Package: fastdid (via r-universe)

October 29, 2024

Type Package

Title Fast Staggered Difference-in-Difference Estimators

Version 1.0.2

Date 2024-10-25

Maintainer Lin-Tung Tsai <tsaidondon@gmail.com>

Description A fast and flexible implementation of Callaway and Sant'Anna's (2021)<doi:10.1016/j.jeconom.2020.12.001> staggered Difference-in-Differences (DiD) estimators, 'fastdid' reduces the computation time from hours to seconds, and incorporates extensions such as time-varying covariates and multiple events.

License MIT + file LICENSE

Imports data.table (>= 1.15.0), stringr, BMisc, collapse, dreamerr (>= 1.4.0), parglm, ggplot2

Suggests did, knitr, parallel, rmarkdown, tinytest

Encoding UTF-8

RoxygenNote 7.3.2

URL https://github.com/TsaiLintung/fastdid, https://tsailintung.github.io/fastdid/

BugReports https://github.com/TsaiLintung/fastdid/issues

VignetteBuilder knitr

NeedsCompilation no

Author Lin-Tung Tsai [aut, cre, cph], Maxwell Kellogg [ctb], Kuan-Ju Tseng [ctb]

Repository CRAN

Date/Publication 2024-10-28 12:50:04 UTC

Contents

`fastdid` *Fast Staggered DID Estimation*

Description

Performs Difference-in-Differences (DID) estimation.

Usage

```
fastdid(
  data,
  timevar,
  cohortvar,
  unitvar,
  outcomevar,
  control_option = "both",
  result_type = "group_time",
  balanced_event_time = NA,
  control_type = "ipw",
  allow_unbalance_panel = FALSE,
  boot = FALSE,
  biters = 1000,
  cband = FALSE,
  alpha = 0.05,
  weightvar = NA,
  clustervar = NA,
  covariatesvar = NA,
  varycovariatesvar = NA,
  copy = TRUE,
  validate = TRUE,
  anticipation = 0,
  base_period = "universal",
  exper = NULL,
  full = FALSE,
  parallel = FALSE,
  cohortvar2 = NA,
  event_specific = TRUE,
  double_control_option = "both"
)
```

Arguments

`data`	data.table, the dataset.
`timevar`	character, name of the time variable.
`cohortvar`	character, name of the cohort (group) variable.

`unitvar` character, name of the unit (id) variable.

`outcomevar` character vector, name(s) of the outcome variable(s).

`control_option` character, control units used for the DiD estimates, options are "both", "never", or "notyet".

`result_type` character, type of result to return, options are "group_time", "time", "group", "simple", "dynamic" (time since event), "group_group_time", or "dynamic_stagger".

`balanced_event_time` number, max event time to balance the cohort composition.

`control_type` character, estimator for controlling for covariates, options are "ipw" (inverse probability weighting), "reg" (outcome regression), or "dr" (doubly-robust).

`allow_unbalance_panel` logical, allow unbalance panel as input or coerce dataset into one.

`boot` logical, whether to use bootstrap standard error.

`biters` number, bootstrap iterations. Default is 1000.

`cband` logical, whether to use uniform confidence band or point-wise.

`alpha` number, the significance level. Default is 0.05.

`weightvar` character, name of the weight variable.

`clustervar` character, name of the cluster variable.

`covariatesvar` character vector, names of time-invariant covariate variables.

`varycovariatesvar` character vector, names of time-varying covariate variables.

`copy` logical, whether to copy the dataset.

`validate` logical, whether to validate the dataset.

`anticipation` number, periods with anticipation.

`base_period` character, type of base period in pre-preiods, options are "universal", or "varying".

`exper` list, arguments for experimental features.

`full` logical, whether to return the full result (influence function, call, weighting scheme, etc,.).

`parallel` logical, whether to use parallization on unix system.

`cohortvar2` character, name of the second cohort (group) variable.

`event_specific` logical, whether to recover target treatment effect or use combined effect.

`double_control_option` character, control units used for the double DiD, options are "both", "never", or "notyet".

Details

‘balanced_event_time‘ is only meaningful when ‘result_type == "dynamic‘.

‘result_type‘ as ‘group-group-time‘ and ‘dynamic staggered‘ is only meaningful when using double did.

‘biter‘ and ‘clustervar‘ is only used when ‘boot == TRUE‘.

Value

A data.table containing the estimated treatment effects and standard errors or a list of all results when ‘full == TRUE‘.

Examples

```
# simulated data
simdt <- sim_did(1e+02, 10, cov = "cont", second_cov = TRUE, second_outcome = TRUE, seed = 1)
dt <- simdt$dt

#basic call
result <- fastdid(data = dt, timevar = "time", cohortvar = "G",
                  unitvar = "unit", outcomevar = "y",
                  result_type = "group_time")
```

`plot_did_dynamics` *Plot event study*

Description

Plot event study results.

Usage

```
plot_did_dynamics(x, margin = "event_time")
```

Arguments

`x`	A data table generated with [fastdid] with one-dimensional index.
`margin`	character, the x-axis of the plot

Value

A ggplot2 object

Examples

```
# simulated data
simdt <- sim_did(1e+02, 10, seed = 1)
dt <- simdt$dt

#estimation
result <- fastdid(data = dt, timevar = "time", cohortvar = "G",
                  unitvar = "unit", outcomevar = "y",
                  result_type = "dynamic")

#plot
plot_did_dynamics(result)
```

sim_did *Simulate a Difference-in-Differences (DiD) dataset*

Description

Simulates a dataset for a Difference-in-Differences analysis with various customizable options.

Usage

```
sim_did(
  sample_size,
  time_period,
  untreated_prop = 0.3,
  epsilon_size = 0.001,
  cov = "no",
  hetero = "all",
  second_outcome = FALSE,
  second_cov = FALSE,
  vary_cov = FALSE,
  na = "none",
  balanced = TRUE,
  seed = NA,
  stratify = FALSE,
  treatment_assign = "latent",
  second_cohort = FALSE,
  confound_ratio = 1,
  second_het = "all"
)
```

Arguments

sample_size	The number of units in the dataset.
time_period	The number of time periods in the dataset.
untreated_prop	The proportion of untreated units.
epsilon_size	The standard deviation for the error term in potential outcomes.
cov	The type of covariate to include ("no", "int", or "cont").
hetero	The type of heterogeneity in treatment effects ("all" or "dynamic").
second_outcome	Whether to include a second outcome variable.
second_cov	Whether to include a second covariate.
vary_cov	include time-varying covariates
na	Whether to generate missing data ("none", "y", "x", or "both").
balanced	Whether to balance the dataset by random sampling.
seed	Seed for random number generation.

`stratify`	Whether to stratify the dataset based on a binary covariate.
`treatment_assign`	The method for treatment assignment ("latent" or "uniform").
`second_cohort`	include confounding events
`confound_ratio`	extent of event confoundedness
`second_het`	heterogeneity of the second event

Value

A list containing the simulated dataset (dt) and the treatment effect values (att).

Examples

```
# Simulate a DiD dataset with default settings
data <- sim_did(sample_size = 100, time_period = 5)
```

Index