



Title

rddensity — Manipulation Testing using Local-Polynomial Density Estimation.

Syntax

```
rddensity var [if] [in] [, c(#) p(#) q(#) fitselect(fitmethod) kernel(kernelfn)  
             h(# #) bwselect(bwmeth) vce(vcemeth) all plot plot_range(# #) plot_n(# #)  
             plot_grid(gridmethod) genvars(varname) level(#) graph_options(...) ]
```

Description

rddensity implements manipulation testing procedures using the local polynomial density estimators proposed in Cattaneo, Jansson and Ma (2019). For a review on 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](#).

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

<https://sites.google.com/site/rdpackages/>

Options

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

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

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

fitselect(*fitmethod*) specifies whether restrictions should be imposed. Options are:

- unrestricted** for density estimation without any restrictions (two-sample, unrestricted inference). This is the default option.
- restricted** for density estimation assuming equal c.d.f. and higher-order derivatives.

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

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 **rdbandwidth**. 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(*bwmeth*) specifies the bandwidth selection procedure to be used. Options are:

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

vce(*vcemethod*) specifies the procedure used to compute the variance-covariance matrix estimator. Options are:
plugin for asymptotic plug-in standard errors.
jackknife for jackknife standard errors. This is the default option.

all if specified, **rddensity** reports two testing procedures (given choices **fitselect**(*fitmethod*) and **bwselect**(*bwmethod*)):
 1. Conventional test statistic (not valid when using MSE-optimal bandwidth choice).
 2. Robust bias-corrected statistic. This is the default option.

plot if specified, **rddensity** plots density around the cutoff (this feature depends on a companion package **lpdensity**). Note that additional estimation (computing time) is need.

plot_range(# #) specifies the lower and upper bound of the plotting region. By default it is three bandwidths around the cutoff.

plot_n(# #) specifies the number of grid points used for plotting on the two sides of the cutoff. By default it is **plot_n(10 10)** (that is, 10 points are used on each side).

plot_grid(*gridmethod*) specifies how the grids are positioned. Options are:
es evenly spaced.
qs quantile spaced.

genvars(*VarName*) specifies if new variables should be generated to store estimation results for plotting. If *VarName* is provided, the following new variables will be generated:
VarName_grid (grid points),
VarName_bw (bandwidth),
VarName_f (point estimate with polynomial order **p**(#)),
VarName_cil and *VarName_cir* (confidence interval constructed with polynomial order **q**(#)),
VarName_group (equals 1 if the grid point is to the right of the cutoff).

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

graph_options(...) specifies options for plotting.

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

Manipulation test using all three method available:
. 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, plot plot_range(-50 50) plot_n(100 100) genvars(temp)
graph_options(title("RDDENSITY PLOT: Senate Data") xtitle("margin"))
```

Saved results

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

Macros

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

References

- Cattaneo, M. D., Frandsen, B., 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., Michael Jansson, and Xinwei Ma. 2018. Manipulation Testing based on Density Discontinuity. *Stata Journal* 18(1): 234-261.
- Cattaneo, M. D., Michael Jansson, and Xinwei Ma. 2019. Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, forthcoming.
- McCrory, 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.