Universidad de San Andrés

Departamento de Economía

Maestría en Economía

Extrapolation in Multiple-Cutoff Regression Discontinuity Designs

Gonzalo Vázquez Baré

DNI: 32479425

Mentor: Matias D. Cattaneo

Buenos Aires

13 de Agosto, 2019

Tesis de Maestría en Economía de

Gonzalo VÁZQUEZ BARÉ

"Extrapolación en diseños de regresión discontinua con múltiples cutoffs"

Resumen

Si bien la mayor parte de la literatura econométrica sobre diseños de regresión discontinua se enfoca en el caso de un solo cutoff (valor de corte), frecuentemente se analizan programas o políticas en los que distintas observaciones enfrentan distintos umbrales. El presente trabajo introduce la idea de selección a cutoffs y estudia identificación de efectos de tratamiento cuando unidades enfrentando distintos cutoffs pueden diferir en variables observables y no observables. El resultado principal provee condiciones bajo las cuales es posible identificar y estimar un efecto de tratamiento fuera del cutoff. Los resultados son ilustrados usando datos de Oportunidades, un programa anti-pobreza de gran escala en México en el cual elegibilidad a nivel de hogar fue basada en un índice de pobreza con distintos cutoffs según la región geográfica.

Palabras clave: extrapolación, diseños de regresión discontinua, efectos de tratamiento, inferencia causal.

"Extrapolation in multiple-cutoff regression discontinuity designs"

Abstract

Although the bulk of the econometric literature on Regression Discontinuity Designs (RDDs) focuses on the case of a single cutoff, researchers frequently study programs or policies in which different observations face different thresholds. This paper introduces the idea of selection into cutoffs and studies identification of treatment effects when units facing different cutoffs are allowed to differ in both observable and unobservable characteristics. The main result provides conditions under which it is possible to identify and estimate a treatment effect away from the cutoff. The findings are illustrated using data from Oportunidades, a large-scale anti-poverty program in Mexico in which household eligibility was based on a poverty index with cutoffs varying across geographical regions.

Keywords: extrapolation, regression discontinuity designs, treatment effects, causal inference.

Códigos JEL: C10, C13, C18.

1 Introduction

The Regression Discontinuity Design (henceforth, RDD) is currently one of the most widely used methodologies to estimate the effect of programs and policies in the social sciences, and is applied in a broad range of fields including Economics, Political Science, Education and Criminology and others. Originally proposed by Thistlethwaite and Campbell (1960), this methodology exploits a discontinuity in the probability of assignment to a program or treatment around a known threshold, inducing variation in participation that is (locally) unrelated to potential confounders. In general, the RDD is an appealing method whenever the probability of treatment assignment jumps discontinuously when a score exceeds a threshold.

Although the most basic application of RDDs involves a single cutoff, frequently researchers study more complicated programs or policies where assignment rules are based on multiple thresholds. For example, Brollo, Nannicini, Perotti, and Tabellini (2013), De la Mata (2012) and Cerqua and Pellegrini (2014) use RDD approaches to analyze the impact of programs where the cutoff changes over geographical sites. On the other hand, Benavente, Crespi, Figal Garone, and Maffioli (2012) use data from a subsidy policy implemented over a period of 15 years in Chile where the cutoff changes every year. Black, Galdo, and Smith (2007) study a reemployment program in Kentucky in which the cutoff changes by week and local employment office. Angrist and Lavy (1999) exploit discontinuities generated by class size caps at each multiple of 40 to estimate the effect of class size on academic achievement. In Political Economy and Political Science, the analysis of incumbency advantage represents a case where multiple cutoffs are pervasive. Starting with Lee (2008), researchers have estimated the effect of incumbency advantage on different outcomes by looking at elections where the margin of victory (i.e. difference in vote shares) is small. Since different elections have different number of competing parties, the threshold varies across elections.

While this situation is common in practice, the methodological and econometric literature provides little, if any, guidance on how to adapt the RDD to the case of multiple cutoffs. Applied researchers have used a variety of approaches to deal with this situation in practice. A common strategy is to reduce the problem to a single cutoff by looking at distance to the cutoff and pooling all the samples together (see Cattaneo, Keele, Titiunik, and Vazquez-Bare, 2016, for a comprehensive list of studies adopting this strategy). Other studies try to account for heterogeneity in treatment effects by calculating estimates at each cutoff separately (Brollo, Nannicini, Perotti, and Tabellini, 2013; De la Mata, 2012). In other cases, researchers calculate the treatment effect at each threshold and then construct some type of weighted average of the estimates (see e.g. Black, Galdo, and Smith, 2007; Cerqua and Pellegrini, 2014). In all cases, the question remains of whether there is a better way to exploit variability in cutoffs.

The purpose of this paper is to derive identification results in RDDs with multiple cutoffs, henceforth, multi-cutoff RDDs. We extend the results in Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) and study conditions under which the availability of multiple cutoffs can be exploited to estimate treatment effects over different parts of the distribution of the running variable. The main idea of the paper is that, although multiple thresholds provide valuable information to uncover heterogeneity in treatment effects, direct comparisons between estimates at each cutoff may not be sensible when units have control on which cutoff they face. We introduce the idea of selection into cutoffs and study conditions under which potential biases introduced by this phenomenon can be dealt with.

The remainder of the paper is organized as follows. Section 2 discusses the econometric framework and briefly reviews the existing literature on multiple-cutoff RDDs. Section 3 discusses parameters of interest and provides the main identification results. Section 4 illustrates the identification results using data from a well-known poverty-alleviation program in Mexico (Oportunidades). Section 5 concludes.

2 Framework and existing literature

The statistical framework in this paper will be the usual potential outcomes framework, and we will follow closely the notation in Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016). Let X_i be the running variable (score) for unit i, which is assumed to be continuous with a continuous density $f_X(x)$. Unlike a single-cutoff RDD, the cutoff faced by unit i will be a random variable C_i taking values in a finite set $\mathcal{C} = \{c_0, c_1, ..., c_J\}$. Intuitively, C_i can be thought of as a group indicator. For example, consider a policy that is implemented in regions A and B, where the cutoff is 100 in region A and 120 in region B. Then $C_i = 100$ for all units in region A and $C_i = 120$ for all units in region B. Importantly, the cutoffs are mutually exclusive, in the sense that each unit can face only one cutoff. The conditional density of the running variable at each cutoff will be denoted $f_{X|C}(x|c)$ for each $c \in \mathcal{C}$. The treatment indicator is $D_i \in \{0, 1\}$, which is a function of both the running variable and the cutoff, $D_i = D_i(X_i, C_i)$. For instance, in a sharp RDD, $D_i = \mathbf{1}(X_i \geq C_i)$ where $\mathbf{1}(.)$ is the indicator function. The re-centered running variable will be $\tilde{X}_i = X_i - C_i$. Finally, Y_{1i} and Y_{0i} denote the potential outcomes of unit i under treatment and control, respectively, and $Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i)$ is the observed outcome. The treatment effect for unit i is defined as $\boldsymbol{\tau}_i = Y_{1i} - Y_{0i}$.

In this setup, Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) study identification when the running variable is re-centered to normalize all cutoffs to zero, thus reducing the problem to a single cutoff. This approach, common among applied researchers, is referred to as the "pooling approach". The authors show that the pooled RD estimand identifies a weighted average of average treatment effects at different cutoffs, where the weights depend on the conditional density of the running variable at each cutoff and the proportion of the sample facing each cutoff. Formally, the pooled estimand is defined as:

$$\tau^p = \lim_{x \searrow 0} \mathbb{E}[Y_i \mid \tilde{X}_i = x] - \lim_{x \nearrow 0} \mathbb{E}[Y_i \mid \tilde{X}_i = x]$$

Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) assume the following conditions:

1. **(sharp RD)**: $\lim_{x \searrow c} \mathbb{E}[D_i \mid X_i = x, C_i = c] = 1$ and $\lim_{x \nearrow c} \mathbb{E}[D_i \mid X_i = x, C_i = c] = 0$, for all $c \in \mathcal{C}$.

2. **(continuity of the regression functions)**: $\mathbb{E}[Y_{0i} \mid X_i = x, C_i = c]$ and $\mathbb{E}[Y_{1i} \mid X_i = x, C_i = c]$ are continuous in x at $x = c$ for all $c \in \mathcal{C}$.

3. **(continuity of the conditional density)**: $f_{X|C}(x|c)$ is continuous in x at $x = c$ for all $c \in \mathcal{C}$.

Under these assumptions, Lemma 1 in the paper shows that the pooled estimand can be written as:

$$\tau^p = \sum_{c \in \mathcal{C}} \mathbb{E}[\tau_i \mid X_i = c, C_i = c]\, \omega(c) \tag{1}$$

where

$$\omega(c) = \frac{f_{X|C}(c|c)\mathbb{P}[C_i = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i = c]} \tag{2}$$

One important difference with the single-cutoff RDD is worth discussing. While in single-cutoff RDD the the regression functions depend on X_i alone, in a multi-cutoff RDD the regression functions are allowed to depend on the cutoff C_i. This additional notation is conceptually relevant as it highlights two sources of heterogeneity: the treatment effect may change accross cutoffs because it is evaluated at different values of the running variable, but also because individuals exposed to different cutoffs can have different responses, for example, when units can self-select into cutoffs. This will be one of the main points of this paper and will be discussed in detail in the following sections.

This second source of endogeneity, namely, that units exposed to different cutoffs may be heterogeneous in both observed and unobserved characteristics, has been studied by Hotz, Imbens, and Mortimer (2005) and later by Allcott (2015) in the context of randomized trials implemented across different sites. Their setting is similar to ours, since RDDs can be seen as local randomized experiments (Lee, 2008; Cattaneo, Frandsen, and Titiunik, 2015). The approach followed by the authors is to assume unconfoundedness, i.e. that sites can be made comparable by controlling for observed differences. We will consider selection on observables in section 3.2 and extend the results to cases in which this assumption fails in section 3.3.

A somewhat recent and closely related literature analyzes RDD under multiple running variables, which occurs, for example, when a scholarship is assigned based on the results in two different exams (Keele and Titiunik, 2015; Imbens and Zajonc, 2011; Papay, Willett, and Murnane, 2011; Wong, Steiner, and Cook, 2013). This setup is conceptually different to the one considered in this paper, although as shown in Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016), both designs can be subsumed under a single general framework.

3 Identification results

As suggested in the previous section, in a multi-cutoff RDD pooling all subsamples ignores variability in cutoffs and may hide important heterogeneity in average treatment effects across groups. In other words, the pooling approach wastes valuable information that may be useful to identify average treatment effects at different points of the distribution of the running variable. At first glance, this problem seems easy to overcome, as one can always estimate a single-cutoff RDD for each available cutoff so long as the sample size allows it. It might be tempting in this case to compare these different estimates to see how the treatment effect changes with the value of the score. However, we argue that the comparison between estimates across cutoffs may not be a fair comparison when selection into cutoffs is possible.

To understand this point, it is instructive first to define some parameters of interest in the context of multi-cutoff RDDs. The most obvious magnitude of interest is the average treatment effect at some cutoff $c \in \mathcal{C}$ for the subpopulation facing that particular cutoff:

$$\tau(c, c) := \mathbb{E}[\tau_i \mid X_i = c, C_i = c] \tag{3}$$

This parameter can be recovered by estimating a single-cutoff regression discontinuity for each group, which will be consistent under the usual RDD conditions given for example in Hahn, Todd, and van der Klaauw (2001) or Lee (2008).

Another relevant parameter is the average treatment effect for the group facing cutoff $c \in \mathcal{C}$ but with the score evaluated at some other cutoff $c' \in \mathcal{C}$:

$$\tau(c', c) := \mathbb{E}[\tau_i \mid X_i = c', C_i = c] \tag{4}$$

Intuitively, this magnitude captures the effect that the subpopulation facing cutoff c would have experienced had it faced cutoff c' instead. Because this is a counterfactual magnitude, in a single-cutoff RDD, this parameter is not identified unless one is willing to impose strong parametric assumptions on the shape of the conditional expectation. However, in section 3.3 we will provide conditions under which this quantity is nonparametrically identified when multiple cutoffs are available.

Finally, the difference between 3 and 4 captures the change in the average treatment effect for the group facing cutoff c when the score is switched from c' to c:

$$\delta(c, c') = \mathbb{E}[\tau_i \mid X_i = c, C_i = c] - \mathbb{E}[\tau_i \mid X_i = c', C_i = c] \tag{5}$$

Now, the difference between two estimates at cutoffs $c, c' \in \mathcal{C}$ is consistent for:

$$\Delta(c, c') = \mathbb{E}[\tau_i \mid X_i = c, C_i = c] - \mathbb{E}[\tau_i \mid X_i = c', C_i = c'] \tag{6}$$

Hence, the above parameter is capturing the difference between average treatment effects at different values of the running variable, but for different subpopulations. Although depending on the context this could be a policy-relevant parameter, it has to be interpreted with caution as the comparison is not *ceteris paribus*. For instance, consider an anti-poverty program implemented in two different regions where each region has a different cutoff, and suppose that the running variable is a poverty index. In this case, each estimate captures the average treatment effect with the poverty index evaluated at the cutoff for each of the two regions. If the two regions differ in, say, the proportion of ethnic minorities, the treatment effects will vary not only because they are evaluated at different values of the poverty index but also because of the different ethnic composition in each region. In terms of the above parameters, we have that:

$$\begin{aligned}
\Delta(c, c') &= \mathbb{E}[\tau_i \mid X_i = c, C_i = c] - \mathbb{E}[\tau_i \mid X_i = c', C_i = c'] \\
&= \mathbb{E}[\tau_i \mid X_i = c, C_i = c] - \mathbb{E}[\tau_i \mid X_i = c', C_i = c] \\
&\quad + \mathbb{E}[\tau_i \mid X_i = c', C_i = c] - \mathbb{E}[\tau_i \mid X_i = c', C_i = c'] \\
&= \delta(c, c') + \mathbb{E}[\tau_i \mid X_i = c', C_i = c] - \mathbb{E}[\tau_i \mid X_i = c', C_i = c']
\end{aligned}$$

This shows that $\Delta(c, c')$ can be written as $\delta(c, c')$ plus an additional term that captures the difference in average treatment effects between subpopulations facing cutoffs c and c'.

Although the fact that different RDDs are not directly comparable is not new (see e.g. De la Mata, 2012, footnote 21), to our knowledge it has not been addressed rigorously before. To formalize these ideas, we will define selection into cutoffs in the following way:

Definition 1 (selection into cutoffs) *Let $c, c' \in \mathcal{C}$ and define:*

$$B(x, c, c') := \mathbb{E}[Y_{0i} \mid X_i = x, C_i = c] - \mathbb{E}[Y_{0i} \mid X_i = x, C_i = c']$$

There is selection into cutoffs if $B(x, c, c') \neq 0$ for some $c, c' \in \mathcal{C}$, $c \neq c'$ and for some x in the support of X_i.

The idea of selection into cutoffs is analogous to site selection bias in the context of multiple randomized trials suggested by Allcott (2015). Intuitively, it means that for the same value of the running variable, average potential outcomes under control status differ across cutoffs because sorting generates imbalances in characteristics between units exposed to different thresholds. As a simple example, consider the separable model:

$$Y_{0i} = g(X_i) + u_i$$

where $g(.)$ is some unknown continuous function and u_i includes all individual unobserved characteristics besides the score. For simplicity, suppose u_i is independent of X_i. Then selection into cutoffs means that

$$\mathbb{E}[u_i \mid C_i = c] \neq \mathbb{E}[u_i \mid C_i = c']$$

That is, unobserved individual characteristics differ across units exposed to different cutoffs. This would occur, for example, if more informed or more motivated individuals could self-select into lower thresholds to increase the likelihood of receiving the treatment.

In what follows, we will consider different settings under which variability in cutoffs can be exploited to obtain parameters of interest at different parts of the distribution of the running variable. For the sake of simplicity, we will assume that there are only two cutoffs, $\mathcal{C} = \{c_0, c_1\}$ with $c_0 < c_1$. Additionally, we will consider a sharp design at each cutoff:

Assumption 1 (sharp design) $\lim_{x \searrow c} \mathbb{E}[D_i \mid X_i = x, C_i = c] = 1$ *and* $\lim_{x \nearrow c} \mathbb{E}[D_i \mid X_i = x, C_i = c] = 0$*, for all $c \in \mathcal{C}$.*

We also assume continuity of regression functions at each cutoff, which means that the treatment effect can be identified at each cutoff separately using a single-cutoff RDD:

Assumption 2 (continuity of regression functions) $\mathbb{E}[Y_{0i} \mid X_i = x, C_i = c]$ *and* $\mathbb{E}[Y_{1i} \mid X_i = x, C_i = c]$ *are continuous in x at $x = c$ for all $c \in \mathcal{C}$.*

Assumption 1 and 2 are sufficient to identify $\mathbb{E}[\tau_i \mid X_i = c, C_i = c]$ for all $c \in \mathcal{C}$ (see e.g. Hahn, Todd, and van der Klaauw, 2001). The next two subsections provide conditions under which we can identify a cutoff-free parameter, in the sense that C_i can be dropped from the conditioning, while the last subsection discusses what parameters can be identified when these conditions do not hold.

3.1 Mean-independent cutoffs

The simplest scenario occurs when cutoffs can be seen as randomly assigned conditional on X_i, meaning that there is no selection into cutoffs (except for possible selection through X_i). Formally, assume the following:

Assumption 3 (mean-independent cutoffs)

$$\mathbb{E}[Y_{di} \mid X_i = x, C_i = c_0] = \mathbb{E}[Y_{di} \mid X_i = x, C_i = c_1] = \mathbb{E}[Y_{id} \mid X_i = x]$$

for any $x \in [c_0, c_1]$ *and* $d \in \{0, 1\}$.

This condition can be interpreted intuitively as a randomized experiment stratified on X_i, that is, an experiment that randomly assigns units with the same score to the low or high cutoff. When this assumption holds, the difference between the treatment effects across cutoffs is:

$$\begin{aligned}\Delta(c_0, c_1) &= \mathbb{E}[\tau_i \mid X_i = c_0, C_i = c_0] - \mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_1] \\ &= \mathbb{E}[\tau_i \mid X_i = c_0] - \mathbb{E}[\tau_i \mid X_i = c_1]\end{aligned}$$

which is independent of C_i. Hence, under assumption 3 the selection term is zero and this difference can be interpreted as the change in the average treatment effect when the cutoff is reduced from c_1 to c_0.

Assumption 3 gives a much stronger result because it assumes that mean independence holds for all $x \in [c_0, c_1]$, not only at the endpoints. The usefulness of this condition is that units with the same score but facing different cutoffs can be compared to identify an average treatment effect at any value of the score between the two cutoffs. This intuition is formalized in the following result:

Proposition 1 *Under assumptions 1 and 3, for any* $x_0 \in (c_0, c_1)$*,*

$$\mathbb{E}[\tau_i \mid X_i = x_0] = \mathbb{E}[Y_i \mid X_i = x_0, C_i = c_0] - \mathbb{E}[Y_i \mid X_i = x_0, C_i = c_1]$$

The proof of all the propositions in this paper can be found in the appendix. Note that in fact this result does not exploit in any sense the RDD, so continuity of the regression functions does not play any role whenever the endpoints c_0 and c_1 are omitted. If assumption 2 holds as well, it follows that the treatment effect $\mathbb{E}[\tau_i \mid X = x]$ is identified over the whole closed interval $[c_0, c_1]$: the effect in the open interval (c_0, c_1) comes from proposition 1, the boundary points from a single-cutoff RDD at each cutoff.

3.2 Selection on observables

In most relevant applications, the assumption of mean-independent cutoffs may be overly strong. In the previous anti-poverty program example, assumption 3 would require that all characteristics are balanced, at least on average, across regions. One straightforward way to relax this condition is to assume that the researcher can control for the variables that are unbalanced across groups. In this case, mean independence of cutoffs can be recovered by conditioning on a set of covariates:

Assumption 4 (selection on observables) *There exists a vector of observed covariates Z_i such that*

$$\mathbb{E}[Y_{di} \mid X_i = c, C_i = c_0, Z_i] = \mathbb{E}[Y_{di} \mid X_i = c, C_i = c_1, Z_i] = \mathbb{E}[Y_{di} \mid X_i = c, Z_i]$$

for $c \in \{c_0, c_1\}$ and $d \in \{0, 1\}$.
Moreover, for each $c \in \{c_0, c_1\}$ there is a neighborhood $(c - h_c, c + h_c)$ in which the conditional density of X_i given C_i and Z_i is positive, i.e.

$$f_{X|C,Z}(x|c, z) > 0$$

for all $x \in (c - h_c, c + h_c)$ and for all z in the support of Z_i.

The first part of assumption 4 is the usual conditional independence assumption (see e.g. Imbens, 2004). The requirement that the density be positive can be seen as an overlap condition imposing that, for each possible value of Z_i, there are observations at both sides of the cutoff. In practice, it can be the case that overlap fails for some values of the covariates. For instance, suppose Z_i is a gender indicator. If for some subpopulation facing a particular cutoff all women are above the cutoff and all men are below, clearly the effect conditional on Z_i cannot be identified at that cutoff. When overlap fails, it may be possible in some cases to overcome the problem by focusing on the part of the support where overlap holds, although this changes the estimand. See Crump, Hotz, Imbens, and Mitnik (2009) for a detailed discussion of overlap issues.

Because the idea under selection on observables is to condition on Z_i, assumption 2 needs to be strengthened to hold conditionally on Z_i. We restate the assumption as follows:

Assumption 2' (continuity of regression functions) $\mathbb{E}[Y_{0i} \mid X_i = x, C_i = c, Z_i = z]$ *and* $\mathbb{E}[Y_{1i} \mid X_i = x, C_i = c, Z_i = z]$ *are continuous in* x *at* $x = c$ *for all* $c \in \mathcal{C}$ *and for all* z *in the support of* Z_i.

Under selection on observables, the following results holds:

Proposition 2 *Under assumptions 1, 2' and 4, for* $c \in \{c_0, c_1\}$*:*

$$\mathbb{E}[\tau_i \mid X_i = c, Z_i = z] = \lim_{x \searrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i = z] - \lim_{x \nearrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i = z]$$

for any z *in the support of* Z_i*. Moreover,*

$$\mathbb{E}[\tau_i \mid X_i = c] = \mathbb{E}\left\{ \lim_{x \searrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i] - \lim_{x \nearrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i] \middle| X_i = c \right\}$$

where the outer expectation is with respect to the density of Z_i *conditional on* $X_i = c$.

This result shows that when selection into cutoffs is based on observable characteristics, it is possible to recover a cutoff-free parameter by conditioning on covariates and then averaging with respect to the appropriate distribution. As in the previous subsection, assumption 4 can be strengthened to hold in $[c_0, c_1]$ in order to identify the treatment effect over the whole closed interval.

Interestingly, the selection-on-observables assumption has testable implications that can provide empirical support for its validity. Under assumption 4,

$$\begin{aligned} \mathbb{E}[Y_{0i} \mid X_i = c_0] &= \mathbb{E}\{\mathbb{E}[Y_{0i} \mid X_i = c_0, Z_i] \mid X_i = c_0\} \\ &= \mathbb{E}\{\mathbb{E}[Y_{0i} \mid X_i = c_0, C_i = c_1, Z_i] \mid X_i = c_0\} \\ &= \mathbb{E}\{\mathbb{E}[Y_i \mid X_i = c_0, C_i = c_1, Z_i] \mid X_i = c_0\} \end{aligned}$$

and thus if selection on observables holds, it must be true that:

$$\mathbb{E}\{\mathbb{E}[Y_i \mid X_i = c_0, C_i = c_1, Z_i] \mid X_i = c_0\} = \mathbb{E}\left\{ \lim_{x \nearrow c_0} \mathbb{E}[Y_i \mid X_i = x, C_i = c_0, Z_i] \middle| X_i = c_0 \right\} \quad (7)$$

Intuitively, since under assumption 4 the cutoffs in the conditional expectations are interchangeable, there are two alternative control groups: units facing the low cutoff c_0 with score slightly below c_0, and units facing the high cutoff c_1 with scores around c_0 and the same values of Z_i. This fact suggests a way to assess the validity of the selection-on-observables assumption by simply testing the equality in equation 7. The idea of comparing control groups was suggested by Hotz, Imbens, and Mortimer (2005) in the context of multiple randomized trials.

Regarding implementation, the results from proposition 2 require the nonparametric estimation of several regression functions that depend on the running variable and the covariates. These magnitudes can also be written using inverse probability weighting methods. This idea is analogous to the approach in Abadie (2005) for the case of semiparametric difference-in-differences estimators. Start by defining the propensity score as:

$$p_0(x, z) := \mathbb{P}(C_i = c_0 \mid X_i = x, Z_i = z) \tag{8}$$

Note that the second part of assumption 4 can be seen as a common support condition for the propensity score around the cutoffs, since by Bayes' rule it requires $p_0(x, z)$ to be non-zero at least in a neighborhood of the cutoffs. This propensity score will be assumed to be continuous:

Assumption 5 (continuity of the propensity score) *$p_0(x, z)$ is continuous in x at $x = c$ for $c \in \{c_0, c_1\}$ for all z in the support of Z_i and $c \in \{c_0, c_1\}$.*

Assumption 5 can be interpreted as a McCrary test (McCrary, 2008) that ensures that the probability of facing the low cutoff is does not jump at $X_i = c_0$ for all values of the covariates. Although it is not technically a necessary condition for identification, a jump in the probability of facing the low cutoff could be taken as evidence of selection beyond observable variables, and the validity of the strategy would be hard to defend in this context.

Additionally, let S_i indicate whether unit i faces the low cutoff, $S_i = \mathbf{1}(C_i = c_0)$. The following result gives the reweighted version of the estimand of interest.

Proposition 3 *Under assumptions 1, 2', 4 and 5,*

$$\mathbb{E}[\tau_i \mid X_i = c_0] = \lim_{x \searrow c_0} \mathbb{E}\left[\frac{Y_i S_i}{p_0(c_0, Z_i)} \middle| X_i = x\right] - \lim_{x \nearrow c_0} \mathbb{E}\left[\frac{Y_i S_i}{p_0(c_0, Z_i)} \middle| X_i = x\right]$$

and

$$\mathbb{E}[\tau_i \mid X_i = c_1] = \lim_{x \searrow c_1} \mathbb{E}\left[\frac{Y_i(1-S_i)}{1-p_0(c_1, Z_i)}\middle| X_i = x\right] - \lim_{x \nearrow c_1} \mathbb{E}\left[\frac{Y_i(1-S_i)}{1-p_0(c_1, Z_i)}\middle| X_i = x\right]$$

This result shows that under selection on observables, the treatment effect at each cutoff can be recovered by using two single-cutoff RDDs after the observations have been reweighted by the inverse of the selection probability. In practice, the propensity score is generally unknown and needs to be estimated in a previous step.

Finally, the reweighted version of equation 7 is:

$$\mathbb{E}\left[\frac{Y_i(1-S_i)}{1-p_0(c_1, Z_i)}\middle| X_i = c_0\right] = \lim_{x \nearrow c_0} \mathbb{E}\left[\frac{Y_i S_i}{p_0(c_0, Z_i)}\middle| X_i = x\right] \tag{9}$$

3.3 Extrapolating across cutoffs

Although assumption 4 relaxes the condition of randomized cutoffs, it is still a strong assumption as it requires the researcher to observe all relevant variables driving the differences between subpopulations facing different cutoffs. Access to this information may be unlikely in many applications. When the selection-on-observables assumption is not credible, it may not be possible to identify cutoff-free parameters as in the previous section. As discussed above, in this scenario cross-cutoff comparisons of treatment effects may not be valid. However, under some conditions it may be possible to exploit information from different cutoffs to extrapolate the treatment effect to different values of the running variable. Intuitively, the idea is to use a "diff-in-diff-like" strategy to eliminate the selection bias. This strategy will rely on the fact that the selection bias term in definition 1 is observed for any value of the score below the lowest cutoff.

To simplify the notation, since there are only two cutoffs, write $B(x) := B(x, c_0, c_1)$. Then, for any $x < c_0$, we have that $\mathbb{E}[Y_i \mid X_i = x, C_i = c] = \mathbb{E}[Y_{0i} \mid X_i = x, C_i = c]$ for $c \in \{c_0, c_1\}$. In particular, the selection term at the low cutoff can be written as:

$$B(c_0) = \lim_{x \nearrow c_0} \mathbb{E}[Y_i \mid X_i = x, C_i = c_0] - \mathbb{E}[Y_i \mid X_i = c_0, C_i = c_1] \tag{10}$$

The parameter of interest in this section will be $\mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0]$, that is, the average treatment effect for the subpopulation facing cutoff c_0 but with the running variable evaluated at c_1. Conceptually, this magnitude captures the treatment effect that the group facing the low cutoff would have experienced had it been exposed to the high cutoff instead. For instance, in the previous anti-poverty program example, suppose that region A is exposed to cutoff c_0 and region B is exposed to cutoff c_1. Then, $\mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0]$ is the average effect of the program for units from region A, had they been exposed to cutoff c_1. In general, this parameter is unobservable since an RDD only identifies the effect at the cutoff to which each group was actually exposed.

To see how to recover this parameter, note that:

$$\begin{aligned}\mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0] &= \mathbb{E}[Y_{1i} \mid X_i = c_1, C_i = c_0] - \mathbb{E}[Y_{0i} \mid X_i = c_1, C_i = c_0] \\ &= \mathbb{E}[Y_{1i} \mid X_i = c_1, C_i = c_0] - \mathbb{E}[Y_{0i} \mid X_i = c_1, C_i = c_1] \\ &\quad + \mathbb{E}[Y_{0i} \mid X_i = c_1, C_i = c_1] - \mathbb{E}[Y_{0i} \mid X_i = c_1, C_i = c_0] \\ &= \mathbb{E}[Y_{1i} \mid X_i = c_1, C_i = c_0] - \mathbb{E}[Y_{0i} \mid X_i = c_1, C_i = c_1] - B(c_1)\end{aligned}$$

Now, the first two terms on the right-hand side can be expressed in terms of observables, although $B(c_1)$ is unobservable. However, if we assume that the bias at the high cutoff is equal to the bias in the low cutoff, which can be obtained as described above, then $\mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0]$ is identified. The following assumption states this condition:

Assumption 6 (constant bias across cutoffs) *The selection terms are equal across cutoffs, i.e.,*

$$B(c_0) = B(c_1)$$

The main result of this section can be summarized as follows:

Proposition 4 *Under assumptions 1, 2 and 6, $B(c_0)$ is identified as in equation 10 and*

$$\mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0] = \mathbb{E}[Y_i \mid X_i = c_1, C_i = c_0] - \lim_{x \nearrow c_1} \mathbb{E}[Y_i \mid X_i = x, C_i = c_1] - B(c_0)$$

The result from proposition 4 is depicted graphically in Figure 1. The picture shows the four regression functions $m_d(x, c) := \mathbb{E}[Y_{di} \mid X_i = x, C_i = c]$ with $d \in \{0, 1\}$ and $c \in \{c_0, c_1\}$ for a hypothetical case with $c_0 = 4$ and $c_1 = 7$. The dotted lines represent the part of each regression function that is unobserved, and the dashed and solid lines are observed. The parameter of interest is $\tau(c_1, c_0) := \mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0]$. The idea is to replace the unobserved term $B(c_1)$ with the observed term $B(c_0)$ and add this to $\mathbb{E}[Y_{0i} \mid X_i = c_1, C_i = c_1]$ to obtain $\mathbb{E}[Y_{0i} \mid X_i = c_1, C_i = c_0]$.

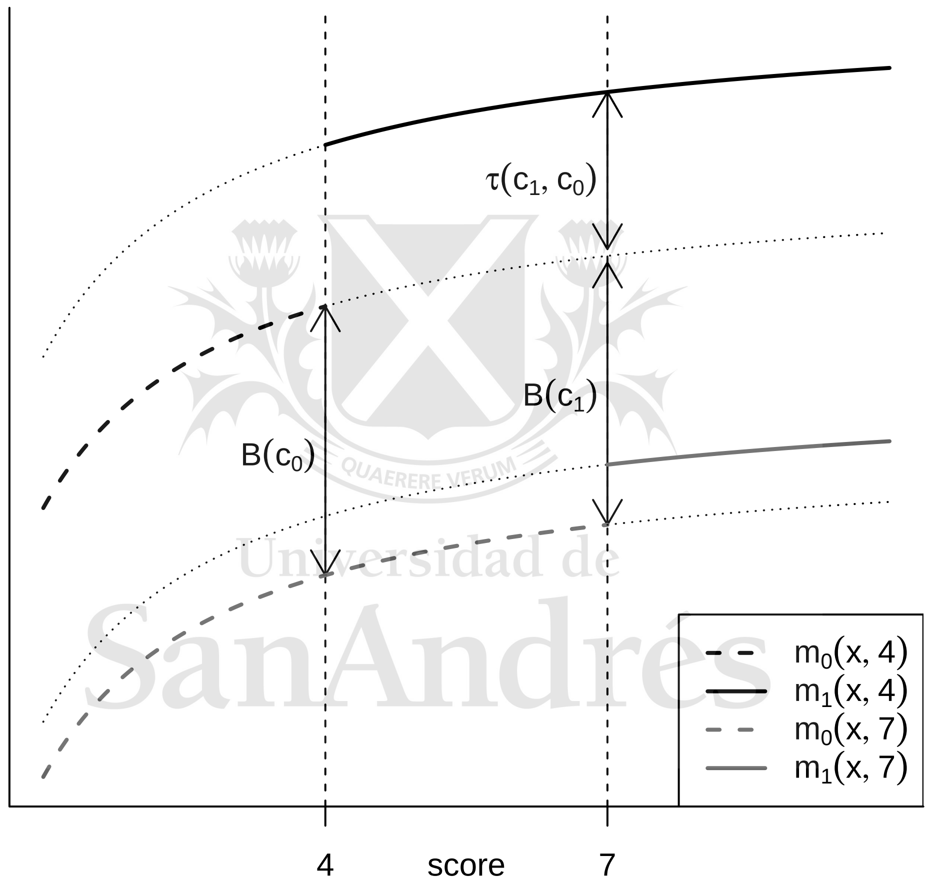

Figure 1: extrapolating across cutoffs.

This approach is similar in spirit to the "difference-in-discontinuities" proposed by Grembi, Nannicini, and Troiano (2016). Their strategy consists on using the difference between RDD estimates at different points time periods to correct for discontinuities in the conditional expectation functions.

Clearly, assumption 6 is an identification assumption and hence it is not testable. However, some evidence can be provided to reinforce its validity in a similar way in which it is done for difference-in-difference models. A sufficient (but not necessary) condition for assumption 6 is that the regression functions $\mathbb{E}[Y_{0i} \mid X_i = x, C = c_0]$ and $\mathbb{E}[Y_{0i} \mid X_i = x, C = c_1]$ are parallel over the whole range of the running variable. Now, as noted previously, both regression functions are estimable for any $x \leq c_0$. If these two regression functions are parallel for $x \leq c_0$, the assumption that they will still be parallel at $x = c_1$ seems more likely to hold. Thus, a simple graphical strategy to support assumption 6 is to estimate the regression functions for units facing cutoffs c_0 and c_1 separately and plot them together to check whether they are parallel.

4 Empirical illustration

To illustrate the ideas presented in this paper, we will use data from the Oportunidades program (formerly known as Progresa), a large-scale poverty alleviation program in Mexico. This program consists of a series of interventions aiming at improving health, nutrition and schooling outcomes in poor households through conditional cash transfers. Because assignment was randomized at the locality level, the impacts of Oportunidades has been studied extensively (see the supplemental appendix in Calonico, Cattaneo, and Titiunik, 2014b, and references thereof). The key feature that makes this program interesting to illustrate our methods is the fact that household eligibility was based on whether a poverty index exceeded a certain cutoff, and the cutoff was different across regions.[1] As explained in Buddelmeyer and Skoufias (2004) and Calonico, Cattaneo, and Titiunik (2014b), this design leads to multiple sharp RDDs that identify an intention-to-treat parameter. The cutoffs and sample sizes are shown in table 1.

The empirical illustration in this paper will initially consider the four rural regions with largest sample sizes (regions 3, 4, 5 and 6). All the empirical analysis will exclude observations with score outside the range (600,900). Cutoffs for these regions are 759.4, 753, 751.5 and 751, respectively. The outcome of interest will be food consumption in $t + 1$, that is, one year after the program started.

[1] The program started in 1998 targeting rural regions and was expanded to urban areas in 2003. See Calonico, Cattaneo, and Titiunik (2014b) and references thereof for more details. This application will only consider rural areas.

Table 1: Cutoffs and sample sizes

	Cutoff	Obs.	Treated	Control
Region 3	759.4	933	618	315
Region 4	753	1189	810	379
Region 5	751.5	3116	2003	1113
Region 6	751	541	441	100
Region 12	569	78	40	38
Region 27	691	828	614	214
Region 28	853.3	175	157	18

4.1 Pooled RDD

We start the empirical analysis by considering the pooling approach explained in section 2 and analyzed more detailedly by Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016). We pool the four regions under study by centering the running variable and estimate a single parameter. Table 2 shows the RD estimates[2] for the pooled sample and for each region separately, together with the weights assigned by the pooled estimate to each individual estimate (see equation 2).

Table 2: Pooled and separate RD estimates

	Estimate	s.e.	Obs.	Weights
Pooled	24.66	7.66	734	
Region 3	31.64	12.91	119	.16
Region 4	15.12	17.05	270	.21
Region 5	24.42	9.54	474	.52
Region 6	27.98	24.65	63	.11

The results reveal substantial heterogeneity in the estimates for each region, with statistically significant effects at the 5 percent level in regions 3 and 5. In terms of weights, region 5 accounts for over half of the pooled estimate. To see what is driving these weights, we can decompose them as in equation 2 into two factors: the proportion of the sample at each cutoff and the conditional density of the re-centered score at 0. Figure 2 shows the kernel density estimate of the centered running variable for each region. On the other hand, Table 3 shows the number and proportion of observations at each cutoff. The difference in weights seems to be driven mainly by the different sample sizes at each cutoff, as the conditional densities have similar values around the cutoffs whereas the proportion of the sample at region 5 is notably larger.

[2]See the appendix for further details about implementation.

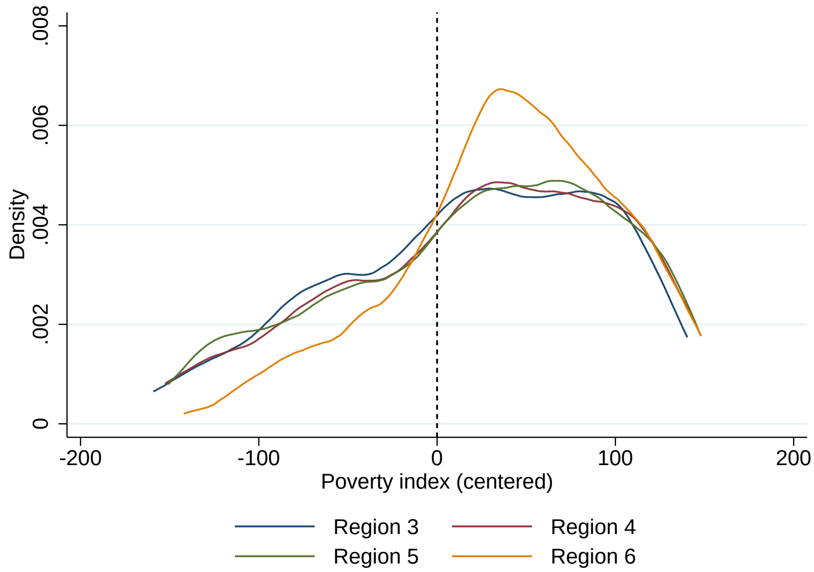

Figure 2: density of the re-centered score.

Table 3: Sample sizes

	Obs.	Prop.
Region 3	607	.15
Region 4	850	.21
Region 5	2169	.53
Region 6	448	.11
Total	4074	1

As explained in the paper, the different estimates are not directly comparable since regions may differ in both observable and unobservable characteristics. Table 4 shows the mean of a list of covariates for each region, together with the p-value from an F-test of the regression of each covariate against region indicators (omitting region 3). The table reveals significant differences in most observed covariates. This issue will be addressed in the next subsections.

4.2 Selection on observables

To account for selection into cutoffs based on observable variables, we implement the reweighted version of the estimand in proposition 3. In what follows, we restrict the analysis to regions 3 and 5 (see the appendix for a discussion on the choice of regions). In this case the low cutoff corresponds to region 5 ($C = 751.5$) and the high cutoff to region 3 ($C = 759.4$). We start by estimating the propensity score:

$$p_0(x, z) = \mathbb{P}[C_i = c_0 \mid X_i = x, Z_i = z]$$

Table 4: Covariate means by region

	Region 3	Region 4	Region 5	Region 6	p-value
Head's age	43	43.25	44.77	43.59	0
Head is male	.92	.93	.94	.93	.04
Head's education	2.58	3.01	3.43	2.63	0
Head is empployed	.91	.92	.91	.93	.335
Children 0 to 5	1.17	1.09	1.03	1.43	0
HH size	6.27	6.19	6.25	6.63	0
Owns the house	.96	.93	.96	.96	.001
Number of rooms	1.76	1.75	1.96	1.49	0
HH has water	.52	.45	.32	.36	0
HH has toilet	.6	.66	.7	.25	0
HH has cement floor	.39	.38	.42	.26	0
HH has electricity	.67	.7	.69	.77	0

using a flexible logit specification and the covariates in table 4, including various squares, cubes and interactions. The results for the propensity score estimation are shown in table 10 in the appendix. Using the predicted probabilities from this model, we estimate the weights:

$$\hat{\theta}_{0i} = \frac{S_i}{\hat{p}_0(X_i, Z_i)} \quad \text{and} \quad \hat{\theta}_{1i} = \frac{1 - S_i}{1 - \hat{p}_0(X_i, Z_i)}$$

The estimates are obtained as differences of simple averages of the reweighted outcome on a window around the cutoff. The window length is set to 23, which is the minimum (rounded to the nearest integer) between the bandwidths chosen for regions 3 and 5 by the method explained in Calonico, Cattaneo, and Titiunik (2014b). The standard errors are bootstrapped to account for the variation of the estimated propensity score. Table 5 displays the results. As indicated in the third column, the estimated parameters are $\hat{\mathbb{E}}[\tau_i \mid X_i = c_0] = 23.86$ and $\hat{\mathbb{E}}[\tau_i \mid X_i = c_1] = 41.44$, both statistically significant at the 5 percent level.

Table 5: Estimated means

	Estimate	s.e.	Difference	s.e.
$\mathbb{E}[\theta_{0i} Y_i \mid X_i = c_0^+]$	122.11	9.44		
$\mathbb{E}[\theta_{0i} Y_i \mid X_i = c_0^-]$	98.37	4.75	23.86	10.79
$\mathbb{E}[\theta_{1i} Y_i \mid X_i = c_0^+]$	154.11	18.94		
$\mathbb{E}[\theta_{1i} Y_i \mid X_i = c_0^-]$	111.61	6.86	41.44	19.88

Equation 7 provides a specification test for the selection-on-observables assumption. The estimated means and their difference are shown in table 6. It is clear from the table that selection on observables does not hold in this context, at least for the set of covariates used in the estimation of the propensity score. In this case, the estimates shown in table 5 may not be valid.

Table 6: Specification test

	Estimate	Difference	s.e.
$\mathbb{E}[\theta_{1i} Y_i \mid X_i = c_0]$	143.19		
$\mathbb{E}[\theta_{0i} Y_i \mid X_i = c_0^-]$	98.37	44.83	3.74

4.3 Extrapolating across cutoffs

Figure 3 displays the estimated regression functions for regions 3 and 5 above and below their respective cutoffs over the range $(700, 800)$ of the score (that is, approximately 50 points below and above the cutoffs). The regressions are estimated using a polynomial of degree 4 separately above and below the cutoff for each region. The graph suggests positive effects on food consumption for both regions separately, an observation that is confirmed by table 2.

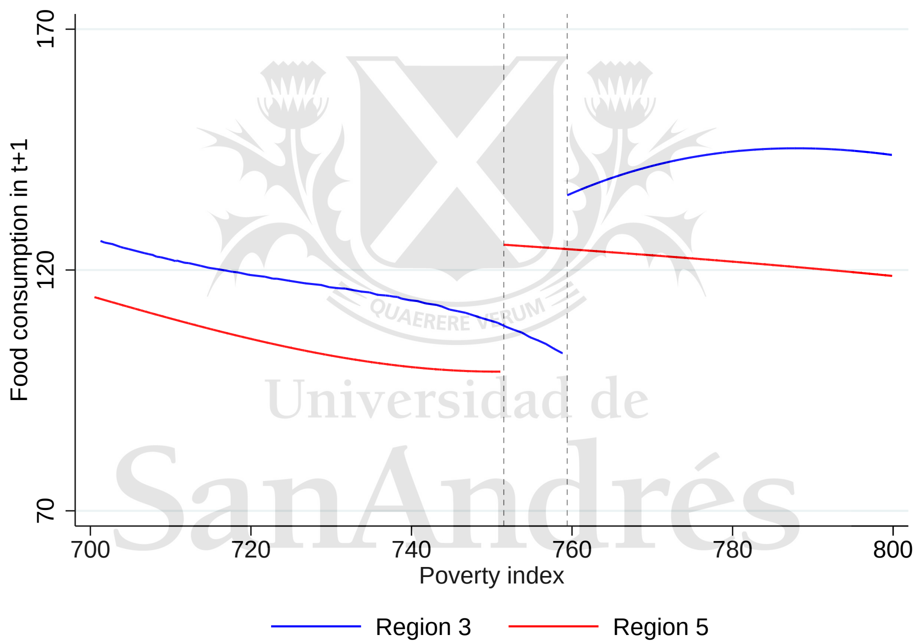

Figure 3: regression functions above and below cutoffs. Global polynomial of degree 4.

The parameter of interest in this section is $\mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0]$. Under the assumptions stated in the paper, after some rearranging this expression can be written as:

$$\mathbb{E}\tau_i \mid X_i = c_1, C_i = c_0] = \mathbb{E}[Y_i \mid X_i = c_1, C_i = c_0,] - \lim_{x \to c_1^-} \mathbb{E}[Y_i \mid X_i = x, C_i = c_1] \\ + \mathbb{E}[Y_i \mid X_i = c_0, C_i = c_1] - \lim_{x \to c_0^-} \mathbb{E}[Y_i \mid X_i = x, C_i = c_0]$$

which can be estimated as a linear combination of four means. The same window length of 23 is used in the calculations. The results are shown in table 7.

Table 7: Estimated means

	Estimate	s.e.	Difference	s.e.
$\mathbb{E}[Y_i \mid X_i = c_1, C_i = c_0]$	124.04	6.37		
$\mathbb{E}[Y_i \mid X_i = c_1^-, C_i = c_0]$	110.00	17.38	14.04	18.51
$\mathbb{E}[Y_i \mid X_i = c_0, C_i = c_1]$	139.58	7.48		
$\mathbb{E}[Y_i \mid X_i = c_0^-, C_i = c_0]$	97.22	9.24	42.36	11.89
$\mathbb{E}[\tau_i \mid X_i = c_1, C_i = c_0]$	56.40	22.00		

The results show that the estimated treatment effect at the high cutoff for the low-cutoff region is 56.40, significantly different from zero at the 5 percent level.

As suggested in section 3.3, a possible specification check is to test whether the observed outcomes of each group are parallel below the lowest cutoff. One simple parametric way to test this is to run the regression:

$$Y_i = \beta_0 + \beta_1(X_i - c_0) + \beta_2\mathbf{1}(C_i = c_0) + \beta_3(X_i - c_0) * \mathbf{1}(C_i = c_0) + u_i$$

using only observations below c_0 (and possibly in a small window) and test $\beta_3 = 0$. We run this regression using observations with $700 < score < 751.5$.[3] The results of this specification check are shown in table 8.

Table 8: Specification check

	Food consumption in t+1
Score	-0.594
	(0.485)
Low cutoff	-5.087
	(13.33)
Interaction	0.261
	(0.538)
Constant	100.8***
	(11.67)
Observations	412

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The table shows that the coefficient corresponding to the interaction is not significant, suggesting that the outcomes follow parallel trends, although this can be due to lack of power. Nevertheless, the results are consistent with the visual evidence provided in figure 3.

[3] The range of the data is restricted because trends need not be parallel over the whole range but just on a small window below the cutoff. However, the regression using observations on the range (600,751.5) yields the same qualitative results, with the p-value for the interaction term above 0.9.

5 Concluding remarks

Although the theoretical and methodological literature on RDD focuses almost exclusively on designs with a single cutoff, empirical researchers are increasingly facing applications where different units face different thresholds. This paper extends the results in Cattaneo, Keele, Titiunik, and Vazquez-Bare (2016) and suggests ways to exploit the variability in cutoffs in order to identify parameters that are unidentifiable when only one cutoff is available. We introduce the idea of selection into cutoffs, which emphasizes the possibility that units facing different thresholds differ in observable and unobservable characteristics. Two scenarios are considered. First, one in which selection into cutoffs is based on observable variables. In this case, it is possible to identify overall cutoff-free treatment effects by properly reweighting observations based on a propensity score. Second, when selection potentially depends on unobservable variables, we suggest a "diff-in-diff-like" strategy to extrapolate the treatment effect of a subpopulation outside the cutoff it faces.

Future work will extend these results in many dimensions. First, identification results will be extended to the case of an arbitrary number of cutoffs, and possibly continuous cutoffs, which occurs for example when using vote shares in party elections and the winning cutoffs depend on the vote share of the third contestant. The findings will also be adapted to fuzzy designs. Finally, estimation and inference in multi-cutoff RDDs raise many issues such as the choice of tuning parameters (like the bandwidth) that need to be addressed more carefully in future research.

References

ABADIE, A. (2005): "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72(1), 1–19.

ALLCOTT, H. (2015): " Site Selection Bias in Program Evaluation," *The Quarterly Journal of Economics*, 130(3), 1117–1165.

ANGRIST, J. D., AND V. LAVY (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 114(2), 533–575.

BENAVENTE, J. M., G. CRESPI, L. FIGAL GARONE, AND A. MAFFIOLI (2012): "The impact of national research funds: A regression discontinuity approach to the Chilean FONDECYT," *Research Policy*, 41(8), 1461–1475.

BLACK, D. A., J. GALDO, AND J. A. SMITH (2007): "Evaluating the Worker Profiling and Reemployment Services System Using a Regression Discontinuity Approach," *American Economic Review*, 97(2), 104–107.

BROLLO, F., T. NANNICINI, R. PEROTTI, AND G. TABELLINI (2013): "The Political Resource Curse," *American Economic Review*, 103(5), 1759–96.

BUDDELMEYER, H., AND E. SKOUFIAS (2004): "An evaluation of the performance of regression discontinuity design on PROGRESA," *The World Bank Policy Research Working Paper Series*, (3386).

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014a): "Robust data-driven inference in the regression-discontinuity design," *Stata Journal*, 14(4), 909–946.

——— (2014b): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

CATTANEO, M. D., B. FRANDSEN, AND R. TITIUNIK (2015): "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate," *Journal of Causal Inference*, 3(1), 1–24.

CATTANEO, M. D., L. KEELE, R. TITIUNIK, AND G. VAZQUEZ-BARE (2016): "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *Journal of Politics*, 78(4), 1229–1248.

CERQUA, A., AND G. PELLEGRINI (2014): "Do subsidies to private capital boost firms' growth? A multiple regression discontinuity design approach," *Journal of Public Economics*, 109, 114–126.

CRUMP, R. K., V. J. HOTZ, G. W. IMBENS, AND O. A. MITNIK (2009): “Dealing with limited overlap in estimation of average treatment effects,” *Biometrika*, 96(1), 187–199.

DE LA MATA, D. (2012): “The Effect Of Medicaid Eligibility On Coverage, Utilization, And Children’S Health,” *Health Economics*, 21(9), 1061–1079.

GREMBI, V., T. NANNICINI, AND U. TROIANO (2016): “Do Fiscal Rules Matter?,” *American Economic Journal: Applied Economics*, 8(3), 1–30.

HAHN, J., P. TODD, AND W. VAN DER KLAAUW (2001): “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 69(1), 201–209.

HOTZ, J. V., G. W. IMBENS, AND J. H. MORTIMER (2005): “Predicting the efficacy of future training programs using past experiences at other locations,” *Journal of Econometrics*, 125(1-2), 241–270.

IMBENS, G. (2004): “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review,” *Review of Economics and Statistics*, 86(1), 4–29.

IMBENS, G., AND T. ZAJONC (2011): “Regression discontinuity design with multiple forcing variables,” *Working paper.*

KEELE, L. J., AND R. TITIUNIK (2015): “Geographic Boundaries as Regression Discontinuities,” *Political Analysis*, 23(1), 127–155.

LEE, D. S. (2008): “Randomized Experiments from Non-random Selection in U.S. House Elections,” *Journal of Econometrics*, 142(2), 675–697.

MCCRARY, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142(2), 698–714.

PAPAY, J. P., J. B. WILLETT, AND R. J. MURNANE (2011): “Extending the regression-discontinuity approach to multiple assignment variables,” *Journal of Econometrics*, 161(2), 203–207.

THISTLETHWAITE, D. L., AND D. T. CAMPBELL (1960): “Regression-discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment,” *Journal of Educational Psychology*, 51(6), 309–317.

WONG, V. C., P. M. STEINER, AND T. D. COOK (2013): “Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables A Comparative Study of Four Estimation Methods,” *Journal of Educational and Behavioral Statistics*, 38(2), 107–141.

Appendices

A Proofs

A.1 Proposition 1

By assumption 1, if $x_0 \in (c_0, c_1)$, $D_i(x_0, c_0) = 1$ and $D_i(x_0, c_1) = 0$. Then, using assumption 3,

$$\begin{aligned}\mathbb{E}[Y_{1i} - Y_{0i} \mid X_i = x_0] &= \mathbb{E}[Y_{1i} \mid X_i = x_0, C_i = c_0] - \mathbb{E}[Y_{0i} \mid X_i = x_0, C_i = c_1] \\ &= \mathbb{E}[Y_i \mid X_i = x_0, C_i = c_0] - \mathbb{E}[Y_i \mid X_i = x_0, C_i = c_1] \quad \square\end{aligned}$$

A.2 Proposition 2

By assumptions 2' and 4,

$$\begin{aligned}\mathbb{E}[Y_{1i} \mid X_i = c, Z_i = z] &= \mathbb{E}[Y_{1i} \mid X_i = c, C_i = c, Z_i] \\ &= \lim_{x \searrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i]\end{aligned}$$

and similarly,

$$\mathbb{E}[Y_{0i} \mid X_i = c, Z_i = z] = \lim_{x \nearrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i]$$

from which

$$\mathbb{E}[\tau_i \mid X_i = c, Z_i = z] = \lim_{x \searrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i = z] - \lim_{x \nearrow c} \mathbb{E}[Y_i \mid X_i = x, C_i = c, Z_i = z]$$

and the remaining result follows from $\mathbb{E}[\tau_i \mid X_i = c] = \mathbb{E}\{\mathbb{E}[\tau_i \mid X_i = c, Z_i] \mid X_i = c\}$. $\square$

A.3 Proposition 3

By the law of iterated expectations,

$$\mathbb{E}[Y_i S_i \mid X_i, Z_i] = \mathbb{E}\left[Y_i \mid X_i, Z_i, C_i = c_0\right] p_0(X_i, Z_i)$$

from which

$$\mathbb{E}\left[\frac{Y_i S_i}{p_0(X_i, Z_i)} \middle| X_i, Z_i\right] = \mathbb{E}[Y_i \mid X_i, Z_i, C_i = c_0]$$

Hence, for $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{E}[Y_{1i} \mid X_i = c_o + \varepsilon] &= \mathbb{E}\{\mathbb{E}[Y_{1i} \mid X_i = c_o + \varepsilon, Z_i] \mid X_i = c_0 + \varepsilon\} \\
&= \mathbb{E}\{\mathbb{E}[Y_{1i} \mid X_i = c_o + \varepsilon, Z_i, C_i = c_0] \mid X_i = c_0 + \varepsilon\} \\
&= \mathbb{E}\{\mathbb{E}[Y_i \mid X_i = c_o + \varepsilon, Z_i, C_i = c_0] \mid X_i = c_0 + \varepsilon\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\frac{Y_i S_i}{p_0(c_0 + \varepsilon, Z_i)}\middle| X_i = c_0 + \varepsilon, Z_i\right]\middle| X_i = c_0 + \varepsilon\right\} \\
&= \mathbb{E}\left[\frac{Y_i S_i}{p_0(c_0 + \varepsilon, Z_i)}\middle| X_i = c_0 + \varepsilon\right]
\end{aligned}$$

and by continuity,

$$\mathbb{E}[Y_{1i} \mid X_i = c_0] = \lim_{x \searrow c_0} \mathbb{E}\left[\frac{Y_i S_i}{p_0(c_0, Z_i)}\middle| X_i = x\right]$$

A similar argument provides the remaining equalities. □

A.4 Proposition 4

This follows immediately from calculations done in the paper under the stated assumptions. □

B Implementation issues

B.1 Pooled RDD estimates

Estimation and window selection is performed using the `rdrobust` Stata command (Calonico, Cattaneo, and Titiunik, 2014a). The estimation is performed using local constant regression. Table 9 shows the bandwidth and sample sizes around the cutoff used for estimation.

Table 9: Bandwidth and sample size

	Bandwidth	Treated	Control
Pooled	22.7	471	263
Region 3	22.91	76	43
Region 4	40.4	174	96
Region 5	28.3	291	183
Region 6	17.75	49	14

To calculate the weights shown in table 2, we use that under the assumption stated in the paper,

$$\omega(c) = \frac{f_{X|C}(c|c)\mathbb{P}[C = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C = c]} = \mathbb{P}[C = c \mid \tilde{X} = 0] \tag{11}$$

The rightmost expression is easier to estimate since it is simply the probability of facing cutoff c for units with the re-centered score at 0. Thus, we calculate the proportion of the sample at each cutoff on a window of the re-centered score around 0 where the bandwidth is the one chosen by the `rdrobust` command for the pooled estimate.

B.2 Choice of regions for empirical analysis

For the ease of exposition, and because all the results in the paper are derived for the simple case of two cutoffs, we will restrict the empirical analysis to only two regions. To select these two regions, we will look at the estimated regression functions. The regression functions are estimated using global polynomial regression of degrees 3 to 5 and cubic spline using the `mkspline` command. The results are depicted in figure 4.

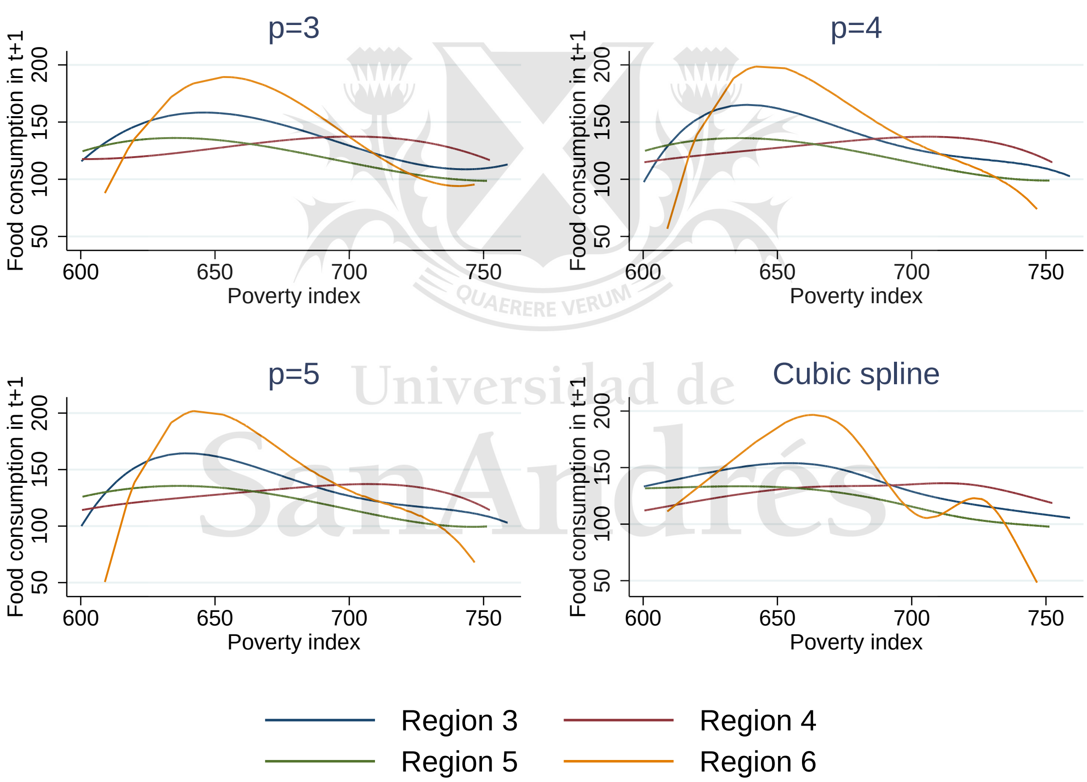

Figure 4: regression functions below the cutoff. Global polynomials and cubic spline.

From figure 4, the regression functions for regions 3 and 5 look fairly parallel in all four specifications. This suggests to use these two regions as it reinforces the validity the extrapolation estimate, as explained in section 3.3.

B.3 Estimation of the propensity score

The propensity score is estimated using a logit model with the covariates in table 4 and includes square, cubes and interactions terms. The results are shown in table 10.

Table 10: Probability of facing low cutoff

	Low cutoff
Score	-0.000605
	(0.000819)
Head's age	0.0143
	(0.102)
Head is male	0.619***
	(0.222)
Head's education	0.401***
	(0.139)
Head is employed	-0.625***
	(0.220)
Head's age squared	0.00145
	(0.00207)
Head's age cubed	-0.0000164
	(0.0000135)
Head's education squared	0.0135
	(0.0159)
Head's education cubed	-0.000819
	(0.000680)
Head's education x age	-0.00697***
	(0.00207)
Children 0 to 5 = 1	0.105
	(0.502)
Children 0 to 5 = 2	0.171
	(0.500)
Children 0 to 5 = 3	-0.00156
	(0.500)
Children 0 to 5 = 4	0.0479
	(0.522)
HH size = 4	0.153
	(0.172)
HH size = 5	0.189

	(0.160)
HH size = 6	0.331**
	(0.161)
HH size = 7	0.275
	(0.176)
Owns the house	0.127
	(0.252)
Has one room	-0.375**
	(0.151)
Has two rooms	0.117
	(0.144)
Has water	-1.317***
	(0.104)
Has toilet	0.458***
	(0.104)
Has cement floor	-0.0481
	(0.108)
Has electricity	-0.434***
	(0.127)
Constant	-0.267
	(1.912)
Observations	2755

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$