Package ‘bndovb’

July 30, 2021

Title  Bounding Omitted Variable Bias Using Auxiliary Data
Version  1.1
Description  Functions to implement a Hwang(2021) \<doi:10.2139/ssrn.3866876>\ estimator, which bounds an omitted variable bias using auxiliary data.
License  GPL-3
Encoding  UTF-8
LazyData  true
Depends  R (>= 2.10)
RoxygenNote  7.1.1
Imports  np, pracma, stats, utils, MASS, dplyr, factormodel, nnet
Suggests  knitr, rmarkdown
VignetteBuilder  knitr
NeedsCompilation  no
Author  Yujung Hwang [aut, cre] (<https://orcid.org/0000-0002-8136-8987>)
Maintainer  Yujung Hwang <yujungghwang@gmail.com>
Repository  CRAN
Date/Publication  2021-07-30 17:40:02 UTC

R topics documented:

auxdat_mecont .................................................. 2
auxdat_medisc .................................................. 2
auxdat_nome .................................................. 3
bndovb ......................................................... 3
bndovbme ....................................................... 5
maindat_mecont ................................................ 7
maindat_medisc ............................................... 7
maindat_nome ................................................ 8

Index  9
A simulated auxiliary data to show how to use 'bndovbme' function

with continuous proxy variables

Description

A simulated auxiliary data to show how to use 'bndovbme' function with continuous proxy variables

Usage

auxdat_medisc

Format

A data frame with 3000 rows and 5 variables:

- **w1**: A common covariate in both main and auxiliary data
- **x**: A common covariate in both main and auxiliary data
- **z1**: A continuous proxy variable
- **z2**: A continuous proxy variable
- **z3**: A continuous proxy variable

Source

This dataset was simulated by simulatePackageData.R in data-raw folder

A simulated auxiliary data to show how to use 'bndovbme' function

with discrete proxy variables

Description

A simulated auxiliary data to show how to use 'bndovbme' function with discrete proxy variables

Usage

auxdat_medisc

Format

A data frame with 3000 rows and 5 variables:

- **w1**: A common covariate in both main and auxiliary data
- **x**: A common covariate in both main and auxiliary data
- **z1**: A discrete proxy variable
- **z2**: A discrete proxy variable
- **z3**: A discrete proxy variable
**Source**

This dataset was simulated by simulatePackageData.R in data-raw folder

---

### auxdat_nome

**A simulated auxiliary data to show how to use 'bndovb' function**

**Description**

A simulated auxiliary data to show how to use 'bndovb' function

**Usage**

auxdat_nome

**Format**

A data frame with 50000 rows and 3 variables:

- **x1** An omitted variable in the main data
- **x2** A common covariate in both main and auxiliary data
- **x3** A common covariate in both main and auxiliary data

**Source**

This dataset was simulated by simulatePackageData.R in data-raw folder

---

### bndovb

**bndovb**

**Description**

This function runs a two sample least squares when auxiliary data contains every right-hand side regressor and main data contains a dependent variable and every right-hand side regressor but one omitted variable.

**Usage**

bndovb(
    maindat,
    auxdat,
    depvar,
    ovar,
    comvar,
    method = 1,
    mainweights = NULL,
    auxweights = NULL,
    signres = NULL
)
**Arguments**

- **maindat**: Main data set. It must be a data frame.
- **auxdat**: Auxiliary data set. It must be a data frame.
- **depvar**: A name of a dependent variable in main dataset
- **ovar**: A name of an omitted variable in main dataset which exists in auxiliary data
- **comvar**: A vector of the names of common regressors existing in both main data and auxiliary data
- **method**: CDF and Quantile function estimation method. Users can choose either 1 or 2. If the method is 1, the CDF and quantile function is estimated assuming a parametric normal distribution. If the method is 2, the CDF and quantile function is estimated using a nonparametric estimator in Li and Racine(2008) doi: 10.1198/073500107000000250, Li, Lin, and Racine(2013) doi: 10.1080/07350015.2012.738955. Default is 1.
- **mainweights**: An optional weight vector for the main dataset. The length must be equal to the number of rows of 'maindat'.
- **auxweights**: An optional weight vector for the auxiliary dataset. The length must be equal to the number of rows of 'auxdat'.
- **signres**: An option to impose a sign restriction on a coefficient of an omitted variable. Set either NULL or pos or neg. Default is NULL. If NULL, there is no sign restriction. If 'pos', the estimator imposes an extra restriction that the coefficient of an omitted variable must be positive. If 'neg', the estimator imposes an extra restriction that the coefficient of an omitted variable must be negative.

**Value**

Returns a list of 4 components:

- **hat_beta_l**: lower bound estimates of regression coefficients
- **hat_beta_u**: upper bound estimates of regression coefficients
- **mu_l**: lower bound estimate of E[ovar*depvar]
- **mu_u**: upper bound estimate of E[ovar*depvar]

**Author(s)**

Yujung Hwang, <yujungghwang@gmail.com>

**References**

**Examples**

```
data(maindat_nome)
data(auxdat_nome)

bndovb(maindat=maindat_nome, auxdat=auxdat_nome, depvar="y", ovar="x1", comvar=c("x2","x3"), method=1)
```
ptype
Either 1 (continuous) or 2 (discrete). Whether proxy variables are continuous or
discrete. Default is 1 (continuous).

comvar
A vector of the names of the common regressors existing in both main data and
auxiliary data

sbar
A cardinality of the support of the discrete proxy variables. Default is 2. If
proxy variables are continuous, this variable is irrelevant.

mainweights
An optional weight vector for the main dataset. The length must be equal to the
number of rows of 'maindat'.

auxweights
An optional weight vector for the auxiliary dataset. The length must be equal to
the number of rows of 'auxdat'.

normalize
Whether to normalize the omitted variable to have mean 0 and standard devia-
tion 1. Set TRUE or FALSE. Default is TRUE. If FALSE, then the scale of the
omitted variable is anchored with the first proxy variable in pvar list.

signres
An option to impose a sign restriction on a coefficient of an omitted variable.
Set either NULL or pos or neg. Default is NULL. If NULL, there is no sign
restriction. If 'pos', the estimator imposes an extra restriction that the coefficient
of an omitted variable must be positive. If 'neg', the estimator imposes an extra
restriction that the coefficient of an omitted variable must be negative.

Value
Returns a list of 4 components :

hat_beta_l lower bound estimates of regression coefficients
hat_beta_u upper bound estimates of regression coefficients
mu_l lower bound estimate of \( E[ovar*depvar] \)
mu_u upper bound estimate of \( E[ovar*depvar] \)

Author(s)
Yujung Hwang, <yujungghwang@gmail.com>

References
Hwang, Yujung (2021) Bounding Omitted Variable Bias Using Auxiliary Data. Available at SSRN.
doi: 10.2139/ssrn.3866876

Examples
## load example data
data(maindat_mecont)
data(auxdat_mecont)

## set ptype=1 for continuous proxy variables
pvar<-c("z1", "z2", "z3")
cvar<-c("x", "w1")
bndovbme(maindat=maindat_mecont, auxdat=auxdat_mecont, depvar="y", pvar=pvar, ptype=1, comvar=cvar)
## set ptype=2 for discrete proxy variables

```r
data(maindat_medisc)
data(auxdat_medisc)
bndovbme(maindat=maindat_medisc, auxdat=auxdat_medisc, depvar="y", pvar=pvar, ptype=2, comvar=cvar)
```

---

### maindat_mecont

A simulated main data to show how to use 'bndovbme' function with continuous proxy variables

#### Description

A simulated main data to show how to use 'bndovbme' function with continuous proxy variables

#### Usage

```r
maindat_mecont
```

#### Format

A data frame with 3000 rows and 3 variables:

- **w1**: A common covariate in both main and auxiliary data
- **x**: A common covariate in both main and auxiliary data
- **y**: A dependent variable

#### Source

This dataset was simulated by simulatePackageData.R in data-raw folder

---

### maindat_medisc

A simulated main data to show how to use 'bndovbme' function with discrete proxy variables

#### Description

A simulated main data to show how to use 'bndovbme' function with discrete proxy variables

#### Usage

```r
maindat_medisc
```
**Format**

A data frame with 3000 rows and 3 variables:

- **w1** A common covariate in both main and auxiliary data
- **x** A common covariate in both main and auxiliary data
- **y** A dependent variable

**Source**

This dataset was simulated by simulatePackageData.R in data-raw folder

---

**maindat_nome**

A simulated main data to show how to use 'bndovb' function

**Description**

A simulated main data to show how to use 'bndovb' function

**Usage**

`maindat_nome`

**Format**

A data frame with 100000 rows and 3 variables:

- **x2** A common covariate in both main and auxiliary data
- **x3** A common covariate in both main and auxiliary data
- **y** A dependent variable

**Source**

This dataset was simulated by simulatePackageData.R in data-raw folder
Index

* datasets
  auxdat_mecont, 2
  auxdat_medisc, 2
  auxdat_nome, 3
  maindat_mecont, 7
  maindat_medisc, 7
  maindat_nome, 8

auxdat_mecont, 2
auxdat_medisc, 2
auxdat_nome, 3

bndovb, 3
bndovbme, 5

maindat_mecont, 7
maindat_medisc, 7
maindat_nome, 8