Package: bndovb (via r-universe)

August 23, 2024

Title Bounding Omitted Variable Bias Using Auxiliary Data

Version 1.2

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)

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Imports np, pracma, stats, utils, MASS, dplyr, factormodel, nnet,
doParallel, foreach, parallel

Suggests knitr, rmarkdown

VignetteBuilder knitr

Repository https://yujunghwang.r-universe.dev

RemoteUrl https://github.com/yujunghwang/bndovb

RemoteRef HEAD

RemoteSha dc4479bacc5e598282233f51c7c04670ccaba530

Contents

---

auxdat_mecont | *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_mecont
```

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

---

auxdat_medisc | *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,
  ci = FALSE,
  nboot = 100,
  scale = -1/2,
  tau = 0.05,
  seed = 210823,
  display = TRUE
)
```

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 nonparaemtric 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.
ci	An option to compute an equal-tailed confidence interval. Default is FALSE. It may take some time to compute CI from bootstrap.
nboot	Number of bootstraps to compute the confidence interval. Default is 100.
scale	A tuning parameter for rescaled numerical bootstrap. The value must be between -1/2 and 0. (main data sample size)^scale is the tuning parameter epsilon_n in Hwang (2021). Default is -1/2 (that is, standard bootstrap).
tau	Significance level. (1-tau)% confidence interval is computed. Default is 0.05.
seed	Seed for random number generation. Default is 210823.
display	It must be either TRUE or FALSE. Whether to display progress and messages. Default is TRUE.

Value

Returns a list of 12 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]

hat_beta_l_cil (1-tau)% confidence interval lower bound for hat_beta_l

hat_beta_l_ciu (1-tau)% confidence interval upper bound for hat_beta_l

hat_beta_u_cil (1-tau)% confidence interval lower bound for hat_beta_u

hat_beta_u_ciu (1-tau)% confidence interval upper bound for hat_beta_u

mu_l_cil (1-tau)% confidence interval lower bound for mu_l

mu_l_ciu (1-tau)% confidence interval upper bound for mu_l

mu_u_cil (1-tau)% confidence interval lower bound for mu_u

mu_u_ciu (1-tau)% confidence interval upper bound for mu_u

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

```
data(maindat_nome)
data(auxdat_nome)

bndovb(maindat=maindat_nome,auxdat=auxdat_nome,depvar="y",ovar="x1",comvar=c("x2","x3"),method=1)
```

---

bndovbme   *bndovbme*

Description

This function runs a two sample least squares when main data contains a dependent variable and
every right hand side regressor but one omitted variable. The function requires an auxiliary data
which includes every right hand side regressor but one omitted variable, and enough proxy variables
for the omitted variable. When the omitted variable is continuous, the auxiliary data must contain
at least two continuous proxy variables. When the omitted variable is discrete, the auxiliary data
must contain at least three continuous proxy variables.

Usage

```
bndovbme(
  maindat,
  auxdat,
  depvar,
  pvar,
  ptype = 1,
  comvar,
  sbar = 2,
  mainweights = NULL,
  auxweights = NULL,
  normalize = TRUE,
  signres = NULL,
  ci = FALSE,
  nboot = 100,
  scale = -1/2,
  tau = 0.05,
  seed = 210823,
  display = TRUE
)
```

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

pvar	A vector of the names of the proxy variables for the omitted variable. When proxy variables are continuous, the first proxy variable is used as an anchoring variable. When proxy variables are discrete, the first proxy variable is used for initialization (For details, see a documentation for "dproxyme" function).
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 deviation 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.
ci	An option to compute an equal-tailed confidence interval. Default is FALSE. It may take some time to compute CI from bootstrap.
nboot	Number of bootstraps to compute the confidence interval. Default is 100.
scale	A tuning parameter for rescaled numerical bootstrap. The value must be between -1/2 and 0. (main data sample size)^scale is the tuning parameter epsilon_n in Hwang (2021). Default is -1/2 (that is, standard bootstrap).
tau	Significance level. (1-tau)% confidence interval is computed. Default is 0.05.
seed	Seed for random number generation. Default is 210823.
display	It must be either TRUE or FALSE. Whether to display progress and messages. Default is TRUE.

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]

hat_beta_l_cil (1-tau)% confidence interval lower bound for hat_beta_l

hat_beta_l_ciu (1-tau)% confidence interval upper bound for hat_beta_l

hat_beta_u_cil (1-tau)% confidence interval lower bound for hat_beta_u

hat_beta_u_ciu (1-tau)% confidence interval upper bound for hat_beta_u

mu_l_cil (1-tau)% confidence interval lower bound for mu_l

mu_l_ciu (1-tau)% confidence interval upper bound for mu_l

mu_u_cil (1-tau)% confidence interval lower bound for mu_u

mu_u_ciu (1-tau)% confidence interval upper bound for mu_u

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

---

bndovbme_tuning　　　*bndovbme_tuning*

Description

This function computes an optimal tuning parameter to compute the confidence interval for bndovbme function The function returns an optimal tuning parameter using double bootstrap procedure

Usage

```
bndovbme_tuning(
  maindat,
  auxdat,
  depvar,
  pvar,
  ptype = 1,
  comvar,
  sbar = 2,
  mainweights = NULL,
  auxweights = NULL,
  normalize = TRUE,
  signres = NULL,
  nboot = 100,
  scalegrid = c(-1/2, -1/3, -1/4, -1/5, -1/6),
  tau = 0.05,
  seed = 210823,
  parallel = TRUE
)
```

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
pvar	A vector of the names of the proxy variables for the omitted variable. When proxy variables are continuous, the first proxy variable is used as an anchoring variable. When proxy variables are discrete, the first proxy variable is used for initialization (For details, see a documentation for "dproxyme" function).
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 deviation 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.

nboot   Number of bootstraps to compute the confidence interval. Default is 100.

scalegrid   Tuning parameter grid to search. It must be a vector of numbers between -1/2 and 0. Default is c(-1/2,-1/3,-1/4,-1/5,-1/6).

tau   Significance level. (1-tau)% confidence interval is computed. Default is 0.05.

seed   Seed for random number generation. Default is 210823.

parallel   Either TRUE or FALSE. Whether to compute in parallel. Default is TRUE.

Value

Returns a list of 3 components :

optimal_scale An optimal scale parameter which gives coverage rates closest to (1-tau)

cover_beta_l A matrix of coverage rates of the lower bound parameters under different scale parameters

cover_beta_u A matrix of coverage rates of the lower bound parameters under different scale parameters

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")

# To shorten computation time, I set the number of bootstrap small in an example below.
# In practice, please set it a large number
bndovbme_tuning(maindat_mecont,auxdat_mecont,depvar="y",pvar=pvar,ptype=1,comvar=cvar,nboot=2)
```

---

bndovb_tuning *bndovb_tuning*

Description

This function computes an optimal tuning parameter to compute the confidence interval for bndovb function The function returns an optimal tuning parameter using double bootstrap procedure

Usage

```
bndovb_tuning(
  maindat,
  auxdat,
  depvar,
  ovar,
  comvar,
  method = 1,
  mainweights = NULL,
  auxweights = NULL,
  signres = NULL,
  nboot = 100,
  scalegrid = c(-1/2, -1/3, -1/4, -1/5, -1/6),
  tau = 0.05,
  seed = 210823,
  parallel = TRUE
)
```

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 nonparaemtric 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.
nboot	Number of bootstraps to compute the confidence interval. Default is 100.
scalegrid	Tuning parameter grid to search. It must be a vector of numbers between -1/2 and 0. Default is c(-1/2,-1/3,-1/4,-1/5,-1/6).
tau	Significance level. (1-tau)% confidence interval is computed. Default is 0.05.
seed	Seed for random number generation. Default is 210823.
parallel	Either TRUE or FALSE. Whether to compute in parallel. Default is TRUE.

Value

Returns a list of 3 components :

optimal_scale An optimal scale parameter which gives coverage rates closest to (1-tau)

cover_beta_l A matrix of coverage rates of the lower bound parameters under different scale parameters

cover_beta_u A matrix of coverage rates of the lower bound parameters under different scale parameters

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

```
data(maindat_nome)
data(auxdat_nome)

# To shorten computation time, I set the number of bootstrap small in an example below.
# In practice, please set it a large number
bndovb_tuning(maindat_nome,auxdat_nome,depvar="y",ovar="x1",comvar=c("x2","x3"),method=1,nboot=2)
```

---

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

```
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

```
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