Package ‘rddensity’

July 13, 2019

Type Package

Title Manipulation Testing Based on Density Discontinuity

Description Density discontinuity test (a.k.a. manipulation test) is commonly employed in regression discontinuity designs and other program evaluation settings to detect whether there is evidence of perfect self-selection (manipulation) around a cutoff where a treatment/policy assignment changes. This package provides tools for conducting the aforementioned statistical test: rd-density() to construct local polynomial based density discontinuity test given a prespecified cutoff, rdbwdensity() to perform bandwidth selection, and rdplotdensity() to construct density plot near the cutoff.

Version 1.0

Author Matias D. Cattaneo, Michael Jansson, Xinwei Ma

Maintainer Xinwei Ma <x1ma@ucsd.edu>

Imports ggplot2, lpdensity

Depends R (>= 3.1.0)

License GPL-2

Encoding UTF-8

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-07-12 22:14:48 UTC

R topics documented:

  rddensity-package .................................................. 2
  rdbwdensity .......................................................... 2
  rddensity ............................................................ 4
  rddensity_senate ..................................................... 6
  rdplotdensity ......................................................... 7

Index 10
rddensity-package

rddensity: Manipulation Testing Based on Density Discontinuity

Description

Density discontinuity test (a.k.a. manipulation test) is commonly employed in regression discontinuity designs and other program evaluation settings to detect whether there is evidence of perfect self-selection (manipulation) around a cutoff where a treatment/policy assignment changes.

This package provides tools for conducting the aforementioned statistical test: rddensity to construct local polynomial based density discontinuity test given a prespecified cutoff, rdbwdensity to perform data-driven bandwidth selection, and rdplotdensity to construct density plot near the cutoff. For a review on manipulation testing see McCrary (2008).

For more details, and related Stata and R packages useful for analysis of RD designs, visit https://sites.google.com/site/rdpackages.

Author(s)

Matias D. Cattaneo, Princeton University <cattaneo@princeton.edu>.  
Michael Jansson, University of California Berkeley. <mjansson@econ.berkeley.edu>.  
Xinwei Ma (maintainer), University of California San Diego. <x1ma@ucsd.edu>.

References


rdbwdensity

Bandwidth Selection for Manipulation Testing Using Local Polynomial Density Estimation

Description

rdbwdensity implements several data-driven bandwidth selection methods for the manipulation testing procedure of Cattaneo, Jansson and Ma (2019).

Companion command: rddensity for density discontinuity (manipulation) testing. A companion Stata package is described in Cattaneo, Jansson and Ma (2018).

Related Stata and R packages useful for inference in regression discontinuity (RD) designs are described at https://sites.google.com/site/rdpackages.
Usage

```
rdbwdenisty(x, c = 0, p = 2, kernel = "", fitselect = "", vce = ")
```

Arguments

- **x**: Numeric vector or one dimensional matrix / data frame, the running variable.
- **c**: Numeric, specifies the threshold or cutoff value in the support of $x$, which determines the two samples (e.g., control and treatment units in RD settings). Default is 0.
- **p**: Integer, specifies the order of the local-polynomial used to construct the density point estimators. Default is 2 (local quadratic approximation).
- **kernel**: String, specifies the kernel function used to construct the local-polynomial estimator(s). Options are: "triangular", "epanechnikov", and "uniform". Default is "triangular".
- **fitselect**: String, 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.
- **vce**: String, 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.

Value

- **h**: Bandwidths for density discontinuity test, left and right to the cutoff, and asymptotic variance and bias.
- **N**: full: full sample size; left/right: sample size to the left/right of the cutoff.
- **opt**: Collects the options used, includes: fitselect, kernel, vce, c, p. See options for rdbwdenisty.
- **X_min**: Smallest observations to the left and right of the cutoff.
- **X_max**: Largest observations to the left and right of the cutoff.

Author(s)

Matias D. Cattaneo, Princeton University <cattaneo@princeton.edu>.
Michael Jansson, University of California Berkeley. <mjansson@econ.berkeley.edu>.
Xinwei Ma (maintainer), University of California San Diego. <x1ma@ucsd.edu>.

References


rddensity

See Also

rddensity

Examples

# Generate a random sample
set.seed(42); x <- rnorm(2000, mean = -0.5)

# Construct bandwidth
summary(rdbwdensity(x, vce="jackknife"))

rddensity Manipulation Testing Using Local Polynomial Density Estimation

Description

rddensity implements manipulation testing procedures using the local polynomial density estimator proposed in Cattaneo, Jansson and Ma (2019). For a review on manipulation testing see McCrary (2008).

Companion command: rdbwdensity for data-driven bandwidth selection, and rdplotdensity for density plot. A companion Stata package is described in Cattaneo, Jansson and Ma (2018). Related Stata and R packages useful for inference in regression discontinuity (RD) designs are described at https://sites.google.com/site/rdpackages.

Usage

rddensity(X, c = 0, p = 2, q = 0, kernel = "", fitselect = "", h = c(), bwselect = "", vce = "", all = FALSE)

Arguments

X Numeric vector or one dimensional matrix/data frame, the running variable.

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

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

q Integer, specifies the order of the local-polynomial used to construct the bias-corrected density point estimators. Default is p+1 (local cubic approximation).

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

**h** Numeric, specifies the bandwidth used to construct the density estimators on the two sides of the cutoff. If not specified, the bandwidth is computed by the companion command `rdbwden`. 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, specifies the bandwidth selection procedure to be used. Options are: "each" bandwidth selection based on MSE of each density separately (two distinct bandwidths). "diff" bandwidth selection based on MSE of difference of densities (one common bandwidth). "sum" bandwidth selection based on MSE of sum of densities (one common bandwidth). "comb" (this is the default option) 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).

**vce** String, 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** Boolean, if specified, `rddensity` reports two testing procedures (given choices `fitselect` and `bwselect`): Conventional test statistic (not valid when using MSE-optimal bandwidth choice). Robust bias-corrected statistic.

### Value

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

**sd_asy** left/right: standard error for the estimated density to the left/right of the cutoff; diff: standard error for the difference in estimated densities. (Based on asymptotic formula.)

**sd_jk** left/right: standard error for the estimated density to the left/right of the cutoff; diff: standard error for the difference in estimated densities. (Based on the jackknife method.)

**test** t_asy/t_jk: t-statistic for the density discontinuity test, with standard error based on asymptotic formula or the jackknife; p_asy/p_jk: p-value for the density discontinuity test, with standard error based on asymptotic formula or the jackknife.

**hat_p** Same as hat, without bias correction (only available when all=TRUE).

**sd_asy_p** Same as sd_asy, without bias correction (only available when all=TRUE).

**sd_jk_p** Same as sd_jk, without bias correction (only available when all=TRUE).

**test_p** Same as test, without bias correction (only available when all=TRUE).

**N** full: full sample size; left/right: sample size to the left/right of the cutoff; eff_left/eff_right: effective sample size to the left/right of the cutoff (this depends on the bandwidth).
rddensity_senate

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

opt
Collects the options used, includes: fitselect, kernel, bwselect1, bwselect, hscale, vce, c, p, q, all. See options for rddensity.

X_min
left/right: the smallest observation to the left/right of the cutoff.

X_max
left/right: the largest observation to the left/right of the cutoff.

Author(s)
Matias D. Cattaneo, Princeton University <cattaneo@princeton.edu>.
Michael Jansson, University of California Berkeley. <mjansson@econ.berkeley.edu>.
Xinwei Ma (maintainer), University of California San Diego. <x1ma@ucsd.edu>.

References

See Also
rdbwdensity, rdplotdensity

Examples
# Continuous Density
set.seed(42)
x <- rnorm(2000, mean = -0.5)
summary(rddensity(x = x, vce="jackknife"))

# Discontinuous density
x[x>0] <- x[x>0] * 2
summary(rddensity(x = x, vce="jackknife"))

rddensity_senate
RD Senate Data

Description
Extract of the dataset constructed by Cattaneo, Frandsen, and Titiunik (2015), which include measures of incumbency advantage in the U.S. Senate for the period 1914-2010.
rdplotdensity

Format

Numeric vector containing 1390 observations:

margin Numeric vector. See Cattaneo, Frandsen and Titiunik (2015) regarding details about this dataset.

Source


rdplotdensity | Density Plot

Description

rdplotdensity constructs density plot near the cutoff. It is based on the local polynomial density estimator proposed in Cattaneo, Jansson and Ma (2019).

Companion command: rddensity for manipulation testing. A companion Stata package is described in Cattaneo, Jansson and Ma (2018).

Related Stata and R packages useful for inference in regression discontinuity (RD) designs are described at https://sites.google.com/site/rdpackages.

Usage

rdplotdensity(rdd, X, plotRange = NULL, plotN = 10, plotGrid = c("es", "qs"), alpha = 0.05, type = NULL, Ctype = NULL, title = "", xlabel = "", ylabel = "", lty = NULL, lwd = NULL, lcol = NULL, pty = NULL, pwd = NULL, pcol = NULL, Cshade = NULL, Cicol = NULL, legendTitle = NULL, legendGroups = NULL)

Arguments

rdd Object returned by rddensity
X Numeric vector or one dimensional matrix / data frame, the running variable.
plotRange Numeric, specifies the lower and upper bound for density plot. By default it is three bandwidths around the cutoff.
plotN Numeric, specifies the number of grid points used for density plot on each side. If more than one is provided, they will be applied to the two sides accordingly. By default 10 points are used on each side.
plotGrid String, specifies the position of grid points. Can be either evenly spaced (default, "es"), or quantile spaced ("qs").
alpha
Numeric scalar between 0 and 1, the significance level for plotting confidence regions. If more than one is provided, they will be applied to the two sides accordingly.

type
String, one of "line" (default), "points" or "both", how the point estimates are plotted. If more than one is provided, they will be applied to the two sides accordingly.

CItypex
String, one of "region" (shaded region, default), "line" (dashed lines), "ebar" (error bars), "all" (all of the previous) or "none" (no confidence region), how the confidence region should be plotted. If more than one is provided, they will be applied to the two sides accordingly.

title, xlabel, ylabel
Strings, title of the plot and labels for x- and y-axis.

1ty
Line type for point estimates, only effective if type is "line" or "both". 1 for solid line, 2 for dashed line, 3 for dotted line. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to the two sides accordingly.

1wd
Line width for point estimates, only effective if type is "line" or "both". Should be strictly positive. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to the two sides accordingly.

1col
Line color for point estimates, only effective if type is "line" or "both". 1 for black, 2 for red, 3 for green, 4 for blue. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to the two sides accordingly.

pty
Scatter plot type for point estimates, only effective if type is "points" or "both". For options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to the two sides accordingly.

pwd
Scatter plot size for point estimates, only effective if type is "points" or "both". Should be strictly positive. If more than one is provided, they will be applied to the two sides accordingly.

pcol
Scatter plot color for point estimates, only effective if type is "points" or "both". 1 for black, 2 for red, 3 for green, 4 for blue. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to the two sides accordingly.

CIfade
Numeric, opaqueness of the confidence region, should be between 0 (transparent) and 1. Default is 0.2. If more than one is provided, they will be applied to the two sides accordingly.

Cicol
Color for confidence region. 1 for black, 2 for red, 3 for green, 4 for blue. For other options, see the instructions for ggplot2 or par. If more than one is provided, they will be applied to the two sides accordingly.

legendTitle
String, title of legend.

legendGroups
String Vector, group names used in legend.
Value

Est1, Estr: Matrices containing estimation results on the two side, with (1) grid (grid points), (2) bw (bandwidths), (3) nh (effective/local sample sizes), (4) f_p (point estimates with p-th order local polynomial), (5) f_q (point estimates with q-th order local polynomial, only if option q is nonzero), (6) se_p (standard error corresponding to f_p), and (7) se_q (standard error corresponding to f_q).

Estplot: A standard ggplot object is returned, hence can be used for further customization.

Author(s)

Matias D. Cattaneo, Princeton University <cattaneo@princeton.edu>.
Michael Jansson, University of California Berkeley. <mjansson@econ.berkeley.edu>.
Xinwei Ma (maintainer), University of California San Diego. <xma@ucsd.edu>.

References


See Also

rddensity

Examples

# Generate a random sample with a discontinuous density
set.seed(42)
x <- rnorm(2000, mean = -0.5); x[x>0] <- x[x>0] * 2

# Estimation
rdd <- rddensity(X = x)

# Plot
plot <- rdplotdensity(rdd, x, plotRange = c(-2, 2), plotN = 25)
Index

_PACKAGE (rddensity-package), 2

ggplot2, 8

margin (rddensity_senate), 6

par, 8

rdbwdensity, 2, 2, 4–6
rddensity, 2, 4, 4, 7, 9
rddensity-package, 2
rddensity_senate, 6
rdplotdensity, 2, 4, 6, 7