The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–18

rdhte: Conditional Average Treatment Effects in RD Designs

Sebastian Calonico
UC Davis
Davis, CA
scalonico@ucdavis.edu

Matias D. Cattaneo
Princeton University
Princeton, NJ
cattaneo@princeton.edu

Max H. Farrell
UC Santa Barbara
Santa Barbara, CA
mhfarrell@gmail.com

Filippo Palomba
Princeton University
Princeton, NJ
fpalomba@princeton.edu

Rocio Titiunik
Princeton University
Princeton, NJ
titiunik@princeton.edu

Abstract. Understanding causal heterogeneous treatment effects based on pre-treatment covariates is a crucial aspect of empirical work. Building on Calonico et al. (2025), this article discusses the software package `rdhte` for estimation and inference of heterogeneous treatment effects in sharp regression discontinuity (RD) designs. The package includes three main commands: `rdhte` conducts estimation and robust bias-corrected inference for heterogeneous RD treatment effects, for a given choice of the bandwidth parameter; `rdbwhte` implements automatic bandwidth selection methods; and `rdhte_lincom` computes point estimates and robust bias-corrected confidence intervals for linear combinations, a post-estimation command specifically tailored to `rdhte`. We also provide an overview of heterogeneous effects for sharp RD designs, give basic details on the methodology, and illustrate using an empirical application. Finally, we discuss how the package `rdhte` complements, and in specific cases recovers, the canonical RD package `rdrobust` (Calonico et al. 2017).

Keywords: st0001, regression discontinuity designs, heterogeneous treatment effects, covariate adjustment.

 st0001

1 Introduction

Studying causal heterogeneous treatment effects based on pretreatment covariates is an essential part of modern empirical work, as it helps uncover fairness concerns, differential impacts, and informs targeted policy interventions. While the regression discontinuity (RD) design has become a widely used tool for causal inference, existing methods primarily focus on estimating a single overall average treatment effect, leaving a gap in rigorous approaches for studying conditional average treatment effects. Calonico et al. (2025) addressed this gap by developing a unified, theoretically grounded methodology for heterogeneity analysis in RD designs, allowing researchers to systematically examine how treatment effects vary across subpopulations based on pretreatment characteristics. Cattaneo et al. (2019, 2023a) provide a practical introduction to RD designs, estimation, and inference. For a review of recent literature on RD, see Cattaneo and Titiunik (2022).

This article discusses the general-purpose software package `rdhte`, which implements the main methodological results in Calonico et al. (2025). The package is composed by three commands:

- `rdhte`. Given a choice of bandwidth for localization near the cutoff determining treatment assignment, this command implements local polynomial regression estimation and inference methods allowing for interactions with pretreatment covariates. More precisely, estimation is implemented using a local weighted least squares regression incorporating semi-linear interactions with the covariates used for heterogeneity analysis, while inference is conducted using robust bias-correction (Calonico et al. 2014, 2018, 2022). The package also allows for inclusion of other pretreatment covariates for efficiency gains purposes (Calonico et al. 2019). A noteworthy feature of this command is that its implementation relies on base commands (`regress` in `Stata`) for generic linear-in-parameters least squares estimation and inference, making `rdhte` more numerically efficient and robust. This feature implies that `rdhte` may not perfectly match `rdrobust` (see Section 5), and may not yield identical numerical results across different statistical software platforms.

- `rdbwhte`. This command implements bandwidth selection tailored towards inclusion of covariates for heterogeneity analysis and efficiency improvements. See Calonico et al. (2020) for a review on the state-of-the art on bandwidth selection methods for regression discontinuity designs. Whenever a bandwidth choice is not manually supplied, the command `rdhte` internally relies on `rdbwhte` to implement data-driven bandwidth selection as a first step.

- `rdhte_lincom`. This post-estimation command implements both estimation and robust bias corrected inference for linear combinations of the treatments effects estimated by `rdhte`, inspired by the built-in command `lincom`, which cannot itself be used directly. The `Stata` post-estimation command `test` can be used directly for testing multiple linear hypotheses, but does not produce point estimates (e.g., for contrasts).

Section 2 introduces the sharp RD setup, highlighting the role of pretreatment covariates, both heterogeneity analysis (which is our main focus) and for efficiency improvements. We detail how heterogeneous treatment effects are recovered using local polynomial regression methods with semi-linear interactions. The methods are adapted from Calonico et al. (2025), and we refer to that paper for the relevant econometric theory. Cattaneo et al. (2023b) offer a broader discussion on the role of covariate-adjustments in RD designs.

Within the heterogeneity framework of Section 2, there are two distinct cases depending on the type of covariates used: (i) dummy variables (or, more generally, factor variables) that identify orthogonal (mutually exclusive) subgroups; or (ii) generic covariates (discrete or continuous). The first case, treated in detail in Section 3, arises when the covariates, and any included covariate interactions, correspond to non-overlapping subsets of the data, that is, when only binary orthogonal variables are used. This case brings the familiar subgroup analysis to RD designs, which naturally arises via indicator variables for different categories and from collapsing continuous variables into distinct bins. In this case, the treatment effect is different for each subgroup in an unrestricted way.

In the second case (Section 4), the covariates can be arbitrary, covering discrete, continuous, or mixed, and allowing for general interactions or transformations such as polynomials or other basis expansions thereof. For this generic setting, Calonico et al. (2025) provide formal identification, estimation, and robust bias-corrected inference methods for heterogeneity analysis based on standard local polynomial regression methods with semi-linear interactions. The treatment effect heterogeneity is defined as a varying coefficient linear function of the pretreatment covariates. As an alternative, see Reguly (2021) and Alcantara et al. (2025) for heterogeneous treatment effect estimation leveraging machine learning methods, and Hsu and Shen (2019, 2021) for specification testing methods.

The package `rdhte` complements the popular package `rdrobust` (Calonico et al. 2017), which focuses on estimation and robust bias-correction inference for the local average treatment effect at the cutoff in RD designs. Section 5 compares the two RD packages, and explains precisely their differences. In certain specific cases, the package `rdhte` matches the analysis based on `rdrobust`.

Throughout Sections 3, 4, and 5, we illustrate the features of `rdhte` using the data from Granzier et al. (2023), who studied coordination behavior in French two-round elections. We add to their results by studying treatment effect heterogeneity. The running variable is the vote margin in the first round election, and the outcome of interest is a binary indicator for running in the second round or not. We explore heterogeneity by party ideology and party strength, as defined in more detail below.

Section 6 concludes. Finally, the latest version of the package `rdhte`, replication files, and other related materials, are available at:

https://rdpackages.github.io/rdhte/.

2 Setup

Extending RD analysis to heterogeneous treatment effects poses a nonparametric challenge, often requiring researchers to rely on semiparametric, yet parsimonious models. Following practice, Calonico et al. (2025) examines the common approach of using local least squares regression with linear interactions, clarifying the conditions under which this method yields meaningful causal interpretations. The authors established that when potential outcomes follow a local linear-in-parameters, functional coefficient model, heterogeneous effects are identifiable and interpretable, particularly for binary orthogonal covariates (subgroup effects). Calonico et al. (2025) further developed formal econometric methods for estimation and (robust bias-corrected) inference, including optimal bandwidth selection and standard error estimators robust to both heteroskedasticity and clustering. These results aim to add rigor and consistency to empirical practice, improving the applicability, and replicability, of RD heterogeneity analysis.

The observed random sample is $(Y_i, T_i, X_i, \mathbf{W}_i', \mathbf{Z}_i')'$, for $i = 1, \ldots, n$, where:

- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ is the observed outcome, and $Y_i(0)$ and $Y_i(1)$ are the underlying potential outcomes under control and treatment status, respectively;
- $T_i = \mathbb{1}(X_i \geq c)$ is the treatment assignment indicator with X_i the continuous score and c a known cutoff;
- $\mathbf{W}_i$ is a d-dimensional vector of pretreatment covariates used for heterogeneity analysis; and
- $\mathbf{Z}_i$ is a d_z-dimensional vector of pretreatment covariates used for efficiency improvements.

Without loss of generality, we set the cutoff $c = 0$ to streamline the presentation. The canonical RD design corresponds to the case where neither $\mathbf{W}_i$ nor $\mathbf{Z}_i$ are included in the analysis.

2.1 Average Treatment Effect & Efficiency Covariates

The canonical sharp RD average treatment effect (at the cutoff $X_i = c = 0$) is

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = 0]. \tag{1}$$

It is common practice to employ the following least squares local polynomial RD estimator:

$$\dot{\tau} = \mathbf{e}_0' \dot{\boldsymbol{\beta}}, \tag{2}$$

where $\mathbf{e}_\ell$ denotes the conformable unit vector with a 1 in its $(\ell + 1)$th element, and

$$\begin{bmatrix} \dot{\boldsymbol{\alpha}} \\ \dot{\boldsymbol{\beta}} \end{bmatrix} = \underset{\boldsymbol{\alpha}, \boldsymbol{\beta}}{\arg\min} \sum_{i=1}^{n} \Big(Y_i - \mathbf{r}_p(X_i)' \boldsymbol{\alpha} - T_i \mathbf{r}_p(X_i)' \boldsymbol{\beta} \Big)^2 K_h(X_i), \tag{3}$$

with $\mathbf{r}_p(u) = (1, u, \ldots, u^p)'$ denoting the pth order polynomial expansion, and $K_h(u) = K(u/h)/h$ for a kernel (i.e., weighting) function $K(\cdot)$ and bandwidth h.

For the local-linear case ($p = 1$), the underlying implementation employing the `Stata` command `reg` is

```
reg y t##c.x
```
(4)

properly localized to $c = 0$ and weighted using the kernel $K(\cdot)$. Then, $\dot{\tau}$ corresponds to the coefficient estimate associated with `t`.

The classical RD treatment effect estimator $\dot{\tau}$ is consistent for τ under standard regularity conditions. Similarly, $\mathbf{e}_1'\dot{\boldsymbol{\beta}}$ is an estimator useful in the Kink RD design. Calonico et al. (2014) studied MSE-optimal point estimation and robust bias-corrected inference for this case, while Calonico et al. (2020) study optimal bandwidth selection for both point estimation and inference.

The covariates $\mathbf{Z}_i$ can be included in the regression estimation to improve efficiency in the estimation of RD treatment effect τ. Calonico et al. (2019) studied MSE-optimal point estimation and robust bias-corrected inference for this case, and recommended the estimator

$$\widetilde{\tau} = \mathbf{e}_0'\widetilde{\boldsymbol{\beta}},$$

where

$$\begin{bmatrix} \widetilde{\boldsymbol{\alpha}} \\ \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\gamma}} \end{bmatrix} = \underset{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}}{\arg\min} \sum_{i=1}^{n} \Big(Y_i - \mathbf{r}_p(X_i)'\boldsymbol{\alpha} - T_i\mathbf{r}_p(X_i)'\boldsymbol{\beta} - \mathbf{Z}_i'\boldsymbol{\gamma} \Big)^2 K_h(X_i). \tag{5}$$

The covariate-adjusted, possibly more efficient RD estimator $\widetilde{\tau}$ is consistent for τ under regularity conditions. The estimated coefficients $\widetilde{\boldsymbol{\gamma}}$ do not have a causal interpretation; they are fitted to improve the precision of $\widetilde{\boldsymbol{\beta}}$. Standard Kink RD designs with covariate-adjustment for efficiency gains consider the estimator $\mathbf{e}_1'\widetilde{\boldsymbol{\beta}}$.

For the local-linear case ($p = 1$), the underlying implementation employing the `Stata` command `reg` is

```
reg y t##c.x z
```
(6)

again properly localized to $c = 0$ and weighted using the kernel $K(\cdot)$. Then, $\widetilde{\tau}$ corresponds to the coefficient associated with `t`.

The package `rdrobust` (Calonico et al. 2017) provides software implementation for estimation and inference on the RD average treatment effects, both with and without adding covariates for efficiency gains.

2.2 Heterogeneous Average Treatment Effects

To complement prior literature focusing on the RD average treatment effect τ, Calonico et al. (2025) considers the (local to $X_i = c = 0$) RD conditional average treatment effect

(CATE) function

$$\kappa(\mathbf{w}) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = 0, \mathbf{W}_i = \mathbf{w}],$$

which employs the covariates $\mathbf{W}_i$ for RD treatment effect heterogeneity. The covariates $\mathbf{W}_i$ are distinct from those used for efficiency improvements, $\mathbf{Z}_i$, but importantly both are predetermined.

When $\mathbf{W}_i$ is continuous and/or high-dimensional, the RD CATE function $\kappa(\mathbf{w})$ can be difficult to estimate nonparametrically without further restrictions. Thus, it is common practice to employ the semilinear least squares local polynomial estimation procedure:

$$\widehat{\kappa}(\mathbf{w}) = \mathbf{e}_0'\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\xi}}' \begin{bmatrix} \mathbf{I}_d \\ \mathbf{0}_{sd\times d} \end{bmatrix} \mathbf{w}, \tag{7}$$

where $\mathbf{I}_d$ denotes the $(d\times d)$ identity matrix, $\mathbf{0}_{sd\times d}$ denotes the $(sd\times d)$ matrix of zeros, $\mathbf{w}$ takes values on the support of $\mathbf{W}_i$, and

$$\begin{aligned} \begin{bmatrix} \widehat{\boldsymbol{\alpha}} \\ \widehat{\boldsymbol{\lambda}} \\ \widehat{\boldsymbol{\xi}} \end{bmatrix} = \underset{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\xi}}{\arg\min} \sum_{i=1}^{n} \Big(& Y_i - \mathbf{r}_p(X_i)'\boldsymbol{\alpha} - T_i\mathbf{r}_p(X_i)'\boldsymbol{\beta} \\ & - [\mathbf{r}_s(X_i)\otimes\mathbf{W}_i]'\boldsymbol{\lambda} - T_i[\mathbf{r}_s(X_i)\otimes\mathbf{W}_i]'\boldsymbol{\xi}\Big)^2 K_h(X_i), \end{aligned} \tag{8}$$

with $\otimes$ denoting the Kronecker product. Calonico et al. (2025) study this estimation procedure and give easy-to-interpret sufficient conditions to ensure that $\widehat{\kappa}(\mathbf{w})$ is consistent for $\kappa(\mathbf{w})$. The most important of these is the assumption that the CATE function can be written as a (local) functional coefficient linear form: $\kappa(\mathbf{w}) = \theta(x) + \boldsymbol{\xi}(x)'\mathbf{w}$, see below. This is without loss of generality for binary orthogonal covariates (Section 3). Calonico et al. (2025) also establish MSE-optimal bandwidth selection and point estimation, and valid robust bias-corrected inference for uncertainty quantification.

The generic form of $\widehat{\kappa}(\mathbf{w})$, and its underlying least squares fit (8), is notationally cumbersome but easy to understand. It arises from adding to the local polynomial estimation the interaction between the covariates $\mathbf{W}_i$ and the polynomial approximation $\mathbf{r}_s(X_i)$ (which may be a different polynomial than the main effect). For example, if $p = s = 1$, so that $\mathbf{r}_s(X_i) = (1, X_i)'$, and $\mathbf{W}_i$ is a continuous variable, the `Stata` implementation would be:

```
reg y t##c.x##c.w                                                    (9)
```

properly localized to $c = 0$ and weighted using the kernel $K(\cdot)$. The estimate $\widehat{\kappa}(w_0)$ of Equation (7), for a specific value w_0 is obtained as the coefficient on `t` plus the product of w_0 and the coefficient on the interaction of `t` and `c.w`.

Covariate-adjustment based on $\mathbf{Z}_i$ for efficiency gains is also allowed: following Calonico et al. (2019), in the regression (8), $\mathbf{Z}_i$ and $\mathbf{Z}_i\otimes\mathbf{W}_i$ are included but without interaction with treatment assignment variable T_i. The resulting coefficients (on $\mathbf{Z}_i$ and

$\mathbf{Z}_i \otimes \mathbf{W}_i$) do not have a causal interpretation, but the resulting estimator $\widehat{\kappa}(\mathbf{w})$ can exhibit efficiency gains. We omit further details for brevity, but the local-linear command is:

```
reg y t##c.x##c.w z##c.w
```
(10)

properly localized to $c = 0$ and weighted using the kernel $K(\cdot)$. The CATE estimate is obtained exactly as before.

The causal interpretation of the probability limit of the estimator $\widehat{\kappa}(\mathbf{w})$ must be carefully considered and depends on the type of covariates as follows.

- Binary Orthogonal $\mathbf{W}_i$. This case corresponds to situations where each component is binary, $\mathbf{W}_i \in \{0,1\}^d$, and at most one component takes the value one, $\mathbf{W}_i'\mathbf{W}_i \leq 1$. By implication, the data is partitioned in disjoint groups as determined by $\mathbf{W}_i$, and thus this case corresponds to subset analysis. The vector $\mathbf{W}_i$ consists of indicator variables (for categories, levels, bins, etc) and their interactions. In this case, the estimator $\widehat{\kappa}(\mathbf{w})$ is fully nonparametric and consistent for generic $\kappa(\mathbf{w})$ under minimal regularity conditions, for all $\mathbf{w} \in \{0,1\}^d$. In this special setting, the assumption $\kappa(\mathbf{w}) = \theta(x) + \boldsymbol{\xi}(x)'\mathbf{w}$ is without loss of generality, i.e., $\kappa(\mathbf{w})$ is unrestricted.
- Generic $\mathbf{W}_i$. This case corresponds to situations where some components of $\mathbf{W}_i$ are continuous, take several discrete values (which are *not* "dummied out"), or are mixed. In this case, the functional coefficient linear model is used by Calonico et al. (2025) to restore a causal interpretation to the local least squares estimation procedure (8). Absent this assumption, $\widehat{\kappa}(\mathbf{w})$ is not consistent for a causal object, but rather its probability limit has the usual local best linear-in-parameters approximation interpretation. The identifying assumption is flexible in that it allows for nonlinear heterogeneity by including in $\mathbf{W}_i$ transformations of the original covariates, such as polynomial expansions or other basis transformations.

Regardless of the specific structure of $\mathbf{W}_i$, Calonico et al. (2025) develop MSE-optimal estimation and robust bias-correction inference for $\kappa(\mathbf{w})$, expanding prior results established for τ. See Arai and Ichimura (2018) and Calonico et al. (2020) for more discussion on bandwidth selection for classical RD designs, Calonico et al. (2014) and Calonico et al. (2018, 2022) for foundational theoretical analysis of robust bias-correction inference. Hyytinen et al. (2018) and De Magalhães et al. (2025) offer comprehensive empirical validation of those estimation and inference methods.

3 Binary Orthogonal Covariates

We begin the illustration of heterogeneity analysis with `rdhte` with the first case: binary orthogonal covariates (subgroup analysis). Recall the context of Granzier et al. (2023), where the outcome is $Y_i \in \{0,1\}$ indicates running in the second round election, and the running variable X_i is the vote margin in the first round. We will show the features of `rdhte` using different measures of party ideology and strength. Unless noted otherwise, all results will use standard errors clustered by district (stored as `cluster_var`) following their original analysis. Because `rdhte` is based on `regress`, clustered standard errors are obtained using the standard `vce` options: `vce(cluster cluster_var)` yields HC1 clustered standard errors, while `vce(hc2 cluster_var)` yields HC2, which is preferred. Currently, HC3 is not available, but it would provide a more robust option. We therefore recommend setting `vce(hc2 cluster_var)`, which we use throughout. In other software, HC3 is available and thus set as the default for both clustered and independent data.

3.1 Single Binary Variable

The most basic form for such analysis is the case of a single binary variable, so that $d = 1$ and $W_i \in \{0, 1\}$. To illustrate, we create the dummy variable `left`, taking the value one if the candidate belongs to the left or far left and zero otherwise. Running `rdhte` yields a point estimate and robust bias corrected inference for each category.

```
. rdhte y x, covs_hte(w_left) vce(hc2 cluster_var)

Sharp RD Heterogeneous Treatment Effects: Subgroups.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =      mserd
     Number of obs |     19723       19811           Kernel        = Triangular
    Order est. (p) |         1           1           VCE method    =        HC2
    Order bias (q) |         2           2


Outcome: y. Running variable: x.
-----------------------------------------------------------------------------------------------------------------
             |     Point      | Robust Inference
   i.w_left  |     Estimate   | z-stat      P>|z|      [95% Conf. Interval]      Nh-      Nh+       h-       h+
-------------+---------------------------------------------------------------------------------------------------
   0.w_left  |     0.021      | 1.522       0.128        -0.003     0.027       4352     4422      0.077    0.077
   1.w_left  |     0.089      | 6.817       0.000        0.062      0.112       5036     4840      0.117    0.117
-----------------------------------------------------------------------------------------------------------------
   (Std. err. adjusted for 4892 clusters in cluster_var)
```

We see that the RD treatment effect is larger for left-of-center candidates (0.089 versus 0.021) and is statistically significantly different from zero, which is not the case for the others. In this case, `rdhte` automatically detects that `w_left` is an indicator for subgroups because `w_left` takes only the values zero and one. Any other coding (e.g., if it takes values 1 and 2) will require use of the `i.`*varname* syntax. This output is omitted to save space but is available in the replication code.

The package, by default, automatically computes the optimal bandwidth for each group separately. The behavior can be overridden using the option `bwjoint`. However, this is generally not advisable because the bias-variance trade-off that determines the optimal bandwidth need not be the same across subsets of the data, just as it need not be the same across different datasets. See the replication code for an example and the help file for `rdbwhte` for more details on bandwidth selection.

A natural question following these results is whether the two subgroups are statistically significantly different. This is addressed using the featured post-estimation command `rdhte_lincom`, which computes the point estimate of the difference (in this case $\widehat{\kappa}(1) - \widehat{\kappa}(0)$) and the robust bias-corrected confidence interval and p-value.

```
. rdhte_lincom 1.w_left - 0.w_left

RD Heterogeneous Treatment Effects: Linear combinations of parameters

( 1)  1.w_left - 0.w_left = 0

--------------------------------------------------------------------------------
                Y1 |    Coefficient   [95% Conf. Interval]    P>|z|
-------------------+------------------------------------------------------------
               (1) |     .06807       0.046       0.105       0.000
--------------------------------------------------------------------------------
```

We see that, indeed, the effect for left-of-center candidates is significantly higher at the 5% level. In the next subsection, we also demonstrate post-estimation hypothesis testing using the Stata built-in command `test`.

3.2 "Dummied Out" Factor Variable – Unordered

Moving beyond the case of a single binary variable, we now explore other instances where the set of covariates $\mathbf{W}_i$ are binary orthogonal variables and therefore identify subgroups of the data.

We obtain a more nuanced view of the heterogeneity by party ideology using the four-level factor variable `w_ideology`. In this case, the four-level factor is "dummied out" to obtain an indicator for each group. This happens automatically using the `i.` syntax.

```
. rdhte y x, covs_hte(i.w_ideology) vce(hc2 cluster_var) labels

Sharp RD Heterogeneous Treatment Effects: Subgroups.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =      mserd
     Number of obs |      19723       19811           Kernel        = Triangular
    Order est. (p) |          1           1           VCE method    =        HC2
    Order bias (q) |          2           2


Outcome: y. Running variable: x.
------------------------------------------------------------------------------------------------------------------------
                 |    Point      | Robust Inference
    i.w_ideology |    Estimate   | z-stat     P>|z|      [95% Conf. Interval]      Nh-      Nh+        h-        h+
-----------------+------------------------------------------------------------------------------------------------------
            left |    0.089      | 6.817      0.000        0.062      0.112        5036     4840       0.117     0.117
          center |    -0.024     | -0.831     0.406       -0.198      0.080         181      156       0.056     0.056
           right |    0.026      | 1.897      0.058       -0.001      0.032        3816     4202       0.086     0.086
        farright |    0.007      | 0.881      0.378       -0.012      0.032         713      384       0.090     0.090
------------------------------------------------------------------------------------------------------------------------
   (Std. err. adjusted for 4949 clusters in cluster_var)
```

For post-estimation, we use the built-in `Stata` command `test` to find that all the non-left categories are statistically indistinguishable from each other and from zero.

```
. test 4.w_ideology = 3.w_ideology = 2.w_ideology = 0

 ( 1)  - 3.w_ideology + 4.w_ideology = 0
 ( 2)  - 2.w_ideology + 4.w_ideology = 0
```

```
 ( 3)  4.w_ideology = 0

           chi2(  3) =    5.07
         Prob > chi2 =    0.1667
```

3.3 “Dummied Out” Factor Variable – Ordered

We bring in another pre-treatment variable for heterogeneity analysis: the strength at the national level, defined as the average of first-round vote shares of all candidates of the same orientation at the national level. The raw measure is stored as `w_strength` and is considered in the next subsection. Here, we categorize strength into the four quartiles and obtain the effect for each. Compared to the case of `w_ideology`, this variable has a clear ordering. Note here that the `i.` syntax is required; otherwise `rdhte` will yield a linear fit treating the variable as continuous (see below).

```
. rdhte y x, covs_hte(i.w_strength_qrt) vce(hc2 cluster_var)

Sharp RD Heterogeneous Treatment Effects: Subgroups.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =      mserd
     Number of obs |      19723       19811          Kernel        = Triangular
    Order est. (p) |          1           1          VCE method    =        HC2
    Order bias (q) |          2           2


Outcome: y. Running variable: x.
------------------------------------------------------------------------------------------------------------------------
                     |    Point      | Robust Inference
   i.w_strength_qrt  |    Estimate   | z-stat     P>|z|     [95% Conf. Interval]     Nh-     Nh+       h-       h+
---------------------+--------------------------------------------------------------------------------------------------
   0.w_strength_qrt  |    0.016      | 1.801      0.072       -0.001     0.032      2583    2283      0.094    0.094
   1.w_strength_qrt  |    0.039      | 2.218      0.027        0.003     0.055      2227    2300      0.085    0.085
   2.w_strength_qrt  |    0.054      | 3.486      0.000        0.025     0.090      2040    2329      0.105    0.105
   3.w_strength_qrt  |    0.094      | 6.649      0.000        0.069     0.127      3436    3486      0.143    0.143
------------------------------------------------------------------------------------------------------------------------
   (Std. err. adjusted for 5085 clusters in cluster_var)
```

We find that the causal effect is increasing with strength, an interesting if not surprising finding. This analysis depends on binning, and can be viewed intuitively (though not formally) as a four-piece approximation to an unknown, nonparametric $\kappa(w)$, following Cattaneo et al. (2024).

3.4 Binary Interactions

Finally, dummying out factor variables, and then interacting them, also yields binary orthogonal variables, one for each unique combination, provided the model is saturated. To illustrate, here we interact the dummy for left-of-center with an indicator for a “strong” candidate, defined as having above-median national strength.

```
. rdhte y x, covs_hte(i.w_left#i.w_strong) vce(hc2 cluster_var)

Sharp RD Heterogeneous Treatment Effects: Subgroups.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =      mserd
     Number of obs |      19723       19811          Kernel        = Triangular
    Order est. (p) |          1           1          VCE method    =        HC2
```

```
    Order bias (q) |         2           2


Outcome: y. Running variable: x.
-----------------------------------------------------------------------------------------------------------------------
                       |    Point      | Robust Inference
   i.w_left#i.w_strong  |    Estimate   | z-stat      P>|z|      [95% Conf. Interval]      Nh-      Nh+        h-        h+
-----------------------+-----------------------------------------------------------------------------------------------
   0.w_left#1.w_strong  |    0.003      | 0.004       0.997       -0.016     0.016       2348     2257       0.070     0.070
   0.w_left#2.w_strong  |    0.044      | 3.011       0.003        0.014     0.065       2316     2643       0.104     0.104
   1.w_left#1.w_strong  |    0.061      | 3.834       0.000        0.029     0.090       2066     1960       0.097     0.097
   1.w_left#2.w_strong  |    0.114      | 6.457       0.000        0.082     0.153       3107     3055       0.143     0.143
-----------------------------------------------------------------------------------------------------------------------
   (Std. err. adjusted for 4988 clusters in cluster_var)
```

As is typical for interaction effects, this adds nuance to the above findings, as we see that not only do left-of-center candidates have larger treatment effects, but this is even more pronounced when the candidates are stronger than the national median.

4 Generic Covariates

We now discuss the case of generic covariates: $\mathbf{W}_i$ is not a set of binary orthogonal variables, and so instead of obtaining heterogeneity by subgroups, we obtain a linear-in-parameters estimate of the function $\kappa(\mathbf{w})$. As discussed above, Calonico et al. (2025) use a functional coefficient model assumption to obtain a causal interpretation of the probability limit of $\widehat{\kappa}(\mathbf{w})$. In this case, instead of selecting an optimal bandwidth for each subset, a single bandwidth is selected.

4.1 Continuous Variables and Replication with Linear Regression

To illustrate we use the measure of strength, `w_strength`. Before, when grouped by quantiles, we saw that the treatment effect was increasing in strength. It may be that a linear fit is an appropriate and parsimonious way of capturing this relationship. Using `w_strength`, we obtain the following. For later comparison with results from `regress`, here we use the uniform kernel to obtain an unweighted local least squares fit. Note that the output of `rdhte` is different relative to the subset analysis, reflecting that there is only one bandwidth and the coefficients here represent qualitatively different objects.

```
. rdhte y x, covs_hte(w_strength) kernel(uni) vce(hc2 cluster_var)

Sharp RD Heterogeneous Treatment Effects: Continuous.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =      mserd
     Number of obs |      19723       19811          Kernel        =    Uniform
Eff. Number of obs |       9514        9529          VCE method    =        HC2
    Order est. (p) |          1           1
    Order bias (q) |          2           2
       BW est. (h) |      0.096       0.096


Outcome: y. Running variable: x.
-----------------------------------------------------------------------------------------
                           |     Point      | Robust Inference
                w_strength  |    Estimate    | z-stat     P>|z|      [95% Conf. Interval]
```

```
----------------------------+-----------------------------------------------------------
                          T  |     -0.055     | -2.504     0.012        -0.129    -0.016
             T#c.w_strength  |     0.262      | 3.922      0.000        0.150     0.451
----------------------------------------------------------------------------------------
   (Std. err. adjusted for 4777 clusters in cluster_var)
```

In Section 3, each row of the output table showed the treatment effect estimate (and robust bias-corrected inference) for each subset of the data. Here, instead, we have the "intercept" and "slope" terms of the estimate $\widehat{\kappa}(\mathbf{w})$ of Equation (7). That is, for a specific level of candidate strength w_0, we would obtain $\widehat{\kappa}(w_0) = -0.055 + 0.262 \times w_0$. The value -0.055 represents $\widehat{\kappa}(0)$, which may or may not have conceptual meaning, depending on the context and the definition of $\mathbf{W}_i$.

To build intuition, this interpretation is identical to interpreting coefficients in a linear regression, because here (by default) we are using `rdhte` with a polynomial order of $p = s = 1$. Recall that identical estimates can be obtained using `regress`, properly localized and weighted, as shown in Equation (9). Here, we have used the uniform kernel, so we must only localize using the same bandwidth in order to obtain the same point estimates. We extract the bandwidth from the ereturns of `rdhte`, generate the treatment indicator, and then run the regression.

```
. local bw = e(h)[1,1]

. gen T = (x>0)

. reg y T##c.x##c.w_strength if abs(x)<=`bw´

      Source |       SS           df       MS      Number of obs   =    19,043
-------------+----------------------------------   F(7, 19035)     =    116.12
       Model |  23.5289217         7  3.36127453   Prob > F        =    0.0000
    Residual |  551.005028    19,035  .028946941   R-squared       =    0.0410
-------------+----------------------------------   Adj R-squared   =    0.0406
       Total |  574.533949    19,042  .030171933   Root MSE        =    .17014

------------------------------------------------------------------------------------
                 y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------------+----------------------------------------------------------------
               1.T |  -.0549442   .0246365    -2.23   0.026    -.1032339   -.0066545
                 x |  -.1455088   .3001954    -0.48   0.628    -.7339183    .4429008
                   |
             T#c.x |
                 1 |   .1504868      .4825     0.31   0.755    -.7952559    1.096229
                   |
        w_strength |  -.2659247   .0391581    -6.79   0.000     -.342678   -.1891715
                   |
    T#c.w_strength |
                 1 |   .2617644   .0588325     4.45   0.000     .1464474    .3770813
                   |
  c.x#c.w_strength |   .7627849   .7198203     1.06   0.289    -.6481266    2.173696
                   |
T#c.x#c.w_strength |
                 1 |  -.7652365   1.144337    -0.67   0.504    -3.008238    1.477765
                   |
             _cons |   1.054821   .0163093    64.68   0.000     1.022853    1.086789
------------------------------------------------------------------------------------
```

Here, the relevant coefficients are those on `1.T` and `T#c.w_strength`.

It is important to remember that this only replicates the point estimates, for *in-*

ference, robust bias correction is required. The inference measures (standard errors, t-statistics, confidence intervals, and p-values) in this `regress` output are not valid.

4.2 Interactions

As a final illustration, consider the analog of the binary interaction at the close of Section 3. There, we obtained an estimate for each of the four categories reflecting left or not and above/below median strength. Here, we consider the interaction of the binary indicator `w_left` with the continuous measure of strength. By fully saturating the model, we obtain a separate intercept and slope for left-of-center and for center-and-right candidates.

```
. rdhte y x, covs_hte(i.w_left##c.w_strength) vce(hc2 cluster_var)

Sharp RD Heterogeneous Treatment Effects: Continuous.

      Cutoff c = 0 | Left of c  Right of c            Number of obs =      39534
-------------------+----------------------            BW type       =      mserd
     Number of obs |      19723       19811           Kernel        = Triangular
Eff. Number of obs |       9533        9547           VCE method    =        HC2
    Order est. (p) |          1           1
    Order bias (q) |          2           2
       BW est. (h) |      0.097       0.097


Outcome: y. Running variable: x.
---------------------------------------------------------------------------------------------------
                                     |    Point      | Robust Inference
             i.w_left##c.w_strength  |    Estimate   | z-stat     P>|z|     [95% Conf. Interval]
-------------------------------------+-------------------------------------------------------------
                                  T  |    -0.033     | -1.360     0.174       -0.095    0.017
                        T#1.w_left   |    -0.059     | -0.272     0.786       -0.309    0.233
                    T#c.w_strength   |    0.141      | 1.775      0.076       -0.015    0.295
            T#1.w_left#c.w_strength  |    0.274      | 0.735      0.462       -0.401    0.882
---------------------------------------------------------------------------------------------------
   (Std. err. adjusted for 4782 clusters in cluster_var)
```

This output reveals the same qualitative conclusion as the binary interaction, but expressed in a different way. It is worth noting that, because this model is fully saturated, the same results can be obtained studying `w_strength` separately for each category of `w_left`. This is shown in the replication code, but omitted here to save space.

5 Comparison with `rdrobust`

In this section, we briefly illustrate how `rdhte` does, and does not, match the output from the popular command `rdrobust` (Calonico et al. 2017). For simplicity, all analyses in this section do not use clustered standard errors, and so for `rdhte` the HC3 option is used. We show replication of average and of heterogeneous treatment effects. Both packages can use covariates for efficiency (using the `covs_eff` option in `rhdte`), but this is not a point of comparison here.

5.1 Implementation Differences

The discrepancies arise for three main reasons, all of which are driven by the fact that `rdhte` (in all platforms) relies on built-in base commands for least squares regression.

1. `rdrobust` allows for, and by default uses, a different bandwidth for bias correction, b, than for the main estimation, h, with $\rho = h/b$. In contrast, `rdhte` forces $h = b$, i.e. $\rho = 1$. This choice has known optimality properties and is reliable in practice (Calonico et al. 2020, 2022).

2. `rdrobust` uses a custom nearest-neighbor variance estimator by default, and also allows for HC0–HC3, as well as for HC and NN cluster-robust variance estimators. In contrast, `rdhte` uses HC3 by default, and also allows for other variance estimators (heteroskedasticity-robust and cluster-robust) available via the corresponding base least squares command (when applicable).

3. `rdrobust` takes a two-sample approach for estimation and inference, while `rdhte` takes a one-sample approach (i.e., single linear regression fit with interactions). This discrepancy in implementation approaches leads to different degrees of freedom adjustments, which in some cases can lead to slightly different standard error estimators (e.g., when implementing HC1). In addition, this discrepancy can lead to different matrix inversions and, by implication, potentially different covariates may be dropped due to multi-collinearity.

5.2 Average Treatment Effect

First, consider the RD average treatment effect τ of Equation (1).

Both `rdrobust` and `rdhte` can be used to obtain the corresponding estimator $\dot{\tau}$ of (2) along with robust bias-corrected inference. For `rdhte`, the average is estimated if no heterogeneity variables are specified. However, the default implementations differ, and so the output will not match, as shown here.

```
. rdhte y x

Sharp RD Average Treatment Effect.

      Cutoff c = 0 | Left of c  Right of c            Number of obs =      39534
-------------------+----------------------            BW type       =      mserd
     Number of obs |      19723       19811            Kernel        = Triangular
Eff. Number of obs |       9536        9550            VCE method    =        HC3
    Order est. (p) |          1           1
    Order bias (q) |          2           2
       BW est. (h) |      0.097       0.097


Outcome: y. Running variable: x.
----------------------------------------------------------------------------
               |    Point      | Robust Inference
               |    Estimate   | z-stat     P>|z|     [95% Conf. Interval]
---------------+------------------------------------------------------------
             T |    0.051      | 7.071      0.000       0.035     0.062
----------------------------------------------------------------------------
```

```
. rdrobust y x

Sharp RD estimates using local polynomial regression.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =      mserd
     Number of obs |      19723       19811           Kernel        = Triangular
Eff. Number of obs |       9545        9559           VCE method    =         NN
    Order est. (p) |          1           1
    Order bias (q) |          2           2
       BW est. (h) |      0.097       0.097
       BW bias (b) |      0.159       0.159
         rho (h/b) |      0.608       0.608

Outcome: y. Running variable: x.
--------------------------------------------------------------------------------
                   | Point          |                Robust Inference
                   | Estimate       |    z-stat      P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
         RD Effect |  .05132        |    8.8951    5.837e-19   .039464     .061771
--------------------------------------------------------------------------------
```

If all these settings are synchronized, the two commands report the same results (up to numerical differences in computation). We illustrate this point by setting $h = 0.1$ (which automatically forces $h = b$ or $\rho = 1$ in `rdrobust`), and variance estimation to HC3.

```
. rdhte y x, h(0.1) vce(hc3)

Sharp RD Average Treatment Effect.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =     Manual
     Number of obs |      19723       19811           Kernel        = Triangular
Eff. Number of obs |       9829        9843           VCE method    =        HC3
    Order est. (p) |          1           1
    Order bias (q) |          2           2
       BW est. (h) |      0.100       0.100


Outcome: y. Running variable: x.
----------------------------------------------------------------------------
               |    Point      | Robust Inference
               |    Estimate   | z-stat     P>|z|     [95% Conf. Interval]
---------------+------------------------------------------------------------
            T  |    0.051      | 7.194      0.000       0.035     0.062
----------------------------------------------------------------------------


. rdrobust y x, h(0.1) rho(1) vce(hc3)

Sharp RD estimates using local polynomial regression.

      Cutoff c = 0 | Left of c  Right of c           Number of obs =      39534
-------------------+----------------------           BW type       =     Manual
     Number of obs |      19723       19811           Kernel        = Triangular
Eff. Number of obs |       9829        9843           VCE method    =        HC3
    Order est. (p) |          1           1
    Order bias (q) |          2           2
       BW est. (h) |      0.100       0.100
       BW bias (b) |      0.100       0.100
         rho (h/b) |      1.000       1.000
```

```
Outcome: y. Running variable: x.
-------------------------------------------------------------------------------
                   | Point          |              Robust Inference
                   | Estimate       |    z-stat      P>|z|     [95% Conf. Interval]
-------------------+-----------------------------------------------------------
         RD Effect |  .05141        |    7.1944    6.275e-13   .035387     .061888
-------------------------------------------------------------------------------
```

5.3 Subgroup Analysis

The same equivalence can be obtained when conducting heterogeneity analysis by subgroup, because this amounts to estimating the average effect for a subset of the data. Here we illustrate that, again enforcing common settings, `rdrobust` obtains the same results as `rdhte` for left-of-center candidates (the other group is omitted to save space).

```
. rdhte y x, covs_hte(w_left) h(0.078 0.116)

Sharp RD Heterogeneous Treatment Effects: Subgroups.

      Cutoff c = 0 | Left of c  Right of c            Number of obs =      39534
-------------------+----------------------            BW type       =     Manual
     Number of obs |     19723       19811            Kernel        = Triangular
    Order est. (p) |         1           1            VCE method    =        HC3
    Order bias (q) |         2           2


Outcome: y. Running variable: x.
------------------------------------------------------------------------------------------------------------------
             |    Point      | Robust Inference
   i.w_left  |    Estimate   | z-stat     P>|z|     [95% Conf. Interval]     Nh-     Nh+       h-       h+
-------------+----------------------------------------------------------------------------------------------------
   0.w_left  |    0.021      | 1.522      0.128      -0.003    0.027       4388    4453      0.078    0.078
   1.w_left  |    0.089      | 7.435      0.000       0.064    0.110       5010    4799      0.116    0.116
------------------------------------------------------------------------------------------------------------------


. rdrobust y x if w_left==1, h(0.116) rho(1) vce(hc3)

Sharp RD estimates using local polynomial regression.

      Cutoff c = 0 | Left of c  Right of c            Number of obs =      17596
-------------------+----------------------            BW type       =     Manual
     Number of obs |      9436        8160            Kernel        = Triangular
Eff. Number of obs |      5010        4799            VCE method    =        HC3
    Order est. (p) |         1           1
    Order bias (q) |         2           2
       BW est. (h) |     0.116       0.116
       BW bias (b) |     0.116       0.116
         rho (h/b) |     1.000       1.000

Outcome: y. Running variable: x.
------------------------------------------------------------------------------
                   | Point          |              Robust Inference
                   | Estimate       |    z-stat      P>|z|    [95% Conf. Interval]
-------------------+----------------------------------------------------------
         RD Effect |  .08863        |    7.4354    1.042e-13   .064238     .110226
------------------------------------------------------------------------------
```

6 Conclusion

This article illustrated the main functionalities of the package `rdhte` for heterogeneous RD treatment effects estimation and robust bias-corrected inference. We also discussed how this package complements the popular RD package `rdrobust`. The methods implemented by the package `rdhte` can also be used in the context of multi-cutoff and multi-score RD designs by discretizing the analysis along the multi-dimensional assignment rule; see Cattaneo et al. (2020) for more discussion. Furthermore, `rdhte` can also be used to implement other RD designs involving comparisons across subgroups, such as in difference-in-cutoffs designs (Grembi et al. 2016) or dynamic designs (Hsu and Shen 2024). We do not discuss these connections further to conserve space.

7 Acknowledgments

Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation through grants SES-2019432 and SES-2241575.

8 References

Alcantara, R., P. R. Hahn, C. Carvalho, and H. Lopes. 2025. Learning Conditional Average Treatment Effects in Regression Discontinuity Designs Using Bayesian Additive Regression Trees. *arXiv preprint arXiv:2503.00326* .

Arai, Y., and H. Ichimura. 2018. Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator. *Quantitative Economics* 9(1): 441–482.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2018. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association* 113(522): 767–779.

———. 2020. Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs. *Econometrics Journal* 23(2): 192–210.

———. 2022. Coverage Error Optimal Confidence Intervals for Local Polynomial Regression. *Bernoulli* 28(4): 2998–3022.

Calonico, S., M. D. Cattaneo, M. H. Farrell, F. Palomba, and R. Titiunik. 2025. Treatment Effect Heterogeneity in Regression Discontinuity Designs. *arXiv preprint arXiv:2503.13696* .

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2017. `rdrobust`: Software for Regression Discontinuity Designs. *Stata Journal* 17(2): 372–404.

———. 2019. Regression Discontinuity Designs using Covariates. *Review of Economics and Statistics* 101(3): 442–451.

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

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng. 2024. On Binscatter. *American Economic Review* 114(5): 1488–1514.

Cattaneo, M. D., N. Idrobo, and R. Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press.

———. 2023a. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Cambridge Elements: Quantitative and Computational Methods for Social Science, Cambridge University Press, arXiv:2301.08958.

Cattaneo, M. D., L. Keele, and R. Titiunik. 2023b. Covariate Adjustment in Regression Discontinuity Designs. In *Handbook of Matching and Weighting in Causal Inference*, ed. D. S. S. J. R. Zubizarreta, E. A. Stuart and P. R. Rosenbaum, chap. 8, 153–168. Boca Raton, FL: Chapman & Hall.

Cattaneo, M. D., and R. Titiunik. 2022. Regression Discontinuity Designs. *Annual Review of Economics* 14: 821–851.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare. 2020. Analysis of Regression Discontinuity Designs with Multiple Cutoffs or Multiple Scores. *Stata Journal* 20(4): 866–891.

De Magalhães, L., D. Hangartner, S. Hirvonen, J. Meriläinen, N. A. Ruiz, and J. Tukiainen. 2025. When Can We Trust Regression Discontinuity Design Estimates from Close Elections? Evidence from Experimental Benchmarks. *Political Analysis* .

Granzier, R., V. Pons, and C. Tricaud. 2023. Coordination and Bandwagon Effects: How Past Rankings Shape the Behavior of Voters and Candidates. *American Economic Journal: Applied Economics* 15(4): 177–217.

Grembi, V., T. Nannicini, and U. Troiano. 2016. Do fiscal rules matter? *American Economic Journal: Applied Economics* 8(3): 1–30.

Hsu, Y.-C., and S. Shen. 2019. Testing treatment effect heterogeneity in regression discontinuity designs. *Journal of Econometrics* 208(2): 468–486.

———. 2021. Testing Monotonicity of Conditional Treatment Effects under Regression Discontinuity Designs. *Journal of Applied Econometrics* 36(3): 346–366.

———. 2024. Dynamic regression discontinuity under treatment effect heterogeneity. *Quantitative Economics* 15(4): 1035–1064.

Hyytinen, A., J. Meriläinen, T. Saarimaa, O. Toivanen, and J. Tukiainen. 2018. When does regression discontinuity design work? Evidence from random election outcomes. *Quantitative Economics* 9(2): 1019–1051.

Reguly, Á. 2021. Heterogeneous Treatment Effects in Regression Discontinuity Designs. *arXiv preprint arXiv:2106.11640* .

9 About the Authors

Sebastian Calonico is an Assistant Professor at the University of California at Davis.

Matias D. Cattaneo is a Professor at Princeton University.

Filippo Palomba is a Ph.D. candidate at Princeton University.

Max H. Farrell is an Associate Professor at the University of California at Santa Barbara.

Rocio Titiunik is a Professor at Princeton University.