



## Title

**rdrobust** — Local Polynomial Regression Discontinuity Estimation with Robust Bias-Corrected Confidence Intervals and Inference Procedures.

## Syntax

```
rdrobust depvar runvar [if] [in] [, c(#) fuzzy(fuzzyvar [sharpbw]) deriv(#)
scalepar(#) p(#) q(#) h(# #) b(# #) rho(#) covs(covars)
covs_drop(covsdroptoption) kernel(kernelfn) weights(weightsvar)
bwselect(bwmethod) scaleregul(#) masspoints(masspointsoption) bwcheck(#)
bwrestrict(bwropt) stdvars(stdopt) vce(vce type [vceopt1 vceopt2]) level(#) all
]
```

## 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 \(2014b\)](#), and [Calonico, Cattaneo, Farrell and Titiunik \(2017\)](#). A companion R package is also described in [Calonico, Cattaneo and Titiunik \(2015b\)](#).

Related Stata and R packages useful for inference in RD designs are described in the following website:

<https://rdpackages.github.io/>

## Options

---

Estimand

---

**c**(#) specifies the RD cutoff for *indepvar*. Default is **c**(0).

**fuzzy**(fuzzyvar [*sharpbw*]) 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. If the option *sharpbw* is set, the fuzzy RD estimation is performed 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.

**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.

**scalepar**(#) specifies scaling factor for RD parameter of interest. This option is useful when the estimator of interest requires a known multiplicative factor rescaling (e.g., Sharp Kink RD). Default is **scalepar**(1) (no rescaling).

---

Local Polynomial Regression

---

**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 ( $h$ ) 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 ( $b$ ) 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  $b=h/\rho$ . Default is **rho(1)** if  $h$  is specified but  $b$  is not.

**covs(covars)** specifies additional covariates to be used for estimation and inference.

**covs\_drop(covsdropoption)** assess collinearity in additional covariates used for estimation and inference. Options **pinv** (default choice) and **invsym** drops collinear additional covariates, differing only in the type of inverse function used. Option **off** only checks collinear additional covariates but does not drop them.

**kernel(kernelfn)** specifies the kernel function used to construct the local-polynomial estimator(s). Options are: **triangular**, **epanechnikov**, and **uniform**. Default is **kernel(triangular)**.

**weights(weightsvar)** is the variable used for optional weighting of the estimation procedure. The unit-specific weights multiply the kernel function.

---

#### Bandwidth Selection

---

**bwselect(bwmethod)** 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  $\min(\text{mserd}, \text{msesum})$ .
- msecomb2** for  $\text{median}(\text{msetwo}, \text{mserd}, \text{msesum})$ , for each side of the cutoff separately.
- cerd** 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  $\min(\text{cerd}, \text{cersum})$ .
- cercomb2** for  $\text{median}(\text{certwo}, \text{cerd}, \text{cersum})$ , 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 \(2017\)](#), [Calonico, Cattaneo and Farrell \(2020\)](#), and [Calonico, Cattaneo, Farrell and Titiunik \(2019\)](#), and the companion software articles.

**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)**.

**masspoints**(*masspointsoption*) checks and controls for repeated observations in the running variable. Options are:  
**off** ignores the presence of mass points.  
**check** looks for and reports the number of unique observations at each side of the cutoff.  
**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**(*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**(*bwropt*) if set **on**, computed bandwidths are restricted to lie within the range of *runvar*. Default is **on**.  
**stdvars**(*stdopt*) if set **on**, *depvar* and *runvar* are standardized before computing the bandwidths. Default is **off**.

---

#### Variance-Covariance Estimation

---

**vce**(*vcetype* [*vceopt1* *vceopt2*]) specifies the procedure used to compute the variance-covariance matrix estimator. Options are:  
**vce(nn** [*nnmatch*]) for heteroskedasticity-robust nearest neighbor variance estimator with *nnmatch* indicating the minimum number of neighbors to be used.  
**vce(hc0)** for heteroskedasticity-robust plug-in residuals variance estimator without weights.  
**vce(hc1)** for heteroskedasticity-robust plug-in residuals variance estimator with *hc1* weights.  
**vce(hc2)** for heteroskedasticity-robust plug-in residuals variance estimator with *hc2* weights.  
**vce(hc3)** for heteroskedasticity-robust plug-in residuals variance estimator with *hc3* weights.  
**vce(nncluster** *clustervar* [*nnmatch*]) for cluster-robust nearest neighbor variance estimation using with *clustervar* indicating the cluster ID variable and *nnmatch* matches indicating the minimum number of neighbors to be used.  
**vce(cluster** *clustervar*) for cluster-robust plug-in residuals variance estimation with degrees-of-freedom weights and *clustervar* indicating the cluster ID variable.  
 Default is **vce(nn 3)**.  
**level**(#) specifies confidence level for confidence intervals. Default is **level(95)**.

---

#### Other Options

---

**all** if specified, **rdrobust** reports three different procedures:  
 (i) conventional RD estimates with conventional variance estimator.  
 (ii) bias-corrected RD estimates with conventional variance estimator.  
 (iii) bias-corrected RD estimates with robust variance estimator.

---

#### Example: Cattaneo, Frandsen and Titiunik (2015) Incumbency Data

```
Setup
. use rdrobust_senate.dta

Robust RD Estimation using MSE bandwidth selection procedure
. rdrobust vote margin

Robust RD Estimation with both bandwidths set to 15
. rdrobust vote margin, h(15)
```

Other generic examples (**y** outcome variable, **x** running variable, **t** treatment take-up indicator):

Estimation for Sharp RD designs  
**. rdrobust y x, deriv(0)**

Estimation for Sharp Kink RD designs  
**. rdrobust y x, deriv(1)**

Estimation for Fuzzy RD designs  
**. rdrobust y x, fuzzy(t)**

Estimation for Fuzzy Kink RD designs  
**. rdrobust y x, fuzzy(t) deriv(1)**

### Stored results

**rdrobust** stores the following in **e()**:

#### Scalars

<b>e(N)</b>	original number of observations
<b>e(N_l)</b>	original number of observations to the left of the cutoff
<b>e(N_r)</b>	original number of observations to the right of the cutoff
<b>e(N_h_l)</b>	effective number of observations (given by the bandwidth <b>h_l</b> ) used to the left of the cutoff
<b>e(N_h_r)</b>	effective number of observations (given by the bandwidth <b>h_r</b> ) used to the right of the cutoff
<b>e(N_b_l)</b>	effective number of observations (given by the bandwidth <b>b_l</b> ) used to the left of the cutoff
<b>e(N_b_r)</b>	effective number of observations (given by the bandwidth <b>b_r</b> ) used to the right of the cutoff
<b>e(c)</b>	cutoff value
<b>e(p)</b>	order of the polynomial used for estimation of the regression function
<b>e(q)</b>	order of the polynomial used for estimation of the bias of the regression function estimator
<b>e(h_l)</b>	bandwidth used for estimation of the regression function below the cutoff
<b>e(h_r)</b>	bandwidth used for estimation of the regression function above the cutoff
<b>e(b_l)</b>	bandwidth used for estimation of the bias of the regression function estimator below the cutoff
<b>e(b_r)</b>	bandwidth used for estimation of the bias of the regression function estimator above the cutoff
<b>e(tau_cl)</b>	conventional local-polynomial RD estimate
<b>e(tau_cl_l)</b>	conventional local-polynomial left estimate
<b>e(tau_cl_r)</b>	conventional local-polynomial right estimate
<b>e(tau_bc)</b>	bias-corrected local-polynomial RD estimate
<b>e(tau_bc_l)</b>	bias-corrected local-polynomial left estimate
<b>e(tau_bc_r)</b>	bias-corrected local-polynomial right estimate
<b>e(se_tau_cl)</b>	conventional standard error of the local-polynomial RD estimator
<b>e(se_tau_rb)</b>	robust standard error of the local-polynomial RD estimator
<b>e(bias_l)</b>	estimated bias for the local-polynomial RD estimator below the cutoff
<b>e(bias_r)</b>	estimated bias for the local-polynomial RD estimator above the cutoff

#### Macros

<b>e(runningvar)</b>	name of running variable
<b>e(outcomevar)</b>	name of outcome variable
<b>e(clustvar)</b>	name of cluster variable
<b>e(covs)</b>	name of covariates
<b>e(vce_select)</b>	vcetype specified in <b>vce()</b>
<b>e(bwselect)</b>	bandwidth selection choice
<b>e(kernel)</b>	kernel choice

## Matrices

<b>e(beta_p_r)</b>	conventional p-order local-polynomial estimates to the right of the cutoff
<b>e(beta_p_l)</b>	conventional p-order local-polynomial estimates to the left of the cutoff
<b>e(beta_covs)</b>	coefficients of the additional covariates, only returned when <b>covs()</b> are used
<b>e(V_cl_r)</b>	conventional variance-covariance matrix to the right of the cutoff
<b>e(V_cl_l)</b>	conventional variance-covariance matrix to the left of the cutoff
<b>e(V_rb_r)</b>	robust variance-covariance matrix to the right of the cutoff
<b>e(V_rb_l)</b>	robust variance-covariance matrix to the left of the cutoff

**References**

- 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, 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, 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, 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. 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 Party Advantages in the U.S. Senate. *Journal of Causal Inference* 3(1): 1-24.

**Authors**

- Sebastian Calonico, Columbia University, New York, NY.  
[sebastian.calonico@columbia.edu](mailto:sebastian.calonico@columbia.edu).
- Matias D. Cattaneo, Princeton University, Princeton, NJ. [cattaneo@princeton.edu](mailto:cattaneo@princeton.edu).
- Max H. Farrell, University of Chicago, Chicago, IL. [max.farrell@chicagobooth.edu](mailto:max.farrell@chicagobooth.edu).
- Rocio Titiunik, Princeton University, Princeton, NJ. [titiunik@princeton.edu](mailto:titiunik@princeton.edu).