

Title

rdplot — Data-Driven Regression Discontinuity Plots.

Syntax

rdplot depvar indepvar [if] [in] [, c(#) nbins(# #) binselect(binmethod) scale(#
 #) support(# #) p(#) h(# #) kernel(kernelfn) weights(weightsvar) covs(covars)
 covs_eval(covars_eval) covs_drop(covsdropoption) masspoints(masspointsoption)
 ci(cilevel) shade graph_options(gphopts) hide genvars]

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 <u>Calonico</u>, <u>Cattaneo and Titiunik (2014)</u>, and <u>Calonico</u>, <u>Cattaneo</u>, <u>Farrell and Titiunik (2017)</u>. A companion R package is also described in <u>Calonico</u>, <u>Cattaneo and Titiunik (2015b)</u>.

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 in indepvar. Default is c(0).

Bin Selection

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 (binmethod) specifies the data-driven procedure to select the number of bins. This option is available only if J- and J+ are not set manually using **nbins(.)**. Options are:

es IMSE-optimal evenly-spaced method using spacings estimators.
espr IMSE-optimal evenly-spaced method using polynomial regression.
esmv mimicking variance evenly-spaced method using spacings estimators.

esmy mimicking variance evenly-spaced method using spacings estimators.

esmyr mimicking variance evenly-spaced method using polynomial regression.

 ${\tt qs}$ IMSE-optimal quantile-spaced method using spacings estimators. ${\tt qspr}$ IMSE-optimal quantile-spaced method using polynomial regression.

qsmv mimicking variance quantile-spaced method using spacings estimators. **qsmvpr** mimicking variance quantile-spaced method using polynomial regression. Default is **binselect(esmv)**.

Note: procedures involving spacing estimators are not invariant to rearrangements of depvar when there are repeated values (i.e., mass points in the running variable).

scale(# #) specifies multiplicative factors, denoted s- and s+, respectively, to
 adjust the number of bins selected. Specifically, the number of bins used for
 the treatment and control groups will be ceil(s- * J-) and ceil(s+ * J+),
 where J- and J+ denote the optimal numbers of bins originally computed for
 each group. Default is scale(1 1).

support(# #) sets an optional extended support of the running variable to be used in the construction of the bins. Default is the sample range.

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 sets binselect (binmethod) as polynomial regression when mass points are present.

Default option is masspoints (adjust).

☐ Polynomial Fit ☐

- p(#) specifies the order of the (global) polynomial fit used to approximate the population conditional expectation functions for control and treated units. Default is p(4).
- $h\left(\text{\# \#} \right)$ 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.
- kernel (kernelfn) 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 (weightsvar) is the variable used for optional weighting of the estimation procedure. The unit-specific weights multiply the kernel function.
- covs (covars) additional covariates used to construct the local-polynomial estimator(s).
- covs_eval(covars_eval) sets the evaluation points for the additional covariates, when included in the estimation. Options are: 0 (default) and mean.
- 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.

Plot Options

ci(cilevel) graphical option to display confidence intervals of level cilevel for

shade graphical option to replace confidence intervals with shaded areas.

graph_options(gphopts) graphical options to be passed on to the underlying graph command.

hide omits the RD plot.

Generate Variables ^l

```
genvars generates new variables storing the following results.
        rdplot id unique bin ID for each observation. Negative natural numbers are
            assigned to observations to the left of the cutoff, and positive natural
            numbers are assigned to observations to the right of the cutoff.
        rdplot_N number of observations in the corresponding bin for each observation.
        rdplot_min_bin lower end value of the bin for each observation.
        rdplot_max_bin upper end value of the bin for each observation.
        rdplot_mean_bin middle point of the corresponding bin for each observation.
        rdplot mean x sample mean of the running variable within the corresponding bin
            for each observation.
        rdplot_mean_y sample mean of the outcome variable within the corresponding bin
            for each observation.
        rdplot_se_y standard deviation of the mean of the outcome variable within the
            corresponding bin for each observation.
        {\tt rdplot\_ci\_1} lower end value of the confidence interval for the sample mean of
            the outcome variable within the corresponding bin for each observation.
        rdplot ci r upper end value of the confidence interval for the sample mean of
            the outcome variable within the corresponding bin for each observation.
        rdplot_hat_y predicted value of the outcome variable given by the global
            polynomial estimator.
Example: Cattaneo, Frandsen and Titiunik (2015) Incumbency Data
    Setup
        . use rdrobust_senate.dta
    Basic specification with title
        . rdplot vote margin, graph_options(title(RD Plot))
    Quadratic global polynomial with confidence bands
        . rdplot vote margin, p(2) ci(95) shade
```

Stored results

```
rdplot stores the following in e():
```

```
Scalars
  e(N 1)
                       original number of observations to the left of the cutoff
  e(N_r)
                       original number of observations to the right of the cutoff
                      cutoff value
  e (c)
                       selected number of bins to the left of the cutoff
  e(J_star_1)
  e(J_star_r)
                      selected number of bins to the right of the cutoff
Macros
  e(binselect)
                     method used to compute the optimal number of bins
Matrices
  e(coef_1)
                      coefficients of the p-th order polynomial estimated to the
                        left of the cutoff
                      coefficients of the p{\text -}{\text th} order polynomial estimated to the
  e(coef_r)
                        right of the cutoff
```

References

- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2017. rdrobust:
 Stata Journal 17(2): 372-404.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. <u>Robust Data-Driven</u>
 <u>Inference in the Regression-Discontinuity Design</u>. *Stata Journal* 14(4): 909-946.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015a. <u>Optimal Data-Driven Regression Discontinuity Plots</u>. 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>U.S. Senate</u>. Journal of Causal Inference 3(1): 1-24.

<u>Authors</u>

- 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.