rddensity

Release 2.2.0

Matias Cattaneo, Rajita Chandak, Michael Jansson, Xinwei Ma

May 27, 2022

CONTENTS:

Density discontinuity testing (a.k.a. manipulation testing) is commonly employed in regression discontinuity designs and other program evaluation settings to detect perfect self-selection (manipulation) around a cutoff where treatment/policy assignment changes. This package implements manipulation testing procedures using the local polynomial density estimators: `rddensity()` to construct test statistics and p-values given a prespecified cutoff, `rdbwdensity()` to perform data-driven bandwidth selection, and `rdplotdensity()` to construct density plots.

Additional information regarding rddensity is available here.

Companion `Stata` aand `R` packages and additional regression discontinuity packages are available at rdpackages.github.io.

Install `rddensity` by running `pip install rddensity`.

Import functions by running

```
>>> from rddensity import rddensity
```

```
>>> from rddensity import rdbwdensity
```

```
>>> from rddensity import rdplotdensity
```

Source code and replication files are available in the rddensity repository.

CHAPTER

ONE

REFERENCES

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. 2022. lpdensity: Local Polynomial Density Estimation and Inference.. *Journal of Statistical Software* Forthcoming.

CHAPTER

TWO

AUTHORS

Matias D. Cattaneo, Princeton University. (cattaneo@princeton.edu).

Rajita Chandak (maintainer), Princeton University. (rchandak@princeton.edu).

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

Xinwei Ma (maintainer), University of California San Diego. (x1ma@ucsd.edu).

2.1 rddensity

2.1.1 rddensity

Manipulation testing using local polynomial density estimation

Description

`rddensity` implements manipulation testing procedures using the local polynomial density estimators proposed in Cattaneo, Jansson and Ma (2020), and implements graphical procedures with valid confidence bands using the results in Cattaneo, Jansson and Ma (2021a,b). 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).

Companion commands: `rdbwdensity` for bandwidth selection and `rdplotdensity` for plotting estimation results.

References

Cattaneo M. D., B. Grandsen, 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. 2020. Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 115(531): 1449-1455.

Cattaneo M. D., M. Jansson, and X. Ma. 2022. lpdensity: Local Polynomial Density Estimation and Inference.. *Journal of Statistical Software* Forthcoming.

Cattaneo M. D., R. Titiunik, and F. Vazquez-Bare. 2017. Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start of 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. (cattaneo@princeton.edu).

Rajita Chandak (maintainer), Princeton University. (rchandak@princeton.edu).

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

Xinwei Ma (maintainer), University of California San Diego. (x1ma@ucsd.edu).

class rddensity.rddensity.**CJMrddensity**(*hat*, *sd_asy*, *sd_jk*, *test*, *hat_p*, *sd_asy_p*, *sd_jk_p*, *test_p*, *n*, *h*, *fitselect*, *kernel*, *vce*, *c*, *p*, *q*, *regularize*, *nLocalMin*, *bino_flag*, *nUniqueMin*, *massPoints*, *massPoints_flag*, *bwselectl*, *bwselect*, *binoN*, *binoW*, *binoNStep*, *binoWStep*, *binoNW*, *binoP*, *useall*, *X_min*, *X_max*, *bino*)

Class of rddensity function outputs. Object type returned by *rddensity()*.

rddensity.rddensity.**rddensity**(*X*, *c=0*, *p=2*, *q=0*, *fitselect='unrestricted'*, *kernel='triangular'*, *vce='jackknife'*, *h=[]*, *bwselect='comb'*, *useall=False*, *massPoints=True*, *regularize=True*, *nLocalMin=None*, *nUniqueMin=None*, *bino_flag=True*, *binoW=None*, *binoN=None*, *binoWStep=None*, *binoNStep=None*, *binoNW=[10]*, *binoP=[0.5]*)

Parameters

X: Numeric vector or one dimentional matrix/dataframe the running variable.

c: Numeric Specifies the threshold or cutoff value in the support of *X*. Default is *0*.

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

fitselect: String specifies the density estimation method. *unrestricted* (Default) for density estimation without any restrictions (two-sample, unrestricted inference). *restricted* for density estimation assuming equal distribution function and higher order dericatives.

kernel: String specifies the kernel function used to construct the local polynomial estimators. Accepted kernels: *triangular* (Default), *epanechnikov* or *uniform*.

vce: String specifies the procedure used to compute the variance-covariance matrix estimatior. *jackknife* (Default) for jackknife standard errors or *plugin* for asymptotic plug-in standard errors.

massPoints: Boolean, Default *True*. Specifies wether to adjust for mass points in the data.

useall: Boolean, Default *False*. If specified, will report two testing procedures: conventional test statistic (not valie when useing mse-optimal bandwidth) and robust bias-corrected statistic.

h: Numeric Specifies the bandwidth used to construct the density estimators on the two sides of the cutoff. If not specified, the bandwidth h is computed using the companion function, *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: String. Specified the bandwidth selection procedure to be used. *each*-based on MSE of each density estimator separately (two distinct bandwidths), *diff* - based on MSE of difference of two density estimators, gives one common bandwidth, *sum*-based on MSE of sum of two density estimators, gives one common bandwidth. *comb* (default)-bandiwdth 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)*

regularize: Boolean, Default *True*. Specifies whether to conduct local sample size checking. When True, the bandwidth is chosen such that the local region includes at least *nLocalMin* observations and at least *nUniqueMin* unique observations.

nLocalMin: Nonnegative integer Specifies the minimum number of observations in each local neighbourhood. This option will be ignored if set to *0* or if *regularize=False*. Default is *20+p+1*.

nUniqueMin: Nonnegative integer Specifies the minimum number of unqieu observations in each local neighbourhood. This option will be ignored if set to *0* or if *regularize=False*. Default is *20+p+1*.

bino: Boolean (Default True). Specifies whether to conduct binomial tests. By default the initial (smallest) window contains 20 observations, and its length is also used as the increment for subsequent windows.

binoW: Numeric. 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.

binoN: Nonnegative integer. Specifies the number of observations (closest to the cutoff) used for the binomial test. This is ignored if *binoW* is provided.

binoWStep: Numeric. Specifies the increment in half lengths.

binoNStep: Nonnegative integer. Specifies the increment in sample size. This is ignored if *binoWStep* is provided.

binoNW: Nonnegative integer. Specifies the total number of windows. Default is *10*.

binoP: Numeric. Specifies the null hypothesis of the binomial test. Default is *0.5*.

Returns

hat: left/right: density estimate to the left/right of the cutoff. diff: difference in estimated densities on the two sides of the cutoff.

sd_asy: left/right: standard error for the estimated density to the left/right of the cutoff, diff: standard error for difference in estimated densities on the two sides of the cutoff. (based on asymptotic method)

sd_jk: left/right: standard error for the estimated density to the left/right of the cutoff, diff: standard error for difference in estimated densities on the two sides of the cutoff. (based on jackknife method)

test: t_asy/t_jk: t statistic for the density discontinuity test. p_asy/p_jk: p-value for the density discontinuity test.

hat_p: Same as hat, without bias correction.

bino: Binomial test results.

h: bandwidth used to the left/right of the cutoff.

n: full: full sample size, left/right: sample size to the left/right of the cutoff.

X_min: Smallest observations to the left and right of the cutoff.

X_max: Largest observations to the left and right of the cutoff.

options: other options passed to the function are also stored within the object.

See also:

rddensity.rdbwdensity

rddensity.rdplotdensity

Example

```
>>> import numpy as np
>>> from rddensity import rddensity
>>> data = np.random.normal(-0.5,1,2000)
>>> rdd = rddensity(X=data, vce="jackknife")
>>> print(repr(rdd))
```

2.1.2 rdbwdensity

Bandwidth selection for manipulation testing

Description

`rdbwdensity` implements several data-driven bandwidth selection methods useful to construct manipulation testing procedures using the local polynomial density estimators proposed in Cattaneo, Jansson and Ma (2020).

Related `Stata` and `R` useful for inference in regression discontinuity (RD) designs are available on the rdpackages website.

Companion commands: `rddensity` for estimation and `rdplotdensity` for plotting estimation results.

References

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. 2020. Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 115(531): 1449-1455.

Authors

Matias D. Cattaneo, Princeton University. (cattaneo@princeton.edu).

Rajita Chandak (maintainer), Princeton University. (rchandak@princeton.edu).

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

Xinwei Ma (maintainer), University of California San Diego. (x1ma@ucsd.edu).

class rddensity.rdbwdensity.**bw_output**(*h*, *n*, *fitselect*, *kernel*, *vce*, *c*, *p*, *regularize*, *nLocalMin*, *nUniqueMin*, *massPoints*, *massPoints_flag*, *X_min*, *X_max*)

Class of rdbwdensity function outputs. Object type returned by *rdbwdensity()*.

rddensity.rdbwdensity.**rdbwdensity**(*X*, *c=0*, *p=2*, *fitselect='unrestricted'*, *kernel='triangular'*, *vce='jackknife'*, *massPoints=True*, *regularize=True*, *nLocalMin=None*, *nUniqueMin=None*)

Parameters

X: Numeric vector or one dimentional matrix/dataframe the running variable.

c: Numeric specifies the threshold or cutoff value in the support of *X*. Default is *0*.

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

fitselect: String specifies the density estimation method. *unrestricted* (Default) for density estimation without any restrictions (two-sample, unrestricted inference). *restricted* for density estimation assuming equal distribution function and higher order dericatives.

kernel: String specifies the kernel function used to construct the local polynomial estimators. Accepted kernels: *triangular* (Default), *epanechnikov* or *uniform*.

vce: String specifies the procedure used to compute the variance-covariance matrix estimatior. *jackknife* (Default) for jackknife standard errors or *plugin* for asymptotic plug-in standard errors.

massPoints: Boolean, Default *True*. Specifies wether to adjust for mass points in the data.

regularize: Boolean, Default *True*. Specifies whether to conduct local sample size checking. When True, the bandwidth is chosen such that the local region includes at least *nLocalMin* observations and at least *nUniqueMin* unique observations.

nLocalMin: Nonnegative integer specifies the minimum number of observations in each local neighbourhood. This option will be ignored if set to *0* or if *regularize=False*. Default is *20+p+1*.

nUniqueMin: Nonnegative integer specifies the minimum number of unqieu observations in each local neighbourhood. This option will be ignored if set to *0* or if *regularize=False*. Default is *20+p+1*.

Returns

h Bandwidths for density discontinuity test, left and right of the cutoff, asymptotic variance and bias.

n full-full sample size, *left*/*right*: sample size to the left/right of the cutoff.

X_min Smallest observations to the left and right of the cutoff.

X_max Largest observations to the left and right of the cutoff.

options other options passed to the function are also stored within the object.

See also:

rddensity.rddensity

rddensity.rdplotdensity

Example

```
>>> import numpy as np
>>> from rddensity import rdbwdensity
>>> data = np.random.normal(-0.5,1,2000)
>>> est = rdbwdensity(data=data, vce="jackknife")
>>> print(repr(est))
```

2.1.3 rdplotdensity

Density plotting for mainpulation testing

Description

`rdplotdensity` constructs density plots. It is based on the local polunomial density estimator proposed in Cattaneo, Jansson and Ma (2020, 2021a).

Companion commands: `rdbwdensity` for bandwidth selection and `rddensity` for estimation.

Details

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 an 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 bandwdith 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 (2020a, 2020b) 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.

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., 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. 2020. Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 115(531): 1449-1455.

Cattaneo, M. D., M. Jansson, and X. Ma. 2021a. Local Regression Distribution Estimators *Journal of Econometrics*, forthcoming.

Cattaneo M. D., M. Jansson, and X. Ma. 2022. lpdensity: Local Polynomial Density Estimation and Inference. *Journal of Statistical Software* Forthcoming.

Authors

Matias D. Cattaneo, Princeton University. (cattaneo@princeton.edu).

Rajita Chandak (maintainer), Princeton University. (rchandak@princeton.edu).

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

Xinwei Ma (maintainer), University of California San Diego. (x1ma@ucsd.edu).

rddensity.rdplotdensity.**rdplotdensity**(*rdd*, *X*, *plotRange=None*, *plotN=[10]*, *plotGrid=['es', 'qs']*, *alpha=0.05*, *plottype='line'*, *CItype='region'*, *CIuniform=False*, *CIsimul=2000*, *CIshade=0.2*, *bwselect=None*, *hist=True*, *histBreaks=None*, *histfillshade=0.1*, *histlinecol='white'*, *title=None*, *xlabel=None*, *ylabel=None*, *legendtitle=None*, *legendgroups=None*)

Parameters

rdd [*rddensity* object] returned by *rddensity*.

X: Numeric vector or one dimentional matrix/dataframe the running variable.

plotRange: Numeric. Specifies the lower and upper bound of the plotting region. Default is *[c-3hl, c+3hr]*

plotN: Numeric. Specifies the bumber of grid points used for plotting on the two sides of the cutoff. Default is 10 on each side.

plotGrid: String. Specifies position of grid points. *es* for evenly spaced, *qs* for quantile spaced.

alpha: Numeric scalar between 0 and 1. The significance level for plotting confidence regions.

plottype: String. *"line"*, *"points"* or *"both"* specifies how the estiamtes are plotted.

CItype: String. *"region"* (default), *"line"* or *"ebar"*, how the confidence region will be plotted.

CIuniform: Boolean (default False) plotting pointwise confidence intervals or uniform confidence bands.

CIsimul: Positive integer. Number of simulations used to construct confidence intervals (default 2000). Ignored if *CIuniform* is False.

CIshade: Numeric, between 0 and 1. Opaquness of confidence region. Default is *0.2*.

bwselect: String. Method for data-driven bandwidth selection. Default uses bandwidth from *rdd*. *"mse-dpi"*- mean squared error-optimal bandwidth selected for each grid points, *"imse-dpi"*- integrated MSE-optimal bandiwdth, common for all grid points, *"mse-rot"*- rule-of-thumb bandiwdth with Gaussian reference model, *"imse-rot"*-integrated rule-of-thumb bandiwdht with Gaussian reference model.

hist: Boolean (default True). Adds histgram in background of plot.

histBreaks: Numeric vector. breakpoints between histogram bars.

histfillshade: Numeric between 0 and 1. Opaqueness of histrogram. Default is *0.1*.

title: String. Title of the plot

xlabel: String. Label for x-axis.

ylabel: String. Label for y-axis.

legendTitle: String. Title of legend.

Returns

plot: plotnine object. Can be customized further with plotnine options.

See also:

rddensity.rddensity

rddensity.rdbwdensity

Example

```
>>> import numpy as np
>>> from rddensity import rddensity
>>> from rddensity import rdplotdensity
>>> data = np.random.normal(-0.5,1,2000)
>>> rdd = rddensity(X=data, vce="jackknife")
>>> plot1 = rdplotdensity(rdd, data)
```

PYTHON MODULE INDEX

r

INDEX