Statistics/Data Analysis

User: rdpermute
Project: rdpermute

Title

rdpermute —— Permutation Test for RD and RK designs

Syntax

rdpermute depvar runvar [*if*], placebo_disconts(*numlist*) true_discont(*string*) [position_true_discont(#) deriv_discont(#) bw(#) reg(#) linear quad cubic skip_install filename(#) save_path(#) dgp(#) bw_manual(#) fg_bandwidth_scaling(# #) fg_bias_porder(#) fg_f_0(#) fg_density_porder(#) fg_num_bins(#) cct_bw_par(#) cct_reg_par(#) silent]

Description

rdpermute implements permutation tests for regression discontinuity (RD) or regression kink (RK) designs developed in Ganong and Jäger (2017). The code calculates RD or RK estimates at a list of pre-specified placebo discontinuities or kinks and computes both asymptotic and randomization-based p-values. It tests for the sharp null hypothesis of no effect of the policy on the outcome and can accommodate several bandwidth choice, estimation, and inference procedures including **rdrobust** developed by Calonico, Cattaneo and Titiunik (2014a,b).

Options

Required

placebo_disconts defines the locations of placebo kinks.

true_discont defines the integer at which the true kink or discontinuity is located. This value has to appear in the set **placebo_disconts**. If **placebo_disconts** is not generated manually, but automatically (for example by loops), it may happen that the binary representations of **true_discont** differs from its corresponding value in **placebo_disconts**. In this case it is possible to use the parameter **position_true_discont** instead. Unless rdpermute prints an error message, this modification is not necessary.

Optional

position_true_discont(*integer -1***)** Position of the expected discontinuity **true_discont** in the vector **placebo_disconts**. This parameter replaces **true_discont** in the case of binary representation errors.

deriv_discont(*integer 1***)** specifies whether a regression discontinuity (0) or a regression kink (1) design is implemented. Default is the implementation of a regression kink design.

bw(*string***)** defines the bandwidth choice method. **fg_aic** is used as default if no alternative is specified. The possible bandwidth choices are:
- cct: uses the procedures and functions in the **rdbwselect** package developed in Calonico, Cattaneo and Titiunik (2014a,b) as a subroutine. The parameters of **rdbwselect** can be altered with the parameter **cct_bw_par.**
- fg: Bandwidth choice as proposed by Fan and Gijbels (1996). Additional parameters (**fg_bias_p_order**, **fg_density_p_order**, **fg_num_bins**, **fg_f0**, and **fg_bandwidth_scaling**) can be used to alter the calculations.
- fg_aic: Fan and Gijbels (1996) bandwidth choice with automatic selection of **fg_bias_p_order**. Additional parameters (**fg_density_p_order**, **fg_num_bins**, **fg_f0**, and **fg_bandwidth_scaling**) can be used to alter the the calculations.
- manual: Manual choice of a constant bandwidth. The bandwidth can be set with the parameter **manual_bw.**

reg(*string***)** defines the estimation procedure. **regress** is used as default. Valid procedures are **regress** and **cct**.
- cct: uses the function **rdrobust** developed by Calonico, Cattaneo and Titiunik (2014a) as subroutine. The parameters of rdrobust can be altered with **cct_reg_par**.
- regress: uses the Stata regression enviroment **regress**.

linear/quad/cubic specifies that a linear, quadratic, or cubic model be used. **rdpermute** will calculate the p-values for each specified model. If neither linear, quad nor cubic are specified, **rdpermute** will calculate the p-values for all of them automatically.

skip_install skips the installation of required packages. **rdpermute** will try to install all dependent packages automatically using stable, predefined versions. This may not always be possible or desired. **skip_install** suppresses the installation. Attention: Some subroutines and parts of our code may not work if the dependent packages are not installed.

filename(*string***)** Name for final .dta output. Only if **filename** is provided, will the data be saved.

save_path(*string***)** Path for final .dta output. If no **save_path** is provided, the results will automatically be placed in the working directory.

dgp(*string***)** adds a column with an index variable to .dta output

bw_manual(*real 1***)** is a numerical value for the method choice **reg(manual)**. The value will be used as bandwidth for the computation of the p-values for all placebo_disconts.

fg_bandwidth_scaling(*numlist***)** specifies the model-dependent constants for the rule-of-thumb bandwidth calculation formula by Fan and Gijbels (1996). It may be necessary to use other values than our presets for linear, quadratic, and cubic regressions. **fg_bandwidth_scaling[1]** describes the prefactor, **fg_bandwidth_scaling[2]** the used exponents. The parameter **fg_bandwidth_scaling** has to contain values for both entries. All other entries in **fg_bandwidth_scaling** are omitted. A detailed description of the formula can be found in Fan and Gijbels (1996).

fg_bias_porder(*integer 4***)** specifies the maximal order of the polynomial used to estimate m^2 m^3 and m^4 for bandwidth choice **bw(fg)**. This parameter is only necessary if the chosen method is **fg** and not **bw(fg_aic)**. Warning: A high **fg_bias_p_order** may result in the instability of the used regressions, without indication by STATA. The choice bw="fg_aic" will automatically prevent such errors and is therefore set as default.

fg_f_0(*real 0***)** specifies the placement of bins for the choice bw(fg). If not set, 50 equally spaced bins on the range of the running variable will be used. We recommend to leave this parameter empty for an automatic estimation of **bw(fg_f_0)**. If you wish to use a manual value, you can define a numerical value in fg_f_0.

fg_density_porder(*integer 3***)** specifies the polynomial order for density estimation meaning that it chooses the maximal exponent of x^p for the estimation of **bw(fg_f_0)** by regression. Warning: A high **fg_density_p_order** may lead to the same problems as in **fg_bias_p_order**. We recommend to use the preset value.

fg_num_bins(*integer 50***)** specifies the number of equally spaced bins for the choice **bw(fg)** and **fg_f_0(0)** that is used to estimate **fg_f_0**.

cct_bw_par(*string***)** specifies additional or alternative parameters for the subroutine **rdbwselect** for the choice **bw(cct)**. All parameters of **rdbwselect** can be altered except for: y, x, p, q, deriv. To alter an option, define the intended values within html-Tags within the string. Example: **cct_bw_par**("<kernel>epa</kernel><bwselect>cerrd</bwselect>").

cct_reg_par(*string***)** specifies additional or alternative parameters for the subroutine **rdrobust** for the choice **reg(cct)**. All parameters of rdrobust can be altered except for: y, x, p, q, deriv, h. Altering is done as in **cct_bw_par**.

silent generates less output while running.

Examples

```
rdpermute y x, placebo_disconts(-0.9(0.1)0.9) true_discont(0) linear quad silent
    bw(fg) save_path(~/Data/working/) filename(placebo_pvalues) dgp(1)
    fg_density_porder(1)

rdpermute y x, placebo_disconts(-100(10)200) true_discont(20) linear silent
    bw(manual) save_path(~/Data/working/) filename(placebo_pvalues)
    bw_manual(10)

rdpermute y x, placebo_disconts(1960(0.25)2017) true_discont(2000) linear quad
    bw(cct) reg(regress) cct_bw_par(<bwselect>cerrd</bwselect>)
```

Stored Results

rdpermute stores the following in **e()**:

e(kink_beta_linear)
e(kink_se_linear)
e(bw_linear)
e(pval_linear)
e(kink_beta_quadratic)
e(kink_se_quadratic)
e(bw_quadratic)
e(pval_quadratic)
e(kink_beta_cubic)
e(kink_se_cubic)
e(bw_cubic)
e(pval_cubic)

With N as number of placebo kinks, matrices kink* and bw* are Nx2. Column 1 reports output using the "cct" bandwidth choice. Column 2 reports results for the "fg" bandwidth choice.

Matrices pval* are 2 x 2. Row 1 is asymptotic p-value. Row 2 is randomization p-value.

Optional .dta output: collapses all of the above into a single file.

References

Calonico, S., Cattaneo, M. D., and Titiunik, R. "Robust data-driven inference in the regression-discontinuity design." *Stata Journal* 14.4: 909-946 (2014a).

Calonico, S., Cattaneo, M. D., and Titiunik, R. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6):2295-2326 (2014b).

Fan, J. and Gijbels, I. *Local Polynomial Modelling and Its Applications*, volume 66. Chapman and Hall (1996).

Ganong, P. and Jäger, S. "A Permutation Test for the Regression Kink Design." *Journal of the American Statistical Association* (2017).

Nichols, A. "rd 2.0: Revised Stata module for regression discontinuity estimation." (2011).

Online References and Dependent Code:

rdbwselect - Bandwidth Selection Procedures for Local Polynomial Regression Discontinuity Estimators

rdrobust - Local Polynomial Regression Discontinuity Estimation with Robust Bias-Corrected Confidence Intervals and Inference Procedures

rdplot - Data-Driven Regression Discontinuity Plots

rd - Regression discontinuity (RD) estimator

All dependent packages will automatically download at the first run of rdpermute. See **skip_install** for suppressing the installation.

Authors

Peter Ganong, University of Chicago, ganong@uchicago.edu.

Simon Jäger, MIT, sjaeger@mit.edu.

Acknowledgments

Dennis Kubitza and Michael Schöner provided excellent research assistance to develop the Stata package.