Package: RegCombin (via r-universe)

November 3, 2026

Title Partially Linear Regression under Data Combination

Version 0.4.1

Description We implement linear regression when the outcome of interest and some of the covariates are observed in two different datasets that cannot be linked, based on D'Haultfoeuille, Gaillac, Maurel (2022) <doi:10.3386/w29953>. The package allows for common regressors observed in both datasets, and for various shape constraints on the effect of covariates on the outcome of interest. It also provides the tools to perform a test of point identification. See the associated vignette <https://github.com/cgaillac/RegCombin/blob/master/RegCombin_vignette.pdf> for theory and code examples.

License GPL-3

Encoding UTF-8

Suggests knitr, rmarkdown

RoxygenNote 7.2.3

Imports dplyr,kableExtra,snowfall,RationalExp,Hmisc,geometry,pracma

NeedsCompilation no

Author Xavier D'Haultfoeuille [aut], Christophe Gaillac [aut, cre], Arnaud Maurel [aut]

Maintainer Christophe Gaillac <christophe.gaillac@economics.ox.ac.uk>

Config/pak/sysreqs cmake libfontconfig1-dev libfreetype6-dev libfribidi-dev make libharfbuzz-dev libicu-dev libpng-dev libuv1-dev libxml2-dev

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

Date/Publication 2026-11-03 06:02:00 UTC

RemoteUrl https://github.com/cranhaven/cranhaven.r-universe.dev

RemoteRef package/RegCombin

RemoteSha e3b03a57f0d85a1cfbae0f027e12709336feb36c

RemoteSubdir RegCombin

Contents

AS_bounds . . . 2
AStest . . . 3
compute_bnds_betac . . . 4
compute_constraints . . . 5
compute_radial . . . 6
compute_radial_test . . . 8
compute_ratio . . . 10
compute_ratio_variance . . . 11
compute_stat_variance . . . 12
compute_support . . . 13
compute_support_paral . . . 16
create_values . . . 18
DGM_bounds . . . 18
DGM_bounds_test . . . 21
ewcdf . . . 24
objective_support . . . 25
point_ident_test . . . 27
regCombin . . . 29
regCombin_profile . . . 32
sampling_rule . . . 34
select_epsilon . . . 34
select_epsilon_test . . . 36
summary_regCombin . . . 38
tabulate_values . . . 39
Variance_bounds . . . 40

Index **42**

`AS_bounds` *This function finds the boundary of the identified set in one specified direction using the AS test and Newton's method.*

Description

This function finds the boundary of the identified set in one specified direction using the AS test and Newton's method.

Usage

```
AS_bounds(start, Yp, Xb, N_max = 30, tol = 10^(-4), tuningParam = NULL)
```

Arguments

start	the starting points for the bissection method
Yp	the observations of the outcome variable.
Xb	the observations of the noncommon regressor (possibly conditional on Xc).
N_max	the maximal number of iterations. Default is 30.
tol	the tolerance of the method. Default is e-4.
tuningParam	the list of tuning parameters. For the details see the function "test" in the package RationalExp.

Value

a list containing, in order: - the value of estimated radial function in this direction - value of the objective function - the number of iterations

---

AStest — *This function computes the AS test using DGM implementation in the package RationalExp*

Description

This function computes the AS test using DGM implementation in the package RationalExp

Usage

```
AStest(lamb, YY, XX, tuningParam = NULL)
```

Arguments

lamb	the point under the form lambda q to be tested.
YY	the observations of the outcome variable.
XX	the observations of the regressor X'q variable.
tuningParam	the list of tuning parameters. For the details see the function "test" in the package RationalExp.

Value

the result of the test at level 5

compute_bnds_betac *Function to compute the bounds on the coefficients of the common regressors.*

Description

Function to compute the bounds on the coefficients of the common regressors.

Usage

```
compute_bnds_betac(
  sample1 = NULL,
  info0,
  values,
  constraint = NULL,
  c_sign0,
  nc_sign0,
  refs0,
  c_var,
  nc_var,
  sam0,
  info1 = NULL,
  constr = TRUE,
  R2bound = NULL,
  values_sel = NULL
)
```

Arguments

sample1	if NULL compute the point estimate, if a natural number then evaluate a bootstrap or subsampling replication.
info0	the results of the estimates (point and bootstrap/subsampling replications) for betanc. No default.
values	the different unique points of support of the common regressor Xc.
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.#'
c_sign0	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign0	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
refs0	indicating the positions in the vector values corresponding to the components of betac.

c_var	label of the commonly observed regressors Xc.
nc_var	label of the non commonly observed regressors Xnc.
sam0	the directions q where the radial function has been computed.
info1	the results of the point estimates for betac. Default is NULL.
constr	if sign constraints imposed. Default is TRUE.
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.

Value

a matrix containing the bounds on the coefficients associated to the common regressor.

`compute_constraints` *Compute the indexes of the values of the common regressors Xc used in the various shape constraints*

Description

Compute the indexes of the values of the common regressors Xc used in the various shape constraints

Usage

```
compute_constraints(
  constraint,
  values,
  values_sel,
  indexes_k = NULL,
  nbV,
  grouped0,
  ind = NULL,
  c_sign = NULL
)
```

Arguments

constraint	the current shape constraint
values	the different unique points of support of the common regressor Xc.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
indexes_k	indexes of the constraints
nbV	indexes of the constraints
grouped0	boolean indexing if the values of Xc have been changed
ind	index
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.

Value

a vector containing:

- the matrix R where each line is a constraint

- the matrices pp0 and pp1, which contains the indexes of the values of Xc in values_sel which enters the various constraints.

`compute_radial` *Function to compute the DGM bounds on the noncommon regressor Xnc*

Description

Function to compute the DGM bounds on the noncommon regressor Xnc

Usage

```
compute_radial(
  sample1 = NULL,
  Xc_x,
  Xnc,
  Xc_y,
  Y,
  values,
  dimXc,
  dimXnc,
  nb_pts,
  sam0,
  eps_default0,
  grid = NULL,
  lim = 10,
  weights_x = NULL,
  weights_y = NULL,
  constraint = NULL,
  c_sign = NULL,
  nc_sign = NULL,
  refs0 = NULL,
  type = "both",
  meth = "adapt",
  version = "first",
  R2bound = NULL,
  values_sel = NULL,
  ties = FALSE,
  modeNA = FALSE
)
```

Arguments

sample1	if NULL compute the point estimate, if a natural number then evaluate a bootstrap or subsampling replication.
Xc_x	the common regressor on the dataset (Xnc,Xc). Default is NULL.
Xnc	the noncommon regressor on the dataset (Xnc,Xc). No default.
Xc_y	the common regressor on the dataset (Y,Xc). Default is NULL.
Y	the outcome variable. No default.
values	the different unique points of support of the common regressor Xc.
dimXc	the dimension of the common regressors Xc.
dimXnc	the dimension of the noncommon regressors Xnc.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
sam0	the directions q to compute the variance bounds on the radial function.
eps_default0	the matrix containing the directions q and the selected epsilon(q).
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to kp.
lim	the limit number of observations under which we do no compute the conditional variance.
weights_x	the sampling weights for the dataset (Xnc,Xc).
weights_y	the sampling weights for the dataset (Y,Xc).
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no contraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no contraints.
refs0	indicating the positions in the vector values corresponding to the components of betac.
type	equal to "both", "up", or "low".
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
version	version of the computation of the ratio, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.
modeNA	indicates if NA introduced if the interval is empty. Default is FALSE.

Value

a list containing:

- upper: the upper bound in the specified directions, possibly with sign constraints

- lower: the lower bound in the specified directions, possibly with sign constraints

- unconstr: the bounds without sign constraints in the specified directions

* If common regressors, upper_agg, lower_agg, and unconstr_agg reports the same values but aggregated over the values of Xc (see the parameter theta0 in the paper)

- Ykmean: the means of Y|Xc for the considered sample

- Xkmean: the means of Xnc|Xc for the considered sample

- DYk: the difference of means of Y|Xc =k - Y|Xc =0 for the considered sample

- DXk: the difference of means of Xnc|Xc =k - Xnc|Xc =0 for the considered sample

- tests: the pvalues of the tests H0 : DXk =0

- ratio_ref: the ratio R in the radial function computed for the initial sample

---

`compute_radial_test` *Function to compute the DGM bounds on the noncommon regressor Xnc, adapted to the point identification test.*

Description

Function to compute the DGM bounds on the noncommon regressor Xnc, adapted to the point identification test.

Usage

```
compute_radial_test(
  sample1 = NULL,
  Xc_x,
  Xnc,
  Xc_y,
  Y,
  values,
  dimXc,
  dimXnc,
  nb_pts,
  sam0,
  eps_default0,
  grid = NULL,
  lim = 10,
  weights_x = NULL,
  weights_y = NULL,
  constraint = NULL,
  c_sign = NULL,
```

```
  nc_sign = NULL,
  refs0 = NULL,
  type = "both",
  meth = "adapt",
  version = "first",
  R2bound = NULL,
  values_sel = NULL,
  ties = FALSE
)
```

Arguments

sample1	if NULL compute the point estimate, if a natural number then evaluate a bootstrap or subsampling replication.
Xc_x	the common regressor on the dataset (Xnc,Xc). Default is NULL.
Xnc	the noncommon regressor on the dataset (Xnc,Xc). No default.
Xc_y	the common regressor on the dataset (Y,Xc). Default is NULL.
Y	the outcome variable. No default.
values	the different unique points of support of the common regressor Xc.
dimXc	the dimension of the common regressors Xc.
dimXnc	the dimension of the noncommon regressors Xnc.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
sam0	the directions q to compute the variance bounds on the radial function.
eps_default0	the matrix containing the directions q and the selected epsilon(q).
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to kp.
lim	the limit number of observations under which we do no compute the conditional variance.
weights_x	the sampling weights for the dataset (Xnc,Xc).
weights_y	the sampling weights for the dataset (Y,Xc).
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no contraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no contraints.
refs0	indicating the positions in the vector values corresponding to the components of betac.

type	equal to "both", "up", or "low".
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
version	version of the computation of the ratio, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.

Value

a list contaning:

- upper: the upper bound in the specified directions, possibly with sign constraints
- lower: the lower bound in the specified directions, possibly with sign constraints
- unconstr: the bounds without sign constraints in the specified directions
- Ykmean: the means of Y|Xc for the considered sample
- Xkmean: the means of Xnc|Xc for the considered sample
- DYk: the difference of means of Y|Xc =k - Y|Xc =0 for the considered sample
- DXk: the difference of means of Xnc|Xc =k - Xnc|Xc =0 for the considered sample
- tests: the pvalues of the tests H0 : DXk =0
- ratio_ref: the ratio R in the radial function computed for the initial sample

---

compute_ratio *Function to compute the main statistic for the point estimate*

Description

Function to compute the main statistic for the point estimate

Usage

```
compute_ratio(
  x_eps0,
  Xp,
  Yp,
  for_critY,
  dimXnc,
  weights_xp,
  weights_yp,
  version = "first",
  grid_I = NULL,
  ties = FALSE
)
```

Arguments

x_eps0	a matrix containing the directions to compute the radial function, and the associated choice epsilon(q).
Xp	the observations of the noncommon regressor (possibly conditional on Xc).
Yp	the observations of the outcome variable.
for_critY	the numerator of the ratio R for the point estimate of the radial function, on the grid grid_I;
dimXnc	the dimension of the noncommon regressors
weights_xp	the sampling or bootstrap weights for the dataset (Xnc,Xc).
weights_yp	the sampling or bootstrap weights for the dataset (Y,Xc).
version	version of the computation of the ratio, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
grid_I	the grid of alpha on which we evaluate the ratio R to compute the point estimate of the radial function.
ties	binary value handling the ties, default is FALSE.

Value

the value of the point estimate of the radial function using the DGM method.

---

compute_ratio_variance

Function to compute the variance bounds

Description

Function to compute the variance bounds

Usage

```
compute_ratio_variance(x, Xp, Yp, dimX2, weights_xp, weights_yp)
```

Arguments

x	a matrix containing the directions to compute the variance bounds on the radial function.
Xp	the observations of the noncommon regressor (possibly conditional on Xc).
Yp	the observations of the outcome variable.
dimX2	the dimension of the noncommon regressors Xnc.
weights_xp	the sampling or bootstrap weights for the dataset (Xnc,Xc).
weights_yp	the sampling or bootstrap weights for the dataset (Y,Xc).

Value

the value of the ratio of the variance entering the variance bounds.

`compute_stat_variance` *Function to compute the Variance bounds on the noncommon regressor Xnc*

Description

Function to compute the Variance bounds on the noncommon regressor Xnc

Usage

```
compute_stat_variance(
  sample1 = NULL,
  X1_x,
  X2,
  X1_y,
  Y,
  values,
  refs0,
  dimX1,
  dimX2,
  nb_pts,
  sam0,
  lim = 1,
  weights_x = NULL,
  weights_y = NULL,
  constraint = NULL,
  c_sign = NULL,
  nc_sign = NULL,
  values_sel = NULL
)
```

Arguments

`sample1`	if NULL compute the point estimate, if a natural number then evaluate a bootstrap or subsampling replication.
`X1_x`	the common regressor on the dataset (Xnc,Xc). Default is NULL.
`X2`	the noncommon regressor on the dataset (Xnc,Xc). No default.
`X1_y`	the common regressor on the dataset (Y,Xc). Default is NULL.
`Y`	the outcome variable. No default.
`values`	the different unique points of support of the common regressor Xc.
`refs0`	indicating the positions in the vector values corresponding to the components of betac.
`dimX1`	the dimension of the common regressors Xc.
`dimX2`	the dimension of the noncommon regressors Xnc.

nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
sam0	the directions q to compute the variance bounds on the radial function.
lim	the limit number of observations under which we do no compute the conditional variance.
weights_x	the sampling weights for the dataset (Xnc,Xc).
weights_y	the sampling weights for the dataset (Y,Xc).
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
values_sel	the selected values of Xc for the conditioning. Default is NULL.

Value

a list containing:

- upper: the upper bound in the specified directions, possibly with sign constraints

- lower: the lower bound in the specified directions, possibly with sign constraints

- unconstr: the bounds without sign constraints in the specified directions

- Ykmean: the means of Y|Xc for the considered sample

- Xkmean: the means of Xnc|Xc for the considered sample

- DYk: the difference of means of Y|Xc =k - Y|Xc =0 for the considered sample

- DXk: the difference of means of Xnc|Xc =k - Xnc|Xc =0 for the considered sample

- tests: the pvalues of the tests H0 : DXk =0

- ratio_ref: the ratio R in the radial function computed for the initial sample

---

compute_support *Compute the support function for the projections of the identified set*

Description

Compute the support function for the projections of the identified set

Usage

```
compute_support(
  sample1 = NULL,
  Xc_x,
  Xnc,
  Xc_y,
  Y,
  values,
  dimXc,
  dimXnc,
  nb_pts,
  sam0,
  eps_default0,
  grid,
  lim = 30,
  weights_x = NULL,
  weights_y = NULL,
  constraint = NULL,
  c_sign = NULL,
  nc_sign = NULL,
  refs0 = NULL,
  type = "both",
  meth = "adapt",
  bc = FALSE,
  version = "first",
  R2bound = NULL,
  values_sel = NULL,
  ties = FALSE,
  modeNA = FALSE
)
```

Arguments

sample1	if NULL compute the point estimate, if a natural number then evaluate a bootstrap or subsampling replication.
Xc_x	the common regressor on the dataset (Xnc,Xc). Default is NULL.
Xnc	the noncommon regressor on the dataset (Xnc,Xc). No default.
Xc_y	the common regressor on the dataset (Y,Xc). Default is NULL.
Y	the outcome variable. No default.
values	the different unique points of support of the common regressor Xc.
dimXc	the dimension of the common regressors Xc.
dimXnc	the dimension of the noncommon regressors Xnc.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
sam0	the directions q to compute the variance bounds on the radial function.

eps_default0	the matrix containing the directions q and the selected epsilon(q).
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to kp.
lim	the limit number of observations under which we do no compute the conditional variance.
weights_x	the sampling weights for the dataset (Xnc,Xc).
weights_y	the sampling weights for the dataset (Y,Xc).
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.#' @param nc_sign if sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
refs0	indicating the positions in the vector values corresponding to the components of betac.
type	Equal to "both".
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
bc	if TRUE compute also the bounds on betac. Default is FALSE.
version	version of the computation of the ratio, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.
modeNA	indicates if NA introduced if the interval is empty. Default is FALSE.

Value

a matrix containing the considered directions and the computed value of the support function.

`compute_support_paral` *Function to minimize to compute the function sigma for the projections of the identified set*

Description

Function to minimize to compute the function sigma for the projections of the identified set

Usage

```
compute_support_paral(
  dir_nb,
  sam0,
  Xnc,
  eps_default0,
  grid,
  dimXc,
  dimXnc,
  Xc_xb = NULL,
  Xncb,
  Xc_yb = NULL,
  Yb,
  values,
  weights_x,
  weights_y,
  constraint = NULL,
  c_sign,
  nc_sign,
  refs0,
  meth,
  T_xy,
  bc,
  version,
  R2bound = NULL,
  values_sel = NULL,
  ties = FALSE,
  modeNA = FALSE
)
```

Arguments

`dir_nb`	the reference for the considered direction e in sam0
`sam0`	the directions q to compute the radial function.
`Xnc`	the noncommon regressor on the dataset (Xnc,Xc). No default
`eps_default0`	the matrix containing the directions q and the selected epsilon(q)

grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to kp.
dimXc	the dimension of the common regressors Xc.
dimXnc	the dimension of the noncommon regressors Xnc.
Xc_xb	the possibly bootstraped/subsampled common regressor on the dataset (Xnc,Xc). Default is NULL.
Xncb	the possibly bootstraped/subsampled noncommon regressor on the dataset (Xnc,Xc). No default.
Xc_yb	the possibly bootstraped/subsampled common regressor on the dataset (Y,Xc). Default is NULL.
Yb	the possibly bootstraped/subsampled outcome variable on the dataset (Y,Xc). No default.
values	the different unique points of support of the common regressor Xc.
weights_x	the bootstrap or sampling weights for the dataset (Xnc,Xc).
weights_y	the bootstrap or sampling weights for the dataset (Y,Xc).
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.#' @param nc_sign if sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
refs0	indicating the positions in the vector values corresponding to the components of betac.
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
T_xy	the apparent sample size the taking into account the difference in the two datasets.
bc	if TRUE compute also the bounds on betac. Default is FALSE.
version	version of the computation of the ratio, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.
modeNA	indicates if NA introduced if the interval is empty. Default is FALSE.

Value

the value of the support function in the specifed direction dir_nb.

`create_values` *Function to create the matrix of the support points for the common regressors Xc*

Description

Function to create the matrix of the support points for the common regressors Xc

Usage

```
create_values(dimX, c_var, Rdata)
```

Arguments

`dimX`	the dimension of the common regressors Xc.
`c_var`	the label of these regressors.
`Rdata`	dataset containing (Xnc,Xc) where Xnc are the non commonly observed regressors, Xc are potential common regressors.

Value

a matrix of the values of the support points for the common regressors Xc

`DGM_bounds` *This function compute the DGM bounds for all the different coefficients.*

Description

This function compute the DGM bounds for all the different coefficients.

Usage

```
DGM_bounds(
  Ldata,
  Rdata,
  values,
  sam0,
  refs0,
  out_var,
  nc_var,
  c_var = NULL,
  constraint = NULL,
  nc_sign = NULL,
  c_sign = NULL,
```

```
  nbCores = 1,
  eps_default = 0.5,
  nb_pts = 1,
  Bsamp = 1000,
  grid = 30,
  weights_x = NULL,
  weights_y = NULL,
  outside = FALSE,
  meth = "adapt",
  modeNA = FALSE,
  version = "second",
  version_sel = "second",
  alpha = 0.05,
  projections = FALSE,
  R2bound = NULL,
  values_sel = NULL,
  ties = FALSE,
  mult = NULL,
  seed = 2131
)
```

Arguments

Ldata	dataset containing (Y,Xc) where Y is the outcome, Xc are potential common regressors.
Rdata	dataset containing (Xnc,Xc) where Xnc are the non commonly observed regressors, Xc are potential common regressors.
values	the different unique points of support of the common regressor Xc.
sam0	the directions q to compute the radial function.
refs0	indicating the positions in the vector values corresponding to the components of betac.
out_var	label of the outcome variable Y.
nc_var	label of the non commonly observed regressors Xnc.
c_var	label of the commonly observed regressors Xc.
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.#' @param nc_sign if sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nbCores	number of cores for the parallel computation. Default is 1.

eps_default	If grid =NULL, then epsilon is taken equal to eps_default.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
Bsamp	the number of bootstrap/subsampling replications. Default is 1000.
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to kp.
weights_x	the sampling weights for the dataset (Xnc,Xc). Default is NULL.
weights_y	the sampling weights for the dataset (Y,Xc). Default is NULL.
outside	if TRUE indicates that the parallel computing has been launched outside of the function. Default is FALSE.
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
modeNA	indicates if NA introduced if the interval is empty. Default is FALSE.
version	version of the computation of the ratio, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
version_sel	version of the selection of the epsilon, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
alpha	for the level of the confidence region. Default is 0.05.
projections	if FALSE compute the identified set along some directions or the confidence regions. Default is FALSE
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.
mult	a list of multipliers of our selected epsilon to look at the robustness of the point estimates with respect to it. Default is NULL
seed	set a seed to fix the subsampling replications

Value

a list containing, in order: - ci : a list with all the information on the confidence intervals

* upper: upper bound of the confidence interval on the radial function S in the specified direction at level alpha, possibly with sign constraints

* lower: lower bound upper bound of the confidence interval on the radial function S, possibly with sign constraints

* unconstr: confidence interval on the radial function S, without sign constraints

* If common regressors, upper_agg, lower_agg, and unconstr_agg reports the same values but aggregated over the values of Xc (see the parameter theta0 in the paper)

* betac_ci: confidence intervals on each coefficients related to the common regressor, possibly with sign constraints

* betac_ci_unc: confidence intervals on each coefficients related to the common regressor without sign constraints

If projection is TRUE:

* support: confidence bound on the support function in each specified direction

- point : a list with all the information on the point estimates

* upper: the upper bounds on betanc, possibly with sign constraints

* lower: the lower bounds on betanc, possibly with sign constraints

* unconstr: bounds on betanc without sign constraints

* If common regressors, upper_agg, lower_agg, and unconstr_agg reports the same values but aggregated over the values of Xc (see the parameter theta0 in the paper)

* betac_pt: bounds on betanc, possibly with sign constraints

* betac_pt_unc: bounds on betanc without sign constraints If projection ==TRUE:

* support: point estimate of the support function in each specified direction

- epsilon : the values of the selected epsilon(q)

Examples

```
n=200
Xnc_x = rnorm(n,0,1.5)
Xnc_y = rnorm(n,0,1.5)
epsilon = rnorm(n,0,1)

## true value
beta0 =1
Y = Xnc_y*beta0 + epsilon
out_var = "Y"
nc_var = "Xnc"

# create the datasets
Ldata<- as.data.frame(Y)
colnames(Ldata) <- c(out_var)
Rdata <- as.data.frame(Xnc_x)
colnames(Rdata) <- c(nc_var)
 values = NULL
s= NULL
refs0 = NULL

sam0 <- rbind(-1,1)
eps0 = 0
############## Estimation #############
output <- DGM_bounds(Ldata,Rdata,values,sam0,refs0,out_var,nc_var)
```

`DGM_bounds_test` *This function compute the DGM bounds for all the different coefficients, adapted to the point identification test.*

Description

This function compute the DGM bounds for all the different coefficients, adapted to the point identification test.

Usage

```
DGM_bounds_test(
  Ldata,
  Rdata,
  values,
  sam0,
  refs0,
  out_var,
  nc_var,
  c_var = NULL,
  constraint = NULL,
  nc_sign = NULL,
  c_sign = NULL,
  nbCores = 1,
  eps_default = 0.5,
  nb_pts = 1,
  Bsamp = 1000,
  grid = 30,
  weights_x = NULL,
  weights_y = NULL,
  outside = FALSE,
  meth = "adapt",
  modeNA = FALSE,
  version = "first",
  version_sel = "first",
  alpha = 0.05,
  projections = FALSE,
  R2bound = NULL,
  values_sel = NULL,
  ties = FALSE,
  seed = 2131
)
```

Arguments

Ldata	dataset containing (Y,Xc) where Y is the outcome, Xc are potential common regressors.
Rdata	dataset containing (Xnc,Xc) where Xnc are the non commonly observed regressors, Xc are potential common regressors.
values	the different unique points of support of the common regressor Xc.
sam0	the directions q to compute the radial function.
refs0	indicating the positions in the vector values corresponding to the components of betac.

out_var	label of the outcome variable Y.
nc_var	label of the non commonly observed regressors Xnc.
c_var	label of the commonly observed regressors Xc.
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nbCores	number of cores for the parallel computation. Default is 1.
eps_default	If grid =NULL, then epsilon is taken equal to eps_default.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
Bsamp	the number of bootstrap/subsampling replications. Default is 1000.
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to kp.
weights_x	the sampling weights for the dataset (Xnc,Xc). Default is NULL.
weights_y	the sampling weights for the dataset (Y,Xc). Default is NULL.
outside	if TRUE indicates that the parallel computing has been launched outside of the function. Default is FALSE.
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
modeNA	indicates if NA introduced if the interval is empty. Default is FALSE.
version	version of the computation of the ratio, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
version_sel	version of the selection of the epsilon, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
alpha	for the level of the confidence region. Default is 0.05.
projections	if FALSE compute the identified set along some directions or the confidence regions. Default is FALSE
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.
seed	set a seed to fix the subsampling replications.

Value

a list containing, in order: - ci : a list with all the information on the confidence intervals

* upper: upper bound of the confidence interval on the radial function S in the specified direction at level alpha, possibly with sign constraints

* lower: lower bound upper bound of the confidence interval on the radial function S, possibly with sign constraints

* unconstr: confidence interval on the radial function S, without sign constraints

* If common regressors, upper_agg, lower_agg, and unconstr_agg reports the same values but aggregated over the values of Xc (see the parameter theta0 in the paper)

* betac_ci: confidence intervals on each coefficients related to the common regressor, possibly with sign constraints

* betac_ci_unc: confidence intervals on each coefficients related to the common regressor without sign constraints

If projection is TRUE:

* support: confidence bound on the support function in each specified direction

- point : a list with all the information on the point estimates

* upper: the upper bounds on betanc, possibly with sign constraints

* lower: the lower bounds on betanc, possibly with sign constraints

* unconstr: bounds on betanc without sign constraints

* If common regressors, upper_agg, lower_agg, and unconstr_agg reports the same values but aggregated over the values of Xc (see the parameter theta0 in the paper)

* betac_pt: bounds on betanc, possibly with sign constraints

* betac_pt_unc: bounds on betanc without sign constraints If projection ==TRUE:

* support: point estimate of the support function in each specified direction

- epsilon : the values of the selected epsilon(q)

`ewcdf` *Compute the weighted empirical cumulative distribution*

Description

Compute the weighted empirical cumulative distribution

Usage

```
ewcdf(x, weights = rep(1/length(x), length(x)))
```

Arguments

`x`	the sample
`weights`	the associated weights if any. Default is uniform.

Value

a vector containing:

- the weighted empirical cumulative distribution function

- the cumulated weights associated to the ordered values of the random variable.

`objective_support` *Internal function to minimize to compute the function sigma for the projections of the identified set*

Description

Internal function to minimize to compute the function sigma for the projections of the identified set

Usage

```
objective_support(
  x,
  dir_nb,
  sam0,
  eps1,
  Xc_xb,
  Xncb,
  Xc_yb,
  Yb,
  values,
  grid,
  weights_x,
  weights_y,
  constraint,
  c_sign,
  nc_sign,
  refs0,
  meth = "adapt",
  T_xy,
  bc = FALSE,
  version = "first",
  R2bound = NULL,
  values_sel = NULL,
  ties = FALSE,
  modeNA = FALSE
)
```

Arguments

`x`	value at which the function is evaluated.
`dir_nb`	the index of the considered direction.

sam0	the set of directions e where to compute the support function
eps1	the matrix of directions q, along the canonical axis, and the selected epsilon(q)
Xc_xb	the possibly bootstraped/subsampled common regressor on the dataset (Xnc,Xc). Default is NULL.
Xncb	the possibly bootstraped/subsampled noncommon regressor on the dataset (Xnc,Xc). No default.
Xc_yb	the possibly bootstraped/subsampled common regressor on the dataset (Y,Xc). Default is NULL.
Yb	the possibly bootstraped/subsampled outcome variable on the dataset (Y,Xc). No default.
values	the different unique points of support of the common regressor Xc.
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to kp.
weights_x	the bootstrap or sampling weights for the dataset (Xnc,Xc).
weights_y	the bootstrap or sampling weights for the dataset (Y,Xc).
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
refs0	indicating the positions in the vector values corresponding to the components of betac.
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
T_xy	the apparent sample size the taking into account the difference in the two datasets.
bc	if TRUE compute also the bounds on betac. Default is FALSE.
version	version of the computation of the ratio, "first" is a degraded version but fast; "second" is a correct version but slower. Default is "second".
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.
modeNA	indicates if NA introduced if the interval is empty. Default is FALSE.

Value

the value the support function

point_ident_test *Function performing the test of point identification on a validation sample.*

Description

Function performing the test of point identification on a validation sample.

Usage

```
point_ident_test(
  validation,
  Ldata = NULL,
  Rdata = NULL,
  out_var,
  nc_var,
  c_var = NULL,
  alpha = 0.05,
  constraint = NULL,
  nc_sign = NULL,
  c_sign = NULL,
  weights_validation = NULL,
  weights_x = NULL,
  weights_y = NULL,
  nbCores = 1,
  grid = 10,
  eps_default = 0.5,
  R2bound = NULL,
  unchanged = FALSE,
  ties = FALSE
)
```

Arguments

validation	dataset containing the joint distribution (Y,Xnc,Xc) where Y is the outcome, Xnc are the non commonly observed regressors, Xc are potential common regressors.
Ldata	dataset containing (Y,Xc) where Y is the outcome, Xc are potential common regressors. Default is NULL
Rdata	dataset containing (Xnc,Xc) where Xnc are the non commonly observed regressors, Xc are potential common regressors. Default is NULL.
out_var	label of the outcome variable Y.
nc_var	label of the non commonly observed regressors Xnc.
c_var	label of the commonly observed regressors Xc.
alpha	the level of the confidence intervals. Default is 0.05.

constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
nc_sign	if sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
c_sign	if sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
weights_validation	the sampling weights for the full dataset (Y, Xnc,Xc). Default is NULL.
weights_x	the sampling weights for the dataset (Xnc,Xc). Default is NULL.
weights_y	the sampling weights for the dataset (Y,Xc). Default is NULL.
nbCores	number of cores for the parallel computation. Default is 1.
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to eps_default.
eps_default	If grid =NULL, then epsilon is taken equal to eps_default.
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
unchanged	Boolean indicating if the categories based on Xc must be kept unchanged (TRUE). Otherwise (FALSE), a thresholding approach is taken imposing that each value appears more than 10 times in both datasets and 0.01 per cent is the pooled one. Default is FALSE.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.

Value

a list containing, in order: - S: the point estimation used the statistic for the test

- S_ci: the CI on the upper bound

- stat: the statistic of the test

- the critical value at level alpha

- the p_value of the test

- the fit with the OLS on this sample

- n the sample size

- epsilon, the choice of epsilon we made

- r2long the r2 on the long regression

-r2short the r2 on the short regression

Examples

```
### Simulating joint distribution according to this DGP
n=200
Xnc = rnorm(n,0,1.5)
epsilon = rnorm(n,0,1)
```

```
## true value
beta0 =1
Y = Xnc*beta0 + epsilon
out_var = "Y"
nc_var = "Xnc"

# create the datasets
validation<- as.data.frame(cbind(Y,Xnc))
colnames(validation) <- c(out_var,nc_var)


############## Estimation #############
test = point_ident_test (validation, Ldata=NULL,Rdata=NULL,out_var,nc_var)
```

---

regCombin *Function computing all the different bounds : DGM and/or Variance*

Description

Function computing all the different bounds : DGM and/or Variance

Usage

```
regCombin(
  Ldata,
  Rdata,
  out_var,
  nc_var,
  c_var = NULL,
  constraint = NULL,
  nc_sign = NULL,
  c_sign = NULL,
  weights_x = NULL,
  weights_y = NULL,
  nbCores = 1,
  methods = c("DGM"),
  grid = 10,
  alpha = 0.05,
  eps_default = 0.5,
  R2bound = NULL,
  projections = FALSE,
  unchanged = FALSE,
  ties = FALSE,
  seed = 2131,
  mult = NULL
)
```

Arguments

Ldata	a dataset including Y and possibly X_c=(X_c1,...,X_cq). X_c must be finitely supported.
Rdata	a dataset including X_nc and the same variables X_c as in Ldata.
out_var	the label of the outcome variable Y.
nc_var	the labels of the regressors X_nc.
c_var	the labels of the regressors X_c (if any).
constraint	a vector of size q indicating the type of constraints (if any) on the function f(x_c1,...,x_cq) for k=1,...,q: "convex", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NA for no constraint. Default is NULL, namely no constraints at all.
nc_sign	a vector of size p indicating sign restrictions on each of the p coefficients of X_nc. For each component, -1 corresponds to a minus sign, 1 to a plus sign and 0 to no constraint. Default is NULL, namely no constraints at all.
c_sign	same as nc_sign but for X_c (accordingly, it is a vector of size q).
weights_x	the sampling weights for the dataset Rdata. Default is NULL.
weights_y	the sampling weights for the dataset Ldata. Default is NULL.
nbCores	number of cores for the parallel computation. Default is 1.
methods	method used for the bounds: "DGM" (Default) and/or "Variance".
grid	the number of points for the grid search on epsilon. If NULL, then grid search is not performed and epsilon is taken as eps_default. Default is 10.
alpha	one minus the nominal coverage of the confidence intervals. Default is 0.05.
eps_default	a pre-specified value of epsilon used only if the grid search for selecting the value of epsilon is not performed, i.e, when grid is NULL. Default is 0.5.
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
projections	a boolean indicating if the identified set and confidence intervals on beta_0k for k=1,...,p are computed (TRUE), rather than the identified set and confidence region of beta_0 (FALSE). Default is FALSE.
unchanged	a boolean indicating if the categories based on X_c must be kept unchanged (TRUE). Otherwise (FALSE), a thresholding approach is taken imposing that each value appears more than 10 times in both datasets and represents more than 0.01 per cent of the pooled dataset (of size n_X+n_Y). Default is FALSE.
ties	a boolean indicating if there are ties in the dataset. If not (FALSE), computation is faster. Default is FALSE.
seed	to avoid fixinx the seed for the subsampling, set to NULL. Otherwise 2131.
mult	a list of multipliers of our selected epsilon to look at the robustness of the point estimates with respect to it. Default is NULL

Value

Use summary_regCombin for a user-friendly print of the estimates. Returns a list containing, in order: - DGM_complete or Variance_complete : the complete outputs of the functions DGM_bounds or Variance_bounds.

and additional pre-treated outputs, replace below "method" by either "DGM" or "Variance":

- methodCI: the confidence region on the betanc without sign constraints

- methodpt: the bounds point estimates on the betanc without sign constraints

- methodCI_sign: the confidence region on the betanc with sign constraints

- methodpt_sign: the bounds point estimates on the betanc with sign constraints

- methodkp: the values of epsilon(q)

- methodbeta1: the confidence region on the betac corresponding to the common regressors Xc without sign constraints

- methodbeta1_pt: the bounds point estimates on the betac corresponding to the common regressors Xc without sign constraints

- methodbeta1_sign: the confidence region on the betac corresponding to the common regressors Xc with sign constraints

- methodbeta1_sign_pt: the bounds point estimates on the betac corresponding to the common regressors Xc with sign constraints

Examples

```
### Simulating according to this DGP
n=200
Xnc_x = rnorm(n,0,1.5)
Xnc_y = rnorm(n,0,1.5)
epsilon = rnorm(n,0,1)

## true value
beta0 =1
Y = Xnc_y*beta0 + epsilon
out_var = "Y"
nc_var = "Xnc"

# create the datasets
Ldata<- as.data.frame(Y)
colnames(Ldata) <- c(out_var)
Rdata <- as.data.frame(Xnc_x)
colnames(Rdata) <- c(nc_var)


############# Estimation #############
output <- regCombin(Ldata,Rdata,out_var,nc_var)
```

`regCombin_profile` *Computing the DGM bounds for different values of epsilon, proportional to the data-driven selected one*

Description

Computing the DGM bounds for different values of epsilon, proportional to the data-driven selected one

Usage

```
regCombin_profile(
  Ldata,
  Rdata,
  out_var,
  nc_var,
  c_var = NULL,
  constraint = NULL,
  nc_sign = NULL,
  c_sign = NULL,
  weights_x = NULL,
  weights_y = NULL,
  nbCores = 1,
  methods = c("DGM"),
  grid = 10,
  alpha = 0.05,
  eps_default = 0.5,
  R2bound = NULL,
  projections = FALSE,
  unchanged = FALSE,
  ties = FALSE,
  multipliers = c(0.25, 0.5, 1, 1.5, 2)
)
```

Arguments

`Ldata`	dataset containing (Y,Xc) where Y is the outcome, Xc are potential common regressors.
`Rdata`	dataset containing (Xnc,Xc) where Xnc are the non commonly observed regressors, Xc are potential common regressors.
`out_var`	label of the outcome variable Y.
`nc_var`	label of the non commonly observed regressors Xnc.
`c_var`	label of the commonly observed regressors Xc.
`constraint`	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave",

	"nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
nc_sign	if sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
c_sign	if sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
weights_x	the sampling weights for the dataset (Xnc,Xc). Default is NULL.
weights_y	the sampling weights for the dataset (Y,Xc). Default is NULL.
nbCores	number of cores for the parallel computation. Default is 1.
methods	method used for the bounds: "DGM" (Default) and/or "Variance".
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to eps_default.
alpha	the level of the confidence intervals. Default is 0.05.
eps_default	If grid =NULL, then epsilon is taken equal to eps_default.
R2bound	the lower bound on the R2 of the long regression if any. Default is NULL.
projections	if FALSE compute the identified set along some directions or the confidence regions. Default is FALSE
unchanged	Boolean indicating if the categories based on Xc must be kept unchanged (TRUE). Otherwise (FALSE), a thresholding approach is taken imposing that each value appears more than 10 times in both datasets and 0.01 per cent is the pooled one. Default is FALSE.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.
multipliers	different multipliers of our selected epsilon to compute the bounds. Default is 0.25,0.5,1,1.5,2.

Value

a list containing, in order: - details: a list with all the detailled results of the estimation for the different multipliers. see "regCombin".

- Profile_point : a matrix with the profile of the bounds without constraints for different values of the multiplier.

- Profile_point_sign : a matrix with the profile of the bounds with constraints for different values of the multiplier.

Examples

```
### Simulating according to this DGP
n=200
Xnc_x = rnorm(n,0,1.5)
Xnc_y = rnorm(n,0,1.5)
epsilon = rnorm(n,0,1)

## true value
beta0 =1
Y = Xnc_y*beta0 + epsilon
```

```
out_var = "Y"
nc_var = "Xnc"

# create the datasets
Ldata<- as.data.frame(Y)
colnames(Ldata) <- c(out_var)
Rdata <- as.data.frame(Xnc_x)
colnames(Rdata) <- c(nc_var)


############## Estimation #############
profile = regCombin_profile(Ldata,Rdata,out_var,nc_var, multipliers = seq(0.1,3,length.out=3))
```

`sampling_rule` *The subsampling rule*

Description

The subsampling rule

Usage

```
sampling_rule(n)
```

Arguments

`n`	sample size.

Value

the subsampling size

`select_epsilon` *Function for the data-driven selection of the epsilon tuning parameter*

Description

Function for the data-driven selection of the epsilon tuning parameter

Usage

```
select_epsilon(
  sam1,
  eps_default,
  Xc_x,
  Xnc,
  Xc_y,
  Y,
  values,
  dimXc,
  dimXnc,
  nb_pts,
  lim,
  weights_x,
  weights_y,
  refs0,
  grid = 30,
  constraint = NULL,
  c_sign = NULL,
  nc_sign = NULL,
  meth = "adapt",
  nbCores = 1,
  version_sel = "first",
  alpha = 0.05,
  ties = FALSE
)
```

Arguments

sam1	the matrix containing the directions q on which to compute the selected rule for epsilon(q)
eps_default	If grid =NULL, then epsilon is taken equal to eps_default.
Xc_x	the common regressor on the dataset (Xnc,Xc). Default is NULL.
Xnc	the noncommon regressor on the dataset (Xnc,Xc). No default.
Xc_y	the common regressor on the dataset (Y,Xc). Default is NULL.
Y	the outcome variable. No default.
values	the different unique points of support of the common regressor Xc.
dimXc	the dimension of the common regressors Xc.
dimXnc	the dimension of the noncommon regressors Xnc.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
lim	the lim number of observations under which we do no compute the conditional variance.
weights_x	the sampling weights for the dataset (Xnc,Xc).

weights_y	the sampling weights for the dataset (Y,Xc).
refs0	indicating the positions in the vector values corresponding to the components of betac.
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to eps_default.
constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
nbCores	number of cores for the parallel computation. Default is 1.
version_sel	version of the selection of the epsilon, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
alpha	the level for the confidence regions. Default is 0.05.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.

Value

a matrix containing the values of the selected epsilon(q) for q directions in sam1.

---

select_epsilon_test *Function for the data-driven selection of the epsilon tuning parameter, adapted to the point identification test.*

Description

Function for the data-driven selection of the epsilon tuning parameter, adapted to the point identification test.

Usage

```
select_epsilon_test(
  sam1,
  eps_default,
  Xc_x,
  Xnc,
  Xc_y,
  Y,
```

```
  values,
  dimXc,
  dimXnc,
  nb_pts,
  lim,
  weights_x,
  weights_y,
  refs0,
  grid = 30,
  constraint = NULL,
  c_sign = NULL,
  nc_sign = NULL,
  meth = "adapt",
  nbCores = 1,
  version_sel = "first",
  alpha = 0.05,
  ties = FALSE
)
```

Arguments

sam1	the matrix containing the directions q on which to compute the selected rule for epsilon(q)
eps_default	If grid =NULL, then epsilon is taken equal to eps_default.
Xc_x	the common regressor on the dataset (Xnc,Xc). Default is NULL.
Xnc	the noncommon regressor on the dataset (Xnc,Xc). No default.
Xc_y	the common regressor on the dataset (Y,Xc). Default is NULL.
Y	the outcome variable. No default.
values	the different unique points of support of the common regressor Xc.
dimXc	the dimension of the common regressors Xc.
dimXnc	the dimension of the noncommon regressors Xnc.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
lim	the lim number of observations under which we do no compute the conditional variance.
weights_x	the sampling weights for the dataset (Xnc,Xc).
weights_y	the sampling weights for the dataset (Y,Xc).
refs0	indicating the positions in the vector values corresponding to the components of betac.
grid	the number of points for the grid search on epsilon. Default is 30. If NULL, then epsilon is taken fixed equal to eps_default.

constraint	a vector indicating the different constraints in a vector of the size of X_c indicating the type of constraints, if any on f(X_c) : "concave", "concave", "nondecreasing", "nonincreasing", "nondecreasing_convex", "nondecreasing_concave", "nonincreasing_convex", "nonincreasing_concave", or NULL for none. Default is NULL, no contraints at all.
c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
meth	the method for the choice of epsilon, either "adapt", i.e. adapted to the direction or "min" the minimum over the directions. Default is "adapt".
nbCores	number of cores for the parallel computation. Default is 1.
version_sel	version of the selection of the epsilon, "first" indicates no weights, no ties, same sizes of the two datasets; "second" otherwise. Default is "second".
alpha	the level for the confidence regions. Default is 0.05.
ties	Boolean indicating if there are ties in the dataset. Default is FALSE.

Value

a matrix containing the values of the selected epsilon(q) for q directions in sam1.

---

summary_regCombin *Produce the final summary table for the output of the felogit function*

Description

Produce the final summary table for the output of the felogit function

Usage

```
summary_regCombin(output, format = NULL)
```

Arguments

output	the output of the felogit function
format	can take value "latex" to print the latex table

Value

a kableExtra or xtable table plotted respectively in the R viewer or terminal

Examples

```
### Simulating according to this DGP
n=200
Xnc_x = rnorm(n,0,1.5)
Xnc_y = rnorm(n,0,1.5)
epsilon = rnorm(n,0,1)

## true value
beta0 =1
Y = Xnc_y*beta0 + epsilon
out_var = "Y"
nc_var = "Xnc"

# create the datasets
Ldata<- as.data.frame(Y)
colnames(Ldata) <- c(out_var)
Rdata <- as.data.frame(Xnc_x)
colnames(Rdata) <- c(nc_var)


############# Estimation #############
output <- regCombin(Ldata,Rdata,out_var,nc_var)
mat = summary_regCombin(output)
```

---

tabulate_values *Function to tabulate the values the common regressors Xc whatever the dimension.*

Description

Function to tabulate the values the common regressors Xc whatever the dimension.

Usage

```
tabulate_values(k, values, Xc0, dimXc)
```

Arguments

k	the considered value in "values"
values	the different unique points of support of the common regressor Xc.
Xc0	dataset containing Xc, the common regressors.
dimXc	the dimension of Xc

Value

a matrix of the number of times the kth value in the vector values appears.

`Variance_bounds` *Function to compute the variance bounds for Xnc*

Description

Function to compute the variance bounds for Xnc

Usage

```
Variance_bounds(
  Ldata,
  Rdata,
  out_var,
  c_var,
  nc_var,
  constraint = NULL,
  c_sign = NULL,
  nc_sign = NULL,
  projections = TRUE,
  values,
  sam0,
  refs0,
  nb_pts,
  eps_default,
  nbCores,
  Bsamp = 2000,
  weights_x = NULL,
  weights_y = NULL,
  outside = FALSE,
  alpha = 0.05,
  values_sel = NULL,
  seed = 21
)
```

Arguments

`Ldata`	dataset containing (Y,Xc) where Y is the outcome, Xc are potential common regressors
`Rdata`	dataset containing (Xnc,Xc) where Xnc are the non commonly observed regressors, Xc are potential common regressors
`out_var`	label of the outcome variable Y.
`c_var`	label of the commonly observed regressors Xc.
`nc_var`	label of the non commonly observed regressors Xnc.
`constraint`	vector of the size of X_c indicating the type of constraint if any on f(X_c) : "monotone", "convex", "sign", or "none". Default is NULL, no contraints at all.

c_sign	sign restrictions on the commonly observed regressors: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
nc_sign	sign restrictions on the non-commonly observed regressors Xnc: -1 for a minus sign, 1 for a plus sign, 0 otherwise. Default is NULL, i.e. no constraints.
projections	if FALSE compute the identified set along some directions or the confidence regions. Default is FALSE.
values	the different unique points of support of the common regressor Xc.
sam0	the directions q to compute the radial function.
refs0	indicating the positions in the vector values corresponding to the components of betac.
nb_pts	the constant C in DGM for the epsilon_0, the lower bound on the grid for epsilon, taken equal to nb_pts*ln(n)/n. Default is 1 without regressors Xc, 3 with Xc.
eps_default	If data_k =NULL, then epsilon is taken equal to eps_default.
nbCores	number of cores for the parallel computation. Default is 1.
Bsamp	the number of bootstrap/subsampling replications. Default is 1000.
weights_x	the sampling weights for the dataset (Xnc,Xc).
weights_y	the sampling weights for the dataset (Y,Xc).
outside	if TRUE indicates that the parallel computing has been launched outside of the function. Default is FALSE.
alpha	for the level of the confidence region. Default is 0.05.
values_sel	the selected values of Xc for the conditioning. Default is NULL.
seed	set a seed to fix the subsampling replications

Value

a list containing, in order: - ci : a list with all the information on the confidence intervals

- upper: upper bound of the confidence interval on betanc at level alpha, possibly with sign constraints

- lower: lower bound upper bound of the confidence interval on betanc, possibly with sign constraints

- unconstr: confidence interval on betanc, without sign constraints

- betac_ci: confidence intervals on each coefficients related to the common regressor, possibly with sign constraints

- betac_ci_unc: confidence intervals on each coefficients related to the common regressor without sign constraints

- point : a list with all the information on the point estimates

- upper: the upper bounds on betanc, possibly with sign constraints

- lower: the lower bounds on betanc, possibly with sign constraints

-unconstr: bounds on betanc without sign constraints

-betac_pt: bounds on betanc, possibly with sign constraints

-betac_pt_unc: bounds on betanc without sign constraints

Index

AS_bounds, 2
AStest, 3

compute_bnds_betac, 4
compute_constraints, 5
compute_radial, 6
compute_radial_test, 8
compute_ratio, 10
compute_ratio_variance, 11
compute_stat_variance, 12
compute_support, 13
compute_support_paral, 16
create_values, 18

DGM_bounds, 18
DGM_bounds_test, 21

ewcdf, 24

objective_support, 25

point_ident_test, 27

regCombin, 29
regCombin_profile, 32

sampling_rule, 34
select_epsilon, 34
select_epsilon_test, 36
summary_regCombin, 38

tabulate_values, 39

Variance_bounds, 40