Package ‘bate’

October 12, 2022

Title Computes Bias-Adjusted Treatment Effect

Version 0.1.0

Description Compute bounds for the treatment effect after adjusting for the presence of omitted variables in linear econometric models, according to the method of Basu (2022) <arXiv:2203.12431>. You supply the data, identify the outcome and treatment variables and additional regressors. The main functions will compute bounds for the bias-adjusted treatment effect. Many plot functions allow easy visualization of results.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.2

LazyData true

Imports ggplot2, concaveman, dplyr, stats, magrittr, tidyselect, purrr, latex2exp, vtable

Suggests rmarkdown, knitr

VignetteBuilder knitr


NeedsCompilation no

Author Deepankar Basu [aut, cre], Evan Wasner [aut]

Maintainer Deepankar Basu <dbasu@umass.edu>

Repository CRAN

Date/Publication 2022-03-28 07:30:05 UTC

R topics documented:

  collect_par .............................. 2
  cplotbias ................................ 3
collect_par

Collect parameters from the short, intermediate and auxiliary regressions

**Description**

Collect parameters from the short, intermediate and auxiliary regressions

**Usage**

```r
collect_par(data, outcome, treatment, control, other_regressors = NULL)
```

**Arguments**

- **data**
  A data frame.

- **outcome**
  The name of the outcome variable (must be present in the data frame).

- **treatment**
  The name of the treatment variable (must be present in the data frame).

- **control**
  Control variables to be added to the intermediate regression.

- **other_regressors**
  Subset of control variables to be added in the short regression (default is NULL).
Value

A data frame with the following columns:

- \( \beta_0 \)  Treatment effect in the short regression
- \( R_0 \)  R-squared in the short regression
- \( \hat{\beta} \)  Treatment effect in the intermediate regression
- \( \hat{R} \)  R-squared in the intermediate regression
- \( \sigma_y \)  Standard deviation of outcome variable
- \( \sigma_x \)  Standard deviation of treatment variable
- \( \tau_x \)  Standard deviation of residual in auxiliary regression

Examples

```r
## Load data set
data("NLSY_IQ")

## Set age and race as factor variables
NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

## Collect parameters from the short, intermediate and auxiliary regressions
parameters <- collect_par(
data = NLSY_IQ, outcome = "iq_std",
treatment = "BF_months",
control = c("age","sex","income","motherAge","motherEDU","mom_married","race"),
other_regressors = c("sex","age"))

## See results
(parameters)
```

Description

Create contour plot of bias

Usage

```r
cplotbias(data)
```

Arguments

- **data**  A data frame that is the output from the "ovbias" function.
delfplot

Plot graph of function $\delta=f(R_{\text{max}})$

Description

Plot graph of function $\delta=f(R_{\text{max}})$
Usage
delfplot(parameters)

Arguments
parameters A vector of parameters that is generated after estimating the short, intermediate and auxiliary regressions.

Value
A plot object created with ggplot

Examples
## Load data set
data("NLSY_IQ")

## Set age and race as factor variables
NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

## Collect parameters from the short, intermediate and auxiliary regressions
parameters <- collect_par(
data = NLSY_IQ, outcome = "iq_std",
treatment = "BF_months",
control = c("age","sex","income","motherAge","motherEDU","mom_married","race"),
other_regressors = c("sex","age"))

## Set limits for the bounded box
Rlow <- parameters$Rtilde
Rhigh <- 0.61
deltalow <- 0.01
deltahigh <- 0.99
e <- 0.01

## Oster's method: Plot of delta = f(Rmax)
p4 <- delfplot(parameters = parameters)
print(p4)
Arguments

data  A data frame that is the output from the "ovbias" function.

Value

A plot object created with ggplot

Examples

## Load data set
data("NLSY_IQ")

## Set age and race as factor variables
NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

## Collect parameters from the short, intermediate and auxiliary regressions
parameters <- collect_par(
data = NLSY_IQ, outcome = "iq_std",
treatment = "BF_months",
control = c("age","sex","income","motherAge","motherEDU","mom_married","race"),
other_regressors = c("sex","age"))

## Set limits for the bounded box
Rlow <- parameters$Rtilde
Rhigh <- 0.61
deltalow <- 0.01
deltahigh <- 0.99
e <- 0.01

## Not run:
## Compute bias and bias-adjusted treatment effect
OVB <- ovbias(
parameters = parameters,
deltalow=deltalow,
deltahigh=deltahigh, Rhigh=Rhigh,
e=e)

## Histogram and density Plot of bstar distribution
p3 <- dplotbate(OVB$Data)
print(p3)

## End(Not run)
**get_border**

**Description**

Extend border of bounded box by +/- e

**Usage**

`expand_border(parameters, deltalo, deltahigh, Rlo, Rhi, e)`

**Arguments**

- **parameters**: A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.
- **deltalo**: The lower limit of delta.
- **deltahigh**: The upper limit of delta.
- **Rlo**: The lower limit of Rmax.
- **Rhi**: The upper limit of Rmax.
- **e**: The step size.

**Value**

Data frame.

---

**get_border**  
*Identify all border points in a region*

**Description**

Identify all border points in a region

**Usage**

`get_border(region, e)`

**Arguments**

- **region**: A data frame containing the x and y coordinates of the region.
- **e**: The step size of the grid in the x and y directions.

**Value**

A data frame containing the x and y coordinates of the border points of the region.
mycubic  
*Compute roots of the cubic equation*

**Description**

Compute roots of the cubic equation

**Usage**

mycubic(parameters, mydelta, Rmax)

**Arguments**

- **parameters**: A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.
- **mydelta**: Value of delta (real number).
- **Rmax**: Value of Rmax (real number).

**Value**

A vector containing the three roots of the cubic equation defined by the parameters, delta and Rmax.

mydisc  
*Evaluates discriminant of the cubic equation*

**Description**

Evaluates discriminant of the cubic equation

**Usage**

mydisc(parameters, mydelta, Rmax)

**Arguments**

- **parameters**: A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.
- **mydelta**: The value of delta (real number).
- **Rmax**: The value of Rmax (real number)

**Value**

Returns a value of 0 or 1; 0 (if discriminant is positive) and 1 (if discriminant is nonpositive)
Description

NLSY data to analyse the effect of maternal behaviour on children’s birth weight. Natality detail files are from 2001 and 2002. Data is from the NLSY Children and Young Adults panel.

Usage

NLSY_BW

Format

A data frame with 7686 observations on 13 variables:

birth_wt  birth weight, grams
BF_months  months of breast feeding
mom_drink_preg_all  did the mother drink at all during pregnancy
lbw_preterm  low birth weight + preterm
age  age of child
female  child female
black  mother black
motherAge  age of mother
motherEDU  years of schooling of mother
mom_married  is the mother married?
income  annual income of mother
sex  years of schooling of mother
race  race of mother
gesweek  gestation week
any_smoke  did the mother smoke at all during pregnancy

Source: https://drive.google.com/file/d/1O1W9dP8F3B10nAZGBegpoqCfysUrn7Uc/view?usp=sharing

Examples

```r
## Load data set
data("NLSY_BW")
## See names of variables
names(NLSY_BW)
```
Description

NLSY data to analyse the effect of maternal behaviour on children’s IQ score. Natality detail files are from 2001 and 2002. Data is from the NLSY Children and Young Adults panel.

Usage

NLSY_IQ

Format

A data frame with 6514 observations on 13 variables:

- iq_std: standardized IQ score, PIAT score
- BF_months: months of breast feeding
- mom_drink_preg_all: did mother drink at all during pregnancy
- lbw_preterm: low birth weight + preterm
- age: age of child
- female: child female
- black: mother black
- motherAge: age of mother
- motherEDU: years of schooling of mother
- mom_married: is the mother married?
- income: annual income of mother
- sex: child sex
- race: race of mother

Source: https://drive.google.com/file/d/1O1W9dP8F3B1DnAZGBegpoqCfysUrn7Uc/view?usp=sharing

Examples

```r
## Load data set
data("NLSY_IQ")
## See names of variables
names(NLSY_IQ)
```
**osterbds**

*Computes identified set according to Oster (2019)*

---

**Description**

Computes identified set according to Oster (2019)

**Usage**

`osterbds(parameters, Rmax)`

**Arguments**

- `parameters`: A vector of parameters that is generated after estimating the short, intermediate and auxiliary regressions.
- `Rmax`: A real number which lies between $R_{\tilde{\text{r}}}^2$ (R-squared for the intermediate regression) and 1.

**Value**

A data frame with three columns:

- **Discriminant**: The value of the discriminant of the quadratic equation that is solved to generate the identified set
- **Interval1**: The interval formed with the first root of the quadratic equation
- **Interval2**: The interval formed with the first root of the quadratic equation

**Examples**

```r
## Load data set
data("NLSY_IQ")

## Set age and race as factor variables
NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

## Collect parameters from the short, intermediate and auxiliary regressions
parameters <- collect_par(
data = NLSY_IQ, outcome = "iq_std",
treatment = "BF_months",
control = c("age","sex","income","motherAge","motherEDU","mom_married","race"),
other_regressors = c("sex","age"))

## Oster's method: bounding sets when Rmax=0.61
osterbds(parameters = parameters, Rmax=0.61)
```
**Description**

Computes delta* according to Oster (2019)

**Usage**

```r
osterdelstar(parameters, Rmax)
```

**Arguments**

- `parameters`: A vector of parameters that is generated after estimating the short, intermediate and auxiliary regressions.
- `Rmax`: A real number that lies between Rtilde (R-squared for the intermediate regression) and 1.

**Value**

A data frame with three columns:

- `delstar`: The value of delta for the chosen value of Rmax.
- `discontinuity`: Indicates whether the point of discontinuity is within the interval formed by Rtilde and 1.
- `slope`: Slope of the function, delta=f(Rmax)

**Examples**

```r
## Load data set
data("NLSY_IQ")

## Set age and race as factor variables
NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

## Collect parameters from the short, intermediate and auxiliary regressions
parameters <- collect_par(
data = NLSY_IQ, outcome = "iq_std",
treatment = "BF_months",
control = c("age","sex","income","motherAge","motherEDU","mom_married","race"),
other_regressors = c("sex","age"))

## Oster's method: delta* (for Rmax=0.61)
osterdelstar(parameters = parameters, Rmax=0.61)
```
ovbias  

Description

Compute bias adjusted treatment effect taking parameter vector as input.

Usage

`ovbias(parameters, deltalow, deltahigh, Rhigh, e)`

Arguments

- `parameters`: A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.
- `deltalow`: The lower limit of delta.
- `deltahigh`: The upper limit of delta.
- `Rhigh`: The upper limit of Rmax.
- `e`: The step size.

Value

List with three elements:

- **Data**: Data frame containing the bias ($bias$) and bias-adjusted treatment effect ($bstar$) for each point on the grid
- **bias_Distribution**: Quantiles (2.5,5.0,50,95,97.5) of the empirical distribution of bias
- **bstar_Distribution**: Quantiles (2.5,5.0,50,95,97.5) of the empirical distribution of the bias-adjusted treatment effect

Examples

```r
## Load data set
data("NLSY_IQ")

## Set age and race as factor variables
NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

## Collect parameters from the short, intermediate and auxiliary regressions
parameters <- collect_par(
data = NLSY_IQ, outcome = "iq_std",
treatment = "BF_months",
control = c("age","sex","income","motherAge","motherEDU","mom_married","race"),
parameters)
```
other_regressors = c("sex","age"))

## Set limits for the bounded box
Rlow <- parameters$Rtilde
Rhight <- 0.61
deltalow <- 0.01
deltahigh <- 0.99
e <- 0.01

## Not run:
## Compute bias and bias-adjusted treatment effect
OVB <- ovbias(
  parameters = parameters,
  deltalow=deltalow,
  deltahigh=deltahigh, Rhight=Rhight,
  e=e)

## Default quantiles of bias
(OVB$bias_Distribution)

## Chosen quantiles of bias
quantile(OVB$Data$bias, c(0.01,0.05,0.1,0.9,0.95,0.975))

## Default quantiles of bias-adjusted treatment effect
(OVB$bstar_Distribution)

## Chosen quantiles of bias-adjusted treatment effect
quantile(OVB$Data$bstar, c(0.01,0.05,0.1,0.9,0.95,0.975))

## End(Not run)

---

ovbias_lm

*ovbias_lm*  
*Compute bias adjusted treatment effect taking three lm objects as input.*

**Description**

Compute bias adjusted treatment effect taking three lm objects as input.

**Usage**

`ovbias_lm(lm_shrt, lm_int, lm_aux, deltalow, deltahigh, Rhight, e)`

**Arguments**

- `lm_shrt` lm object corresponding to the short regression
- `lm_int` lm object corresponding to the intermediate regression
- `lm_aux` lm object corresponding to the auxiliary regression
deltalow  The lower limit of delta
deltahigh The upper limit of delta
Rhigh    The upper limit of Rmax
e        The step size

Value

List with three elements:

Data    Data frame containing the bias and bias-adjusted treatment effect for each point on the grid
bias_Distribution Quantiles (2.5,5.0,50,95,97.5) of the empirical distribution of bias
bstar_Distribution Quantiles (2.5,5.0,50,95,97.5) of the empirical distribution of the bias-adjusted treatment effect

Examples

```r
## Load data set
data("NLSY_IQ")

## Set age and race as factor variables
NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

## Short regression
reg_s <- lm(iq_std ~ BF_months + factor(age) + sex, data = NLSY_IQ)

## Intermediate regression
reg_i <- lm(iq_std ~ BF_months + factor(age) + sex + income + motherAge +
            motherEDU + mom_married + factor(race), data = NLSY_IQ)

## Auxiliary regression
reg_a <- lm(BF_months ~ factor(age) +
            sex + income + motherAge + motherEDU +
            mom_married + factor(race), data = NLSY_IQ)

## Set limits for the bounded box
Rlow <- summary(reg_i)$r.squared
Rhigh <- 0.61
deltalow <- 0.01
deltahigh <- 0.99
e <- 0.01

## Not run:
## Compute bias and bias-adjusted treatment effect
ovb_lm <- ovbias_lm(lm_shrt = reg_s, lm_int = reg_i,
```

ovbias_par

ovbias_par = reg_a, deltalow=deltalow, deltahigh=deltahigh, Rhigh=Rhigh, e=e)

## Default quantiles of bias
ovb_lm$bias_Distribution

# Default quantiles of bias-adjusted treatment effect
ovb_lm$bstar_Distribution

## End(Not run)

ovbias_par

Compute bias adjusted treatment effect taking data frame as input.

Description

Compute bias adjusted treatment effect taking data frame as input.

Usage

ovbias_par(
  data, outcome, treatment, control, other_regressors = NULL, deltalow, deltahigh, Rhigh, e
)

Arguments

data Data frame.
outcome Outcome variable.
treatment Treatment variable.
control Control variables to add in the intermediate regression.
other_regressors Subset of control variables to add in the short regression (default is NULL).
deltalow The lower limit of delta.
deltahigh The upper limit of delta.
Rhigh The upper limit of Rmax.
e The step size.
Value

List with three elements:

Data Data frame containing the bias and bias-adjusted treatment effect for each point on the grid
bias_Distribution Quantiles (2.5,5.0,50,95,97.5) of the empirical distribution of bias
bstar_Distribution Quantiles (2.5,5.0,50,95,97.5) of the empirical distribution of the bias-adjusted treatment effect

Examples

## Load data set
data("NLSY_IQ")

## Set parameters for bounded box
Rhigh <- 0.61
deltalow <- 0.01
deltahigh <- 0.99
e <- 0.01

## Not run:
## Compute bias and bias-adjusted treatment effect
OVB_par <- ovbias_par(data=NLSY_IQ,
                      outcome="iq_std", treatment="BF_months",
                      control=c("age","sex","income","motherAge","motherEDU","mom_married","race"),
                      other_regressors = c("sex","age"), deltalow=deltalow,
                      deltahigh=deltahigh, Rhigh=Rhigh, e=e)

## Default quantiles of bias
OVB_par$bias_Distribution

# Default quantiles of bias-adjusted treatment effect
OVB_par$bstar_Distribution

## End(Not run)

partocoef Returns coefficients of the cubic equation

Description

Returns coefficients of the cubic equation

Usage

partocoef(parameters, mydelta, Rmax)
selectroot

Arguments

parameters A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.

mydelta The value of delta (real number).

Rmax The value of Rmax (real number).

Value

A data frame with the coefficients of the cubic equation.

Description

Select root of the cubic based on the root of a nearest point

Usage

selectroot(parameters, mydelta, Rmax, closest_bias)

Arguments

parameters A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.

mydelta The value of delta (real number).

Rmax The value of Rmax (real number).

closest_bias The value of bias at the nearest point.

Value

Data frame
split_nurr

**Split a region into two parts**

**Description**
Split a region into two parts

**Usage**

```
split_nurr(region1, region2, epsilon, parameters, e)
```

**Arguments**
- `region1`: Data frame with coordinates for region 1
- `region2`: Data frame with coordinates for region 2
- `epsilon`: Closest distance
- `parameters`: A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.
- `e`: The step size of the grid in the x and y directions.

**Value**
List, where first element is region within epsilon distance of region 1 and second element which is region which is not within epsilon distance of region 1.

urrplot

**Region plot to demarcate URR and NURR for the bounded box**

**Description**
Region plot to demarcate URR and NURR for the bounded box

**Usage**

```
urrplot(parameters, deltalow, deltahigh, Rlow, Rhigh, e)
```

**Arguments**
- `parameters`: A vector of parameters (real numbers) that is generated by estimating the short, intermediate and auxiliary regressions.
- `deltalow`: The lower limit for delta.
- `deltahigh`: The upper limit for delta.
- `Rlow`: The lower limit for Rmax.
- `Rhigh`: The upper limit for Rmax.
- `e`: The step size of the grid in the x and y directions.
Value

A plot object created by ggplot

Examples

```r
data("NLSY_IQ")

NLSY_IQ$age <- factor(NLSY_IQ$age)
NLSY_IQ$race <- factor(NLSY_IQ$race)

parameters <- collect_par(
    data = NLSY_IQ, outcome = "iq_std",
    treatment = "BF_months",
    control = c("age","sex","income","motherAge","motherEDU","mom_married","race"),
    other_regressors = c("sex","age"))

Rlow <- parameters$Rtilde
Rhigh <- 0.61
deltalow <- 0.01
deltahigh <- 0.99
e <- 0.01

p1 <- urrplot(parameters, deltalow, deltahigh, Rlow, Rhigh, e=e)

print(p1)
```
Index

* datasets
  NLSY_BW, 9
  NLSY_IQ, 10

collect_par, 2
clplotbias, 3

delfplot, 4
dplotbate, 5

expand_border, 6
get_border, 7

mycubic, 8
mydisc, 8

NLSY_BW, 9
NLSY_IQ, 10

osterbds, 11
osterdelstar, 12
ovbias, 13
ovbias_lm, 14
ovbias_par, 16

partocoeff, 17

selectroot, 18
split_nurr, 19

urrrplot, 19