Package 'rdrobust'

May 16, 2026

Type Package

Title Robust Data-Driven Statistical Inference in Regression-Discontinuity Designs

Version 4.0.0

Date 2026-05-15

Description Regression-discontinuity (RD) designs are quasi-experimental research designs popular in social, behavioral and natural sciences. The RD design is usually employed to study the (local) causal effect of a treatment, intervention or policy. This package provides tools for data-driven graphical and analytical statistical inference in RD designs: rdrobust() to construct local-polynomial point estimators and robust confidence intervals for average treatment effects at the cutoff in Sharp, Fuzzy and Kink RD settings, rdbwselect() to perform bandwidth selection for the different procedures implemented, and rdplot() to conduct exploratory data analysis (RD plots).

Encoding UTF-8

Depends R (>= 3.6.0)

License GPL-3

URL https://github.com/rdpackages/rdrobust

BugReports https://github.com/rdpackages/rdrobust/issues

Imports ggplot2, MASS

Suggests broom, gridExtra, knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Sebastian Calonico [aut, cre],
Matias D. Cattaneo [aut],
Max H. Farrell [aut],
Rocio Titiunik [aut]

Maintainer Sebastian Calonico <scalonico@ucdavis.edu>

Repository CRAN

Date/Publication 2026-05-16 08:10:02 UTC

Contents

`rdrobust-package` *Robust Data-Driven Statistical Inference in RD Designs*

Description

Regression-discontinuity (RD) designs are quasi-experimental research designs popular in social, behavioral and natural sciences. The RD design is usually employed to study the (local) causal effect of a treatment, intervention or policy. This package provides tools for data-driven graphical and analytical statistical inference in RD designs: `rdrobust` to construct local-polynomial point estimators and robust confidence intervals for average treatment effects at the cutoff in Sharp, Fuzzy and Kink RD settings, `rdbwselect` to perform bandwidth selection for the different procedures implemented, and `rdplot` to conduct exploratory data analysis (RD plots).

Details

Package:	rdrobust
Type:	Package
Version:	4.0.0
Date:	2026-05-15
License:	GPL-3

Function for statistical inference: `rdrobust`
Function for bandwidths selection: `rdbwselect`
Function for exploratory data analysis (RD plots): `rdplot`

Author(s)

Sebastian Calonico, University of California, Davis, CA. <scalonico@ucdavis.edu>.

Matias D. Cattaneo, Princeton University, Princeton, NJ. <matias.d.cattaneo@gmail.com>.

Max H. Farrell, University of California, Santa Barbara, CA. <mhfarrell@gmail.com>.

Rocio Titiunik, Princeton University, Princeton, NJ. <rocio.titiunik@gmail.com>.

plot.rdrobust *Plot Method for rdrobust Objects*

Description

Produces a visual summary of an rdrobust result. The main panel shows a binned scatter plot of the outcome against the running variable within the estimated bandwidth, overlaid with the local-polynomial fit curves and (optionally) pointwise confidence bands. Bin sizes are scaled to the average kernel weight of the observations they contain, so bins closest to the cutoff appear largest. An optional effect panel below the main plot displays the conventional point estimate with its robust bias-corrected confidence interval and significance stars.

Usage

```
## S3 method for class 'rdrobust'
plot(x, y, x_run, nbins = 20,
     ci = TRUE, show_effect = FALSE,
     title = NULL,
     x.label = "Running Variable",
     y.label = "Outcome",
     x.lim = NULL, y.lim = NULL,
     col.l = "#3B7DD8", col.r = "#D95F3B",
     base_size = 14, ...)
```

Arguments

x	an object of class "rdrobust", as returned by rdrobust.
y	numeric vector of outcome values (same length and order as used in the original rdrobust call, before any subset was applied).
x_run	numeric vector of running variable values (same length and order as y).
nbins	number of bins per side used for the binned scatter plot. Default is 20.
ci	logical. If TRUE (default), shaded pointwise confidence bands are drawn around each polynomial fit curve. The bands use the conventional variance-covariance matrices V_cl_l and V_cl_r stored in the rdrobust object, evaluated at the confidence level stored in the rdrobust object.
show_effect	logical. If TRUE, a second panel is printed below the main plot showing the conventional point estimate with its robust bias-corrected (RBC) confidence interval as a horizontal error bar, together with significance stars based on the robust p-value. Combining both panels requires the **gridExtra** package; if unavailable the two plots are printed sequentially. Default is FALSE.
title	character string for the plot title. Default is NULL (no title).
x.label	label for the horizontal axis. Default is "Running Variable".
y.label	label for the vertical axis. Default is "Outcome".
x.lim	numeric vector of length 2 specifying the horizontal axis limits. Default is NULL (automatic).

y.lim	numeric vector of length 2 specifying the vertical axis limits. Default is NULL (automatic).
col.l	color for the left-side (control) elements. Default is "#3B7DD8".
col.r	color for the right-side (treated) elements. Default is "#D95F3B".
base_size	base font size in points passed to the underlying **ggplot2** theme. All text elements scale relative to this value. Default is 14.
...	additional arguments (currently unused).

Details

Only observations within the main estimation bandwidth [c - h_l, c + h_r] are shown. Each side is divided into nbins equal-width bins; the plotted point for each bin is the bin mean of both x_run and y. The size of each bin's dot is proportional to the average kernel weight of the observations it contains, computed from the kernel stored in x$kernel (triangular, Epanechnikov, or uniform) evaluated at the normalized distance (x_i - c) / h.

The polynomial fit curve on each side is the fitted conditional mean implied by the local-polynomial coefficients beta_Y_p_l and beta_Y_p_r. The pointwise confidence bands (ci = TRUE) are computed as

$$\hat{\mu}(t) \pm z_{\alpha/2}\sqrt{\mathbf{x}(t)^\top V \mathbf{x}(t)}$$

where $\mathbf{x}(t) = (1, t-c, \ldots, (t-c)^p)^\top$ and V is the conventional variance-covariance matrix of the polynomial coefficients.

When show_effect = TRUE the effect panel reports:

- Point estimate: conventional (coef["Conventional"]).
- Interval: robust bias-corrected (ci["Robust",]).
- Stars: based on pv["Robust",]: * p<0.10, ** p<0.05, *** p<0.01.

Value

Invisibly returns the **ggplot2** plot object when show_effect = FALSE, or a named list list(rd_plot, effect_plot) of two **ggplot2** objects when show_effect = TRUE. In both cases the plot(s) are printed as a side-effect.

Author(s)

Sebastian Calonico, University of California, Davis, CA. <scalonico@ucdavis.edu>.

Matias D. Cattaneo, Princeton University, Princeton, NJ. <matias.d.cattaneo@gmail.com>.

Max H. Farrell, University of California, Santa Barbara, CA. <mhfarrell@gmail.com>.

Rocio Titiunik, Princeton University, Princeton, NJ. <rocio.titiunik@gmail.com>.

References

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6): 2295-2326.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2019. Regression Discontinuity Designs using Covariates. *Review of Economics and Statistics*, 101(3): 442-451.

See Also

rdrobust, rdbwselect, rdplot

Examples

```
x <- runif(500, -1, 1)
y <- 5 + 3*x + 2*(x >= 0) + rnorm(500)
r <- rdrobust(y, x)

## Main RD plot with CI bands
plot(r, y, x)

## Add effect panel
plot(r, y, x, show_effect = TRUE)
```

---

rdbwselect *Bandwidth Selection Procedures for Local Polynomial Regression Discontinuity Estimators*

Description

rdbwselect implements bandwidth selectors for local polynomial Regression Discontinuity (RD) point estimators and inference procedures developed in Calonico, Cattaneo and Titiunik (2014a), Calonico, Cattaneo and Farrell (2018), Calonico, Cattaneo, Farrell and Titiunik (2019) and Calonico, Cattaneo and Farrell (2020).

Companion commands are: rdrobust for point estimation and inference procedures, and rdplot for data-driven RD plots (see Calonico, Cattaneo and Titiunik (2015a) for details).

A detailed introduction to this command is given in Calonico, Cattaneo and Titiunik (2015b) and Calonico, Cattaneo, Farrell and Titiunik (2019). A companion Stata package is described in Calonico, Cattaneo and Titiunik (2014b).

For more details, and related Stata and R packages useful for analysis of RD designs, visit https://rdpackages.github.io/

Usage

```
rdbwselect(y, x, c = NULL,  fuzzy = NULL,
              deriv = NULL, p = NULL, q = NULL,
              covs = NULL, covs_drop = TRUE, ginv.tol = 1e-20,
              kernel = "tri", weights = NULL, bwselect = "mserd",
              vce = "nn",  cluster = NULL, nnmatch = 3,
              scaleregul = 1, sharpbw = FALSE,
              all = NULL, subset = NULL,
              masspoints = "adjust", bwcheck = NULL,
              bwrestrict = TRUE, stdvars = FALSE, data = NULL)
```

Arguments

y	is the dependent variable.
x	is the running variable (a.k.a. score or forcing variable).
c	specifies the RD cutoff in x; default is c = 0.
fuzzy	specifies the treatment status variable used to implement fuzzy RD estimation (or Fuzzy Kink RD if deriv=1 is also specified). Default is Sharp RD design and hence this option is not used.
deriv	specifies the order of the derivative of the regression functions to be estimated. Default is deriv=0 (for Sharp RD, or for Fuzzy RD if fuzzy is also specified). Setting deriv=1 results in estimation of a Kink RD design (up to scale), or Fuzzy Kink RD if fuzzy is also specified.
p	specifies the order of the local-polynomial used to construct the point-estimator; default is p = 1 (local linear regression).
q	specifies the order of the local-polynomial used to construct the bias-correction; default is q = 2 (local quadratic regression).
covs	specifies additional covariates to be used for estimation and inference. One of: • a one-sided formula, e.g. ~ z1 + z2 + factor(g) + I(z3^2): processed with model.matrix, so factors are expanded to contrasts, interactions are supported, and transformations such as I(...), log(), and poly() work. The intercept column is dropped automatically. Symbols are looked up in data first (if supplied), then in the caller's environment; • a character vector of column names (requires data =), e.g. c("z1", "z2"): selected as data[, covs] and coerced to a matrix. Useful for programmatic specifications; • a numeric vector, matrix, or data frame: passed through unchanged (backwards compatible).
covs_drop	if TRUE, it checks for collinear additional covariates and drops them. Default is TRUE.
ginv.tol	tolerance used to invert matrices involving covariates when covs_drop=TRUE.
kernel	is the kernel function used to construct the local-polynomial estimator(s). Options are triangular (default option), epanechnikov and uniform.
weights	is the variable used for optional weighting of the estimation procedure. The unit-specific weights multiply the kernel function.
bwselect	specifies the bandwidth selection procedure to be used. Options are: mserd one common MSE-optimal bandwidth selector for the RD treatment effect estimator. msetwo two different MSE-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator. msesum one common MSE-optimal bandwidth selector for the sum of regression estimates (as opposed to difference thereof). msecomb1 for min(mserd,msesum). msecomb2 for median(msetwo,mserd,msesum), for each side of the cutoff separately.

cerrd one common CER-optimal bandwidth selector for the RD treatment effect estimator.

certwo two different CER-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator.

cersum one common CER-optimal bandwidth selector for the sum of regression estimates (as opposed to difference thereof).

cercomb1 for min(cerrd,cersum).

cercomb2 for median(certwo,cerrd,cersum), for each side of the cutoff separately.

Note: MSE = Mean Square Error; CER = Coverage Error Rate. Default is bwselect=mserd. For details on implementation see Calonico, Cattaneo and Titiunik (2014a), Calonico, Cattaneo and Farrell (2018), and Calonico, Cattaneo, Farrell and Titiunik (2017), and the companion software articles.

vce specifies the procedure used to compute the variance-covariance matrix estimator. Options are:

nn for heteroskedasticity-robust nearest neighbor variance estimator with nnmatch the (minimum) number of neighbors to be used.

hc0 for heteroskedasticity-robust plug-in residuals variance estimator without weights.

hc1 for heteroskedasticity-robust plug-in residuals variance estimator with hc1 weights.

hc2 for heteroskedasticity-robust plug-in residuals variance estimator with hc2 weights.

hc3 for heteroskedasticity-robust plug-in residuals variance estimator with hc3 weights.

cr1 for CR1 cluster-robust variance estimator with degrees-of-freedom correction at the cluster level. Requires cluster to be specified.

cr2 for CR2 cluster-robust variance estimator with leverage adjustments (Bell-McCaffrey). Requires cluster to be specified.

cr3 for CR3 cluster-robust variance estimator with leave-one-cluster-out jackknife (Pustejovsky-Tipton); most conservative. Requires cluster to be specified.

Default is vce=nn. When cluster is specified, only vce=cr1, vce=cr2, and vce=cr3 are valid; other options are automatically mapped to the corresponding cluster variant with a warning.

The CR2/CR3 leverage correction applies to both the conventional and the robust bias-corrected standard errors, including when the point-estimation bandwidth h differs from the bias-correction bandwidth b; in that case the cluster leverage is computed from the bias (b) regression.

cluster indicates the cluster ID variable used for cluster-robust variance estimation. Must be combined with vce=cr1, vce=cr2, or vce=cr3.

nnmatch to be combined with for vce=nn for heteroskedasticity-robust nearest neighbor variance estimator with nnmatch indicating the minimum number of neighbors to be used. Default is nnmatch=3

scaleregul	specifies scaling factor for the regularization term added to the denominator of the bandwidth selectors. Setting scaleregul = 0 removes the regularization term from the bandwidth selectors; default is scaleregul = 1.
sharpbw	option to perform fuzzy RD estimation using a bandwidth selection procedure for the sharp RD model. This option is automatically selected if there is perfect compliance at either side of the threshold.
all	if specified, rdbwselect reports all available bandwidth selection procedures.
subset	an optional vector specifying a subset of observations to be used.
masspoints	checks and controls for repeated observations in the running variable. Options are: (i) off: ignores the presence of mass points; (ii) check: looks for and reports the number of unique observations at each side of the cutoff. (iii) adjust: controls that the preliminary bandwidths used in the calculations contain a minimal number of unique observations. By default it uses 10 observations, but it can be manually adjusted with the option bwcheck. Default option is masspoints=adjust.
bwcheck	if a positive integer is provided, the preliminary bandwidth used in the calculations is enlarged so that at least bwcheck unique observations are used.
bwrestrict	if TRUE, computed bandwidths are restricted to lie within the range of x; default is bwrestrict = TRUE.
stdvars	if TRUE, x and y are standardized before computing the bandwidths; default is stdvars = FALSE.
data	an optional data frame. When supplied, y, x, covs, cluster, fuzzy, weights, and subset may be given as bare variable names referring to columns of data.

Value

N	vector with sample sizes to the left and to the right of the cutoff.
N_h	vector with effective sample sizes to the left and to the right of the cutoff, computed at the selected bandwidth h.
M	vector with the number of unique observations to the left and to the right of the cutoff (when masspoints is not off).
c	cutoff value.
p	order of the local-polynomial used to construct the point-estimator.
q	order of the local-polynomial used to construct the bias-correction estimator.
bws	matrix containing the estimated bandwidths for each selected procedure. Columns are h (left), h (right), b (left), b (right); rows correspond to bandwidth selection methods.
bwselect	bandwidth selection procedure employed. The bandwidth methods reported in bws are available as rownames(bws).
kernel	kernel function used to construct the local-polynomial estimator(s).
vce	variance estimation method used.
masspoints	mass-points option used.

Author(s)

Sebastian Calonico, University of California, Davis, CA. <scalonico@ucdavis.edu>.

Matias D. Cattaneo, Princeton University, Princeton, NJ. <matias.d.cattaneo@gmail.com>.

Max H. Farrell, University of California, Santa Barbara, CA. <mhfarrell@gmail.com>.

Rocio Titiunik, Princeton University, Princeton, NJ. <rocio.titiunik@gmail.com>.

References

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2018. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, 113(522): 767-779.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2020. Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs. *Econometrics Journal*, 23(2): 192-210.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2017. rdrobust: Software for Regression Discontinuity Designs. *Stata Journal* 17(2): 372-404.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2019. Regression Discontinuity Designs using Covariates. *Review of Economics and Statistics*, 101(3): 442-451.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6): 2295-2326.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. Robust Data-Driven Inference in the Regression-Discontinuity Design. *Stata Journal* 14(4): 909-946.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015a. Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association* 110(512): 1753-1769.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015b. rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. *R Journal* 7(1): 38-51.

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

See Also

`rdrobust`, `rdplot`, `plot.rdrobust`

Examples

```
x <- runif(1000,-1,1)
y <- 5 + 3*x + 2*(x>=0) + rnorm(1000)
rdbwselect(y,x)
```

rdplot *Data-Driven Regression Discontinuity Plots*

Description

rdplot implements several data-driven Regression Discontinuity (RD) plots, using either evenly-spaced or quantile-spaced partitioning. Two types of RD plots are constructed: (i) RD plots with binned sample means tracing out the underlying regression function, and (ii) RD plots with binned sample means mimicking the underlying variability of the data. For technical and methodological details see Calonico, Cattaneo and Titiunik (2015a).

Companion commands are: rdrobust for point estimation and inference procedures, and rdbwselect for data-driven bandwidth selection.

A detailed introduction to this command is given in Calonico, Cattaneo and Titiunik (2015b) and Calonico, Cattaneo, Farrell and Titiunik (2017). A companion Stata package is described in Calonico, Cattaneo and Titiunik (2014).

For more details, and related Stata and R packages useful for analysis of RD designs, visit https://rdpackages.github.io/

Usage

```
rdplot(y, x, c = 0, p = 4, nbins = NULL, binselect = "esmv",
          scale = NULL, kernel = "uni", weights = NULL, h = NULL,
          covs = NULL, covs_eval = "mean", covs_drop = TRUE, ginv.tol = 1e-20,
          support = NULL, subset = NULL, masspoints = "adjust",
          hide = FALSE, ci = NULL, shade = FALSE, title = NULL,
          x.label = NULL, y.label = NULL, x.lim = NULL, y.lim = NULL,
          col.dots = NULL, col.lines = NULL, data = NULL)
```

Arguments

y	is the dependent variable.
x	is the running variable (a.k.a. score or forcing variable).
c	specifies the RD cutoff in x; default is c = 0.
p	specifies the order of the global-polynomial used to approximate the population conditional mean functions for control and treated units; default is p = 4.
nbins	specifies the number of bins used to the left of the cutoff, denoted J_-, and to the right of the cutoff, denoted J_+, respectively. If not specified, J_+ and J_- are estimated using the method and options chosen below.
binselect	specifies the procedure to select the number of bins. This option is available only if J_- and J_+ are not set manually. Options are: es: IMSE-optimal evenly-spaced method using spacings estimators. espr: IMSE-optimal evenly-spaced method using polynomial regression. esmv: mimicking variance evenly-spaced method using spacings estimators. This is the default option.

esmvpr: mimicking variance evenly-spaced method using polynomial regression.

qs: IMSE-optimal quantile-spaced method using spacings estimators.

qspr: IMSE-optimal quantile-spaced method using polynomial regression.

qsmv: mimicking variance quantile-spaced method using spacings estimators.

qsmvpr: mimicking variance quantile-spaced method using polynomial regression.

`scale` specifies a multiplicative factor to be used with the optimal numbers of bins selected. Specifically, the number of bins used for the treatment and control groups will be `scale`$\times\hat{J}_+$ and `scale`$\times\hat{J}_-$, where $\hat{J}_\cdot$ denotes the estimated optimal numbers of bins originally computed for each group; default is `scale = 1`.

`kernel` specifies the kernel function used to construct the local-polynomial estimator(s). Options are: `triangular`, `epanechnikov`, and `uniform`. Default is `kernel=uniform` (i.e., equal/no weighting to all observations on the support of the kernel).

`weights` is the variable used for optional weighting of the estimation procedure. The unit-specific weights multiply the kernel function.

`h` specifies the bandwidth used to construct the (global) polynomial fits given the kernel choice `kernel`. If not specified, the bandwidths are chosen to span the full support of the data. If two bandwidths are specified, the first bandwidth is used for the data below the cutoff and the second bandwidth is used for the data above the cutoff.

`covs` specifies additional covariates to be used in the polynomial regression. One of:

- a one-sided `formula`, e.g. `~ z1 + z2 + factor(g) + I(z3^2)`: processed with `model.matrix`, so factors are expanded to contrasts, interactions are supported, and transformations such as `I(...)`, `log()`, and `poly()` work. The intercept column is dropped automatically. Symbols are looked up in `data` first (if supplied), then in the caller's environment;
- a character vector of column names (requires `data =`), e.g. `c("z1", "z2")`: selected as `data[, covs]` and coerced to a matrix. Useful for programmatic specifications;
- a numeric vector, matrix, or data frame: passed through unchanged (backwards compatible).

`covs_eval` sets the evaluation points for the additional covariates, when included in the estimation. Options are: `covs_eval = 0` and `covs_eval = "mean"` (default)

`covs_drop` if TRUE, it checks for collinear additional covariates and drops them. Default is TRUE.

`ginv.tol` tolerance used to invert matrices involving covariates when `covs_drop=TRUE`.

`support` specifies an optional extended support of the running variable to be used in the construction of the bins; default is the sample range.

`subset` an optional vector specifying a subset of observations to be used.

`masspoints` checks and controls for repeated observations in the running variable. Options are:

(i) `off`: ignores the presence of mass points;

	(ii) check: looks for and reports the number of unique observations at each side of the cutoff. (iii) adjust: sets binselect() as polynomial regression when mass points are present. Default option is masspoints=adjust.
hide	logical. If TRUE, it omits the RD plot; default is hide = FALSE.
ci	optional graphical option to display confidence intervals of selected level for each bin.
shade	optional graphical option to replace confidence intervals with shaded areas.
title	optional title for the RD plot.
x.label	optional label for the x-axis of the RD plot.
y.label	optional label for the y-axis of the RD plot.
x.lim	optional setting for the range of the x-axis in the RD plot.
y.lim	optional setting for the range of the y-axis in the RD plot.
col.dots	optional setting for the color of the dots in the RD plot.
col.lines	optional setting for the color of the lines in the RD plot.
data	an optional data frame. When supplied, y, x, covs, weights, and subset may be given as bare variable names referring to columns of data.

Value

binselect	method used to compute the optimal number of bins.
N	sample sizes used to the left and right of the cutoff.
N_h	effective sample sizes used to the left and right of the cutoff.
c	cutoff value.
p	order of the global polynomial used.
h	bandwidth used to the left and right of the cutoff.
kernel	kernel used.
J	selected number of bins to the left and right of the cutoff.
J_IMSE	IMSE optimal number of bins to the left and right of the cutoff.
J_MV	Mimicking variance number of bins to the left and right of the cutoff.
coef	matrix containing the coefficients of the p^{th} order global polynomial estimated both sides of the cutoff.
coef_covs	coefficients of the additional covariates, only returned when covs() are used.
scale	selected scale value.
rscale	implicit scale value.
bin_avg	average bin length.
bin_med	median bin length.
vars_bins	data frame containing the variables used to construct the bins: bin id, cutoff values, mean of x and y within each bin, cutoff points and confidence interval bounds.
vars_poly	data frame containing the variables used to construct the global polynomial plot.
rdplot	a standard ggplot object that can be used for further customization.

Author(s)

Sebastian Calonico, University of California, Davis, CA. <scalonico@ucdavis.edu>.

Matias D. Cattaneo, Princeton University, Princeton, NJ. <matias.d.cattaneo@gmail.com>.

Max H. Farrell, University of California, Santa Barbara, CA. <mhfarrell@gmail.com>.

Rocio Titiunik, Princeton University, Princeton, NJ. <rocio.titiunik@gmail.com>.

References

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2017. rdrobust: Software for Regression Discontinuity Designs. *Stata Journal* 17(2): 372-404.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. Robust Data-Driven Inference in the Regression-Discontinuity Design. *Stata Journal* 14(4): 909-946.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015a. Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association* 110(512): 1753-1769.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015b. rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. *R Journal* 7(1): 38-51.

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

See Also

rdbwselect, rdrobust, plot.rdrobust

Examples

```
x <- runif(1000,-1,1)
y <- 5 + 3*x + 2*(x>=0) + rnorm(1000)
rdplot(y,x)
```

---

rdrobust *Local-Polynomial RD Estimation with Robust Confidence Intervals*

Description

rdrobust implements local polynomial Regression Discontinuity (RD) point estimators with robust bias-corrected confidence intervals and inference procedures developed in Calonico, Cattaneo and Titiunik (2014a), Calonico, Cattaneo and Farrell (2018), Calonico, Cattaneo, Farrell and Titiunik (2019), and Calonico, Cattaneo and Farrell (2020). It also computes alternative estimation and inference procedures available in the literature.

Companion commands are: rdbwselect for data-driven bandwidth selection, and rdplot for data-driven RD plots (see Calonico, Cattaneo and Titiunik (2015a) for details).

A detailed introduction to this command is given in Calonico, Cattaneo and Titiunik (2015b), and Calonico, Cattaneo, Farrell and Titiunik (2017). A companion Stata package is described in Calonico, Cattaneo and Titiunik (2014b).

For more details, and related Stata and R packages useful for analysis of RD designs, visit https://rdpackages.github.io/

Usage

```
rdrobust(y, x, c = NULL,  fuzzy = NULL,
          deriv = NULL, p = NULL, q = NULL,
          h = NULL, b = NULL, rho = NULL,
          covs = NULL,  covs_drop = TRUE, ginv.tol = 1e-20,
          kernel = "tri", weights = NULL, bwselect = "mserd",
          vce = "nn", cluster = NULL,
          nnmatch = 3, level = 95, scalepar = 1, scaleregul = 1,
          sharpbw = FALSE, subset = NULL,
          masspoints = "adjust", bwcheck = NULL,
          bwrestrict = TRUE, stdvars = FALSE, data = NULL)
```

Arguments

y	is the dependent variable.
x	is the running variable (a.k.a. score or forcing variable).
c	specifies the RD cutoff in x; default is c = 0.
fuzzy	specifies the treatment status variable used to implement fuzzy RD estimation (or Fuzzy Kink RD if deriv=1 is also specified). Default is Sharp RD design and hence this option is not used.
deriv	specifies the order of the derivative of the regression functions to be estimated. Default is deriv=0 (for Sharp RD, or for Fuzzy RD if fuzzy is also specified). Setting deriv=1 results in estimation of a Kink RD design (up to scale), or Fuzzy Kink RD if fuzzy is also specified.
p	specifies the order of the local-polynomial used to construct the point-estimator; default is p = 1 (local linear regression).
q	specifies the order of the local-polynomial used to construct the bias-correction; default is q = 2 (local quadratic regression).
h	specifies the main bandwidth used to construct the RD point estimator. If not specified, bandwidth h is computed by the companion command rdbwselect. If two bandwidths are specified, the first bandwidth is used for the data below the cutoff and the second bandwidth is used for the data above the cutoff.
b	specifies the bias bandwidth used to construct the bias-correction estimator. If not specified, bandwidth b is computed by the companion command rdbwselect. If two bandwidths are specified, the first bandwidth is used for the data below the cutoff and the second bandwidth is used for the data above the cutoff.
rho	specifies the value of rho, so that the bias bandwidth b equals h/rho. Default is rho = 1 if h is specified but b is not.
covs	additional covariates to be used for efficiency improvements. One of: • a one-sided formula, e.g. ~ z1 + z2 + factor(g) + I(z3^2): processed with model.matrix, so factors are expanded to contrasts, interactions are supported, and transformations such as I(...), log(), and poly() work.

The intercept column is dropped automatically. Symbols are looked up in `data` first (if supplied), then in the caller's environment;

- a character vector of column names (requires `data =`), e.g. `c("z1", "z2")`: selected as `data[, covs]` and coerced to a matrix. Useful for programmatic specifications;
- a numeric vector, matrix, or data frame: passed through unchanged (backwards compatible).

`covs_drop` if TRUE, it checks for collinear additional covariates and drops them. Default is TRUE.

`ginv.tol` tolerance used to invert matrices involving covariates when `covs_drop=TRUE`.

`kernel` is the kernel function used to construct the local-polynomial estimator(s). Options are `triangular` (default option), `epanechnikov` and `uniform`.

`weights` is the variable used for optional weighting of the estimation procedure. The unit-specific weights multiply the kernel function.

`bwselect` specifies the bandwidth selection procedure to be used. By default it computes both `h` and `b`, unless `rho` is specified, in which case it only computes `h` and sets `b=h/rho`.

Options are:

`mserd` one common MSE-optimal bandwidth selector for the RD treatment effect estimator.

`msetwo` two different MSE-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator.

`msesum` one common MSE-optimal bandwidth selector for the sum of regression estimates (as opposed to difference thereof).

`msecomb1` for min(`mserd`,`msesum`).

`msecomb2` for median(`msetwo`,`mserd`,`msesum`), for each side of the cutoff separately.

`cerrd` one common CER-optimal bandwidth selector for the RD treatment effect estimator.

`certwo` two different CER-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator.

`cersum` one common CER-optimal bandwidth selector for the sum of regression estimates (as opposed to difference thereof).

`cercomb1` for min(`cerrd`,`cersum`).

`cercomb2` for median(`certwo`,`cerrd`,`cersum`), for each side of the cutoff separately.

Note: MSE = Mean Square Error; CER = Coverage Error Rate. Default is `bwselect=mserd`. For details on implementation see Calonico, Cattaneo and Titiunik (2014a), Calonico, Cattaneo and Farrell (2018), and Calonico, Cattaneo, Farrell and Titiunik (2019), and the companion software articles.

`vce` specifies the procedure used to compute the variance-covariance matrix estimator. Options are:

`nn` for heteroskedasticity-robust nearest neighbor variance estimator with `nnmatch` the (minimum) number of neighbors to be used.

`hc0` for heteroskedasticity-robust plug-in residuals variance estimator without weights.

hc1 for heteroskedasticity-robust plug-in residuals variance estimator with hc1 weights.

hc2 for heteroskedasticity-robust plug-in residuals variance estimator with hc2 weights.

hc3 for heteroskedasticity-robust plug-in residuals variance estimator with hc3 weights.

cr1 for CR1 cluster-robust variance estimator with degrees-of-freedom correction at the cluster level. Requires cluster to be specified.

cr2 for CR2 cluster-robust variance estimator with leverage adjustments (Bell-McCaffrey). Requires cluster to be specified.

cr3 for CR3 cluster-robust variance estimator with leave-one-cluster-out jackknife (Pustejovsky-Tipton); most conservative. Requires cluster to be specified.

Default is vce=nn. When cluster is specified, only vce=cr1, vce=cr2, and vce=cr3 are valid; other options are automatically mapped to the corresponding cluster variant with a warning.

The CR2/CR3 leverage correction applies to both the conventional and the robust bias-corrected standard errors, including when the point-estimation bandwidth h differs from the bias-correction bandwidth b; in that case the cluster leverage is computed from the bias (b) regression.

`cluster` indicates the cluster ID variable used for cluster-robust variance estimation. Must be combined with vce=cr1, vce=cr2, or vce=cr3.

`nnmatch` to be combined with for vce=nn for heteroskedasticity-robust nearest neighbor variance estimator with nnmatch indicating the minimum number of neighbors to be used. Default is nnmatch=3

`level` sets the confidence level for confidence intervals; default is level = 95.

`scalepar` specifies scaling factor for RD parameter of interest. This option is useful when the population parameter of interest involves a known multiplicative factor (e.g., sharp kink RD). Default is scalepar = 1 (no scaling).

`scaleregul` specifies scaling factor for the regularization term added to the denominator of the bandwidth selectors. Setting scaleregul = 0 removes the regularization term from the bandwidth selectors; default is scaleregul = 1.

`sharpbw` option to perform fuzzy RD estimation using a bandwidth selection procedure for the sharp RD model. This option is automatically selected if there is perfect compliance at either side of the cutoff.

`subset` an optional vector specifying a subset of observations to be used.

`masspoints` checks and controls for repeated observations in the running variable. Options are:

(i) off: ignores the presence of mass points;

(ii) check: looks for and reports the number of unique observations at each side of the cutoff.

(iii) adjust: controls that the preliminary bandwidths used in the calculations contain a minimal number of unique observations. By default it uses 10 observations, but it can be manually adjusted with the option bwcheck.

Default option is masspoints=adjust.

bwcheck	if a positive integer is provided, the preliminary bandwidth used in the calculations is enlarged so that at least bwcheck unique observations are used.
bwrestrict	if TRUE, computed bandwidths are restricted to lie within the range of x; default is bwrestrict = TRUE.
stdvars	if TRUE, x and y are standardized before computing the bandwidths; default is stdvars = FALSE.
data	an optional data frame. When supplied, y, x, covs, cluster, fuzzy, weights, and subset may be given as bare variable names referring to columns of data.

Value

N	vector with the sample sizes used to the left and to the right of the cutoff.
N_h	vector with the effective sample sizes used to the left and to the right of the cutoff.
N_b	vector with the effective sample sizes used to the left and to the right of the cutoff for bias estimation.
M	vector with the number of unique observations to the left and to the right of the cutoff (when masspoints is not off).
c	cutoff value.
p	order of the polynomial used for estimation of the regression function.
q	order of the polynomial used for estimation of the bias of the regression function.
bws	matrix containing the bandwidths used: row h for the estimation bandwidth, row b for the bias bandwidth; columns left and right.
tau_cl	conventional local-polynomial estimate to the left and to the right of the cutoff.
tau_bc	bias-corrected local-polynomial estimate to the left and to the right of the cutoff.
coef	vector containing conventional and bias-corrected local-polynomial RD estimates.
se	vector containing conventional and robust standard errors of the local-polynomial RD estimates.
bias	estimated bias for the local-polynomial RD estimator below and above the cutoff.
beta_Y_p_l	conventional p-order local-polynomial estimates to the left of the cutoff for the outcome variable.
beta_Y_p_r	conventional p-order local-polynomial estimates to the right of the cutoff for the outcome variable.
beta_T_p_l	conventional p-order local-polynomial estimates to the left of the cutoff for the first stage (fuzzy RD).
beta_T_p_r	conventional p-order local-polynomial estimates to the right of the cutoff for the first stage (fuzzy RD).
coef_covs	coefficients of the additional covariates, only returned when covs is specified.
V_cl_l	conventional variance-covariance matrix estimated below the cutoff.
V_cl_r	conventional variance-covariance matrix estimated above the cutoff.

V_rb_l	robust variance-covariance matrix estimated below the cutoff.
V_rb_r	robust variance-covariance matrix estimated above the cutoff.
z	vector containing the z-statistics associated with conventional, bias-corrected and robust local-polynomial RD estimates.
pv	vector containing the p-values associated with conventional, bias-corrected and robust local-polynomial RD estimates.
ci	matrix containing the confidence intervals associated with conventional, bias-corrected and robust local-polynomial RD estimates.
kernel	kernel function used.
vce	variance estimation method used.
bwselect	bandwidth selection method used.
level	confidence level used.
masspoints	mass points option used.
rdmodel	character string describing the model estimated: design type (sharp/fuzzy/kink), whether covariates were included, and whether standard errors are clustered.
n_clust	vector with the number of clusters to the left and to the right of the cutoff. NULL when cluster is not specified.
tau_T, se_T, z_T, pv_T, ci_T	first-stage estimates, standard errors, z-statistics, p-values, and confidence intervals for the treatment indicator (fuzzy RD only).

Author(s)

Sebastian Calonico, University of California, Davis, CA. <scalonico@ucdavis.edu>.

Matias D. Cattaneo, Princeton University, Princeton, NJ. <matias.d.cattaneo@gmail.com>.

Max H. Farrell, University of California, Santa Barbara, CA. <mhfarrell@gmail.com>.

Rocio Titiunik, Princeton University, Princeton, NJ. <rocio.titiunik@gmail.com>.

References

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2018. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, 113(522): 767-779.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2020. Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs. *Econometrics Journal*, 23(2): 192-210.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2017. rdrobust: Software for Regression Discontinuity Designs. *Stata Journal*, 17(2): 372-404.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2019. Regression Discontinuity Designs using Covariates. *Review of Economics and Statistics*, 101(3): 442-451.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6): 2295-2326.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. Robust Data-Driven Inference in the Regression-Discontinuity Design. *Stata Journal* 14(4): 909-946.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015a. Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association* 110(512): 1753-1769.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015b. rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. *R Journal* 7(1): 38-51.

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

See Also

rdbwselect, rdplot, plot.rdrobust

Examples

```
x <- runif(1000,-1,1)
y <- 5 + 3*x + 2*(x>=0) + rnorm(1000)
rdrobust(y,x)

# Using a data frame with covariates via the `data =` argument
df <- data.frame(y = y, x = x, z1 = rnorm(1000), z2 = rnorm(1000))
rdrobust(y, x, covs = ~ z1 + z2, data = df)
```

---

rdrobust_RDsenate *RD Senate Data*

Description

Extract of the dataset constructed by Cattaneo, Frandsen, and Titiunik (2015), which includes measures of incumbency advantage in the U.S. Senate for the period 1914-2010. The dataset contains the running variable (Democratic vote margin at election t), the main outcome (Democratic vote share at election $t + 2$), additional outcomes and lagged vote shares useful as covariates, and a state identifier suitable for cluster-robust inference.

Usage

```
data(rdrobust_RDsenate)
```

Format

A data frame with 1390 observations on the following 17 variables.

`state` character. U.S. state name; can be used as a cluster variable for cluster-robust variance estimation.

`year` numeric. Election year.

`margin` numeric. Democratic vote margin at election t (running variable); equals `demmv` in the original dataset.

`vote` numeric. Democratic vote share at election $t + 2$ (main outcome); equals `demvoteshfor2` in the original dataset.

`class` numeric. U.S. Senate class (1, 2, or 3).

`termshouse` numeric. Number of terms served in the U.S. House.

`termssenate` numeric. Number of terms served in the U.S. Senate.

`dopen` numeric. Indicator for open-seat election.

`population` numeric. State population.

`presdemvoteshlag1` numeric. Lagged Democratic presidential vote share.

`demvoteshlag1` numeric. Democratic vote share at election t (same cycle as running variable); useful as a covariate.

`demvoteshlag2` numeric. Democratic vote share at election $t - 2$; useful as a covariate.

`demvoteshfor1` numeric. Democratic vote share at election $t + 1$.

`demwinprv1` numeric. Indicator for Democratic win at election $t - 1$.

`demwinprv2` numeric. Indicator for Democratic win at election $t - 2$.

`dmidterm` numeric. Indicator for midterm election year.

`dpresdem` numeric. Indicator for Democratic president in office.

Source

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

References

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

Index