Statistics/Data Analysis

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(***#***) p(***#***) q(***#***) h(***# #***) b(***# #***) rho(***#***) covs(***covars***) kernel(***kernelfn***) weights(***weightsvar***) bwselect(***bwmethod***) vce(***vcetype [vceopt1 vceopt2]***) level(***#***) scalepar(***#***) scaleregul(***#***) 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 (2017), and Calonico, Cattaneo, Farrell and Titiunik (2018). 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://sites.google.com/site/rdpackages/

Options

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.

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.

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.

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(**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 (2017), and Calonico, Cattaneo, Farrell and Titiunik (2018), and the companion software articles.

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)**.

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).

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)**.

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)

Saved results

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

Scalars

e(N)	original number of observations
e(N_l)	original number of observations to the left of the cutoff
e(N_r)	original number of observations to the right of the cutoff
e(N_h_l)	effective number of observations (given by the bandwidth h_l) used to the left of the cutoff
e(N_h_r)	effective number of observations (given by the bandwidth h_r) used to the right of the cutoff
e(N_b_l)	effective number of observations (given by the bandwidth b_l) used to the left of the cutoff
e(N_b_r)	effective number of observations (given by the bandwidth b_r) used to the right of the cutoff
e(c)	cutoff value
e(p)	order of the polynomial used for estimation of the regression function
e(q)	order of the polynomial used for estimation of the bias of the regression function estimator
e(h_l)	bandwidth used for estimation of the regression function below the cutoff
e(h_r)	bandwidth used for estimation of the regression function above the cutoff
e(b_l)	bandwidth used for estimation of the bias of the regression function estimator below the cutoff
e(b_r)	bandwidth used for estimation of the bias of the regression function estimator above the cutoff

e(tau_cl)	conventional local-polynomial RD estimate
e(tau_cl_l)	conventional local-polynomial left estimate
e(tau_cl_r)	conventional local-polynomial right estimate
e(tau_bc)	bias-corrected local-polynomial RD estimate
e(tau_bc_l)	bias-corrected local-polynomial left estimate
e(tau_bc_r)	bias-corrected local-polynomial right estimate
e(se_tau_cl)	conventional standard error of the local-polynomial RD estimator
e(se_tau_rb)	robust standard error of the local-polynomial RD estimator
e(bias_l)	estimated bias for the local-polynomial RD estimator below the cutoff
e(bias_r)	estimated bias for the local-polynomial RD estimator above the cutoff

Macros

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

Matrices

e(beta_p_r)	conventional p-order local-polynomial estimates to the right of the cutoff
e(beta_p_l)	conventional p-order local-polynomial estimates to the left of the cutoff
e(V_cl_r)	conventional variance-covariance matrix to the right of the cutoff
e(V_cl_l)	conventional variance-covariance matrix to the left of the cutoff
e(V_rb_r)	robust variance-covariance matrix to the right of the cutoff
e(V_rb_l)	robust variance-covariance matrix to the left of the cutoff

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, 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. 2018. Regression Discontinuity Designs using Covariates. *Review of Economics and Statistics*, forthcoming.

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, University of Miami, Coral Gables, FL.
scalonico@bus.miami.edu.

Matias D. Cattaneo, University of Michigan, Ann Arbor, MI. cattaneo@umich.edu.

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

Rocio Titiunik, University of Michigan, Ann Arbor, MI. titiunik@umich.edu.