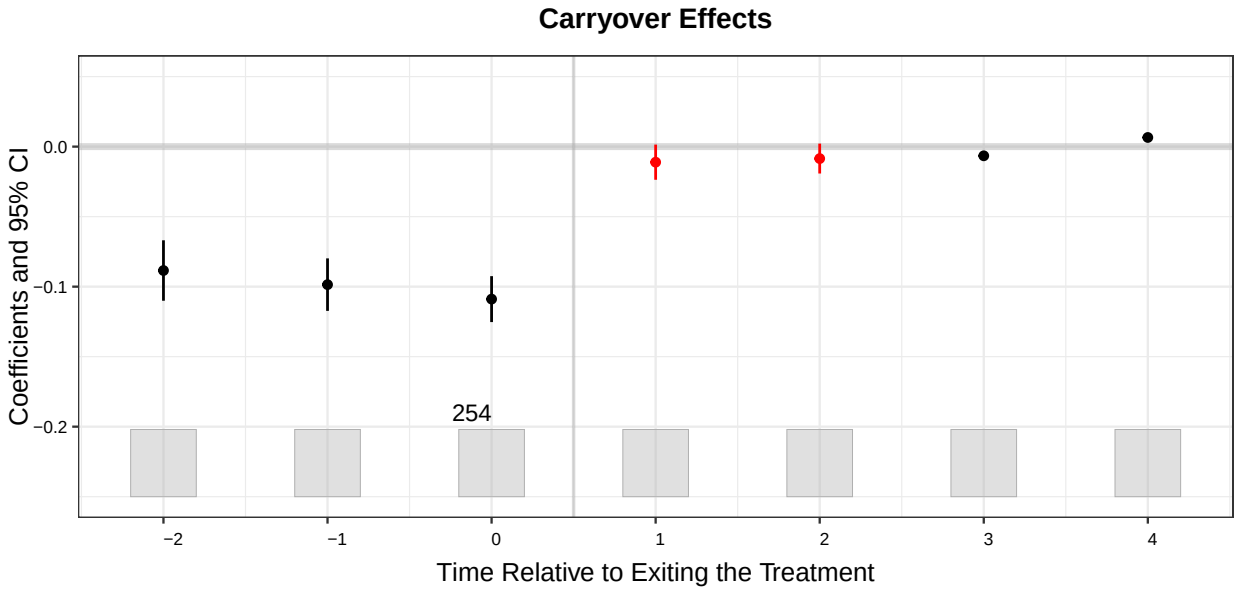
```
## OLS estimation, Dep. Var.: relTurnoutW
## Observations: 3,983
## Fixed-effects: muni.: 569,  year: 7
## Standard-errors: Clustered (muni.)
##                Estimate Std. Error    t value   Pr(>|t|)
## neg_treatment -0.091834   0.006821 -13.464255  < 2.2e-16 ***
## log.elVotersT -0.034926   0.020345  -1.716669 0.08658494 .
## relElVotersW   0.003002   0.000830   3.617621 0.00032382 ***
## log.repsT      0.071615   0.023963   2.988580 0.00292395 **
## log.pop       -0.063700   0.024387  -2.612099 0.00923687 **
## log.popDens    0.047052   0.017069   2.756672 0.00602697 **
## relPopW        0.001625   0.002945   0.551587 0.58144853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.059816     Adj. R2: 0.757046
##                  Within R2: 0.152607
```



Diagnostic Tests

Based on FEct, we conduct the test for (no) carryover effects.





Test Statistics

```
##                                    p-value
## F test                                  NA
## Equivalence test (default)              NA
## Equivalence test (threshold=ATT)        NA
## Placebo test                            NA
## Carryover effect test                0.064
```

Summary

Since the exit timing of treated units does not vary, TWFE is robust to HTE, and the imputation estimator (FEct) produces a very similar estimate. Because there are no pre-treatment periods, we cannot assess the PT assumption using pre-treatment data. Using post-treatment data, however, the no-carryover-effect assumption appear plausible.

Trounstine (2020)

Contents

A First Look at Data 1
Point Estimates 2
Dynamic Treatment Effects 4
Diagnostic Tests 7
Sensitivity Analysis 7
Summary 10

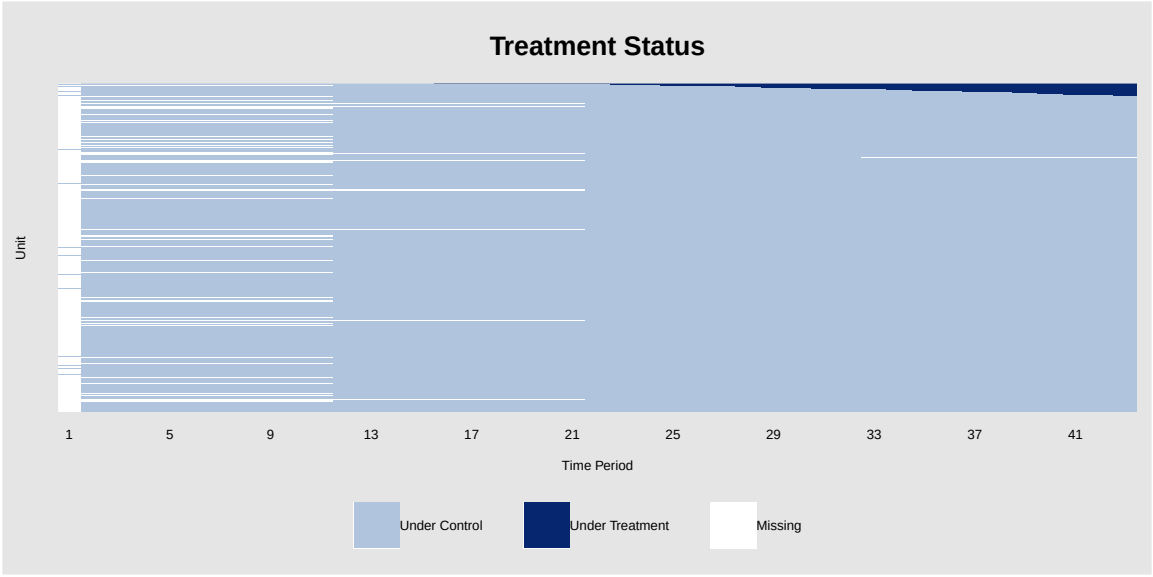
A First Look at Data

The paper investigates the effects of land use change (measured by the Fair Housing Act Lawsuit) on racial segregation, using US city-year level panel data during 1968-2011. One of the main findings of this paper is that "when cities are threatened or forced by the court to liberalize their land use laws they see growth in their population of people of color" (p451).

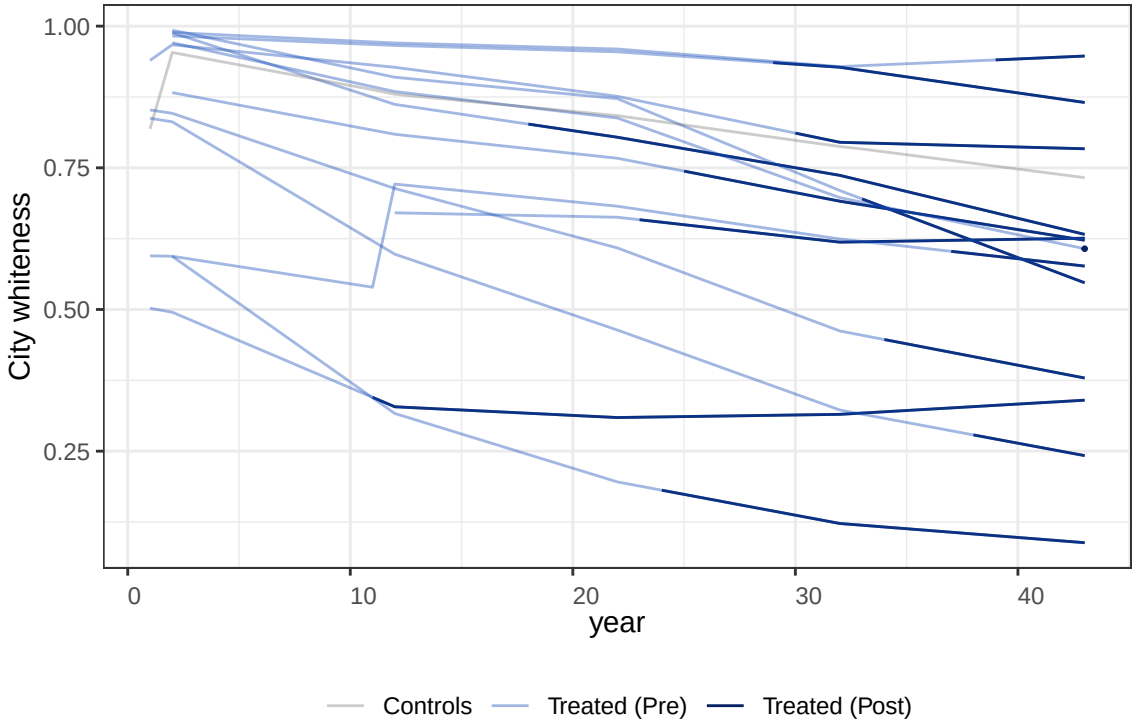
Model. We focus on **Model 1 of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the city level.

Replication Summary	
Unit of analysis	City × Year
Treatment	Land use change
Outcome	City Whiteness
Treatment type	Staggered
Outcome type	Binary
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a city). We see that a small number of units are treated at various time points.



Plotting the outcome variable. We plot the outcome variable for each city. The ever-treated units are highlighted in blue.



Point Estimates

We first present the regression result following the authors' original specification and conduct a Goodman-bacon decomposition using the original specification. We then drop the always-treated units (there is none in this case) and apply TWFE, Stacked DID, IW (Sun & Abraham) estimator, CS (Callaway & Sant'anna)

estimator, DIDM estimator and the imputation method (FEct) to the data. The point estimates and their 95% CIs are shown in the figure below. Throughout the analysis, we use blue and black bars to represent confidence intervals (CIs) based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

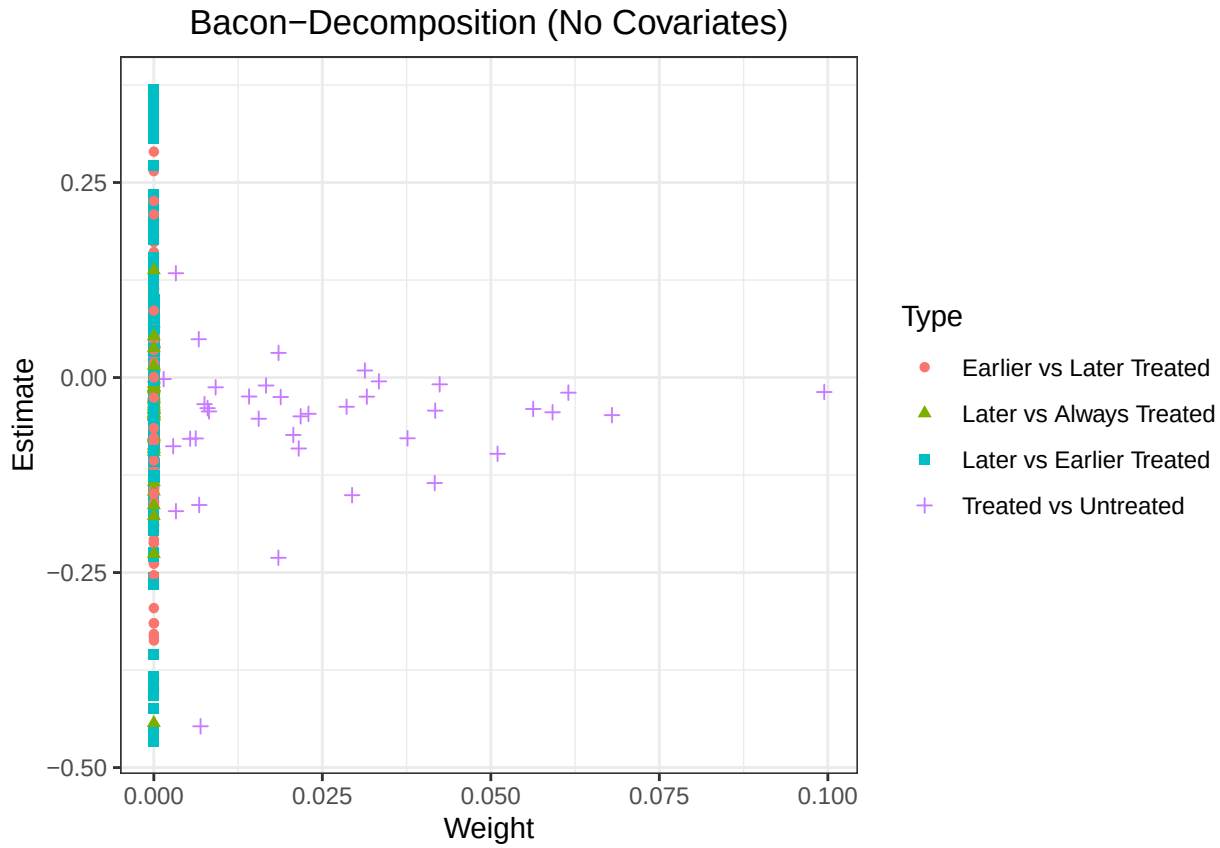
Original Finding

```
sol <- feols(pctnhwht_cityi~fairhousingimp_v1|geo_id2+year,
             data = df,cluster = "geo_id2")
summary(sol)
```

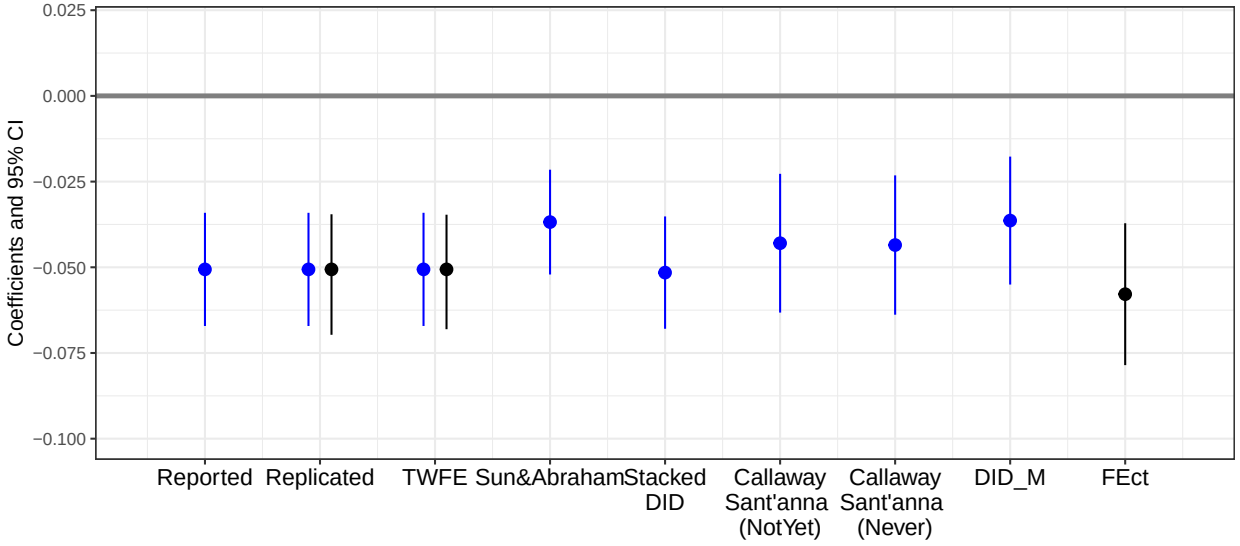
```
## OLS estimation, Dep. Var.: pctnhwht_cityi
## Observations: 182,809
## Fixed-effects: geo_id2: 4,568,  year: 43
## Standard-errors: Clustered (geo_id2)
##                     Estimate Std. Error  t value   Pr(>|t|)
## fairhousingimp_v1 -0.050611   0.008424 -6.00814 2.0225e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.068444     Adj. R2: 0.886308
##                  Within R2: 0.003812
```

Goodman-Bacon Decomposition

In the Goodman-Bacon decomposition, we see that the TWFE estimate is dominated by comparisons between treated and untreated units (purple crosses).




```
##   |                                                                      |
```



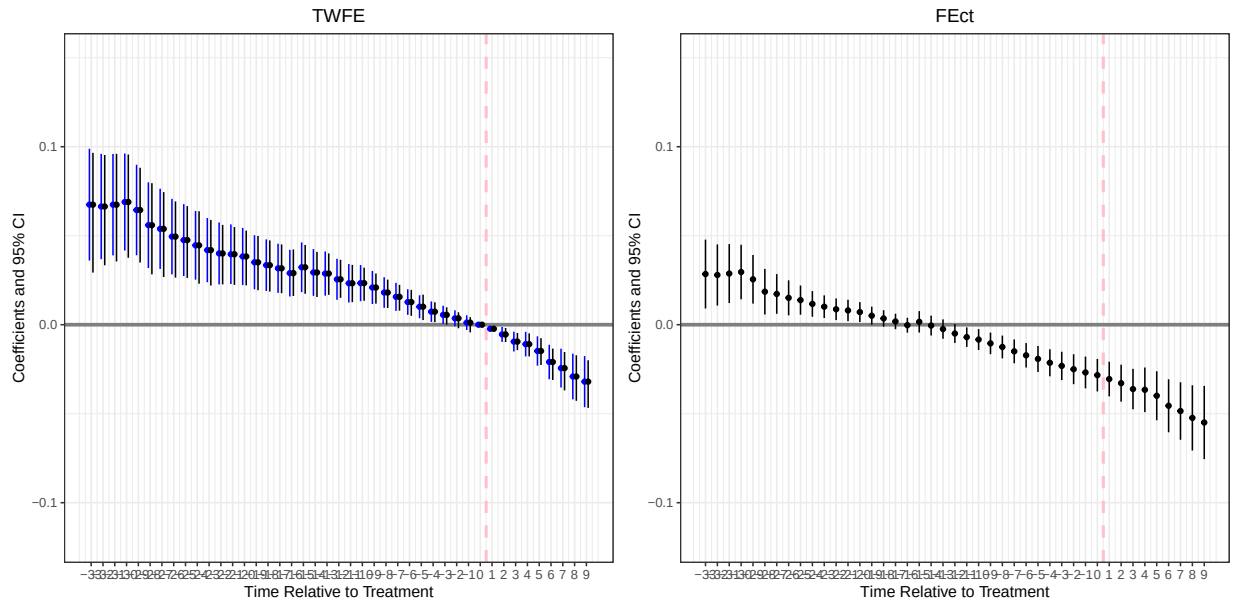
Estimates from all estimators we deploy are broadly consistent.

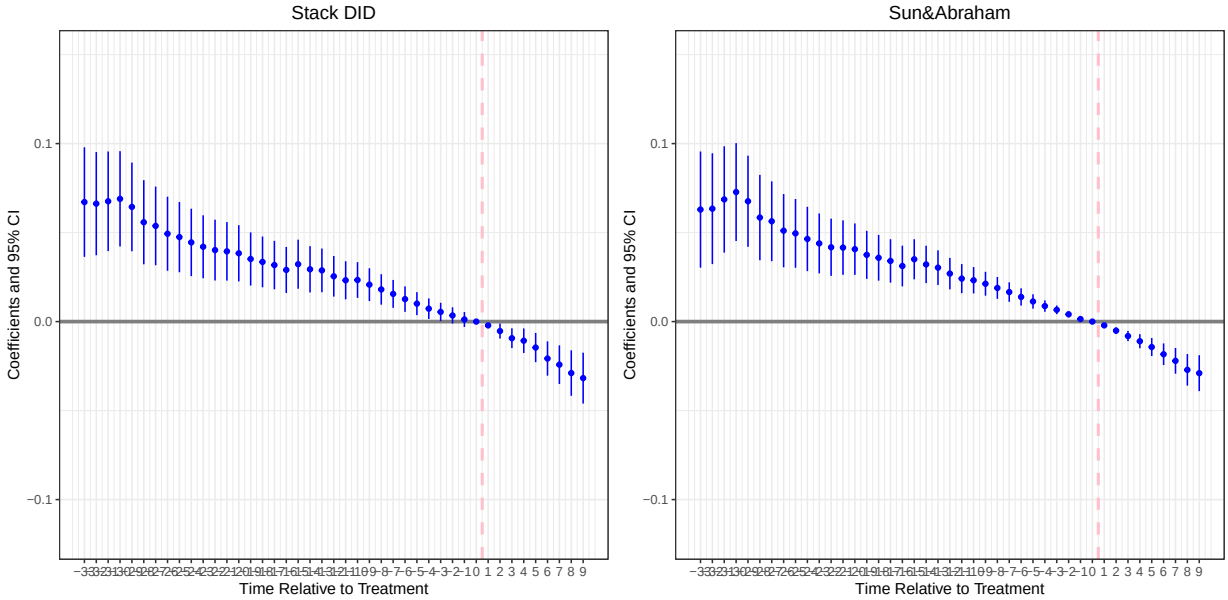
| |

Dynamic Treatment Effects

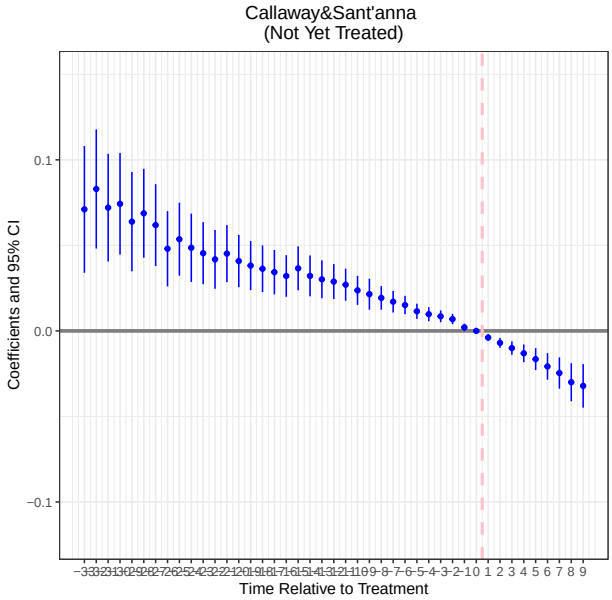
We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. The results are shown below.

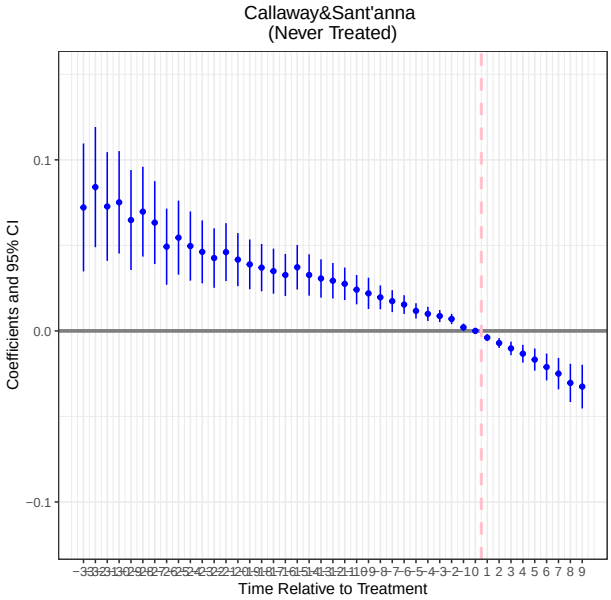
| |

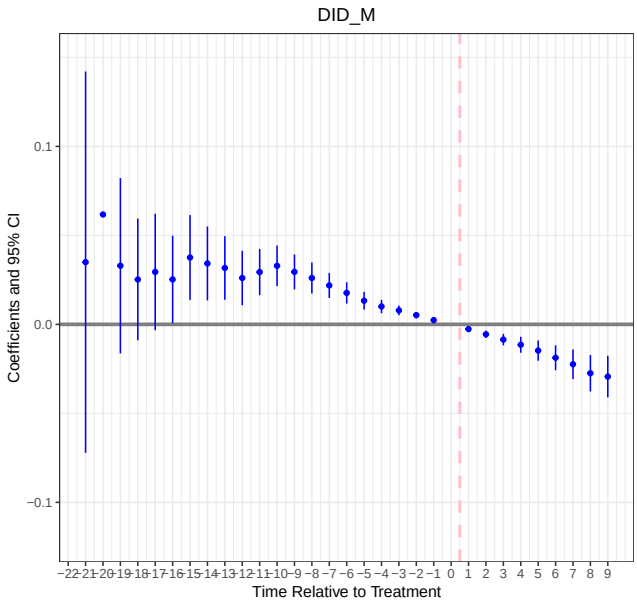



Stack DID
Coefficients and 95% CI
0.1
0.0
−0.1
Time Relative to Treatment

Sun&Abraham
Coefficients and 95% CI
0.1
0.0
−0.1
Time Relative to Treatment


Callaway&Sant'anna
(Not Yet Treated)
Coefficients and 95% CI
0.1
0.0
−0.1
Time Relative to Treatment

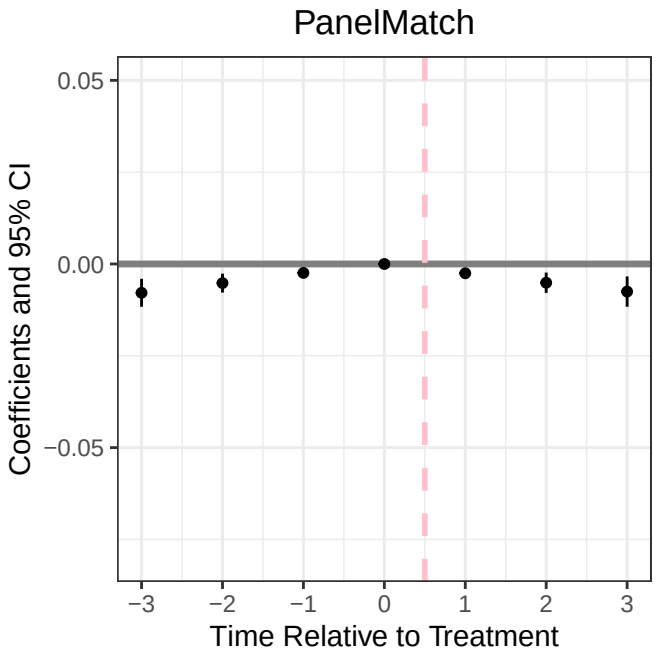

Callaway&Sant'anna
(Never Treated)
Coefficients and 95% CI
0.1
0.0
−0.1
Time Relative to Treatment



Estimates are broadly consistent with each other. Most estimated DTEs are statistically significant in pre-treatment periods and are often similar or larger in magnitude than post-treatment DTEs.

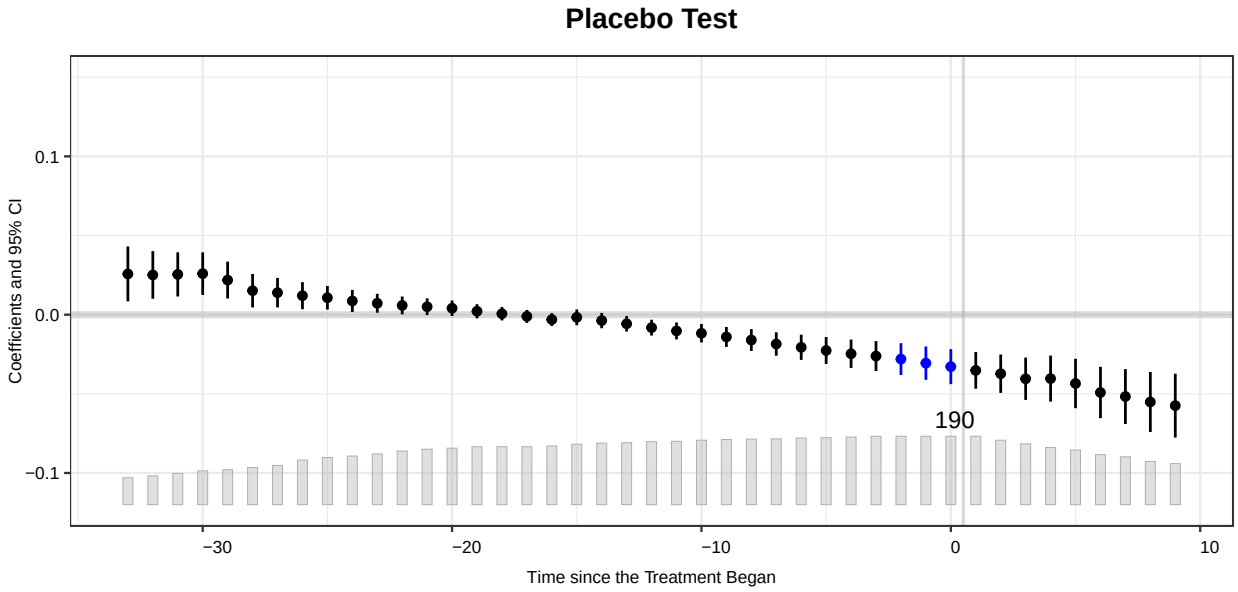
DTE of a Balanced Subsample

We also estimate the DTE using PanelMatch (lead = 3 and lag = 4). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.



Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend and a placebo test.



Test Statistics

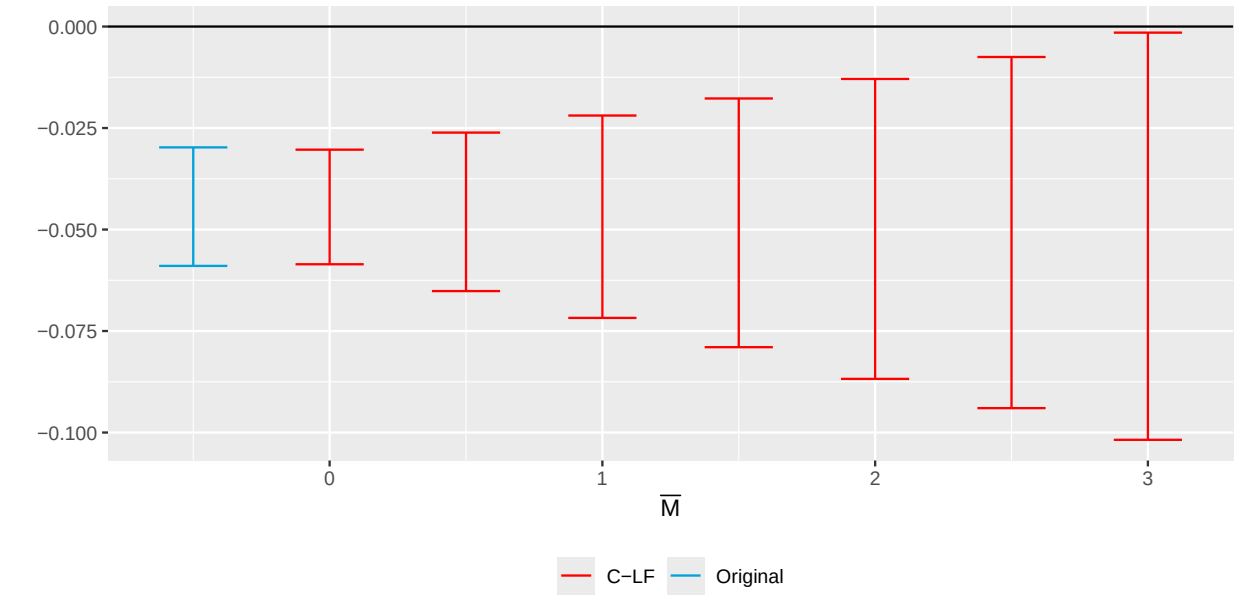
```
##                                  p-value
## F test                             0.001
## Equivalence test (default)         0.772
## Equivalence test (threshold=ATT)   0.001
## Placebo test                       0.000
## Carryover effect test                 NA
```

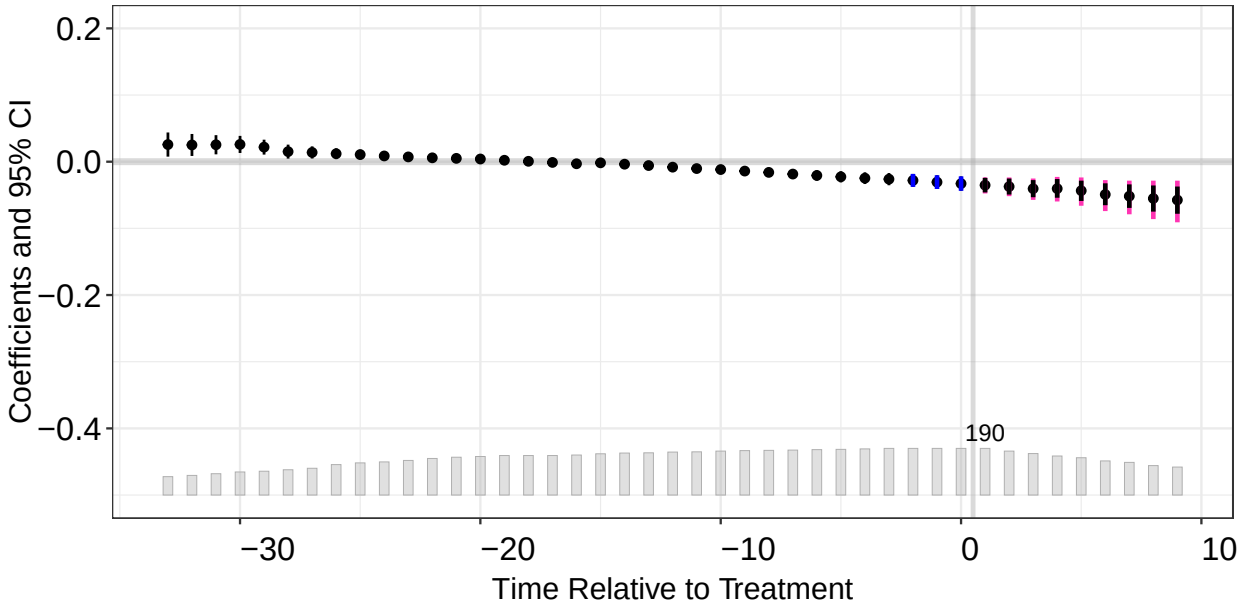
We find evidence of violations of the parallel trends (PT) violation: both the F test and placebo test reject.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We assess the robustness of this significant ATT using the method proposed by Rambachan and Roth (2022). We utilize the three placebo periods to establish the baseline deviation from the parallel trend observed during the pre-treatment periods. The critical value for a significant effect is determined to be $\bar{M} \approx 3$. This finding allows us to reject the null hypothesis of no treatment effects, provided that we accommodate for parallel trend deviations in post-treatment periods up to 3 times the maximum deviation observed between consecutive periods in the placebo periods.

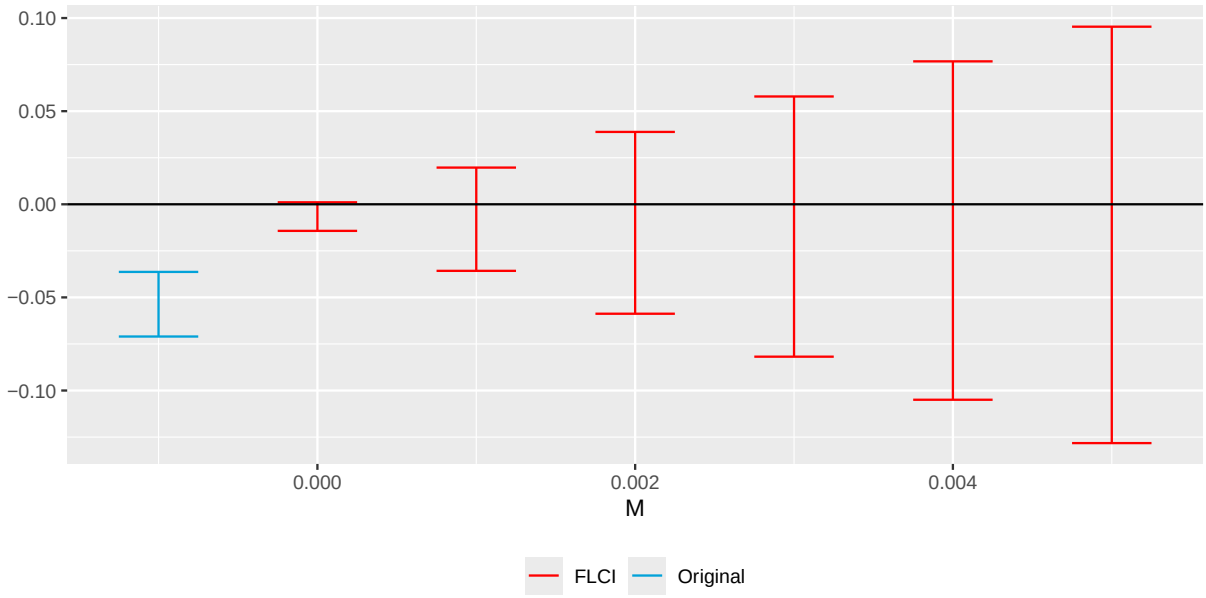
We also assess the robustness of the estimated dynamic treatment effects under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for parallel trend deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



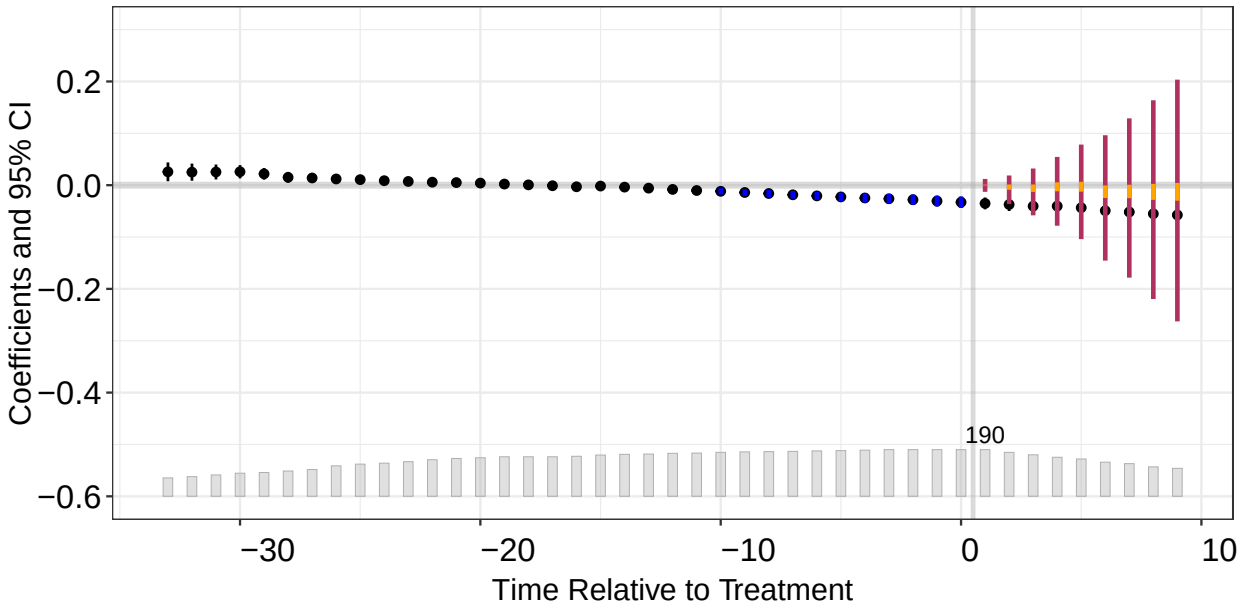


As we observe a linear differential pre-trend using all methods, we also apply the smoothness restriction following Rambachan and Roth (2022). Specifically, we impose that the slope of the pre-trend can change by no more than M across consecutive periods in the post-treatment periods. After holding out six pre-treatment periods, the estimated ATT derived from the FEct estimator is significant. We use these six placebo periods to benchmark the linear differential pre-trend. Even under $M = 0$, meaning the slope of the pre-trend is required to remain constant in the post-treatment periods, the estimated ATT is not significant.

We also evaluate the robustness of the estimated dynamic treatment effects under $M = 0$ and $M = 0.004$. The period-wise robust confidence intervals for the DTE, which allow the slope of the pre-trend to change by no more than 0.004 across consecutive periods, are shown in maroon. In contrast, the confidence intervals for the DTE that assume the slope of the pre-trend remains constant in the post-treatment periods are depicted in orange. Both of them cover zero.




```
## Warning in seq_len(length.out - 2L) * (del/n1):          1
##      c()   as.vector()
## Warning in seq_len(length.out - 2L) * (del/n1):          1
##      c()   as.vector()
## Warning in seq_len(length.out - 2L) * (del/n1):          1
##      c()   as.vector()
## Warning in seq_len(length.out - 2L) * (del/n1):          1
##      c()   as.vector()
## Warning in seq_len(length.out - 2L) * (del/n1):          1
##      c()   as.vector()
## Warning in seq_len(length.out - 2L) * (del/n1):          1
##      c()   as.vector()
```



Summary

Overall, the main result of the chosen model remains statistically significant when HTE-robust estimators are applied. However, we find strong evidence for violations of the PT. The ATT estimate is not robust to PT violations (e.g., caused by a linear differential trend).

Weschle (2021)

Contents

A First Look at Data 1
Point Estimates 3
Dynamic Treatment Effects 4
Diagnostic Tests 6
Sensitivity Analysis 7
Summary 7

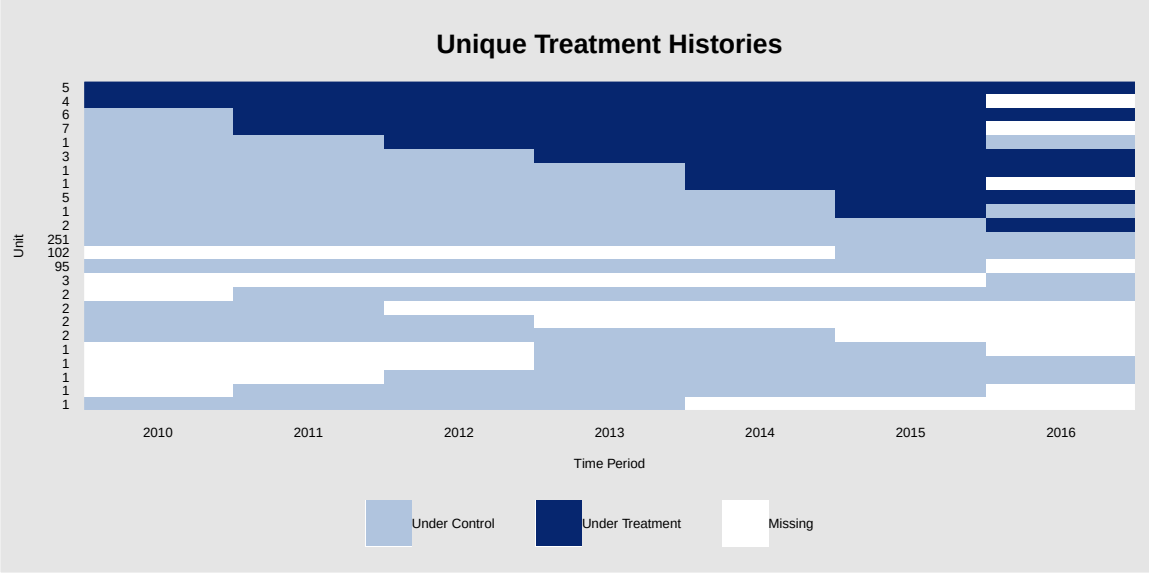
A First Look at Data

The paper investigates the effects of holding different parliamentary positions on private sector earnings, using UK House of Commons MP-year level panel data, between 2010 and 2016. One of the main findings of this paper is that "Former ministers … substantially increase their private sector earnings soon after leaving their position" (p715).

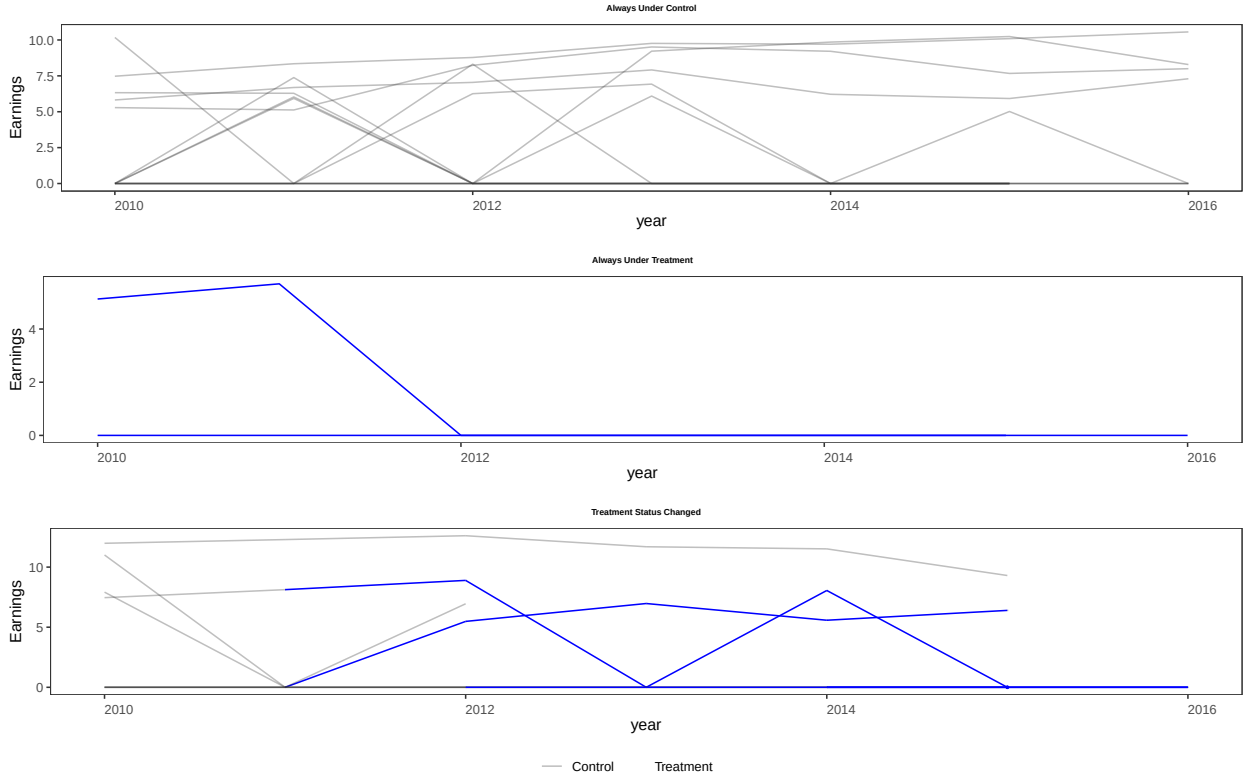
Model. We focus on **Model 1 of Table 1** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the unit level.

Replication Summary	
Unit of analysis	MP × year
Treatment	Holing parliamentary position
Outcome	Private sector Earnings
Treatment type	General
Outcome type	Continous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (an MP). We see that a small number of units are treated at various time points and there are treatment reversals. There is some missingness.



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



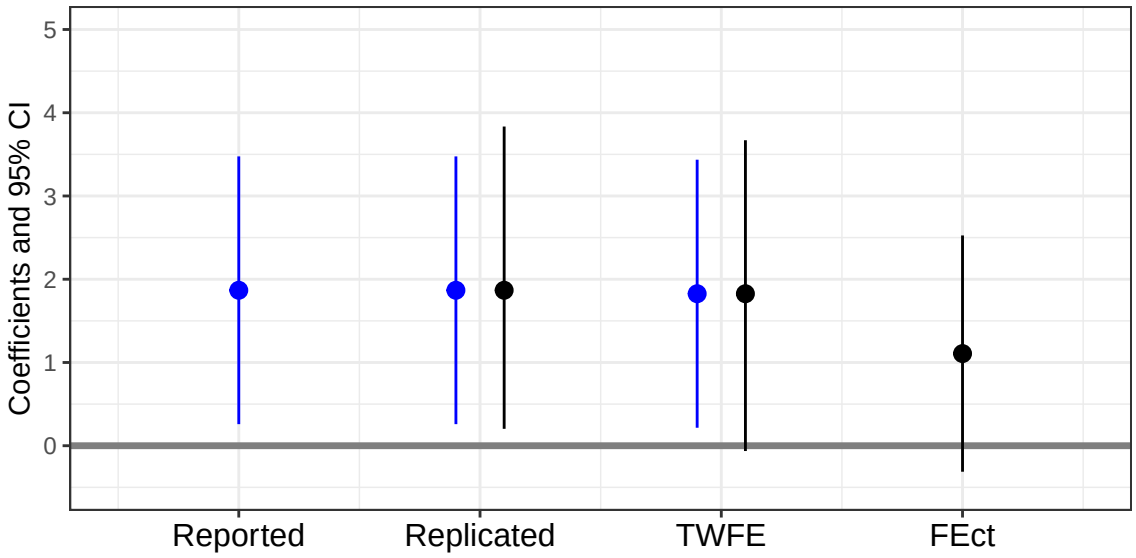
Point Estimates

We first replicate the authors' using the original specification—we are able to successfully replicate the point estimate. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs) are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

```
sol <- feols(earnings.all.infl.log.1~minister.post+ minister+undersec+shadow.cabinet+
               frontbench.team+com.chair+com.member+minister.state+
               minister.state.post+undersec.post+shadow.cabinet.post+frontbench.team.post+
               com.chair.post+com.member.post+enter+leave |id + year,data = df,cluster = "id")
summary(sol)
```

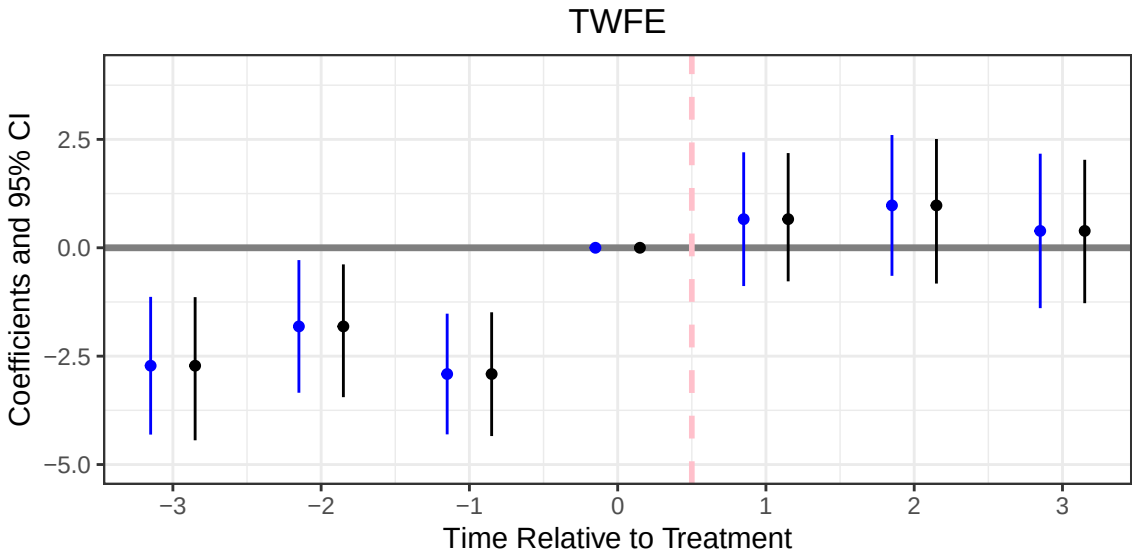
```
## OLS estimation, Dep. Var.: earnings.all.infl.log.1
## Observations: 4,714
## Fixed-effects: id: 845,  year: 7
## Standard-errors: Clustered (id)
##                       Estimate Std. Error   t value   Pr(>|t|)
## minister.post         1.867321   0.820476  2.275900 2.3102e-02 *
## minister             -1.556140   0.515502 -3.018686 2.6151e-03 **
## undersec             -2.059293   0.342968 -6.004329 2.8520e-09 ***
## shadow.cabinet       -0.333409   0.197272 -1.690099 9.1378e-02 .
## frontbench.team      -0.282766   0.326767 -0.865344 3.8710e-01
## com.chair            -0.368415   0.491769 -0.749162 4.5397e-01
## com.member            0.031214   0.414455  0.075313 9.3998e-01
## minister.state       -1.779338   0.448492 -3.967383 7.8839e-05 ***
## minister.state.post   0.585634   0.810810  0.722282 4.7032e-01
## undersec.post        -0.326898   0.572947 -0.570556 5.6845e-01
## shadow.cabinet.post  -0.019456   0.293099 -0.066381 9.4709e-01
## frontbench.team.post  0.814468   0.510518  1.595374 1.1100e-01
## com.chair.post       -1.690731   0.767155 -2.203898 2.7801e-02 *
## com.member.post      -0.140500   0.441880 -0.317959 7.5059e-01
## enter                -0.502514   0.182151 -2.758781 5.9273e-03 **
## leave                -1.078801   0.235901 -4.573106 5.5247e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2.1202     Adj. R2: 0.652488
##                Within R2: 0.081754
```

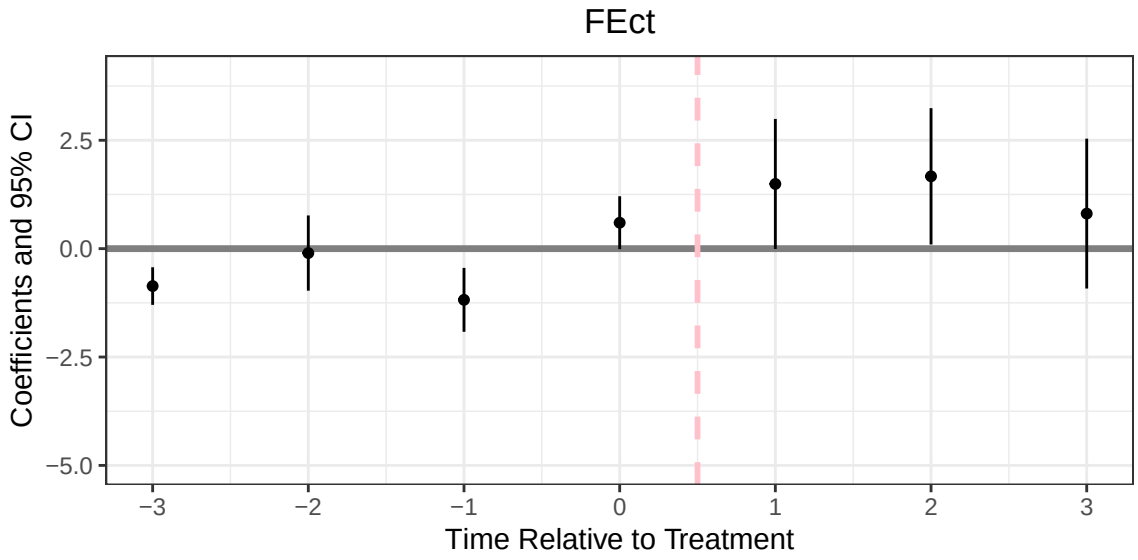


The TWFE and FEct estimates are consistent with each other, although the FEct estimate is not statistically significant.

Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.

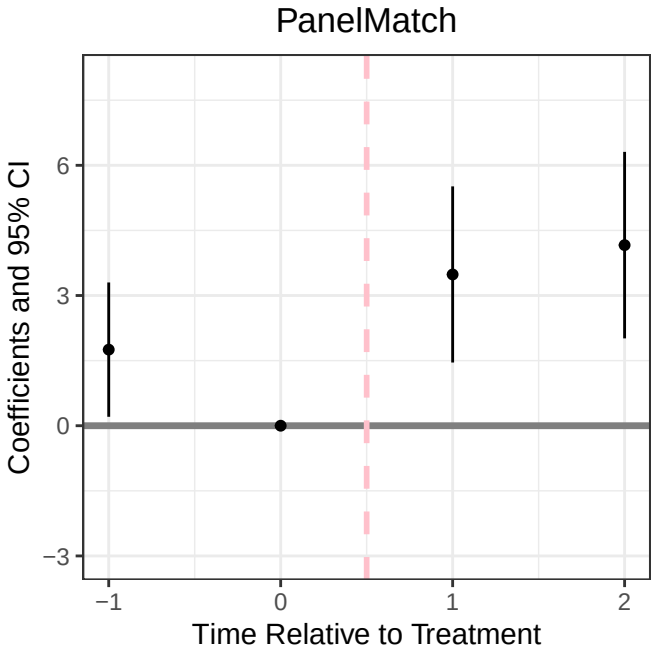




The TWFE and FEct estimates are consistent with each other.

DTE of a Balanced Subsample

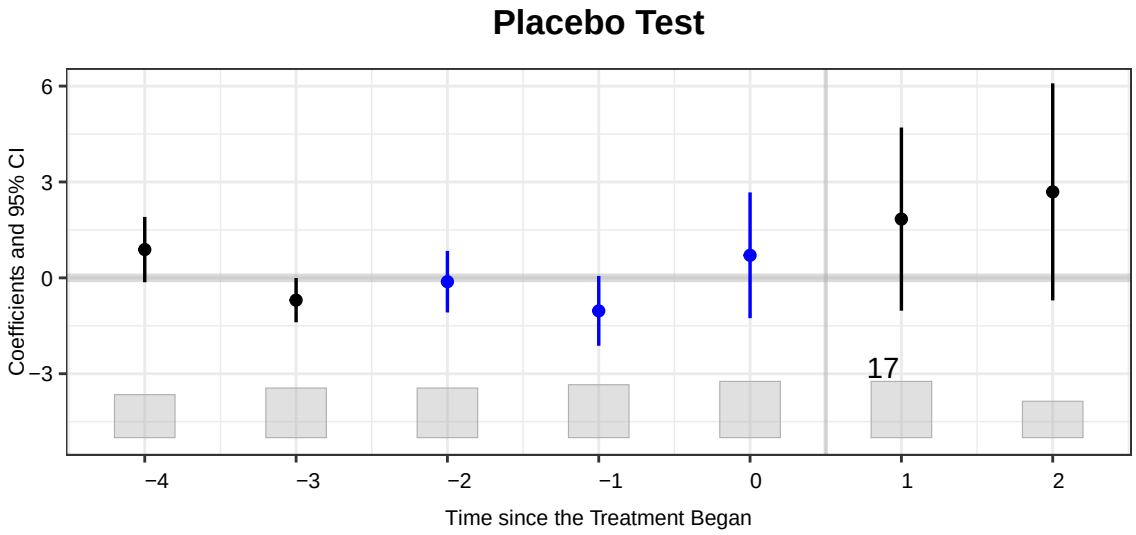
We also estimate the DTE using PanelMatch (lead = 2 and lag = 2). This method estimates treatment effects only for treated units with complete post-treatment (lead) and pre-treatment (lag) periods surrounding the treatment. Note that PanelMatch focuses on a subgroup of treated units and does not incorporate time-varying covariates into the outcome model. Therefore, its estimated treatment effect is not directly comparable to other HTE-robust estimators.

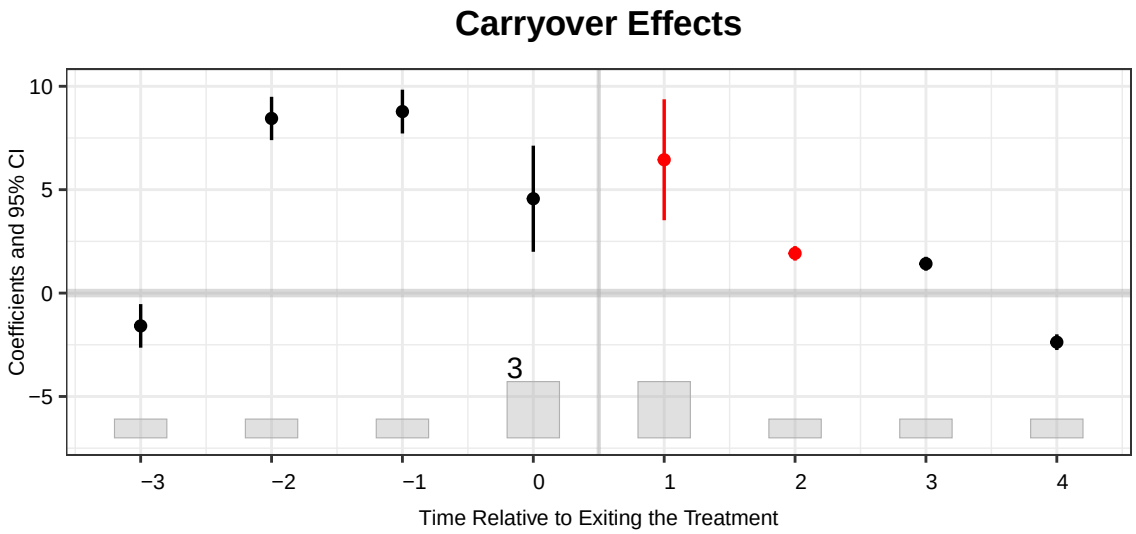


FEct and PanelMatch estimates are broadly consistent with each other.

Diagnostic Tests

Based on FEct, we conduct several diagnostic tests, including testing for (no) pre-trend, a placebo test, and a test for (no) carryover effects.





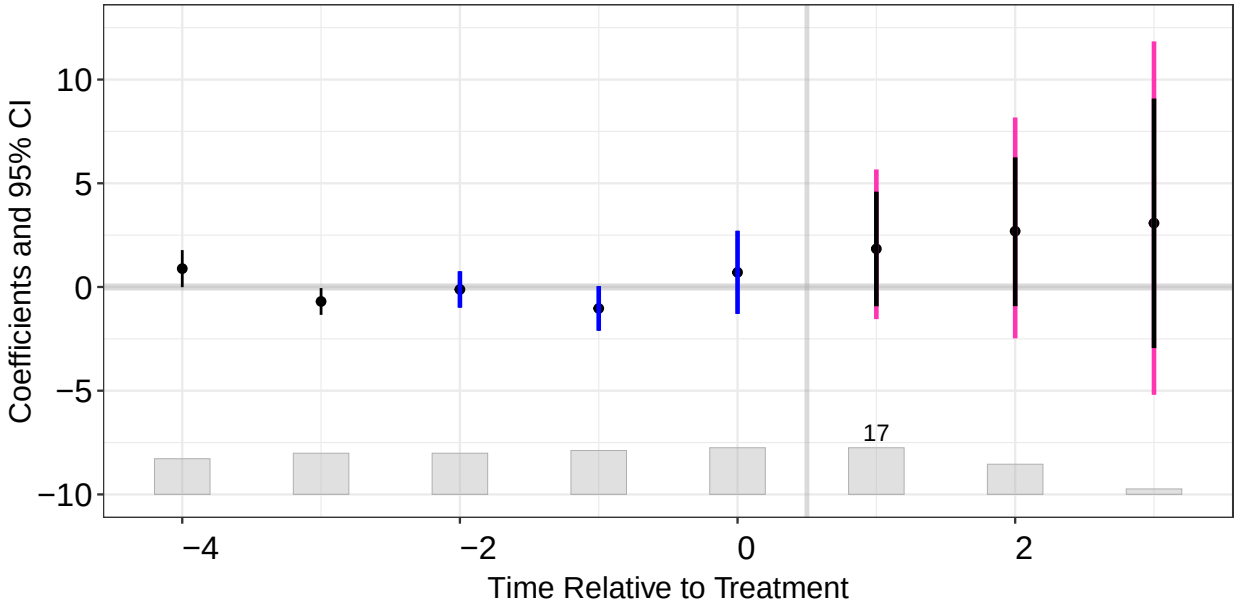
Test Statistics

```
##                                  p-value
## F test                             0.003
## Equivalence test (default)         0.820
## Equivalence test (threshold=ATT)   0.578
## Placebo test                       0.819
## Carryover effect test              0.000
```

We find some evidence of the parallel trends (PT) violation: the F test rejects. The equivalence test can reject the null that the residuals in pre-treatment periods exceed the estimated ATT. The test indicates a potential violation of the no-carryover-effect assumption.

Sensitivity Analysis

After holding out three pre-treatment periods, the estimated ATT derived from the FEct estimator is not significant, so we only assess the robustness of the estimated dynamic treatment effects using the method proposed by Rambachan and Roth (2022) under $\bar{M} = 0.5$. The period-wise robust confidence intervals for the DTE, which allow for PT deviations up to 0.5 times the maximum observed deviation in the placebo periods, are depicted in pink.



Summary

Overall, the main result of the chosen model appears to be not robust to HTE-robust estimator. We find some evidence of PT violations. Due to limited power, the estimated ATT is no longer statistically significant after we hold out three periods for the placebo test, which is the basis of the sensitivity analysis.

Zhang et al. (2021)

Contents

A First Look at Data 1

Point Estimates 2

Dynamic Treatment Effects 3

Diagnostic Tests 4

Summary 5

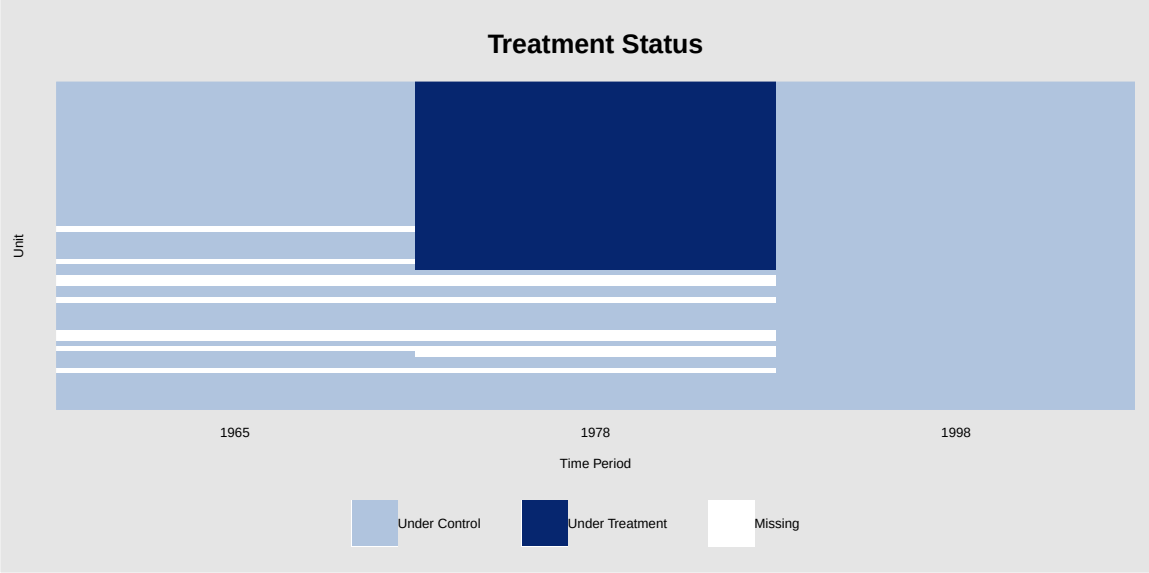
A First Look at Data

The paper investigates the effects of marginal faction counties on private economic activities using county-year level panel data from counties in two Chinese provinces during 1965-1998. One of the main findings of this paper is that "counties governed by political elites from the marginalized faction in Zhejiang province had developed more vibrant non-state industry than counties governed by political elites from the dominant faction during the Cultural Revolution" (p1017).

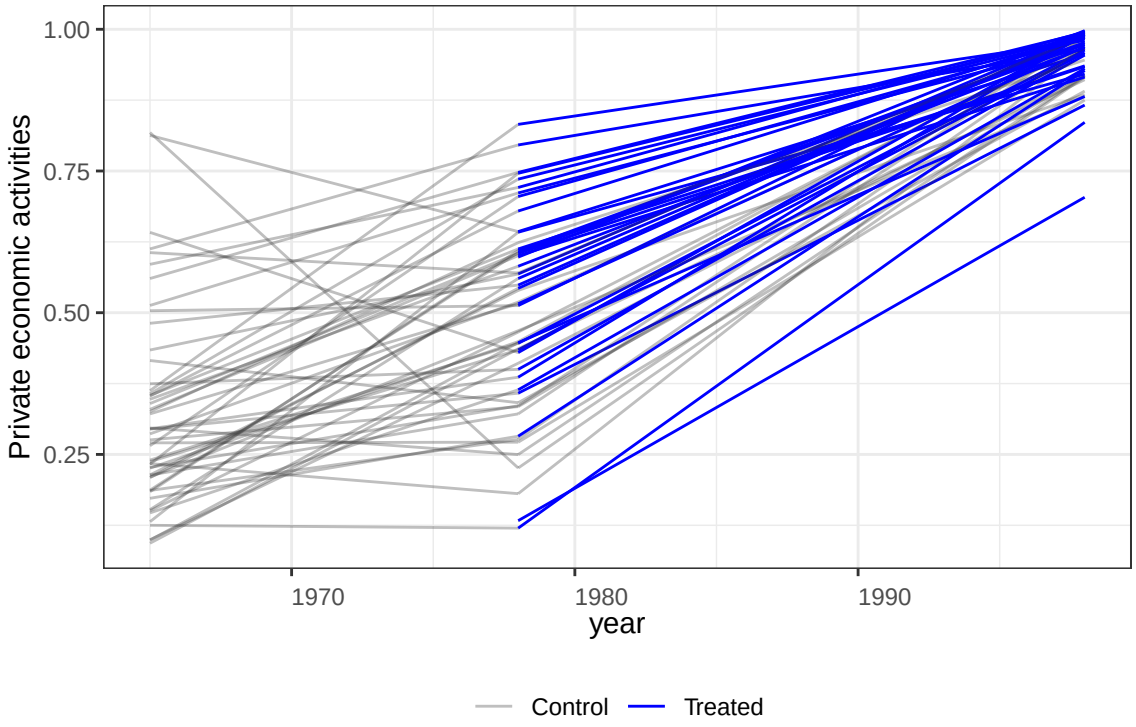
Model. We focus on **Model 1 of Table 2** in the paper. The authors use a two-way fixed effects (TWFE) model and report robust standard errors clustered at the county level.

Replication Summary	
Unit of analysis	County × year
Treatment	MFC
Outcome	Private economic activities
Treatment type	General
Outcome type	Continuous
Fixed Effects	Unit+Time

Plotting treatment status. First, we plot the treatment status in the data. In the figure below, each column represents a time period (a year) and each row represents a unit (a county).



Plotting the outcome variable. We plot the trajectory of the outcome variable for each city. The observations under treated status are marked in blue.



Point Estimates

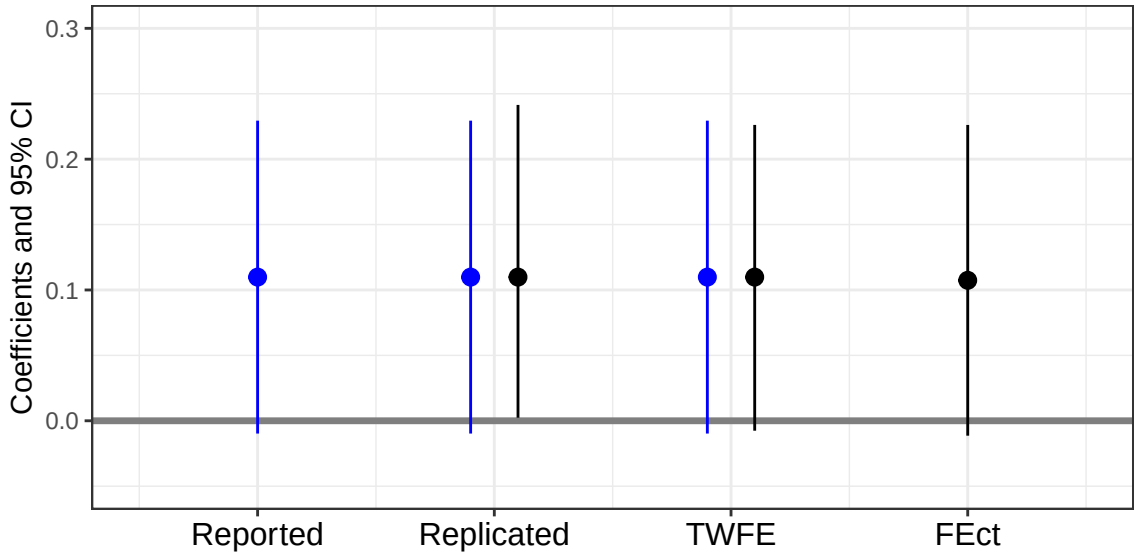
We first replicate the authors' using the original specification—we are able to successfully replicate the point estimate. We then drop the always-treated units (there is none in this case) and apply two estimators: TWFE and the imputation method (FEct). The point estimates and their 95% confidence intervals (CIs)

are shown in the figure below. Throughout the analysis, we use blue and black bars to represent CIs based on cluster-robust SEs and cluster-bootstrapped CIs, respectively.

Original Finding

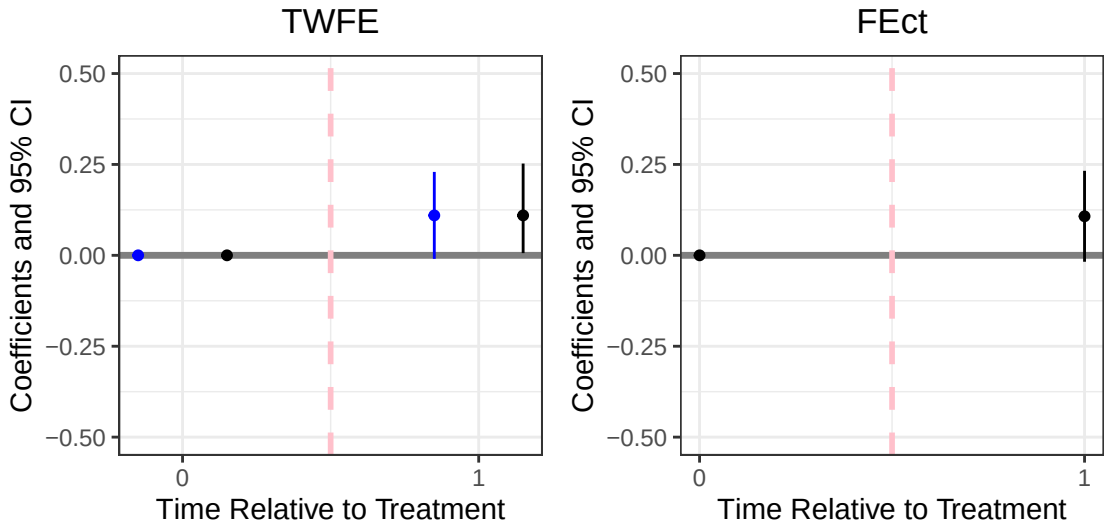
```
sol <- feols(nonsoe~mfc1978+mfc1998|countyid+year,data = df,cluster = "countyid")
summary(sol)
```

```
## OLS estimation, Dep. Var.: nonsoe
## Observations: 166
## Fixed-effects: countyid: 61,  year: 3
## Standard-errors: Clustered (countyid)
##          Estimate Std. Error   t value Pr(>|t|)
## mfc1978  0.109831   0.061015  1.800054  0.07688 .
## mfc1998 -0.037727   0.049372 -0.764123  0.44779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.092564     Adj. R2: 0.8476
##                  Within R2: 0.089504
```

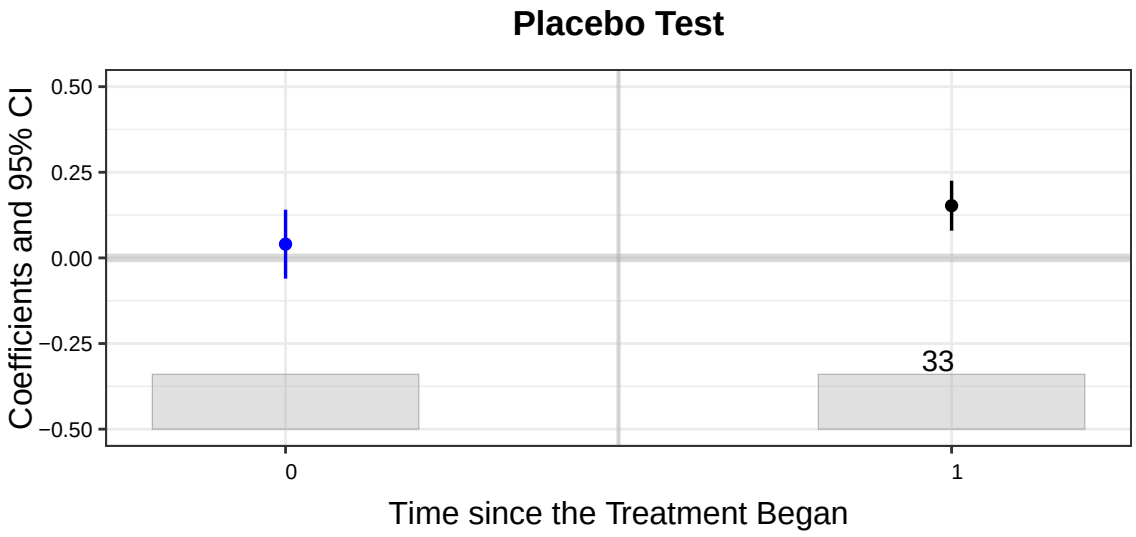


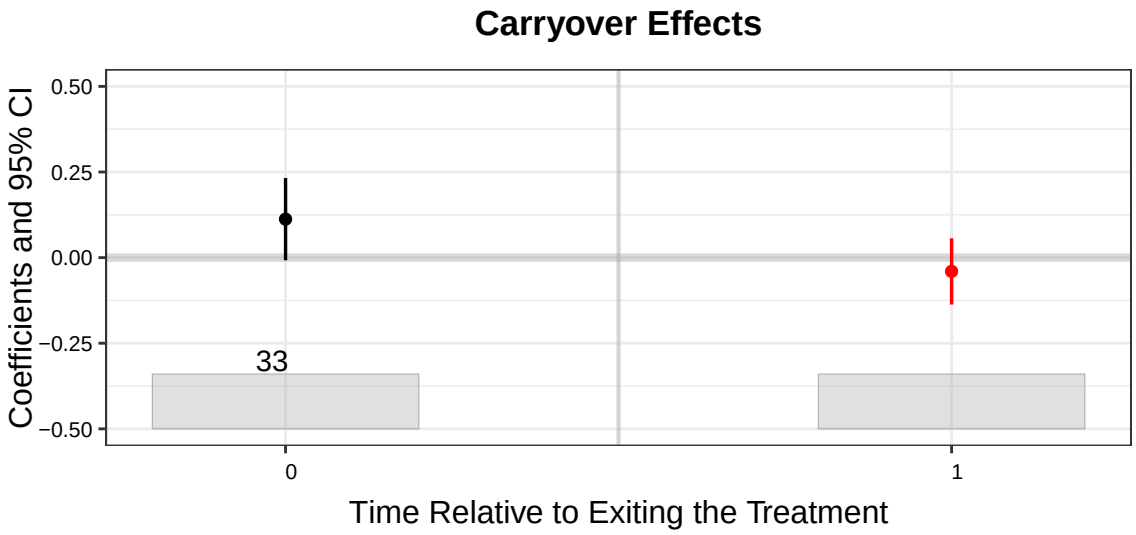
Dynamic Treatment Effects

We then move onto estimating dynamic treatment effects (DTEs) and obtaining the following DTE/event-study plots. We use two estimators, TWFE and FEct. The results are shown below.



Diagnostic Tests






```
##                                    p-value
## F test                                  NA
## Equivalence test (default)              NA
## Equivalence test (threshold=ATT)        NA
## Placebo test                         0.434
## Carryover effect test                0.416
```

Summary

The study appears underpowered, although the imputation estimator (FEct) yields a result similar to TWFE. Because there is one pre-treatment period, it is difficult for us to evaluate whether the parallel trends assumption is plausible.