Package ‘DRDID’

November 13, 2021

Type Package

Title Doubly Robust Difference-in-Differences Estimators

Version 1.0.3

Description Implements the locally efficient doubly robust difference-in-differences (DiD) estimators for the average treatment effect proposed by Sant'Anna and Zhao (2020) <doi:10.1016/j.jeconom.2020.06.003>. The estimator combines inverse probability weighting and outcome regression estimators (also implemented in the package) to form estimators with more attractive statistical properties. Two different estimation methods can be used to estimate the nuisance functions.


License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 4.0)

Imports stats, trust, BMisc (>= 1.4.1)

RoxygenNote 7.1.2

Suggests knitr, rmarkdown, spelling, testthat

Date 2021-11-012

Language en-US

BugReports https://github.com/pedrohcgs/DRDID/issues

NeedsCompilation no

Author Pedro H. C. Sant'Anna [aut, cre], Jun Zhao [aut]

Maintainer Pedro H. C. Sant'Anna <pedro.h.santanna@vanderbilt.edu>

Repository CRAN

Date/Publication 2021-11-13 07:40:02 UTC
The `DRDID` package implements different estimators for the average treatment effect on the treated in difference-in-differences (DID) setups where the parallel trends assumption holds after you condition on a vector of pre-treatment covariates. The main estimators implemented here are the locally efficient, doubly-robust DID estimators proposed by Sant'Anna and Zhao (2020) <https://arxiv.org/abs/1812.01723>. A number of other DID estimators discussed in Sant’Anna and Zhao (2020) are also implemented.

**References**

drdid is used to compute the locally efficient doubly robust estimators for the ATT in difference-in-differences (DiD) setups. It can be used with panel or stationary repeated cross section data. Data should be stored in "long" format.

Usage

```r
drdid(
  yname,
  tname,
  idname,
  dname,
  xformla = NULL,
  data,
  panel = TRUE,
  estMethod = c("imp", "trad"),
  weightsname = NULL,
  boot = FALSE,
  boot.type = c("weighted", "multiplier"),
  nboot = 999,
  inffunc = FALSE
)
```

Arguments

- `yname`: The name of the outcome variable.
- `tname`: The name of the column containing the time periods.
- `idname`: The name of the column containing the unit id name.
- `dname`: The name of the column containing the treatment group (=1 if observation is treated in the post-treatment, =0 otherwise).
- `xformla`: A formula for the covariates to include in the model. It should be of the form \( \sim X_1 + X_2 \) (intercept should not be listed as it is always automatically included). Default is NULL which is equivalent to `xformla=~1`.
- `data`: The name of the data.frame that contains the data.
- `panel`: Whether or not the data is a panel dataset. The panel dataset should be provided in long format – that is, where each row corresponds to a unit observed at a particular point in time. The default is TRUE. When `panel = TRUE`, the variable `idname` must be set. When `panel = FALSE`, the data is treated as stationary repeated cross sections.
estMethod the method to estimate the nuisance parameters. The default is "imp" which uses weighted least squares to estimate the outcome regressions and inverse probability tilting to the estimate the propensity score, leading to the improved locally efficient DR DID estimator proposed by Sant’Anna and Zhao (2020). The other alternative is "trad", which then uses OLS to estimate outcome regressions and maximum likelihood to estimate propensity score. This leads to the "traditional" locally efficient DR DID estimator proposed by Sant’Anna and Zhao (2020).

weightsname The name of the column containing the sampling weights. If NULL, then every observation has the same weights.

boot Logical argument to whether bootstrap should be used for inference. Default is FALSE and analytical standard errors are reported.

boot.type Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".

nboot Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.

inffunc Logical argument to whether influence function should be returned. Default is FALSE.

Details
When panel data are available (panel = TRUE), the drdid function implements the locally efficient doubly robust difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.1) in Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and of a linear regression model for the outcome evolution among the comparison units.

When only stationary repeated cross-section data are available (panel = FALSE), the drdid function implements the locally efficient doubly robust difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.4) in Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and of (separate) linear regression models for the outcome of both treated and comparison units, in both pre and post-treatment periods.

When one sets estMethod = "imp" (the default), the nuisance parameters (propensity score and outcome regression parameters) are estimated using the methods described in Sections 3.1 and 3.2 of Sant’Anna and Zhao (2020). In short, the propensity score parameters are estimated using the inverse probability tilting estimator proposed by Graham, Pinto and Pinto (2012), and the outcome regression coefficients are estimated using weighted least squares, where the weights depend on the propensity score estimates; see Sant’Anna and Zhao (2020) for details.

When one sets estMethod = "trad", the propensity score parameters are estimated using maximum likelihood, and the outcome regression coefficients are estimated using ordinary least squares.

The main advantage of using estMethod = "imp" is that the resulting estimator is not only locally efficient and doubly robust for the ATT, but it is also doubly robust for inference; see Sant’Anna and Zhao (2020) for details.

Value
A list containing the following components:
**drdid**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>The DR DID point estimate</td>
</tr>
<tr>
<td>se</td>
<td>The DR DID standard error</td>
</tr>
<tr>
<td>uci</td>
<td>Estimate of the upper bound of a 95% CI for the ATT</td>
</tr>
<tr>
<td>lci</td>
<td>Estimate of the lower bound of a 95% CI for the ATT</td>
</tr>
<tr>
<td>boots</td>
<td>All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL</td>
</tr>
<tr>
<td>att.inf.func</td>
<td>Estimate of the influence function. Default is NULL</td>
</tr>
<tr>
<td>ps.flag</td>
<td>Convergence Flag for the propensity score estimation (only active if estMethod = &quot;imp&quot;): =0 if trust algorithm converged, =1 if IPT (original) algorithm converged (in case it was used), =2 if GLM logit estimator was used (i.e., if both trust and IPT did not converged).</td>
</tr>
<tr>
<td>call.param</td>
<td>The matched call.</td>
</tr>
<tr>
<td>argu</td>
<td>Some arguments used in the call (panel, estMethod, boot, boot.type, nboot, type=&quot;dr&quot;)</td>
</tr>
</tbody>
</table>

**References**


**Examples**

```r
# -----------------------------------------------
# Panel data case
# -----------------------------------------------
# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw_long, nsw_long$treated == 0 | nsw_long$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(unique(eval_lalonde_cps$id), 5000)
eval_lalonde_cps <- eval_lalonde_cps[eval_lalonde_cps$id %in% unit_random,]
# -----------------------------------------------
# Implement improved DR locally efficient DID with panel data
drdid(yname="re", tname = "year", idname = "id", dname = "experimental",
  xformla=~ age+ educ+ black+ married+ ndegree+ hisp+ re74,
  data = eval_lalonde_cps, panel = TRUE)
```

```r
# Implement "traditional" DR locally efficient DID with panel data
drdid(yname="re", tname = "year", idname = "id", dname = "experimental",
  xformla=~ age+ educ+ black+ married+ ndegree+ hisp+ re74,
  data = eval_lalonde_cps, panel = TRUE, estMethod = "trad")
```

# Repeated cross section case
# -----------------------------------------------
# use the simulated data provided in the package
#Implement "improved" DR locally efficient DID with repeated cross-section data
drdid(yname="y", tname = "post", idname = "id", dname = "d",
    xformla= ~ x1 + x2 + x3 + x4,
    data = sim_rc, panel = FALSE, estMethod = "imp")

#Implement "traditional" DR locally efficient DID with repeated cross-section data
drdid(yname="y", tname = "post", idname = "id", dname = "d",
    xformla= ~ x1 + x2 + x3 + x4,
    data = sim_rc, panel = FALSE, estMethod = "trad")

drdid_imp_panel

---

**Improved locally efficient doubly robust DiD estimator for the ATT, with panel data**

**Description**

drdid_imp_panel is used to compute the locally efficient doubly robust estimators for the ATT in difference-in-differences (DiD) setups with panel data. The resulting estimator is also doubly robust for inference; see Section 3.1 of Sant’Anna and Zhao (2020).

**Usage**

```r
drdid_imp_panel(
    y1, y0, D, covariates, i.weights = NULL, boot = FALSE, boot.type = "weighted", nboot = NULL, inffunc = FALSE)
```

**Arguments**

- `y1` An n x 1 vector of outcomes from the post-treatment period.
- `y0` An n x 1 vector of outcomes from the pre-treatment period.
- `D` An n x 1 vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
- `covariates` An n x k matrix of covariates to be used in the propensity score and regression estimation. If covariates = NULL, this leads to an unconditional DID estimator.
- `i.weights` An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.
The `drdid_imp_panel` function implements the locally efficient doubly robust difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.1) in Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and of a linear regression model for the outcome evolution among the comparison units.

The nuisance parameters (propensity score and outcome regression parameters) are estimated using the methods described in Sections 3.1 of Sant’Anna and Zhao (2020). In short, the propensity score parameters are estimated using the inverse probability tilting estimator proposed by Graham, Pinto and Pinto (2012), and the outcome regression coefficients are estimated using weighted least squares, where the weights depend on the propensity score estimates; see Sant’Anna and Zhao (2020) for details.

The resulting estimator is not only locally efficient and doubly robust for the ATT, but it is also doubly robust for inference; see Sant’Anna and Zhao (2020) for details.

**Value**

A list containing the following components:

- **ATT**
  - The DID point estimate.
- **se**
  - The DID standard error.
- **uci**
  - The upper bound of the 95% CI for the ATT.
- **lci**
  - The lower bound of the 95% CI for the ATT.
- **boots**
  - All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.
- **ps.flag**
  - Convergence Flag for the propensity score estimation: =0 if `trust` algorithm converged, =1 if IPW algorithm converged (in case it was used), =2 if GLM logit estimator was used (i.e., if both `trust` and IPT did not converged).
- **att.inf.func**
  - Estimate of the influence function. Default is NULL.
- **call.param**
  - The matched call.
- **argu**
  - Some arguments used (explicitly or not) in the call (panel = TRUE, estMethod = "imp", boot, boot.type, nboot, type="dr")
References


Examples

# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw, nsw$treated == 0 | nsw$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(1:nrow(eval_lalonde_cps), 5000)
eval_lalonde_cps <- eval_lalonde_cps[unit_random,]
# Select some covariates
covX = as.matrix(cbind(eval_lalonde_cps$age, eval_lalonde_cps$educ,
                         eval_lalonde_cps$black, eval_lalonde_cps$married,
                         eval_lalonde_cps$nodegree, eval_lalonde_cps$hisp,
                         eval_lalonde_cps$re74))

# Implement improved DR locally efficient DID with panel data
drdid_imp_panel(y1 = eval_lalonde_cps$re78, y0 = eval_lalonde_cps$re75,
                 D = eval_lalonde_cps$experimental,
                 covariates = covX)

drdid_imp_rc
Improved locally efficient doubly robust DiD estimator for the ATT, with repeated cross-section data

Description

drdid_imp_rc is used to compute the locally efficient doubly robust estimators for the ATT in difference-in-differences (DiD) setups with stationary repeated cross-sectional data. The resulting estimator is also doubly robust for inference; see Section 3.2 of Sant’Anna and Zhao (2020).

Usage

drdid_imp_rc(
  y,
  post,
  D,
  covariates,
  i.weights = NULL,
  boot = FALSE,
  boot.type = "weighted",
  nboot = NULL,
Arguments

- **y**: An \( n \times 1 \) vector of outcomes from the both pre and post-treatment periods.
- **post**: An \( n \times 1 \) vector of Post-Treatment dummies (\( \text{post} = 1 \) if observation belongs to post-treatment period, and \( \text{post} = 0 \) if observation belongs to pre-treatment period.)
- **D**: An \( n \times 1 \) vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
- **covariates**: An \( n \times k \) matrix of covariates to be used in the propensity score and regression estimation. If \( \text{covariates} = \text{NULL} \), this leads to an unconditional DID estimator.
- **i.weights**: An \( n \times 1 \) vector of weights to be used. If \( \text{NULL} \), then every observation has the same weights.
- **boot**: Logical argument to whether bootstrap should be used for inference. Default is \( \text{FALSE} \).
- **boot.type**: Type of bootstrap to be performed (not relevant if \( \text{boot} = \text{FALSE} \)). Options are "weighted" and "multiplier". If \( \text{boot} = \text{TRUE} \), default is "weighted".
- **nboot**: Number of bootstrap repetitions (not relevant if \( \text{boot} = \text{FALSE} \)). Default is 999.
- **inffunc**: Logical argument to whether influence function should be returned. Default is \( \text{FALSE} \).

Details

The `drdid_imp_rc` function implements the locally efficient doubly robust difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.4) in Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and of (separate) linear regression models for the outcome of both treated and comparison units, in both pre and post-treatment periods.

The nuisance parameters (propensity score and outcome regression parameters) are estimated using the methods described in Sections 3.2 of Sant’Anna and Zhao (2020). In short, the propensity score parameters are estimated using the inverse probability tilting estimator proposed by Graham, Pinto and Pinto (2012), and the outcome regression coefficients are estimated using weighted least squares, where the weights depend on the propensity score estimates; see Sant’Anna and Zhao (2020) for details.

The resulting estimator is not only locally efficient and doubly robust for the ATT, but it is also doubly robust for inference; see Sant’Anna and Zhao (2020) for details.

Value

A list containing the following components:

- **ATT**: The DR DID point estimate
- **se**: The DR DID standard error
- **uci**: Estimate of the upper bound of a 95\% CI for the ATT
$drdid\_imp\_rc1$

$lci$ Estimate of the lower bound of a 95% CI for the ATT

$boots$ All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL

$ps.flag$ Convergence Flag for the propensity score estimation: =0 if trust algorithm converged, =1 if IPW algorithm converged (in case it was used), =2 if GLM logit estimator was used (i.e., if both trust and IPT did not converged).

$att.inf.func$ Estimate of the influence function. Default is NULL

$call.param$ The matched call.

$argu$ Some arguments used (explicitly or not) in the call (panel = FALSE, estMethod = "imp", boot, boot.type, nboot, type="dr")

References


Examples

# use the simulated data
covX = as.matrix(sim_rc[,5:8])

# Implement the improved, locally efficient DR DID estimator
drdid_imp_rc(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d, covariates= covX)

$drdid\_imp\_rc1$ Improved doubly robust DiD estimator for the ATT, with repeated cross-section data

Description

$drdid\_imp\_rc1$ is used to compute the doubly robust estimators for the ATT in difference-in-differences (DiD) setups with stationary repeated cross-sectional data. The resulting estimator is also doubly robust for inference, though it is not locally efficient; see Section 3.2 of Sant'Anna and Zhao (2020).

Usage

$drdid\_imp\_rc1($

y,
post,
D,
covariates,
library(drdid)

# Arguments

Arguments

\[ y \]
An \( n \times 1 \) vector of outcomes from the both pre and post-treatment periods.

\[ post \]
An \( n \times 1 \) vector of Post-Treatment dummies (\( post = 1 \) if observation belongs to post-treatment period, and \( post = 0 \) if observation belongs to pre-treatment period.)

\[ D \]
An \( n \times 1 \) vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).

\[ covariates \]
An \( n \times k \) matrix of covariates to be used in the propensity score and regression estimation. If covariates = NULL, this leads to an unconditional DID estimator.

\[ i.weights \]
An \( n \times 1 \) vector of weights to be used. If NULL, then every observation has the same weights.

\[ boot \]
Logical argument to whether bootstrap should be used for inference. Default is FALSE.

\[ boot.type \]
Type of bootstrap to be performed (not relevant if \( boot = FALSE \)). Options are "weighted" and "multiplier". If \( boot = TRUE \), default is "weighted".

\[ nboot \]
Number of bootstrap repetitions (not relevant if \( boot = FALSE \)). Default is 999.

\[ inffunc \]
Logical argument to whether influence function should be returned. Default is FALSE.

Details

The \texttt{drdid_imp_rc1} function implements the doubly robust difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.3) in Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and of (separate) linear regression models for the outcome among the comparison units in both pre and post-treatment time periods. Importantly, this estimator is not locally efficient for the ATT.

The nuisance parameters (propensity score and outcome regression parameters) are estimated using the methods described in Sections 3.2 of Sant’Anna and Zhao (2020). In short, the propensity score parameters are estimated using the inverse probability tilting estimator proposed by Graham, Pinto and Pinto (2012), and the outcome regression coefficients are estimated using weighted least squares, where the weights depend on the propensity score estimates; see Sant’Anna and Zhao (2020) for details.

The resulting estimator is not only doubly robust for the ATT, but it is also doubly robust for inference. However, we stress that it is not locally efficient; see Sant’Anna and Zhao (2020) for details.
Value

A list containing the following components:

- **ATT**: The DR DID point estimate
- **se**: The DR DID standard error
- **uci**: Estimate of the upper bound of a 95% CI for the ATT
- **lci**: Estimate of the lower bound of a 95% CI for the ATT
- **boots**: All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.
- **ps.flag**: Convergence Flag for the propensity score estimation: =0 if trust algorithm converged, =1 if IPW algorithm converged (in case it was used), =2 if GLM logit estimator was used (i.e., if both trust and IPT did not converged).
- **att.inf.func**: Estimate of the influence function. Default is NULL
- **call.param**: The matched call.
- **argu**: Some arguments used (explicitly or not) in the call (panel = FALSE, estMethod = "imp2", boot, boot.type, nboot, type="dr")

References


Examples

```r
# use the simulated data provided in the package
covX = as.matrix(sim_rc[,5:8])
# Implement the improved DR DID estimator (but not locally efficient!)
drdid_imp_rc1(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d,
covariates= covX)
```

---

**drdid_panel**

Locally efficient doubly robust DiD estimator for the ATT, with panel data

Description

`drdid_panel` is used to compute the locally efficient doubly robust estimators for the ATT in difference-in-differences (DiD) setups with panel data.
Usage

drdid_panel(
  y1,
  y0,
  D,
  covariates,
  i.weights = NULL,
  boot = FALSE,
  boot.type = "weighted",
  nboot = NULL,
  inffunc = FALSE
)

Arguments

y1 An n x 1 vector of outcomes from the post-treatment period.
y0 An n x 1 vector of outcomes from the pre-treatment period.
D An n x 1 vector of Group indicators (=1 if observation is treated in the post-
treatment, =0 otherwise).
covariates An n x k matrix of covariates to be used in the propensity score and regression
estimation. If covariates = NULL, this leads to an unconditional DID estimator.
i.weights An n x 1 vector of weights to be used. If NULL, then every observation has the
same weights.
boot Logical argument to whether bootstrap should be used for inference. Default is
FALSE.
boot.type Type of bootstrap to be performed (not relevant if boot = FALSE). Options are
"weighted" and "multiplier". If boot = TRUE, default is "weighted".
nboot Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
inffunc Logical argument to whether influence function should be returned. Default is
FALSE.

Details

The drdid_panel function implements the locally efficient doubly robust difference-in-differences
(DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.1) in
Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the
probability of being in the treated group, and of a linear regression model for the outcome evolution
among the comparison units.

The propensity score parameters are estimated using maximum likelihood, and the outcome regression
coefficients are estimated using ordinary least squares.

Value

A list containing the following components:

ATT The DR DID point estimate.
The DR DID standard error.

Estimate of the upper bound of a 95% CI for the ATT.

Estimate of the lower bound of a 95% CI for the ATT.

All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.

Estimate of the influence function. Default is NULL.

The matched call.

Some arguments used (explicitly or not) in the call (panel = TRUE, estMethod = "trad", boot, boot.type, nboot, type="dr")

**References**


**Examples**

# Form the Lalonde sample with CPS comparison group (data in wide format)
eval_lalonde_cps <- subset(nsw, nsw$treated == 0 | nsw$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(1:nrow(eval_lalonde_cps), 5000)
eval_lalonde_cps <- eval_lalonde_cps[unit_random,
# Select some covariates
covX = as.matrix(cbind(eval_lalonde_cps$age, eval_lalonde_cps$educ,
eval_lalonde_cps$black, eval_lalonde_cps$married,
 eval_lalonde_cps$nodegree, eval_lalonde_cps$hisp,
  eval_lalonde_cps$re74))

# Implement traditional DR locally efficient DID with panel data
drdid_panel(y1 = eval_lalonde_cps$re78, y0 = eval_lalonde_cps$re75,
D = eval_lalonde_cps$experimental,
covariates = covX)

---

drdid_rc

Locally efficient doubly robust DiD estimator for the ATT, with repeated cross-section data

**Description**

`drdid_rc` is used to compute the locally efficient doubly robust estimators for the ATT in difference-in-differences (DiD) setups with stationary repeated cross-sectional data.
Usage

drdid_rc(
  y,
  post,
  D,
  covariates,
  i.weights = NULL,
  boot = FALSE,
  boot.type = "weighted",
  nboot = NULL,
  inffunc = FALSE
)

Arguments

  y  An n x 1 vector of outcomes from the both pre and post-treatment periods.
  post  An n x 1 vector of Post-Treatment dummies (post = 1 if observation belongs to post-treatment period, and post = 0 if observation belongs to pre-treatment period.)
  D  An n x 1 vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
  covariates  An n x k matrix of covariates to be used in the propensity score and regression estimation. If covariates = NULL, this leads to an unconditional DID estimator.
  i.weights  An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.
  boot  Logical argument to whether bootstrap should be used for inference. Default is FALSE.
  boot.type  Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
  nboot  Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
  inffunc  Logical argument to whether influence function should be returned. Default is FALSE.

Details

The drdid_rc function implements the locally efficient doubly robust difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.4) in Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and of (separate) linear regression models for the outcome of both treated and comparison units, in both pre and post-treatment periods.

The propensity score parameters are estimated using maximum likelihood, and the outcome regression coefficients are estimated using ordinary least squares; see Sant’Anna and Zhao (2020) for details.
**Value**

A list containing the following components:

- **ATT**: The TR-DR DID point estimate
- **se**: The TR-DR DID standard error
- **uci**: Estimate of the upper bound of a 95% CI for the ATT
- **lci**: Estimate of the lower bound of a 95% CI for the ATT
- **boots**: All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.
- **att.inf.func**: Estimate of the influence function. Default is NULL.
- **call.param**: The matched call.
- **argu**: Some arguments used (explicitly or not) in the call (panel = TRUE, estMethod = "trad", boot, boot.type, nboot, type="dr")

**References**


**Examples**

```r
# use the simulated data provided in the package
covX = as.matrix(sim_rc[,5:8])
# Implement the 'traditional' locally efficient DR DID estimator
drdid_rc(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d,
        covariates= covX)
```

**Description**

drdid_rc1 is used to compute the doubly robust estimators for the ATT in difference-in-differences (DiD) setups with stationary repeated cross-sectional data. The resulting estimator is not locally efficient; see Section 3.2 of Sant’Anna and Zhao (2020).

**Usage**

```r
drdid_rc1(  
    y,  
    post,  
    D,  
    covariates,  
    i.weights = NULL,
)```
Arguments

y
An n x 1 vector of outcomes from the both pre and post-treatment periods.

post
An n x 1 vector of Post-Treatment dummies (post = 1 if observation belongs to post-treatment period, and post = 0 if observation belongs to pre-treatment period.)

D
An n x 1 vector of Group indicators ( =1 if observation is treated in the post-treatment, =0 otherwise).

covariates
An n x k matrix of covariates to be used in the propensity score and regression estimation. If covariates = NULL, this leads to an unconditional DID estimator.

i.weights
An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.

boot
Logical argument to whether bootstrap should be used for inference. Default is FALSE.

boot.type
Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".

nboot
Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.

inffunc
Logical argument to whether influence function should be returned. Default is FALSE.

Details

The drdid_rc1 function implements the doubly robust difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (3.3) in Sant’Anna and Zhao (2020). This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and of (separate) linear regression models for the outcome among the comparison units in both pre and post-treatment time periods. Importantly, this estimator is not locally efficient for the ATT.

The propensity score parameters are estimated using maximum likelihood, and the outcome regression coefficients are estimated using ordinary least squares.

The resulting estimator is not locally efficient; see Sant’Anna and Zhao (2020) for details.

Value

A list containing the following components:

ATT
The DR DID point estimate

se
The DR DID standard error

uci
Estimate of the upper bound of a 95% CI for the ATT

lci
Estimate of the lower bound of a 95% CI for the ATT
boots  All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.
att.inf.func  Estimate of the influence function. Default is NULL.
call.param  The matched call.
argu  Some arguments used (explicitly or not) in the call (panel = FALSE, estMethod = "trad2", boot, boot.type, nboot, type="dr")

References


Examples

# use the simulated data provided in the package
covX = as.matrix(sim_rc[,5:8])
# Implement the 'traditional' DR DID estimator (not locally efficient!)
drdid_rc1(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d, covariates= covX)

ipwdid  Inverse probability weighted DiD estimators for the ATT

Description

ipwdid computes the inverse probability weighted estimators for the average treatment effect on the treated in difference-in-differences (DiD) setups. It can be used with panel or stationary repeated cross-sectional data, with or without normalized (stabilized) weights. See Abadie (2005) and Sant'Anna and Zhao (2020) for details.

Usage

ipwdid(
  yname,
  tname,
  idname,
  dname,
  xformula = NULL,
  data,
  panel = TRUE,
  normalized = TRUE,
  weightsname = NULL,
  boot = FALSE,
  boot.type = c("weighted", "multiplier"),
  nboot = 999,
  inffunc = FALSE
)
Arguments

- **yname**: The name of the outcome variable.
- **tname**: The name of the column containing the time periods.
- **idname**: The name of the column containing the unit id name.
- **dname**: The name of the column containing the treatment group (=1 if observation is treated in the post-treatment, =0 otherwise)
- **xformla**: A formula for the covariates to include in the model. It should be of the form ~ X1 + X2 (intercept should not be listed as it is always automatically included). Default is NULL which is equivalent to xformla=~1.
- **data**: The name of the data.frame that contains the data.
- **panel**: Whether or not the data is a panel dataset. The panel dataset should be provided in long format – that is, where each row corresponds to a unit observed at a particular point in time. The default is TRUE. When panel = TRUE, the variable idname must be set. When panel = FALSE, the data is treated as stationary repeated cross sections.
- **normalized**: Logical argument to whether IPW weights should be normalized to sum up to one. Default is TRUE.
- **weightsname**: The name of the column containing the sampling weights. If NULL, then every observation has the same weights.
- **boot**: Logical argument to whether bootstrap should be used for inference. Default is FALSE and analytical standard errors are reported.
- **boot.type**: Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
- **nboot**: Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
- **inffunc**: Logical argument to whether influence function should be returned. Default is FALSE.

Details

The ipwdid function implements the inverse probability weighted (IPW) difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) proposed by Abadie (2005) (normalized = FALSE) or Hajek-type version defined in equations (4.1) and (4.2) in Sant’Anna and Zhao (2020), when either panel data or stationary repeated cross-sectional data are available. This estimator makes use of a logistic propensity score model for the probability of being in the treated group, and the propensity score parameters are estimated via maximum likelihood.

Value

A list containing the following components:

- **ATT**: The IPW DID point estimate
- **se**: The IPW DID standard error
- **uci**: Estimate of the upper bound of a 95% CI for the ATT
- **lci**: Estimate of the lower bound of a 95% CI for the ATT
boots

All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.

att.inf.func

Estimate of the influence function. Default is NULL.

call.param

The matched call.

argu

Some arguments used in the call (panel, normalized, boot, boot.type, nboot, type="ipw")

References


Examples

```r
# -----------------------------------------------
# Panel data case
# -----------------------------------------------
# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw_long, nsw_long$treated == 0 | nsw_long$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(unique(eval_lalonde_cps$id), 5000)
eval_lalonde_cps <- eval_lalonde_cps[eval_lalonde_cps$id %in% unit_random,]
# Implement IPW DID with panel data (normalized weights)
ipwdid(yname="re", tname = "year", idname = "id", dname = "experimental",
      xformla= ~ age+ educ+ black+ married+ nodegree+ hisp+ re74,
      data = eval_lalonde_cps, panel = TRUE)

# -----------------------------------------------
# Repeated cross section case
# -----------------------------------------------
# Use the simulated data provided in the package
# Implement IPW DID with repeated cross-section data (normalized weights)
# Use Bootstrap to make inference with 199 bootstrap draws (just for illustration)
ipwdid(yname="y", tname = "post", idname = "id", dname = "d",
       xformla= ~ x1 + x2 + x3 + x4,
       data = sim_rc, panel = FALSE,
       boot = TRUE, nboot = 199)
```

ipw_did_panel

Inverse probability weighted DiD estimator, with panel data
**Description**

`ipw_did_panel` is used to compute inverse probability weighted (IPW) estimators for the ATT in difference-in-differences (DiD) setups with panel data. IPW weights are not normalized to sum up to one, that is, the estimator is of the Horwitz-Thompson type.

**Usage**

```r
ipw_did_panel(
  y1,
  y0,
  D,
  covariates,
  i.weights = NULL,
  boot = FALSE,
  boot.type = "weighted",
  nboot = NULL,
  inffunc = FALSE
)
```

**Arguments**

- `y1`  
  An \( n \times 1 \) vector of outcomes from the post-treatment period.

- `y0`  
  An \( n \times 1 \) vector of outcomes from the pre-treatment period.

- `D`  
  An \( n \times 1 \) vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).

- `covariates`  
  An \( n \times k \) matrix of covariates to be used in the propensity score estimation. If `covariates = NULL`, this leads to an unconditional DID estimator.

- `i.weights`  
  An \( n \times 1 \) vector of weights to be used. If `NULL`, then every observation has the same weights.

- `boot`  
  Logical argument to whether bootstrap should be used for inference. Default is `FALSE`.

- `boot.type`  
  Type of bootstrap to be performed (not relevant if `boot = FALSE`). Options are "weighted" and "multiplier". If `boot = TRUE`, default is "weighted".

- `nboot`  
  Number of bootstrap repetitions (not relevant if `boot = FALSE`). Default is 999.

- `inffunc`  
  Logical argument to whether influence function should be returned. Default is `FALSE`.

**Value**

A list containing the following components:

- `ATT`  
  The IPW DID point estimate.

- `se`  
  The IPW DID standard error

- `uci`  
  Estimate of the upper bound of a 95% CI for the ATT

- `lci`  
  Estimate of the lower bound of a 95% CI for the ATT
boots 

All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.

att.inf.func 

Estimate of the influence function. Default is NULL.

call.param 

The matched call.

argu 

Some arguments used (explicitly or not) in the call (panel = TRUE, normalized = FALSE, boot, boot.type, nboot, type="ipw")

References


Examples

# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw, nsw$treated == 0 | nsw$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(1:nrow(eval_lalonde_cps), 5000)
eval_lalonde_cps <- eval_lalonde_cps[unit_random,]
# Select some covariates
covX = as.matrix(cbind(eval_lalonde_cps$age, eval_lalonde_cps$educ,
                      eval_lalonde_cps$black, eval_lalonde_cps$married,
                      eval_lalonde_cps$nodegree, eval_lalonde_cps$hisp,
                      eval_lalonde_cps$re74))
# Implement (unnormalized) IPW DID with panel data
ipw_did_panel(y1 = eval_lalonde_cps$re78, y0 = eval_lalonde_cps$re75,
               D = eval_lalonde_cps$experimental,
               covariates = covX)

Description

ipw_did_rc is used to compute inverse probability weighted (IPW) estimators for the ATT in difference-in-differences (DiD) setups with stationary cross-sectional data. IPW weights are not normalized to sum up to one, that is, the estimator is of the Horwitz-Thompson type.
Usage

ipw_did_rc(
  y, post, D, covariates, i.weights = NULL, boot = FALSE, boot.type = "weighted", nboot = NULL, inffunc = FALSE
)

Arguments

y An n x 1 vector of outcomes from the both pre and post-treatment periods.
post An n x 1 vector of Post-Treatment dummies (post = 1 if observation belongs to post-treatment period, and post = 0 if observation belongs to pre-treatment period.)
D An n x 1 vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
covariates An n x k matrix of covariates to be used in the propensity score estimation. If covariates = NULL, this leads to an unconditional DID estimator.
i.weights An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.
boot Logical argument to whether bootstrap should be used for inference. Default is FALSE.
boot.type Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
nboot Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
inffunc Logical argument to whether influence function should be returned. Default is FALSE.

Value

A list containing the following components:

ATT The IPW DID point estimate.
se The IPW DID standard error
uci Estimate of the upper bound of a 95% CI for the ATT
lci Estimate of the lower bound of a 95% CI for the ATT
boots All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL
att.inf.func Estimate of the influence function. Default is NULL
call.param The matched call.
Some arguments used (explicitly or not) in the call (panel = FALSE, normalized = FALSE, boot, boot.type, nboot, type="ipw")

References


Examples

# use the simulated data provided in the package
covX = as.matrix(sim_rc[,5:8])
# Implement unnormalized IPW DID estimator
ipw_did_rc(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d, covariates= covX)

nsw

National Supported Work Demonstration dataset

Description

nsw contains all the subsamples of from the National Supported Work (NSW) Demonstration analyzed used by Smith and Todd (2005) in their paper “Does matching overcome LaLonde’s critique of nonexperimental estimators?”.

Usage

nsw

Format

A data frame in "wide" format with 19204 observations on the following and 14 variables:

treated an indicator variable for treatment status. Missing if not part of the NSW experimental sample
age age in years.
educ years of schooling.
black indicator variable for blacks.
marricd indicator variable for marital status.
nodegree indicator variable for high school diploma.
dwincl indicator variable for inclusion in Dehejia and Wahba sample. Missing if not part of the experimental sample
re74 real earnings in 1974 (pre-treatment).
**nsw_long**

- **re75** real earnings in 1975 (pre-treatment).
- **re78** real earnings in 1978 (post-treatment).
- **hisp** indicator variable for Hispanics.
- **early_ra** indicator variable for inclusion in the early random assignment sample in Smith and Todd (2005). Missing if not part of the experimental sample.
- **sample** 1 if NSW (experimental sample), 2 if CPS comparison group, 3 if PSID comparison group.
- **experimental** 1 if in experimental sample, 0 otherwise.

**Source**


**References**


---

**nsw_long**

*National Supported Work Demonstration dataset, in long format*

**Description**

nsw_long is the same dataset as nsw but in a long format.

**Usage**

nsw_long

**Format**

A data frame in "long" format with 38408 observations on the following and 15 variables:

- **id** unique identifier for each cross-sectional unit (worker).
- **year** year. 1975 is the pre-treatment and 1978 is the post-treatment.
- **treated** an indicator variable for treatment status. Missing if not part of the NSW experimental sample.
- **age** age in years.
- **educ** years of schooling.
- **black** indicator variable for blacks.
- **married** indicator variable for marital status.
- **nodegree** indicator variable for high school diploma.
**ordid**

**Description**

`ordid` computes the outcome regressions estimators for the average treatment effect on the treated in difference-in-differences (DiD) setups. It can be used with panel or repeated cross section data. See Sant'Anna and Zhao (2020) for details.

**Usage**

```r
ordid(
  yname,
  tname,
  idname,
  dname,
  xformula = NULL,
  data,
  panel = TRUE,
  weightsname = NULL,
  boot = FALSE,
  boot.type = c("weighted", "multiplier"),
  nboot = 999,
inffunc = FALSE
)
```

**Source**


**References**


**ordid**

Outcome regression DiD estimators for the ATT

**Description**

`ordid` computes the outcome regressions estimators for the average treatment effect on the treated in difference-in-differences (DiD) setups. It can be used with panel or repeated cross section data. See Sant’Anna and Zhao (2020) for details.

**Usage**

`ordid(
  yname,
  tname,
  idname,
  dname,
  xformula = NULL,
  data,
  panel = TRUE,
  weightsname = NULL,
  boot = FALSE,
  boot.type = c("weighted", "multiplier"),
  nboot = 999,
inffunc = FALSE
)"
The `ordid` function implements outcome regression difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (2.2) of Sant’Anna and Zhao (2020). The estimator follows the same spirit of the nonparametric estimators proposed by Heckman, Ichimura and Todd (1997), though here the the outcome regression models are assumed to be linear in covariates (parametric). The nuisance parameters (outcome regression coefficients) are estimated via ordinary least squares.

**Value**

A list containing the following components:

- **ATT** The IPW DID point estimate
- **se** The IPW DID standard error
- **uci** Estimate of the upper bound of a 95% CI for the ATT
- **lci** Estimate of the lower bound of a 95% CI for the ATT
- **boots** All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.

**Arguments**

- **yname** The name of the outcome variable.
- **tname** The name of the column containing the time periods.
- **idname** The name of the column containing the unit id name.
- **dname** The name of the column containing the treatment group (=1 if observation is treated in the post-treatment, =0 otherwise).
- **xformla** A formula for the covariates to include in the model. It should be of the form ~ X1 + X2. (intercept should not be listed as it is always automatically included). Default is NULL which is equivalent to `xformla=~1`.
- **data** The name of the data.frame that contains the data.
- **panel** Whether or not the data is a panel dataset. The panel dataset should be provided in long format – that is, where each row corresponds to a unit observed at a particular point in time. The default is TRUE. When `panel = TRUE`, the variable `idname` must be set. When `panel = FALSE`, the data is treated as stationary repeated cross sections.
- **weightsname** The name of the column containing the sampling weights. If NULL, then every observation has the same weights.
- **boot** Logical argument to whether bootstrap should be used for inference. Default is FALSE and analytical standard errors are reported.
- **boot.type** Type of bootstrap to be performed (not relevant if `boot = FALSE`). Options are "weighted" and "multiplier". If `boot = TRUE`, default is "weighted".
- **nboot** Number of bootstrap repetitions (not relevant if `boot = FALSE`). Default is 999.
- **inffunc** Logical argument to whether influence function should be returned. Default is FALSE.
reg_did_panel

Outcome regression DiD estimator for the ATT, with panel data

Description

reg_did_panel computes the outcome regressions estimators for the average treatment effect on the treated in difference-in-differences (DiD) setups with panel data.

argu

Some arguments used in the call (panel, normalized, boot, boot.type, nboot, type="or")

References


Examples

# -----------------------------------------------
# Panel data case
# -----------------------------------------------
# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw_long, nsw_long$treated == 0 | nsw_long$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(unique(eval_lalonde_cps$id), 5000)
eval_lalonde_cps <- eval_lalonde_cps[eval_lalonde_cps$id %in% unit_random,]
# Implement OR DID with panel data
ordid(yname="re", tname = "year", idname = "id", dname = "experimental",
     xformla= ~ age+ educ+ black+ married+ nodegree+ hisp+ re74,
     data = eval_lalonde_cps, panel = TRUE)

# -----------------------------------------------
# Repeated cross section case
# -----------------------------------------------
# use the simulated data provided in the package
# Implement OR DID with repeated cross-section data
# use Bootstrap to make inference with 199 bootstrap draws (just for illustration)
ordid(yname="y", tname = "post", idname = "id", dname = "d",
      xformla= ~ x1 + x2 + x3 + x4,
      data = sim_rc, panel = FALSE,
      boot = TRUE, nboot = 199)
Usage

```r
reg_did_panel(
  y1, y0, D, covariates, i.weights = NULL, boot = FALSE, boot.type = "weighted", nboot = NULL, inffunc = FALSE
)
```

Arguments

- `y1`: An $n \times 1$ vector of outcomes from the post-treatment period.
- `y0`: An $n \times 1$ vector of outcomes from the pre-treatment period.
- `D`: An $n \times 1$ vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
- `covariates`: An $n \times k$ matrix of covariates to be used in the regression estimation. If covariates = NULL, this leads to an unconditional DID estimator.
- `i.weights`: An $n \times 1$ vector of weights to be used. If NULL, then every observation has the same weights.
- `boot`: Logical argument to whether bootstrap should be used for inference. Default is FALSE.
- `boot.type`: Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
- `nboot`: Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
- `inffunc`: Logical argument to whether influence function should be returned. Default is FALSE.

Details

The `reg_did_panel` function implements outcome regression difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (2.2) of Sant'Anna and Zhao (2020) when panel data are available. The estimator follows the same spirit of the non-parametric estimators proposed by Heckman, Ichimura and Todd (1997), though here the the outcome regression models are assumed to be linear in covariates (parametric).

The nuisance parameters (outcome regression coefficients) are estimated via ordinary least squares.

Value

A list containing the following components:

- `ATT`: The Reg DID point estimate
- `se`: The Reg DID standard error
reg_did_rc

uci

Estimate of the upper bound of a 95% CI for the ATT

lci

Estimate of the lower bound of a 95% CI for the ATT

boots

All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL

att.inf.func

Estimate of the influence function. Default is NULL

call.param

The matched call.

argu

Some arguments used (explicitly or not) in the call (panel = TRUE, boot, boot.type, nboot, type="or")

References


Examples

# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw, nsw$treated == 0 | nsw$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(1:nrow(eval_lalonde_cps), 5000)
eval_lalonde_cps <- eval_lalonde_cps[unit_random,]
# Select some covariates
covX = as.matrix(cbind(eval_lalonde_cps$age, eval_lalonde_cps$educ,
eval_lalonde_cps$black, eval_lalonde_cps$married,
eval_lalonde_cps$nodegree, eval_lalonde_cps$hisp,
eval_lalonde_cps$re74))
# Implement OR DID with panel data
reg_did_panel(y1 = eval_lalonde_cps$re78, y0 = eval_lalonde_cps$re75,
D = eval_lalonde_cps$experimental,
covariates = covX)

reg_did_rc

Outcome regression DiD estimator for the ATT, with repeated cross-section data

Description

reg_did_rc computes the outcome regressions estimators for the average treatment effect on the treated in difference-in-differences (DiD) setups with stationary repeated cross-sectional data.
Usage

    reg_did_rc(
        y,  # An n x 1 vector of outcomes from the both pre and post-treatment periods.
        post,  # An n x 1 vector of Post-Treatment dummies (post = 1 if observation belongs to post-treatment period, and post = 0 if observation belongs to pre-treatment period.)
        D,  # An n x 1 vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
        covariates,  # An n x k matrix of covariates to be used in the regression estimation. If covariates = NULL, this leads to an unconditional DID estimator.
        i.weights = NULL,  # An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.
        boot = FALSE,  # Logical argument to whether bootstrap should be used for inference. Default is FALSE.
        boot.type = "weighted",  # Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
        nboot = NULL,  # Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
        inffunc = FALSE  # Logical argument to whether influence function should be returned. Default is FALSE.
    )

Arguments

- **y**: An n x 1 vector of outcomes from the both pre and post-treatment periods.
- **post**: An n x 1 vector of Post-Treatment dummies (post = 1 if observation belongs to post-treatment period, and post = 0 if observation belongs to pre-treatment period.)
- **D**: An n x 1 vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
- **covariates**: An n x k matrix of covariates to be used in the regression estimation. If covariates = NULL, this leads to an unconditional DID estimator.
- **i.weights**: An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.
- **boot**: Logical argument to whether bootstrap should be used for inference. Default is FALSE.
- **boot.type**: Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
- **nboot**: Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
- **inffunc**: Logical argument to whether influence function should be returned. Default is FALSE.

Details

The `reg_did_rc` function implements outcome regression difference-in-differences (DID) estimator for the average treatment effect on the treated (ATT) defined in equation (2.2) of Sant’Anna and Zhao (2020) when stationary repeated cross-sectional data are available. The estimator follows the same spirit of the nonparametric estimators proposed by Heckman, Ichimura and Todd (1997), though here the outcome regression models are assumed to be linear in covariates (parametric).

The nuisance parameters (outcome regression coefficients) are estimated via ordinary least squares.

Value

A list containing the following components:
### ATT
The Reg DID point estimate

### se
The Reg DID standard error

### uci
Estimate of the upper bound of a 95% CI for the ATT

### lci
Estimate of the lower bound of a 95% CI for the ATT

### boots
All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL

### att.inf.func
Estimate of the influence function. Default is NULL

### call.param
The matched call.

### argu
Some arguments used (explicitly or not) in the call (panel = FALSE, boot, boot.type, nboot, type="or")

### References


### Examples

```r
# use the simulated data provided in the package
covX = as.matrix(sim_rc[,5:8])
# Implement OR DID estimator
reg_did_rc(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d, covariates= covX)
```

---

**sim_rc**  
*Simulated repeated cross-section data*

### Description

sim_rc contains a simulated dataset following the DGP1 in Sant’Anna and Zhao (2020).

### Usage

sim_rc
std_ipw_did_panel

Format

A data frame in "long" format with 1000 observations on the following and 8 variables:

- **id**: unique identifier for each cross-sectional unit.
- **post**: an indicator variable for post-treatment period (1 if post, 0 if pre treatment period).
- **y**: outcome of interest
- **d**: an indicator variable for treatment group. Equal to 1 if experience treatment in the post-treatment period; equal to 0 if never experience treatment.
- **x1**: Covariate z1 in Sant’Anna and Zhao(2020)
- **x2**: Covariate z2 in Sant’ Anna and Zhao(2020)
- **x3**: Covariate z3 in Sant’Anna and Zhao(2020)
- **x4**: Covariate z4 in Sant’ Anna and Zhao(2020)

Source

Sant’Anna and Zhao (2020)

References


Description

std_ipw_did_panel is used to compute inverse probability weighted (IPW) estimators for the ATT in difference-in-differences (DiD) setups with panel data. IPW weights are normalized to sum up to one, that is, the estimator is of the Hajek type.

Usage

```r
std_ipw_did_panel(
  y1,
  y0,
  D,
  covariates,
  i.weights = NULL,
  boot = FALSE,
  boot.type = "weighted",
  nboot = NULL,
  inffunc = FALSE
)
```
Arguments

\( y_1 \) An \( n \times 1 \) vector of outcomes from the post-treatment period.
\( y_0 \) An \( n \times 1 \) vector of outcomes from the pre-treatment period.
\( D \) An \( n \times 1 \) vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).

\texttt{covariates} An \( n \times k \) matrix of covariates to be used in the propensity score estimation. If \texttt{covariates} = NULL, this leads to an unconditional DID estimator.
\texttt{i.weights} An \( n \times 1 \) vector of weights to be used. If NULL, then every observation has the same weights.
\texttt{boot} Logical argument to whether bootstrap should be used for inference. Default is FALSE.
\texttt{boot.type} Type of bootstrap to be performed (not relevant if \texttt{boot} = FALSE). Options are "weighted" and "multiplier". If \texttt{boot} = TRUE, default is "weighted".
\texttt{nboot} Number of bootstrap repetitions (not relevant if \texttt{boot} = FALSE). Default is 999.
\texttt{inffunc} Logical argument to whether influence function should be returned. Default is FALSE.

Value

A list containing the following components:

\texttt{ATT} The IPW DID point estimate.
\texttt{se} The IPW DID standard error
\texttt{uci} Estimate of the upper bound of a 95% CI for the ATT
\texttt{lci} Estimate of the lower bound of a 95% CI for the ATT
\texttt{boots} All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL
\texttt{att.inf.func} Estimate of the influence function. Default is NULL
\texttt{call.param} The matched call.
\texttt{argu} Some arguments used (explicitly or not) in the call (panel = TRUE, normalized = TRUE, boot, boot.type, nboot, type=\texttt{"ipw"} )

References


Examples

# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw, nsw$treated == 0 | nsw$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(1:nrow(eval_lalonde_cps), 5000)
eval_lalonde_cps <- eval_lalonde_cps[unit_random,]
# Select some covariates
covX = as.matrix(cbind(eval_lalonde_cps$age, eval_lalonde_cps$educ,
    eval_lalonde_cps$black, eval_lalonde_cps$married,
    eval_lalonde_cps$nodegree, eval_lalonde_cps$hisp,
    eval_lalonde_cps$re74))
# Implement normalized IPW DID with panel data
std_ipw_did_panel(y1 = eval_lalonde_cps$re78, y0 = eval_lalonde_cps$re75,
    D = eval_lalonde_cps$experimental,
    covariates = covX)

std_ipw_did_rc

Standardized inverse probability weighted DiD estimator, with repeated cross-section data

Description

std_ipw_did_rc is used to compute inverse probability weighted (IPW) estimators for the ATT in DID setups with stationary repeated cross-sectional data. IPW weights are normalized to sum up to one, that is, the estimator is of the Hajek type.

Usage

std_ipw_did_rc(
  y, 
  post, 
  D, 
  covariates, 
  i.weights = NULL, 
  boot = FALSE, 
  boot.type = "weighted", 
  nboot = NULL, 
  inffunc = FALSE
)

Arguments

y An n x 1 vector of outcomes from the both pre and post-treatment periods.

post An n x 1 vector of Post-Treatment dummies (post = 1 if observation belongs to post-treatment period, and post = 0 if observation belongs to pre-treatment period.)
D  An \( n \times 1 \) vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).

covariates  An \( n \times k \) matrix of covariates to be used in the propensity score estimation. If covariates = NULL, this leads to an unconditional DID estimator.

i.weights  An \( n \times 1 \) vector of weights to be used. If NULL, then every observation has the same weights.

boot  Logical argument to whether bootstrap should be used for inference. Default is FALSE.

boot.type  Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".

nboot  Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.

inffunc  Logical argument to whether influence function should be returned. Default is FALSE.

Value

A list containing the following components:

ATT  The IPW DID point estimate.

se  The IPW DID standard error

uci  Estimate of the upper bound of a 95% CI for the ATT

lci  Estimate of the lower bound of a 95% CI for the ATT

boots  All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL

att.inf.func  Estimate of the influence function. Default is NULL

call.param  The matched call.

argu  Some arguments used (explicitly or not) in the call (panel = FALSE, normalized = TRUE, boot, boot.type, nboot, type="ipw")

References


Examples

# use the simulated data provided in the package
covX = as.matrix(sim_rc[,5:8])
# Implement normalized IPW DID estimator
std_ipw_did_rc(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d, covariates= covX)
Description

twfe_did_panel is used to compute linear two-way fixed effects estimators for the ATT in difference-in-differences (DiD) setups with panel data. As illustrated by Sant’Anna and Zhao (2020), this estimator generally do not recover the ATT. We encourage empiricists to adopt alternative specifications.

Usage

twfe_did_panel(
  y1,
  y0,
  D,
  covariates,
  i.weights = NULL,
  boot = FALSE,
  boot.type = "weighted",
  nboot = NULL,
  inffunc = FALSE
)

Arguments

- **y1**: An n x 1 vector of outcomes from the post-treatment period.
- **y0**: An n x 1 vector of outcomes from the pre-treatment period.
- **D**: An n x 1 vector of Group indicators (=1 if observation is treated in the post-treatment, =0 otherwise).
- **covariates**: An n x k matrix of covariates to be used in the regression estimation.
- **i.weights**: An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.
- **boot**: Logical argument to whether bootstrap should be used for inference. Default is FALSE.
- **boot.type**: Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
- **nboot**: Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
- **inffunc**: Logical argument to whether influence function should be returned. Default is FALSE.
Value
A list containing the following components:

- ATT: The TWFE DID point estimate
- se: The TWFE DID standard error
- uci: Estimate of the upper bound of a 95% CI for the TWFE parameter.
- lci: Estimate of the lower bound of a 95% CI for the TWFE parameter.
- boots: All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL
- att.inf.func: Estimate of the influence function. Default is NULL

Examples

```r
# Form the Lalonde sample with CPS comparison group
eval_lalonde_cps <- subset(nsw, nsw$treated == 0 | nsw$sample == 2)
# Further reduce sample to speed example
set.seed(123)
unit_random <- sample(1:nrow(eval_lalonde_cps), 5000)
eval_lalonde_cps <- eval_lalonde_cps[unit_random,]
# Select some covariates
covX = as.matrix(cbind(eval_lalonde_cps$age, eval_lalonde_cps$educ,
                        eval_lalonde_cps$black, eval_lalonde_cps$married,
                        eval_lalonde_cps$nodegree, eval_lalonde_cps$hisp,
                        eval_lalonde_cps$re74))
# Implement TWFE DID with panel data
twfe_did_panel(y1 = eval_lalonde_cps$re78, y0 = eval_lalonde_cps$re75,
                D = eval_lalonde_cps$experimental,
                covariates = covX)
```

**twfe_did_rc**

Two-way fixed effects DiD estimator, with repeated cross-section data

Description
twfe_did_rc is used to compute linear two-way fixed effects estimators for the ATT in difference-in-differences (DiD) setups with stationary repeated cross-sectional data. As illustrated by Sant’Anna and Zhao (2020), this estimator generally do not recover the ATT. We encourage empiricists to adopt alternative specifications.

Usage
twfe_did_rc(
  y,
  post,
  D,
  covariates = NULL,
)
i.weights = NULL,
boot = FALSE,
boot.type = "weighted",
nboot = NULL,
inffunc = FALSE
)

Arguments

y  An n x 1 vector of outcomes from the both pre and post-treatment periods.
post  An n x 1 vector of Post-Treatment dummies (post = 1 if observation belongs to post-treatment period, and post = 0 if observation belongs to pre-treatment period.)
D  An n x 1 vector of Group indicators (=1 if observation is treated in the post-treatment period, =0 otherwise).
covariates  An n x k matrix of covariates to be used in the regression estimation.
i.weights  An n x 1 vector of weights to be used. If NULL, then every observation has the same weights.
boot  Logical argument to whether bootstrap should be used for inference. Default is FALSE.
boot.type  Type of bootstrap to be performed (not relevant if boot = FALSE). Options are "weighted" and "multiplier". If boot = TRUE, default is "weighted".
nboot  Number of bootstrap repetitions (not relevant if boot = FALSE). Default is 999.
inffunc  Logical argument to whether influence function should be returned. Default is FALSE.

Value

A list containing the following components:

ATT  The TWFE DID point estimate
se  The TWFE DID standard error
uci  Estimate of the upper bound of a 95% CI for the TWFE parameter.
lci  Estimate of the lower bound of a 95% CI for the TWFE parameter.
boots  All Bootstrap draws of the ATT, in case bootstrap was used to conduct inference. Default is NULL.
att.inf.func  Estimate of the influence function. Default is NULL.

Examples

# use the simulated data provided in the package
covX = as.matrix(sim_rc[,5:8])
# Implement TWFE DID estimator (you probably should consider something else....)
twfe_did_rc(y = sim_rc$y, post = sim_rc$post, D = sim_rc$d, covariates= covX)
Index

* datasets
  nsw, 24
  nsw_long, 25
  sim_rc, 32

drdid, 3
DRDID-package, 2
drdid_imp_panel, 6
drdid_imp_rc, 8
drdid_imp_rcl, 10
drdid_panel, 12
drdid_rc, 14
drdid_rcl, 16

ipw_did_panel, 20
ipw_did_rc, 22
ipwdid, 18

nsw, 24
nsw_long, 25

ordid, 26

reg_did_panel, 28
reg_did_rc, 30

sim_rc, 32
std_ipw_did_panel, 33
std_ipw_did_rc, 35

twfe_did_panel, 37
twfe_did_rc, 38