

Package ‘MIRDD’

May 16, 2026

Type Package

Title Diagnostic Tool by Multiple Imputation for Regression
Discontinuity Designs

Version 0.2.4

Date 2026-05-17

Maintainer Masayoshi Takahashi <mtakahashi615@g.chuo-u.ac.jp>

Description Estimates average treatment effects at the cutoff based on sharp regression discontinuity designs (RDD) and multiple imputation regression discontinuity designs (MIRDD). It provides diagnostic tools for RDD by comparing results with those from MIRDD, as proposed in Takahashi (2023) <[doi:10.1080/03610918.2021.1960374](https://doi.org/10.1080/03610918.2021.1960374)>. The package includes datasets from Takahashi (2023) and Takahashi (2026) <[doi:10.1016/j.softx.2026.102707](https://doi.org/10.1016/j.softx.2026.102707)>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Imports Amelia, rdrobust, stats, graphics

NeedsCompilation no

Depends R (>= 3.5.0)

Author Masayoshi Takahashi [aut, cre] (ORCID:
<<https://orcid.org/0000-0002-2616-0614>>)

Repository CRAN

Date/Publication 2026-05-16 16:20:02 UTC

Contents

MIRDD-package	2
lee2008	2
LudwigMiller2007Modified	3
MIdiagRDD	4

Index	7
--------------	----------

MIRDD-package

MIRDD: Diagnostic Tool by Multiple Imputation for Regression Discontinuity Designs

Description

This package implements the method proposed in Takahashi (2023), which provides a novel framework for the regression discontinuity design (RDD) by reinterpreting the estimation of treatment effects as a missing data problem. While standard RDD relies on observations near the cutoff, MIRDD utilizes multiple imputation to account for missing potential outcomes, offering a diagnostic tool to assess the validity and robustness of RDD estimates.

Details

The main function in this package is [MIdiagRDD](#).

Author(s)

Maintainer: Masayoshi Takahashi <mtakahashi615@g.chuo-u.ac.jp> ([ORCID](#))

References

Takahashi, M. 2023. Multiple imputation regression discontinuity designs: Alternative to regression discontinuity designs to estimate the local average treatment effect at the cutoff. *Communications in Statistics - Simulation and Computation* 52 (9): 4293-4312. doi:[10.1080/03610918.2021.1960374](https://doi.org/10.1080/03610918.2021.1960374)

Takahashi, M. 2026. MIRDD: An R package for multiple imputation regression discontinuity design. *SoftwareX* 34(102707): 1-6. doi:[10.1016/j.softx.2026.102707](https://doi.org/10.1016/j.softx.2026.102707)

lee2008

Lee (2008) Dataset

Description

A dataset used in Takahashi (2023).

Usage

```
data(lee2008)
```

Format

A data frame with 6558 observations on the following 3 variables.

y1 a numeric vector (dependent variable). Democrat vote share election at $t + 1$.

x1 a numeric vector (running variable). The cutoff point is where $x1 = 0$. Democratic vote share at t .

x2 a numeric vector (additional covariate). Democratic vote share at $t - 1$

Source

Lee, D. S. 2008. Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* 142 (2): 675-697.

References

Takahashi, M. 2023. Multiple imputation regression discontinuity designs: Alternative to regression discontinuity designs to estimate the local average treatment effect at the cutoff. *Communications in Statistics - Simulation and Computation* 52 (9): 4293-4312. doi:[10.1080/03610918.2021.1960374](https://doi.org/10.1080/03610918.2021.1960374)

LudwigMiller2007Modified

Ludwig and Miller (2007) Modified Dataset

Description

A modified dataset used in Takahashi (2026).

Usage

```
data(LudwigMiller2007Modified)
```

Format

A data frame with 1037 observations on the following 11 variables.

y1 a numeric vector (dependent variable)

x1 a numeric vector (running variable). The cutoff point is where $x1 = \log(59.1984)$.

z1 a numeric vector (additional covariate)

z2 a numeric vector (additional covariate)

z3 a numeric vector (additional covariate)

z4 a numeric vector (additional covariate)

z5 a numeric vector (additional covariate)

z6 a numeric vector (additional covariate)

z7 a numeric vector (additional covariate)

z8 a numeric vector (additional covariate)

z9 a numeric vector (additional covariate)

Source

Ludwig, J., and D. L. Miller. 2007. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics* 122 (1): 159-208.

References

Takahashi, M. 2026. MIRDD: An R package for multiple imputation regression discontinuity design. *SoftwareX* 34(102707): 1-6. doi:[10.1016/j.softx.2026.102707](https://doi.org/10.1016/j.softx.2026.102707)

MIdiagRDD

Diagnostic Tool by Multiple Imputation for Regression Discontinuity Designs

Description

Estimates average treatment effects at the cutoff based on sharp regression discontinuity designs (RDD) and multiple imputation regression discontinuity designs (MIRDD). It provides diagnostic tools for RDD by comparing results with those from MIRDD.

Usage

```
MIdiagRDD(
  y,
  x,
  cut,
  seed = NULL,
  M1 = 100,
  M2 = 5,
  M3 = 1,
  p2s1 = 1,
  emp = 0,
  bw = "mserd",
  ker = "triangular",
  h = NULL,
  type = "Conventional",
  p1 = 1,
  conf = 95,
  upper = 1,
  covs1 = NULL,
  up = NULL,
  lo = NULL
)
```

Arguments

y	A numeric vector of the outcome variable.
x	A numeric vector of the running variable (forcing variable).
cut	A numeric value indicating the cutoff point in x. The user must supply a specific number.
seed	A seed number for reproducibility. Default is NULL.
M1	Number of imputations for MIRDD. Default is 100.
M2	Number of imputations for visualization (plots 3, 4, 9, and 10). Default is 5. These datasets are the subsets of M1 imputed datasets. Thus, M2 cannot be larger than M1.

M3	Number of imputed datasets for plots 5 to 10. Default is 1. These datasets are the subsets of M1 imputed datasets. Thus, M3 cannot be larger than M1.
p2s1	Integer for Amelia's p2s argument (0 or 1), where 0 for no screen printing and 1 for screen printing of multiple imputation process. Default is 1.
emp	Amelia's empirical (ridge) prior. Default is 0. A reasonable upper bound is 0.1.
bw	Bandwidth selection method for rdrobust. Options are "mserd" (default), "msexum", "cerrd", and "cersum". "mserd" is one common MSE-optimal bandwidth selector. "msexum" 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	Kernel function for rdrobust. Options are "triangular" (default option), "epanechnikov", and "uniform".
h	Number for bandwidth. Default is NULL (data-driven).
type	Inference type: "Conventional" (default), "Bias-Corrected", or "Robust".
p1	Polynomial order (1 or 2) for rdrobust and MIRDD. Default is 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	Confidence level (0-100). Default is 95.
upper	If 1 (default), treatment is $x \geq \text{cut}$. If 0, treatment is $x < \text{cut}$.
covs1	Optional covariates. If two additional covariates z1 and z2 need to be used, then <code>covs1 = data.frame(z1, z2)</code> .
up	Optional upper bound for imputed values.
lo	Optional lower bound for imputed values.

Value

Estimate	Estimated quantities of the average treatment effects (ATE) 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 ATE at the cutoff.
bandwidth	Length of the bandwidth used for RDD analysis.

In addition to the data frame, a series of diagnostic plots are generated:

- 1. MIRDD, RDD, Naive** A diagnostic plot to visualize the relationship among the three estimators. Red vertical line is RDD, black solid line is naive, and histogram is MI.
- 2. MIRDD 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. 2023. Multiple imputation regression discontinuity designs: Alternative to regression discontinuity designs to estimate the local average treatment effect at the cutoff. *Communications in Statistics - Simulation and Computation* 53(9): 4293-4312. doi:10.1080/03610918.2021.1960374
- Takahashi, M. 2026. MIRDD: An R package for multiple imputation regression discontinuity design. *SoftwareX* 34(102707): 1-6. doi:10.1016/j.softx.2026.102707
- Calonico, S., Cattaneo, M.D., and Titiunik, R. 2015. rdrobust: An R Package for robust non-parametric inference in regression-discontinuity designs. *R Journal* 7(1): 38-51. doi:10.32614/RJ2015004
- Honaker, J., King, G., and Blackwell, M. 2011. Amelia II: A program for missing data. *Journal of Statistical Software* 45(7): 1-47. doi:10.18637/jss.v045.i07

Examples

```
# Example usage with dummy data
x <- runif(100, -1, 1)
y <- 0.5 * x + (x >= 0) + rnorm(100, 0, 0.1)
MIdiagRDD(y = y, x = x, cut = 0)
```

Index

* datasets

lee2008, [2](#)

LudwigMiller2007Modified, [3](#)

lee2008, [2](#)

LudwigMiller2007Modified, [3](#)

MIdiagRDD, [2](#), [4](#)

MIRDD (MIRDD-package), [2](#)

MIRDD-package, [2](#)