# Package 'didimputation'

October 13, 2022

**Type** Package

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

**Version** 0.3.0

**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** 7.2.1

**LinkingTo** Rcpp, RcppArmadillo

**Depends** R (>= 2.10), fixest (>= 0.10.0), data.table (>= 1.10.0)

**Imports** Matrix, magrittr, Rcpp, broom, dplyr, glue, stringr, purrr, tidyr

**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** 2022-08-25 20:02:33 UTC

## R topics documented:

## 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
```

### 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*

### 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.

**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,
  wname = NULL,
  wtr = NULL,
  horizon = NULL,
  pretrends = NULL,
  cluster_var = NULL
)
```

## Arguments

| | |
|---|---|
| data | A data.frame |
| yname | String. Variable name for outcome. Use fixest c() syntax for multiple lhs. |
| 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. |
| wname | 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. |
| cluster_var | String. Varaible name for clustering groups. If not supplied, then idname is used as default. |

## 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 x 6
#>   lhs     term  estimate std.error conf.low conf.high
#>   <chr>   <chr>    <dbl>     <dbl>    <dbl>     <dbl>
#> 1 dep_var treat     2.00    0.0182     1.97      2.04
```

**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 x 6
#>    lhs     term  estimate std.error conf.low conf.high
#>    <chr>   <chr>    <dbl>     <dbl>    <dbl>     <dbl>
#>  1 dep_var 0         1.97    0.0425     1.89      2.05
#>  2 dep_var 1         2.05    0.0434     1.97      2.14
#>  3 dep_var 2         2.03    0.0432     1.95      2.12
#>  4 dep_var 3         1.97    0.0428     1.88      2.05
#>  5 dep_var 4         1.97    0.0420     1.88      2.05
#>  6 dep_var 5         2.03    0.0423     1.95      2.11
#>  7 dep_var 6         2.04    0.0450     1.95      2.13
#>  8 dep_var 7         2.00    0.0437     1.91      2.08
#>  9 dep_var 8         2.02    0.0440     1.93      2.10
#> 10 dep_var 9         1.96    0.0440     1.87      2.04
#> # ... 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 = "c(l_homicide, l_assault)", gname = "effyear",
               first_stage = ~ 0 | sid + year,
               tname = "year", idname = "sid")
#> # A tibble: 2 x 6
#>   lhs        term  estimate std.error conf.low conf.high
#>   <chr>      <chr>    <dbl>     <dbl>    <dbl>     <dbl>
#> 1 l_homicide treat   0.0798    0.0609  -0.0395     0.199
#> 2 l_assault  treat   0.0496    0.0513  -0.0510     0.150
```

# Index