package ‘did’

February 18, 2020

Title  Treatment Effects with Multiple Periods and Groups
Version 1.2.3
Description The standard Difference-in-Differences (DID) setup involves two periods and two groups -- a treated group and untreated group. Many applications of DID methods involve more than two periods and have individuals that are treated at different points in time. This package contains tools for computing average treatment effect parameters in Difference in Differences models with more than two periods and with variation in treatment timing using the methods developed in Callaway and Sant’Anna (2019) <https://ssrn.com/abstract=3148250>. The main parameters are group-time average treatment effects which are the average treatment effect for a particular group at a particular time. These can be aggregated into a fewer number of treatment effect parameters, and the package deals with the cases where there is selective treatment timing, dynamic treatment effects, calendar time effects, or combinations of these. There are also functions for testing the Difference in Differences assumption, and plotting group-time average treatment effects.

Depends R (>= 2.10)
License GPL-2
Encoding UTF-8
LazyData true
Imports BMisc (>= 1.3.1), MASS, pbapply, stats, ggplot2, knitr, utils, gridExtra
RoxygenNote 7.0.2
VignetteBuilder knitr
Suggests rmarkdown
NeedsCompilation no
Author Brantly Callaway [aut, cre], Pedro H.C. Sant’Anna [aut]
Maintainer Brantly Callaway <bmcallaw@olemiss.edu>
Repository CRAN
Date/Publication 2020-02-18 00:00:02 UTC
R topics documented:

AGGTE .................................................. 2
did ..................................................... 3
expf ....................................................... 4
ggdid ..................................................... 4
indicator ................................................... 5
MP .......................................................... 6
mp.spatt .................................................. 7
MP.TEST ................................................... 11
mpdta ...................................................... 12
onefun ...................................................... 13
summary.AGGTE .......................................... 13
summary.MP ............................................... 14
summary.MP.TEST ........................................ 14

Index .................................................... 15

| AGGTE | AGGTE |

Description

AGGTE class for aggregate treatment effects

Usage

AGGTE(
simple.att = NULL,
simple.se = NULL,
selective.att = NULL,
selective.se = NULL,
selective.att.g = NULL,
selective.se.g = NULL,
dynamic.att = NULL,
dynamic.se = NULL,
dynamic.att.e = NULL,
dynamic.se.e = NULL,
calendar.att = NULL,
calendar.se = NULL,
calendar.att.t = NULL,
calendar.se.t = NULL,
dynsel.att.e1 = NULL,
dynsel.se.e1 = NULL,
dynsel.att.ee1 = NULL,
dynsel.se.ee1 = NULL,
groups = NULL,
)
times = NULL
)

Arguments

simple.att       simple weighted average of group-time average treatment effects
simple.se        the standard error for simple.att
selective.att    aggregated group-time average treatment effects when there is selective treatment timing
selective.se     the standard error for selective.att
selective.att.g  aggregated group-time average treatment effects when there is selective treatment timing for each particular group
selective.se.g   the standard error for selective.att.g
dynamic.att      aggregated group-time average treatment effects when there are dynamic treatment effects
dynamic.se       the standard error for dynamic.att
dynamic.att.e    aggregated group-time average treatment effects for each length of exposure to treatment
dynamic.se.e     the standard error for dynamic.att.e
calendar.att     the aggregated group-time average treatment effects when there are calendar time effects
calendar.se      the standard error for calendar.att
calendar.att.t   the aggregated group-time average treatment effects when there are calendar time effects for each time period
calendar.se.t    the standard error for calendar.att.t
dynsel.att.e1    aggregated group-time average treatment effects when there are dynamic treatment effects and selective treatment timing. Here, e1 is the number of periods that a group is required to be treated in order to be included in the results.
dynsel.se.e1     the standard error for dynsel.att.e1
dynsel.att.ee1   aggregated group-time average treatment effects when there are dynamic treatment effects and selective treatment timing. Here, e1 is the number of periods that a group is required to be treated in order to be included in the results and for each length of exposure to treatment
dynsel.se.ee1    the standard error for dynsel.att.ee1
groups           vector of all groups
times            vector of all times

did

did: Difference in Differences with multiple periods and variation in treatment timing

Description

did: Difference in Differences with multiple periods and variation in treatment timing
expf

Description

exponential weighting function

Usage

expf(X, u)

Arguments

X matrix of X’s from the data
u a particular value to multiply times the X’s

Value

numeric vector

Examples

data(mpdta)
dta <- subset(mpdta, year==2007)
X <- model.matrix(~lpop, data=dta)
X <- expf(X, X[1,])

ggdid

Description

Function to plot MP objects

Usage

ggidid(
  mobj,
  type = c("attgt", "dynamic", "selective", "calendar", "dynsel"),
  ylim = NULL,
  xlab = NULL,
  ylab = NULL,
  title = "Group",
  xgap = 1,
  ncol = 1,
  e1 = 1
)

Arguments

- `mpobj` : an MP object
- `type` : the type of plot, should be one of "attgt", "dynamic", "selective", "calendar", "dynsel". "attgt" is the default and plots all group-time average treatment effects separately by group (including pre-treatment time periods); "dynamic" plots dynamic treatment effects – these are the same as event studies; "selective" plots average effects of the treatment separately by group (which allows for selective treatment timing); "calendar" plots average treatment effects by time period; and "dynsel" plots dynamic effects allowing for selective treatment timing (this also requires setting the additional parameter `e1`)
- `ylim` : optional y limits for the plot; setting here makes the y limits the same across different plots
- `xlab` : optional x-axis label
- `ylab` : optional y-axis label
- `title` : optional plot title
- `xgap` : optional gap between the labels on the x-axis. For example, `xgap=3` indicates that the labels should show up for every third value on the x-axis. The default is 1.
- `ncol` : The number of columns to include in the resulting plot. The default is 1.
- `e1` : only used when plot type is "dynsel", this specifies the number of post-treatment periods that need to be available for particular groups to be included in the resulting plot when there are dynamic treatment effects and selective treatment timing

Examples

```r
## Not run:
data(mpdata)
out <- mp.spatt(lemp ~ treat, xformla=~lpop, data=mpdata,
  panel=TRUE, first.treat.name="first.treat",
  idname="countyreal", tname="year",
  bstrap=FALSE, se=TRUE, cband=FALSE)
ggdid(out)
## End(Not run)
```

Description

`indicator` weighting function
Usage

indicator(X, u)

Arguments

X  matrix of X's from the data
u  a particular value to compare X's to

Value

numeric vector

Examples

data(mpdta)
dta <- subset(mpdta, year==2007)
X <- model.matrix(~lpop, data=dta)
X <- indicator(X, X[1,])

Description

multi-period object

Usage

MP(
    group,
    t,
    att,
    V,
    c,
    inffunc,
    n = NULL,
    W = NULL,
    Wpval = NULL,
    aggte = NULL
)
mp.spatt

Arguments

- **group**: which group (defined by period first treated) an group-time average treatment effect is for
- **t**: which time period a group-time average treatment effect is for
- **att**: the group-average treatment effect for group group and time period t
- **V**: the variance matrix for group-time average treatment effects
- **c**: critical value if one is obtaining uniform confidence bands
- **inffunc**: the influence function for estimating group-time average treatment effects
- **n**: the number of observations
- **W**: the Wald statistic for pre-testing the common trends assumption
- **Wpval**: the p-value of the Wald statistic for pre-testing the common trends assumption
- **aggte**: an aggregate treatment effects object

Value

MP object

mp.spatt

Description

mp.spatt computes the ATT in the case where there are more than two periods of data and allowing for treatment to occur at different points in time extending the method of Abadie (2005). This method relies on once individuals are treated they remain in the treated state for the duration.

Usage

mp.spatt(
  formula,
  xformula = NULL,
  data,
  tname,
  aggte = TRUE,
  w = NULL,
  panel = FALSE,
  idname = NULL,
  first.treat.name,
  alp = 0.05,
  method = "logit",
  se = TRUE,
  bstrap = FALSE,
  biters = 100,
  clustervars = NULL,
cband = FALSE, 
citers = 100, 
seedvec = NULL, 
pl = FALSE, 
cores = 2, 
printdetails = TRUE )

Arguments

formula The formula y ~ d where y is the outcome and d is the treatment indicator (d should be binary)
xformula A optional one sided formula for additional covariates that will be adjusted for. E.g ~ age + education. Additional covariates can also be passed by name using the x paramater.
data The name of the data.frame that contains the data
tname The name of the column containing the time periods
aggtte boolean for whether or not to compute aggregate treatment effect parameters, default TRUE
w A vector of weights for each observation (not implemented)
panel Boolean indicating whether the data is panel or repeated cross sections
idname The individual (cross-sectional unit) id name
first.treat.name The name of the variable in data that contains the first period when a particular observation is treated. This should be a positive number for all observations in treated groups. It should be 0 for observations in the untreated group.
alp the significance level, default is 0.05
method The method for estimating the propensity score when covariates are included
se Boolean whether or not to compute standard errors
bstrap Boolean for whether or not to compute standard errors using the multiplier bootstrap. If standard errors are clustered, then one must set `bstrap=TRUE`. 
biters The number of bootstrap iterations to use. The default is 100, and this is only applicable if `bstrap=TRUE`.
clustervars A vector of variables to cluster on. At most, there can be two variables (otherwise will throw an error) and one of these must be the same as `idname` which allows for clustering at the individual level.
cband Boolean for whether or not to compute a uniform confidence band that covers all of the group-time average treatment effects with fixed probability 1 - alp. The default is FALSE and the resulting standard errors will be pointwise.
citers Computing uniform confidence bands requires the bootstrap, if `cband = TRUE`, then this is the number of bootstrap iterations to compute the confidence band. The default is 100.
seedvec Optional value to set random seed; can possibly be used in conjunction with bootstrapping standard errors# (not implemented)
mp.spatt.test

pl
Boolean for whether or not to use parallel processing
cores
The number of cores to use for parallel processing
printdetails
Boolean for showing detailed results or not

Value

MP object

References


Examples

data(mpdta)

## with covariates
out1 <- mp.spatt(lamp ~ treat, xformla=~lpop, data=mpdta,
                  panel=TRUE, first.treat.name="first.treat",
                  idname="countyreal", tname="year",
                  bstrap=FALSE, se=TRUE, cband=FALSE)

## summarize the group-time average treatment effects
summary(out1)
## summarize the aggregated treatment effect parameters
summary(out1$aggte)

## without any covariates
out2 <- mp.spatt(lamp ~ treat, xformla=NULL, data=mpdta,
                  panel=TRUE, first.treat.name="first.treat",
                  idname="countyreal", tname="year",
                  bstrap=FALSE, se=TRUE, cband=FALSE)

summary(out2)

Description

integrated moments test for conditional common trends holding in all pre-treatment time periods across all groups
Usage

mp.spatt.test(
  formla = NULL,
  xformlalist = NULL,
  data = NULL,
  tname = NULL,
  weightfun = NULL,
  w = NULL,
  panel = FALSE,
  idname = NULL,
  first.treat.name = NULL,
  alp = 0.05,
  method = "logit",
  biters = 100,
  clustervarlist = NULL,
  pl = FALSE,
  cores = 2
)

Arguments

formula  The formula y ~ d where y is the outcome and d is the treatment indicator (d should be binary)

xformlalist  A list of formulas for the X variables. This allows to test using different specifications for X, if desired

data  The name of the data.frame that contains the data

tname  The name of the column containing the time periods

weightfun  A function that takes in two arguments, X and u, to compute the weighting function for the test. The default is 1*(X <= u)

w  A vector of weights for each observation (not implemented)

panel  Boolean indicating whether the data is panel or repeated cross sections

idname  The individual (cross-sectional unit) id name

first.treat.name  The name of the variable in data that contains the first period when a particular observation is treated. This should be a positive number for all observations in treated groups. It should be 0 for observations in the untreated group.

alp  the significance level, default is 0.05

method  The method for estimating the propensity score when covariates are included

biters  The number of bootstrap iterations to use. The default is 100, and this is only applicable if bootstrap=TRUE.

clustervarlist  A list of cluster variables. This allows to conduct the test using different levels of clustering, if desired.

pl  Boolean for whether or not to use parallel processing

cores  The number of cores to use for parallel processing
**Value**

list containing test results

**References**


**Examples**

```r
## Not run:
data(mpdta)
mptest <- mp.spatt.test(lemp ~ treat, xformlalist=list(~lpop), data=mpdta, 
   panel=TRUE, first.treat.name="first.treat", 
   idname="countyreal", tname="year", clustervarlist=list(NULL))
summary(mptest[[1]])

## End(Not run)
data(mpdta)
mptest <- mp.spatt.test(lemp ~ treat, xformlalist=list(NULL), data=mpdta, 
   panel=TRUE, first.treat.name="first.treat", 
   idname="countyreal", tname="year", clustervarlist=list(NULL))
summary(mptest[[1]])
```

**Description**

MP.TEST objects

**Usage**

```r
MP.TEST(
   CvM, 
   CvMb, 
   CvMcval, 
   CvMpval, 
   KS, 
   KSb, 
   KScval, 
   KSpval, 
   clustervars, 
   xformula 
)
```
Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CvM</td>
<td>Cramer von Mises test statistic</td>
</tr>
<tr>
<td>CvMb</td>
<td>a vector of boostrapped Cramer von Mises test statistics</td>
</tr>
<tr>
<td>CvMcval</td>
<td>CvM critical value</td>
</tr>
<tr>
<td>CvMpval</td>
<td>p-value for CvM test</td>
</tr>
<tr>
<td>KS</td>
<td>Kolmogorov-Smirnov test statistic</td>
</tr>
<tr>
<td>KSb</td>
<td>a vector of boostrapped KS test statistics</td>
</tr>
<tr>
<td>KScval</td>
<td>KS critical value</td>
</tr>
<tr>
<td>KSfval</td>
<td>p-value for KS test</td>
</tr>
<tr>
<td>clustervars</td>
<td>vector of which variables were clustered on for the test</td>
</tr>
<tr>
<td>xformula</td>
<td>formula for the X variables used in the test</td>
</tr>
</tbody>
</table>

mpdta

County teen employment

Description

A dataset containing (the log of) teen employment in 500 counties in the U.S. from 2004 to 2007. This is a subset of the dataset used in Callaway and Sant’Anna (2018). See that paper for additional descriptions.

Usage

mpdta

Format

A data frame with 2000 rows and 5 variables:

- **year** the year of the observation
- **countyreal** a unique identifier for a particular county
- **lpop** the log of 1000s of population for the county
- **lemp** the log of teen employment in the county
- **first.treat** the year that the state where the county is located raised its minimum wage, it is set equal to 0 for counties that have minimum wages equal to the federal minimum wage over the entire period.
- **treat** whether or not a particular county is treated in that year

Source

Callaway and Sant’Anna (2018)
onefun

**onefun**

**Description**

just return the value 1

**Usage**

```r
onefun(X, u)
```

**Arguments**

- `X` matrix of X’s from the data
- `u` a particular value to compare X’s to

**Value**

numeric vector

**Examples**

```r
data(mpdta)
dta <- subset(mpdta, year==2007)
X <- model.matrix(~lpop, data=dta)
X <- onefun(X, X[1,])
```

---

summary.AGGTE

**summary.AGGTE**

**Description**

print a summary of an AGGTE object

**Usage**

```r
## S3 method for class 'AGGTE'
summary(
  object,
  type = c("dynamic", "selective", "calendar", "dynsel"),
  e1 = 1,
  ...)
```
Arguments

object  an AGGTE object

type  which type of summary to print, options are "dynamic", "selective", "calendar", and "dynsel"

e1  if the type is "dynsel", this is the number of post-treatment periods required in order for a group to be used to construct aggregated parameters with selective treatment timing and dynamic effects; otherwise not used

...  other variables

Description

prints a summary of a MP object

Usage

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

Arguments

object  an MP object

...  extra arguments

Description

print a summary of test results

Usage

## S3 method for class 'MP.TEST'
summary(object, ...)

Arguments

object  an MP.TEST object

...  other variables
Index

*Topic **datasets**
  mpdta, 12

AGGTE, 2

did, 3

expf, 4

ggdid, 4

indicator, 5

MP, 6

mp.spatt, 7

mp.spatt.test, 9

MP.TEST, 11

mpdta, 12

onefun, 13

summary.AGGTE, 13

summary.MP, 14

summary.MP.TEST, 14