Identification and Inference in Regression Discontinuity Designs with a Manipulated Running Variable[*]

Francois Gerard, Miikka Rokkanen, and Christoph Rothe[†]

Abstract

A key assumption in regression discontinuity analysis is that units cannot manipulate the value of their running variable in a way that guarantees or avoids assignment to the treatment. Standard identification arguments break down if this condition is violated. This paper shows that treatment effects remain partially identified in this case. We derive sharp bounds on the treatment effects, show how to estimate them, and propose ways to construct valid confidence intervals. Our results apply to both sharp and fuzzy regression discontinuity designs. We illustrate our methods by studying the effect of unemployment insurance on unemployment duration in Brazil, where we find strong evidence of manipulation at eligibility cutoffs.

1 INTRODUCTION

The regression discontinuity (RD) design (Thistlethwaite and Campbell, 1960) has become a popular empirical strategy in economics to evaluate the causal impact of treatments using observational data. Its distinct feature is that units are assigned to receive the treatment if and only if their value of a continuous running variable exceeds a fixed cutoff. This structure provides a transparent way to identify and estimate treatment effects for units close to the cutoff. The key idea is that the design creates a form of local randomization: conditional

on a unit being close to the cutoff, it is essentially random whether its value of the running variable falls to the left or to the right of the cutoff. Treatment effects can thus be inferred by comparing the outcomes of units on either side of the cutoff (Lee and Lemieux, 2010).

A concern regarding the validity of this argument in practice is that in many contexts units can influence their value of the running variable through strategic behavior. We refer to any such behavior as *manipulation* of the running variable in this paper. Manipulation is problematic if it leads to units just to the left and right of the cutoff being no longer comparable due to self-selection, in which case treatment effects are no longer point identified. This is indeed an issue that arises in many empirical applications.[1]

A jump in the density of the running variable at the cutoff is a strong indication that an RD design is impacted by manipulation (McCrary, 2008). In the applied literature, it has therefore become current practice to address concerns about manipulation by testing the null hypothesis that the density of the running variable is smooth around the cutoff. If this null hypothesis is not rejected, researchers typically proceed with their empirical analysis under the assumption that no manipulation occurs, while when facing a rejection they often give up on using the cutoff for inference on treatment effects.[2] This practice is problematic for at least two reasons. First, a non-rejection might not be due to the absence of manipulation but to lack of statistical power. The local randomization property could still be violated in such cases, and estimates based on ignoring this possibility be severely biased. Second, even if a test rejects the null hypothesis of no manipulation, the number of "problematic" units could still be relatively modest, and the data thus still be informative to some extent.

---

[1] For instance, Urquiola and Verhoogen (2009) document that schools manipulate enrollment to avoid having to add an additional classroom when faced with class-size caps in Chile. Other examples abound in education (e.g. Card and Giuliano, 2014; Dee, Dobbie, Jacob, and Rockoff, 2014; Scott-Clayton, 2011) as well as in other applied fields (e.g. Camacho and Conover, 2011). Manipulation of running variables around discontinuities (or "notches") in tax and transfer systems has even generated its own literature in public finance (Kleven and Waseem, 2013).

[2] There is a small number of papers that develop tailored solutions that are valid only under strong assumptions in this case. For examples, see Anderson and Magruder (2012), Bajari, Hong, Park, and Town (2011), and Davis, Engberg, Epple, Sieg, and Zimmer (2013).

In this paper, we propose a partial identification approach to dealing with the issue of potentially manipulated running variables in RD designs. We show that while the data are unable to uniquely pin down (or *point identify*) treatment effects under a general manipulation pattern, they are still informative in the sense that they imply bounds on (or *partially identify*) the value of interesting causal parameters. Our main contribution is to derive and explicitly characterize these bounds. We also propose methods for estimating our bounds, and discuss how to construct confidence intervals for treatment effects that have good coverage properties. Our results apply to both sharp and fuzzy RD designs. The approach is illustrated with an application to the Brazilian unemployment insurance system.

Our paper focuses on settings where manipulation of the running variable breaks the local randomization property by creating two unobservable types of units: *always-assigned* units for which the realization of the running variable is always on one particular side of the cutoff (which we normalize to be to the right of the cutoff), and *potentially-assigned* units that behave as postulated by the standard assumptions of an RD design. Such a structure can arise from various patterns of strategic behavior. The most immediate is one where always-assigned units are "manipulators" who have control over the value of the running variable and can ensure a realization that is to the right of the cutoff. However, we also discuss several other examples of settings that fit our framework.

We focus on the causal effect among potentially-assigned units as our parameter of interest. This is because the lack of local randomization makes it difficult to derive meaningful conclusions about causal effects among always-assigned units.[3] Our identification argument consists of two steps. First, building on McCrary (2008), we use the magnitude of the discontinuity in the density of the running variable at the cutoff to identify the proportion of always-assigned units among all units close to the cutoff. Second, we use this information to

---

[3]By definition, we never observe such a unit being assigned *not* to receive the treatment. It is therefore impossible to construct non-trivial bounds on such a unit's expected outcome in the absence of the treatment.

bound treatment effects by considering the "worst case" scenarios, in which always-assigned units are those with either the highest or the lowest values of the outcome variable, and then trimming the respective observations. We also explore how additional assumptions or covariates can be used to shrink the identified set. Our approach shares some similarities with that of Lee (2009) for bounding treatment effects in randomized experiments under sample selection,[4] but it is more involved especially in the case of a fuzzy RD design due its more complicated structure.

Estimates of our bounds are relatively straightforward to obtain in practice since our partial identification results deliver explicit expressions for them. We propose computationally convenient sample analogue estimators that involve nonparametric estimation of density, conditional quantile, and (truncated) conditional expectation functions using local polynomial smoothing (Fan and Gijbels, 1996). Our approach also includes a novel "polynomial truncation" method, which is required due to the particular structure of our problem. Finally, we discuss how to construct valid confidence intervals using existing techniques from the partial identification literature. We recommend the use of such confidence intervals in applications irrespective of the outcome of McCrary's (2008) test in order to ensure that inference is robust against the possibility of manipulation.

Last, we apply our approach to study the effect of unemployment insurance (UI) takeup around an eligibility cutoff in Brazil. The main purpose of this exercise is to illustrate our approach, but the empirical application is also relevant in itself. UI programs have been adopted or considered in a growing number of developing countries, but there is still limited evidence on their impacts. Moreover, providing new evidence may be challenging because manipulation of UI eligibility around existing cutoffs may be more likely in countries where informal employment is prevalent. We find strong evidence of manipulation around the

---

[4]Chen and Flores (2012) extend Lee (2009) to sample selection in randomized experiments with imperfect compliance. Kim (2012) and Dong (2015) extend Lee (2009) to sample selection in RD designs.

eligibility cutoff. Yet, we are able to infer that UI takeup increases the time it takes to return to a formal job by at least 24% for potentially-assigned workers at this cutoff (in the first year after layoff).

Our paper contributes to the growing methodological literature on RD designs. For references, see Hahn, Todd, and Van der Klaauw (2001), Porter (2003), Lee and Card (2008), McCrary (2008), Frandsen, Frölich, and Melly (2012), Imbens and Kalyanaraman (2012), Angrist and Rokkanen (forthcoming), and Calonico, Cattaneo, and Titiunik (2015). Our paper is also related to the extensive literature on partial identification. For references, see Manski (1990), Manski (1997), Horowitz and Manski (1995), Horowitz and Manski (2000), Manski and Tamer (2002), Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Andrews and Soares (2010), and Chernozhukov, Lee, and Rosen (2013).

The remainder of the paper is organized as follows. Section 2 introduces a framework for RD designs with manipulation. Section 3 studies partial identification of treatment effects in both Sharp and Fuzzy RD designs. Sections 4 and 5 discuss estimation and inference, respectively. Section 6 contains the empirical application. Finally, Section 7 concludes. Proofs and additional material can be found in the Appendix.

2 Setup

2.1 Basic RD Design

The aim of an RD design is to study the causal effect of a binary treatment or intervention on some outcome variable. We observe a random sample of n units, indexed by $i = 1, \ldots, n$, from some large population. The effect of the treatment is potentially heterogeneous among these units, which could be individuals or firms for instance. Following Rubin (1974), each unit is therefore characterized by a pair of potential outcomes, $Y_i(1)$ and $Y_i(0)$, which denote the outcome of unit i with and without receiving the treatment, respectively. Out of these two potential outcomes, we only observe the one corresponding to the realized treatment

status. Let $D_i \in \{0, 1\}$ denote the treatment status of unit i, with $D_i = 1$ if unit i receives the treatment, and $D_i = 0$ if unit i does not receive the treatment. The observed outcome can then be written as $Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$.

In an RD design, the treatment assignment of each unit i is a deterministic function of a so-called running variable X_i that is measured prior to, or is not affected by, the treatment. Let $Z_i \in \{0, 1\}$ denote the treatment assignment of unit i, with $Z_i = 1$ if unit i is assigned to receive the treatment, and $Z_i = 0$ if unit i is not assigned to receive the treatment. Then $Z_i = \mathbb{I}(X_i \geq c)$ for some fixed cutoff value c. Let the potential treatment status of unit i as a function of the running variable be $D_i(x)$, so that the observed treatment status is $D_i = D_i(X_i)$. Also define the limits $D_i^+ = D_i(c^+) \equiv \lim_{x \downarrow c} D_i(x)$ and $D_i^- = D_i(c^-) \equiv \lim_{x \uparrow c} D_i(x)$.[5] The extent to which units comply with their assignment distinguishes the two types of RD designs that are commonly distinguished in the literature: the Sharp RD (SRD) design and the Fuzzy RD (FRD) design. In a SRD design, compliance with the treatment assignment is perfect, and thus $D_i^+ = 1$ and $D_i^- = 0$ for all units i. In a FRD design, on the other hand, values of D_i^+ and D_i^- differ across units, but the conditional treatment probability $\mathbb{E}(D_i | X_i = x)$ is discontinuous at $x = c$.

2.2 Manipulation

Identification in standard RD designs relies on the intuition that conditional on the realization of a unit's running variable being close to the cutoff, it is essentially random whether this realization falls to the left or right of the cutoff.[6] Such "local randomization" ensures that units on different sides of the cutoff are comparable except for their treatment assignments. Treatment effects can thus be identified by comparing outcomes (and treatment

---

[5]Throughout the paper, we use the notation that $g(c^+) = \lim_{x \downarrow c} g(x)$ and $g(c^-) = \lim_{x \uparrow c} g(x)$ for a generic function $g(\cdot)$. We also follow the convention that whenever we take a limit we implicitly assume that this limit exists and is finite. Similarly, whenever an expectation or some other moment of a random variable is taken, it is implicitly assumed that the corresponding object exists and is finite.

[6]Such an interpretation is justified under a continuity condition on the distribution of potential outcomes; see below.

probabilities) of units on different sides of the cutoff. This source of identification may break down if at least some units behave strategically and can influence the value of their running variable. Throughout the paper, we refer to any pattern of behavior that fits this rather general description as *manipulation of the running variable.*

Manipulation by itself does not necessarily break the local randomization property of an RD design. For example, if students take a test, the presence of a pass/fail cutoff may increase effort among those who expect that their score will be close to the cutoff relative to those who are confident that they will pass. Despite such clear strategic behavior, the distribution of exerted effort is comparable among students just to the left and right of the cutoff in this case. An RD design thus identifies the causal effect of passing the test in this case. More generally, the RD design remains valid as long as manipulation does not lead to a discontinuity at the cutoff in the distribution of the underlying characteristics of the units.

Our paper focuses on settings where manipulation of the running variable is likely to break the local randomization property by creating two unobservable types of units: *always-assigned* units whose value of the running variable only takes on values to the right of the cutoff, and *potentially-assigned* units who can potentially be observed on either side of the cutoff. Such a structure can arise from several type of behavior. The most immediate is one where some units have control over the value of the running variable to the extent that they can ensure a realization to the right of the cutoff (and assignment to treatment is desirable for all units). We discuss several concrete examples and alternative mechanisms that also fit our framework below.

More formally, let $M_i \in \{0, 1\}$ denote an indicator for the unobserved type of unit i, with $M_i = 1$ if unit i is always-assigned and $M_i = 0$ if unit i is potentially-assigned. We then impose three assumptions for our analysis. The first one implies that the standard conditions from the RD literature (e.g. Hahn, Todd, and Van der Klaauw, 2001) are satisfied among potentially-assigned units.

Assumption 1. *(i)* $\mathbb{P}(D_i = 1|X_i = c^+, M_i = 0) > \mathbb{P}(D_i = 1|X_i = c^-, M_i = 0)$; *(ii)* $\mathbb{P}(D_i^+ \geq D_i^-|X_i = c, M_i = 0) = 1$; *(iii)* $\mathbb{E}(Y_i(d)|D_i^+ = d^1, D_i^- = d^0, X_i = x, M_i = 0)$, $\mathbb{P}(D_i^+ = 1|X_i = x, M_i = 0)$ *and* $\mathbb{P}(D_i^- = 1|X_i = x, M_i = 0)$ *are continuous in* x *at* c *for* $d, d^0, d^1 \in \{0, 1\}$; *(iv)* $F_{X|M=0}(x)$ *is differentiable in* x *at* c, *and the derivative is strictly positive.*

This assumption is stated here for the general case of a Fuzzy RD design. Assumption 1(i) requires that the treatment probability changes discontinuously at the cutoff value of the running variable, with the direction of the change normalized to be positive. Assumption 1(ii) is a monotonicity condition stating that the response of treatment selection to crossing the cutoff is monotone for every unit. Assumption 1(iii) is a continuity condition which roughly speaking requires the mean of potential outcomes and potential treatment status to be the same on both sides of the cutoff. Finally, Assumption 1(iv) implies that the running variable has a positive density at the cutoff, and thus that there are potentially-assigned units close to the cutoff on either side. Note that Assumptions 1(i)-(iii) simplify to the condition that $\mathbb{E}(Y_i(d)|X_i = x, M_i = 0)$ is continuous in x at c for $d \in \{0, 1\}$ for the special case of a Sharp RD design.

Assumption 2. *The derivative of* $F_{X|M=0}(x)$ *is continuous in* x *at* c.

Assumption 2 is a weak regularity condition on the distribution of the running variable among potentially-assigned units. Together with Assumption 1(iv), this assumption implies that the density of X_i among potentially-assigned units is smooth and strictly positive over some open neighborhood of c. Continuity of the running variable's density around the cutoff is a reasonable condition in applications, and is generally considered to be an indication for the absence of manipulation in the literature (McCrary, 2008).

Assumption 3. *(i)* $\mathbb{P}(X_i \geq c|M_i = 1) = 1$, *(ii)* $F_{X|M=1}(x)$ *is right-differentiable in* x *at* c.

Assumption 3 is the only restriction we impose on the properties of always-assigned units in our setup. Its first part, which is key to our analysis, is the defining property of this group. Together with Assumption 1, it implies that the running variable only takes on values to the right of the cutoff among those units that are problematic for the validity of the standard RD design. This is similar to, although much less restrictive than, the assumption of monotonic manipulation in McCrary (2008). The second part rules out mass points in the distribution of X_i around the cutoff. In particular, it rules out that the running variable is exactly equal to the cutoff among always-assigned units. However, the distribution of X_i is allowed to be arbitrarily highly concentrated close to c. In view of Assumption 1(iv), this condition implies that a unit's type cannot simply be inferred from the value of its running variable (without such a condition the analysis would be trivial). It also implies that in the full population, which contains both always-assigned and potentially-assigned units, the observed running variable X_i is continuously distributed, with a density that is generally discontinuous at c. Moreover, Assumption 1(iv) and 3 together imply that all units observed to the left of the cutoff are of the potentially-assigned type, i.e. $\mathbb{P}(M_i = 1|X_i = c^-) = 0$, whereas to the right of the cutoff we observe a mixture of types.

2.3 Discussion

Several types of strategic behavior can generate subgroups of always-assigned and potentially-assigned units. To illustrate this point, consider the example of an income transfer program for which eligibility is based on a cutoff value of a poverty score X_i, and the formula used to calculate the score takes as inputs household characteristics recorded during home visits by local administrators. Programs of this type are found in many developing countries, and various types of manipulation have been documented in this context (Camacho and Conover, 2011). We now give some examples of strategic behavior that fit into our setup.

1. Suppose that the formula for the poverty score is not publicly known. Then neither house-

holds nor local administrators can ensure program assignment through misreporting of input variables within reasonable bounds. There are thus no always-assigned households. Every household is potentially-assigned and a standard RD design is valid.

2. Suppose that some households know the formula for the poverty score, and are well connected to local administrators, who turn a blind eye when the households report inaccurate information. Those households can report combinations of variables such that X_i is to the right of the cutoff. They are thus always-assigned, while all other households are potentially-assigned.[7] Alternatively, suppose that some local administrators refuse to collaborate such that only a fraction of households is able to carry out its intended manipulation. Only those households that succeed in manipulating the running variable would then be always-assigned. The subset of households whose manipulation efforts fail would be counted as potentially-assigned along with those households that never made a manipulation attempt.

3. Suppose that all households report information truthfully, but local administrators fill in combinations of variables such that X_i is to the right of the cutoff irrespective of the reported information if a household strongly supports local elected officials. Such households are then always-assigned, even though they are not engaging in any manipulation themselves.

4. Suppose that households can request a second home visit after learning the outcome of the first one, and that only the most recent score is relevant for program eligibility. Let X_{ji} be the poverty score based on the jth visit of household i, and suppose that households will request a second visit if and only if they were ineligible based on the first visit. Then the observed running variable is $X_i = X_{1i} \cdot \mathbb{I}(X_{1i} \geq c) + X_{2i} \cdot \mathbb{I}(X_{1i} < c)$. All households with $X_{1i} \geq c$ are always-assigned, whereas all households that receive a second visit are

---

[7]Misreporting households should have an incentive not to report information in such a away that their poverty score is exactly equal to the cutoff in order to avoid detection by e.g. central administrators. This makes the assumption of a continuously distributed running variable among always-assigned units palatable.

potentially-assigned. "Manipulation" occurs in this case even though nobody is doing anything illegal or against the terms of the program.

One can easily construct further variants of these examples which still fit into our setup.[8] The examples also have natural analogues in other contexts. Consider for instance an educational program for which students are eligible if their score in a test falls to the right of some cutoff. Teachers could then directly manipulate test scores, or students could retake the test if their score falls to the left of the cutoff. Our setup thus applies to a wide range of empirical settings.

2.4 Parameters of Interest

Our parameter of interest is the average causal effect of the treatment among potentially-assigned units. This is because the lack of local randomization makes it difficult to derive meaningful conclusions about causal effects among always-assigned units. Specifically, we study the identification of

$$\Gamma_0 \equiv \mathbb{E}(Y_i(1) - Y_i(0)|X_i = c, D_i^+ > D_i^-, M_i = 0), \tag{2.1}$$

which can be understood as the local average treatment effect for the subgroup of potentially-assigned "compliers", who receive the treatment if and only if their value of the running variable X_i is to the right of the cutoff (Imbens and Angrist, 1994). In the special case of a sharp design, where $D_i^+ > D_i^-$ for all units, the conditioning on "compliers" in (2.1) is redundant, and Γ_0 can be interpreted as a standard average treatment effect.

One reason for the popularity of the RD design is that in settings without manipulation, where every unit is potentially-assigned, the parameter Γ_0 is likely *policy relevant*. Specifically, it captures the causal effect for units whose program assignment would change following

---

[8]For example, misreporting household information or requesting a second home visit may be costly, with the cost depending on the distance between the cutoff and the true or initial poverty score, respectively.

marginal changes in the level of the cutoff. We can use the same examples as in the previous subsection to discuss the policy relevance of Γ_0 in the presence of always-assigned units.

In the examples involving active manipulation of input variables, the same units are likely to be always-assigned irrespective of marginal changes in the level of the cutoff. Γ_0 would therefore capture the policy-relevant effect in this case given that such marginal changes would only affect program assignment for potentially-assigned households.

In contrast, Γ_0 would not capture the causal effect for all units whose program assignment will change with marginal changes in the level of the cutoff in the example involving a second home visit. Consider a small increase of the cutoff from c to $\tilde{c}$. Households with $X_{1i} < c$ and $X_{2i} \in [c, \tilde{c})$ will still request a second home visit, but will not be assigned to the program anymore. Γ_0 captures the local average treatment effect for these households. However, other households may also lose eligibility for the program. Households with $X_{1i} \in [c, \tilde{c})$ will now request a second home visit, but only those with $X_{2i} > \tilde{c}$ will remain eligible. Moreover, some households that would have become eligible thanks to a second visit may stop requesting such a visit if doing so involves a cost that is, e.g., increasing in the distance between the cutoff and the initial poverty score. We cannot identify the local average treatment effect for these other types of households. Therefore Γ_0 always captures part of the policy-relevant effect, but the *degree* of its policy relevance depends on the context and behaviors at play.

3 Identification under Manipulation

Since we cannot infer whether any given unit is always-assigned or potentially-assigned, the local average treatment effect Γ_0 defined in (2.1) is generally not point identified. In this section, we therefore derive bounds on this quantity. By Assumption 1 and standard arguments from the RD literature (e.g. Hahn, Todd, and Van der Klaauw, 2001), it follows

that $\Gamma_0 = \Delta_0/\Psi_0$, where

$$\Delta_0 \equiv \mathbb{E}(Y_i|X_i = c^+, M_i = 0) - \mathbb{E}(Y_i|X_i = c^-, M_i = 0) \quad \text{and}$$

$$\Psi_0 \equiv \mathbb{E}(D_i|X_i = c^+, M_i = 0) - \mathbb{E}(D_i|X_i = c^-, M_i = 0).$$

When the RD design is sharp this representation can be further simplified, as $\Psi_0 = 1$ and thus $\Gamma_0 = \Delta_0$ in this case. We work with these representations in our identification analysis.

3.1 Proportion of Always-Assigned Units

In order to obtain bounds on treatment effects, we first study identification of two important intermediate quantities. These are $\tau \equiv \mathbb{P}(M_i = 1|X_i = c^+)$, the proportion of always-assigned units among all units just to the right of the cutoff, and $\tau_d \equiv \mathbb{P}(M_i = 1|X_i = c^+, D_i = d)$, the proportion of always-assigned units among units with treatment status $d \in \{0, 1\}$ just to the right of the cutoff. While we cannot observe or infer the type of any given unit, under our assumptions we can point identify τ from the size of the discontinuity in the density of the observed running variable at the cutoff.

Lemma 1. *Suppose Assumptions 1–3 hold. Then $\tau = 1 - f_X(c^-)/f_X(c^+)$, where f_X denotes the density of X_i.*

The two probabilities τ_1 and τ_0 are not point identified but only partially identified in our model. There are two logical restrictions on the range of their plausible values. First, by the law of total probability, any pair of candidate values $(\tau_1, \tau_0) \in [0, 1]^2$ has to satisfy the restriction that

$$\tau = \tau_1 \cdot \mathbb{E}(D_i|X_i = c^+) + \tau_0 \cdot (1 - \mathbb{E}(D_i|X_i = c^+)). \tag{3.1}$$

Second, our monotonicity condition in Assumption 1(i) implies that

$$\mathbb{E}(D_i|X_i = c^+) \cdot \frac{1 - \tau_1}{1 - \tau} > \mathbb{E}(D_i|X_i = c^-). \tag{3.2}$$

13

With $\mathcal{T}$ denoting the set containing those $(\tau_1, \tau_0) \in [0,1]^2$ that satisfy the two restrictions (3.1)–(3.2), we have the following result.

Lemma 2. *Suppose that Assumptions 1–3 hold. Then $\mathcal{T}$ is a sharp identified set for the pair of probabilities (τ_1, τ_0).*

Geometrically, the set $\mathcal{T}$ is a straight line in $[0,1]^2$. For our following analysis, it will be notationally convenient to represent this set in terms of the location of the endpoints of the line. That is, we can write

$$\mathcal{T} = \{(\eta_1(t), \eta_0(t)) : t \in [0,1]\} \quad \text{with} \quad \eta_d(t) = \tau_d^L + t \cdot (\tau_d^U - \tau_d^L)$$

for $d \in \{0,1\}$, where

$$\tau_1^L = \max\left\{0, 1 - \frac{1-\tau}{g^+}\right\}, \qquad \tau_1^U = \min\left\{1 - \frac{(1-\tau)\cdot g^-}{g^+}, \frac{\tau}{g^+}\right\},$$

$$\tau_0^L = \min\left\{1, \frac{\tau}{1-g^+}\right\}, \qquad \tau_0^U = \max\left\{0, \tau - \frac{(1-\tau)\cdot(g^+ - g^-)}{1-g^+}\right\},$$

using the shorthand notation that $g^+ = \mathbb{P}(D_i = 1 | X_i = c^+)$ and $g^- = \mathbb{P}(D_i = 1 | X_i = c^-)$.

3.2 Bounds on Treatment Effects for Potentially-Assigned Units

Using the above results on the proportion of always-assigned units, we can derive bounds on Γ_0. It is instructive to first consider a sharp RD design before studying the more general case of a fuzzy RD design.

3.2.1 Sharp RD Designs

Since $D_i^+ > D_i^-$ for every unit i in a sharp RD design, the causal parameter Γ_0 simplifies to an average treatment effect in this case. Since also $\Psi_0 = 1$, it can be written as

$$\Gamma_0 = \mathbb{E}(Y_i | X_i = c^+, M_i = 0) - \mathbb{E}(Y_i | X_i = c^-, M_i = 0). \tag{3.3}$$

14

As we only observe potentially-assigned units to the left of the cutoff, we have that $\mathbb{E}(Y_i|X_i = c^-, M_i = 0) = \mathbb{E}(Y_i|X_i = c^-)$ is point identified. We thus only need to bound the remaining conditional expectation on the right-hand side of (3.3).

Exploiting the fact that τ is point identified, this can be achieved by following a strategy similar to that in Lee (2009) for sample selection in randomized experiments. An upper bound on the expected outcome of potentially-assigned units just to the right of the cutoff is given by the expected outcome of all units there whose outcome is bigger than the corresponding τ quantile. Similarly, a lower bound is given by the expected outcome of those units with outcomes smaller than the corresponding $1 - \tau$ quantile. These bounds correspond to the two "worst case" scenarios in which the proportion τ of units with either the highest or the lowest outcomes are the always-assigned units. These bounds are sharp in the sense that the corresponding "worst case" scenarios are empirically conceivable, and thus the upper or lower bound could potentially coincide with the parameter of interest. Theorem 1 combines these arguments into bounds on Γ_0.

Theorem 1. *Suppose Assumptions 1–3 hold, that $D_i^+ > D_i^-$ for all $i = 1, \ldots, n$, and that $F_{Y|X}(y|c^+)$ is continuous in y. Let $Q_{Y|X}$ denote the conditional quantile function of Y_i given X_i. Then sharp lower and upper bounds on Γ_0 are given, respectively, by*

$$\Gamma_{0,SRD}^L = \mathbb{E}\left(Y_i|X_i = c^+, Y_i \leq Q_{Y|X}\left(1 - \tau|c^+\right)\right) - \mathbb{E}\left(Y_i|X_i = c^-\right) \ \ and$$

$$\Gamma_{0,SRD}^U = \mathbb{E}\left(Y_i|X_i = c^+, Y_i \geq Q_{Y|X}\left(\tau|c^+\right)\right) - \mathbb{E}\left(Y_i|X_i = c^-\right).$$

3.2.2 Fuzzy RD Designs

We now extend the partial identification result for Γ_0 in the Sharp RD design to the case of a Fuzzy RD design, which requires a more involved argument. Recall that we can write the

15

parameter of interest as

$$\Gamma_0 = \frac{\mathbb{E}(Y_i|X_i = c^+, M_i = 0) - \mathbb{E}(Y_i|X_i = c^-, M_i = 0)}{\mathbb{E}(D_i|X_i = c^+, M_i = 0) - \mathbb{E}(D_i|X_i = c^-, M_i = 0)} = \frac{\Delta_0}{\Psi_0}$$

in the FRD design. The two expectations in this expression that condition on $X_i = c^-$ are again point identified: as only potentially-assigned units are observed to the left of the cutoff, the conditioning on $M_i = 0$ is redundant for these terms. The form of the two remaining conditional expectations is familiar from the previous subsection, but simply applying the techniques used there to each of them separately would be unnecessarily conservative. This is because the sets of units that would be trimmed when maximizing the numerator and minimizing the denominator, respectively, would not necessarily be the same. Thus these bounds would not be jointly achievable.

To derive sharp (that is, best possible) bounds on the parameter of interest, write the unknown term in the definition of Δ_0 as

$$\mathbb{E}(Y_i|X_i = c^+, M_i = 0)$$
$$= \sum_{d \in \{0,1\}} \mathbb{E}(Y_i|X_i = c^+, M_i = 0, D_i = d) \cdot \mathbb{P}(D_i = d|X_i = c^+, M_i = 0),$$

and suppose for a moment that τ_1 and τ_0 were actually known. Then the two conditional expectations on the right-hand side of the previous equation can be bounded sharply by considering the "worst case" scenarios in which the always-assigned units in either treatment status are the units with the highest or the lowest outcomes. That is, an upper bound on the expectation that conditions on $D_i = 1$ can be obtained by trimming the treated units just to the right of the cutoff with outcomes below the corresponding τ_1 quantile, and for the expectation that conditions on $D_i = 0$ by trimming those untreated units just to the right of the cutoff with outcomes below the corresponding τ_0 quantile. Lower bounds can be obtained analogously. Moreover, by Bayes' Theorem, the two probabilities on the right-hand

16

side of the last equation are point identified since

$$\mathbb{P}(D_i = d | X_i = c^+, M_i = 0) = \mathbb{P}(D_i = d | X_i = c^+) \cdot \frac{1 - \tau_d}{1 - \tau} \quad \text{for} \quad d \in \{0, 1\},$$

Given knowledge of τ_1 and τ_0, sharp lower and upper bounds on Δ_0 are thus given by

$$\Delta_0^L(\tau_1, \tau_0) = \sum_{d=0,1} \mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_d | c^+, d), D_i = d)$$

$$\times \mathbb{P}(D_i = d | X_i = c^+) \cdot \frac{1 - \tau_d}{1 - \tau} - \mathbb{E}(Y_i | X_i = c^-) \quad \text{and}$$

$$\Delta_0^U(\tau_1, \tau_0) = \sum_{d=0,1} \mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_d | c^+, d), D_i = d)$$

$$\times \mathbb{P}(D_i = d | X_i = c^+) \cdot \frac{1 - \tau_d}{1 - \tau} - \mathbb{E}(Y_i | X_i = c^-),$$

respectively. Moreover, with knowledge of τ_1 and τ_0 the term Ψ_0 would also be point identified through the relationship

$$\Psi_0 = \mathbb{E}(D_i | X_i = c^+) \cdot \frac{1 - \tau_1}{1 - \tau} - \mathbb{E}(D_i | X_i = c^-) \equiv \Psi_0(\tau_1). \tag{3.4}$$

If we knew the values of τ_1 and τ_0, sharp lower and upper bounds on Γ_0 would thus be given, respectively, by

$$\Gamma_0^L(\tau_1, \tau_0) \equiv \frac{\Delta_0^L(\tau_1, \tau_0)}{\Psi_0(\tau_1)} \quad \text{and} \quad \Gamma_0^U(\tau_1, \tau_0) \equiv \frac{\Delta_0^U(\tau_1, \tau_0)}{\Psi_0(\tau_1)}, \cdot \tag{3.5}$$

These bounds are not practially useful by themselves since τ_1 and τ_0 are in fact not known in our setup; but (following the result in Lemma 2) only partially identified. However, we can use them to obtain sharp bounds on Γ_0 by finding those values of $(\tau_1, \tau_0) \in \mathcal{T}$ that lead to the most extreme values of the quantities defined in (3.5). This "worst case" approach is formalized in the following theorem.

Theorem 2. *Suppose that Assumptions 1–3 hold, and that $F_{Y|XD}(y | c^+, d)$ is continuous in*

17

y for $d \in \{0, 1\}$. Then sharp lower and upper bounds on Γ_0 are given, respectively, by

$$\Gamma_{0,FRD}^{L} = \inf_{(t_1,t_0)\in\mathcal{T}} \frac{\Delta_0^L(t_1, t_0)}{\Psi_0(t_1)} \quad and \quad \Gamma_{0,FRD}^{U} = \sup_{(t_1,t_0)\in\mathcal{T}} \frac{\Delta_0^U(t_1, t_0)}{\Psi_0(t_1)}.$$

In general, it could be the case that $\Gamma_0^L = -\infty$ and/or $\Gamma_0^U = \infty$ because $\Psi_0(t_1)$ is not necessarily bounded away from zero. A simple sufficient condition that ensures finiteness of both the upper and lower bound in Theorem 2 is that

$$\tau < \frac{\mathbb{E}(D_i|X_i = c^+) - \mathbb{E}(D_i|X_i = c^-)}{1 - \mathbb{E}(D_i|X_i = c^-)}.$$

With more excessive levels of manipulation it might not be possible to distinguish empirically between a setting where always-assigned units just to the right of the cutoff have very low treatment probabilities, and a setting where the treatment probability of potentially-assigned units on either side of the cutoff is identical. In the latter setting, Γ_0 would not be identified even if we could observe each unit's type.

3.2.3 Fuzzy RD Designs with Additional Restrictions

The bounds in Theorem 2 can be narrowed if one is willing to impose stronger assumptions on the units' behavior than the (rather weak) ones that we have imposed so far. Consider for instance a setting where always-assigned units obtain values of the running variable to the right of the cutoff by actively misreporting some information. Such units thus make the conscious choice to become eligible for the treatment. It might therefore seem plausible to assume that their probability of actually receiving the treatment conditional on being eligible is relatively high in some appropriate sense. Depending on the exact details of the empirical application, one might be willing to assume, for example, that always-assigned units are at least as likely to get treated as eligible potentially-assigned units. The following Theorem provides expressions for the bounds under this assumption.

Theorem 3. *Suppose that the conditions of Theorem 2 hold, and that $\mathbb{E}(D_i|X_i = c^+, M_i =$*

$1) \geq \mathbb{E}(D_i | X_i = c^+, M_i = 0)$. *Then the set $\mathcal{T}_a$ of possible values of (τ_1, τ_0) that are compatible with the data is given by*

$$\mathcal{T}_a \equiv \{(t_1, t_0) : (t_1, t_0) \in \mathcal{T} \text{ and } t_1 \geq \tau\},$$

and thus sharp lower and upper bounds on Γ_0 are given, respectively, by

$$\Gamma^L_{0, FRD(a)} = \inf_{(t_1, t_0) \in \mathcal{T}_a} \frac{\Delta^L_0(t_1, t_0)}{\Psi_0(t_1)} \quad \text{and} \quad \Gamma^U_{0, FRD(a)} = \sup_{(t_1, t_0) \in \mathcal{T}_a} \frac{\Delta^U_0(t_1, t_0)}{\Psi_0(t_1)}.$$

In some cases, it may be reasonable to drive this line of reasoning further and consider the identifying power of the assumption that always-assigned units *always* receive the treatment. The following theorem provides expressions for the bounds under this assumption.

Theorem 4. *Suppose that the conditions of Theorem 2 hold, and that $\mathbb{E}(D_i | X_i = c^+, M_i = 1) = 1$. Then the values τ_1 and τ_0 are point identified:*

$$\tau_1 = \frac{\tau}{\mathbb{E}(D_i | X_i = c^+)} \quad \text{and} \quad \tau_0 = 0;$$

and sharp lower and upper bounds on Γ_0 are given, respectively, by

$$\Gamma^L_{0, FRD(b)} = \frac{\Delta^L_0(\tau/\mathbb{E}(D_i | X_i = c^+), 0)}{\Psi_0(\tau/\mathbb{E}(D_i | X_i = c^+))} \quad \text{and} \quad \Gamma^U_{0, FRD(b)} = \frac{\Delta^U_0(\tau/\mathbb{E}(D_i | X_i = c^+), 0)}{\Psi_0(\tau/\mathbb{E}(D_i | X_i = c^+))}.$$

Comparing the first part of the Theorem 3 to Lemma 2 and the following discussion, we see the additional restrictions on treatment probabilities increase the lowest possible value of τ_1 from $\max\{0, 1 + (\tau - 1)/\mathbb{E}(D_i | X_i = c^+)\}$ to τ, and correspondingly decrease the largest possible value for τ_0 from $\min\{1, \tau/(1 - \mathbb{E}(D_i | X_i = c^+))\}$ to τ. Hence $\mathcal{T}_a \subset \mathcal{T}$, and we obtain narrower bounds because optimization is carried out over a smaller set. Under the conditions of Theorem 4, the set of plausible values of (τ_1, τ_0) shrinks to a singleton, which means that sharp bounds on Γ_0 can be defined without invoking an optimization operator.

3.3 Further Extensions and Remarks

In this subsection, we discuss a number of extensions and remarks related to the main results that we derived in the previous subsection.

3.3.1 Using Covariates to Tighten the Bounds

It is possible to use covariates that are measured prior to the treatment to obtain bounds on causal effects that are more narrow than the ones derived above, and thus potentially more informative in applications. Let W_i be a vector of such covariates with a support $\mathcal{W}$. By using arguments similar to those in Lee (2009), one can then narrow the bounds on Γ_0 in the SRD and FRD designs as described in Corollaries 1 and 2, respectively. The idea is that if the proportion of always-assigned units varies with the value of W_i, trimming units based on their position in the outcome distribution conditional on W_i will lead to a smaller number of units with extreme outcomes being trimmed overall, which narrows the bounds.

Corollary 1. *Suppose that the assumptions of Theorem 1 hold, mutatis mutandis, conditional on the covariates W_i with probability 1. Then sharp lower and upper bounds on Γ_0 are given by*

$$\Gamma^L_{0,FRD-W} = \int \mathbb{E}(Y_i|X_i = c^+, Y_i \leq Q_{Y|XW}(1-\tau(w)|c^+, w), W_i = w)dF_{W|X}(w|c^-)$$

$$- \mathbb{E}(Y_i|X_i = c^-) \quad and$$

$$\Gamma^U_{0,FRD-W} = \int \mathbb{E}(Y_i|X_i = c^+, Y_i \geq Q_{Y|XW}(\tau(w)|c^+, w), W_i = w)dF_{W|X}(w|c^-)$$

$$- \mathbb{E}(Y_i|X_i = c^-),$$

respectively, where $\tau(w) = 1 - f_{X|W}(c^+|w)/f_{X|W}(c^-|w)$.

Corollary 2. *Suppose that the assumptions of Theorem 2 hold, mutatis mutandis, conditional on the covariates W_i with probability 1. Then sharp lower and upper bounds on Γ_0 are*

given by

$$\Gamma^L_{0,FRD-W} = \inf_{(t_1(w),t_0(w))\in\mathcal{T}(w),w\in\mathcal{W}} \frac{\int \Delta^L_0(t_1(w),t_0(w)|W_i=w)dF_{W|X}(w|c^-)}{\int \Psi_0(t_1(w)|W_i=w)dF_{W|X}(w|c^-)} \quad and$$

$$\Gamma^U_{0,FRD-W} = \sup_{(t_1(w),t_0(w))\in\mathcal{T}(w),w\in\mathcal{W}} \frac{\int \Delta^U_0(t_1(w),t_0(w)|W_i=w)dF_{W|X}(w|c^-)}{\int \Psi_0(t_1(w)|W_i=w)dF_{W|X}(w|c^-)},$$

respectively, where $t_1(w)$, $t_0(w)$, $\mathcal{T}(w)$, $\Delta^L_0(t_1(w),t_0(w)|W_i=w)$, $\Delta^U_0(t_1(w),t_0(w)|W_i=w)$, and $\Psi_0(t_1(w)|W_i=w)$ are as defined above, conditioning throughout on $W_i=w$.

3.3.2 Bounds for Discrete Outcomes

The results in Theorems 1 and 2 are stated for the case in which the outcome variable is continuously distributed. This is for notational convenience only, and our results immediately generalize to the case of a discrete outcome variable. This additional flexibility can be important in many empirical settings. The following corollaries provide these generalizations.

Corollary 3. *Suppose that the assumptions of Theorem 1 hold, and that $supp(Y_i|X_i=c^+)$ is a finite set. Then sharp lower and upper bounds on Γ_0 are given by*

$$\Gamma^L_{0,SRD} = (1-\theta^L)\mathbb{E}(Y_i|X_i=c^+,Y_i<Q_{Y|X}(1-\tau|c^+))$$
$$+ \theta^L Q_{Y|X}(1-\tau|c^+) - \mathbb{E}(Y_i|X_i=c^-) \quad and$$

$$\Gamma^U_{0,SRD} = (1-\theta^U)\mathbb{E}(Y_i|X_i=c^+,Y_i>Q_{Y|X}(\tau|c^+))$$
$$+ \theta^U Q_{Y|X}(\tau|c^+) - \mathbb{E}(Y_i|X_i=c^-),$$

respectively, with

$$\theta^L = \frac{\mathbb{P}(Y_i\geq Q_{Y|X}(1-\tau|c^+)|X_i=c^+)-\tau}{1-\tau} \quad and$$

$$\theta^U = \frac{\mathbb{P}(Y_i\leq Q_{Y|X}(\tau|c^+)|X_i=c^+)-\tau}{1-\tau},$$

using the convention that $\mathbb{E}(A|A<\min supp(A)) = \mathbb{E}(A|A>\max supp(A)) = 0$ for a generic random variable A with finite support.

Corollary 4. *Suppose that the assumptions of Theorem 2 hold, and that* $\text{supp}(Y_i|X_i = c^+)$ *is a finite set. Then sharp lower and upper bounds on* Γ_0 *are given by*

$$\Gamma^L_{0,FRD} = \inf_{(t_1,t_0)\in\mathcal{T}} \frac{\Delta^L_0(t_1,t_0)}{\Psi_0(t_1)} \quad and \quad \Gamma^U_{0,FRD} = \sup_{(t_1,t_0)\in\mathcal{T}} \frac{\Delta^U_0(t_1,t_0)}{\Psi_0(t_1)},$$

respectively, with

$$\Delta^L_0(\tau_1,\tau_0) = \sum_{d=0,1} ((1-\theta^L_d)\mathbb{E}(Y_i|X_i = c^+, Y_i < Q_{Y|XD}(1-\tau_d|c^+,d), D_i = d)$$

$$+ \theta^L_d Q_{Y|XD}(1-\tau_d|c^+,d))\mathbb{P}(D_i = d|X_i = c^+)\frac{1-\tau_d}{1-\tau} - \mathbb{E}(Y_i|X_i = c^-) \quad and$$

$$\Delta^U_0(\tau_1,\tau_0) = \sum_{d=0,1} ((1-\theta^U_d)\mathbb{E}(Y_i|X_i = c^+, Y_i > Q_{Y|XD}(\tau_d|c^+,d), D_i = d)$$

$$+ \theta^U_d Q_{Y|XD}(\tau_d|c^+,d))\mathbb{P}(D_i = d|X_i = c^+)\frac{1-\tau_d}{1-\tau} - \mathbb{E}(Y_i|X_i = c^-),$$

and with

$$\theta^L_d = \frac{\mathbb{P}(Y_i \geq Q_{Y|XD}(1-\tau_d|c^+,d)|X_i = c^+, D_i = d) - \tau_d}{1-\tau_d} \quad and$$

$$\theta^U_d = \frac{\mathbb{P}(Y_i \leq Q_{Y|XD}(\tau_d|c^+,d)|X_i = c^+, D_i = d) - \tau_d}{1-\tau_d}$$

for $d \in \{0,1\}$, *using the convention that* $\mathbb{E}(A|A < \min \text{supp}(A)) = \mathbb{E}(A|A > \max \text{supp}(A)) = 0$ *for a generic random variable* A *with finite support.*

3.3.3 Identifying the Characteristics of Always-Assigned and Potentially-Assigned Units

It is not possible to determine whether any given unit belongs to the group of always-assigned ones or the group of potentially-assigned ones in our setup. This does not mean, however, that it is impossible to give any further characterization of these two subgroups. In particular, if the data include a vector W_i of predetermined covariates, it is possible to identify the distribution of these covariates among always-assigned and potentially-assigned units just to the right of the cutoff as long as the distribution of W_i (conditional on type)

22

does not change discontinuously at c. This could be useful, for instance, to target policies aimed at mitigating manipulation. The following corollary formally states the result.

Corollary 5. *Suppose that Assumptions 1–2 hold, and that* $\mathbb{E}\left(g\left(W_i\right)|X_i = x, M_i = 0\right)$ *is continuous in x at c for some known function $g(\cdot)$. Then*

$$\mathbb{E}\left(g\left(W_i\right)|X_i = c^+, M_i = 1\right) = \frac{1}{\tau}\left(\mathbb{E}\left(g\left(W_i\right)|X_i = c^+\right) - \mathbb{E}\left(g\left(W_i\right)|X_i = c^-\right)\right)$$

$$+ \mathbb{E}\left(g\left(W_i\right)|X_i = c^-\right) \quad and$$

$$\mathbb{E}\left(g\left(W_i\right)|X_i = c^+, M_i = 0\right) = \mathbb{E}\left(g\left(W_i\right)|X_i = c^-\right).$$

By putting $g(w) = w^k$, for example, the corollary shows that the data identify the kth moment of the distribution of W_i among always-assigned and potentially-assigned units just to the right of the cutoff, where $k \in \mathbb{N}$. If W_i is the age of an individual, for instance, it is possible to determine whether always-assigned individuals are older on average than potentially-assigned individuals. By putting $g(w) = \mathbb{I}\left\{w \leq r\right\}$, we obtain an identification result for the conditional CDF of the covariates at any level $r \in \mathbb{R}$. This means that we can identify any feature of the conditional distribution of W_i.

4 ESTIMATION

In this section, we describe how the upper and lower bounds on treatment effects that we derived in the previous section can be estimated in practice. This requires combining a number of intermediate steps in which the right and/or left limits of various density, conditional quantile or (truncated) conditional expectation functions have to be estimated. Following the recent RD literature, we focus on flexible nonparametric methods, and in particular local polynomial smoothing, for this task. Local polynomial estimators are well-known to have good properties in boundary regions, and are thus appealing for RD settings. The main idea is to approximate the respective unknown function with a polynomial of

order $p \in \mathbb{Z}_+$, often taken to be 1 in applications, and then to estimate the coefficients of the approximating polynomial by minimizing some weighted measure of distance. The weighting results in a local approximation of the unknown function in the sense that the weights are decreasing to zero with the distance of an observation to the cutoff. The exact values of these weights are determined by a smoothing parameter, or bandwidth, $h \in \mathbb{R}_+$ and a kernel function $K(\cdot)$, taken to be a symmetric density function on the interval $[-1, 1]$.

4.1 Estimation of Intermediate Quantities

We begin by describing the construction of a number of intermediate estimators. To simplify the exposition, we use the same polynomial order p, bandwidth h and kernel function $K(\cdot)$ in all intermediate estimation steps in this paper. We also use the notation that $\boldsymbol{\pi}_p(x) = (1/0!, x/1!, x^2/2!, \ldots, x^p/p!)'$ and $K_h(x) = K(x/h)/h$ for any $x \in \mathbb{R}$, and define the $(p+1)$-vector $\mathbf{e}_1 = (1, 0, \ldots, 0)'$. The data available to the econometrician is an independent and identically distributed sample $\{(Y_i, D_i, X_i), i = 1, \ldots, n\}$ of size n.

4.1.1 Proportion of Always-Assigned Units

Following the result in Lemma 1, estimating τ requires estimates of the right and left limits of the density at the cutoff. There are a number of nonparametric estimators that can be used to estimate densities at boundary points; see for example Lejeune and Sarda (1992), Jones (1993), Cheng (1997) or Cattaneo, Jansson, and Ma (2015). Here we use a particularly simple procedure that corresponds to a standard Rosenblatt-Parzen estimator using the *equivalent kernel* of a local polynomial regression of order p. Specifically, our estimators of $f_X(c^+)$ and $f_X(c^-)$ are given by

$$\widehat{f}^+ = \frac{1}{n} \sum_{i=1}^n K_h^{p+}(X_i - c)\mathbb{I}\{X_i \geq c\} \quad \text{and} \quad \widehat{f}^- = \frac{1}{n} \sum_{i=1}^n K_h^{p-}(X_i - c)\mathbb{I}\{X_i < c\},$$

respectively, where $K^{p+}(x) = \mathbf{e}_1' S^{-1}(1, x, \ldots, x^p)' K(x)$ with $S = (a_{j+l})_{0 \leq j,l \leq p}$ a matrix of dimension $(p+1) \times (p+1)$ and $a_j = \int_0^\infty u^j K(u) du$ for $j = 1, \ldots, 2p$ being constants that de-

pend on the kernel function only; and $K^{p-}(x)$ is defined analogously with $b_j = \int_{-\infty}^0 u^j K(u) du$ replacing the a_j. In view of Lemma 1, and the fact that by assumption the proportion of always-assigned units among units just to the right of the cutoff has to be non-negative, our estimate of τ is then given by

$$\widehat{\tau} = \max\{\widetilde{\tau}, 0\}, \quad \text{where} \quad \widetilde{\tau} = 1 - \widehat{f}^- / \widehat{f}^+.$$

To see the connection with local polynomial regression, let $\widetilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i = x\}$ be the empirical density function of the sample points $\{X_1, \ldots, X_n\}$. Then we have that

$$\widehat{f}^+ = \mathbf{e}_1' \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \int_c^\infty (\widetilde{f}_n(x) - \boldsymbol{\pi}_p(x - c)'\beta)^2 K_h(x - c) dx \quad \text{and}$$

$$\widehat{f}^- = \mathbf{e}_1' \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \int_{-\infty}^c (\widetilde{f}_n(x) - \boldsymbol{\pi}_p(x - c)'\beta)^2 K_h(x - c) dx.$$

These two estimators can thus be interpreted as local polynomial approximations to the empirical density function (cf. Lejeune and Sarda, 1992).

4.1.2 Conditional Quantile Functions

Next, we construct estimates of the conditional quantile function of the outcome given the running variable, and of the outcome given the running variable and the unit's treatment status. For reasons explained below, we not only estimate these functions themselves, but also their right derivatives with respect to the running variable up to order p. To simplify the notation, we denote the vectors that contain the right limit of the respective conditional quantile functions and their first p derivatives by

$$Q_\nabla^+(t) = (Q_{Y|X}(t, c^+), \partial_x Q_{Y|X}(t, x)|_{x=c^+}, \ldots, \partial_x^p Q_{Y|X}(t, x)|_{x=c^+})' \text{ and}$$

$$Q_\nabla^+(t, d) = (Q_{Y|XD}(t, c^+, d), \partial_x Q_{Y|XD}(t, x, d)|_{x=c^+}, \ldots, \partial_x^p Q_{Y|XD}(t, x, d)|_{x=c^+})'.$$

Note that to keep the notation simple we distinguish these two vectors of functions through their arguments only. For any $t \in (0, 1)$ and $d \in \{0, 1\}$, our estimates of these two objects

are then given by

$$\widehat{Q}^+_\nabla(t) = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n \rho_t(Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)K_h(X_i - c)\mathbb{I}\{X_i \ge c\}, \text{ and}$$

$$\widehat{Q}^+_\nabla(t, d) = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n \rho_t(Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)K_h(X_i - c)\mathbb{I}\{X_i \ge c, D_i = d\},$$

respectively, where $\rho_t(x) = (t - \mathbb{I}\{x < 0\})x$ is the "check" function commonly used in the quantile regression literature (Koenker and Bassett, 1978; Koenker, 2005). See Yu and Jones (1998) for more details on local polynomial quantile regression. From a numerical point of view, these estimation steps can be carried out by any software package able to compute a weighted linear quantile regression estimator.

4.1.3 Conditional Expectations

Our last groups of intermediate estimands are the left and right limits of various (possibly truncated) conditional expectation functions. To simplify the exposition, we use the shorthand notation that

$$Q^+(t) = Q_{Y|X}(t, c^+) \quad \text{and} \quad Q^+(t, d) = Q_{Y|XD}(t, c^+, d)$$

for the conditional quantiles, write

$$m^{L+}(t) = \mathbb{E}(Y_i|X_i = c^+, Y_i \le Q^+(t)),$$

$$m^{U+}(t) = \mathbb{E}(Y_i|X_i = c^+, Y_i \ge Q^+(1 - t)),$$

$$m^{L+}(t_1, t_0, d) = \mathbb{E}(Y_i|X_i = c^+, Y_i \le Q^+(t_d, d), D_i = d),$$

$$m^{U+}(t_1, t_0, d) = \mathbb{E}(Y_i|X_i = c^+, Y_i \ge Q^+(1 - t_d, d), D_i = d),$$

26
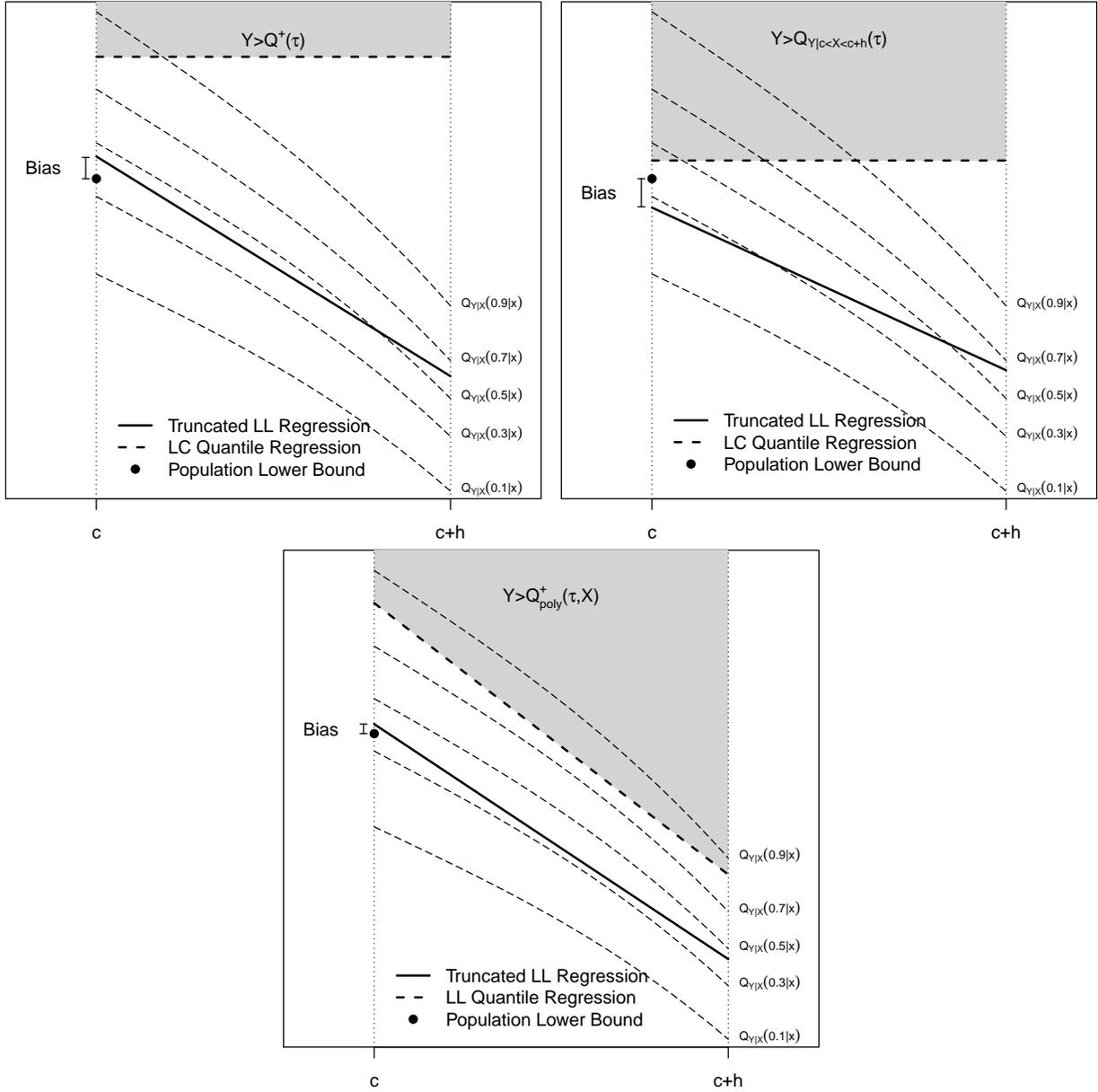
Figure 4.1: Illustration of the benefits of polynomial truncation.

for the various truncated conditional expectation functions, and denote the "non-truncated" conditional expectations that we have to estimate by

$$m^- = \mathbb{E}(Y_i | X_i = c^-), \quad g^+ = \mathbb{E}(D_i | X_i = c^+), \quad g^- = \mathbb{E}(D_i | X_i = c^-).$$

Estimation of the truncated conditional expectations $m^{L+}(\tau)$ and $m^{U+}(\tau)$ involves a subtle issue. At first sight, one might think that a natural way to estimate these objects is to run local polynomial regressions in the subsamples with $Y_i \leq \widehat{Q}^+(\widehat{\tau})$ and $Y_i \geq \widehat{Q}^+(1 - \widehat{\tau})$, respectively. However, such a "constant truncation" rule has several undesirable properties, as illustrated in the top left panel of Figure 4.1 for the case of $m^{L+}(\tau)$ and a linear approximating function, that is $p = 1$. Suppose that the conditional quantiles of Y given $X = x$ are downward sloping in x over the area from c to $c + h$, the right neighborhood of the cutoff. Then first the proportion of units that is truncated from this neighborhood is substantially smaller than $\widehat{\tau}$. And second, those units that are truncated all have values of X_i very close to the cutoff. This leads to an additional bias in the local linear estimator (similar to the bias of the OLS estimator in a standard linear model with fixed censoring).

An alternative, seemingly natural way to estimate $m^{L+}(\tau)$ and $m^{U+}(\tau)$ if a uniform kernel is used would be to run local polynomial regressions in the subsamples with $Y_i \leq \widehat{Q}_{Y|c<X<c+h}(\widehat{\tau})$ and $Y_i \geq \widehat{Q}_{Y|c<X<c+h}(1 - \widehat{\tau})$, respectively, where $\widehat{Q}_{Y|c<X<c+h}$ denotes the empirical quantile function of the outcomes of those units whose value of X_i falls into the right neighborhood of the cutoff. As illustrated in the top right panel of Figure 4.1, this alternative "constant truncation" rule does not resolve the problem, as it again tends to remove a disproportionate number of units very close to the cutoff.

To address the problem, we propose to use a "polynomial truncation" rule, which removes units whose value of Y_i is below the value of a polynomial approximation of the conditional quantile function, before fitting an approximate regression function. The bottom panel of Figure 4.1 illustrates how proceeding like this can remove a substantial amount of bias from the estimates. In a first step, we compute

$$\widehat{Q}^+_{\text{poly}}(t, x) = \boldsymbol{\pi}_p(x - c)' \widehat{Q}^{\pm}_{\nabla}(t),$$

which is the estimated pth order polynomial approximation of the conditional quantile func-

28

tion $x \mapsto Q_{Y|X}(t, x)$ in a local neighborhood to the right of the cutoff. In a second step, we then estimate $m^{L+}(\tau)$ and $m^{U+}(\tau)$ by running local polynomial regression in the subsamples with $Y_i \leq \widehat{Q}^+_{\text{poly}}(\widehat{\tau}, X_i)$ and $Y_i \geq \widehat{Q}^+_{\text{poly}}(1 - \widehat{\tau}, X_i)$, respectively:

$$\widehat{m}^{L+}(\widehat{\tau}) = \mathbf{e}'_1 \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^{n} (Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}\left\{ X_i \geq c, Y_i \leq \widehat{Q}^+_{\text{poly}}(\widehat{\tau}, X_i) \right\}$$

$$\widehat{m}^{U+}(\widehat{\tau}) = \mathbf{e}'_1 \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^{n} (Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}\left\{ X_i \geq c, Y_i \geq \widehat{Q}^+_{\text{poly}}(1 - \widehat{\tau}, X_i) \right\}.$$

The same issue that complicates estimating $m^{L+}(\tau)$ and $m^{U+}(\tau)$ must also be taken into account when estimating $m^{L+}(t_1, t_0)$ and $m^{U+}(t_1, t_0)$. By following the same reasoning as above, we propose similar "polynomially truncated" local polynomial regression estimators for this task. We begin by computing

$$\widehat{Q}^+_{\text{poly}}(t, x, d) = \boldsymbol{\pi}_p(x - c)'\widehat{Q}^+_{\nabla}(t, d),$$

which is the estimated pth order polynomial approximation of the conditional quantile function $x \mapsto Q_{Y|XD}(t, x, d)$ in a local neighborhood to the right of the cutoff, and then estimate $m^{L+}(t_1, t_0, d)$ and $m^{U+}(t_1, t_0, d)$ by

$$\widehat{m}^{L+}(t_1, t_0, d) = \mathbf{e}'_1 \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^{n} (Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}^L_i(t_1, t_0, d),$$

$$\widehat{m}^{U+}(t_1, t_0, d) = \mathbf{e}'_1 \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^{n} (Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}^U_i(t_1, t_0, d)$$

respectively, with

$$\mathbb{I}^L_i(t_1, t_0, d) = \mathbb{I}\left\{ X_i \geq c, Y_i \leq \widehat{Q}^+_{\text{poly}}(t_d, X_i, d), D_i = d \right\},$$

$$\mathbb{I}^U_i(t_1, t_0, d) = \mathbb{I}\left\{ X_i \geq c, Y_i \geq \widehat{Q}^+_{\text{poly}}(1 - t_d, X_i, d), D_i = d \right\}.$$

Estimating the remaining conditional expectations we mentioned above poses no particular

29

difficulties. Estimates of m^-, g^+ and g^- can be obtained by local polynomial regression:

$$\widehat{m}^- = \mathbf{e}_1' \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^n (Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}\{X_i < c\},$$

$$\widehat{g}^+ = \mathbf{e}_1' \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^n (D_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}\{X_i \geq c\},$$

$$\widehat{g}^- = \mathbf{e}_1' \underset{\beta \in \mathbb{R}^{p+1}}{\text{argmin}} \sum_{i=1}^n (D_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}\{X_i < c\},$$

Again, any software package able to solve weighted least squares problems can be used to compute all the above estimates.

4.2 Estimation of Bounds

Using the intermediate estimates defined in the previous subsection, it is now straightforward to construct sample analogue estimators of the lower and upper bounds on the treatment effect Γ_0 in the sharp RD design. Specifically, our estimators of $\Gamma_{0,SRD}^L$ and $\Gamma_{0,SRD}^U$ from Theorem 1 are given by

$$\widehat{\Gamma}_{0,SRD}^L = \widehat{m}^{L+}(\widehat{\tau}) - \widehat{m}^- \quad \text{and} \quad \widehat{\Gamma}_{0,SRD}^U = \widehat{m}^{U+}(\widehat{\tau}) - \widehat{m}^-.$$

These estimators can be shown to be consistent under general conditions. We explore their large sample properties in more detail in our discussion of inference below.

Estimating the bounds on Γ_0 that we derived in Theorem 2 for the fuzzy RD design is a more difficult problem because our estimands take the form of a non-smooth functional of the population distribution of the data. Hirano and Porter (2012) show that it is not possible to construct locally asymptotically unbiased estimators for such objects, and that overly aggressive attempts at bias correction must eventually cause the variance of the estimator to explode. Like all conceivable estimators, our proposed estimators are therefore biased. To describe them, recall the representation of the set $\mathcal{T}$ given after Lemma 2, and note that

the lower and upper bound on Γ_0 from Theorem 2 can be written as

$$\Gamma^L_{0,FRD} = \inf_{t\in[0,1]} \Gamma^L_0(t) \quad \text{and} \quad \Gamma^U_{0,FRD} = \sup_{t\in[0,1]} \Gamma^U_0(t),$$

respectively, where $\Gamma^j_0(t) = \Delta^j_0(\eta_1(t), \eta_0(t))/\Psi_0(\eta_1(t))$ for $j \in \{L, U\}$. This notation, which incorporates the shape of the set $\mathcal{T}$ into the definition of the bounding function, has the advantage that the area over which optimization takes place does no longer depend on nuisance parameters. We then define the corresponding sample analogues

$$\widehat{\Gamma}^j_0(t) = \widehat{\Delta}^j_0(\widehat{\eta}_1(t), \widehat{\eta}_0(t))/\widehat{\Psi}_0(\widehat{\eta}_1(t)),$$

$$\widehat{\Delta}^j_0(t_1, t_0) = \widehat{m}^{j+}(t_1, t_0, 1) \cdot \frac{\widehat{g}^+(1-t_1)}{1-\widehat{\tau}} + \widehat{m}^{j+}(t_1, t_0, 0) \cdot \frac{(1-\widehat{g}^+)(1-t_0)}{1-\widehat{\tau}} - \widehat{m}^-,$$

$$\widehat{\Psi}_0(t_1) = \frac{\widehat{g}^+ \cdot (1-t_1)}{1-\widehat{\tau}} - \widehat{g}^-, \qquad \widehat{\eta}_d(t) = \widehat{\tau}^L_d + t \cdot (\widehat{\tau}^U_d - \widehat{\tau}^L_d);$$

for $j \in \{L, U\}$; and our estimates of the endpoints of the line $\mathcal{T}$ are given by

$$\widehat{\tau}^L_1 = \max\left\{0, 1 - \frac{1-\widehat{\tau}}{\widehat{g}^+}\right\}, \quad \widehat{\tau}^U_1 = \min\left\{1 - \frac{(1-\widehat{\tau}) \cdot \widehat{g}^-}{\widehat{g}^+}, \frac{\widehat{\tau}}{\widehat{g}^+}\right\},$$

$$\widehat{\tau}^L_0 = \min\left\{1, \frac{\widehat{\tau}}{1-\widehat{g}^+}\right\}, \qquad \widehat{\tau}^U_0 = \max\left\{0, \widehat{\tau} - \frac{(1-\widehat{\tau}) \cdot (\widehat{g}^+ - \widehat{g}^-)}{1-\widehat{g}^+}\right\}.$$

Our estimators of $\Gamma^L_{0,FRD}$ and $\Gamma^U_{0,FRD}$ from Theorem 2 are then given by

$$\widehat{\Gamma}^L_{0,FRD} = \inf_{t\in[0,1]} \widehat{\Gamma}^L_0(t) \quad \text{and} \quad \widehat{\Gamma}^U_{0,FRD} = \sup_{t\in[0,1]} \widehat{\Gamma}^U_0(t),$$

respectively. These estimators are consistent under general conditions, but in finite samples $\widehat{\Gamma}^L_{0,FRD}$ is downward biased for $\Gamma^L_{0,FRD}$, and $\widehat{\Gamma}^U_{0,FRD}$ is upward biased for $\Gamma^U_{0,FRD}$. These estimators thus tend to overstate the degree of partial identification, and produce conservative estimates of the identified set that are generally too wide. This has to be taken into account when interpreting results in any empirical setting.

Estimation under the conditions of Theorems 3 and 4 is conceptually similar to that under the conditions of Theorem 2 and 1, respectively, and thus we keep the discussion of

31

the empirical implementation of our refined bounds in these cases rather brief. Following the remark after Theorem 3, estimates $\widehat{\Gamma}^L_{0,FRD(a)}$ and $\widehat{\Gamma}^U_{0,FRD(a)}$ of $\Gamma^L_{0,FRD(a)}$ and $\Gamma^U_{0,FRD(a)}$ can be obtained by proceeding exactly as in the case of the basic fuzzy RD design, after redefining $\widehat{\tau}^L_1 = \widehat{\tau}^L_0 = \widehat{\tau}$. Estimates of the bounds $\Gamma^L_{0,FRD(b)}$ and $\Gamma^U_{0,FRD(b)}$ from Theorem 4 can be constructed as $\widehat{\Gamma}^L_{0,FRD(b)} = \widehat{\Delta}^L_0(\widehat{\tau}/\widehat{g}^+, 0)/\widehat{\Psi}_0(\widehat{\tau}/\widehat{g}^+)$ and $\widehat{\Gamma}^U_{0,FRD(b)} = \widehat{\Delta}^L_0(\widehat{\tau}/\widehat{g}^+, 0)/\widehat{\Psi}_0(\widehat{\tau}/\widehat{g}^+)$, respectively. Implementing one the of various refinements from Section 3.3 can also be accomplished along similar lines.

5 INFERENCE

In this section, we construct "manipulation-robust" confidence intervals for Γ_0 that are valid irrespective of the true value of τ. These confidence intervals apply to both sharp and fuzzy RD designs. We also propose an alternative approach to inference that is particularly suited for settings where researchers have strong prior beliefs that manipulation is either absent or at least very rare in their application, but where τ is only imprecisely estimated.

5.1 Confidence Intervals in the Sharp RD Design

We begin by considering "manipulation-robust" inference on Γ_0 in the sharp RD design. Note that the identified set for Γ_0 is an interval, and that we have estimates $\widehat{\Gamma}^L_{0,SRD}$ and $\widehat{\Gamma}^U_{0,SRD}$ of its lower and upper bounds. If these estimates were jointly asymptotically normal irrespective of the true value of τ, then our setup would be a special case of the one studied by Imbens and Manski (2004) and Stoye (2009), and we could use their proposed construction of a confidence interval. However, in our setup $\widehat{\Gamma}^L_{0,SRD}$ and $\widehat{\Gamma}^U_{0,SRD}$ are only jointly asymptotically normal if $\tau > 0$ (under appropriate regularity conditions; see the appendix). When $\tau = 0$ their limiting distribution is non-Gaussian, as the estimated level of manipulation $\widehat{\tau} = \max\{0, 1 - \widehat{f}^-/\widehat{f}^+\}$ fails to be asymptotically normal in this case.[9]

---

[9]Under standard regularity conditions we have that $\sqrt{nh}(\widehat{\tau} - \tau) \xrightarrow{d} \max\{0, Z\}$ if $\tau = 0$, where Z is a Gaussian random variable with mean zero.

Due to the discontinuity in the limiting distribution at $\tau = 0$, a Gaussian approximation to the distribution of $\widehat{\Gamma}^L_{0,SRD}$ and $\widehat{\Gamma}^U_{0,SRD}$ is poor in finite samples of any size if τ is not well-separated from zero. Moreover, it is well known that the standard bootstrap is unable to consistently estimate the limiting distribution in this case (Andrews, 2000). To obtain a confidence interval that is uniformly valid in the level of manipulation, we therefore propose to estimate the limiting distribution for a level of manipulation that is biased away from its point estimate $\widehat{\tau}$ into the least favorable direction, which is away from zero. This approach is analogous to the idea of moment selection in the moment inequality literature. Such a "biased" limiting distribution is Gaussian, but due to the complexity of its asymptotic variance we suggest estimating it via the bootstrap. Specifically, we construct a bootstrap distribution under which the bootstrap analogue of $\widetilde{\tau} = 1 - \widehat{f}^-/\widehat{f}^+$ is centered around $\max\{\widehat{\tau}, \kappa_n \widehat{s}_{\widetilde{\tau}}\}$, where $\widehat{s}_{\widetilde{\tau}}$ is the standard error of $\widetilde{\tau}$, and κ_n is a sequence of constants that slowly tends to infinity. Following much of the moment inequality literature, we choose $\kappa_n = \log(n)^{1/2}$ in this paper. The algorithm for our bootstrap is as follows.

1. Generate bootstrap samples $\{Y_{i,b}, D_{i,b}, X_{i,b}\}_{i=1}^n$, $b = 1, \ldots, B$ by sampling with replacement from the original data $\{Y_i, D_i, X_i\}_{i=1}^n$; for some large integer B.

2. Calculate $\widetilde{\tau}^*_b = 1 - \widehat{f}^-_b/\widehat{f}^+_b$, and put $\widehat{s}_{\widetilde{\tau}}$ as the sample standard deviation of $\{\widetilde{\tau}^*_b\}_{b=1}^B$.

3. Calculate $\widetilde{\tau}_b = \widetilde{\tau}^*_b - \widetilde{\tau} + \max\{\widehat{\tau}, \kappa_n \widehat{s}_{\widetilde{\tau}}\}$ and $\widehat{\tau}_b = \max\{\widetilde{\tau}_b, 0\}$.

4. For $j \in \{U, L\}$, calculate $\widehat{\Gamma}^j_{0,SRD,b}$ using the redefined estimate $\widehat{\tau}_b$ from the previous step, and put $\widehat{s}^j$ as the sample standard deviation of $\{\widehat{\Gamma}^j_{0,SRD,b}\}_{b=1}^B$.

Now define $\widehat{\Gamma}^{L*}_{0,SRD} = \widehat{m}^{L+}(\widehat{\tau}^*) - \widehat{m}^-$ and $\widehat{\Gamma}^{U*}_{0,SRD} = \widehat{m}^{U+}(\widehat{\tau}^*) - \widehat{m}^-$, with $\widehat{\tau}^* = \max\{\widetilde{\tau}, \kappa_n \widehat{s}_{\widetilde{\tau}}\}$, as estimates of the lower and upper bound on Γ_0 using a conservatively biased estimate of τ. A uniformly valid $1 - \alpha$ confidence interval for Γ_0 is then given by

$$\mathcal{C}^{SRD}_{1-\alpha} \equiv \left[\widehat{\Gamma}^{L*}_{0,SRD} - r_\alpha \cdot \widehat{s}^L, \ \widehat{\Gamma}^{U*}_{0,SRD} + r_\alpha \cdot \widehat{s}^U \right],$$

where r_α is the value that solves the equation

$$\Phi\left(r_\alpha + \frac{\widehat{\Gamma}_{0,SRD}^{U*} - \widehat{\Gamma}_{0,SRD}^{L*}}{\max\{\widehat{s}^L, \widehat{s}^U\}}\right) - \Phi(-r_\alpha) = 1 - \alpha,$$

and $\Phi(\cdot)$ is the CDF of the standard normal distribution. Note that for $\alpha < 0.5$ the critical value r_α satisfies $\Phi^{-1}(1 - \alpha) < r_\alpha < \Phi^{-1}(1 - \alpha/2)$; and that r_α is, all else equal, decreasing in the length $\widehat{\Gamma}_{0,SRD}^{U*} - \widehat{\Gamma}_{0,SRD}^{L*}$ of the estimated identified set. As pointed out by Imbens and Manski (2004), this choice of critical value ensures that the confidence interval has good finite-sample coverage properties irrespective of the length of the identified set.

5.2 Confidence Intervals in the Fuzzy RD Design

Constructing a confidence interval for Γ_0 under the conditions of Theorem 2 involves some further complications due to the presence of optimization operators in the definitions of the upper and lower bounds. Our setup is different from the one in Chernozhukov, Lee, and Rosen (2013), who study inference on *intersection bounds* of the form $[\sup_v \theta(v), \inf_v \theta(v)]$. Our problem is an example of *union bounds*, where the role of the inf and the sup operator in the definition of the identified set is reversed. We are not aware of any general results regarding inference for union bounds, but the problem can be addressed using the intersection-union testing principle of Berger (1982).

The main idea is the following. Recall the definition of the set of feasible values for (τ_1, τ_0), which is $\mathcal{T} = \{\eta(t) \equiv (\eta_1(t), \eta_0(t)) : t \in [0, 1]\}$ with $\eta_d(t) = \tau_d^L + t \cdot (\tau_d^U - \tau_d^L)$, and note that our construction of the bounds on Γ_0 involved showing that if we knew that $(\tau_1, \tau_0) = \eta(t)$ for some fixed value of $t \in [0, 1]$, the sharp lower and upper bound on Γ_0 would be $\Gamma_0^L(t)$ and $\Gamma_0^U(t)$ respectively. Now suppose that for every $t \in [0, 1]$ we had a $1 - \alpha$ confidence interval $\mathcal{C}_{1-\alpha}^{FRD}(t)$ for Γ_0 that was valid if the true value of (τ_1, τ_0) was equal to $\eta(t)$. Then the intersection-union testing principle implies that $\mathcal{C}_{1-\alpha}^{FRD} = \cup_{t \in [0,1]} \mathcal{C}_{1-\alpha}^{FRD}(t)$ is an asymptotically valid $1 - \alpha$ confidence interval for Γ_0. That is, a candidate value for Γ_0 is outside of $\mathcal{C}_{1-\alpha}^{FRD}$ if

and only if it is outside of $\mathcal{C}_{1-\alpha}^{FRD}(t)$ for all $t \in [0,1]$. An important feature of this approach is that both the "fixed t" and the overall confidence interval have level $1 - \alpha$: there is no need for a multiplicity adjustment to account for the fact that we are implicitly testing a continuum of hypotheses.

The problem of constructing a confidence interval for Γ_0 under the conditions of Theorem 2 is thus reduced to constructing a confidence interval for Γ_0 under the assumption that $(\tau_1, \tau_0) = \eta(t)$ for some known $t \in [0,1]$. Such a confidence interval can be based on the estimates $\widehat{\Gamma}_0^L(t)$ and $\widehat{\Gamma}_0^U(t)$. For reasons analogous to those in the previous section, the limiting distribution of these estimators is discontinuous at $\tau = 0$. We therefore consider a modified bootstrap procedure to estimate of the distribution of $\widehat{\Gamma}_0^L(t)$ and $\widehat{\Gamma}_0^U(t)$ in which the value of τ is again biased away from zero. The algorithm is as follows.

1. Generate bootstrap samples $\{Y_{i,b}, D_{i,b}, X_{i,b}\}_{i=1}^n$, $b = 1, \ldots, B$ by sampling with replacement from the original data $\{Y_i, D_i, X_i\}_{i=1}^n$; for some large integer B.

2. Calculate $\widetilde{\tau}_b^* = 1 - \widehat{f}_b^- / \widehat{f}_b^+$, and put $\widehat{s}_{\widetilde{\tau}}$ as the sample standard deviation of $\{\widetilde{\tau}_b^*\}_{b=1}^B$.

3. Calculate $\widetilde{\tau}_b = \widetilde{\tau}_b^* - \widetilde{\tau} + \max\{\widehat{\tau}, \kappa_n \widehat{s}_{\widetilde{\tau}}\}$ and $\widehat{\tau}_b = \max\{\widetilde{\tau}_b, 0\}$.

4. For $j \in \{U, L\}$, calculate $\widehat{\Gamma}_{0,b}^j(t)$ using the redefined estimate $\widehat{\tau}_b$ from the previous step, and put $\widehat{s}^j(t)$ as the sample standard deviation of $\{\widehat{\Gamma}_{0,b}^j(t)\}_{b=1}^B$.

Now define $\widehat{\Gamma}_0^{L*}(t)$ and $\widehat{\Gamma}_0^{U*}(t)$ exactly as $\widehat{\Gamma}_0^L(t)$ and $\widehat{\Gamma}_0^U(t)$, with the exception that $\widehat{\tau}^* = \max\{\widetilde{\tau}, \kappa_n \widehat{s}_{\widetilde{\tau}}\}$ is used in place of $\widehat{\tau}$ in the calculations. Then our "fixed t" confidence interval for Γ_0 with nominal level $1 - \alpha$ is given by

$$\mathcal{C}_{1-\alpha}^{FRD}(t) = \left[\widehat{\Gamma}_0^{L*}(t) - r_\alpha(t) \cdot \widehat{s}^L(t), \ \widehat{\Gamma}_0^{U*}(t) + r_\alpha(t) \cdot \widehat{s}^U(t) \right],$$

where $r_\alpha(t)$ is the value that solves the equation

$$\Phi \left(r_\alpha(t) + \frac{\widehat{\Gamma}_0^{U*}(t) - \widehat{\Gamma}_0^{L*}(t)}{\max\{\widehat{s}^L(t), \widehat{s}^U(t)\}} \right) - \Phi(-r_\alpha(t)) = 1 - \alpha,$$

35

and $\Phi(\cdot)$ is the CDF of the standard normal distribution. The final intersection-union confidence interval for Γ_0 is then given by

$$\mathcal{C}_{1-\alpha}^{FRD} = \left[\inf_{t\in[0,1]} \left(\widehat{\Gamma}_0^L(t) - r_\alpha(t) \cdot \widehat{s}^L(t) \right), \sup_{t\in[0,1]} \left(\widehat{\Gamma}_0^U(t) + r_\alpha(t) \cdot \widehat{s}^U(t) \right) \right].$$

The construction of this confidence interval does not account for the fact that the limiting distribution of the $\widehat{\Gamma}_0^j(t)$ is not only discontinuous at $\tau = 0$, but also at those values of τ under which (i) the denominator $\Psi_0(\eta_1(t))$ in the definition of $\Gamma_0^j(t)$ becomes equal to zero; and (ii) one of the various max and min operators in the definition of the function $\eta_d(\cdot)$ becomes binding. Our confidence interval might thus not have the correct size if the true value of τ is close to either of these discontinuities. On the other hand, our approach should work well if

$$\tau \ll \frac{\mathbb{E}(D_i|X_i = c^+) - \mathbb{E}(D_i|X_i = c^-)}{1 - \mathbb{E}(D_i|X_i = c^-)}, \tag{5.1}$$

as this rules out issue (i), and if either

$$\tau \ll 1 - \mathbb{E}(D_i|X_i = c^+) \text{ or } \tau \gg 1 - \mathbb{E}(D_i|X_i = c^+), \tag{5.2}$$

as this rules out issue (ii). Both conditions appear reasonable for many applications, including the one we study below. To keep the exposition simply, we therefore do not include any "safeguards" against such cases into our bootstrap procedure.

5.3 An Alternative Strategy for Inference

A price one has to be pay for the uniform validity of the above-mentioned confidence intervals over all plausible values of τ is that they might be quite conservative in settings where τ is close or equal to zero, but measured rather imprecisely. Such confidence intervals accurately reflect the uncertainty about Γ_0 from the point of view of a researcher that is *a priori* agnostic about what the true value of τ is, and wants to ensure correct coverage properties of the

confidence interval under all values of τ that are not ruled out with near-certainty by the data. In many applications, however, researchers have strong prior beliefs that manipulation is either completely absent or at least quite rare in their setting. When this view is confirmed by a point estimate of τ that is close to zero, imposing the restriction that τ is equal to $\kappa_n \widehat{s}_\tau$ in the bootstrap world might seem overly pessimistic to them.

We therefore consider an alternative strategy for inference that can be used in such situations. The idea is as follows. Suppose that the main interest of the researcher is in testing the hypothesis that $\Gamma_0 = G$ against the alternative that $\Gamma_0 \neq G$, where G is some constant. We then propose to construct confidence intervals $\mathcal{C}_{1-\alpha}(\tau^*)$ for Γ_0 that are derived under the (counterfactual) assumption that it is known that $\tau = \tau^*$. For $\tau^* = 0$, this yields the usual "no manipulation" confidence interval, and generally $\mathcal{C}_{1-\alpha}(\tau^*)$ becomes wider as τ^* increases. The researcher can then plot the upper and lower boundary of $\mathcal{C}_{1-\alpha}(\tau^*)$ as a function of τ^*, and check graphically for which levels of manipulation the point $\Gamma_0 = G$ is contained in the confidence interval. For example, suppose that $G \notin \mathcal{C}_{1-\alpha}(0)$, but that $G \in \mathcal{C}_{1-\alpha}(\tau^*)$ for $\tau^* \geq 0.1$. Then the researcher can report that in his preferred "no manipulation" specification the null hypothesis that $\Gamma_0 = G$ is rejected at the critical level α, and that at least a 10% level of manipulation around the cutoff would be needed to reverse this result.

We believe that such an exercise is a useful robustness check for every RD study, and a reasonable way to visualize the impact of potential manipulation when the above-mentioned "manipulation-robust" confidence intervals are deemed to be too conservative. We now describe the necessary modifications of our bootstrap algorithms for this approach.

"Fixed Manipulation" Bootstrap in the Sharp RD Design

1. For $\tau^* \in [0,1]$, define $\widehat{\Gamma}_{0,SRD}^L(\tau^*) = \widehat{m}^{L+}(\tau^*) - \widehat{m}^-$ and $\widehat{\Gamma}_{0,SRD}^U(\tau^*) = \widehat{m}^{U+}(\tau^*) - \widehat{m}^-$.

2. Generate bootstrap samples $\{Y_{i,b}, D_{i,b}, X_{i,b}\}_{i=1}^n$, $b = 1, \ldots, B$ by sampling with replace-

ment from the original data $\{Y_i, D_i, X_i\}_{i=1}^n$; for some large integer B.

3. For $j \in \{U, L\}$, calculate $\widehat{\Gamma}_{0,SRD,b}^j(\tau^*)$, and put $\widehat{s}^j(\tau^*)$ as the sample standard deviation of $\{\widehat{\Gamma}_{0,SRD,b}(\tau^*)^j\}_{b=1}^B$.

4. Compute the $1 - \alpha$ confidence interval

$$\mathcal{C}_{1-\alpha}^{SRD}(\tau^*) \equiv \left[\widehat{\Gamma}_{0,SRD}^L(\tau^*) - r_\alpha(\tau^*) \cdot \widehat{s}^L(\tau^*), \ \widehat{\Gamma}_{0,SRD}^U(\tau^*) + r_\alpha(\tau^*) \cdot \widehat{s}^U(\tau^*)\right],$$

where $r_\alpha(\tau^*)$ is the value that solves the equation

$$\Phi\left(r_\alpha(\tau^*) + \frac{\widehat{\Gamma}_{0,SRD}^U(\tau^*) - \widehat{\Gamma}_{0,SRD}^L(\tau^*)}{\max\{\widehat{s}^L(\tau^*), \widehat{s}^U(\tau^*)\}}\right) - \Phi(-r_\alpha(\tau^*)) = 1 - \alpha.$$

"Fixed Manipulation" Bootstrap in the Fuzzy RD Design

1. For $\tau^* \in [0, 1]$ and $t \in [0, 1]$, define $\widehat{\Gamma}_0^L(\tau^*, t)$ and $\widehat{\Gamma}_0^U(\tau^*, t)$ exactly as $\widehat{\Gamma}_0^L(t)$ and $\widehat{\Gamma}_0^U(t)$, with the exception that τ^* is used in place of $\widehat{\tau}$ in the calculations.

2. Generate bootstrap samples $\{Y_{i,b}, D_{i,b}, X_{i,b}\}_{i=1}^n$, $b = 1, \ldots, B$ by sampling with replacement from the original data $\{Y_i, D_i, X_i\}_{i=1}^n$; for some large integer B.

3. For $j \in \{U, L\}$, calculate $\widehat{\Gamma}_{0,b}^j(\tau^*, t)$, and put $\widehat{s}^j(\tau^*, t)$ as the sample standard deviation of $\{\widehat{\Gamma}_{0,b}^j(\tau^*, t)\}_{b=1}^B$.

4. Compute the $1 - \alpha$ confidence interval

$$\mathcal{C}_{1-\alpha}^{FRD}(\tau^*)$$
$$= \left[\inf_{t \in [0,1]} \left(\widehat{\Gamma}_0^L(\tau^*, t) - r_\alpha(\tau^*, t) \cdot \widehat{s}^L(\tau^*, t)\right), \ \sup_{t \in [0,1]} \left(\widehat{\Gamma}_0^U(\tau^*, t) + r_\alpha(\tau^*, t) \cdot \widehat{s}^U(\tau^*, t)\right)\right],$$

where $r_\alpha(\tau^*, t)$ is the value that solves the equation

$$\Phi\left(r_\alpha(\tau^*, t) + \frac{\widehat{\Gamma}_0^U(\tau^*, t) - \widehat{\Gamma}_0^L(\tau^*, t)}{\max\{\widehat{s}^L(\tau^*, t), \widehat{s}^U(\tau^*, t)\}}\right) - \Phi(-r_\alpha(\tau^*, t)) = 1 - \alpha.$$

6 Empirical Application

In this section, we apply the estimation and inference procedures developed above to bound the effect of unemployment insurance (UI) takeup around an eligibility cutoff in Brazil.

UI is a good example of a program for which our approach is useful. There is a whole literature that is interested in the effect of UI on the time it takes for displaced workers to find a new job. RD designs are natural empirical strategies to estimate this effect because UI programs typically feature discontinuities in eligibility based on the level of some running variable, such as the number of months of employment prior to layoff. At the same time, there is another literature studying how UI may increase turnover and layoffs when benefits are not fully experience-rated, i.e. when firms are not paying at the margin for the benefits drawn by their workers (e.g. Feldstein, 1976; Anderson and Meyer, 2000). In practice, UI benefits are never fully experience-rated. The possibility that manipulation of the running variable could invalidate the standard assumption for an RD design is thus a serious concern in the UI context. This concern may be particularly severe in countries with high labor market informality, such as Brazil. The utility costs of being (reported as) laid off when eligible for UI may be relatively low for some workers if they can work informally, possibly in the very same firm that they are supposedly being laid off from. Finally, our key identifying assumption ("one-sided manipulation") is likely to apply: all displaced workers are likely to have a (weak) preference for being eligible for (additional) UI benefits.[10]

The main purpose of our empirical exercise is to illustrate the applicability of our approach. However, its results are relevant for their own sake. The effect of UI on non-formal-employment duration (i.e., the time between two formal jobs), for which we estimate bounds below, is an important input to the evaluation of the optimal design of UI programs.[11]

---

[10]They always have the choice to not take up UI.

[11]See for instance Baily (1978) and Chetty (2008). The literature typically refers to the effect on non-employment duration because it considers countries where all jobs are assumed to be formal.

Specifically, the local average treatment effect at an eligibility cutoff would be an important input for the welfare effect of marginal changes in the location of the cutoff. Moreover, UI programs have been adopted or considered in a number of developing countries. Yet, the existing evidence for developing countries, whose labor markets present distinctive features such as a high prevalence of informal work opportunities, remains limited (Gerard and Gonzaga, 2014).

6.1 Institutional Details, Data, and Sample Selection

In Brazil, a worker who is reported as involuntarily laid off from a private formal job is eligible for UI under two conditions. First, she must have at least six months of continuous job tenure at layoff. Second, there must be at least 16 months between the date of her layoff and the date of the last layoff after which she applied for and drew UI benefits. We focus on the eligibility cutoff created by this second condition.[12] Eligible workers can draw three, four, or five months of UI benefits if they accumulated more than 6, 12, or 24 months of formal employment in the 36 months prior to layoff. Their benefit level depends nonlinearly on their average wage in the three months prior to layoff. The replacement rate is 100% at the bottom of the wage distribution but already down to 60% for a worker who earned three times the minimum wage (the benefit schedule is shown in the Appendix). Finally, UI benefits are not experience-rated in Brazil as they are financed by a tax on firms' total sales.

Our empirical analysis relies on two administrative datasets. The first one is a longitudinal matched employee-employer dataset covering by law the universe of formal employees. Every year, firms must report all workers formally employed at some point during the previous calendar year. The data include information on wage, tenure, age, race, gender, education, sector of activity, and establishment size. Importantly, the data also include hiring and

---

[12]There is evidence of manipulation around the six-month cutoff as well (Gerard and Gonzaga, 2014). The 16-month cutoff is more arbitrary and thus less likely to coincide with other possible discontinuities. For instance, six months of job tenure may be a salient milestone for evaluating employees' performance.

separation dates, as well as the reason for separation. The second dataset is the registry of all UI payments. Individuals can be matched in both datasets as they are identified through the same ID number. Combining the datasets, we know whether any displaced formal employee is eligible for UI, how many UI payments she draws, and when she is formally reemployed.

Our sample of analysis is constructed as follows. First, we consider all workers, 18 to 55 years old, who lost a private full-time formal job and drew UI benefits between 2002 and 2010. Second, workers who find a new formal job before exhausting their benefits are entitled to draw the remaining benefits after a new layoff, even if it occurred before the 16-month cutoff. We thus limit the sample to workers who exhausted their UI benefits (a large majority; see Gerard and Gonzaga, 2014) such that the change in eligibility at the 16-month cutoff is sharp. To do so, we select workers who drew the maximum number of UI benefits because we can only measure the number of UI benefits a worker is eligible for imprecisely in our data (see Gerard and Gonzaga, 2014, for details). Third, we drop workers whose layoff date fell after the 28th of a month. This is because policy rules create bunching in the layoff density at the 16-month cutoff even in the absence of manipulation among these workers. For instance, all workers laid off between October 29th and 31st in 2007, became eligible on February 28th in 2009, because there are only 28 days in February. Fourth, we follow each worker in our sample until her first subsequent layoff. We keep workers who had more than six month of job tenure at layoff (which is the first eligibility condition) and whose new layoff date fell within 120 days of the 16-month eligibility cutoff. Finally, we follow workers for at least one year after the new layoff date. As a result, we drop workers whose new layoff date fell in 2010. Our sample ultimately consists of workers with a relatively high attachment to the labor force, a relatively high turnover rate, and a relatively high ability to find a new formal job rapidly.[13] These are not the characteristics of the average formal employee or UI

---

[13]They were previously eligible for five months of UI, so they had accumulated 24 months of formal employment within a 36-month window. They were laid off twice in 16 months and they had accumulated at least six month of continuous tenure at layoff. Therefore, they found a job relatively quickly after their first

beneficiary in Brazil, but characteristics of workers for whom the 16-month cutoff may be binding.

6.2 Graphical Evidence

Figure 6.2 displays graphically some of the patterns in our data. Observations are aggregated by day between the layoff date and the 16-month eligibility cutoff. Panel A provides evidence of manipulation. The share of observations located just to the right of the cutoff is higher than the share located just to the left of the cutoff by about 9%. Several types of behavior could lead to this pattern. Some workers may provoke their layoff (e.g. shirk) or ask their employer to report their quit as a layoff once they are eligible for UI. As in more developed countries, employers may put some workers on temporary layoff once they are eligible for UI, but possibly keeping them on an informal payroll in the context of developing countries. Alternatively, workers laid off before the 16-month cutoff may lobby their employers to (officially) lay them off after the cutoff, a behavior that could be modeled similarly to the example of a second home visit in Section 2. The manipulation in panel A is likely the result of a combination of these different types of behaviors (and possibly others).

Panels B-F then display similar graphs for various outcome variables. Panel B suggests that workers were at least partially aware of the eligibility rule, a likely condition for manipulation to take place. The share of workers applying for UI benefits is monotonically increasing to the left of the cutoff, but it jumps by about 40%-points at the cutoff. Panel C shows that the eligibility rule was enforced. The share of workers drawing some UI benefits is close to zero to the left of the cutoff, but takeup jumps to about 80% at the cutoff. Eligible workers drew on average 3.2 months of UI benefits (panel D), implying that UI takers drew on average four months of UI benefits (= 3.2/.8). Panel E displays the share of workers who

---

layoff. Indeed, Gerard and Gonzaga (2014) document that about 50% of workers eligible for five months of UI benefits remain without a formal job one year after layoff.
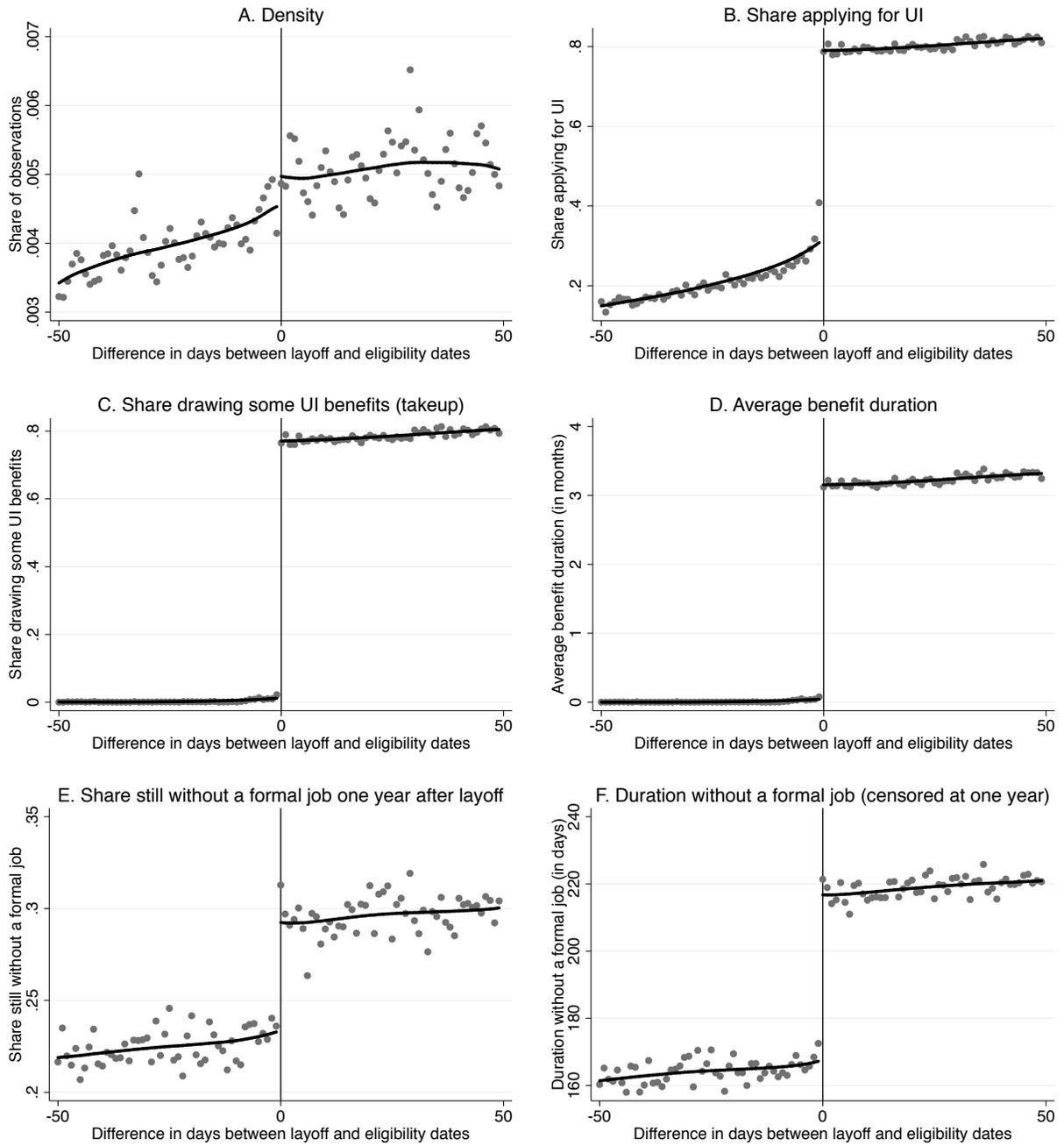
Figure 6.2: Graphical evidence for our empirical application

The figure displays the mean of different variables on each side of the cutoff by day between the layoff and eligibility dates, as well as local linear regressions on each side of the cutoff using an edge kernel and a bandwidth of 30 days.

remain without a formal job for at least one year after layoff. It increases discontinuously from 23.5% to 29% at the cutoff. Relatedly, panel F shows a clear discontinuity in non-formal-employment duration (censored at one year), which increases from about 170 days to 220 days at the cutoff. Note that the discontinuity for the outcome in panel E implies that the effect on non-formal-employment duration censored at a later point (e.g. two years) would be larger (in levels, but not necessarily in percents). Our estimates below could easily be obtained using other censoring points at the cost of a smaller sample.

The discontinuities in panels E and F could be due to a treatment effect, but they could also be due to a selection bias. Workers on each side of the cutoff may have different potential outcomes in the presence of manipulation. For instance, we show below that always-assigned workers have different observable characteristics than potentially-assigned workers. The estimators developed above allow us to bound treatment effects for potentially-assigned workers, despite the possibility of sample selection.

6.3 Estimates

We implement our estimation and inference procedures for the outcome in panel F, non-formal-employment duration. We censor the outcome at one year so that we can measure it for all workers in our sample. The censoring also serves an illustrative purpose. The share of right-censored observations is always larger than the estimated share of always-assigned units. Our lower bound thus always truncates a discrete outcome (whether an observation is right censored or not), while our upper bound essentially truncates a continuous outcome.[14]

Results are presented in Table 1 for an edge kernel (Cheng, Fan, and Marron, 1997) and a bandwidth of 30 days around the cutoff.[15] For bounds in the FRD case that involve numerical optimization, we use a grid search to look for the infimum and supremum using 51

---

[14]The distribution of outcomes just to the left and right of the cutoff is shown in the Appendix.

[15]We do not have theoretical results on the optimal bandwidth for the estimation of our bounds. Our estimates are similar if we use bandwidths of 10 or 50 days (available upon request).

Table 1: The effect of UI on non-formal employment duration at the eligibility cutoff

	Point/Interval Estimate	95% Conf. Interval
A. Basic Inputs		
Share of always-assigned workers (τ)	.0882	[.0665,.1099]
Increase in UI take-up at the cutoff	.7583	[.7511,.7655]
B. ITT/SRD estimates		
ITT - Ignoring Manipulation	49.27	[45.99,52.54]
ITT - Standard Bounds (Th.1)	[34.86,69.58]	[30.37,74.83]
ITT - Bounds using covariates (Cor.1)	[34.87,69.35]	[30.54,74.44]
C. LATE/FRD estimates		
LATE - Ignoring Manipulation	64.97	[60.81,69.13]
LATE - Standard Bounds (Th.2)	[45.09,86.36]	[39.88,92.07]
LATE - Bounds using refinement A (Th.3)	[45.99,85.65]	[40.17,90.77]
LATE - Bounds using refinement B (Th.4)	[47.38,85.48]	[41.58,90.57]
LATE - Bounds using covariates (Cor.2)	[45.09,86.18]	[39.60,91.58]

Total number of observations within our bandwidth of 30 days around the cutoff: 120,862. The counterfactual value of the outcome at the cutoff for potentially-assigned workers ($\widehat{m}^-$) was estimated at 167.4. Confidence intervals have nominal level of 95% and are based on 500 bootstrap samples. Bounds that use covariates only use gender.

values for $t \in [0, 1]$. "Manipulation-robust" confidence intervals are based on 500 bootstrap samples. We note that in our setting the value of τ seems to be well-separated from zero, so that the safeguards that ensure uniform validity of these confidence intervals for small values of τ are not of any practical importance here.

Panel A of Table 1 first reports estimates of key inputs for our bounds. Always-assigned workers are estimated to account for 8.8% of observations just to the right of the cutoff and UI takeup is estimated to increase by 76%-points at the cutoff.

We then report results from two types of exercises. First, we consider a SRD design in which *UI eligibility* is defined as the treatment of interest (panel B). The causal effect on the outcome can be interpreted as an intention-to-treat (ITT) parameter in this case. Second, we consider the usual FRD design with *UI takeup* as the treatment (panel C). Naive RD esti-

mates that assume no manipulation yield a 49.3-day increase in non-formal-employment duration from being eligible for UI (ITT/SRD) and a 65-day increase in non-formal-employment duration from UI takeup (LATE/FRD). Those estimates, however, confound treatment effects and selection bias. Table 1 thus also provides estimates of our bounds for the treatment effects, using the SRD formulas (resp. the FRD formulas) for the ITT (resp. the LATE). The bounds are relatively large, which is not surprising given the extent of manipulation. Point estimates for the standard upper bounds are 69.6 days (ITT) and 86.4 days (LATE). Point estimates for the standard lower bounds are 34.9 days (ITT) and 45.1 days (LATE).

Table 1 also provides estimates that use behavioral assumptions (our refinements A and B) or covariates (here, gender) to tighten our bounds. These refinements have no meaningful additional identifying power in this case, yielding bounds that are qualitatively very similar to the standard ones. This is because, although mean outcomes differ, tails of the outcome distribution do not differ much by takeup status or covariate.

Finally, we present the results of two additional exercises. First, we illustrate the alternative strategy for inference that we recommend when researchers have strong prior beliefs that manipulation is either absent or very rare in their setting. We display point estimates and confidence intervals for our bounds in the SRD and FRD cases for various fixed levels of the extent of manipulation (share of always-assigned units). Figure 6.3 shows that inference on treatment effects can be quite sensitive to the extent of manipulation. For instance, the width of the confidence intervals more than doubles when we assume a small degree of manipulation (a share of always-assigned units of 2.5%) rather than no manipulation. This illustrates the importance of taking into account the possibility of manipulation even when the McCrary (2008) test fails to reject the null hypothesis of no manipulation.

Second, we estimate the characteristics of potentially-assigned and always-assigned workers. This could be useful to target policies aimed at mitigating manipulation in the timing
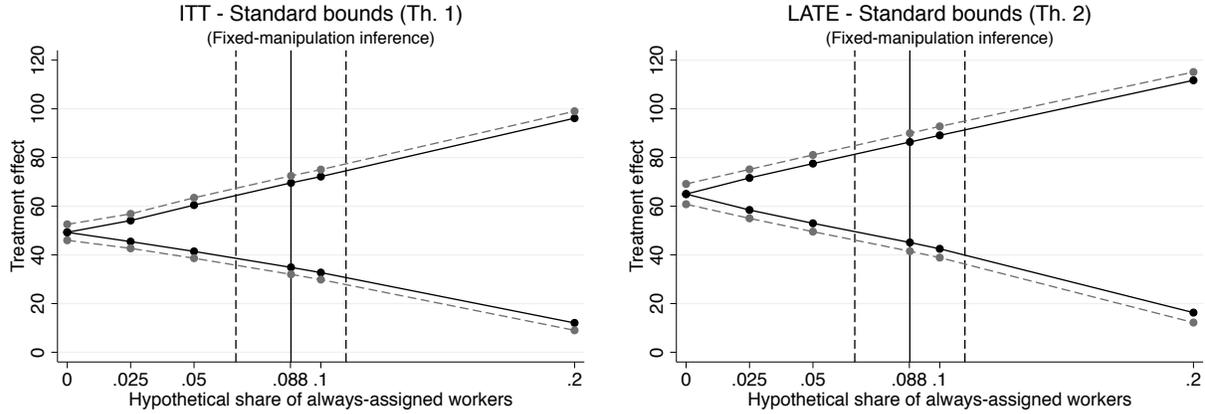
Figure 6.3: Fixed-manipulation inference for our empirical application

The figure displays point estimates and confidence intervals for our bounds for fixed levels of the degree of manipulation. The solid vertical line (resp. dashed vertical lines) corresponds to our point estimate (confidence interval) for the extent of manipulation (see Table 1).

of layoff. Table 2 displays the estimated difference in the mean of workers' characteristics at the cutoff (column 1), as would be typically presented in an RD analysis. The associated graphs are presented in the Appendix. We find significant evidence of selection in terms of gender, wage (and thus replacement rate), sector of activity, and firm size. Columns (2) and (3) then display the estimated means for potentially-assigned and always-assigned workers. Always-assigned workers are 14%-points less likely to be male, earned on average .22 log point less, are 19%-points more likely (resp. 24%-point less likely) to come from the industrial sector (resp. service sector), and are 14%-point more likely to come from a small firm than the potentially-assigned workers. The large difference in gender composition motivated the choice of using gender to construct bounds in Table 1.

In sum, a naive RD design that ignores manipulation estimates that UI takeup increases non-formal-employment in the first year after layoff by about 39% (= 65/167) at the eligibility cutoff. However, we find significant evidence of manipulation and selection at the cutoff. Our bounds imply that the magnitude of this estimate may be heavily affected by selection,

Table 2: Characteristics of always- and potentially-assigned workers

	Difference at cutoff (1)	Potentially-assigned (2)	Always-assigned (3)
Share male	-.0122	.7021	.5642
	[-.0229,-.0014]	[.694,.7103]	[.4415,.6869]
Average age	-.1559	32.59	30.82
	[-.3461,.0342]	[32.44,32.73]	[28.70,32.94]
Average years of education	.0092	9.066	9.171
	[-.0566,.075]	[9.015,9.118]	[8.463,9.878]
Average tenure	.0107	8.513	8.635
	[-.0276,.049]	[8.49,8.536]	[8.203,9.067]
Average log wage (R$)	-.0195	6.682	6.461
	[-.0328,-.0063]	[6.672,6.692]	[6.307,6.614]
Average replacement rate	.0067	.7322	.8085
	[.0026,.0108]	[.7292,.7353]	[.7602,.8567]
Share from	.0044	.3624	.4119
commercial sector	[-.0067,.0155]	[.354,.3708]	[.2896,.5342]
Share from	-.0002	.0996	.0976
construction sector	[-.0079,.0075]	[.0938,.1053]	[.0128,.1825]
Share from	.017	.2376	.4307
industrial sector	[.0058,.0282]	[.2295,.2458]	[.304,.5574]
Share from	-.0212	.3004	.0597
service sector	[-.0331,-.0094]	[.2916,.3092]	[-.0817,.2011]
Share from small firm	.0121	.3931	.5298
(< 10 employees)	[.0010,.0231]	[.3843,.4019]	[.4041,.6555]

Total number of observations within our bandwidth of 30 days around the cutoff: 120862. Numbers in square brackets are 95% confidence intervals calculated by adding ±1.96×standard error to the respective point estimate, where standard errors are calculated via the bootstrap with 500 replications.

but that the treatment effect for potentially-assigned units is an increase of at least 24% (lower limit of the confidence interval). UI takeup thus leads to significant delay in formal reemployment at the existing cutoff. Of course, this estimate is particular to the sample of workers for whom the cutoff may be binding.[16] It must also be put in perspective with the welfare gains from providing these workers with insurance. For example, the welfare gain is arguably lower for workers with high turnover rates, for whom layoff is an expected event.

7 Conclusions

In this paper, we propose a partial identification approach to deal with the issue of potentially manipulated running variables in RD designs. We show that while the data are unable to uniquely pin down treatment effects if manipulation occurs, they are generally still informative in the sense that they imply bounds on the value of interesting causal parameters in both sharp and fuzzy RD designs. Our main contribution is to derive and explicitly characterize these bounds. We also propose methods to estimate our bounds in practice, and discuss how to construct confidence intervals for treatment effects that have good coverage properties. The approach is illustrated with an application to the Brazilian unemployment insurance (UI) system. We recommend the use of our approach in applications irrespective of the outcome of McCrary's (2008) test for manipulation.

Our approach can also be useful for RD designs where running variables are not manipulated. Suppose for example that the probability of missing outcomes changes discontinuously at the cutoff. This could be the case if outcomes are based on surveys, and units are easier to locate and survey if they were assigned to the program. Our approach could be used to partially identify causal effects of the program at the cutoff for units whose outcomes would not be missing in the absence of program assignment. Another possibility is that

---

[16]These workers may be particularly responsive to the incentives created by UI given that they have a relatively high ability to find a new formal job rapidly (based on past employment history).

the characteristics of a program (e.g. the level of UI benefits) change at the cutoff rather than program assignment per se. In this case, a standard RD design will not identify the causal effect of program characteristics conditional on program takeup if program takeup is a function of program characteristics. Our approach could be used to partially identify such a causal effect for units that would takeup the program in absence of the change in program characteristics.

A Proofs of Identification Results

A.1 Proof of Lemma 1

Since the density of X_i is continuous around c among potentially-assigned units by Assumption 2, we have that $f_{X|M=0}(c^-) = f_{X|M=0}(c^+)$, and therefore $f_X(c^+) = (1 - \mathbb{P}(M = 1)) f_{X|M=0}(c^-) + \mathbb{P}(M = 1) f_{X|M=1}(c^+)$. Since $f_{X|M=1}(x) = 0$ for $x < c$ by Assumption 3, we also have that $f_X(c^-) = (1 - \mathbb{P}(M = 1)) f_{X|M=0}(c^-)$. Hence $(f_X(c^+) - f_X(c^-))/f_X(c^+) = f_{X|M=1}(c^+)\mathbb{P}(M = 1)/f_X(c^+) = \tau$, where the last equality follows from Bayes' Theorem. □

A.2 Proof of Lemma 2

By Assumption 1(i) and the law of total probability, our model implies that $\mathbb{E}(D_i|X_i = c^+) \cdot (1 - \tau_1)/(1 - \tau) > \mathbb{E}(D_i|X_i = c^-)$ and $\tau = \tau_1 \cdot \mathbb{E}(D_i|X_i = c^+) + \tau_0 \cdot (1 - \mathbb{E}(D_i|X_i = c^+))$. By construction, any point $(\tau_1, \tau_0) \notin \mathcal{T}$ is incompatible with at least one of these two restrictions. It thus remains to be shown that any point $(\tau_1, \tau_0) \in \mathcal{T}$ is compatible with our model and the observed joint distribution of (Y, D, X). Note that it suffices to consider the latter distribution for $X \in (c - \epsilon, c + \epsilon)$ for some small $\epsilon > 0$, as our model has no implications for the distribution of observables outside that range. Let $(\tilde{Y}(1), \tilde{Y}(0), \tilde{D}^+, \tilde{D}^-, \tilde{M}, \tilde{X})$ be a random vector taking values on the support of $(Y(1), Y(0), D^+, D^-, M, X)$, and define $\tilde{D}$ and $\tilde{Y}$ analogous to D and Y in our Section 2.1. We now construct a particular joint distribution of $(\tilde{Y}(1), \tilde{Y}(0), \tilde{D}^+, \tilde{D}^-, \tilde{M}, \tilde{X})$. For

$x \in (c - \epsilon, c + \epsilon)$, let

$$f_{\tilde{X}}(x) = f_X(x) \quad \text{and} \quad \mathbb{P}(\tilde{M} = 1 | \tilde{X} = x) = \begin{cases} 1 - f_X(c^-)/f_X(x) & \text{if } x \geq c \\ 0 & \text{if } x < c. \end{cases}$$

Moreover, let

$$\mathbb{P}(\tilde{D}^- = 0, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 0) = \begin{cases} \mathbb{P}(D = 1 | X = x) \cdot \frac{1 - \tau_1}{1 - \tau} - \mathbb{P}(D = 1 | X = c^-) & \text{if } x \geq c, \\ \mathbb{P}(D = 1 | X = c^+) \cdot \frac{1 - \tau_1}{1 - \tau} - \mathbb{P}(D = 1 | X = x) & \text{if } x < c, \end{cases}$$

$$\mathbb{P}(\tilde{D}^- = 1, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 0) = \begin{cases} \mathbb{P}(D = 1 | X = c^-) & \text{if } x \geq c, \\ \mathbb{P}(D = 1 | X = x) & \text{if } x < c, \end{cases}$$

$$\mathbb{P}(\tilde{D}^- = 0, \tilde{D}^+ = 0 | \tilde{X} = x, \tilde{M} = 0) = 1 - \mathbb{P}(\tilde{D}^- = 0, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 0)$$
$$- \mathbb{P}(\tilde{D}^- = 1, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 0),$$

$$\mathbb{P}(\tilde{D}^- = 1, \tilde{D}^+ = 0 | \tilde{X} = x, \tilde{M} = 0) = 0,$$

and

$$\mathbb{P}(\tilde{D}^- = 0, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 1) = \begin{cases} \mathbb{P}(D = 1 | X = x) \cdot \frac{\tau_1}{\tau} - h(x) & \text{if } x \geq c, \\ \mathbb{P}(D = 1 | X = c^+) \cdot \frac{\tau_1}{\tau} - h(c^+) & \text{if } x < c, \end{cases}$$

$$\mathbb{P}(\tilde{D}^- = 1, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 1) = \begin{cases} h(x) & \text{if } x \geq c, \\ h(c^+) & \text{if } x < c, \end{cases}$$

$$\mathbb{P}(\tilde{D}^- = 0, \tilde{D}^+ = 0 | \tilde{X} = x, \tilde{M} = 1) = 1 - \mathbb{P}(\tilde{D}^- = 0, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 1),$$
$$- \mathbb{P}(\tilde{D}^- = 1, \tilde{D}^+ = 1 | \tilde{X} = x, \tilde{M} = 1),$$

$$\mathbb{P}(\tilde{D}^- = 1, \tilde{D}^+ = 0 | \tilde{X} = x, \tilde{M} = 1) = 0,$$

where $h(\cdot)$ is an arbitrary continuous function satisfying that $0 \leq h(x) \leq \mathbb{P}(D = 1 | X = x) \cdot \tau_1/\tau$. With these choices, the implied distribution of $(\tilde{D}, \tilde{X}) | \tilde{X} \in (c - \epsilon, c + \epsilon)$ is the same as that of $(D, X) | X \in (c - \epsilon, c + \epsilon)$ for every $(\tau_1, \tau_0) \in \mathcal{T}$. It thus remains to be shown that one can construct

a distribution of $(\tilde{Y}(1), \tilde{Y}(0))$ given $(\tilde{D}^+, \tilde{D}^-, \tilde{X}, \tilde{M})$ that is compatible with our assumptions and such that the distribution of $\tilde{Y}$ given $(\tilde{D}, \tilde{X})$ for $\tilde{X} \in (c - \epsilon, c + \epsilon)$ is the same as the distribution of Y given (D, X) for $X \in (c - \epsilon, c + \epsilon)$ for every $(\tau_1, \tau_0) \in \mathcal{T}$. But this is always possible because our model encompasses the setting in which the label "always-assigned unit" is randomly assigned with probability τ_d to units with treatment status d and running variable to the right of the cutoff. Put differently, the conditional distribution of $(Y(1), Y(0))$ given (D^+, D^-, X, M) implies no restrictions on the feasible values of τ_1 and τ_0. $\qquad\square$

A.3 Proof of Theorem 1

By Assumption 3, we have that $\mathbb{E}(Y_i \mid X_i = c^-) = \mathbb{E}(Y_i | X_i = c^-, M_i = 0)$. Showing that $\mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q_{Y|X}(1 - \tau \mid c^+))$ and $\mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q_{Y|X}(\tau \mid c^+))$ are sharp lower and upper bounds for $\mathbb{E}(Y_i | X_i = c^+, M_i = 0)$ therefore suffices to prove the result. This follows from Corollary 4.1 in Horowitz and Manski (1995) using the following two steps. First, by the total law of probability, we can write $F_{Y|X}(y|c^+) = (1 - \tau)) F_{Y|X,M=0}(y|c^+) + \tau F_{Y|X,M=1}(y|c^+)$. Second, set $Q = F_{Y|X}(y|c^+)$, $P_{11} = F_{Y|X,M=0}(y|c^+)$, $P_{00} = F_{Y|X,M=1}(y|c^+)$, and $p = \tau$ in their notation. $\qquad\square$

A.4 Proof of Theorem 2

If τ_1 and τ_0 are known, it follows along the lines of the main text that $\Psi_0(\tau_1) = \Psi_0$. Let us now show that $\Delta_0^L(\tau_1, \tau_0)$ and $\Delta_0^L(\tau_1, \tau_0)$ are sharp lower and upper bounds for Δ_0. By Assumption 3, $\mathbb{E}(Y_i \mid X_i = c^-) = \mathbb{E}(Y_i \mid X_i = c^-, M_i = 0)$. Following the main text's argument, we thus only need to show that

$$
\begin{aligned}
&\mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_1 \mid c^+, 1), D_i = 1) \cdot \frac{\mathbb{P}(D_i = 1 | X_i = c^+) \cdot (1 - \tau_1)}{1 - \tau} \\
&\quad + \mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_0 \mid c^+, 0), D_i = 0) \cdot \frac{\mathbb{P}(D_i = 0 | X_i = c^+) \cdot (1 - \tau_0)}{1 - \tau} \quad \text{and} \\
&\mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_1 \mid c^+, 1), D_i = 1) \cdot \frac{\mathbb{P}(D_i = 1 | X_i = c^+) \cdot (1 - \tau_1)}{1 - \tau} \\
&\quad + \mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_0 \mid c^+, 0), D_i = 0) \cdot \frac{\mathbb{P}(D_i = 0 | X_i = c^+) \cdot (1 - \tau_0)}{1 - \tau}
\end{aligned}
$$

are sharp lower and upper bounds for $\mathbb{E}(Y_i|X_i = c^+, M_i = 0)$. Using the same reasoning as in the proof of Theorem 1, we find that $\mathbb{E}(Y_i|X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_1 \mid c^+, 1), D_i = 1)$ and $\mathbb{E}(Y_i|X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_0 \mid c^+, 0), D_i = 0)$ are sharp lower bounds for $\mathbb{E}(Y_i|X_i = c^+, M_i = 0, D_i = 1)$ and $\mathbb{E}(Y_i|X_i = c^+, M_i = 0, D_i = 0)$. Similarly, we find that $\mathbb{E}(Y_i|X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_1 \mid c^+, 1), D_i = 1)$ and $\mathbb{E}(Y_i|X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_0 \mid c^+, 0), D_i = 0)$ are sharp upper bounds for $\mathbb{E}(Y_i|X_i = c^+, M_i = 0, D_i = 1)$ and $\mathbb{E}(Y_i|X_i = c^+, M_i = 0, D_i = 0)$. Since $\Psi_0(\tau_1) = \Psi_0$, the result then follows by noting that the conditional bounds are simultaneously attainable since the expectations are over disjoint subsets of the sample space.

We have now shown that given knowledge of τ_1 and τ_0 the terms $\Delta_0^L(\tau_1, \tau_0)/\Psi_0(\tau_1)$ and $\Delta_0^U(\tau_1, \tau_0)/\Psi_0(\tau_1)$ are sharp lower and upper bounds for Γ_0. In general, we only know however that (τ_1, τ_0) is contained in the set $\mathcal{T}$. Due to the sharpness result in Lemma 2, any element of $\mathcal{T}$ is a feasible candidate for the value of (τ_1, τ_0). Thus, the bounds Γ_0^L and Γ_0^U are sharp. $\qquad\square$

A.5 Proof of Theorem 3 and Theorem 4

The proof of Theorem 3 follows the same steps as the proof of Lemma 2, noting that $\mathbb{P}(D_i = 1 \mid X_i = c^+, M_i = 1) \geq \mathbb{P}(D_i = 1 \mid X_i = c^+, M_i = 0)$ implies the additional restriction that $\tau_1 \geq \tau$. This follows from applying Bayes' Theorem to both sides of the inequality and rearranging terms. For Theorem 4, it follows again from Bayes' Theorem that $\mathbb{P}(D_i = 1 \mid X_i = c^+, M_i = 1) = 1$ implies $\tau_1 = \tau/\mathbb{E}(D_i \mid X_i^* = c^-)$. In addition, $\mathbb{P}(D_i = 1 \mid X_i = c^+, M_i = 1) = 1$ implies that $\tau_0 = 0$. $\qquad\square$

B Asymptotic Theory for the SRD Design

In this Appendix, we derive the asymptotic properties of the estimates of the upper and lower bound on Γ_0 in the SRD design. We impose the following regularity conditions, which are mostly standard in the context of local polynomial estimation in an regression discontinuity context.

Assumption B.1. *The data $\{(Y_i, X_i, D_i)\}_{i=1}^n$ are an independent and identically distributed sample from the distribution of some random vector (Y, X, D).*

Assumption B.2. *For some $\kappa > 0$, $\kappa_1 = [c, c + \kappa)$ and $\kappa_0 = (c - \kappa, c]$ the following holds.*

(a) $f_X(x)$ is continuous, bounded, and bounded away from zero for $x \in \kappa_1$ and $x \in \kappa_0$.

(b) $\mathbb{E}(Y_i|X_i = x, Y_i \le Q^+(\tau))$, $\mathbb{E}(Y_i|X_i = x, Y_i \ge Q^+(1-\tau))$, $Q_{Y|X}(\tau, x)$ and $Q_{Y|X}(1-\tau, x)$ are $p+1$ times continuously differentiable for $x \in \kappa_1$.

(c) $\mathbb{E}(Y_i|X_i = x)$ is $p+1$ times continuously differentiable for $x \in \kappa_0$.

(d) $\mathbb{V}(Y_i|X_i = x, Y_i \le Q^+(\tau))$, $\mathbb{V}(Y_i|X_i = x, Y_i \ge Q^+(1-\tau))$ are continuous and bounded away from zero for $x \in \kappa_1$; and $\mathbb{V}(Y_i|X_i = x)$ is continuous and bounded away from zero for $x \in \kappa_0$.

(e) $f_{Y|X}(y, x)$ is continuous, bounded and bounded away from zero in y for $x \in \kappa_1$.

Assumption B.3. *The kernel function K is a symmetric, continuous probability density function with compact support, say $[-1, 1]$.*

Assumption B.4. *The bandwidth $h = h(n)$ is such that $nh^{2p+1} \to 0$ and $nh \to \infty$ as $n \to \infty$.*

Under these conditions, we have the following result.

Theorem B.1. *Suppose that Assumptions B.1–B.4 hold, and that $\tau > 0$. Then*

$$\sqrt{nh} \begin{pmatrix} \widehat{\Gamma}^L_{0,SRD} - \Gamma^L_{0,SRD} \\ \widehat{\Gamma}^U_{0,SRD} - \Gamma^U_{0,SRD} \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

where Σ is a 2×2 matrix described more explicitly in the proof of the theorem.

Proof. Proving this result involves many arguments that are standard when analyzing the properties of kernel-based nonparametric estimators. We therefore only provide a sketch that focuses on the main issues. We begin by defining the following shorthand notation for various derivatives:

$$m_Q^{L+} = \partial_t \mathbb{E}(Y_i|X_i = c^+, Y_i \le t)|_{t=Q^+(\tau)}, \qquad m_\tau^{L+} = \partial_t \mathbb{E}(Y_i|X_i = c^+, Y_i \le Q^+(t))|_{t=\tau},$$

$$m_Q^{U+} = \partial_t \mathbb{E}(Y_i|X_i = c^+, Y_i \ge t)|_{t=Q^+(1-\tau)}, \qquad m_\tau^{U+} = \partial_t \mathbb{E}(Y_i|X_i = c^+, Y_i \ge Q^+(1-t))|_{t=\tau}.$$

We also define the following infeasible estimators of m^{L+} and m^{U+}, in which the true conditional

quantile function is used to truncate the sample instead of the polynomial approximation:

$$\widetilde{m}^{L+} = \mathbf{e}_1' \operatorname*{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} (Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}\left\{X_i > c, Y_i \leq Q_{Y|X}(\tau, X_i)\right\}$$

$$\widetilde{m}^{U+} = \mathbf{e}_1' \operatorname*{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} (Y_i - \boldsymbol{\pi}_p(X_i - c)'\beta)^2 K_h(X_i - c) \mathbb{I}\left\{X_i > c, Y_i \geq Q_{Y|X}(1 - \tau, X_i)\right\}$$

Finally, with $\mathbf{0} = (0, 0, \ldots, 0)'$, $\mathbf{1} = (1, 1, \ldots, 1)'$ and $\mathbf{h}_p = (1, h, h^2, \ldots, h^p)'$, we define the matrix

$$A \equiv \begin{pmatrix} 1 & 0 & \mathbf{1}' \cdot m_Q^{L+} & \mathbf{0}' & m_\tau^{L+} & -1 \\ 0 & 1 & \mathbf{0}' & \mathbf{1}' \cdot m_Q^{U+} & m_\tau^{U+} & -1 \end{pmatrix}$$

and the following random vector of (possibly scaled) intermediate estimators:

$$\mathbf{v} \equiv \begin{pmatrix} \widetilde{m}^{L+} - m^{L+} \\ \widetilde{m}^{U+} - m^{U+} \\ \mathbf{h}_p' \widehat{Q}_\nabla^+(\tau) - \mathbf{h}_p' Q_\nabla^+(\tau) \\ \mathbf{h}_p' \widehat{Q}_\nabla^+(1 - \tau) - \mathbf{h}_p' Q_\nabla^+(1 - \tau) \\ \widehat{\tau} - \tau \\ \widehat{m}^- - m^- \end{pmatrix}.$$

It can then be shown that under Assumptions B.1–B.4 it holds that $\sqrt{nh} \cdot \mathbf{v} \xrightarrow{d} N(0, V)$ for some positive definite matrix V as follows. Using standard arguments (e.g. Kong, Linton, and Xia, 2010), we obtain that each component of the random vector $\mathbf{v}$ can be written as the sum of three terms: (i) a kernel-weighted sum of independent and identically distributed random variables that has mean zero and variance of order $O((nh)^{-1})$; (ii) a deterministic bias term that is of order $O(h^{p+1})$; and (iii) a remainder term of the order $o_P((nh)^{-1/2}) + h^{p+1})$. The result is then implied by the restrictions on h and p; and the multivariate version of Ljapunov's CLT. We do not give an explicit formula for every element of V as the related calculation are tedious but standard, and thus not very insightful.

Next, by using stochastic equicontinuity arguments one can show that the estimators $\widehat{m}^{L+}$ and

$\widehat{m}^{U+}$ satisfy the following stochastic decomposition:

$$\widehat{m}^{L+} = \widetilde{m}^{L+} + m_{Q(\tau)}^{L+}\mathbf{h}'_p(\widehat{Q}^+_\nabla(\tau) - Q^+_\nabla(\tau)) + m_\tau^{L+}(\widehat{\tau} - \tau) + o_P((nh)^{-1/2}) + O(h^{p+1})$$

$$\widehat{m}^{U+} = \widetilde{m}^{U+} + m_{Q(1-\tau)}^{U+}\mathbf{h}'_p(\widehat{Q}^+_\nabla(1-\tau) - Q^+_\nabla(1-\tau)) + m_\tau^{U+}(\widehat{\tau} - \tau) + o_P((nh)^{-1/2}) + O(h^{p+1}).$$

The statement of the theorem then follows from the above arguments with $\Sigma = AVA'$, where A is as defined above. □

C Additional Tables and Graphs

We present here some supporting graphs. Figure C.4 displays the distribution of our outcome variable (duration without a formal job, censored at one year after layoff) on the left and on the right of the cutoff (30-day window around the cutoff). Figure C.5 displays the distribution of our outcome variable on the right of the cutoff for female and male, separetely. Figure C.6 displays the full schedule of the UI benefit level, which is a function of a beneficiary's average monthly wage in the three years prior to her layoff. Figure C.7 displays the mean of different covariates on each side of the cutoff by day between the layoff and eligibility dates.
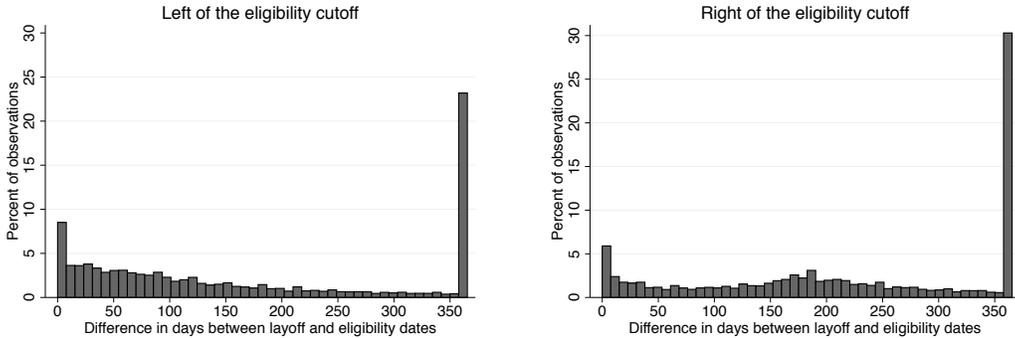
Figure C.4: Distribution of our outcome variable on each side of the cutoff

The figure displays the distribution of our outcome variable (duration without a formal job, censored at one year after layoff) on the left and on the right of the cutoff (30-day window on each side of the cutoff).
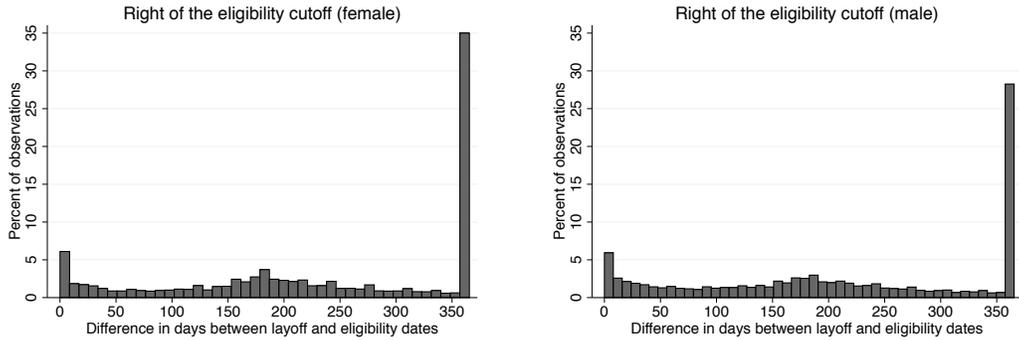
Figure C.5: Distribution of our outcome variable on the right side of the cutoff by gender

The figure displays the distribution of our outcome variable (duration without a formal job, censored at one year after layoff) on the right of the cutoff (30-day window on each side of the cutoff) for female and male, separately.
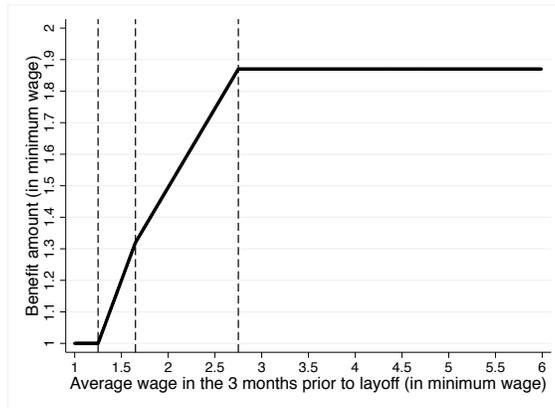
Figure C.6: Monthly UI benefit amount

The figure displays the relationship between a UI beneficiary's average monthly wage in the three months prior to her layoff and her monthly UI benefit level. All monetary values are indexed to the federal minimum wage, which changes every year. The replacement rate is 100% at the bottom of the wage distribution as the minimum benefit level is one minimum wage. The graph displays a slope of 0% until 1.25 minimum wages, then of 80% until 1.65 minimum wages, and finally of 50% until 2.75 minimum wages. The maximum benefit level is 1.87 minimum wages.
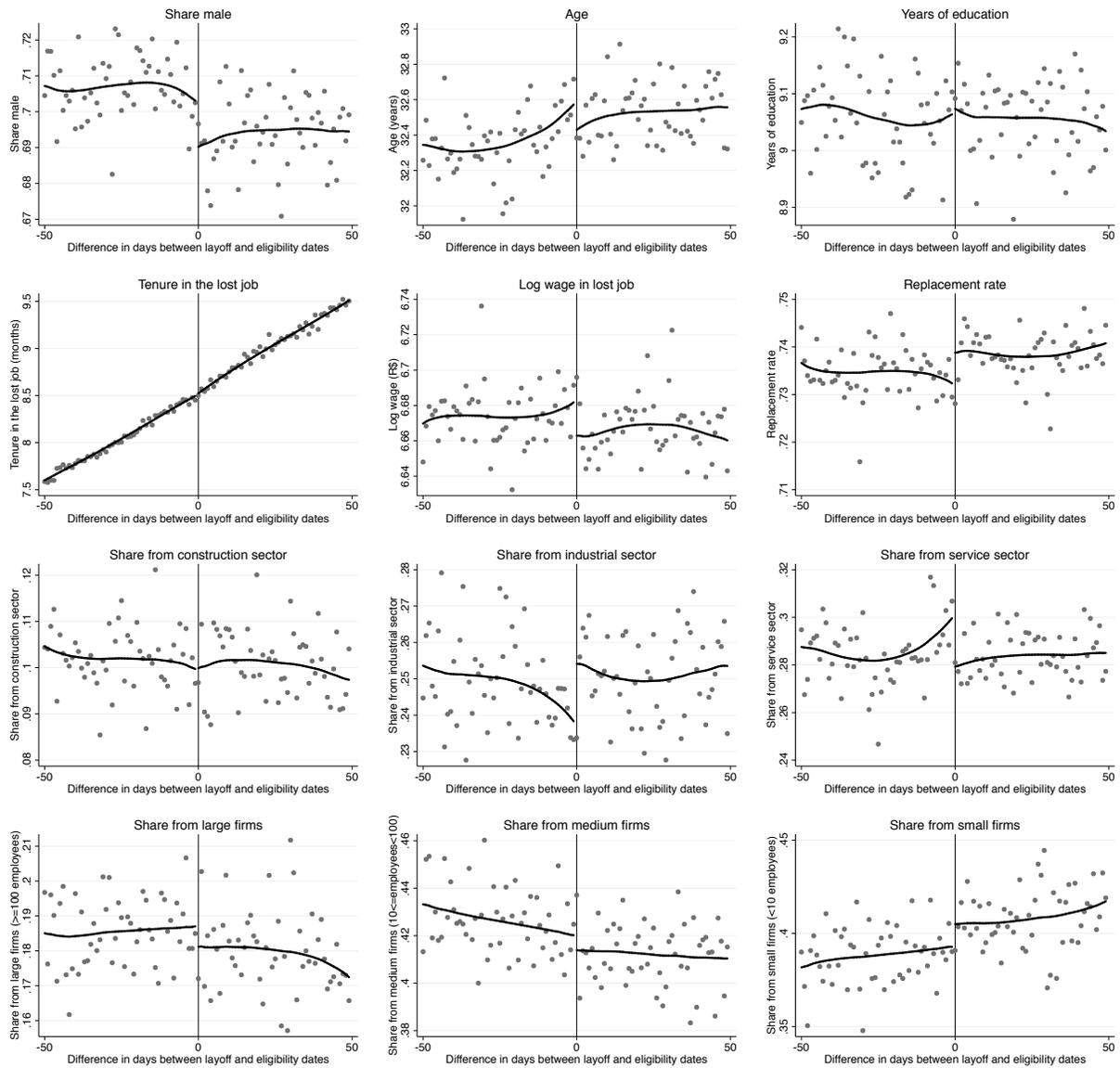
Figure C.7: Graphical evidence for the characteristics of always-assigned units in our empirical application

The figure displays the mean of different covariates on each side of the cutoff by day between the layoff and eligibility dates.

References

Anderson, M., and J. Magruder (2012): "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Economic Journal*, 122(563), 957–989.

Anderson, P., and B. Meyer (2000): "The effects of the unemployment insurance payroll tax on wages, employment, claims and denials," *Journal of Public Economics*, 78, 81–106.

Andrews, D. (2000): "Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space," *Econometrica*, 68(2), 399–405.

Andrews, D., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Angrist, J., and M. Rokkanen (forthcoming): "Wanna Get Away: Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff," *Journal of the American Statistical Association*.

Baily, M. (1978): "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 10, 379–402.

Bajari, P., H. Hong, M. Park, and R. Town (2011): "Regression Discontinuity Designs with an Endogenous Forcing Variable and an Application to Contracting in Health Care," NBER Working Paper 17643.

Berger, R. (1982): "Multiparameter hypothesis testing and acceptance sampling," *Technometrics*, 24(4), 295–300.

Calonico, S., M. Cattaneo, and R. Titiunik (2015): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

Camacho, A., and E. Conover (2011): "Manipulation of Social Program Eligibility," *American Economic Journal: Economic Policy*, 3(2), 41–65.

CARD, D., AND L. GIULIANO (2014): "Does Gifted Education Work? For Which Students?," NBER Working Paper 20453.

CATTANEO, M. D., M. JANSSON, AND X. MA (2015): "A Simple Local Polynomial Density Estimator with an Application to Manipulation Testing Supplemental Appendix," *Working Paper.*

CHEN, X., AND C. FLORES (2012): "Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps," Unpublished manuscript.

CHENG, M.-Y. (1997): "A bandwidth selector for local linear density estimators," *Annals of Statistics*, 25(3), 1001–1013.

CHENG, M.-Y., J. FAN, AND J. MARRON (1997): "On Automatic Boundary Corrections," *Annals of Statistics*, 25, 1691–1708.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2013): "Intersection Bounds: estimation and inference," *Econometrica*, 81(2), 667–737.

CHETTY, R. (2008): "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 116(2), 173–234.

DAVIS, B., J. ENGBERG, D. EPPLE, H. SIEG, AND R. ZIMMER (2013): "Bounding the Impact of a Gifted Program On Student Retention using a Modified Regression Discontinuity Design," Unpublished manuscript.

DEE, T., W. DOBBIE, B. JACOB, AND J. ROCKOFF (2014): "Rules and Discretion in the Evaluation of Students and Schools: The Case of the New York Regents Examinations," Unpublished manuscript.

DONG, Y. (2015): "Regression Discontinuity Designs with Sample Selection," Unpublished manuscript.

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

FELDSTEIN, M. (1976): "Temporary Layoffs in the Theory of Unemployment," *Journal of Political Economy*, 84, 937–958.

FRANDSEN, B. R., M. FRÖLICH, AND B. MELLY (2012): "Quantile treatment effects in the regression discontinuity design," *Journal of Econometrics*, 168(2), 382–395.

GERARD, F., AND G. GONZAGA (2014): "Informal Labor and the Efficiency Cost of Social Programs: Evidence from the Brazilian Unemployment Insurance Program," *Mimeo, Columbia University*.

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.

HIRANO, K., AND J. R. PORTER (2012): "Impossibility results for nondifferentiable functionals," *Econometrica*, pp. 1769–1790.

HOROWITZ, J., AND C. MANSKI (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63(2), 281–302.

——— (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95(449), 77–84.

IMBENS, G., AND J. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

IMBENS, G., AND K. KALYANARAMAN (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79(3), 933–959.

IMBENS, G., AND C. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Jones, M. C. (1993): "Simple boundary correction for kernel density estimation," *Statistics and Computing*, 3(3), 135–146.

Kim, B. M. (2012): "Do Developmental Mathematics Courses Develop the Mathematics?," Unpublished manuscript.

Kleven, H., and M. Waseem (2013): "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics*, 128(2), 669–723.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, pp. 33–50.

Kong, E., O. Linton, and Y. Xia (2010): "Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model," *Econometric Theory*, 26(05), 1529–1564.

Lee, D. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76(3), 1071–1102.

Lee, D., and D. Card (2008): "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 142(2), 655–674.

Lee, D., and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355.

Lejeune, M., and P. Sarda (1992): "Smooth estimators of distribution and density functions," *Computational Statistics & Data Analysis*, 14(4), 457–471.

Manski, C. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review: Papers and Proceedings*, 80(2), 319–323.

——— (1997): "Monotone Treatment response," *Econometrica*, 65(6), 1311–1334.

MANSKI, C., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

MCCRARY, J. (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142(2), 698–714.

PORTER, J. (2003): "Estimation in the Regression Discontinuity Model," Unpublished manuscript.

RUBIN, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66(5), 688–701.

SCOTT-CLAYTON, J. (2011): "On Money and Motivation: A Quasi-Experimental Analysis of Financial Incentives for College Achievement," *Journal of Human Resources*, 46(3), 614–646.

STOYE, J. (2009): "More on confidence intervals for partially identified parameters," *Econometrica*, pp. 1299–1315.

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.

URQUIOLA, M., AND E. VERHOOGEN (2009): "Class-Size Caps, Sorting, and the Regression Discontinuity Design," *American Economic Review*, 99(1), 179–215.

YU, K., AND M. JONES (1998): "Local Linear Quantile Regression," *Journal of the American statistical Association*, 93(441), 228–237.