Package 'didimputation'

August 2, 2021

Type Package

Title Imputation Estimator from Borusyak, Jaravel, and Spiess (2021)

Version 0.1.0

Description Estimates Two-way Fixed Effects difference-in-differences/event-study models using the imputation-based approach proposed by Borusyak, Jaravel, and Spiess (2021).

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

LinkingTo Rcpp, RcppArmadillo

Depends R (>= 2.10), fixest (>= 0.9.0)

Imports Matrix, magrittr, Rcpp, broom, dplyr, glue, methods, rlang, stringr

Suggests haven, testthat (>= 3.0.0)

Config/testthat/edition 3

License MIT + file LICENSE

NeedsCompilation yes

Author Kyle Butts [aut, cre] (<https://orcid.org/0000-0002-9048-8059>)

Maintainer Kyle Butts <kyle.butts@colorado.edu>

Repository CRAN

Date/Publication 2021-08-02 08:10:07 UTC

R topics documented:

df_het — *Simulated data with two treatment groups and heterogenous effects 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)`

Description

Simulated data with two treatment groups and heterogenous effects

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
```

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 treatmet effect = te_m

state State that unit is in

group String name for group

df_hom	*Simulated data with two treatment groups and homogenous effects 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)`

Description

Simulated data with two treatment groups and homogenous effects

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.

unit_fe Unit FE

year_fe Year FE

error Random error component

te Static treatment effect = te

te_dynamic Dynamic treatmet effect = te_m

group String name for group

state State that unit is in

weight Weight from runif()

did_imputation *Borusyak, Jaravel, and Spiess (2021) Estimator*

Description

Treatment effect estimation and pre-trend testing in staggered adoption diff-in-diff designs with an imputation approach of Borusyak, Jaravel, and Spiess (2021)

Usage

```
did_imputation(
  data,
  yname,
  gname,
  tname,
  idname,
  first_stage = NULL,
  weights = NULL,
  wtr = NULL,
  horizon = NULL,
  pretrends = NULL
)
```

Arguments

data	A data.frame
yname	String. Variable name for outcome.
gname	String. Variable name for unit-specific date of treatment (never-treated should be zero or NA)
tname	String. Variable name for calendar period
idname	String. Variable name for unique unit id
first_stage	Formula for Y(0). Formula following fixest::feols. Fixed effects specified after "|". If not specified, then just unit and time fixed effects will be used.
weights	String. Variable name for estimation weights of observations. This is used in estimating Y(0) and also augments treatment effect weights.
wtr	Character vector of treatment weight names (see horizon for standard static and event-study weights)
horizon	Integer vector of event_time or TRUE. This only applies if wtr is left as NULL. if specified, weighted averages/sums of treatment effects will be reported for each of these horizons separately (i.e. tau0 for the treatment period, tau1 for one period after treatment, etc.). If TRUE, all horizons are used. If wtr and horizon are null, then the static treatment effect is calculated.
pretrends	Integer vector or TRUE. Which pretrends to estimate. If TRUE, all pretrends are used.

Details

The imputation-based estimator is a method of calculating treatment effects in a difference-in-differences framework. The method estimates a model for Y(0) using untreated/not-yet-treated observations and predicts Y(0) for the treated observations hat(Y_it(0)). The difference between treated and predicted untreated outcomes Y_it(1) - hat(Y_it(0)) serves as an estimate for the treatment effect for unit i in period t. These are then averaged to form average treatment effects for groups of it.

Value

A `data.frame` containing treatment effect term, estimate, standard error and confidence interval. This is in `tidy` format.

Examples

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

```
# Load Example Dataset
data("df_hom", package="did2s")
```

Static TWFE:

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

```
did_imputation(data = df_hom, yname = "dep_var", gname = "g",
               tname = "year", idname = "unit")
#> # A tibble: 1 × 5
#>   term  estimate std.error conf.low conf.high
#>   <chr>    <dbl>     <dbl>    <dbl>     <dbl>
#> 1 treat     2.02    0.0324     1.96      2.09
```

Event Study:

Or you can use relative-treatment indicators to estimate an event study estimate

```
did_imputation(data = df_hom, yname = "dep_var", gname = "g",
               tname = "year", idname = "unit", horizon=TRUE)
#> # A tibble: 21 × 5
#>    term  estimate std.error conf.low conf.high
#>    <chr>    <dbl>     <dbl>    <dbl>     <dbl>
#>  1 0         2.12    0.0737     1.97      2.26
#>  2 1         1.86    0.0841     1.69      2.02
#>  3 2         1.99    0.0810     1.83      2.15
#>  4 3         2.00    0.0855     1.84      2.17
#>  5 4         1.95    0.0856     1.78      2.12
#>  6 5         2.04    0.0835     1.87      2.20
#>  7 6         2.03    0.0807     1.87      2.19
#>  8 7         2.03    0.0865     1.86      2.19
#>  9 8         1.98    0.0826     1.81      2.14
#> 10 9         2.12    0.0842     1.96      2.29
#> # . . .  with 11 more rows
```

Example from Cheng and Hoekstra (2013):

Here's an example using data from Cheng and Hoekstra (2013)

```
# Castle Data
castle <- haven::read_dta("https://github.com/scunning1975/mixtape/raw/master/castle.dta")

did_imputation(data = castle, yname = "l_homicide", gname = "effyear",
               first_stage = ~ 0 | sid + year,
               tname = "year", idname = "sid")
#> # A tibble: 1 × 5
#>   term  estimate std.error conf.low conf.high
#>   <chr>    <dbl>     <dbl>    <dbl>     <dbl>
#> 1 treat   0.0798    0.0531  -0.0243     0.184
```

Index