Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operations Research

Donut covariate-adjusted regression discontinuity designs

Andrea Busso (597857)

Supervisor:	Jens Klooster
Second assessor:	dr. Eoghan O'Neill
Date final version:	1st July 2024

Abstract

Regression discontinuity (RD) designs are used to evaluate the causal effects of a treatment on an outcome variable based on a cutoff value that determines treatment exposure. The 'donut' RD design excludes observations in a small area around the cutoff to avoid possible data distortion caused by systematic sorting or other data issues. The donut addition is used as a robustness measure in RD designs.

We study and evaluate the use of the donut design in both basic and covariate-adjusted RD designs, the latter of which include additional covariates in the estimation process. Furthermore, we compare the performances of the different covariate-adjusted models using the basic RD design as a benchmark.

We employ two simulation studies and an empirical application based on child mortality in the United States to evaluate the models. We compare bias, standard deviation and root mean squared error as main performance measures. Additionally we provide t-statistics-based tests and a bootstrap estimation to have a comprehensive overview of the models.

Our findings indicate that while donut designs have higher standard deviations, they exhibit relatively low biases, highlighting their effectiveness as a robustness measure when data distortion around the cutoff is present. We also conclude that the covariate-adjusted RD design proposed by Frölich & Huber (2019) generally yields more precise results compared to the other models considered.

1 Introduction

The regression discontinuity (RD) design is a method used to estimate the effect that a certain treatment has on an outcome variable Y. It does so by comparing the treatment group to the control group, where assignment to the treatment group is determined by a running variable X and a fixed cutoff point c for each unit. There are two types of RD designs. The sharp RD (SRD) design, where every unit with $X > c$ is assigned to the treatment group, and fuzzy RD (FRD) design, where each unit has a certain probability of being assigned to the treatment group based on the value of X. A more detailed description of RD designs is provided by Imbens & Lemieux (2008).

The so-called 'donut' RD design was first introduced by Barreca et al. (2011). It estimates treatment effects by excluding observations with X values in a specific range around the cutoff c. The decision to employ the donut RD design is usually driven by concerns about potential distortions in RD treatment effect estimation due to systematic sorting or other data issues. The rationale is that if these concerns are unfounded, using the donut design should not significantly alter the estimates, thereby improving the robustness of estimations.

In addition Calonico et al. (2019) and Frölich & Huber (2019) have developed methods that include additional covariates in the RD design, which are not dependent on the running variable. We will refer to these models throughout the paper as CCFT and FH RD designs respectively. The aim of covariate-adjusted RD designs is to improve estimation accuracy.

The primary scope of this paper is to investigate whether the donut RD design improves the estimation of treatment effects. To do so, we evaluate basic and covariate-adjusted RD designs with and without the use of the donut. Additionally we compare the performance of covariate-adjusted RD designs to determine which performs best.

Our results will be divided in two parts. Firstly we'll consider two simulation studies, of which the first replicates the work of Noack & Rothe (2023). The second one extends it by including the covariate-adjusted designs. Subsequently, we provide an empirical application to analyze the impact of Head Start assistance on child mortality in the United States, utilizing a dataset compiled in 1965 based on 1960 census information. From our findings, we will evaluate the cases for which the donut implementation is most efficient, therefore helping policymakers and researchers in selecting the most appropriate models with a more complete overview on the topic.

To our knowledge, no previous research has evaluated the combination of donuts with covariate-adjusted RD designs, nor has there been a direct comparison of different covariate-adjusted RD designs. We contribute to the literature by addressing these aspects. Our findings support the use of the donut as robustness measure for RD designs.

This paper is structured as follows. We present a literature review on RD designs in Section 2. In Section 3, we describe the methodology for the various models. We illustrate the simulation studies in Section 4 and present the empirical application in Section 5. Lastly, we draw our conclusions in Section 6.

2 Literature Review

RD designs were first introduced by Thistlethwaite & Campbell (1960) as methods of testing causal hypotheses, used to study the effect of student scholarships on career aspiration. Their procedure consists in estimating two different regressions on each side of a cutoff which determines whether the treatment is applied to each unit. The estimates are obtained by subtracting the values of the two regressions when close to the cutoff. Hahn et al. (2001) and Imbens & Lemieux (2008) provide an extensive overview on RD designs and their estimation.

In recent years RD designs popularity has had a high increase, being applied in many different fields. Ludwig & Miller (2007) used RD designs to study the effect of Head Start assistance on child mortality in the United States, Salman et al. (2022) apply the fuzzy RD design to study the effect of Paris Agreement on global environmental efficiency and Ebenstein et al. (2017) study the effects that pollution has in China's life expectancy.

Barreca et al. (2011) estimate effect of very low birth weight classification on infant mortality, in doing so they introduce a 'donut' design, which excludes observations near the cutoff to avoid possible distortions of the data. Noack & Rothe (2023) further study the donut RD design, providing results that incentivize its use as a robustness measure. Even being a very recent method, Salvi et al. (2023) apply it in a study that focuses on the effect of exceeding the deductible on insurees' healthcare consumption in Switzerland.

Additionally, some papers developed some covariate-adjusted RD designs, where additional covariates are added in order to improve their estimation accuracy. The first covariate-adjusted model we consider in the paper is presented by Calonico et al. (2019), it proposes an addition to the local linear RD treatment effect estimator to include the covariates. They evaluate the model by estimating it on the Head Start dataset on child mortality in the United States which, as mentioned previously, was originally studied by Ludwig & Miller (2007) and will be also included in this paper's evaluations. Marshall (2024) applies their design to estimate the effects that certain characteristics of candidates have on winning elections and Carta & Rizzica (2018) uses it to study the effects that an early access to subsidized childcare for 2-year-old children in Italy has on several measures of maternal labor supply and on children's cognitive outcomes.

The second covariate-adjusted design is introduced by Frölich & Huber (2019), they propose a nonparametric approach to estimate the covariate-adjusted effects and evaluate it first on a simulation study, from which we take inspiration for our simulation, and on an empirical dataset. Bonfim et al. (2023) make use of their estimator to evaluate the effects that supporting small firms in Portugal has on the access to a government credit certification program.

This paper will expand the literature by providing a direct comparison of covariate-adjusted designs and by evaluating the use of the donut on these models.

3 Methodology

This section outlines the methodologies used to construct and evaluate our models.

3.1 Sharp regression discontinuity designs

The goal of RD designs is to estimate the causal effects of a treatment on an outcome variable Y. Each unit in the sample either receives the treatment or does not. Denote $Y_i(1)$ and $Y_i(0)$ as the outcomes with and without exposure to the treatment, respectively. We focus on the comparison between $Y_i(1)$ and $Y_i(0)$, often denoted by their difference. Since we cannot observe both $Y_i(1)$ and $Y_i(0)$ for the same unit, our focus is on the average effects of the treatment over the sample.

Let $W_i \in \{0, 1\}$ such that $W_i = 0$ denotes that unit i received no treatment, and $W_i = 1$ denotes that unit i received the treatment. The observed outcome of unit i is then:

$$Y_i = (1 - W_i) \cdot Y_i(0) + W_i \cdot Y_i(1). \tag{1}$$

Additionally, we can observe the pretreatment variables X_i and Z_i, where X_i is a scalar variable and Z_i is an M-vector. Which are not affected by the treatment. For each unit in the sample we observe the following quadruple (Y_i, W_i, X_i, Z_i).

In sharp RD designs (SRD), the assignment variable W_i is a deterministic function of X_i: $W_i = \mathbb{1}\{X_i \geq c\}$, where c is the cutoff over which units are assigned to the treatment group. In SRD designs, we study the discontinuity of the outcome given the covariate in order to estimate the average causal effect of the treatment:

$$\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x], \tag{2}$$

which can be interpreted as:

$$\tau_{SRD} = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = c]. \tag{3}$$

In this paper each model will be estimated with a sharp RD design. Further insights on RD designs are available in Imbens & Lemieux (2008).

We now define some regularity conditions that are needed to allow for (donut) sharp RD designs estimations and evaluations. These assumptions follow from Noack & Rothe

(2023).

Assumption 1: *X_i is continuous, with a distribution that follow the density function f_X, which is bounded away from the cutoff and includes the donut hole.*

Assumption 2: *For all $x \in \mathcal{X}$ and $q > 2$, $\mathbb{E}[(Y_i - \mathbb{E}[Y_i|X_i])^q|X_i = x]$ exists and is bounded uniformly. Furthermore, $\mathbb{V}[Y_i|X_i = x]$ is L-Lipschitz continuous $\forall x \in \mathcal{X} \setminus \{c\}$ and bounded uniformly away from c.*

Assumption 3: *The kernel K is a both symmetric and bounded density function, continuous in some compact set. It is always equal to zero outside that interval.*

These assumptions are useful to allow approximations in the methodology. The triangular and Epanechnikov kernels, as many other kernels, satisfy Assumption 3 and therefore can be used in our estimations.

3.2 Basic RD estimators

As in Noack & Rothe (2023), the estimate of τ using basic RD design will be made through local linear regression with the following formula:

$$\hat{\tau}(h) = e_1^\intercal \arg\min_\beta \sum_{i=1}^{n} K_h(X_i - c)(Y - (W_i, (X_i - c), W_i(X_i - c), 1)^\intercal \beta)^2, \quad (4)$$

where K is a kernel function with compact support and $K_h(x) = K(x/h)/h$. The bandwidth $h > 0$ determines the area of data around the cutoff used to estimate the treatment effects. Lastly $e_1 = (1, 0, 0, 0)^\intercal$ selects the right coefficient from the rest of the formula. This approach fits two different linear specifications for each side of the cutoff using weighted least squares, giving non-zero weights only to values $X_i \in [c - h, c + h]$.

The donut RD estimate is found by excluding the values of X_i close to the cutoff. The constant d determines the excluded area.

$$\hat{\tau}(h, d) = e_1^\intercal \arg\min_\beta \sum_{i=1}^{n} K_h(X_i - c)(Y - (W_i, (X_i - c), W_i(X_i - c), 1)^\intercal \beta)^2 \mathbb{1}\{|X_i - c| \geq d\}. \quad (5)$$

The latter works equally to the previous estimator, the only difference being that it only includes values $X_i \in [c - h, c - d] \cup [c + d, c + h]$. Note that the two estimators are the same when $d = 0$, which means $\hat{\tau}(h) = \hat{\tau}(h, 0)$.

Finally, for this design, we assume that μ^* belongs to the typical smoothness class of functions that are twice continuously differentiable, excluding the cutoff. With $\mu^* = \mathbb{E}\{Y_i|X_i^* = x\}$, where X_i^* is the value that the running variable would have without any data issues. This can be written as

$$\mu^* \in \mathcal{F}(M) = \{m_1(x)\mathbb{1}\{x \geq 0\} + m_0(x)\mathbb{1}\{x < 0\}, \|m_t''(\cdot)\|_\infty \leq M, t \in \{0,1\}\}, \tag{6}$$

where we assume the analyst to know the value of the uniform smoothness bound $M > 0$.

3.3 CCFT RD design

The covariates-adjusted models additionally include the vector of covariates Z_i in the estimation. This is done in order to remove small sample bias and improve the precision. The CCFT RD design was first presented by Calonico et al. (2019), it substitutes the standard local linear RD treatment effect estimator in (4) with:

$$\hat{\tau}_{CCFT}(h) = e_1^\intercal \arg\min_\beta \sum_{i=1}^n K_h(X_i - c)(Y - (W_i, (X_i - c), W_i(X_i - c), Z_i^\intercal, 1)^\intercal \beta)^2. \tag{7}$$

From which we can derive the donut variation

$$\hat{\tau}_{CCFT}(h, d) = e_1^\intercal \arg\min_\beta \sum_{i=1}^n K_h(X_i - c)(Y - (W_i, (X_i - c), W_i(X_i - c), Z_i^\intercal, 1)^\intercal \beta)^2 \mathbb{1}\{|X_i - c| \geq d\}. \tag{8}$$

For this design we need some additional assumptions to allow for the inclusion of covariates, the same used by Calonico et al. (2019). For $t \in \{0,1\}$ and $\forall x \in \mathcal{X}$:

Assumption 4: $\mathbb{E}[Z_i(t)|X_i = x]$ *is thrice continuously differentiable and* $\mathbb{E}[Z_i(t)Y_i(t)|X_i = x]$ *is continuously differentiable.*

Assumption 5: $\mathbb{V}[(Y_i(t), Z_i(t)^\intercal)|X_i = x]$ *is continuously differentiable and invertible.*

Assumption 6: $\mathbb{E}[\|(Y_i(t), Z_i(t)^\intercal)\|^4|X_i = x]$ *is continuous, with* $\|\cdot\|$ *being the Euclidean norm.*

3.4 FH RD design

The last model was presented in Frölich & Huber (2019). Their paper states that under a set of assumptions the treatment effect is nonparametrically identified as

$$\tau_{FH}(h) = \frac{\int (m^+(z,c) - m^-(z,c)) \cdot \frac{f^+(z|c) + f^-(z|c)}{2}\, dx}{\int (d^+(z,c) - d^-(z,c)) \cdot \frac{f^+(z|c) + f^-(z|c)}{2}\, dx}, \tag{9}$$

where $m^+(Z,x) = \lim_{\epsilon\to 0}\mathbb{E}[Y \mid Z, X = x+\epsilon]$, $m^-(Z,x) = \lim_{\epsilon\to 0}\mathbb{E}[Y \mid Z, X = x-\epsilon]$, $d^+(Z,x)$ and $d^-(Z,x)$ defined in the same way with W replacing Y. The estimator used for (9) is

$$\hat{\tau}_{FH}(h) = \frac{\sum_{i=1}^n (\hat{m}^+(Z_i,c) - \hat{m}^-(Z_i,c)) \cdot K_h(X_i - c)}{\sum_{i=1}^n (\hat{d}^+(Z_i,c) - \hat{d}^-(Z_i,c)) \cdot K_h(X_i - c)}, \tag{10}$$

where $\hat{m}$ and $\hat{d}$ are nonparametric estimators, K is a kernel function and $K_h(x) = K(x/h)/h$. For the sharp design it can be simplified to

$$\hat{\tau}_{FH}(h) = \frac{\sum_{i=1}^n (\hat{m}^+(Z_i,c) - \hat{m}^-(Z_i,c)) \cdot K_h(X_i - c)}{\sum_{i=1}^n K_h(X_i - c)}. \tag{11}$$

Including the donut hole this finally becomes:

$$\hat{\tau}_{FH}(h,d) = \frac{\sum_{i=1}^n (\hat{m}^+(Z_i,c) - \hat{m}^-(Z_i,c)) \cdot K_h(X_i - c) \cdot \mathbb{1}\{|X_i - c| \geq d\}}{\sum_{i=1}^n K_h(X_i - c) \cdot \mathbb{1}\{|X_i - c| \geq d\}}. \tag{12}$$

3.5 Performance measures

To assess models performance we compute the bias $b(h,d)$, standard deviation $s(h,d)$, and root mean squared error (RMSE) $r(h,d)$. Given the actual value of the parameter τ^*, these measures for estimate $\hat{\tau}(h,d)$ can be found as

$$\begin{aligned} b(h,d) &= \sum_{i=1}^n \omega_i(h,d)(\mu(X_i) - \tau^*), \\ s(h,d) &= \sqrt{\sum_{i=1}^n \omega_i(h,d)^2 \sigma_i^2}, \\ r(h,d) &= \sqrt{b^2(h,d) + s^2(h,d)}, \end{aligned} \tag{13}$$

where $\mu(x) = \mathbb{E}[Y_i|X_i = x]$, with ω_i being the weights selected from the running variable and with $\sigma_i = \mathbb{V}[Y_i|X_i]$.

3.6 Confidence intervals

We employ 'bias-aware' confidence intervals (Armstrong & Kolesár, 2018), as utilized by Calonico et al. (2019), to evaluate the donut estimator in the replication. To compute them we first need to find the 'worst case' bias of the donut RD estimator:

$$\bar{b}(h,d) = -\frac{M}{2}\sum_{i=1}^n \omega_i(h,d) X_i^2 \operatorname{sign}(X_i). \tag{14}$$

By denoting its conditional variance as

$$\hat{s}^2(h,d) \equiv \sum_{i=1}^{n} \omega_i(h,d)^2 \hat{\sigma}_i^2, \tag{15}$$

where $\hat{\sigma}_i^2$ are the estimates of the conditional variance $\sigma_i^2 = \mathbb{V}(Y_i|X_i)$. We decompose the usual t-statistic as

$$\frac{\hat{\tau}(h,d) - \tau^*}{\hat{s}(h,d)} = \frac{\hat{\tau}(h,d) - \tau^* - b(h,d)}{\hat{s}(h,d)} + \frac{b(h,d)}{\hat{s}(h,d)}. \tag{16}$$

We can therefore find the bias-aware confidence intervals as

$$C_n(d) = \left[\hat{\tau}(h,d) \pm \mathrm{cv}_{1-\frac{\alpha}{2}}\left(\frac{\bar{b}(h,d)}{\hat{s}(h,d)}\right)\hat{s}(h,d)\right], \tag{17}$$

with $\mathrm{cv}_{1-\frac{\alpha}{2}}(r)$ the $1-\frac{\alpha}{2}$ quantile of $N(r,1)$.

3.7 Specification testing

Here, we present methods to assess the statistical significance of model differences. The first two are used in Noack & Rothe (2023) and will be used for the replication part of this paper, while the third one will be used to additionally compare the covariate-adjusted models.

3.7.1 Donut and conventional

Firstly, we focus on comparing the donut estimates with the conventional ones. To do so we define the following null hypothesis

$$H_0 : \mu(x) = \mu^*(x) \text{ for all } |x| < d, \tag{18}$$

and define the difference between the two estimates as

$$\hat{\Delta}(h,d) \equiv \hat{\tau}(h,d) - \hat{\tau}(h,0) = \sum_{i=1}^{n}(\omega_i(h,d) - \omega_i(h,0))Y_i. \tag{19}$$

With the following bias and conditional variance

$$\begin{aligned} b_\Delta(h,d) &= \sum_{i=1}^{n}(\omega_i(h,d) - \omega_i(h,0))\mu(X_i) \text{ and} \\ s_\Delta^2(h,d) &= \sum_{i=1}^{n}(\omega_i(h,d)^2 + \omega_i(h,0)^2 - 2\omega_i(h,d)\omega_i(h,0))\sigma_i^2. \end{aligned} \tag{20}$$

Which can be estimated as

$$\sup_{\mu \in \mathcal{F}(M)} |b_\Delta(h,d)| \equiv \bar{b}_\Delta(h,d) = -\frac{M}{2}\sum_{i=1}^{n} \omega_i(h,d) - \omega_i(h,0))X_i^2 \text{sign}(X_i) \text{ and}$$
$$\hat{s}_\Delta^2 = \sum_{i=1}^{n}(\omega_i(h,d)^2 + \omega_i(h,0)^2 - 2\omega_i(h,d)\omega_i(h,0))\hat{\sigma}_i^2, \tag{21}$$

with $\hat{\sigma}_i^2$ a nearest-neighbor estimate of σ_i^2. We can now formulate the following t-statistic and decision rule.

$$t_\Delta = \frac{\hat{\Delta}(h,d)}{\hat{s}_\Delta(h,d)},$$
$$\text{Reject } H_0 \text{ if } |t_\Delta| > \ \text{cv}_{1-\frac{\alpha}{2}}\left(\frac{\bar{b}_{\Delta(h,d)}}{\hat{s}_{\Delta(h,d)}}\right). \tag{22}$$

3.7.2 Donut and within donut

The next test compares the donut estimator with a basic RD design using d as bandwidth. Therefore making a comparison between the effects inside and outside of the donut area. The test follows the same structure of the previous one, but replacing $\hat{\Delta}(h,d)$ with

$$\hat{\Gamma}(h,d) \equiv \hat{\tau}(h,d) - \hat{\tau}(d,0) = \sum_{i=1}^{n}(\omega_i(h,d) - \omega_i(d,0))Y_i. \tag{23}$$

The rest of the test is formulated in the same way as the previous one, leading to the following t-statistic and decision rule

$$t_\Gamma = \frac{\hat{\Gamma}(h,d)}{\hat{s}_\Gamma(h,d)},$$
$$\text{Reject } H_0 \text{ if } |t_\Gamma| > \ \text{cv}_{1-\frac{\alpha}{2}}\left(\frac{\bar{b}_{\Gamma(h,d)}}{\hat{s}_{\Gamma(h,d)}}\right), \tag{24}$$

with $\text{cv}_{1-\frac{\alpha}{2}}(r)$ the $1-\frac{\alpha}{2}$ quantile of $N(r,1)$.

3.7.3 Covariate-adjusted and conventional

Lastly, we want to compare the performances of the covariate-adjusted models. To do so we use the following steps in order to compare them with the conventional RD and with each other. This method simply consists in a t-test between the estimates of two models $\hat{\tau}_1$ and $\hat{\tau}_2$. We formulate the following null hypothesis

$$H_0 : \hat{\tau}_1 = \hat{\tau}_2, \tag{25}$$

and define the variable

$$\hat{\Omega}(h,d) = \hat{\tau}_1(h,d) - \hat{\tau}_2(h,d), \text{ with } h > d \geq 0. \tag{26}$$

We can now define the t-statistic and the decision rule:

$$\begin{gathered} t_\Omega = \frac{\hat{\Omega}(h,d)}{\hat{s}_\Omega(h,d)}, \\ \text{Reject } H_0 \text{ if } |t_\Omega| > \ \text{cv}_{1-\frac{\alpha}{2}}(0), \end{gathered} \tag{27}$$

where $\text{cv}_{1-\frac{\alpha}{2}}(r)$ is the $1-\frac{\alpha}{2}$ quantile of $N(r,1)$ and $\hat{s}_\Omega(h,d)$ is the conditional variance of $\hat{\Omega}$.

3.8 Empirical application: bootstrap

Lastly, for the empirical application, we will apply a bootstrap method to compare the results, this was also done by Frölich & Huber (2019). The bootstrap method involves randomly selecting observations from the dataset with replacement. This process creates bootstrap samples of the same size of the original dataset, which will be used to compute the estimates. We will repeat this procedure r times, resulting in r estimates for each model.

We will extract the model estimates by taking the simple averages of the outcomes. This method is advantageous because it allows us to estimate 95% bootstrap percentile confidence intervals by selecting the values below and above which lie 2.5% of the bootstrap estimates, resulting in robust results (Hesterberg, 2011).

4 Simulation studies

We utilize two distinct simulated datasets. The first dataset is equivalent to the one used by Noack & Rothe (2023), and is used for the replication of their results. The procedure is delineated as follows:

$$\begin{aligned} &X_i \sim U(-1,1) \\ &Y_i = \lambda_L(X_i) + \epsilon_i, \text{ where } \epsilon_i \sim \mathcal{N}(0, 0.5) \text{ independent of } X_i, \text{ with} \\ &\lambda_L(x) = \text{sign}(x)x^2 - L\text{sign}(x)((x - 0.1 \cdot \text{sign}(x))^2 \\ &\qquad\qquad - 0.1^2 \cdot \text{sign}(x))\mathbb{1}\{|x| < 0.1\} \text{ for } L \in \{0, 10, ..., 40\}. \end{aligned} \tag{28}$$

Here, X_i and Y_i denote the running and outcome variables, respectively. This dataset will be re-estimated for each value of L, which indicates the distortion of the data inside the interval $(-0.1, 0.1)$. The case where $L = 0$ indicates no need for the donut estimator. We set the number of observations to $n = 1000$ and replications to $r = 10000$.

Tables 1, 2 and 3 display results from the replication of Noack & Rothe (2023), these estimations are made by using a triangular kernel, with MSE-optimal bandwidths, $d = 0.1$ and $M = 2$[1]. As we can notice, these tables do not align with the results reported by Noack & Rothe (2023), which could be caused by a possible computation mistake in their estimations.

Table 1: Replication results: point estimation.

	Bias		Std. Dev.		RMSE	
L	Regular	Donut	Regular	Donut	Regular	Donut
0	-0.049	-0.128	0.098	0.152	0.110	0.199
10	-0.118	-0.127	0.099	0.150	0.153	0.197
20	-0.187	-0.128	0.100	0.149	0.212	0.196
30	-0.255	-0.127	0.101	0.152	0.274	0.198
40	-0.322	-0.127	0.103	0.149	0.338	0.196

Table 2: Replication results: confidence intervals.

	CI Coverage		CI Length	
L	Regular	Donut	Regular	Donut
0	0.950	0.946	0.431	0.749
10	0.836	0.950	0.430	0.749
20	0.607	0.950	0.431	0.748
30	0.351	0.949	0.430	0.749
40	0.150	0.950	0.431	0.749

Table 3: Replication results: specification testing.

	Rejection Frequency	
L	$\hat{\Lambda}$	$\hat{\Gamma}$
0	0.027	0.029
10	0.012	0.018
20	0.023	0.062
30	0.057	0.186
40	0.131	0.410

Table 1, shows that the donut estimates consistently have higher standard deviations compared to standard RD. In terms of bias and RMSE, we only see an improvement in performance when $L \geq 20$. In Table 2, we note that the confidence intervals estimated with standard RD are less reliable when $L \geq 10$, while donut confidence intervals always keep a coverage of $\approx 95\%$. Lastly, Table 3 shows that the donut estimates provide a statistically significant improvement compared to conventional estimates solely when $L \geq 30$, and to the within donut estimates when $L \geq 20$. These outcomes suggest that

[1]These estimations are done in R using the RDHonest package

substantial data distortion around the cutoff is necessary to justify donut estimations. However, standard RD remains preferable in scenarios with lower distortions. Proving that the donut estimates can be used as a more robust method, at the cost of some loss of precision.

The second simulated data will be used for the further methods comparisons, it is a combination of the simulations presented by Frölich & Huber (2019) and Noack & Rothe (2023), therefore allowing for both additional covariates and data distortion. It is fully described by Equation (29).

$$\begin{aligned}
&X_i, U_i, V_i, W_i \sim \mathcal{N}(0,1) \text{ independently of each other,} \\
&D_i = \mathbb{1}\{X_i > 0\}, \quad Z_{1i} = \alpha D_i + 0.5U_i, \quad Z_{2i} = \alpha D_i + 0.5V_i, \\
&Y_i = \lambda_L(X_i) + D_i + \beta(Z_{1i} + Z_{2i}) + \frac{\beta}{2}(Z_{1i}^2 + Z_{2i}^2) + W_i, \text{ with} \\
&\lambda_L(x) = 0.5x - 0.25x\mathbb{1}\{x > 0\} + 0.25x^2 - L\text{sign}(x)((x - 0.1 \cdot \text{sign}(x))^2 \\
&\qquad\qquad - 0.1^2 \cdot \text{sign}(x))\mathbb{1}\{|x| < 0.1\} \text{ for } L \in \{0, 10, ..., 40\}.
\end{aligned} \tag{29}$$

Where Y_i and X_i are respectively the outcome and running variables, with Z_{1i} and Z_{2i} being the additional covariates. The variable α defines the strength of the relationship between Z_{1i} and Z_{2i} and β defines their effect on Y_i. In this research, we set them as $\alpha = 0$ and $\beta = 0.4$ such that the covariates remain balanced around the threshold while still having an effect on Y_i. Finally L is defined analogously as in the replication dataset (8) and is needed for the donut estimations. For estimations we will use $M = 2$, $d = 0.1$ with the Epanechnikov kernel (Wang & Van Ryzin, 1981). The number of observations will again be set to $n = 1000$ and the replications to $r = 1000$[2].

Tables 4, 5 and 6 display bias, standard deviation and RMSE for each model. Starting with the basic RD model in Table 4, for $L = 0$ both the conventional and donut estimates have close to no bias. As L increases, the donut estimate demonstrates to consistently keep low biases, having instead an increase in standard deviations, showing an improvement in RMSE only when $L = 40$.

Basic RD design						
	Bias		Std. Dev.		RMSE	
L	Standard	Donut	Standard	Donut	Standard	Donut
0	-0.00	0.00	0.20	0.27	0.20	0.27
10	-0.05	0.01	0.19	0.27	0.20	0.27
20	-0.11	-0.02	0.20	0.27	0.23	0.27
30	-0.14	-0.00	0.20	0.27	0.25	0.27
40	-0.20	-0.00	0.20	0.28	0.29	0.28

Table 4: Simulation results: (donut) basic RD design.

[2]The estimations are computed in R with the packages RDHonest, rdrobust and np

Similarly, the CCFT RD design design (Table 5) does see an improvement in the bias by using the donut estimate when L increases. However, due to its higher standard deviation, the RMSE never improves with the use of donut.

Comparing the CCFT RD estimates with the basic donut RD estimates we find that the covariate-adjusted design demonstrates slightly worse biases and standard deviations when the donut is not included. When the donut hole is included, the standard deviation further diverge with the basic RD design one. Demonstrating consistently higher RMSEs, and therefore worse precision.

CCFT RD design								
	Bias		Std. Dev.		RMSE		$\hat{\Omega}$ Rejection Frequency	
L	Standard	Donut	Standard	Donut	Standard	Donut	Standard	Donut
0	-0.00	0.01	0.21	0.34	0.21	0.34	0.055	0.056
10	-0.05	0.02	0.20	0.33	0.21	0.33	0.058	0.061
20	-0.12	-0.02	0.21	0.35	0.25	0.35	0.049	0.048
30	-0.16	-0.01	0.23	0.34	0.28	0.34	0.051	0.065
40	-0.23	0.01	0.24	0.37	0.34	0.37	0.073	0.043

Table 5: Simulation results: (donut) CCFT RD design.

Lastly, when investigating the FH RD design biases, standard deviations and RMSEs (Table 6) we again have similar findings. The donut bias does have an improvement whenever L increases, with a fairly low standard deviation. The RMSE sees an improvement when $L \geq 20$.

FH RD design does see an improvement when compared to basic RD biases, displaying slightly lower values. Standard deviation tends to increase with L, resulting in lower RMSEs when $L \leq 30$ and slightly higher otherwise. When the donut hole is included, this model shows to be the best performing, keeping both low biases and standard deviations. With consistently lower RMSEs than the other two models.

These findings align to our predictions and maintain consistency between each other. It is clear that the inclusion of the donut hole in RD designs does improve their biases when a distortion is included. On the other hand, this increases the standard deviations of the estimates, which lowers the precision of the results. Considering the RMSE results we find that the donut estimate improves estimations strictly when the values of L get bigger, never being able to improve the results when the distortion is not present. These findings provide an indication that the donut design can be justified when a possible distortion of the data is expected or taken into consideration.

Finally Table 7 shows the t-test performed between the two covariate-adjusted models with and without the donut hole. The two models show to have a small statistically significant difference in performance, with all values > 0.05. Indicating FH to be a better model for all values of L both with and without the donut hole.

FH RD design								
	Bias		Std. Dev.		RMSE		$\hat{\Omega}$ Rejection Frequency	
L	Standard	Donut	Standard	Donut	Standard	Donut	Standard	Donut
0	-0.01	-0.01	0.17	0.22	0.17	0.22	0.047	0.036
10	-0.04	0.00	0.17	0.21	0.18	0.21	0.059	0.048
20	-0.09	-0.01	0.20	0.21	0.22	0.21	0.045	0.056
30	-0.13	-0.00	0.23	0.21	0.26	0.21	0.052	0.052
40	-0.19	0.01	0.25	0.21	0.32	0.21	0.056	0.048

Table 6: Simulation results: (donut) FH RD design.

Surprisingly, from the results we find the CCFT RD design to have the worse performances, even lower than the basic RD design. The FH RD design turns out to be the most precise design, it seems to perform better compared to the other models both when the distortion of the data is low or when the donut hole is included. Making it clear that when both some distortion or no distortion is expected, FH RD design could be the best choice, respectively with the donut hole or without. The only case in which the basic RD would have better performance is if the distortion is present but not expected, leading to a model without the donut hole even if needed.

	$\hat{\Omega}$ Rejection Frequency	
L	Standard	Donut
0	0.061	0.052
10	0.059	0.063
20	0.053	0.064
30	0.056	0.060
40	0.063	0.051

Table 7: Simulation results: comparison of the covariate-adjusted models.

5 Empirical application

Lastly, we consider a real dataset for a further comparison of the models. We follow Calonico et al. (2019)'s approach and study the effect of Head Start assistance on child mortality in the United States. We use as running variable the county-level poverty index, which was constructed in 1965 by the federal government based on 1960 census information, with cutoff $\bar{x} = 59.2$, and as outcome variable the child mortality rate due to causes affected by Head Start's health services components. The dataset contains nine additional variables, for faster computations we will only include two of those as additional covariates. We select the variables with the highest absolute correlation to the outcome variable. All the correlations are shown in Appendix A. Table 8 shows summary statistics of the variables we use in our designs.

	Mean	Std. Dev.	Median	Min.	Max.	Skewness	Kurtosis
mort_HS	2.26	5.73	0	0	136.1	10.0	177.9
povrate60	36.7	15.3	33.6	15.2	81.6	0.6	-0.6
pctsch534	0.55	0.06	0.55	0.20	0.75	-0.4	1.8
pctblack	10.6	16.8	1.5	0.0	83.4	1.8	2.4

Table 8: Summary statistics of Head Start dataset. Here *mort_HS* is the outcome variable, it is the United States morality rates per 10000 children between 5 and 9 years old between 1973-1983. The running variable is *povrate60*, the county's poverty rate in 1960 relative to the 300$^{\text{th}}$ poorest county. Lastly *pctsch534* and *pctblack* are the additional covariates, they respectively are the percentage of children aged 3 to 5 and the percentage of black population. Observations with missing values were discarded prior to the calculations.

We estimate the models using the bootstrap samples with $r = 200$, the Epanechnikov kernel (Wang & Van Ryzin, 1981), $M = 1$ and $d = 0.2$. The means and confidence interval lengths of the resulting estimates are displayed in Table 9, together with the point estimates obtained with the full dataset. We additionally provide plots that show the distribution of the estimates and their confidence intervals in Figure 1. These are added to allow for a direct visual comparison of the models.

	Basic RD		CCFT RD design		FH RD design	
	Standard	Donut	Standard	Donut	Standard	Donut
Point estimate	-3.28	-3.20	-1.96	-1.86	-1.67	-1.39
Bootstrap estimate	-3.48	-3.32	-2.76	-2.71	-1.38	-1.30
Bootstrap interval length	5.22	5.85	5.41	5.93	7.37	8.36

Table 9: Empirical application results.

The bootstrap and standard estimates seem to have close values for basic RD and FH RD designs, with the CCFT RD design having the highest differences. This could be a sign of being more sensitive to duplicate observations. For what regards the use of the donut hole, we find similar results to the simulation study. the donut hole has a small effect on the estimates, while having an higher effect on the confidence intervals due to the higher standard deviations.

When comparing the models we find that the confidence intervals overlap, indicating that the estimates have similar values. With the CCFT RD design having slightly larger confidence intervals compared to the basic RD. Surprisingly, the FH results display significantly larger confidence intervals, giving an indication of the FH RD design having less accuracy compared to the other models. This could be a sign of the FH RD design being more sensitive to the different bootstrap samples. This can also be noticed in the estimates plots, where basic and CCFT RD designs have many values near the average, being closer to a normal distribution. The FH RD results' estimates are instead more uniform across the interval. These results do not completely align to the simulation results, where FH standard deviations were closer, and often lower, to the basic RD designs'.

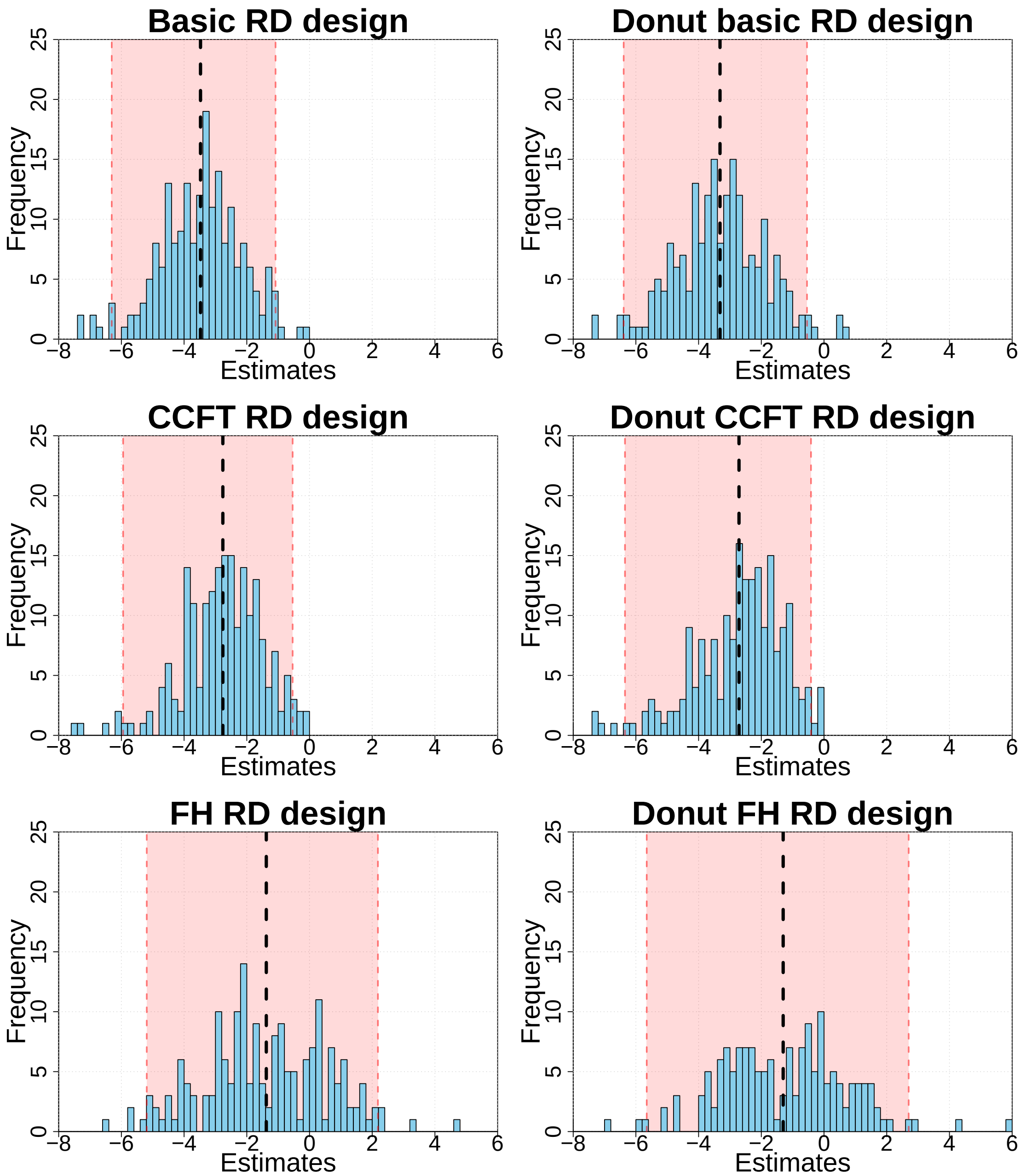

Figure 1: Plots of bootstrap results. For each model, the blue bars correspond to the frequency of each resulting estimate. The black dashed lines represents the average bootstrap results, the red dashed lines display the confidence intervals. The tables are all set in the same scale to allow for a direct comparison of the plots.

6 Conclusion

The scope of this research was to assess the performance of basic and covariate-adjusted regression discontinuity designs when estimating the causal effect of a treatment on an outcome variable. Donut RD designs serve as a robustness measure in our models by excluding a small area of observations suspected of distortion. Our models include the basic and two covariate-adjusted RD designs, each estimated with and without the donut hole. This approach allowed us to assess the donut design performance and to compare covariate-adjusted models across an empirical dataset and two simulated datasets.

Our simulation findings have indicated donut designs having higher standard deviations compared to standard designs, leading to slightly inferior performances when the data distortion is minimal or not present. On the other hand it significantly improves biases and root mean squared errors under conditions of high distortion, demonstrating its effectiveness in ensuring robustness when distortion is suspected.

Regarding the comparison of covariate-adjusted models, we surprisingly find that Calonico et al. (2019)'s model does not outperform the basic RD design, consistently resulting in inferior estimates. Frölich & Huber (2019)'s model instead proves to be a significant improvement to the basic RD for small amounts of distortion and over all amounts of distortion when the donut hole is included. These findings suggest Frölich & Huber (2019)'s model performs best overall, making it the recommended choice in the presence of covariates.

The empirical application indicates similar conclusions for what regards the donut implementation and Calonico et al. (2019)'s model performance, it however shows lower precision by Frölich & Huber (2019)'s design. This could be due to a higher sensitivity to the bootstrap samples.

For future research we make the following suggestions. Firstly, we did not account for different bandwidth optimization or kernels, these could provide a more complete view of the results with additional insights. We then suggest comparing different donut values, at the moment the donut size is an arbitrary choice, a research on finding an optimal donut value could help for further improvement of the models. Lastly, we propose evaluating the various models under conditions where covariates are correlated, to investigate potential impacts on the estimates.

References

Armstrong, T. B. & Kolesár, M. (2018). Optimal inference in a class of regression models. *Econometrica*, *86*(2), 655–683.

Barreca, A. I., Guldi, M., Lindo, J. M. & Waddell, G. R. (2011). Saving babies? revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics*, *126*(4), 2117–2123.

Bonfim, D., Custódio, C. & Raposo, C. (2023). Supporting small firms through recessions and recoveries. *Journal of Financial Economics*, *147*(3), 658–688.

Calonico, S., Cattaneo, M. D., Farrell, M. H. & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, *101*(3), 442–451.

Carta, F. & Rizzica, L. (2018). Early kindergarten, maternal labor supply and children's outcomes: evidence from italy. *Journal of Public Economics*, *158*, 79–102.

Ebenstein, A., Fan, M., Greenstone, M., He, G. & Zhou, M. (2017). New evidence on the impact of sustained exposure to air pollution on life expectancy from china's huai river policy. *Proceedings of the National Academy of Sciences*, *114*(39), 10384–10389.

Frölich, M. & Huber, M. (2019). Including covariates in the regression discontinuity design. *Journal of Business & Economic Statistics*, *37*(4), 736–748.

Hahn, J., Todd, P. & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(6), 497–526.

Imbens, G. W. & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635.

Ludwig, J. & Miller, D. L. (2007). Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly Journal of Economics*, *122*(1), 159–208.

Marshall, J. (2024). Can close election regression discontinuity designs identify effects of winning politician characteristics? *American Journal of Political Science*, *68*(2), 494–510.

Noack, C. & Rothe, C. (2023). Donut regression discontinuity designs. *arXiv preprint arXiv:2308.14464*.

Salman, M., Long, X., Wang, G. & Zha, D. (2022). Paris climate agreement and global environmental efficiency: New evidence from fuzzy regression discontinuity design. *Energy Policy*, *168*, 113128.

Salvi, I., Cordier, J., Kuklinski, D., Vogel, J. & Geissler, A. (2023). *Price sensitivity and demand for healthcare services: Investigating demand-side financial incentives using anonymised claims data from switzerland* (Tech. Rep.). St.Gallen: Working Paper Series in Health Economics, Management and Policy.

Thistlethwaite, D. L. & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*(6), 309.

Wang, M.-C. & Van Ryzin, J. (1981). A class of smooth estimators for discrete distributions. *Biometrika*, *68*(1), 301–309.

A Covariates correlations

The Head Start dataset contains nine possible additional covariates for our models. These variables are levels and percentages of the population population for children aged 3 to 5, children aged 14 to 17 and adults older than 25. The total population and the percentages of black and urban populations are also included. To have easier computations we only select two, based on their absolute correlation with the outcome variable. This is done to increase the effects these covariates have on the estimates. Figure 2 displays the correlation of every covariate with the outcome variable. Even though all values show low correlations, we find *pctsch534* and *pctblack* to have the highest absolute correlations.

Figure 2: Correlations between each covariate and the outcome variable, red and blue bars represent negative and positive correlations respectively.