# Package ‘rdrobust’

August 10, 2021

**Type** Package

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

**Version** 1.0.5

**Date** 2021-08-10

**Author** Sebastian Calonico <sebastian.calonico@columbia.edu>, Matias D. Cattaneo <cattaneo@princeton.edu>, Max H. Farrell <max.farrell@chicagobooth.edu>, Rocio Titiunik <titiunik@princeton.edu>

**Maintainer** Sebastian Calonico <sebastian.calonico@columbia.edu>

**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).

**Depends** R (>= 3.1.1)

**License** GPL-2

**Imports** ggplot2, MASS

## R topics documented:

---

`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: | 1.0.5 |
| Date: | 2021-08-10 |
| License: | GPL-2 |

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

## Author(s)

Sebastian Calonico, Columbia University, New York, NY. <sebastian.calonico@columbia.edu>.

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

Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

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

---

`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, prchk=TRUE)
```

## 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. |
| 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:<br>mserd one common MSE-optimal bandwidth selector for the RD treatment effect estimator.<br>msetwo two different MSE-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect estimator.<br>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.

Default is vce=nn.

cluster
: indicates the cluster ID variable used for cluster-robust variance estimation with degrees-of-freedom weights. By default it is combined with vce=nn for cluster-robust nearest neighbor variance estimation. Another option is plug-in residuals combined with vce=hc0.

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.

prchk — internal check function.

## Value

N — vector with sample sizes to the left and to the righst of the cutoff.

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.

bwselect — bandwidth selection procedure employed.

kernel — kernel function used to construct the local-polynomial estimator(s).

## Author(s)

Sebastian Calonico, Columbia University, New York, NY. <sebastian.calonico@columbia.edu>.

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

Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

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

## 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`

### Examples

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

---

## `rdbwselect_2014` *Deprecated Bandwidth Selection Procedures for Local-Polynomial Regression-Discontinuity Estimators.*

---

### Description

`rdbwselect_2014` is a deprecated command implementing three bandwidth selectors for local polynomial Regression Discontinuity (RD) point estimators and inference procedures, as described in Calonico, Cattaneo and Titiunik (2014).

This command is no longer supported or updated, and it is made available only for backward compatibility purposes. Please use `rdbwselect` instead.

The latest version of the rdrobust package includes the following commands: `rdrobust` for point estimation and inference procedures. `rdbwselect` for data-driven bandwidth selection. `rdplot` for data-driven RD plots.

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

### Usage

```
rdbwselect_2014(y, x, subset = NULL, c = 0, p = 1, q = 2, deriv = 0,
               rho = NULL, kernel = "tri", bwselect = "CCT", scaleregul = 1,
               delta = 0.5, cvgrid_min = NULL, cvgrid_max = NULL,
               cvgrid_length = NULL, cvplot = FALSE, vce = "nn", matches = 3,
               all = FALSE, precalc = TRUE )
```

### Arguments

| | |
|---|---|
| `y` | is the dependent variable. |
| `x` | is the running variable (a.k.a. score or forcing variable). |
| `subset` | an optional vector specifying a subset of observations to be used. |
| `c` | specifies the RD cutoff in `x`; default is `c = 0`. |

| | |
|---|---|
| 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). |
| deriv | specifies the order of the derivative of the regression function to be estimated; default is deriv = 0 (Sharp RD, or Fuzzy RD if fuzzy is also specified). Setting it equal to 1 results in estimation of a Kink RD design (or Fuzzy Kink RD if fuzzy is also specified). |
| rho | if specified, sets the pilot bandwidth b equal to h/rho, where h is computed using the method and options chosen below. |
| kernel | is the kernel function used to construct the local-polynomial estimator(s). Options are triangular (default option), epanechnikov and uniform. |
| bwselect | selects 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:<br>CCT for bandwidth selector proposed by Calonico, Cattaneo and Titiunik (2014) (default option).<br>IK for bandwidth selector proposed by Imbens and Kalyanaraman (2012) (only available for Sharp RD design).<br>CV for cross-validation method proposded by Ludwig and Miller (2007) (only available for Sharp RD design). |
| scaleregul | specifies scaling factor for the regularization terms of CCT and IK bandwidth selectors. Setting scaleregul = 0 removes the regularization term from the bandwidth selectors; default is scaleregul = 1. |
| delta | sets the quantile that defines the sample used in the cross-validation procedure. This option is used only if bwselect(CV) is specified; default is delta = 0.5, that is, the median of the control and treated samples. |
| cvgrid_min | sets the minimum value of the bandwidth grid used in the cross-validation procedure. This option is used only if bwselect = "CV" is specified. |
| cvgrid_max | sets the maximum value of the bandwidth grid used in the cross-validation procedure. This option is used only if bwselect = "CV" is specified. |
| cvgrid_length | sets the bin length of the (evenly-spaced) bandwidth grid used in the cross-validation procedure. This option is used only if bwselect = "CV" is specified. |
| cvplot | generates a graph of the CV objective function. This option is used only if bwselect = "CV" is specified. |
| vce | specifies the procedure used to compute the variance-covariance matrix estimator. This option is used only if CCT or IK bankdwith procedures are employed. Options are:<br>nn for nearest-neighbor matches residuals using matches number of matches. This is the default option (with matches = 3, see below).<br>resid for estimated plug-in residuals using h bandwidth. |
| matches | specifies the number of matches in the nearest-neighbor based variance-covariance matrix estimator. This options is used only when nearest-neighbor matches residuals are employed; default is matches = 3. |
| all | if specified, rdbwselect_2014 reports three different procedures:<br>CCT for bandwidth selector proposed by Calonico, Cattaneo and Titiunik (2014).<br>IK for bandwidth selector proposed by Imbens and Kalyanaraman (2012).<br>CV for cross-validation method proposed by Ludwig and Miller (2007). |
| precalc | internal option. |

### Value

| | |
|---|---|
| bws | matrix containing the estimated bandwidths for each selected procedure. |
| bwselect | bandwidth selection procedure employed. |
| kernel | kernel function used to construct the local-polynomial estimator(s). |
| 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. |

### Author(s)

Sebastian Calonico, Columbia University, New York, NY. <sebastian.calonico@columbia.edu>.

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

Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

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

### References

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

### See Also

rdrobust, rdplot

### Examples

```
x<-runif(1000,-1,1)
y<-5+3*x+2*(x>=0)+rnorm(1000)
rdbwselect_2014(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 type 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)
```

## 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:<br>`es`: IMSE-optimal evenly-spaced method using spacings estimators.<br>`espr`: IMSE-optimal evenly-spaced method using polynomial regression.<br>`esmv`: mimicking variance evenly-spaced method using spacings estimators. This is the default option.<br>`esmvpr`: mimicking variance evenly-spaced method using polynomial regression.<br>`qs`: IMSE-optimal quantile-spaced method using spacings estimators.<br>`qspr`: IMSE-optimal quantile-spaced method using polynomial regression.<br>`qsmv`: mimicking variance quantile-spaced method using spacings estimators.<br>`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 $\hat{J}.$ 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. |

| | |
|---|---|
| covs_eval | sets the evaluation points for the additional covariates, when included in the estimation. Options are: covs_eval = 0 (default) and covs_eval = "mean" |
| 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:<br>(i) off: ignores the presence of mass points;<br>(ii) check: looks for and reports the number of unique observations at each side of the cutoff.<br>(iii) adjust: sets binselect() as polynomial regression when mass points are present.<br>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. |

## Value

| | |
|---|---|
| binselect | method used to compute the optimal number of bins. |
| N | sample sizes used to the left and right of the cutoff. |
| Nh | 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. |
| 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, Columbia University, New York, NY. <sebastian.calonico@columbia.edu>.

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

Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

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

## 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

## 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, all = NULL, subset = NULL,
          masspoints = "adjust", bwcheck = NULL,
          bwrestrict = TRUE, stdvars = FALSE)
```

## 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` specifies additional covariates to be used for estimation and inference.
- `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.<br>hc3 for heteroskedasticity-robust plug-in residuals variance estimator with hc3 weights.<br>Default is vce=nn. |
| cluster | indicates the cluster ID variable used for cluster-robust variance estimation with degrees-of-freedom weights. By default it is combined with vce=nn for cluster-robust nearest neighbor variance estimation. Another option is plug-in residuals combined with vce=hc0. |
| 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. |
| all | if specified, rdrobust reports three different procedures:<br>(i) conventional RD estimates with conventional standard errors.<br>(ii) bias-corrected estimates with conventional standard errors.<br>(iii) bias-corrected estimates with robust standard errors. |
| 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:<br>(i) off: ignores the presence of mass points;<br>(ii) check: looks for and reports the number of unique observations at each side of the cutoff.<br>(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).<br>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. |

## 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. |

| | |
|---|---|
| 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. |
| 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_p_l | conventional p-order local-polynomial estimates to the left of the cutoff. |
| beta_p_r | conventional p-order local-polynomial estimates to the right of the cutoff. |
| 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. |
| 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. |

## Author(s)

Sebastian Calonico, Columbia University, New York, NY. <sebastian.calonico@columbia.edu>.

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

Max H. Farrell, University of Chicago, Chicago, IL. <max.farrell@chicagobooth.edu>.

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

## 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

### Examples

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

---

## rdrobust_RDsenate *RD Senate Data*

---

### Description

Extract of the dataset constructed by Cattaneo, Frandsen, and Titiunik (2015), which include measures of incumbency advantage in the U.S. Senate for the period 1914-2010.

### Usage

```
data(rdrobust_RDsenate)
```

### Format

A data frame with 1390 observations on the following 2 variables.

`margin` a numeric vector.

`vote` a numeric vector.

### 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