R-Package 'MIdiagRDD'

January 14, 2021

Type Package

Title Diagnostic Tool by Multiple Imputation for Regression Discontinuity Designs

Version 0.0.1

Date January 14, 2021

Author Masayoshi Takahashi

Description `R`-Package MIdiagRDD estimates the local average treatment effects based on a regression discontinuity design (RDD) and a multiple imputation discontinuity design (MIDD), and diagnoses RDD by comparing the results from MIDD.

Depends Amelia, rdrobust

URL https://github.com/m-takahashi123/MIdiagRDD

Repository GitHub

R topics documented:

MIdiagRDD-package	*Diagnostic Tool by Multiple Imputation for Regression Discontinuity Designs*

Description

To use this package, click Code and Download ZIP from https://github.com/review538/MIdiagRDD. After downloading the package, set the working directory in R, and read R-Package MIdiagRDD using R-function source as follows.

```
setwd("C:/Folder")
source("MIdiagRDD.R")
```

datalee	*Simulated dataset based on Lee (2008)*

Description

This is a simulated dataset based on the data used in Lee (2008).

Usage

```
data1<-read.csv("datalee.csv", header=TRUE)
attach(data1)
```

Format

A data frame with 6558 observations on the following four variables.

`y1` the variable of interest (the dependent variable)

`x1` the running variable. The cutoff point is where `x1=0`.

`x2` and `x3` additional covariates

References

Imbens, G., and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. The Review of Economic Studies, 79 (3), pp.933-959.

Lee, D. S. (2008). Randomized Experiments from Non-Random Selection in U.S. House Elections. Journal of Econometrics, 142, pp.675-697.

`MIdiagRDD`	*R-function to compute the local average treatment effects and to diagnose a regression discontinuity design by a multiple imputation discontinuity design.*

Description

`MIdiagRDD` estimates the local average treatment effects based on a regression discontinuity design (RDD) and a multiple imputation discontinuity design (MIDD), and diagnoses RDD by comparing the results from MIDD.

Usage

```
MIdiagRDD(y, x, cut, seed=1, M1=100, M2=5, p2s1=1, emp=0,
          bw="mserd", ker="triangular", bwidth=1, p1=1,
          conf=95, upper=1, covs1=NULL)
```

Arguments

`y`	the variable of interest (the dependent variable)
`x`	the running variable (forcing variable) that determines the cutoff point.
`cut`	specifies the RDD cutoff point in `x`. The user must supply a specific number.
`seed`	sets the seed value for random numbers. Default is 1.
`M1`	the number of imputed datasets to create. Default is 100.
`M2`	the number of imputed datasets to display in graphs. Default is 5. These datasets are the subsets of `M1` imputed datasets. Thus, `M2` cannot be larger than `M1`.
`p2s1`	an integer value taking either 0 or 1, where 0 for no screen output and 1 for screen printing of multiple imputation process. Default is 1.
`emp`	number indicating level of the empirical (ridge) prior. Default is 0. A reasonable upper bound is 0.1.
`bw`	specifies the bandwidth selection procedure for the regression discontinuity design. Choice is `mserd`, `msesum`, `cerrd`, and `cersum`. Default is `mserd`. `mserd` is one common MSE-optimal bandwidth selector. `msesum` is one common MSE-optimal bandwidth selector for the sum of regression estimates. `cerrd` is one common CER-optimal bandwidth selector.

`cersum` is one common CER-optimal bandwidth selector for the sum of regression estimates.
MSE is Mean Squared Error. CER is Coverage Error Rate.

`ker`	is the kernel function used to construct the local-polynomial estimator for the RDD. Options are `triangular` (default option), `epanechnikov`, and `uniform`.
`bwidth`	a number to adjust the size of the chosen bandwidth. Default is 1.
`p1`	specifies the order of the local-polynomial used to construct the point-estimator for the RDD and the MIDD. Default is `p1=1` (local linear regression). Can take either 1 (local linear regression) or 2 (local quadratic regression). When specified larger than 2, it will be considered 2.
`conf`	is the confidence level for the confidence interval. Default is 95.
`upper`	specifies which part of the running variable is the treatment group. If the upper part is the treatment group, `upper=1` (default). If the lower part is the treatment group, `upper=0`. Default is 1.
`covs1`	specifies additional covariates to be used for estimation and inference in the RDD and multiple imputation. Adding covariates can be done by specifying `covs1=data.frame(covariate1, covariate2, ..., covariatep)`.

Value

`Estimate`	Estimated quantities of the local average treatment effects (LATE) at the cutoff.
`Std.Error`	Standard error of the estimate.
`CI.LL`	Lower limit of the 95% confidence interval.
`CI.UL`	Upper limit of the 95% confidence interval.
`size`	Sub-sample size to estimate the LATE at the cutoff.
`ratio`	Ratio of the sub-sample size to the original sample size.
`1.MI, RDD, Naive`	A diagnostic plot to visualize the relationship among the three estimators. Red vertical line is RDD, black solid line is naïve, and histogram is MI.
`2.MI and RDD`	A diagnostic plot to visualize the relationship between the two estimators. Red vertical line is RDD and histogram is MI.

`3.Densities (Control)`	A diagnostic plot to visualize the densities of observed and imputed data. Gray solid curve is the density of observed data in the control group. Blue solid curve is the density of observed data in the treatment group. Red dashed lines are the densities of imputed data in the control group.
`4.Densities (Treatment)`	A diagnostic plot to visualize the densities of observed and imputed data. Gray solid curve is the density of observed data in the control group. Blue solid curve is the density of observed data in the treatment group. Red dashed lines are the densities of imputed data in the treatment group.
`5.Observed Values`	A diagnostic plot to visualize the scatterplot of observed data. Gray circles are observed data in the control group. Blue triangles are observed data in the treatment group.
`6.Observed & Imputed Values`	A diagnostic plot to visualize the scatterplot of observed and imputed data. Red circles are imputed data in the control group. Red triangles are imputed data in the treatment group. These imputed data are overlaid on the observed data in Figure 5.
`7.Observed & Imputed (Control)`	A diagnostic plot to clearly visualize the scatterplot of observed and imputed data in the control group only.
`8.Observed & Imputed (Treatment)`	A diagnostic plot to clearly visualize the scatterplot of observed and imputed data in the treatment group only.
`9.Around Cutoff (Control)`	A diagnostic plot to clearly visualize the scatterplot, around the cutoff point, of observed and imputed data in the control group only. Five solid lines are the estimated linear regression lines based on multiply imputed data.
`10.Around Cutoff (Treatment)`	A diagnostic plot to clearly visualize the scatterplot, around the cutoff point, of observed and imputed data in the treatment group only. Five solid lines are the estimated linear regression lines based on multiply

	imputed data.
`11.Local Slope (Control)`	A diagnostic plot to visualize the distribution of the coefficients of the estimated linear regression models around the cutoff point in the control group.
`12.Local Slope (Treatment)`	A diagnostic plot to visualize the distribution of the coefficients of the estimated linear regression models around the cutoff point in the treatment group.

References

Takahashi, M. (2021). Multiple Imputation Discontinuity Design: An Alternative to Regression Discontinuity Designs to Estimate the Local Average Treatment Effect at the Cutoff.

Calonico, S., Cattaneo, M.D., and Titiunik, R. (2015). rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. R Journal 7(1), pp.38-51.

Honaker, J., King, G., and Blackwell, M. (2011). Amelia II: A Program for Missing Data. Journal of Statistical Software, 45(7), pp.1-47.

Example 1

```
setwd("C:/Folder")
source("MIdiagRDD.R")
data1<-read.csv("datalee.csv", header = TRUE)
attach(data1)
MIdiagRDD(y = y1, x = x1, cut = 0)
```

Example 2

```
MIdiagRDD(y = y1, x = x1, cut = 0, M1 = 50, M2 = 3, bw =
          "cerrd", ker = "uniform", bwidth = 0.5, p1 = 2,
          conf = 90, covs1 = data.frame(x2,x3))
```