Package: unitdid (via r-universe)

September 9, 2024

Type Package

Title Unit-level Difference-in-Difference Estimator

Version 0.0.6.1

Description The package provides a function to estimate the unit-level difference-in-difference estimator proposed by Arkhangelsky, Yanagimoto, and Zohar (2024).

License MIT + file LICENSE

Depends R (>= 4.1.0)

Imports dplyr (>= 1.1.0), fixest, magrittr, purrr, rlang, stats

Suggests ggplot2, haven, here, knitr, rmarkdown, testthat (>= 3.0.0), tools, altdoc

Config/testthat/edition 3

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

URL https://kazuyanagimoto.com/unitdid/, https://github.com/kazuyanagimoto/unitdid/

VignetteBuilder knitr

Repository https://kazuyanagimoto.r-universe.dev

RemoteUrl https://github.com/kazuyanagimoto/unitdid

RemoteRef HEAD

RemoteSha ff7e24f9f6baacaff30263e0b40bf4f241a5fa66

Contents

`aggregate_unitdid` *Aggregate the mean and variance of the estimated unit-level DiD effects*

Description

Aggregate the mean and variance of the estimated unit-level DiD effects

Usage

```
aggregate_unitdid(
  object,
  agg = "full",
  na.rm = TRUE,
  by = NULL,
  normalized = NULL,
  allow_negative_var = FALSE,
  only_full_horizon = TRUE
)
```

Arguments

`object`	`unitdid` object
`agg`	Aggregation method. One of `c("full", "event", "event_age")` and the default is `full`. If `by` is provided in the model, all the options will separately aggregate by its group. The `event` option aggregates by the group of the event timing. The `event_age` option aggregates by the group of the age at the event time. `event_age` requires the `bname` to be provided in the model.
`na.rm`	Logical. If `TRUE`, remove `NA` values for the aggregation. The default is `TRUE`.
`by`	A character vector of variables to aggregate separately by. Default is inherited from the `unitdid` object but you can override it here. You can estimate the unit-level DiD effects separately by `by` in `unitdid` but you can also aggregate the estimates by (higher-level) `by` here. You can use "rel_time" as the highest level of aggregation.
`normalized`	Logical. If `TRUE`, the function will normalize the aggregated mean and variance by the mean of the imputed outcome variable. Default is inherited from the `unitdid` object.
`allow_negative_var`	Logical. If `FALSE`, the function will return the estimated variance trimmed at zero. Default is `FALSE`.

only_full_horizon

Logical. If TRUE, when you aggregate the unit-level treatment effect, only the event year (ename) with full horizon (k_min:k_max) will be included. This is recommended in the case that you do not want to change the composition of the event year (or age for the child penalties) for each estimated point in k_min:k_max. Default is TRUE.

Value

A tibble with the aggregated mean and variance of the estimated unit-level DiD effects

---

base_heterocp *Simulated Individual Child Panalty Data*

Description

Simulated Individual Child Panalty Data

Usage

```
base_heterocp
```

Format

base_heterocp**:**

A dataframe with 1000 individuals for each birth year from 1965 to 1984:

id Individual identifier

year Year of observation

byear Birth year

cage Age at first birth

rel_time Relative time to first birth

y Outcome variable

Source

Generated by gen_heterocp() with seed 1234

gen_heterocp *Generate Sample Heterogenous Child Penalty Data*

Description

Generate Sample Heterogenous Child Penalty Data

Usage

```
gen_heterocp(size_cohort = 300)
```

Arguments

size_cohort	n_obsumber of individuals per birth year

Value

A sample dataframe with heterogenous child penalty over the age at first birth

Examples

```
set.seed(1234)
base_heterocp <- gen_heterocp()
```

get_unitdid *Get unit-level Difference-in-Differences estimates*

Description

Get unit-level Difference-in-Differences estimates

Usage

```
get_unitdid(
  object,
  normalized = NULL,
  export = TRUE,
  only_full_horizon = FALSE
)
```

Arguments

object	unitdid object
normalized	Logical. If TRUE, the function will normalize them by the mean of the imputed outcome variable. Default is inherited from the unitdid object.
export	Logical. If TRUE, the function will not export the columns with the zz000 prefix, which are used in the internal computation.
only_full_horizon	Logical. If TRUE, only the event year (ename) with full horizon (k_min:k_max) will be exported. This is recommended in the case that you do not want to change the composition of the event year (or age for the child penalties) for each estimated point in k_min:k_max for aggregation. Default is FALSE.

Value

A dataframe with a new column of the unit-level DiD estimates

---

summary.unitdid *Aggregate the mean and variance of the estimated unit-level DiD effects*

Description

Aggregate the mean and variance of the estimated unit-level DiD effects

Usage

```
## S3 method for class 'unitdid'
summary(object, ...)
```

Arguments

object	unitdid object
...	aggregate_unitdid arguments

Value

A tibble with the summary statistics

Examples

```
library(unitdid)
mdl_base <- base_heterocp |>
  unitdid(yname = "y",
          iname = "id",
          tname = "year",
          ename = "cyear",
          bname = "byear")
summary(mdl_base, agg = "event_age")
```

unitdid *A function estimates unit-level difference-in-differences*

Description

A function estimates unit-level difference-in-differences

Usage

```
unitdid(
  data,
  yname,
  iname,
  tname,
  ename,
  first_stage = NULL,
  wname = NULL,
  k_min = 0,
  k_max = 5,
  compute_varcov = "none",
  by = NULL,
  bname = NULL,
  normalized = FALSE,
  newnames = NULL
)
```

Arguments

data	The dataframe containing all the variables
yname	Outcome variable
iname	Unit identifier
tname	Time variable
ename	Event timing variable
first_stage	Formula for Y(0). Formula follows fixest::feols. If not specified, unit (iname) and time (tname) fixed effects will be used.
wname	Optional. The name of the weight variable.
k_min	Relative time to treatment at which treatment starts. Default is 0.
k_max	Relative time to treatment at which treatment ends. Default is 5.
compute_varcov	One of c("none", "var", "cov") and Default is "none". If "var", the function will estimate the unit-level variance of the outcome variable. If "cov", the function will estimate the unit-level covariance of the outcome variable for each pair within k_min:k_max.
by	A character vector of variables to estimate separately by. Default is NULL.

bname — Birth year variable. Default is NULL. Necessary to aggregate the estimates by age at event.

normalized — Logical. If TRUE, the function will normalize the outcome variable scale. Default is FALSE.

newnames — Optional. A list of new names for the output variables. `ytildename` is the name of the imputed outcome variable. Default is `paste0(yname, "_tilde")`. `yvarname` is the name of the unit-level variance of the outcome variable. Default is `paste0(yname, "_var")`. `yvarrawname` is the name of the raw unit-level variance of the outcome variable, which is the variance before subtracting the variance of the measurement error. Default is `paste0(yname, "_varraw")`. `yvarerrname` is the name of the unit-level variance of the measurement error. Default is `paste0(yname, "_varerr")`. `ycovname` is the name of the unit-level covariance of the outcome variable. Default is `paste0(yname, "_cov")`. `ycovrawname` is the name of the raw unit-level covariance of the outcome variable, which is the covariance before subtracting the covariance of the measurement error. Default is `paste0(yname, "_covraw")`. `ycoverrname` is the name of the unit-level covariance of the measurement error. Default is `paste0(yname, "_coverr")`. `kprimename` is the name of the relative time to treatment. This is used for the second column name of the relative time of the unit-level covariance estimation. Default is "kprime".

Value

A `unitdid` class object.

Index