DidToolbox: Real Data Example

Dr. Ralf Elsas-Nicolle, LMU Munich, Germany

Last change: 10/03/2025

Table of Contents

0 Overview 1
1 Load data and look at descriptive statistics 2
2 Standard TWFE and Bacon decomposition 3
3 Staggered DiD estimation 5
3.1 Overview 5
3.2 Wooldridge and BJS 5
3,3 de Chaisemartin / d'Haultfoeuille (CH) 6
3.4 Callaway / Sant'Anna (CS) 6
4 Conclusions 8
Literature 9

0 Overview

This example illustrates using the Matlab DiD Toolbox in a real world application - a test of the so-called "Castle Doctrine" (a change in U.S. self-defense law) on the homicides before and after the change in law. This analysis is based on the example in Scott Cunningham's diff-in-diff blog and the study by *Cheng & Hoekstra (2013)*.

Background

From 2000 to 2010, more than 20 states passed the so-called “Castle Doctrine” law. For example, in 2005, the US state of Florida introduced a variant law dubbed “Stand Your Ground”. This expanded the existing “castle” doctrine – a legal doctrine in the US concerning lethal self-defense within one’s home. Whereas in other cases a victim has the duty to retreat from danger, within their own home, castle doctrine allows the application of lethal self-defense under some conditions. “Stand Your Ground” in Florida expanded the applicability of castle doctrine to include public spaces, a controversial change, especially due to concerns about an increase in violence. Other federal states followed, with all law changes occurring in the period from 2005-2009.

As “Stand Your Ground” laws, such as the one in Florida, state that use of violent force is permissible if the user “reasonably believes” it is necessary to avert death, injury, or another violent crime, many feared that this would result in a lowered barrier to using lethal force and possibly vigilantism.

For this reason, the issue came under scrutiny from various sources, one of these being a paper by Cheng & Hoekstra, published in 2013. They investigate the question of whether “Stand Your Ground” laws resulted in an increase in homicides, as feared by many of their critics. It is important to note that the key variable of interest “homicide” (measured in logs per 100,000 inhabitants) does not (!) capture self-defense killings but only murder and non-negligent manslaughter, so killings that even under the castle doctrine are illegal. Cheng’s & Hoekstra’s data set included details on crime rates, socioeconomic factors, and particularly homicide rates in all US states between the years of 2000 and 2010, using the introduction of “Stand Your Ground” laws as a treatment.

Also, estimator settings are such that they resemble the default settings in Stata 18, so that direct comparisons (using the same data) are possible. A Stata log-file with results for the Castle Doctrine data is available here.

1 Load data and look at descriptive statistics

In the first step we load the relevant data.

```
clear; close all;
tbl =
readtable("d:\Projekte\basic_econometrics\Diff_in_Diff\Castle\castleData.xlsx");
```

```
Warning: Column headers from the file were modified to make them valid MATLAB identifiers before creating
variable names for the table. The original column headers are saved in the VariableDescriptions property.
Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable names.
```

```
tbl = sortrows(tbl,["sid","year"]);
```

First descriptive statistic is the list of federal states and their respective treatment year:

```
firstTreatYear = varfun(@min, tbl(tbl.treatPost==1,:), ...
    'InputVariables','year', 'GroupingVariables','state');
firstTreatYear.Properties.VariableNames{'min_year'} = 'firstYear'
```

```
firstTreatYear = 21×3 table
```

	state	GroupCount	firstYear
1	'Alabama'	4	2007
2	'Alaska'	4	2007
3	'Arizona'	4	2007
4	'Florida'	5	2006
5	'Georgia'	4	2007
6	'Indiana'	4	2007
7	'Kansas'	4	2007
8	'Kentucky'	4	2007
9	'Louisiana'	4	2007
10	'Michigan'	4	2007
11	'Mississippi'	4	2007
12	'Missouri'	3	2008
13	'Montana'	1	2010
14	'North Dakota'	3	2008
15	'Ohio'	2	2009
16	'Oklahoma'	4	2007
17	'South Carolina'	4	2007
18	'South Dakota'	4	2007

	state	GroupCount	firstYear
19	'Tennessee'	3	2008
20	'Texas'	3	2008
21	'West Virginia'	2	2009

Second, we use the toolbox to better understand the data:

```
% Descriptive Statistics
descript = did.dataDesc(tbl, idVar="sid", timeVar="year",yVar="l_homicide",
dVar="treatPost");
```

```
[DataDesc] Units = 50 | Periods = 11 | Nobs = 550 | Balanced Units = 50 (100.0%)
[DataDesc] Ever-treated = 21 (42.0%) | Never-treated = 29 (58.0%) | Leavers = 0 (0.0% of ever)
Cohorts (first 6 rows):
    cohort_label    N_units    Share    N_leavers    ShareLeaversAmongCohort    FirstTreat_time
    ____________    _______    _____    _________    _______________________    _______________

       "2006"         1       0.02         0                  0                      2006
       "2007"        13       0.26         0                  0                      2007
       "2008"         4       0.08         0                  0                      2008
       "2009"         2       0.04         0                  0                      2009
       "2010"         1       0.02         0                  0                      2010

Outcome Y (overall) – N/Mean/SD/Min/P25/Median/P75/Max:
     N       Mean       SD         Min         P25       Median      P75       Max
    ___    _______    _______    ________    _______    _______    ______    ______

    550    1.4058     0.59015    -0.43594    0.97188    1.5373     1.8609    2.6759

Outcome Y by cohort (including 'never'):
    cohort_int    cohort_time    cohort_label      N       Mean        SD          Min          P25       Median
    __________    ___________    ____________    ___    _______    ________    _________    _______    _______

         0             NaN          "never"       319    1.2267     0.57215     -0.17885    0.82622    1.1735
         7            2006          "2006"         11     1.746    0.083851       1.6253      1.702    1.7194
         8            2007          "2007"        143    1.7919     0.43815    -0.039177     1.5925    1.8462
         9            2008          "2008"         44    1.4603     0.74346     -0.43594     1.2017    1.8274
        10            2009          "2009"         22     1.396     0.20724      0.82169     1.3086    1.4626
        11            2010          "2010"         11     1.041     0.26116      0.59264    0.83982    1.2019
```

2 Standard TWFE and Bacon decomposition

To try understand the impact of the Castle Doctrine, we first run a standard two-way fixed effects (TWFE) estimation. This ignores the staggered introduction of the law in different states over time.

```
resTWFE = did.fit("twfe",tbl, idVar="sid", timeVar="year",yVar="l_homicide",
dVar="treatPost",  vcov="clustered", ...
    clusters=["gvar","year"]);
```

```
[TWFE] ATT(treatPost) = 0.0694
[TWFE]
  1×5 table

      Name          Estimate       SE         tStat       pValue
    ____________    ________    ________    _______    _________

    "treatPost"     0.069398    0.028034    2.4756     0.056146
```

The result is an average average effect on the treated (ATT) of 0.069, i.e. homicide rates go up on average by about 7%.

But as we know about the staggered introduction, we conduct the Bacon decomposition, which basically shows how the different comparisons between treated and non-treated states enter the TWFE ATT estimate.

```
resBacon = did.fit("bacon", tbl,idVar="sid", timeVar="year",yVar="l_homicide",
dVar="treatPost");
```

```
---- Bacon: Overall (TWFE) -----
    0.0694

Bacon: ---- Comparisons -----
                type                group1    group2      weight      estimate
    ____________________________    ______    ______    __________    _________

    "Treated vs Never Treated"       2006         0       0.050306      0.14503
    "Treated vs Never Treated"       2007         0        0.61039     0.059254
    "Treated vs Never Treated"       2008         0        0.16098      0.09201
    "Treated vs Never Treated"       2009         0       0.060368      0.18195
    "Treated vs Never Treated"       2010         0       0.016769      0.07399
    "Treated earlier vs later"       2006      2007      0.0045102     0.042003
    "Treated earlier vs later"       2006      2008      0.0027755      0.09126
    "Treated earlier vs later"       2006      2009      0.0020816     0.055197
    "Treated earlier vs later"       2006      2010      0.0013878     -0.05417
    "Treated earlier vs later"       2007      2008       0.021048     -0.02196
    "Treated earlier vs later"       2007      2009       0.021048     0.015093
    "Treated earlier vs later"       2007      2010       0.015786    -0.071019
    "Treated earlier vs later"       2008      2009      0.0037007     -0.15385
    "Treated earlier vs later"       2008      2010      0.0037007     -0.21834
    "Treated earlier vs later"       2009      2010      0.0010408    -0.040521
    "Treated later vs earlier"       2007      2006      0.0030068    -0.043492
    "Treated later vs earlier"       2008      2006      0.0013878     0.053698
    "Treated later vs earlier"       2008      2007      0.0090205     0.043797
    "Treated later vs earlier"       2009      2006     0.00069388      0.13323
    "Treated later vs earlier"       2009      2007      0.0060136      0.14955
    "Treated later vs earlier"       2009      2008     0.00092518     -0.14972
    "Treated later vs earlier"       2010      2006     0.00023129    -0.019355
    "Treated later vs earlier"       2010      2007      0.0022551     0.006548
    "Treated later vs earlier"       2010      2008     0.00046259      -0.2148
    "Treated later vs earlier"       2010      2009     0.00011565     -0.02279

Bacon: ---- Comparisons agg. by Type ----
                Type                  Weight     Estimate
    ____________________________    ________    _________

    "Treated earlier vs later"      0.077079    -0.028577
    "Treated later vs earlier"      0.024112     0.045635
    "Treated vs Never Treated"       0.89881     0.078438
```

The Bacon decomposition shows all pairwise comparisons over different time period combination, for each of the five cohorts whose treatment started in 2006, 2007, 2008, 2009 and 2010, respectively. It shows the cohort-year specific coefficient contribution as well as its weight, i.e. how much the estimate enters the TWFE ATT estimate.

Already the pairwise comparison table shows that TWFE implicitly contains "forbidden" comparisons - in particular "Treated later vs earlier" and therefore potentially introduces bias into the ATT estimate if treatment effects differ between cohorts. The table showing aggregated coefficient by group of comparisons and the respective weights illustrates heterogeneity in cohort effects. However, the weight of potentially troublesome comparisons is relatively low (about 10%) which implies that any bias may not be too severe. This impact depends on the specific setting being analyzed.

Goodman-Bacon shows in his analysis though that weights of bad comparisons can be even negative and absolutely large (depending on cohort sizes and differences in cohort ATTs).

3 Staggered DiD estimation

3.1 Overview

In the next step, we analyze different estimators that are specifically designed to deal with staggered DiD settings. As DiD analysis always aim to estimate a counterfactual component (what would have been the outcome for a treated unit if it would not have been treated), assumption on how to use a control groups pre-treatment characteristics to estimate the counterfactual (the famous "parallel trends" assumption).

The estimators mostly differ in implicit assumption about anticipation of treatment effects and and characteristics of the used control group. Except for one, all estimators also assume that the treatment is absorptive, i.e. it is switched on but never (in the data) goes away. Only the *de Chaisemartin/d'Haultfoeuille (2020)* estimator (CH in the toolbox) can take "on/off"-type of treatments into account. It also differs from the other estimators in measuring treatment effects only in the very first period treatment for each cohort.

3.2 Wooldridge and BJS

Two estimators basically just use an extended OLS framework - *Wooldridge (2021)* (*Wooldridge* in the toolbox) and *Borusyak et al. (2024)*. The Wooldridge estimator specifies one particular regression equation while Borusiak (*BJS* in the toolbox) first runs a regression using only untreated observations and from that estimates treatment effects in a second step.

```
resWool = did.fit("wooldridge", tbl, idVar="sid",
timeVar="year",yVar="l_homicide", dVar="treatPost", ...
    vcov="clustered", clusters="gvar", details=false);
```

```
[wooldridge] ATT by cohort (mean of cohorts coefficients):
    Cohort    #TimePeriods     ATT(k)         SE          tStat         pValue
    ______    ____________    ________    __________    _________    __________

       7           5           0.12768      0.014324        8.9141    0.00029586
       8           4          0.048677      0.011045        4.4071     0.0069753
       9           3          0.083532      0.010422        8.0151    0.00048856
      10           2           0.17615     0.0079329        22.205    3.4407e-06
      11           1           0.07399     2.1721e-09    3.4063e+07    4.1387e-37

[wooldridge] Overall ATT (cohort-share weighted):
    Estimate       SE        tStat       pValue
    ________    _________    ______    __________

    0.072424    0.0099368    7.2884    0.00076085

[wooldridge] ATT by event time (mean/median across cohorts):
    k    ATT_hat_mean    ATT_hat_median    nCells
    _    ____________    ______________    ______

    0      0.088246         0.086855         5
    1        0.1423          0.13335         4
    2       0.08942          0.12652         3
    3      0.098753         0.098753         2
    4      0.033817         0.033817         1
```

In comparison, we could also have used the BJS estimator, that conducts ATT estimation in two steps.

```
resBJS = did.fit("BJS", tbl, idVar="sid", timeVar="year",yVar="l_homicide",
dVar="treatPost");
```

```
[BJS] Untreated-only fit on ALL D==0: N_untreated = 476 | Units = 50 | Periods = 11
[BJS] Treated identified: 74/74 (100.0%)
[BJS] Overall ATT = 0.0669 (SE=0.0593; 95% CI [-0.0493, 0.1831])
ATT by Cohort
```

cohort	ATT_obs	SD_obs	N_obs	SE_obs	CI_lo_obs	CI_hi_obs
7	0.12768	0.033517	5	0.014989	0.086066	0.1693
8	0.048677	0.26921	52	0.037333	-0.026272	0.12363
9	0.083532	0.25941	12	0.074886	-0.081291	0.24835
10	0.17615	0.15191	4	0.075955	-0.06557	0.41787
11	0.07399	0	1	0	0.07399	0.07399

3,3 de Chaisemartin / d'Haultfoeuille (CH)

The CH estimator only measures treatment effects at the first period of treatment per cohort. Also it allows for on/off-treatments. Here we do not illustrate the latter issue but show estimates for the Castle Doctrine data (which has no leavers). Accordingly, in the output the "Leavers-only" row (summary table) and the DID_minus column (cohort aggregation) are empty. The overall effect is 0.01, which is statistically insignificant.

CH also reports a placebo test, where two untreated periods of the cohorts are compared (if available in the data). It basically asks "*Before treatment starts, do the groups that will be treated later already evolve differently from the control groups?*" This placebo test actually raises concerns about the Castle Doctrine data as it fails to be insignificant, thereby showing a potential violation of the parallel trends assumption.

```
resCH = did.fit("DID_M", tbl, idVar="sid", timeVar="year",yVar="l_homicide",
dVar="treatPost");
```

```
=== DID_M summary (table) ===
```

Effect	Estimate	SE	t	p
"DID_M"	0.010336	0.082549	0.12521	0.90087
"Joiners-only"	0.010336	0.082549	0.12521	0.90087
"Leavers-only"	0	0	NaN	NaN
"Overall (cohort-weighted)"	0.010336	0.069657	0.14838	0.88265
"Placebo DID_M"	0.10258	0.043071	2.3816	0.021167

```
=== DID_M by Cohort (table) ===
```

t	DID_plus	DID_minus
2006	0.19373	NaN
2007	0.052498	NaN
2008	-0.22137	NaN
2009	0.21859	NaN
2010	-0.21088	NaN

3.4 Callaway / Sant'Anna (CS)

The CS estimator deals with cohorts and times separately by analyzing ATT(g,t) and then aggregating the results in flexible ways. It is particularly suited to handle covariates, as it uses (up to) two versions of modeling their influence on treated und untreated individuals.

- *Outcome Regression (RA/OR)*: estimates the untreated counterfactual outcome evolution conditional on covariates. This can work even with weak overlap, because it extrapolates the outcome model to treated units whose covariate profiles are not observed in the control group.
- *Inverse Probability Weighting (IPW)*: reweights untreated units so their covariate distribution matches that of the treated group. This directly addresses compositional bias from covariate imbalance.
- *Doubly Robust (DR)*: combines OR and IPW so that the ATT estimate is consistent if either the outcome model or the propensity score model is correctly specified. This provides “extra protection” against model misspecification, and even when both models are slightly wrong, DR typically performs better asymptotically than either method alone.

We use either of the two methods to calculate standard errors ("multiplier" and "clustered"). Clustering is over cohorts and the control group is chosen to be both never as well as not-yet treated units. Note that if one chose the control as being "never" the standard errors become (weirdly) extremely tiny, probably due to the small sample and group sizes (the same effect shows in the Stata analysis).

```
resCS = did.fit("cs", tbl,idVar="sid", timeVar="year",yVar="l_homicide",
dVar="treatPost", ...
                Comparison="notyet", Delta=0, Approach="unconditional",
ClusterVar="gvar",...
                SEMethod="clustered");
```

```
[CS] R=15 cells; Approach=unconditional; Comp=notyet; δ=0; SE=clustered.
[CS] Overall ATT
  1×7 table
```

Estimate	SE	tStat	pValue	LB	UB	crit
0.017412	0.0041567	4.1889	2.8035e-05	0.0092648	0.025559	1.96

```
[CS] Estimates
  15×11 table
```

Name	Effect	Estimate	g	t	SE	tStat	pValue	gYear
"ATT(g=7, t=7)"	"ATT(g,t)"	0.19373	7	7	0.036272	5.3411	9.2395e-08	2006
"ATT(g=7, t=8)"	"ATT(g,t)"	0.30161	7	8	0.0127	23.749	0	2006
"ATT(g=7, t=9)"	"ATT(g,t)"	0.25927	7	9	0.024361	10.643	0	2006
"ATT(g=7, t=10)"	"ATT(g,t)"	0.23841	7	10	0.034641	6.8825	5.8822e-12	2006
"ATT(g=7, t=11)"	"ATT(g,t)"	0.23222	7	11	3.4694e-18	6.6933e+16	0	2006
"ATT(g=8, t=8)"	"ATT(g,t)"	0.052498	8	8	0.022107	2.3747	0.017561	2007
"ATT(g=8, t=9)"	"ATT(g,t)"	-0.039399	8	9	0.0064293	-6.1281	8.8953e-10	2007
"ATT(g=8, t=10)"	"ATT(g,t)"	0.018147	8	10	0.0040529	4.4776	7.5489e-06	2007
"ATT(g=8, t=11)"	"ATT(g,t)"	-0.019152	8	11	1.1943e-17	-1.6036e+15	0	2007
"ATT(g=9, t=9)"	"ATT(g,t)"	-0.22137	9	9	0.018034	-12.275	0	2008
"ATT(g=9, t=10)"	"ATT(g,t)"	0.11019	9	10	0.023126	4.7645	1.8933e-06	2008
"ATT(g=9, t=11)"	"ATT(g,t)"	0.01415	9	11	1.7347e-17	8.157e+14	0	2008
"ATT(g=10, t=10)"	"ATT(g,t)"	0.21859	10	10	0.0051232	42.667	0	2009
"ATT(g=10, t=11)"	"ATT(g,t)"	0.033923	10	11	5.2042e-18	6.5185e+15	0	2009
"ATT(g=11, t=11)"	"ATT(g,t)"	-0.21088	11	11	3.4694e-18	-6.0781e+16	0	2010

```
[CS] By Cohort
  5×8 table
```

g	Estimate	SE	tStat	pValue	LB	UB	gYear
7	0.24505	0.017749	13.806	0	0.21026	0.27984	2006
8	0.0030236	0.0056803	0.5323	0.59452	-0.0081098	0.014157	2007

9	-0.032344	0.0137	-2.3608	0.018235	-0.059196	-0.0054913	2008
10	0.12626	0.0025616	49.288	0	0.12124	0.13128	2009
11	-0.21088	3.4694e-18	-6.0781e+16	0	-0.21088	-0.21088	2010

```
resCS = did.fit("cs", tbl, ...
    idVar="sid", timeVar="year", yVar="l_homicide", dVar="treatPost", ...
    Comparison="notyet", Delta=0, Approach="unconditional", ...
    SEMethod="multiplier", ...
    B=999, rngSeed=42, multiplier="mammen");
```

```
[CS] R=15 cells; Approach=unconditional; Comp=notyet; δ=0; SE=multiplier. Bootstrap B=999, crit(95%)=2.703
[CS] Overall ATT
  1×6 table
```

Estimate	SE	tStat	pValue	LB	UB
0.017412	0.035535	0.49	0.62062	-0.053353	0.088177

```
[CS] Estimates
  15×11 table
```

Name	Effect	Estimate	g	t	SE	tStat	pValue	gYear	tYe
"ATT(g=7, t=7)"	"ATT(g,t)"	0.19373	7	7	0.028105	6.8933	0	2006	200
"ATT(g=7, t=8)"	"ATT(g,t)"	0.30161	7	8	0.035921	8.3964	0	2006	200
"ATT(g=7, t=9)"	"ATT(g,t)"	0.25927	7	9	0.052913	4.8999	0	2006	200
"ATT(g=7, t=10)"	"ATT(g,t)"	0.23841	7	10	0.041531	5.7406	0	2006	200
"ATT(g=7, t=11)"	"ATT(g,t)"	0.23222	7	11	0.039392	5.8951	0	2006	201
"ATT(g=8, t=8)"	"ATT(g,t)"	0.052498	8	8	0.047606	1.1028	0.27528	2007	200
"ATT(g=8, t=9)"	"ATT(g,t)"	-0.039399	8	9	0.050917	-0.77379	0.45045	2007	200
"ATT(g=8, t=10)"	"ATT(g,t)"	0.018147	8	10	0.054796	0.33118	0.74474	2007	200
"ATT(g=8, t=11)"	"ATT(g,t)"	-0.019152	8	11	0.047437	-0.40374	0.6977	2007	201
"ATT(g=9, t=9)"	"ATT(g,t)"	-0.22137	9	9	0.24704	-0.89608	0.40841	2008	200
"ATT(g=9, t=10)"	"ATT(g,t)"	0.11019	9	10	0.072726	1.5151	0.12513	2008	200
"ATT(g=9, t=11)"	"ATT(g,t)"	0.01415	9	11	0.10423	0.13576	0.87588	2008	201
"ATT(g=10, t=10)"	"ATT(g,t)"	0.21859	10	10	0.10365	2.109	0.019019	2009	200
"ATT(g=10, t=11)"	"ATT(g,t)"	0.033923	10	11	0.047383	0.71594	0.49449	2009	201
"ATT(g=11, t=11)"	"ATT(g,t)"	-0.21088	11	11	0.033503	-6.2942	0	2010	201

```
[CS] By Cohort
  5×8 table
```

g	Estimate	SE	tStat	pValue	LB	UB	gYear
7	0.24505	0.030342	8.0761	0	0.17095	0.31914	2006
8	0.0030236	0.032812	0.092151	0.93193	-0.0771	0.083147	2007
9	-0.032344	0.12957	-0.24963	0.75676	-0.34874	0.28405	2008
10	0.12626	0.068494	1.8433	0.069069	-0.041001	0.29351	2009
11	-0.21088	0.033503	-6.2942	0	-0.29269	-0.12907	2010

4 Conclusions

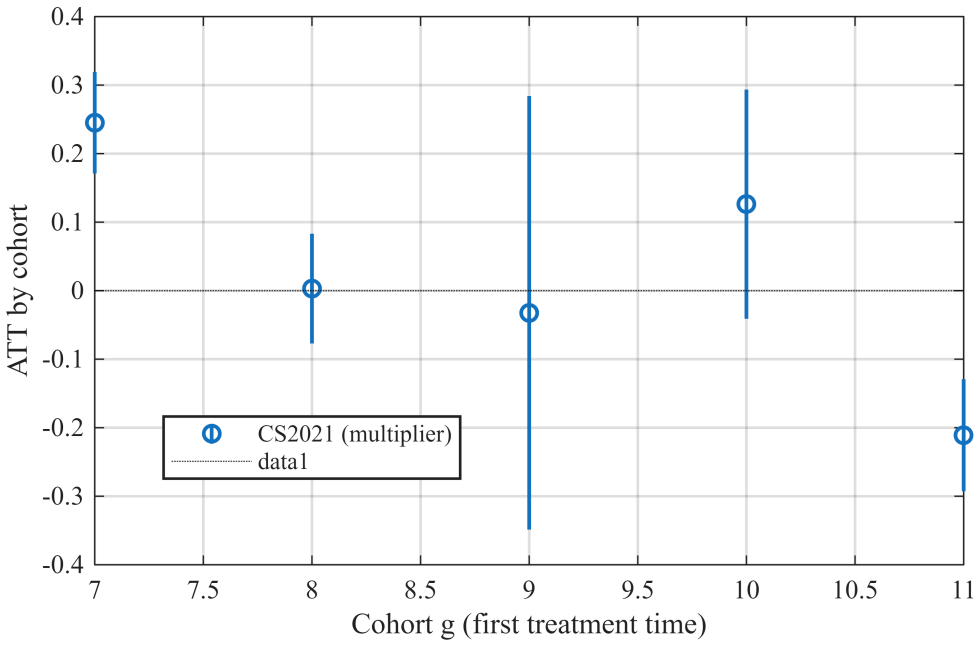
This example served to illustrate the DiD Toolbox for Matlab in a real world analysis setting, looking at the impact of the Castle Doctrine on homicides in the U.S.

In terms of economic results, the original paper documents an increase in homicides between 5% - 8%, depending on their model specification. When using the nowadays available modern estimators for staggered DiD analyses, the evidence becomes less strong. To illustrate, the following graphs compare the estimated cohort effects for *Callayway/Sant'Anna* (CS) and *Borusyak et al.* (BJS) showing that only for two cohorts (2006 and 2010) effects appear to be statistically significant (with a lot of heterogeneity across

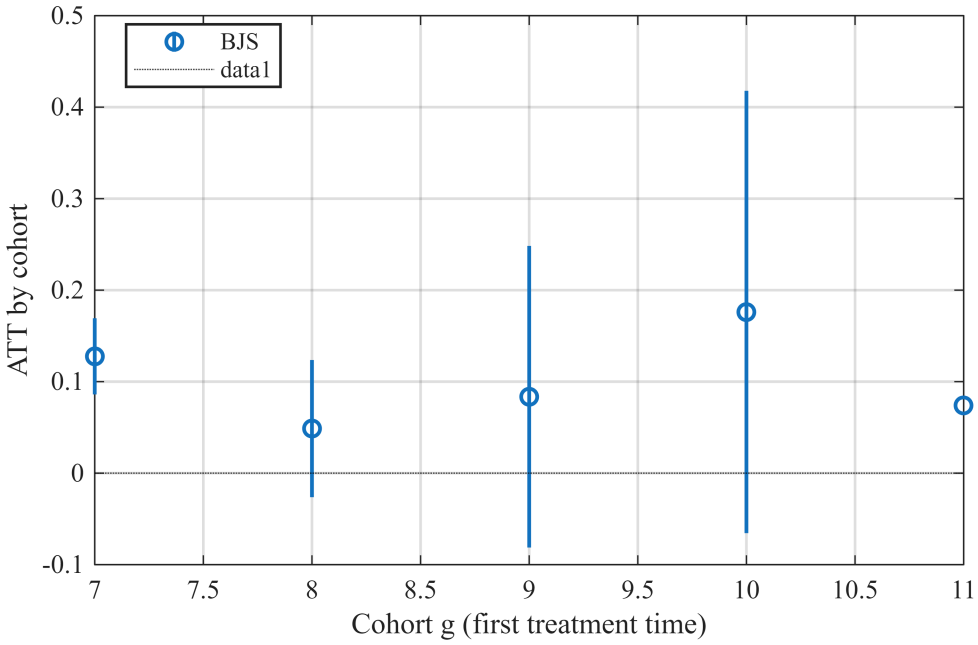
cohorts in term of point estimates). Both cohorts, however, comprise only two federal states, Florida and Montana, respectively.

Also, the *de Chaisemartin/d'Haultfoeuille* (CH) estimator indicates potential violations of the parallel trend assumption, as the placebo test (comparing untreated periods of cohorts) has a significant finding, while this doesn't hold for the actual analysis.

```
did.did_plot(resCS, "cohort");
```




```
did.did_plot(resBJS,"cohort");
```



Literature

Cheng Cheng and Mark Hoekstra (2013): Does Strengthening Self-Defense Law Deter Crime or Escalate Violence? Evidence from Expansions to Castle Doctrine,The Journal of Human Resources, Vol. 48, No. 3, pp. 821-853.

Callaway, Brantly, and Pedro H. C. Sant'Anna. (2021): "Difference-in-Differences with Multiple Time Periods", *Journal of Econometrics* 225 (2): 200–230.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. American Economic Review, 110(9), 2964–2996.

Sant'Anna, Pedro H. C., and Jun Zhao (2020): "Doubly Robust Difference-in-Differences Estimators", *Journal of Econometrics* 219 (1): 101–122.