Package 'rdlocrand'

Title Local Randomization Methods for RD Designs

Version 0.8

Author Matias D. Cattaneo, Rocio Titiunik, Gonzalo Vazquez-Bare

Maintainer Gonzalo Vazquez-Bare <gvazquez@econ.ucsb.edu>

Description

The regression discontinuity (RD) design is a popular quasi-experimental design for causal inference and policy evaluation. Under the local randomization approach, RD designs can be interpreted as randomized experiments inside a window around the cutoff. This package provides tools to perform randomization inference for RD designs under local randomization: rdrandinf() to perform hypothesis testing using randomization inference, rdwinselect() to select a window around the cutoff in which randomization is likely to hold, rdsensitivity() to assess the sensitivity of the results to different window lengths and null hypotheses and rdrbounds() to construct Rosenbaum bounds for sensitivity to unobserved confounders. See Cattaneo, Titiunik and Vazquez-Bare (2016) <https://rdpackages.github.io/references/Cattaneo-Titiunik-VazquezBare_2016_Stata.pdf> for further methodological details.

Imports AER,
sandwich

Depends R (>= 3.1)

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

R topics documented:

---

rdlocrand-package   *rdlocrand: Local Randomization Methods for RD Designs*

Description

The regression discontinuity (RD) design is a popular quasi-experimental design for causal inference and policy evaluation. Under the local randomization approach, RD designs can be interpreted as randomized experiments inside a window around the cutoff. The rdlocrand package provides tools to analyze RD designs under local randomization: rdrandinf to perform hypothesis testing using randomization inference, rdwinselect to select a window around the cutoff in which randomization is likely to hold, rdsensitivity to assess the sensitivity of the results to different window lengths and null hypotheses and rdrbounds to construct Rosenbaum bounds for sensitivity to unobserved confounders. For more details, and related Stata and R packages useful for analysis of RD designs, visit https://rdpackages.github.io/.

Author(s)

Matias Cattaneo, Princeton University. <cattaneo@princeton.edu>

Rocio Titiunik, Princeton University. <titiunik@princeton.edu>

Gonzalo Vazquez-Bare, UC Santa Barbara. <gvazquez@econ.ucsb.edu>

References

Cattaneo, M.D., B. Frandsen and R. Titiunik. (2015). Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate. *Journal of Causal Inference* 3(1): 1-24.

Cattaneo, M.D., R. Titiunik and G. Vazquez-Bare. (2016). Inference in Regression Discontinuity Designs under Local Randomization. *Stata Journal* 16(2): 331-367.

Cattaneo, M.D., R. Titiunik and G. Vazquez-Bare. (2017). Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality. *Journal of Policy Analysis and Management* 36(3): 643-681.

Rosenbaum, P. (2002). Observational Studies. Springer.

---

rdrandinf   *Randomization Inference for RD Designs under Local Randomization*

Description

rdrandinf implements randomization inference and related methods for RD designs, using observations in a specified or data-driven selected window around the cutoff where local randomization is assumed to hold.

Usage

```
rdrandinf(
  Y,
  R,
  cutoff = 0,
  wl = NULL,
  wr = NULL,
  statistic = "diffmeans",
  p = 0,
  evall = NULL,
  evalr = NULL,
  kernel = "uniform",
  fuzzy = NULL,
  nulltau = 0,
  d = NULL,
  dscale = NULL,
  ci,
  interfci = NULL,
  bernoulli = NULL,
  reps = 1000,
  seed = 666,
  quietly = FALSE,
  covariates,
  obsmin = NULL,
  wmin = NULL,
  wobs = NULL,
  wstep = NULL,
  wsymmetric = FALSE,
  wmasspoints = FALSE,
  nwindows = 10,
  dropmissing = FALSE,
  rdwstat = "diffmeans",
  approx = FALSE,
  rdwreps = 1000,
  level = 0.15,
  plot = FALSE,
  obsstep = NULL
)
```

Arguments

Y	a vector containing the values of the outcome variable.
R	a vector containing the values of the running variable.
cutoff	the RD cutoff (default is 0).
wl	the left limit of the window. The default takes the minimum of the running variable.
wr	the right limit of the window. The default takes the maximum of the running variable.
statistic	the statistic to be used in the balance tests. Allowed options are `diffmeans` (difference in means statistic), `ksmirnov` (Kolmogorov-Smirnov statistic) and `ranksum` (Wilcoxon-Mann-Whitney standardized statistic). Default option is

	diffmeans. The statistic ttest is equivalent to diffmeans and included for backward compatibility.
p	the order of the polynomial for outcome transformation model (default is 0).
evall	the point at the left of the cutoff at which to evaluate the transformed outcome is evaluated. Default is the cutoff value.
evalr	specifies the point at the right of the cutoff at which the transformed outcome is evaluated. Default is the cutoff value.
kernel	specifies the type of kernel to use as weighting scheme. Allowed kernel types are uniform (uniform kernel), triangular (triangular kernel) and epan (Epanechnikov kernel). Default is uniform.
fuzzy	indicates that the RD design is fuzzy. fuzzy can be specified as a vector containing the values of the endogenous treatment variable, or as a list where the first element is the vector of endogenous treatment values and the second element is a string containing the name of the statistic to be used. Allowed statistics are ar (Anderson-Rubin statistic) and tsls (2SLS statistic). Default statistic is ar. The tsls statistic relies on large-sample approximation.
nulltau	the value of the treatment effect under the null hypothesis (default is 0).
d	the effect size for asymptotic power calculation. Default is 0.5 * standard deviation of outcome variable for the control group.
dscale	the fraction of the standard deviation of the outcome variable for the control group used as alternative hypothesis for asymptotic power calculation. Default is 0.5.
ci	calculates a confidence interval for the treatment effect by test inversion. ci can be specified as a scalar or a vector, where the first element indicates the value of alpha for the confidence interval (typically 0.05 or 0.01) and the remaining elements, if specified, indicate the grid of treatment effects to be evaluated. This option uses rdsensitivity to calculate the confidence interval. See corresponding help for details. Note: the default tlist can be narrow in some cases, which may truncate the confidence interval. We recommend the user to manually set a large enough tlist.
interfci	the level for Rosenbaum's confidence interval under arbitrary interference between units.
bernoulli	the probabilities of treatment for each unit when assignment mechanism is a Bernoulli trial. This option should be specified as a vector of length equal to the length of the outcome and running variables.
reps	the number of replications (default is 1000).
seed	the seed to be used for the randomization test.
quietly	suppresses the output table.
covariates	the covariates used by rdwinselect to choose the window when wl and wr are not specified. This should be a matrix of size n x k where n is the total sample size and k is the number of covariates.
obsmin	the minimum number of observations above and below the cutoff in the smallest window employed by the companion command rdwinselect. Default is 10.
wmin	the smallest window to be used (if minobs is not specified) by the companion command rdwinselect. Specifying both wmin and obsmin returns an error.
wobs	the number of observations to be added at each side of the cutoff at each step.

wstep	the increment in window length (if obsstep is not specified) by the companion command rdwinselect. Specifying both obsstep and wstep returns an error.
wsymmetric	requires that windows be symmetrized around the cutoff when (wobs is specified).
wmasspoints	specifies that the running variable is discrete and each masspoint should be used as a window.
nwindows	the number of windows to be used by the companion command rdwinselect. Default is 10.
dropmissing	drop rows with missing values in covariates when calculating windows.
rdwstat	the statistic to be used by the companion command rdwinselect (see corresponding help for options). Default option is ttest.
approx	forces the companion command rdwinselect to conduct the covariate balance tests using a large-sample approximation instead of finite-sample exact randomization inference methods.
rdwreps	the number of replications to be used by the companion command rdwinselect. Default is 1000.
level	the minimum accepted value of the p-value from the covariate balance tests to be used by the companion command rdwinselect. Default is .15.
plot	draws a scatter plot of the minimum p-value from the covariate balance test against window length implemented by the companion command rdwinselect.
obsstep	the minimum number of observations to be added on each side of the cutoff for the sequence of fixed-increment nested windows. Default is 2. This option is deprecated and only included for backward compatibility.

Value

sumstats	summary statistics
obs.stat	observed statistic(s)
p.value	randomization p-value(s)
asy.pvalue	asymptotic p-value(s)
window	chosen window
ci	confidence interval (only if ci option is specified)
interf.ci	confidence interval under interferecen (only if interfci is specified)

Author(s)

Matias Cattaneo, Princeton University. <cattaneo@princeton.edu>

Rocio Titiunik, Princeton University. <titiunik@princeton.edu>

Gonzalo Vazquez-Bare, UC Santa Barbara. <gvazquez@econ.ucsb.edu>

References

Cattaneo, M.D., R. Titiunik and G. Vazquez-Bare. (2016). Inference in Regression Discontinuity Designs under Local Randomization. *Stata Journal* 16(2): 331-367.

Examples

```
# Toy dataset
X <- array(rnorm(200),dim=c(100,2))
R <- X[1,] + X[2,] + rnorm(100)
Y <- 1 + R -.5*R^2 + .3*R^3 + (R>=0) + rnorm(100)
# Randomization inference in window (-.75,.75)
tmp <- rdrandinf(Y,R,wl=-.75,wr=.75)
# Randomization inference in window (-.75,.75), all statistics
tmp <- rdrandinf(Y,R,wl=-.75,wr=.75,statistic='all')
# Randomization inference with window selection
# Note: low number of replications to speed up process.
# The user should increase the number of replications.
tmp <- rdrandinf(Y,R,statistic='all',covariates=X,wmin=.5,wstep=.125,rdwreps=500)
```

---

rdrbounds	*Rosenbaum bounds for RD designs under local randomization*

Description

rdrbounds calculates lower and upper bounds for the randomization p-value under different degrees of departure from a local randomized experiment, as suggested by Rosenbaum (2002).

Usage

```
rdrbounds(
  Y,
  R,
  cutoff = 0,
  wlist,
  gamma,
  expgamma,
  bound = "both",
  statistic = "ranksum",
  p = 0,
  evalat = "cutoff",
  kernel = "uniform",
  fuzzy = NULL,
  nulltau = 0,
  prob,
  fmpval = FALSE,
  reps = 1000,
  seed = 666
)
```

Arguments

Y	a vector containing the values of the outcome variable.
R	a vector containing the values of the running variable.

cutoff	the RD cutoff (default is 0).
wlist	the list of window lengths to be evaluated. By default the program constructs 10 windows around the cutoff, the first one including 10 treated and control observations and adding 5 observations to each group in subsequent windows.
gamma	the list of values of gamma to be evaluated.
expgamma	the list of values of exp(gamma) to be evaluated. Default is `c(1.5,2,2.5,3)`.
bound	specifies which bounds the command calculates. Options are upper for upper bound, `lower` for lower bound and `both` for both upper and lower bounds. Default is `both`.
statistic	the statistic to be used in the balance tests. Allowed options are `diffmeans` (difference in means statistic), `ksmirnov` (Kolmogorov-Smirnov statistic) and `ranksum` (Wilcoxon-Mann-Whitney standardized statistic). Default option is `diffmeans`. The statistic `ttest` is equivalent to `diffmeans` and included for backward compatibility.
p	the order of the polynomial for outcome adjustment model. Default is 0.
evalat	specifies the point at which the adjusted variable is evaluated. Allowed options are `cutoff` and `means`. Default is `cutoff`.
kernel	specifies the type of kernel to use as weighting scheme. Allowed kernel types are `uniform` (uniform kernel), `triangular` (triangular kernel) and epan (Epanechnikov kernel). Default is `uniform`.
fuzzy	indicates that the RD design is fuzzy. `fuzzy` can be specified as a vector containing the values of the endogenous treatment variable, or as a list where the first element is the vector of endogenous treatment values and the second element is a string containing the name of the statistic to be used. Allowed statistics are ar (Anderson-Rubin statistic) and `tsls` (2SLS statistic). Default statistic is `ar`. The `tsls` statistic relies on large-sample approximation.
nulltau	the value of the treatment effect under the null hypothesis. Default is 0.
prob	the probabilities of treatment for each unit when assignment mechanism is a Bernoulli trial. This option should be specified as a vector of length equal to the length of the outcome and running variables.
fmpval	reports the p-value under fixed margins randomization, in addition to the p-value under Bernoulli trials.
reps	number of replications. Default is 1000.
seed	the seed to be used for the randomization tests.

Value

gamma	list of gamma values.
expgamma	list of exp(gamma) values.
wlist	window grid.
p.values	p-values for each window (under gamma = 0).
lower.bound	list of lower bound p-values for each window and gamma pair.
upper.bound	list of upper bound p-values for each window and gamma pair.

Author(s)

Matias Cattaneo, Princeton University. `<cattaneo@princeton.edu>`

Rocio Titiunik, Princeton University. `<titiunik@princeton.edu>`

Gonzalo Vazquez-Bare, UC Santa Barbara. `<gvazquez@econ.ucsb.edu>`

References

Cattaneo, M.D., R. Titiunik and G. Vazquez-Bare. (2016). Inference in Regression Discontinuity Designs under Local Randomization. *Stata Journal* 16(2): 331-367.

Rosenbaum, P. (2002). Observational Studies. Springer.

Examples

```
# Toy dataset
R <- runif(100,-1,1)
Y <- 1 + R -.5*R^2 + .3*R^3 + (R>=0) + rnorm(100)
# Rosenbaum bounds
# Note: low number of replications and windows to speed up process.
# The user should increase these values.
rdrbounds(Y,R,expgamma=c(1.5,2),wlist=c(.3),reps=100)
```

| rdsensitivity | *Sensitivity analysis for RD designs under local randomization* |

Description

rdsensitivity analyze the sensitivity of randomization p-values and confidence intervals to different window lengths.

Usage

```
rdsensitivity(
  Y,
  R,
  cutoff = 0,
  wlist,
  wlist_left,
  tlist,
  statistic = "diffmeans",
  p = 0,
  evalat = "cutoff",
  kernel = "uniform",
  fuzzy = NULL,
  ci = NULL,
  ci_alpha = 0.05,
  reps = 1000,
  seed = 666,
  nodraw = FALSE,
  quietly = FALSE
)
```

Arguments

Y	a vector containing the values of the outcome variable.
R	a vector containing the values of the running variable.
cutoff	the RD cutoff (default is 0).
wlist	the list of windows to the right of the cutoff. By default the program constructs 10 windows around the cutoffwith 5 observations each.
wlist_left	the list of windows to the left of the cutoff. If not specified, the windows are constructed symmetrically around the cutoff based on the values in wlist.
tlist	the list of values of the treatment effect under the null to be evaluated. By default the program employs ten evenly spaced points within the asymptotic confidence interval for a constant treatment effect in the smallest window to be used.
statistic	the statistic to be used in the balance tests. Allowed options are diffmeans (difference in means statistic), ksmirnov (Kolmogorov-Smirnov statistic) and ranksum (Wilcoxon-Mann-Whitney standardized statistic). Default option is diffmeans. The statistic ttest is equivalent to diffmeans and included for backward compatibility.
p	the order of the polynomial for outcome adjustment model. Default is 0.
evalat	specifies the point at which the adjusted variable is evaluated. Allowed options are cutoff and means. Default is cutoff.
kernel	specifies the type of kernel to use as weighting scheme. Allowed kernel types are uniform (uniform kernel), triangular (triangular kernel) and epan (Epanechnikov kernel). Default is uniform.
fuzzy	indicates that the RD design is fuzzy. fuzzy can be specified as a vector containing the values of the endogenous treatment variable, or as a list where the first element is the vector of endogenous treatment values and the second element is a string containing the name of the statistic to be used. Allowed statistics are ar (Anderson-Rubin statistic) and tsls (2SLS statistic). Default statistic is ar. The tsls statistic relies on large-sample approximation.
ci	returns the confidence interval corresponding to the indicated window length. ci has to be a two-dimensional vector indicating the left and right limits of the window. Default alpha is .05 (95% level CI).
ci_alpha	Specifies value of alpha for the confidence interval. Default alpha is .05 (95% level CI).
reps	number of replications. Default is 1000.
seed	the seed to be used for the randomization tests.
nodraw	suppresses contour plot.
quietly	suppresses the output table.

Value

tlist	treatment effects grid
wlist	window grid
results	table with corresponding p-values for each window and treatment effect pair.
ci	confidence interval (if ci is specified).

Author(s)

Matias Cattaneo, Princeton University. `<cattaneo@princeton.edu>`

Rocio Titiunik, Princeton University. `<titiunik@princeton.edu>`

Gonzalo Vazquez-Bare, UC Santa Barbara. `<gvazquez@econ.ucsb.edu>`

References

Cattaneo, M.D., R. Titiunik and G. Vazquez-Bare. (2016). Inference in Regression Discontinuity Designs under Local Randomization. *Stata Journal* 16(2): 331-367.

Examples

```
# Toy dataset
R <- runif(100,-1,1)
Y <- 1 + R -.5*R^2 + .3*R^3 + (R>=0) + rnorm(100)
# Sensitivity analysis
# Note: low number of replications to speed up process.
# The user should increase the number of replications.
tmp <- rdsensitivity(Y,R,wlist=seq(.75,2,by=.25),tlist=seq(0,5,by=1),reps=500)
```

---

rdwinselect   *Window selection for RD designs under local randomization*

Description

rdwinselect implements the window-selection procedure based on balance tests for RD designs under local randomization. Specifically, it constructs a sequence of nested windows around the RD cutoff and reports binomial tests for the running variable runvar and covariate balance tests for covariates covariates (if specified). The recommended window is the largest window around the cutoff such that the minimum p-value of the balance test is larger than a prespecified level for all nested (smaller) windows. By default, the p-values are calculated using randomization inference methods.

Usage

```
rdwinselect(
  R,
  X,
  cutoff = 0,
  obsmin = NULL,
  wmin = NULL,
  wobs = NULL,
  wstep = NULL,
  wsymmetric = FALSE,
  wmasspoints = FALSE,
  dropmissing = FALSE,
  nwindows = 10,
  statistic = "diffmeans",
  p = 0,
```

```
    evalat = "cutoff",
    kernel = "uniform",
    approx = FALSE,
    level = 0.15,
    reps = 1000,
    seed = 666,
    plot = FALSE,
    quietly = FALSE,
    obsstep = NULL
)
```

Arguments

R	a vector containing the values of the running variable.
X	the matrix of covariates to be used in the balancing tests. The matrix is optional but the recommended window is only provided when at least one covariate is specified. This should be a matrix of size n x k where n is the total sample size and k is the number of covariates.
cutoff	the RD cutoff (default is 0).
obsmin	the minimum number of observations above and below the cutoff in the smallest window. Default is 10.
wmin	the smallest window to be used.
wobs	the number of observations to be added at each side of the cutoff at each step. Default is 5.
wstep	the increment in window length.
wsymmetric	requires that windows be symmetrized around the cutoff when (wobs is specified).
wmasspoints	specifies that the running variable is discrete and each masspoint should be used as a window.
dropmissing	drop rows with missing values in covariates when calculating windows.
nwindows	the number of windows to be used. Default is 10.
statistic	the statistic to be used in the balance tests. Allowed options are diffmeans (difference in means statistic), ksmirnov (Kolmogorov-Smirnov statistic), ranksum (Wilcoxon-Mann-Whitney standardized statistic) and hotelling (Hotelling's T-squared statistic). Default option is diffmeans. The statistic ttest is equivalent to diffmeans and included for backward compatibility.
p	the order of the polynomial for outcome adjustment model (for covariates). Default is 0.
evalat	specifies the point at which the adjusted variable is evaluated. Allowed options are cutoff and means. Default is cutoff.
kernel	specifies the type of kernel to use as weighting scheme. Allowed kernel types are uniform (uniform kernel), triangular (triangular kernel) and epan (Epanechnikov kernel). Default is uniform.
approx	forces the command to conduct the covariate balance tests using a large-sample approximation instead of finite-sample exact randomization inference methods.
level	the minimum accepted value of the p-value from the covariate balance tests. Default is .15.
reps	number of replications. Default is 1000.

seed	the seed to be used for the randomization tests.
plot	draws a scatter plot of the minimum p-value from the covariate balance test against window length.
quietly	suppress output
obsstep	the minimum number of observations to be added on each side of the cutoff for the sequence of fixed-increment nested windows. This option is deprecated and only included for backward compatibility.

Value

window	recommended window (NA is covariates are not specified)
wlist	list of window lengths
results	table including window lengths, minimum p-value in each window, corresponding number of the variable with minimum p-value (i.e. column of covariate matrix), Binomial test p-value and sample sizes to the left and right of the cutoff in each window.
summary	summary statistics.

Author(s)

Matias Cattaneo, Princeton University. `<cattaneo@princeton.edu>`

Rocio Titiunik, Princeton University. `<titiunik@princeton.edu>`

Gonzalo Vazquez-Bare, UC Santa Barbara. `<gvazquez@econ.ucsb.edu>`

References

Cattaneo, M.D., R. Titiunik and G. Vazquez-Bare. (2016). Inference in Regression Discontinuity Designs under Local Randomization. *Stata Journal* 16(2): 331-367.

Examples

```
# Toy dataset
X <- array(rnorm(200),dim=c(100,2))
R <- X[1,] + X[2,] + rnorm(100)
# Window selection adding 5 observations at each step
# Note: low number of replications to speed up process.
tmp <- rdwinselect(R,X,obsmin=10,wobs=5,reps=500)
# Window selection setting initial window and step
# The user should increase the number of replications.
tmp <- rdwinselect(R,X,wmin=.5,wstep=.125,reps=500)
# Window selection with approximate (large sample) inference and p-value plot
tmp <- rdwinselect(R,X,wmin=.5,wstep=.125,approx=TRUE,nwin=80,quietly=TRUE,plot=TRUE)
```

Index