



Title

rddensity — Manipulation Testing Using Local Polynomial Density Estimation.

Syntax

```
rddensity Var [if] [in] [,
    c(#) p(#) q(#) fitselect(FitMethod) kernel(KernelFn) vce(VceMethod)
    nomasspoints level(#) all
    h(# #) bwselect(BwMethod) nlocalmin(#) nunique(#) noregularize
    bino_n(#) bino_nstep(#) bino_w(# #) bino_wstep(# #) bino_nw(#) bino_p(#)
    nobinomial
    plot plot_range(# #) plot_n(# #) plot_grid(GridMethod)
    plot_bwselect(BwMethod)
    plot_ciuniform plot_cisimul(# #)
    graph_opt(GraphOpt) genvars(NewVarName)
    plotl_estype(EstType) esll_opt(LineOpt) espl_opt(PtOpt)
    plotr_estype(EstType) eslr_opt(LineOpt) espr_opt(PtOpt)
    plotl_citype(CIType) cirl_opt(AreaOpt) cill_opt(LineOpt)
    cibl_opt(EbarOpt)
    plotr_citype(CIType) cirr_opt(AreaOpt) cilr_opt(LineOpt)
    cibr_opt(EbarOpt)
    hist_range(# #) hist_n(# #) hist_width(# #) histl_opt(BarOpt)
    histr_opt(BarOpt) nohistogram
]
```

Description

rddensity implements manipulation testing procedures using the local polynomial density estimators proposed in [Cattaneo, Jansson and Ma \(2020a\)](#), and implements graphical procedures with valid confidence bands using the results in [Cattaneo, Jansson and Ma \(2020b\)](#). In addition, the command provides complementary manipulation testing based on finite sample exact binomial testing following the results in [Cattaneo, Frandsen and Titiunik \(2015\)](#) and [Cattaneo, Frandsen and Vazquez-Bare \(2017\)](#). For an introduction to manipulation testing see McCrary (2008).

A detailed introduction to this Stata command is given in [Cattaneo, Jansson and Ma \(2018\)](#). Companion R functions are also available [here](#).

Companion function is [rdbwdensity](#). For graphical procedures, the [lpdensity](#) package is required.

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

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

Options

Density Estimation

c(#) specifies the threshold or cutoff value in the support of *Var*, which determines the two samples (e.g., control and treatment units in RD settings). Default is **c(0)**.

p(#) specifies the local polynomial order used to construct the density estimators. Default is **p(2)** (local quadratic approximation).

q(#) specifies the local polynomial order used to construct the bias-corrected density estimators. Default is **q(p(#)+1)** (local cubic approximation for default **p(2)**).

fitselect(*FitMethod*) specifies the density estimation method.
unrestricted for density estimation without any restrictions (two-sample, unrestricted inference). This is the default option.
restricted for density estimation assuming equal distribution function and higher-order derivatives.

kernel(*KernelFn*) specifies the kernel function used to construct the local polynomial estimators.

triangular $K(u) = (1 - |u|) * (|u| \leq 1)$. This is the default option.
epanechnikov $K(u) = 0.75 * (1 - u^2) * (|u| \leq 1)$.
uniform $K(u) = 0.5 * (|u| \leq 1)$.

vce(*VceMethod*) specifies the procedure used to compute the variance-covariance matrix estimator.

plugin for asymptotic plug-in standard errors.
jackknife for jackknife standard errors. This is the default option.

nomasspoints will not adjust for mass points in the data.

level(#) specifies the level of the confidence interval, which should be between 0 and 100. Default is **level(95)**.

all if specified, **rddensity** reports two testing procedures:
 Conventional test statistic (not valid when using MSE-optimal bandwidth choice).
 Robust bias-corrected statistic. This is the default option.

Bandwidth Selection

h(# #) specifies the bandwidth (*h*) used to construct the density estimators on the two sides of the cutoff. If not specified, the bandwidth *h* is computed by the companion command **rdbwdensity**. 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.

bwselect(*BwMethod*) specifies the bandwidth selection procedure to be used.

each based on MSE of each density estimator separately (two distinct bandwidths, *hl* and *hr*).
diff based on MSE of difference of two density estimators (one common bandwidth, *hl=hr*).
sum based on MSE of sum of two density estimators (one common bandwidth, *hl=hr*).
comb bandwidth is selected as a combination of the alternatives above. This is the default option.
 For **fitselect(unrestricted)**, it selects median(**each**,**diff**,**sum**).
 For **fitselect(restricted)**, it selects min(**diff**,**sum**).

nlocalmin(#) specifies the minimum number of observations in each local neighborhood. This option will be ignored if set to 0, or if **noregularize** is used. Default is **20+p(#)+1**.

nuniquemin(#) specifies the minimum number of unique observations in each local neighborhood. This option will be ignored if set to 0, or if **noregularize** is used. Default is **20+p(#)+1**.

noregularize suppresses local sample size checking.

Binomial Test

bino_w(# #) specifies the half length(s) of the initial window. If two values are provided, they will be used for the data below and above the cutoff separately.

bino_n(#) specifies the sample size in the initial window. This option will be ignored if **bino_w**(# #) is provided.

bino_wstep(# #) specifies the increment in half length(s).

bino_nstep(#) specifies the increment in sample size. This option will be ignored if **bino_wstep**(#) is provided.

bino_nw(#) specifies the total number of windows. Default is 10.

bino_p(#) specifies the null hypothesis of the binomial test. Default is 0.5.

nobinomial suppresses the binomial test. By default, the initial (smallest) window contains 20 observations, and its length is also used as the increment for subsequent windows.

Plotting

plot if specified, **rddensity** plots density estimates and confidence intervals/bands around the cutoff (this feature depends on a companion package **lpdensity**). Note that additional estimation (computing time) is needed.

plot_range(# #) specifies the lower and upper bound of the plotting region. Default is $[c-3*hl, c+3*hr]$ (three bandwidths around the cutoff).

plot_n(# #) specifies the number of grid points used for plotting on the two sides of the cutoff. Default is **plot_n(10 10)** (i.e., 10 points are used on each side).

plot_grid(*GridMethod*) specifies how the grid points are positioned. Options are **es** (evenly spaced) and **qs** (quantile spaced).

plot_bwselect(*BwMwthod*) specifies the method for data-driven bandwidth selection. Options are **mse-dpi**, **imse-dpi**, **mse-rot**, and **imse-rot**. See **lpdensity** for additional details. If this option is omitted, the same bandwidth(s) used for manipulation testing will be employed.

plot_ciuniform plots uniform confidence bands instead of pointwise confidence intervals. The companion option, **plot_cisimul**(#), specifies the number of simulations used to construct critical values. Default is 2000.

graph_opt(*GraphOpt*) specifies additional options for plotting, such as legends and labels.

genvars(*NewVarName*) specifies if new variables should be generated to store estimation results.

Remark. Bias correction is only used for the construction of confidence intervals/bands, but not for point estimation. The point estimates, denoted by f_p , are constructed using local polynomial estimates of order $p(\#)$, while the centering of the confidence intervals/bands, denoted by f_q , are constructed using local polynomial estimates of order $q(\#)$. The confidence intervals/bands take the form: $[f_q - cv * SE(f_q), f_q + cv * SE(f_q)]$, where cv denotes the appropriate critical value and $SE(f_q)$ denotes a standard error estimate for the centering of the confidence interval/band. As a result, the confidence intervals/bands may not be centered at the point estimates because they have been bias-corrected. Setting $q(\#)$ and $p(\#)$ to be equal results on centered at the point estimate confidence intervals/bands, but requires undersmoothing for valid inference (i.e., (I)MSE-optimal bandwidth for the density point estimator cannot be used). Hence the bandwidth would need to be specified manually when $q(\#) = p(\#)$, and the point estimates will not be (I)MSE optimal. See Cattaneo, Jansson and Ma (2020b, 2020c) for details, and also Calonico, Cattaneo, and Farrell (2018, 2020) for robust bias correction methods.

Sometimes the density point estimates may lie outside of the confidence intervals/bands, which can happen if the underlying distribution exhibits high curvature at some evaluation point(s). One possible solution in this case is to increase the polynomial order $p(\#)$ or to employ a smaller bandwidth.

Additional Plotting Options: Histogram

hist_range(# #) specifies the lower and upper bound of the histogram plot.
Default is $[c-3*hl, c+3*hr]$ (three bandwidths around the cutoff).

hist_n(# #) specifies the number of histogram bars. Default is $\min[\sqrt{N}, 10*\log(N)/\log(10)]$, where N is the number of observations within the range specified by **hist_range**(# #).

hist_width(# #) specifies the width of histogram bars. This option will be ignored if **hist_range**(# #) is provided.

nohistogram suppresses the histogram in the background of the plot.

Additional Plotting Options: Below the Cutoff

plotl_estype(*EstType*) specifies the plotting style of point estimates.

line a curve. This is the default option.
points individual points.
both both of the above.
none will not plot point estimates.

esll_opt(*LineOpt*) specifies additional **twoway line** options for plotting point estimates.

espl_opt(*PtOpt*) specifies additional **twoway scatter** options for plotting point estimates.

plotl_citype(*EstType*) specifies the plotting style of confidence intervals/bands.

region shaded region. This is the default option.
line upper and lower bounds.
ebar error bars.
all all of the above.
none will not plot confidence intervals/bands.

cirl_opt(*AreaOpt*) specifies additional **twoway rarea** options for plotting confidence intervals/regions.

cill_opt(*LineOpt*) specifies additional **twoway rline** options for plotting confidence intervals/regions.

cibl_opt(*EbarOpt*) specifies additional **twoway rcap** options for plotting confidence intervals/regions.

histl_opt(*BarOpt*) specifies additional **twoway bar** options for histogram.

Additional Plotting Options: Above the Cutoff

plotr_estype(*EstType*) specifies the plotting style of point estimates.

line a curve. This is the default option.
points individual points.
both both of the above.
none will not plot point estimates.

eslr_opt(*LineOpt*) specifies additional **twoway line** options for plotting point estimates.

espr_opt(*PtOpt*) specifies additional **twoway scatter** options for plotting point estimates.

plotr_citype(*EstType*) specifies the plotting style of confidence intervals/bands.

region shaded region. This is the default option.
line upper and lower bounds.
ebar error bars.
all all of the above.
none will not plot confidence intervals/bands.

cirr_opt(*AreaOpt*) specifies additional **twoway rarea** options for plotting confidence intervals/regions.

cilr_opt(*LineOpt*) specifies additional **twoway rline** options for plotting confidence intervals/regions.

cibr_opt(*EbarOpt*) specifies additional **twoway rcap** options for plotting confidence intervals/regions.

histr_opt(*BarOpt*) specifies additional **twoway bar** options for histogram.

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

Load dataset (cutoff is 0 in this dataset):

```
. use rddensity_senate.dta
```

Manipulation test using default options:

```
. rddensity margin
```

Reporting both conventional and robust bias-corrected statistics:

```
. rddensity margin, all
```

Manipulation test using manual bandwidths choices and plug-in standard errors:

```
. rddensity margin, h(10 20) vce(plugin)
```

Plot density and save results to variables:

```
. capture drop temp_*
. rddensity margin, pl plot_range(-50 50) plot_n(100 100) genvars(temp)
```

Saved results

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

Macros

e(c)	cutoff value
e(p)	order of the polynomial used for density estimation
e(q)	order of the polynomial used for bias-correction estimation
e(N_l)	sample size to the left of the cutoff
e(N_r)	sample size to the right of the cutoff
e(N_h_l)	effective sample size (within bandwidth) to the left of the cutoff
e(N_h_r)	effective sample size (within bandwidth) to the right of the cutoff
e(h_l)	bandwidth used to the left of the cutoff
e(h_r)	bandwidth used to the right of the cutoff
e(f_ql)	bias-corrected density estimate to the left of the cutoff
e(f_qr)	bias-corrected density estimate to the right of the cutoff
e(se_ql)	standard error for bias-corrected density estimate to the left of the cutoff
e(se_qr)	standard error for bias-corrected density estimate to the right of the cutoff
e(se_q)	standard error for bias-corrected density test
e(T_q)	bias-corrected t-statistic
e(pv_q)	p-value for bias-corrected density test
e(runningvar)	running variable used
e(kernel)	kernel used
e(fitmethod)	model used
e(bwmethod)	bandwidth selection method used
e(vce)	standard errors estimator used

Only available if **all** is specified:

e(f_pl)	density estimate to the left of the cutoff without bias correction
e(f_pr)	density estimate to the right of the cutoff without bias correction
e(se_pl)	standard error for density estimate to the left of the cutoff without bias correction
e(se_pr)	standard error for density estimate to the right of the cutoff without bias correction
e(se_p)	standard error for density test without bias correction
e(T_p)	t-statistic without bias correction
e(pv_p)	p-value for density test without bias correction

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, and M. H. Farrell. 2020. Coverage Error Optimal Confidence Intervals for Local Polynomial Regression. Working paper.
- Cattaneo, M. D., B. Frandsen, and R. Titiunik. 2015. Randomization Inference in the Regression Discontinuity Design: An Application to the Study of Party Advantages in the U.S. Senate. *Journal of Causal Inference* 3(1): 1-24.
- Cattaneo, M. D., M. Jansson, and X. Ma. 2018. Manipulation Testing based on Density Discontinuity. *Stata Journal* 18(1): 234-261.
- Cattaneo, M. D., M. Jansson, and X. Ma. 2020a. Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, forthcoming.
- Cattaneo, M. D., M. Jansson, and X. Ma. 2020b. Local Regression Distribution Estimators. Working paper.
- Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2020c. lpdensity: Local Polynomial Density Estimation and Inference. Working paper.
- Cattaneo, M. D., Titiunik, R. and G. Vazquez-Bare. 2017. Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality. *Journal of Policy Analysis and Management* 36(3): 643-681.
- McCrary, J. 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics* 142(2): 698-714.

Authors

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

Michael Jansson, University of California Berkeley, Berkeley, CA. mjansson@econ.berkeley.edu.

Xinwei Ma, University of California San Diego, La Jolla, CA. xlma@ucsd.edu.