Statistics/Data Analysis

Title

rdbwselect —— Bandwidth Selection Procedures for Local Polynomial Regression Discontinuity Estimators.

Syntax

rdbwselect *depvar indepvar* [*if*] [*in*] [**, c(***#***) fuzzy(***fuzzyvar [sharpbw]***) deriv(***#***) p(***#***) q(***#***) covs(***covars***) kernel(***kernelfn***) weights(***weightsvar***) bwselect(***bwmethod***) vce(***vcetype [vceopt1 vceopt2]***) scaleregul(***#***) all**]

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), and Calonico, Cattaneo, Farrell and Titiunik (2018).

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 (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).

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: **tri**angular, **epa**nechnikov, and **uni**form. 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.
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 (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)**.

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, **rdbwselect** reports all available bandwidth selection procedures.

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

Setup
```
. use rdrobust_senate.dta
```

MSE bandwidth selection procedure
```
. rdbwselect vote margin
```

All bandwidth bandwidth selection procedures
```
. rdbwselect vote margin, all
```

Saved results

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

Scalars

e(N_l)	number of observations to the left of the cutoff
e(N_r)	number of observations 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_mserd)	MSE-optimal bandwidth selector for the RD treatment effect estimator.
e(h_msetwo_l)	MSE-optimal bandwidth selectors below the cutoff for the RD treatment effect estimator.
e(h_msetwo_r)	MSE-optimal bandwidth selectors above the cutoff for the RD treatment effect estimator.
e(h_msesum)	MSE-optimal bandwidth selector for the sum of regression estimates.
e(h_msecomb1)	for min(**mserd**,**msesum**).
e(h_msecomb2_l)	for median(**msetwo**,**mserd**,**msesum**), below the cutoff.
e(h_msecomb2_r)	for median(**msetwo**,**mserd**,**msesum**), above the cutoff.
e(h_cerrd)	CER-optimal bandwidth selector for the RD treatment effect estimator.
e(h_certwo_l)	CER-optimal bandwidth selectors below the cutoff for the RD treatment effect estimator.
e(h_certwo_r)	CER-optimal bandwidth selectors above the cutoff for the RD treatment effect estimator.
e(h_cersum)	CER-optimal bandwidth selector for the sum of regression estimates.
e(h_cercomb1)	for min(**cerrd**,**cersum**).
e(h_cercomb2_l)	for median(**certwo_l**,**cerrd**,**cersum**), below the cutoff.
e(h_cercomb2_r)	for median(**certwo_r**,**cerrd**,**cersum**), above the cutoff.
e(b_mserd)	MSE-optimal bandwidth selector for the bias of the RD treatment effect estimator.
e(b_msetwo_l)	MSE-optimal bandwidth selectors below the cutoff for the bias of the RD treatment effect estimator.
e(b_msetwo_r)	MSE-optimal bandwidth selectors above the cutoff for the bias of the RD treatment effect estimator.
e(b_msesum)	MSE-optimal bandwidth selector for the sum of regression estimates for the bias of the RD treatment effect estimator.
e(b_msecomb1)	for min(**mserd**,**msesum**).
e(b_msecomb2_l)	for median(**msetwo**,**mserd**,**msesum**), below the cutoff.
e(b_msecomb2_r)	for median(**msetwo**,**mserd**,**msesum**), above the cutoff.
e(b_cerrd)	CER-optimal bandwidth selector for the bias of the RD treatment effect estimator.
e(b_certwo_l)	CER-optimal bandwidth selectors below the cutoff for the bias of the RD treatment effect estimator.
e(b_certwo_r)	CER-optimal bandwidth selectors above the cutoff for the bias of the RD treatment effect estimator.
e(b_cersum)	CER-optimal bandwidth selector for the sum of regression estimates for the bias of the RD treatment effect estimator.
e(b_cercomb1)	for min(**cerrd**,**cersum**).
e(b_cercomb2_l)	for median(**certwo_l**,**cerrd**,**cersum**), below the cutoff.
e(b_cercomb2_r)	for median(**certwo_r**,**cerrd**,**cersum**), 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

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.