Quant II

Lab 7: DiD

Giacomo Lemoli

March 23, 2023

Today's plan

- Problems with TWFE under staggered treatment and alternatives
  - `did`, `did_multiplegt`, `twowayfeweights`
- Violations of parallel trends
  - `honestDID`

Workhorse example

Mafia and political quality: Daniele & Geys (EJ 2015)

ORGANISED CRIME, INSTITUTIONS AND POLITICAL QUALITY: EMPIRICAL EVIDENCE FROM ITALIAN MUNICIPALITIES*

Gianmarco Daniele and Benny Geys

This article assesses how legal institutions affect the influence of politically active criminal organisations on the human capital of elected politicians using data from over 1,500 Southern Italian municipalities in the period 1985–2011. It exploits municipal government dissolutions imposed by the national government for (presumed) mafia infiltration as a source of exogenous variation in the presence of politically active criminal organisations. The results support theoretical predictions that the average education level of local politicians significantly increases when active mafia infiltration of local politics is remedied through the implementation of a stricter legal-institutional framework.

Workhorse example

- What is the effect of removing mafia connections on the quality of politicians?
- Staggered dissolution of municipal governments in Italian south after crimina investigations
- Outcome: education of government members (average)
- "Unlike the traditional DiD model, which relies on a shock at one point in time across all treated jurisdictions, we can exploit the fact that dissolutions did not take place at the same point in time for each municipality [...] to further strengthen our identification"

Staggered treatment

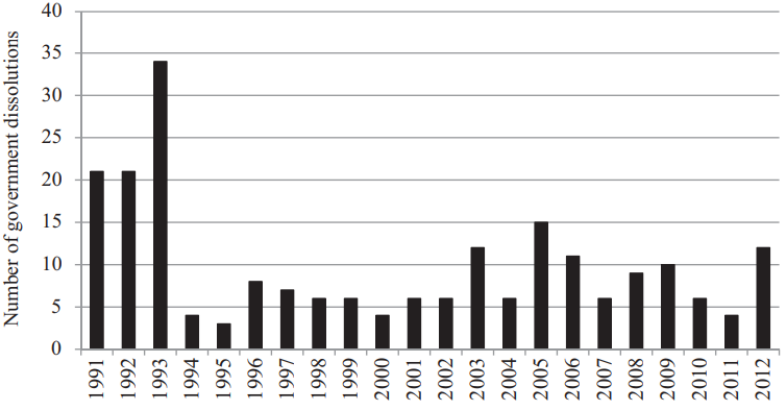


Fig. 1. *Local Governments Dissolved for (Suspected) Mafia Infiltration, 1991–2012*
Notes. Data up to 2007 derive from the website of the Italian Parliament '*Commissione Parlamentare Antimafia*' (available at http://www.camera.it/_bicamerali/leg15/commbicantimafia/documentazionetematica/23/schedabase.asp). The remaining years were manually collected by the authors.

TWFE: Stata implementation

TWFE: R implementation

```
library(tidyverse); library(haven); library(fixest)

dat <- read_dta("Organized Crime and Political Quality.dta")

dat <- mutate(dat, befcomgeneral_recode=(1-befcomgeneral)*mafiaben)

feols(MeanEduPol ~ befcomgeneral_recode + mafiaben | ID_municip[trend] + year, dat) %>% etable()
```

```
##                                         .
## Dependent Var.:              MeanEduPol
##
## befcomgeneral_recode 0.3226** (0.1249)
## Fixed-Effects:       -----------------
## ID_municip                         Yes
## year                               Yes
## Varying Slopes:      -----------------
## trend (ID_municip)                 Yes
## ____________________ _________________
## S.E.: Clustered         by: ID_municip
## Observations                    39,040
## R2                             0.78186
## Within R2                      0.00082
```

Callaway & Sant'Anna

- G_g: group/cohort of units treated at $t = g$
- t: time
- C: group of never-treated units
- Target quantity: $ATT(g, t) = \mathbb{E}[Y_t(g) - Y_t(0) \mid G_g = 1]$
- Assumptions: conditional parallel trends based on "never-treated" or "not-yet-treated" group, overlap

Callaway & Sant'Anna

- First, define treatment groups based on their first year of treatment

```
library(did)

# Group variable
dat <- dat %>% group_by(desc_comune) %>%
  mutate(group = case_when(mafiaben==1 ~ min(year[befcomgeneral_recode==1]),
                           mafiaben==0 ~ 0)) %>%
  ungroup()
```

Callaway & Sant'Anna

Basic implementation: never treated as controls, estimation through outcome regression

$$ATT_{or}^{new}(g,t,\delta) = \mathbb{E}\left[\frac{G_g}{\mathbb{E}[G_g]}(Y_t - Y_{g-\delta-1} - m_{g,t,\delta}^{nev}(X))\right]$$

$$m_{g,t,\delta}^{nev}(X) = \mathbb{E}[Y_t - Y_{g-\delta-1} \mid X, C = 1]$$

Callaway & Sant'Anna

Basic implementation

```
set.seed(1)
out <- att_gt(yname = "MeanEduPol",       # outcome variable
              gname = "group",            # group variable
              idname = "ID_municip",      # unit identifier
              tname = "year",             # year variable
              xformla = ~1,               # covariates (if any)
              data = dat,                 # data
              est_method = "reg",         # estimation method
              allow_unbalanced_panel = T # allow unbalanced panel
              )
```

Callaway & Sant'Anna

Visualize the results:

```
summary(out)
```

```
##
## Call:
## att_gt(yname = "MeanEduPol", tname = "year", idname = "ID_municip",
##     gname = "group", xformla = ~1, data = dat, allow_unbalanced_panel = T,
##     est_method = "reg")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-Differences with Multiple Time Period
##
## Group-Time Average Treatment Effects:
##  Group Time ATT(g,t) Std. Error [95% Simult.  Conf. Band]
##   1991 1986   0.1347     0.2383       -1.1616      1.4310
##   1991 1987   0.4434     0.4952       -2.2508      3.1376
##   1991 1988  -0.9035     0.4919       -3.5798      1.7729
##   1991 1989  -0.0684     0.1580       -0.9278      0.7911
##   1991 1990  -0.3298     0.1211       -0.9888      0.3292
##   1991 1991       NA         NA            NA          NA
##   1991 1992       NA         NA            NA          NA
##   1991 1993   0.7169     0.2749       -0.7785      2.2123
##   1991 1994   0.1444     0.3069       -1.5253      1.8141
##   1991 1995   0.0422     0.2801       -1.4815      1.5659
##   1991 1996  -0.1051     0.3098       -1.7904      1.5802
##   1991 1997   0.0749     0.3350       -1.7477      1.8974
##   1991 1998   0.4527     0.3157       -1.2647      2.1701
##   1991 1999   0.2969     0.3037       -1.3552      1.9490
##   1991 2000   0.2517     0.2727       -1.2318      1.7351
##   1991 2001   0.4887     0.3414       -1.3688      2.3462
##   1991 2002   0.2587     0.3715       -1.7623      2.2798
##   1991 2003   0.3802     0.3540       -1.5459      2.3063
##   1991 2004   0.3284     0.3314       -1.4748      2.1316
```

Callaway & Sant'Anna

Tidy format for result manipulation

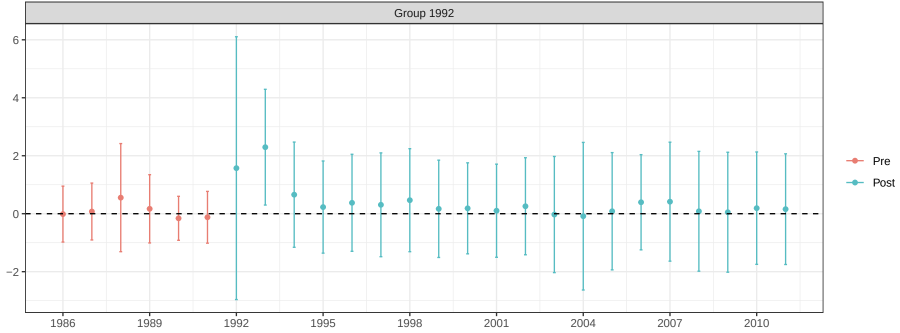
```r
tidy(out) %>% head()
```

```
##             term group time    estimate std.error   conf.low conf.high
## 1 ATT(1991,1986)  1991 1986  0.13472577 0.2382629 -1.1615660 1.4310175
## 2 ATT(1991,1987)  1991 1987  0.44341118 0.4952030 -2.2507877 3.1376101
## 3 ATT(1991,1988)  1991 1988 -0.90345925 0.4919250 -3.5798241 1.7729056
## 4 ATT(1991,1989)  1991 1989 -0.06835298 0.1579756 -0.9278345 0.7911285
## 5 ATT(1991,1990)  1991 1990 -0.32980297 0.1211220 -0.9887788 0.3291728
## 6 ATT(1991,1991)  1991 1991          NA        NA         NA        NA
##   point.conf.low point.conf.high
## 1     -0.3322608      0.60171238
## 2     -0.5271688      1.41399115
## 3     -1.8676145      0.06069602
## 4     -0.3779795      0.24127356
## 5     -0.5671977     -0.09240821
## 6             NA              NA
```

Callaway & Sant'Anna

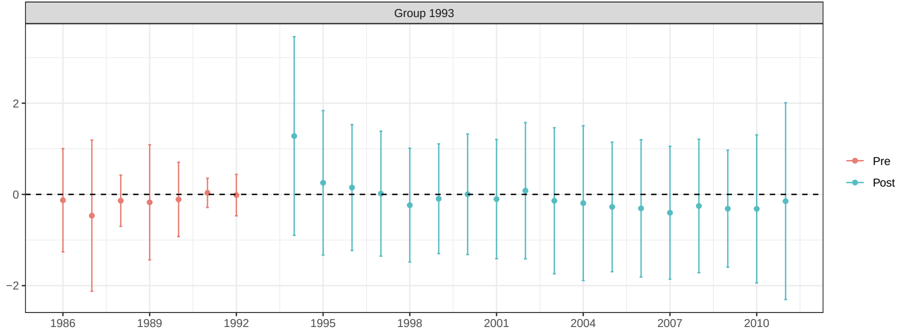
Graphical visualization

```
# By default makes a plot for each group, here we select a few because there are many
ggdid(out, group = 1992, xgap=3, title="", theming=F)+ theme_bw()
```



Graphical visualization

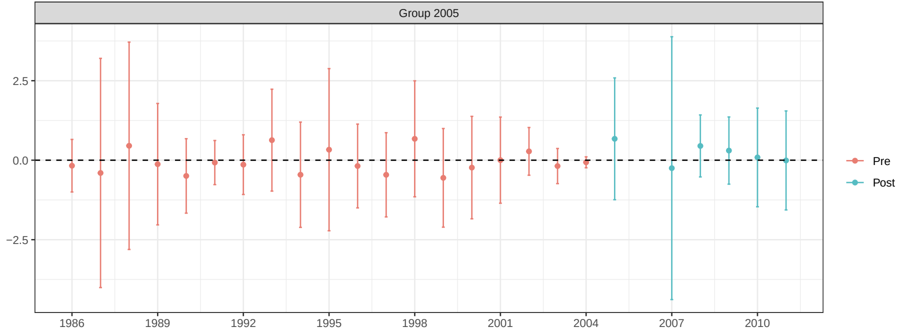
```
# By default makes a plot for each group, here we select a few because there are many
ggdid(out, group = 1993, xgap=3, title="", theming=F)+ theme_bw()
```



Callaway & Sant'Anna

Graphical visualization

```
# By default makes a plot for each group, here we select a few because there are many
ggdid(out, group = 2005, xgap=3, title="", theming=F)+ theme_bw()
```



Callaway & Sant'Anna

- Change the estimation method through the `est_method` argument
- `ipw`: weighting estimator

$$ATT^{new}_{ipw}(g,t,\delta) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1})\right]$$

- `dr`: doubly robust estimator

$$ATT^{new}_{dr}(g,t,\delta) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1} - m^{nev}_{g,t,\delta}(X$$

Aggregate estimates

- Focus on $ATT(g,t)$ fully exploits TE heterogeneity (cost: more uncertainty)
- Aggregate the single estimates to obtain quantities of theoretical interest
- For instance, simple weighted average of all $ATT(g,t)$

```
aggte(out, type ="simple", na.rm=T)
```

```
##
## Call:
## aggte(MP = out, type = "simple", na.rm = T)
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-Differences with Multiple Time Period
##
##
##     ATT    Std. Error     [ 95%  Conf. Int.]
##  0.1961         0.117    -0.0332      0.4255
##
##
## ---
## Signif. codes: `*' confidence band does not cover 0
##
## Control Group:  Never Treated,  Anticipation Periods:  0
## Estimation Method:  Outcome Regression
```

Aggregate estimates

Dynamic ATTs relative to treatment period (event-study)

```
aggte(out, type = "dynamic", na.rm=T) %>% tidy() %>%
  filter(event.time%in%seq(-5, 5, 1)) %>% select(term, event.time, estimate, std.error, conf.low, conf.high)
```

```
##       term event.time     estimate  std.error     conf.low  conf.high
## 1  ATT(-5)         -5 -0.007734229 0.06426679 -0.193430687 0.17796223
## 2  ATT(-4)         -4  0.052855514 0.11742452 -0.286438159 0.39214919
## 3  ATT(-3)         -3 -0.184326353 0.09069612 -0.446389350 0.07773664
## 4  ATT(-2)         -2 -0.005523172 0.05109552 -0.153161720 0.14211538
## 5  ATT(-1)         -1 -0.045999271 0.04540994 -0.177209552 0.08521101
## 6   ATT(0)          0  0.905711129 0.33291270 -0.056227458 1.86764972
## 7   ATT(1)          1  0.789464457 0.23784243  0.102227803 1.47670111
## 8   ATT(2)          2  0.546082202 0.15749660  0.091001762 1.00116264
## 9   ATT(3)          3  0.330062212 0.11474200 -0.001480437 0.66160486
## 10  ATT(4)          4  0.271588055 0.10872579 -0.042570968 0.58574708
## 11  ATT(5)          5  0.161634985 0.11137843 -0.160188764 0.48345873
```

Aggregate estimates

Group-level ATTs

```
aggte(out, type="group", na.rm=T) %>% tidy() %>%
  filter(group%in%c("1992", "1993", "2005")) %>%
  select(term, group, estimate, std.error, conf.low, conf.high)
```

```
##        term group    estimate std.error   conf.low conf.high
## 1 ATT(1992)  1992  1.57183593 0.7678492 -0.5562787 3.6999505
## 2 ATT(1993)  1993 -0.04837113 0.2144896 -0.6428349 0.5460926
## 3 ATT(2005)  2005  0.12960768 0.2082879 -0.4476678 0.7068831
```

De Chaisemartin & D'Haultfoeuille

Main result in de Chaisemartin and D'Haultfœuille (2020)

$$E[\hat{\beta}_{fe}] = \left[\sum_{(g,t):D_{g,t} \neq 0} W_{g,t} TE_{g,t} \right]$$

where

$$TE_{g,t} = \bar{Y}_{g,t}(1) - \bar{Y}_{g,t}(0)$$
$$W_{g,t} = \frac{N_{g,t}}{N_1} \frac{e_{g,t}}{\sum_{(g,t):D_{g,t} \neq 0} \frac{N_{g,t}}{N_1} e_{g,t}}$$
$$e_{g,t} = D_{g,t} - D_{g,.} - (D_{.,t} - D_{.,.})$$

Different from ATT formula, where the weights are $\frac{N_{g,t}}{N_1}$

De Chaisemartin & D'Haultfoeuille

Implications:

- $W_{g,t}$ sum to 1, and some may be negative
- TEs with a negative weight enter the sum with opposite sign
- The estimate $\hat{\beta}_{fe}$ can even have the different sign
- Arises if TEs are correlated with weights or if TE is heterogeneous
- Why: comparison of newly treated ("switchers") to already treated groups

De Chaisemartin & D'Haultfoeuille

Two packages:

- `twowayfeweights`: estimate effect weights in the TWFE regression and diagnose possible problems
- `did_multiplegt`: implement an alternative estimator

$$DID_M = \sum_{t=2}^{T} \left(\frac{N_{1,0,t}}{N_S} DID_{+,t} + \frac{N_{0,1,t}}{N_S} DID_{-,t} \right)$$

where

$$DID_{+,t} = \sum_{g:D_{g,t}=1,D_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=D_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}}(Y_{g,t} - Y_{g,t-1})$$

$$DID_{-,t} = \sum_{g:D_{g,t}=D_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}}(Y_{g,t} - Y_{g,t-1}) - \sum_{g:D_{g,t}=0,D_{g,t-1}=1} \frac{N_{g,t}}{N_{0,1,t}}(Y_{g,t} - Y_{g,t-1})$$

Stata implementation

Honest DiD

- These methods fix estimation problems, not identification
- Rely on (conditional) parallel trends
- How do we address possible violations of the identification assumption?
- Arambachan & Roth (2022): sensitivity analysis approach
- Idea: use PT violations detected in pre-treatment periods to inform possible violations after the treatment

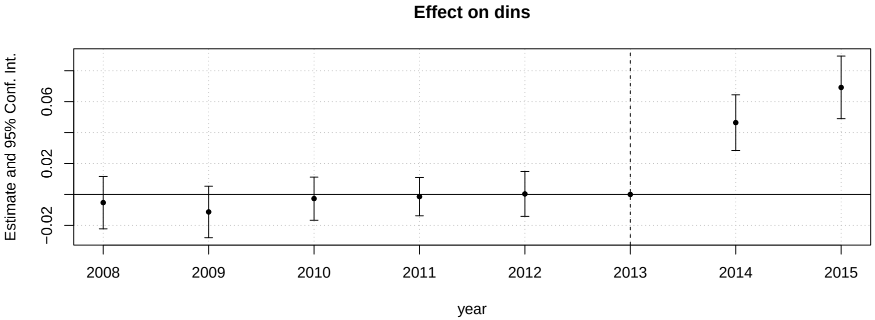
Honest DiD

Simple DiD (non-staggered)

```
#remotes::install_github("asheshrambachan/HonestDiD")

mt <- read_dta("https://raw.githubusercontent.com/Mixtape-Sessions/Advanced-DID/main/Exercises/Data/ehec_data.d

# Make a simple non-staggered version
mt_ns <- mt %>% filter(year < 2016 & (is.na(yexp2)| yexp2 != 2015))

# Create a treatment dummy
mt_ns <- mt_ns %>% mutate(D = case_when(yexp2 == 2014 ~ 1, T ~ 0))

# TWFE model
twfe <- feols(dins ~ i(year, D, ref = 2013) | stfips + year,
                        cluster = "stfips",
                        data = mt_ns)
```

Honest DiD

```
iplot(twfe)
```



Honest DiD

Sensitivity analysis 1: relative magnitude restrictions of PT violations

Call δ the difference in trends between treated and control

$$\Delta^{RM}(\bar{M}) = \{\delta : \forall t \geq 0, |\delta_{t+1} - \delta_t| \leq \bar{M} \times \max_{s<0} |\delta_{s+1} - \delta_s|\}$$

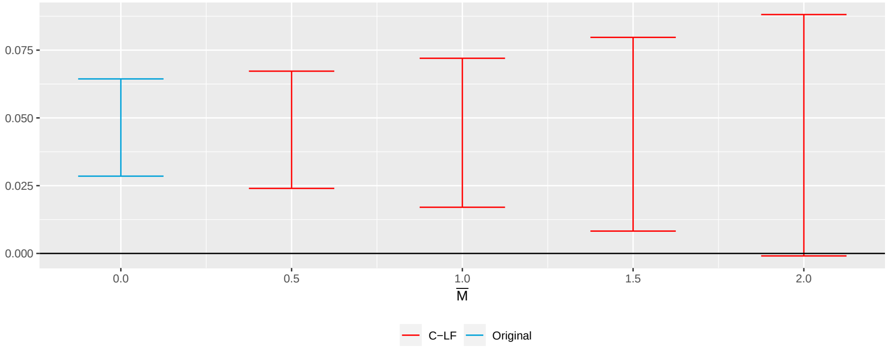
Implementation

```r
# Extract coefficients and vcov matrix
betahat <- summary(twfe)$coefficients
sigma <- summary(twfe)$cov.scaled

(delta_rm_results <-
  HonestDiD::createSensitivityResults_relativeMagnitudes(
    betahat = betahat, #coefficients
    sigma = sigma, #covariance matrix
    numPrePeriods = 5, #num. of pre-treatment coefs
    numPostPeriods = 2, #num. of post-treatment coefs
    Mbarvec = seq(0.5,2,by=0.5) #values of Mbar
    ))
```

```
## # A tibble: 4 x 5
##          lb     ub method Delta    Mbar
##       <dbl>  <dbl> <chr>  <chr>   <dbl>
## 1  0.0240   0.0672 C-LF   DeltaRM   0.5
## 2  0.0170   0.0720 C-LF   DeltaRM   1
## 3  0.00824  0.0797 C-LF   DeltaRM   1.5
## 4 -0.000916 0.0881 C-LF   DeltaRM   2
```

Sensitivity plot

```
originalResults <- HonestDiD::constructOriginalCS(betahat = betahat,
                                                  sigma = sigma,
                                                  numPrePeriods = 5,
                                                  numPostPeriods = 2)

HonestDiD::createSensitivityPlot_relativeMagnitudes(delta_rm_results, originalResults)
```



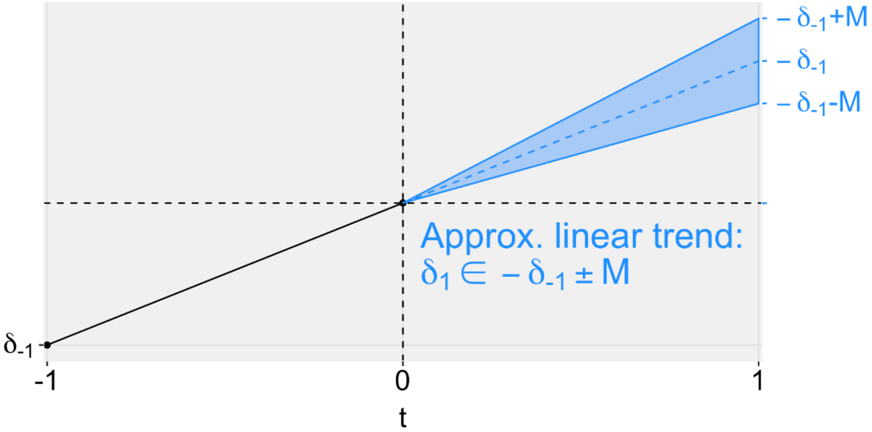
Honest DiD

Sensitivity analysis 2: smoothness restrictions

$$\Delta^{SD}(M) \equiv \{\delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M, \forall t\}$$

Group linear trends correspond to the case where $M = 0$

Honest DiD



$-\delta_{-1}+M$
$-\delta_{-1}$
$-\delta_{-1}-M$
Approx. linear trend:
$\delta_1 \in -\delta_{-1} \pm M$
δ_{-1}
-1
0
1
t

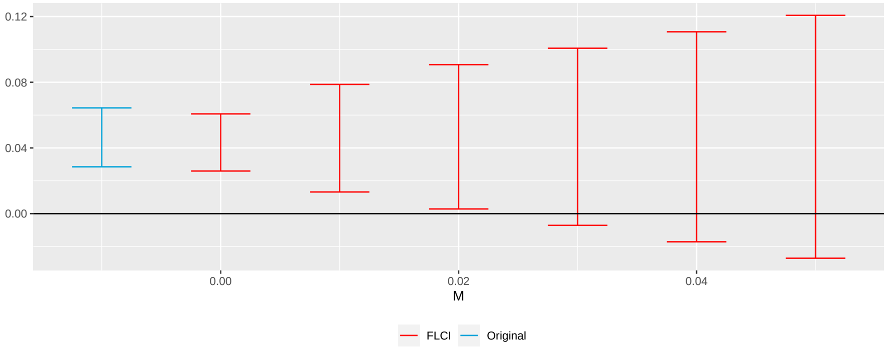
Honest DiD

```r
(delta_sd_results <-
  HonestDiD::createSensitivityResults(betahat = betahat,
                                      sigma = sigma,
                                      numPrePeriods = 5,
                                      numPostPeriods = 2,
                                      Mvec = seq(from = 0, to = 0.05, by =0.01)))
```

```
## # A tibble: 6 x 5
##         lb     ub method Delta       M
##      <dbl>  <dbl> <chr>  <chr>   <dbl>
## 1  0.0259  0.0607 FLCI   DeltaSD  0
## 2  0.0132  0.0787 FLCI   DeltaSD  0.01
## 3  0.00286 0.0907 FLCI   DeltaSD  0.02
## 4 -0.00714 0.101  FLCI   DeltaSD  0.03
## 5 -0.0171  0.111  FLCI   DeltaSD  0.04
## 6 -0.0271  0.121  FLCI   DeltaSD  0.05
```

Honest DiD

```
HonestDiD::createSensitivityPlot(delta_sd_results, originalResults)
```



Honest DiD

Accomodates non-staggered treatment timing, integration in `did` not yet available

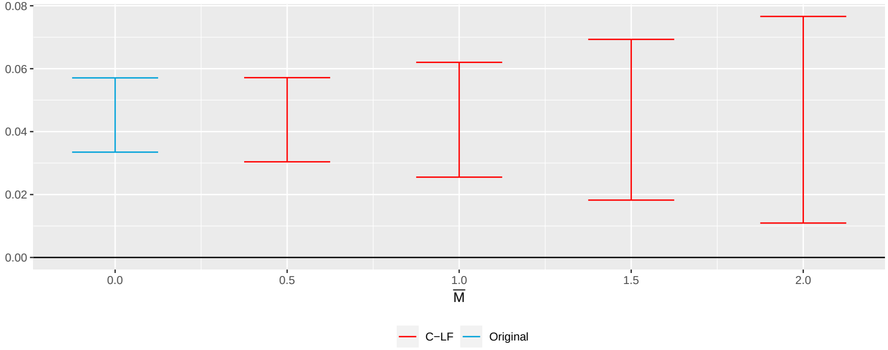
```
## Auxiliary function by Sant'Anna to transport objects from did to honestdid
# (Omitted)
```

Honest DiD

```
cs_results <- att_gt(yname = "dins",
                     tname = "year",
                     idname = "stfips",
                     gname = "yexp2",
                     data = mt %>% mutate(yexp2 = ifelse(is.na(yexp2), 3000, yexp2)),
                     control_group = "notyettreated",
                     base_period = "universal")

es <- aggte(cs_results, type = "dynamic",
            min_e = -5, max_e = 5)

#Run sensitivity analysis for relative magnitudes
sensitivity_results <-
  honest_did.AGGTEobj(es,
                      e =0,
                      type = "relative_magnitude",
                      Mbarvec = seq(from = 0.5, to = 2, by = 0.5))
```

Honest DiD

```
HonestDiD::createSensitivityPlot_relativeMagnitudes(sensitivity_results$robust_ci,
                                                    sensitivity_results$orig_ci)
```



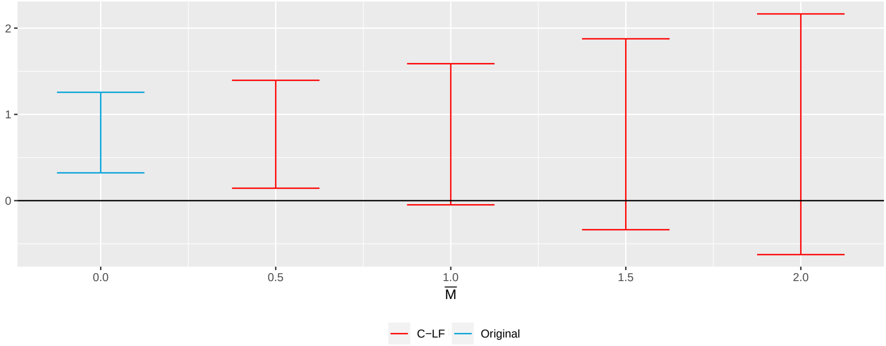
Honest DiD

Back to the mafia example:

```
out2 <- att_gt(yname = "MeanEduPol", gname = "group", idname = "ID_municip",
               tname = "year", xformla = ~1, data = dat,
               est_method = "reg", allow_unbalanced_panel = T,
               base_period = "universal")

es <- aggte(out2, type = "dynamic", na.rm=T, min_e = -5, max_e = 5)

sensitivity_results <-
  honest_did.AGGTEobj(es,
                      e =1,
                      type = "relative_magnitude",
                      Mbarvec = seq(from = 0.5, to = 2, by = 0.5))
```

Honest DiD

```
HonestDiD::createSensitivityPlot_relativeMagnitudes(sensitivity_results$robust_ci,
                                                    sensitivity_results$orig_ci)
```



Other packages

Diagnostics:

- Estimate the weights $W_{g,t}$ in de Chaisemartin and D'Haultfœuille (2020): `TwoWayFEWeights` (R), `twowayfeweights` (Stata)
- Bacon decomposition of TWFE and weights associated: `bacondecomp` (R/Stata)
- Decomposition of TWFE coefficients in event-study design (Sun and Abraham 2020): `eventstudyweights` (Stata)

Estimators:

- de Chaisemartin and D'Haultfœuille (2020): `DIDmultiplegt` (R), `did_multiplegt` (Stata)
- Callaway and Sant'Anna (2021): `did` (R), `csdid` (Stata)
- Sun and Abraham (2020): `fixest::sunab`, `staggered_sa` (R), `eventstudyinteract` (Stata)
- Borusyak et al (2021): `didimputation` (R), `did_imputation` (Stata)
- Wrapper for many of the above: `did2s` (R)

Take-away

- Staggered treatment adoption is one of the most recurring settings in empirical social science
- New estimators address problems linear regression
- Core intuition: decompose the DiD into multiple “clean” DiDs
- Identification assumptions are the most important thing