Statistics/Data analysis

Title

rdhte —— RD Heterogeneous Treatment Effects Estimation and Inference.

Syntax

rdhte *depvar runvar* [*if*] [*in*] [, **covs_hte(***covars***)** **c(***#***)** **p(***#***)** **q(***#***)** **h(***#***)** **h_l(***#***)** **h_r(***#***)** **kernel(***kernelfn***)** **vce(***vcetype***)** **level(***#***)** **covs_eff(***covars***)** **bwjoint** **labels**]

Description

rdhte provides estimation and inference for heterogeneous treatment effects in RD designs using local polynomial regressions, allowing for interactions with pretreatment covariates (Calonico, Cattaneo, Farrell, Palomba and Titiunik, 2025a). Inference is implemented using robust bias-correction methods (Calonico, Cattaneo, and Titiunik, 2014)

Companion commands are: rdbwhte for data-driven bandwidth selection and rdhte_lincom for testing linear restrictions of parameters. More general post-estimation linear hypotheses can be tested with the Stata function test.

A detailed introduction to **rdhte** in Stata is given in Calonico, Cattaneo, Farrell, Palomba and Titiunik (2025b).

Related software packages for analysis and interpretation of RD designs and related methods are available in:

https://rdpackages.github.io/

For background methodology, see Calonico, Cattaneo, Farrell, and Titiunik (2019), Calonico, Cattaneo and Farrell (2020), Cattaneo and Titiunik (2022).

Options

Estimand

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

covs_hte(*covars***)** specifies covariate(s) for heterogeneous treatment effects. Factor variables notation can be used to distinguish between continuous and categorical variables, select reference categories, specify interactions between variables, and include polynomials of continuous variables. If not specified, the RD Average Treatment Effect is computed.

labels displays the final RD estimates using variable labels from **covs_hte(***covars***)**.

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(*#***)**, **h_l(***#***)** and **h_r(***#***)** set the bandwidths to construct the RD estimator. The same choice could be used on each side of the cutoff (via **h(***#***)**), or different to the left and right (using **h_l(***#***)** and **h_r(***#***)**). More than one bandwidth can be specified for categorical covariates. If not specified, bandwidths are computed by the companion command rdbwhte.

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

covs_eff(*covars***)** specifies additional covariates to be used for efficiency improvements.

Data-Driven Bandwidth Selection

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)**.

bwjoint forces all bandwidths to be the same across groups. Default is **level(95)**.

Variance-Covariance Estimation

vce(<u>*vcetype*</u>**)** vcetype may be **robust**, **cluster** *clustvarlist*, **hc2** [*clustvar*], or **hc3**.
Default is **vce(hc3)**.

level(*#***)** specifies confidence level for confidence intervals.

<u>Example:</u>

Setup using <u>Granzier, Pons, and Tricaud (2023)</u> Data
. use rdhte_dataset.dta

RD-HTE Estimation by left/right groups
. rdhte y x, covs_hte(i.w_left)

RD-HTE Estimation by left/right groups with common bandwidth
. rdhte y x, covs_hte(i.w_left) bwjoint

RD-HTE Estimation by left/right groups and strong
. rdhte y x, covs_hte(i.w_left#i.w_strong)

RD-HTE Estimation using a continuous variable
. rdhte y x, covs_hte(c.w_strength)

RD-HTE Estimation using a continuous variable with clustered standard errors
. rdhte y x, covs_hte(c.w_strength) vce(cluster cluster_var)

<u>Stored results</u>

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

Scalars

e(N)	original number of observations
e(c)	cutoff value
e(p)	order of the polynomial used for estimation of the regression function

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(kernel)	kernel choice

Matrices

e(h)	bandwidth
e(tau_hat)	p-order local-polynomial estimates for the outcome variable
e(tau_bc)	bias-corrected local-polynomial estimates for the outcome variable
e(tau_se)	robust standard errors
e(tau_V)	robust variance-covariance matrix
e(tau_t)	robust t-statistics
e(tau_pv)	robust p-values
e(tau_N)	sample size
e(tau_ci_lb)	robust lower bound confidence interval
e(tau_ci_ub)	robust upper bound confidence interval

References

Calonico, Cattaneo, Farrell, Palomba and Titiunik. 2025a. Treatment Effect Heterogeneity in Regression Discontinuity Designs. *Working Paper*.

Calonico, Cattaneo, Farrell, Palomba and Titiunik. 2025b. rdhte: Conditional Average Treatment Effects in RD Designs. *Working Paper*.

Granzier, Pons, and Tricaud. 2023. Coordination and Bandwagon Effects: How Past Rankings Shape the Behavior of Voters and Candidates. *American Economic Journal: Applied Economics*, 15(4): 177?217.

Cattaneo and Titiunik. 2022. Regression Discontinuity Designs. *Annual Review of Economics*, 14: 821-851.

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

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

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

Authors

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

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

Max H. Farrell, University of California, Santa Barbara, CA. maxhfarrell@ucsb.edu.

Filippo Palomba, Princeton University, Princeton, NJ. fpalomba@princeton.edu.

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