Package: rdd (via r-universe)

September 3, 2024

Maintainer Drew Dimmery <drewd@nyu.edu>

Author Drew Dimmery

Version 0.57

License Apache License (== 2.0)

Title Regression Discontinuity Estimation

Description Provides the tools to undertake estimation in Regression Discontinuity Designs. Both sharp and fuzzy designs are supported. Estimation is accomplished using local linear regression. A provided function will utilize Imbens-Kalyanaraman optimal bandwidth calculation. A function is also included to test the assumption of no-sorting effects.

Type Package

Date 2016-03-14

Depends R (>= 2.15.0), sandwich, lmtest, AER, Formula

Collate 'kernelwts.R' 'DCdensity.R' 'IKbandwidth.R' 'RDestimate.R' 'plot.RD.R' 'print.RD.R' 'rdd-package.R' 'summary.RD.R'

RoxygenNote 5.0.1

NeedsCompilation no

Date/Publication 2016-03-14 23:46:03

Repository https://ddimmery.r-universe.dev

RemoteUrl https://github.com/cran/rdd

RemoteRef HEAD

RemoteSha e531ccb64f9de647f18e7a20cfdaae8b202dea89

Contents

`rdd-package` *Regression Discontinuity Estimation Package*

Description

Regression discontinuity estimation package

Details

`rdd` supports both sharp and fuzzy RDD utilizing the **AER** package for 2SLS regression under the fuzzy design. Local linear regressions are performed to either side of the cutpoint using the Imbens-Kalyanamaran optimal bandwidth calculation, `IKbandwidth`.

Author(s)

Drew Dimmery <drewd@nyu.edu>

See Also

`RDestimate`, `DCdensity`, `IKbandwidth`, `summary.RDplot.RD`, `kernelwts`

`DCdensity` *McCrary Sorting Test*

Description

`DCdensity` implements the McCrary (2008) sorting test.

Usage

```
DCdensity(runvar, cutpoint, bin = NULL, bw = NULL, verbose = FALSE,
  plot = TRUE, ext.out = FALSE, htest = FALSE)
```

Arguments

runvar	numerical vector of the running variable
cutpoint	the cutpoint (defaults to 0)
bin	the binwidth (defaults to 2*sd(runvar)*length(runvar)^(-.5))
bw	the bandwidth to use (by default uses bandwidth selection calculation from McCrary (2008))
verbose	logical flag specifying whether to print diagnostic information to the terminal. (defaults to FALSE)
plot	logical flag indicating whether to plot the histogram and density estimations (defaults to TRUE). The user may wrap this function in additional graphical options to modify the plot.
ext.out	logical flag indicating whether to return extended output. When FALSE (the default) DCdensity will return only the p-value of the test. When TRUE, DCdensity will return the additional information documented below.
htest	logical flag indicating whether to return an "htest" object compatible with base R's hypothesis test output.

Value

If ext.out is FALSE, only the p value will be returned. Additional output is enabled when ext.out is TRUE. In this case, a list will be returned with the following elements:

theta	the estimated log difference in heights at the cutpoint
se	the standard error of theta
z	the z statistic of the test
p	the p-value of the test. A p-value below the significance threshhold indicates that the user can reject the null hypothesis of no sorting.
binsize	the calculated size of bins for the test
bw	the calculated bandwidth for the test
cutpoint	the cutpoint used
data	a dataframe for the binning of the histogram. Columns are cellmp (the midpoints of each cell) and cellval (the normalized height of each cell)

Author(s)

Drew Dimmery <<drewd@nyu.edu>>

References

McCrary, Justin. (2008) "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*. 142(2): 698-714. http://dx.doi.org/10.1016/j.jeconom.2007.05.005

Examples

```
#No discontinuity
x<-runif(1000,-1,1)
DCdensity(x,0)

#Discontinuity
x<-runif(1000,-1,1)
x<-x+2*(runif(1000,-1,1)>0&x<0)
DCdensity(x,0)
```

IKbandwidth *Imbens-Kalyanaraman Optimal Bandwidth Calculation*

Description

IKbandwidth calculates the Imbens-Kalyanaraman optimal bandwidth for local linear regression in Regression discontinuity designs.

Usage

```
IKbandwidth(X, Y, cutpoint = NULL, verbose = FALSE, kernel = "triangular")
```

Arguments

X	a numerical vector which is the running variable
Y	a numerical vector which is the outcome variable
cutpoint	the cutpoint
verbose	logical flag indicating whether to print more information to the terminal. Default is FALSE.
kernel	string indicating which kernel to use. Options are "triangular" (default and recommended), "rectangular", "epanechnikov", "quartic", "triweight", "tricube", "gaussian", and "cosine".

Value

The optimal bandwidth

Author(s)

Drew Dimmery <<drewd@nyu.edu>>

References

Imbens, Guido and Karthik Kalyanaraman. (2009) "Optimal Bandwidth Choice for the regression discontinuity estimator," *NBER Working Paper Series*. 14726. http://www.nber.org/papers/w14726

kernelwts *Kernel Weighting function*

Description

This function will calculate the appropriate kernel weights for a vector. This is useful when, for instance, one wishes to perform local regression.

Usage

```
kernelwts(X, center, bw, kernel = "triangular")
```

Arguments

X	input x values. This variable represents the axis along which kernel weighting should be performed.
center	the point from which distances should be calculated.
bw	the bandwidth.
kernel	a string indicating the kernel to use. Options are "triangular" (the default), "epanechnikov", "quartic", "triweight", "tricube", "gaussian", and "cosine".

Value

A vector of weights with length equal to that of the X input (one weight per element of X).

Author(s)

Drew Dimmery <<drewd@nyu.edu>>

Examples

```
require(graphics)

X<-seq(-1,1,.01)
triang.wts<-kernelwts(X,0,1,kernel="triangular")
plot(X,triang.wts,type="l")

cos.wts<-kernelwts(X,0,1,kernel="cosine")
plot(X,cos.wts,type="l")
```

plot.RD *Plot of the Regression Discontinuity*

Description

Plot the relationship between the running variable and the outcome

Usage

```
## S3 method for class 'RD'
plot(x, gran = 400, bins = 100, which = 1, range, ...)
```

Arguments

x	rd object, typically the result of RDestimate
gran	the granularity of the plot. This specifies the number of points to either side of the cutpoint for which the estimate is calculated.
bins	if the dependent variable is binary, include the number of bins within which to average
which	identifies which of the available plots to display. For a sharp design, the only possibility is 1, the plot of the running variable against the outcome variable. For a fuzzy design, an additional plot, 2, may also be displayed, showing the relationship between the running variable and the treatment variable. Both plots may be displayed with which=c(1,2).
range	the range of values of the running variable for which to plot. This should be a vector of length two of the format c(min,max). To plot from the minimum to the maximum value, simply enter c("min","max"). The default is a window 20 times wider than the first listed bandwidth from the rd object, truncated by the min/max values of the running variable from the data.
...	unused

Details

It is important to note that this function will only plot the discontinuity using the bandwidth which is first in the vector of bandwidths passed to RDestimate

Author(s)

Drew Dimmery <<drewd@nyu.edu>>

print.RD *Print the Regression Discontinuity*

Description

Print a very basic summary of the regression discontinuity

Usage

```
## S3 method for class 'RD'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

x	rd object, typically the result of RDestimate
digits	number of digits to print
...	unused

Author(s)

Drew Dimmery <<drewd@nyu.edu>>

RDestimate *Regression Discontinuity Estimation*

Description

RDestimate supports both sharp and fuzzy RDD utilizing the **AER** package for 2SLS regression under the fuzzy design. Local linear regressions are performed to either side of the cutpoint using the Imbens-Kalyanaraman optimal bandwidth calculation, IKbandwidth.

Usage

```
RDestimate(formula, data, subset = NULL, cutpoint = NULL, bw = NULL,
  kernel = "triangular", se.type = "HC1", cluster = NULL,
  verbose = FALSE, model = FALSE, frame = FALSE)
```

Arguments

formula	the formula of the RDD. This is supplied in the format of y ~ x for a simple sharp RDD, or y ~ x | c1 + c2 for a sharp RDD with two covariates. Fuzzy RDD may be specified as y ~ x + z where x is the running variable, and z is the endogenous treatment variable. Covariates are then included in the same manner as in a sharp RDD.
data	an optional data frame

subset	an optional vector specifying a subset of observations to be used
cutpoint	the cutpoint. If omitted, it is assumed to be 0.
bw	a numeric vector specifying the bandwidths at which to estimate the RD. If omitted, the bandwidth is calculated using the Imbens-Kalyanaraman method, and then estimated with that bandwidth, half that bandwidth, and twice that bandwidth. If only a single value is passed into the function, the RD will similarly be estimated at that bandwidth, half that bandwidth, and twice that bandwidth.
kernel	a string specifying the kernel to be used in the local linear fitting. "triangular" kernel is the default and is the "correct" theoretical kernel to be used for edge estimation as in RDD (Lee and Lemieux 2010). Other options are "rectangular", "epanechnikov", "quartic", "triweight", "tricube", "gaussian" and "cosine".
se.type	this specifies the robust SE calculation method to use. Options are, as in vcovHC, "HC3", "const", "HC", "HC0", "HC1", "HC2", "HC4", "HC4m", "HC5". This option is overriden by cluster.
cluster	an optional vector specifying clusters within which the errors are assumed to be correlated. This will result in reporting cluster robust SEs. This option overrides anything specified in se.type. It is suggested that data with a discrete running variable be clustered by each unique value of the running variable (Lee and Card 2008).
verbose	will provide some additional information printed to the terminal.
model	logical. If TRUE, the model object will be returned.
frame	logical. If TRUE, the data frame used in model fitting will be returned.

Details

Covariates are problematic for inclusion in the regression discontinuity design. This package allows their inclusion, but cautions against them insomuch as is possible. When covariates are included in the specification, they are simply included as exogenous regressors. In the sharp design, this means they are simply added into the regression equation, uninteracted with treatment. Likewise for the fuzzy design, in which they are added as regressors in both stages of estimation.

Value

RDestimate returns an object of class "RD". The functions summary and plot are used to obtain and print a summary and plot of the estimated regression discontinuity. The object of class RD is a list containing the following components:

type	a string denoting either "sharp" or "fuzzy" RDD.
est	numeric vector of the estimate of the discontinuity in the outcome under a sharp design, or the Wald estimator in the fuzzy design for each corresponding bandwidth
se	numeric vector of the standard error for each corresponding bandwidth
z	numeric vector of the z statistic for each corresponding bandwidth
p	numeric vector of the p value for each corresponding bandwidth
ci	the matrix of the 95 for each corresponding bandwidth

bw	numeric vector of each bandwidth used in estimation
obs	vector of the number of observations within the corresponding bandwidth
call	the matched call
na.action	the observations removed from fitting due to missingness
model	(if requested) For a sharp design, a list of the lm objects is returned. For a fuzzy design, a list of lists is returned, each with two elements: firststage, the first stage lm object, and iv, the ivreg object. A model is returned for each corresponding bandwidth.
frame	(if requested) Returns the model frame used in fitting.

Author(s)

Drew Dimmery <<drewd@nyu.edu>>

References

Lee, David and Thomas Lemieux. (2010) "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*. 48(2): 281-355. http://www.aeaweb.org/articles.php?doi=10.1257/jel.48.2.281

Imbens, Guido and Thomas Lemieux. (2010) "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*. 142(2): 615-635. http://dx.doi.org/10.1016/j.jeconom.2007.05.001

Lee, David and David Card. (2010) "Regression discontinuity inference with specification error," *Journal of Econometrics*. 142(2): 655-674. http://dx.doi.org/10.1016/j.jeconom.2007.05.003

Angrist, Joshua and Jorn-Steffen Pischke. (2009) *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

See Also

summary.RD, plot.RD, DCdensity IKbandwidth, kernelwts, vcovHC, ivreg, lm

Examples

```
x<-runif(1000,-1,1)
cov<-rnorm(1000)
y<-3+2*x+3*cov+10*(x>=0)+rnorm(1000)
RDestimate(y~x)
# Efficiency gains can be made by including covariates
RDestimate(y~x|cov)
```

summary.RD *Summarizing Regression Discontinuity Designs*

Description

summary method for class "RD"

Usage

```
## S3 method for class 'RD'
summary(object, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

object	an object of class "RD", usually a result of a call to RDestimate
digits	number of digits to display
...	unused

Value

summary.RD returns an object of class "summary.RD" which has the following components:

coefficients	A matrix containing bandwidths, number of observations, estimates, SEs, z-values and p-values for each estimated bandwidth.
fstat	A global F-test of the corresponding model

Author(s)

Drew Dimmery <<drewd@nyu.edu>>

Index