Package ‘did2s’

July 21, 2022

Title Two-Stage Difference-in-Differences Following Gardner (2021)
Version 0.7.0
Description Estimates Two-way Fixed Effects difference-in-differences/event-study models using the approach proposed by Gardner (2021). To avoid the problems caused by OLS estimation of the Two-way Fixed Effects model, this function first estimates the fixed effects and covariates using untreated observations and then in a second stage, estimates the treatment effects.
License MIT + file LICENSE
Encoding UTF-8
LazyData true
RoxygenNote 7.2.0
Depends R (>= 3.5.0), fixest (>= 0.10.1)
Imports data.table, ggplot2, glue, stats, stringr, rlang, cli, Matrix, Rcpp, did, broom, methods, staggered, didimputation, boot
URL http://kylebutts.com/did2s/
Suggests rmarkdown, knitr, haven, testthat (>= 3.0.0)
VignetteBuilder knitr
LinkingTo RcppArmadillo, Rcpp
Config/testthat/edition 3
NeedsCompilation yes
Author Kyle Butts [aut, cre] (<https://orcid.org/0000-0002-9048-8059>), John Gardner [aut] (<https://orcid.org/0000-0002-4028-6862>), Grant McDermott [ctb] (<https://orcid.org/0000-0001-7883-8573>), Laurent Berge [ctb]
Maintainer Kyle Butts <kyle.butts@colorado.edu>
Repository CRAN
Date/Publication 2022-07-21 09:40:02 UTC
R topics documented:

castle ........................................ 2
df_het ......................................... 2
df_hom .......................................... 3
did2s ........................................... 4
event_study .................................... 8
gen_data ........................................ 9

Description

State-wide panel data from 2000-2010 that has information on castle-doctrine, the so-called "stand-your-ground" laws that were implemented by 20 states.

Usage

castle

Format

A data frame with 550 rows and 5 variables:

sid state id, unit of observation
year time in panel data
l_homicide log of the number of homicides per capita
effyear year that castle doctrine is passed
post 0/1 variable for when castle doctrine is active
time_till time relative to castle doctrine being passed into law

df_het

Simulated data with two treatment groups and heterogenous effects

Description

Generated using the following call: did2s::gen_data(panel = c(1990, 2020), g1 = 2000, g2 = 2010, g3 = 0, te1 = 2, te2 = 1, te3 = 0, te_m1 = 0.05, te_m2 = 0.15, te_m3 = 0)

Usage

df_het
df_hom

**Format**

A data frame with 31000 rows and 15 variables:

- **unit** individual in panel data
- **year** time in panel data
- **g** the year that treatment starts
- **dep_var** outcome variable
- **treat** T/F variable for when treatment is on
- **rel_year** year relative to treatment start. Inf = never treated.
- **rel_year_binned** year relative to treatment start, but <=-6 and >=6 are binned.
- **unit_fe** Unit FE
- **year_fe** Year FE
- **error** Random error component
- **te** Static treatment effect = te
- **te_dynamic** Dynamic treatment effect = te_m
- **state** State that unit is in
- **group** String name for group

---

**Description**

Generated using the following call: `did2s::gen_data(panel = c(1990, 2020), g1 = 2000, g2 = 2010, g3 = 0, te1 = 2, te2 = 2, te3 = 0, te_m1 = 0, te_m2 = 0, te_m3 = 0)`

**Usage**

df_hom

**Format**

A data frame with 31000 rows and 15 variables:

- **unit** individual in panel data
- **year** time in panel data
- **g** the year that treatment starts
- **dep_var** outcome variable
- **treat** T/F variable for when treatment is on
- **rel_year** year relative to treatment start. Inf = never treated.
- **rel_year_binned** year relative to treatment start, but <=-6 and >=6 are binned.
did2s

This function calculates two-stage difference-in-differences following Gardner (2021).

**Description**

Calculate two-stage difference-in-differences following Gardner (2021)

**Usage**

```r
did2s(
  data,
  yname,
  first_stage,
  second_stage,
  treatment,
  cluster_var,
  weights = NULL,
  bootstrap = FALSE,
  n_bootstraps = 250,
  return_bootstrap = FALSE,
  verbose = TRUE
)
```

**Arguments**

- `data`: The dataframe containing all the variables
- `yname`: Outcome variable
- `first_stage`: Fixed effects and other covariates you want to residualize with in the first stage. Formula following `fixest::feols`. Fixed effects specified after `|`.
- `second_stage`: Second stage, these should be the treatment indicator(s) (e.g. treatment variable or event-study leads/lags). Formula following `fixest::feols`. Use `i()` for factor variables, see `fixest::i`.
- `treatment`: A variable that = 1 if treated, = 0 otherwise
- `cluster_var`: String name for group
- `weights`: Weight from `runif()`
- `bootstrap`, `n_bootstraps`, `return_bootstrap`, `verbose`
cluster_var  What variable to cluster standard errors. This can be IDs or a higher aggregate level (state for example)
weights  Optional. Variable name for regression weights.
bootstrap  Optional. Should standard errors be calculated using bootstrap? Default is FALSE.
n_bootstraps  Optional. How many bootstraps to run. Default is 250.
return_bootstrap  Optional. Logical. Will return each bootstrap second-stage estimate to allow for manual use, e.g. percentile standard errors and empirical confidence intervals.
verbose  Optional. Logical. Should information about the two-stage procedure be printed back to the user? Default is TRUE.

Value

fixest object with adjusted standard errors (either by formula or by bootstrap). All the methods from fixest package will work, including `fixest::esttable` and `fixest::coefplot`

Examples

Load example dataset which has two treatment groups and homogeneous treatment effects

```r
# Load Example Dataset
data("df_hom")

Static TWFE:
You can run a static TWFE fixed effect model for a simple treatment indicator

```r
code
static <- did2s(df_hom,
    yname = "dep_var", treatment = "treat", cluster_var = "state",
    first_stage = ~ 0 | unit + year,
    second_stage = ~ i(treat, ref=FALSE))
```

```r
#> Running Two-stage Difference-in-Differences
#> • first stage formula `~ 0 | unit + year`
#> • second stage formula `~ i(treat, ref = FALSE)`
#> • The indicator variable that denotes when treatment is on is `treat`
#> • Standard errors will be clustered by `state`
```

```r
fixest::esttable(static)
```

```r
#> static
#> Dependent Var.: dep_var
#> treat = TRUE 2.005*** (0.0202)
#> _______________ _________________
#> S.E. type Custom
#> Observations 46,500
#> R2 0.47520
#> Adj. R2 0.47520
#> ---
#> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
Event Study:
Or you can use relative-treatment indicators to estimate an event study estimate

```r
es <- did2s(df_hom,
  yname = "dep_var", treatment = "treat", cluster_var = "state",
  first_stage = ~ 0 | unit + year,
  second_stage = ~ i(rel_year, ref=c(-1, Inf)))
```

- Running Two-stage Difference-in-Differences
- first stage formula `~ 0 | unit + year`
- second stage formula `~ i(rel_year, ref = c(-1, Inf))`
- The indicator variable that denotes when treatment is on is `treat`
- Standard errors will be clustered by `state`

```r
fixest::esttable(es)
```

<table>
<thead>
<tr>
<th>rel_year</th>
<th>Coef.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-20</td>
<td>0.0043</td>
<td>0.0322</td>
</tr>
<tr>
<td>-19</td>
<td>0.0222</td>
<td>0.0296</td>
</tr>
<tr>
<td>-18</td>
<td>-0.0358</td>
<td>0.0308</td>
</tr>
<tr>
<td>-17</td>
<td>0.0043</td>
<td>0.0337</td>
</tr>
<tr>
<td>-16</td>
<td>-0.0186</td>
<td>0.0353</td>
</tr>
<tr>
<td>-15</td>
<td>-0.0045</td>
<td>0.0346</td>
</tr>
<tr>
<td>-14</td>
<td>-0.0393</td>
<td>0.0384</td>
</tr>
<tr>
<td>-13</td>
<td>0.0453</td>
<td>0.0323</td>
</tr>
<tr>
<td>-12</td>
<td>0.0324</td>
<td>0.0309</td>
</tr>
<tr>
<td>-11</td>
<td>-0.0245</td>
<td>0.0349</td>
</tr>
<tr>
<td>-10</td>
<td>-0.0017</td>
<td>0.0241</td>
</tr>
<tr>
<td>-9</td>
<td>0.0155</td>
<td>0.0242</td>
</tr>
<tr>
<td>-8</td>
<td>-0.0073</td>
<td>0.0210</td>
</tr>
<tr>
<td>-7</td>
<td>-0.0513*</td>
<td>0.0202</td>
</tr>
<tr>
<td>-6</td>
<td>0.0269</td>
<td>0.0237</td>
</tr>
<tr>
<td>-5</td>
<td>0.0136</td>
<td>0.0237</td>
</tr>
<tr>
<td>-4</td>
<td>0.0381</td>
<td>0.0223</td>
</tr>
<tr>
<td>-3</td>
<td>-0.0228</td>
<td>0.0284</td>
</tr>
<tr>
<td>-2</td>
<td>0.0041</td>
<td>0.0228</td>
</tr>
<tr>
<td>0</td>
<td>1.971***</td>
<td>0.0470</td>
</tr>
<tr>
<td>1</td>
<td>2.050***</td>
<td>0.0466</td>
</tr>
<tr>
<td>2</td>
<td>2.033***</td>
<td>0.0441</td>
</tr>
<tr>
<td>3</td>
<td>1.966***</td>
<td>0.0400</td>
</tr>
<tr>
<td>4</td>
<td>1.965***</td>
<td>0.0430</td>
</tr>
<tr>
<td>5</td>
<td>2.030***</td>
<td>0.0456</td>
</tr>
<tr>
<td>6</td>
<td>2.040***</td>
<td>0.0447</td>
</tr>
<tr>
<td>7</td>
<td>1.995***</td>
<td>0.0370</td>
</tr>
<tr>
<td>8</td>
<td>2.019***</td>
<td>0.0485</td>
</tr>
<tr>
<td>9</td>
<td>1.955***</td>
<td>0.0468</td>
</tr>
<tr>
<td>10</td>
<td>1.950***</td>
<td>0.0455</td>
</tr>
<tr>
<td>11</td>
<td>2.117***</td>
<td>0.0664</td>
</tr>
<tr>
<td>12</td>
<td>2.132***</td>
<td>0.0741</td>
</tr>
</tbody>
</table>
#> rel_year = 13 2.019*** (0.0640)
#> rel_year = 14 2.013*** (0.0522)
#> rel_year = 15 1.961*** (0.0605)
#> rel_year = 16 1.916*** (0.0584)
#> rel_year = 17 1.938*** (0.0607)
#> rel_year = 18 2.070*** (0.0666)
#> rel_year = 19 2.066*** (0.0609)
#> rel_year = 20 1.964*** (0.0612)

#> S.E. type Custom
#> Observations 46,500
#> R2 0.47577
#> Adj. R2 0.47533

#> Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# plot rel_year coefficients and standard errors
fixest::coefplot(es, keep = "rel_year::(.*)")

**Example from Cheng and Hoekstra (2013):**
Here’s an example using data from Cheng and Hoekstra (2013)

```r
# Castle Data
castle <- haven::read_dta("https://github.com/scunning1975/mixtape/raw/master/castle.dta")
did2s(
data = castle,
yname = "l_homicide",
first_stage = ~ 0 | sid + year,
second_stage = ~ i(post, ref=0),
treatment = "post",
cluster_var = "state", weights = "popwt"
)
```

#> Running Two-stage Difference-in-Differences
#> • first stage formula `~ 0 | sid + year`
#> • second stage formula `~ i(post, ref = 0)`
#> • The indicator variable that denotes when treatment is on is `post`
#> • Standard errors will be clustered by `state`
#> OLS estimation, Dep. Var.: l_homicide
#> Observations: 550
#> Standard-errors: Custom
#> Estimate Std. Error t value Pr(>|t|)
#> post::1 0.075142 0.03538 2.12387 0.034127 *
#> ---
#> Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#> RMSE: 263.4 Adj. R2: 0.052465
event_study

Estimate event-study coefficients using TWFE and 5 proposed improvements.

Description

Uses the estimation procedures recommended from Borusyak, Jaravel, Spiess (2021); Callaway and Sant’Anna (2020); Gardner (2021); Roth and Sant’Anna (2021); Sun and Abraham (2020)

Usage

event_study(
  data,
  yname,
  idname,
  gname,
  tname,
  xformula = NULL,
  weights = NULL,
  estimator = c("all", "TWFE", "did2s", "did", "impute", "sunab", "staggered")
)

plot_event_study(out, separate = TRUE, horizon = NULL)

Arguments

data The dataframe containing all the variables
yname Variable name for outcome variable
idname Variable name for unique unit id
gname Variable name for unit-specific date of initial treatment (never-treated should be zero or NA)
tname Variable name for calendar period
xformula A formula for the covariates to include in the model. It should be of the form ~ X1 + X2. Default is NULL.
weights Variable name for estimation weights. This is used in estimating Y(0) and also augments treatment effect weights
estimator Estimator you would like to use. Use "all" to estimate all. Otherwise see table to know advantages and requirements for each of these.
out Output from event_study()
separate Logical. Should the estimators be on separate plots? Default is TRUE.
horizon Numeric. Vector of length 2. First element is min and second element is max of event_time to plot
Value

event_study returns a data.frame of point estimates for each estimator
plot_event_study returns a ggplot object that can be fully customized

Examples

```r
code
out = event_study(
data = did2s::df_het, yname = "dep_var", idname = "unit",
tname = "year", gname = "g", estimator = "all"
)
plot_event_study(out)
```

Description

Generate TWFE data

Usage

```r
gen_data(
g1 = 2000,
g2 = 2010,
g3 = 0,
panel = c(1990, 2020),
te1 = 2,
te2 = 2,
te3 = 2,
te_m1 = 0,
te_m2 = 0,
te_m3 = 0,
n = 1500
)
```

Arguments

- `g1`: treatment date for group 1. For no treatment, set g = 0.
- `g2`: treatment date for group 2. For no treatment, set g = 0.
- `g3`: treatment date for group 3. For no treatment, set g = 0.
- `panel`: numeric vector of size 2, start and end years for panel
- `te1`: treatment effect for group 1. Will ignore for that group if g = 0.
- `te2`: treatment effect for group 1. Will ignore for that group if g = 0.
te3 treatment effect for group 1. Will ignore for that group if g = 0. 

te_m1 treatment effect slope per year

te_m2 treatment effect slope per year

n number of individuals in sample

Value

Dataframe of generated data

Examples

# Homogeneous treatment effect
df_hom <- gen_data(panel = c(1990, 2020),
g1 = 2000, g2 = 2010, g3 = 0,
tel = 2, te2 = 2, te3 = 0,
te_m1 = 0, te_m2 = 0, te_m3 = 0)

# Heterogeneous treatment effect
df_het <- gen_data(panel = c(1990, 2020),
g1 = 2000, g2 = 2010, g3 = 0,
tel = 2, te2 = 1, te3 = 0,
te_m1 = 0.05, te_m2 = 0.15, te_m3 = 0)
Index

* datasets
  castle, 2
  df_het, 2
  df_hom, 3

castle, 2

df_het, 2
df_hom, 3
did2s, 4

event_study, 8
event_study(), 8

fixest::coefplot, 5
fixest::esttable, 5
fixest::feols, 4
fixest::i, 4

gen_data, 9

plot_event_study(event_study), 8