

Package: didimputation (via r-universe)

May 17, 2026

Type Package

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

Version 0.5.0

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

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Depends R (>= 3.5.0), fixest (>= 0.13.2), data.table (>= 1.10.0)

Imports Matrix

Suggests haven, testthat (>= 3.0.0)

Config/testthat/edition 3

URL <https://github.com/kylebutts/didimputation>

License MIT + file LICENSE

Repository <https://kylebutts.r-universe.dev>

Date/Publication 2025-12-11 16:20:13 UTC

RemoteUrl <https://github.com/kylebutts/didimputation>

RemoteRef HEAD

RemoteSha 69b4f8dfe16b007474721fc5610859b56a80cdc6

Contents

df_het	2
df_hom	2
did_imputation	3

Index	7
--------------	----------

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

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

<code>data</code>	A <code>data.frame</code>
<code>yname</code>	String. Variable name for outcome. Use <code>fixest c()</code> syntax for multiple lhs, e.g. <code>"c(y1, y2)"</code> .
<code>gname</code>	String. Variable name for unit-specific date of treatment (never-treated should be zero or NA).
<code>tname</code>	String. Variable name for calendar period.
<code>idname</code>	String. Variable name for unique unit id.
<code>first_stage</code>	Formula for $Y(0)$. Formula following <code>fixest::feols</code> . Fixed effects specified after <code>" "</code> . If not specified, then just unit and time fixed effects will be used.
<code>wname</code>	String. Variable name for estimation weights of observations. This is used in estimating $Y(0)$ and also augments treatment effect weights.
<code>wtr</code>	Character vector of treatment weight names (see <code>horizon</code> for standard static and event-study weights)
<code>horizon</code>	Integer vector of <code>event_time</code> or <code>TRUE</code> . This only applies if <code>wtr</code> is left as <code>NULL</code> . If specified, weighted averages/sums of treatment effects will be reported for each of these horizons separately (i.e. <code>tau0</code> for the treatment period, <code>tau1</code> for one period after treatment, etc.). If <code>TRUE</code> , all horizons are used. If <code>wtr</code> and <code>horizon</code> are null, then the static treatment effect is calculated.
<code>pretrends</code>	Integer vector or <code>TRUE</code> . Which pretrends to estimate. If <code>TRUE</code> , all pretrends are used.
<code>cluster_var</code>	String. Variable name for clustering groups. If not supplied, then <code>idname</code> 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 (i, t) .

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="didimputation")
```

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")
#>      term estimate  std.error conf.low conf.high
#> <char>    <num>      <num>    <num>    <num>
#> 1:  treat  2.024639  0.03243596  1.961065  2.088214
```

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)
#>      term estimate  std.error conf.low conf.high
#> <char>    <num>      <num>    <num>    <num>
#> 1:      0  2.117232  0.07368419  1.972811  2.261653
#> 2:      1  1.856536  0.07672104  1.706163  2.006909
#> 3:      2  1.986357  0.07137180  1.846468  2.126246
#> 4:      3  2.004843  0.07653409  1.854836  2.154850
#> 5:      4  1.950228  0.07543636  1.802372  2.098083
#> 6:      5  2.038302  0.07580288  1.889728  2.186875
#> 7:      6  2.031571  0.07223098  1.889999  2.173144
#> 8:      7  2.025286  0.07541719  1.877468  2.173104
#> 9:      8  1.976081  0.07493409  1.829210  2.122951
#> 10:     9  2.121434  0.07268404  1.978974  2.263895
#> 11:    10  2.087984  0.08271442  1.925864  2.250105
#> 12:    11  1.942825  0.11421421  1.718965  2.166685
#> 13:    12  1.940532  0.11200348  1.721005  2.160059
#> 14:    13  1.964569  0.11361969  1.741875  2.187264
#> 15:    14  2.023456  0.11753255  1.793092  2.253820
#> 16:    15  2.235051  0.12110086  1.997693  2.472409
#> 17:    16  2.178438  0.11552325  1.952013  2.404864
#> 18:    17  1.935576  0.11278311  1.714521  2.156631
#> 19:    18  2.134953  0.10993120  1.919488  2.350418
#> 20:    19  2.111984  0.11146282  1.893517  2.330451
#> 21:    20  1.925168  0.11214206  1.705370  2.144967
#>      term estimate  std.error conf.low conf.high
```

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")
#> Key: <lhs>
#>      lhs  term  estimate  std.error  conf.low  conf.high
```

```
#>      <char> <char>      <num>      <num>      <num>      <num>
#> 1:  l_assault  treat 0.04955260 0.05132258 -0.05103966 0.1501449
#> 2:  l_homicide treat 0.07980155 0.06088398 -0.03953105 0.1991341
```

Index

* datasets

df_het, 2

df_hom, 2

df_het, 2

df_hom, 2

did_imputation, 3

fixest::feols, 4