

Partial Identification in Regression Discontinuity Designs with Manipulated Running Variables

Francois Gerard, Miikka Rokkanen, and Christoph Rothe*

PRELIMINARY AND INCOMPLETE

Abstract

In a regression discontinuity design, units are assigned to receive a treatment if their value of a running variable lies above a known cutoff. A key assumption to ensure point identification of treatment effects in this context is that units cannot manipulate the value of their running variable to guarantee or avoid assignment to the treatment. Standard identification arguments break down if this condition is violated. This paper shows that treatment effects remain partially identified even if some units are able to manipulate their running variable. We derive sharp bounds on the treatment effects for the subpopulation of units that do not engage in manipulation, and show how to estimate them in practice. Our results apply to both sharp and fuzzy regression discontinuity designs. We use our method to study the effect of unemployment insurance on unemployment duration in Brazil, where we find strong evidence of manipulation at eligibility cutoffs.

*First Version: August 15, 2015. This Version: September 23, 2015. Contact information: Columbia University, Department of Economics, 420 West 118th Street, New York, NY 10027.

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 lies above a known 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: it is essentially random whether a unit falls just below or just above the cutoff (Lee and Lemieux, 2010). For this argument to be valid, it is necessary that units cannot manipulate the value of their running variable in order to guarantee or avoid assignment to treatment. If such manipulation was possible, units just below and just above the cutoff would be unlikely to be comparable because of self-selection. Thus manipulation is likely to break down local randomization and consequently the identification of treatment effects.

Manipulation is an issue that commonly arises in empirical applications. 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. Scott-Clayton (2011) finds that students manipulate high school credits in order to qualify for the PROMISE scholarship in West Virginia. Card and Giuliano (2014) deal with manipulation of IQ scores used to assign students to a gifted classroom in an unnamed large urban school district in the US. Dee, Dobbie, Jacob, and Rockoff (2014) provide evidence that teachers manipulate test scores so that students meet performance standards in New York City. Manipulation of running variables around kinks and discontinuities in tax and transfer systems has also generated its own literature in public economics (e.g., Saez, 2010; Kleven and Waseem, 2013).

A jump in the density of the running variable at the cutoff is a strong indication that some units have perfect control over the value of their running variable (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 simply give up on using the cutoff for inference on treatment effects.¹ This practice is problematic for at least two reasons. First, a non-rejection might not be due to a total lack of manipulation, but to a lack of statistical power. The local randomization assumption could thus still be violated, and estimates based on ignoring this possibility potentially be severely biased in this case. Second, even if a test rejects the null hypothesis of no manipulation, the number of units engaging in manipulation could still be relatively modest, and the data thus still be informative to some extent.

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 if manipulation occurs, they are generally 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 to estimate our bounds in practice, 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.

We model manipulation by assuming that each unit is of one of two unobserved types, which we call manipulators and non-manipulators, respectively. Here manipulators are units that can perfectly control whether their value of the running variable falls above or below the cutoff, whereas non-manipulators behave as postulated by the standard assumptions of an RD design. Since the lack of randomization makes it difficult to derive meaningful

¹There is a small number of papers that develops tailored solutions that are only valid under strong (often unspecified) assumptions in this case. For examples, see Bajari, Hong, Park, and Town (2011), Davis, Engberg, Epple, Sieg, and Zimmer (2013), and Anderson and Magruder (2012).

conclusions about the effect of the treatment on manipulators, we focus on causal effects on non-manipulators as our parameter of interest. Our identification strategy consists of two steps. First, building on the arguments of McCrary (2008), we use the height of the jump of the running variable’s density at the cutoff to identify the proportion of manipulators among units close to the cutoff. Second, we use this information to bound treatment effects by considering the “worst case” scenarios in which manipulators are the units with either the highest or the lowest value of the outcome variable, and then trimming the respective observations. This approach is similar to that of Lee (2009) for bounding treatment effects in randomized experiments under sample selection,² but is more involved especially in the case of a fuzzy RD design due to the more complicated structure of our model.

Our partial identification results are constructive, in the sense that they deliver explicit expressions for the bounds. Estimates of our bounds are relatively straightforward to compute in an empirical application. In this paper, we propose computationally simple procedures 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 bounds. We also discuss how to construct valid confidence intervals using techniques proposed by Imbens and Manski (2004) and Chernozhukov, Lee, and Rosen (2013). 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.

In this paper, we also provide an empirical study the effect of UI eligibility in Brazil that makes use of our approach. UI is a relevant program to consider. It typically varies benefits based on some discontinuous rules. For instance, eligibility often requires a minimum number

²Chen 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 RD.

of months of employment prior to layoff. Such a rule may lead to changes in the number and timing of layoffs, thus potentially breaking down the point identification of the effects of UI eligibility at the cutoff. Brazil is also a relevant country to study. 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 one such cutoff. Yet, we are able to infer that UI eligibility increases the time it takes to return to a formal job by at least 15.5% for non-manipulators at the cutoff.

Our paper contributes to the methodological literature on RD designs. Earlier papers in this literature have focused on nonparametric identification and estimation of average treatment effects (Hahn, Todd, and Van der Klaauw, 2001; Porter, 2003) and quantile treatment effects (Frandsen, Frölich, and Melly, 2012) as well as optimal bandwidth choice (Imbens and Kalyanaraman, 2012) and bias-correction (Calonico, Cattaneo, and Titiunik, 2015) for local polynomial regression-based estimators, inference with discrete running variables (Lee and Card, 2008), testing for manipulation of the running variable (McCrary, 2008; Cattaneo, Jansson, and Ma, 2015), and extrapolation of treatment effects away from the cutoff (Angrist and Rokkanen, forthcoming; Rokkanen, 2015).

Our paper is also related to the extensive literature on partial identification. Much of this literature studies the impact of various assumptions on the shape of the identified set in some class of models. For references, see Manski (1990), Manski (1997), Horowitz and Manski (1995), Horowitz and Manski (2000), Manski and Tamer (2002), and Blundell, Gosling, Ichimura, and Meghir (2007). Another strand of this literature focuses on inference on either the identified set or the unidentified parameters. For references, see Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), Andrews and Soares (2010), Romano and Shaikh (2010), Andrews and Shi (2013), and Chernozhukov, Lee, and Rosen (2013)

The remainder of the paper is organized as follows. In Section 2, we introduce a framework for RD designs with manipulation. In Section 3, we study partial identification of treatment effects in both Sharp and Fuzzy RD designs. Section 4 discusses estimation and inference. Section 5 contains the empirical application. Finally, Section 6 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 $i = 1, \dots, n$, 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. 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, each unit i is assigned a treatment status according to a deterministic function of a continuous 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 assigned to 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 known 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 $D_i^- = D_i(c^-) \equiv$

$\lim_{x \uparrow c} D_i(x)$.³ The extent to which units comply with their assignment distinguishes the two main types of RD designs that are commonly considered 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 of local randomization; namely that Z_i is as good as randomly assigned around the cutoff (Lee and Lemieux, 2010). This feature allows one to identify treatment effects by comparing outcomes (and treatment probabilities) of units just to the left and right of the cutoff. This source of identification may break down if at least some of the units have perfect control over the realization of the running variable, and are thus able to affect their treatment assignment through this channel. We will refer to this behavior as manipulation of the running variable.⁴ Manipulation is modeled by assuming that each unit is of one of two unobservable types, which we call manipulators and non-manipulators, respectively. Here we think of manipulators as units that exert perfect control over their value of the running variable, whereas among non-manipulators the usual RD design is valid.⁵

³Throughout 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 will 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.

⁴McCrary (2008) refers to this as complete manipulation. He also discusses the concept of partial manipulation in which units have only imperfect control over the running variable. For example, students taking a test might exert more effort if they anticipate that their score will be close to, and possibly below, the cutoff. We ignore partial manipulation because it is unlikely to break down local randomization.

⁵We use the manipulation terminology since in many applications those units with perfect control over their value of the running variable achieve this by violating certain institutional rules, or at least by acting in bad faith. However, this is not necessary for our setup. For example, if a teacher would manipulate the grades of certain students, we would still refer to these students as manipulators, even if they are unaware of the teacher's actions.

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 a manipulator and $M_i = 0$ otherwise. We then impose the following standard assumption from the RD literature (e.g. Hahn, Todd, and Van der Klaauw, 2001; Frandsen, Frölich, and Melly, 2012) regarding the behavior of non-manipulators.

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.

Assumption 1(i) states 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 that states that the response of treatment selection to crossing the threshold is monotone for every unit. Assumption 1(iii) is a continuity condition which roughly speaking requires the mean of potential outcomes and treatment states 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 non-manipulating units close to cutoff on either side.

In addition to Assumption 1, which is standard in the RD literature, we also impose the following weak regularity condition on the distribution of the running variable among non-manipulators.

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

Together with Assumption 1(iv), this assumption implies that the density of X_i among non-manipulators is smooth and strictly positive over some open neighborhood of c . Continuity of the running variable's density around the cutoff would seem like a reasonable

condition in applications, and is generally considered to be an indication for the absence of manipulation in the literature (McCrary, 2008).

The only restriction concerning the behavior of manipulators that we impose in our setup is the following.

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 .

The first part of Assumption 3, which will be key for our following analysis, implies that no unit will manipulate the value of its running variables such that it falls below the cutoff. Such “one-sided” manipulation would seem empirically plausible if eligibility for the treatment is unambiguously desirable from any unit’s point of view; as in the case of a cash transfer or unemployment benefits, for example. The assumption seems also generally plausible in FRD designs, where being above the threshold does not imply mandatory treatment uptake. Note that the condition has no implications for how the manipulators’ values of the running variable would change in a counterfactual scenario in which manipulation suddenly became impossible. For example, we do not require that in the absence of manipulation the running variable of manipulators would fall below the cutoff (such additional structure would also not affect our analysis below).

Assumption 3(ii) rules out mass points in the distribution of X_i for manipulators around the cutoff. In particular, it postulates that manipulators do not choose their running variable to be exactly equal to c with positive probability. However, the distribution of X_i is allowed to be arbitrarily highly concentrated close to the cutoff. In view of Assumption 1(iv), the condition implies that manipulators cannot simply be distinguished from non-manipulators through their values of the observed running variable (without such a condition the analysis would be trivial, as manipulation would effectively be observable). It also implies that in the full population that contains both manipulators and non-manipulators the observed running variable X_i will be continuously distributed, although its density will generally be discontinuous at c . Moreover, taken together Assumptions 2 and 3 imply that all units

observed to the left of the cutoff are non-manipulators, i.e. $\mathbb{P}(M_i = 1|X_i = c^-) = 0$, whereas to the right of the cutoff we observe a mixture of manipulators and non-manipulators.

3. IDENTIFICATION UNDER MANIPULATION

3.1. *Parameters of Interest*

Since the lack of local randomization makes it difficult to derive meaningful conclusions about the causal effect of the treatment among manipulators from observable quantities, we focus on causal effects among non-manipulators as our parameter of interest in this paper. Specifically, we study identification of

$$\Gamma_0 \equiv \mathbb{E}(Y_i(1) - Y_i(0)|X_i = c, D_i^+ > D_i^-, M_i = 0), \quad (3.1)$$

which can be understood as the local average treatment effect for the subgroup of non-manipulating “compliers” who receive the treatment if and only if their value of the running variable X_i is above the cutoff (Imbens and Angrist, 1994). By Assumption 1 and standard arguments from the RD literature (e.g. Hahn, Todd, and Van der Klaauw, 2001), it follows that our parameter of interest can be written as

$$\Gamma_0 = \frac{\Delta_0}{\Psi_0}, \quad \text{where} \quad \Delta_0 \equiv \mathbb{E}(Y_i|X_i = c^+, M_i = 0) - \mathbb{E}(Y_i|X_i = c^-, M_i = 0) \quad \text{and}$$

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

We mainly work with this representation in our identification analysis.

3.2. *Proportion of Manipulators*

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 manipulators 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 manipulators just to the right of the cutoff among units with treatment

status $d \in \{0, 1\}$. While we cannot observe or infer the type of any given unit, under our assumptions we can point identify τ from the height of the jump of 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, which imposes two logical restrictions 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

$$\tau = \tau_1 \cdot \mathbb{E}(D_i|X_i = c^+) + \tau_0 \cdot (1 - \mathbb{E}(D_i|X_i = c^+)). \quad (3.2)$$

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

$$\mathbb{E}(D_i|X_i = c^+) \cdot \frac{1 - \tau_1}{1 - \tau} > \mathbb{E}(D_i|X_i = c^-). \quad (3.3)$$

With \mathcal{T} denoting the set containing those $(\tau_1, \tau_0) \in [0, 1]^2$ that satisfy the two restrictions (3.2)–(3.3), 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

$$\begin{aligned}\tau_1^L &= \max \left\{ 0, 1 - \frac{1 - \tau}{g^+} \right\}, & \tau_1^U &= \min \left\{ 1 - \frac{(1 - \tau) \cdot g^-}{g^+}, \frac{\tau}{g^+} \right\}, \\ \tau_0^L &= \min \left\{ 1, \frac{\tau}{1 - g^+} \right\}, & \tau_0^U &= \max \left\{ 0, \tau - \frac{(1 - \tau) \cdot (g^+ - g^-)}{1 - g^+} \right\},\end{aligned}$$

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

3.3. Bounds on Treatment Effects for Non-Manipulators

3.3.1. Sharp RD Designs

It is instructive to first consider deriving bounds on Γ_0 in a sharp RD design before studying the more general case of a fuzzy design. Since $D_i^+ > D_i^-$ for every unit i in a sharp design, the causal parameter Γ_0 simplifies to an average treatment effect in this case; and since also $\Psi_0 = 1$ it can be written as

$$\Gamma_0 = \mathbb{E}(Y_i | X_i = c^+, M_i = 0) - \mathbb{E}(Y_i | X_i = c^-, M_i = 0). \quad (3.4)$$

Since we only observe non-manipulating units below 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 (3.4). 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 non-manipulators 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; and a lower bound by the expected outcome of those with outcomes smaller than the corresponding $1 - \tau$ quantile. These bounds correspond to two “worst case” scenarios in which the proportion τ of units with either the highest or the lowest outcomes are all manipulators. 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 sharp bounds on Γ_0 .

Theorem 1. *Suppose Assumptions 1–3 hold, that $D_i^+ > D_i^-$ for all $i = 1, \dots, n$, and that $F_{Y|X}(y|c^+)$ is continuous in y . Then sharp lower and upper bounds on Γ_0 are given by*

$$\begin{aligned}\Gamma_{0,SRD}^L &= \mathbb{E}(Y_i|X_i = c^+, Y_i \leq Q_{Y|X}(1 - \tau|c^+)) - \mathbb{E}(Y_i|X_i = c^-) \quad \text{and} \\ \Gamma_{0,SRD}^U &= \mathbb{E}(Y_i|X_i = c^+, Y_i \geq Q_{Y|X}(\tau|c^+)) - \mathbb{E}(Y_i|X_i = c^-),\end{aligned}$$

respectively, where $Q_{Y|X}$ denotes the conditional quantile function of Y_i given X_i .

3.3.2. Fuzzy RD Designs

We now extend the partial identification result for Γ_0 in the SRD design to the case of a fuzzy design, which requires a more involved argument. Recall that we can write the parameter of interest as

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

in the FRD design. The two expectations in this expression that condition on $X_i = c^-$ are again point identified: as only non-manipulators are observed below 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. To derive bounds, write the unknown term in the the definition of Δ_0 as

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

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 could be bounded sharply by considering the “worst case” scenarios in which manipulators of either treatment status

are the units with the highest or the lowest outcomes. That is, an upper bound on the expectation that conditions on $D_i = 1$ could 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 could be obtained analogously. Moreover, by Bayes' Theorem, the two probabilities on the right-hand side of the last equation would be point identified since

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

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

$$\begin{aligned} \Delta_0^L(\tau_1, \tau_0) &= \sum_{d=0,1} \mathbb{E}(Y_i|X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_d|c^+, d), D_i = d) \\ &\quad \times \mathbb{P}(D_i = d|X_i = c^+) \cdot \frac{1 - \tau_d}{1 - \tau} - \mathbb{E}(Y_i|X_i = c^-) \quad \text{and} \\ \Delta_0^U(\tau_1, \tau_0) &= \sum_{d=0,1} \mathbb{E}(Y_i|X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_d|c^+, d), D_i = d) \\ &\quad \times \mathbb{P}(D_i = d|X_i = c^+) \cdot \frac{1 - \tau_d}{1 - \tau} - \mathbb{E}(Y_i|X_i = c^-), \end{aligned}$$

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

$$\Psi_0 = \mathbb{E}(D_i|X_i = c^+) \cdot \frac{1 - \tau_1}{1 - \tau} - \mathbb{E}(D_i|X_i = c^-) \equiv \Psi_0(\tau_1). \quad (3.5)$$

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

$$\Gamma_0^L(\tau_1, \tau_0) \equiv \frac{\Delta_0^L(\tau_1, \tau_0)}{\Psi_0(\tau_1)} \quad \text{and} \quad \Gamma_0^U(\tau_1, \tau_0) \equiv \frac{\Delta_0^U(\tau_1, \tau_0)}{\Psi_0(\tau_1)}, \quad (3.6)$$

respectively. In view of Lemma 2, we then 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 (3.6). 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 y for $d \in \{0, 1\}$. Then sharp lower and upper bounds on Γ_0 are given by*

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

respectively.

Remark 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 manipulators just to the right of the cutoff have very low treatment probabilities, and a setting where the treatment probability of non-manipulators 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.

Remark 2. Due to the presence of optimization operators, the bounds in Theorem 2 are a non-smooth functional of the population distribution of observables. Hirano and Porter (2012) show that for such parameters no locally asymptotically unbiased estimators exist, and that for any estimator a bias-correction procedure that reduces bias too much will eventually cause the variance of the estimator to explode. These impossibility results impose restrictions on the potential accuracy of estimation and inference in such settings. We employ techniques developed by Chernozhukov, Lee, and Rosen (2013) below in order to achieve a reasonable balance between bias reduction and accuracy when taking our results to actual data.

3.3.3. Fuzzy RD Designs with Additional Restrictions on Manipulators' Behavior

The bounds in Theorem 2 can be narrowed if one is willing to impose stronger assumptions on the behavior of manipulators than the (rather weak) ones that we have imposed so far. In settings where the treatment is generally perceived as being attractive, and manipulators exert some effort in order to become eligible, it might seem plausible to assume that their probability of receiving the treatment given eligibility 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 manipulators are at least as likely to get treated as eligible non-manipulators.

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 by

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

respectively, in this case.

Remark 3. Comparing the first part of the theorem to Lemma 2 and the following discussion, we see the conditions of Theorem 3 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 more narrow bounds because optimization is carried out over a smaller area.

One can drive the line of reasoning that motivated the previous result even further, and consider the identifying power of the assumption that manipulators *always* receive the

treatment.

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 by

$$\Gamma_{0,FRD(b)}^L = \frac{\Delta_0^L(\tau/\mathbb{E}(D_i|X_i = c^+), 0)}{\Psi_0(\tau/\mathbb{E}(D_i|X_i = c^+))} \quad \text{and} \quad \Gamma_{0,FRD(b)}^U = \frac{\Delta_0^U(\tau/\mathbb{E}(D_i|X_i = c^+), 0)}{\Psi_0(\tau/\mathbb{E}(D_i|X_i = c^+))},$$

respectively, in this case.

Remark 4. 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. This has the additional advantage that estimation and inference are not subject to the problems pointed out by Hirano and Porter (2012) in this case.

3.4. Further Extensions and Remarks

In this subsection, we discuss a number of extensions and remarks related to our main results that we derived above.

3.4.1. 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 in the previous subsection, and thus potentially more informative in applications. Consider for simplicity the case of a SRD design, and let W_i be a vector of such covariates. Then by using arguments similar to those in Lee (2009), one can narrow the lower and upper bounds on Γ_0 as described in the following corollary (more narrow bounds for the FRD Design than those given in Theorem 2 can be derived in a similar fashion).

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*

$$\begin{aligned}\Gamma_{0,SRD-W}^L &= \int \mathbb{E}(Y_i|X_i = c^+, Y_i \leq Q_{Y|XW}(1 - \tau(w)|c^+, w), W_i = w) dF_{W|X}(x, c^+) \\ &\quad - \mathbb{E}(Y_i|X_i = c^-) \quad \text{and} \\ \Gamma_{0,SRD-W}^U &= \int \mathbb{E}(Y_i|X_i = c^+, Y_i \geq Q_{Y|XW}(\tau(w)|c^+, W_i = w)) dF_{W|X}(x, c^+) \\ &\quad - \mathbb{E}(Y_i|X_i = c^-),\end{aligned}$$

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

3.4.2. Bounds for Discrete Outcomes.

The results in Theorem 1 and 2 are stated for the case that 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 can be important in many empirical settings. For example, if the outcome variable is the duration in months until a worker finds a new job after being laid off (and the treatment is being eligible for unemployment benefits, for instance), then in practice the outcome will only take a relatively small number of distinct values. The following corollary generalizes the statement of Theorem 1 to the case of a discrete outcome (an extension of Theorem 2 could be obtained analogously).

Corollary 2. *Suppose that Assumptions 1–3 hold, that $D_i^+ > D_i^-$ for all $i = 1, \dots, n$, 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*

$$\begin{aligned}\Gamma_{0,SRD}^L &= (1 - \theta^L) \mathbb{E}(Y_i|X_i = c^+, Y_i < Q_{Y|X}(1 - \tau|c^+)) \\ &\quad + \theta^L Q_{Y|X}(1 - \tau|c^+) - \mathbb{E}(Y_i|X_i = c^-) \quad \text{and} \\ \Gamma_{0,SRD}^U &= (1 - \theta^U) \mathbb{E}(Y_i|X_i = c^+, Y_i > Q_{Y|X}(\tau|c^+)) \\ &\quad + \theta^U Q_{Y|X}(\tau|c^+) - \mathbb{E}(Y_i|X_i = c^-),\end{aligned}$$

respectively, with

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

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

and 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.

3.4.3. Identifying the Characteristics of Manipulators and Non-Manipulators.

It is not possible in our setup to determine whether any given unit is a manipulator or not. 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 pretreatment covariates, it is possible to identify the distribution of these covariates among manipulators and non-manipulators just to the right of the cutoff if the distribution of W_i does not change discontinuously at c . The following corollary formally states the result.

Corollary 3. *Suppose that Assumptions 1–2 hold, and that $\mathbb{E}(g(W_i) | X_i = x, M_i = 0)$ is continuous in x at c for some known function $g(\cdot)$. Then*

$$\begin{aligned} \mathbb{E}(g(W_i) | X_i = c^+, M_i = 1) &= \frac{1}{\tau} (\mathbb{E}(g(W_i) | X_i = c^+) - \mathbb{E}(g(W_i) | X_i = c^-)) \\ &\quad + \mathbb{E}(g(W_i) | X_i = c^-) \quad \text{and} \\ \mathbb{E}(g(W_i) | X_i = c^+, M_i = 0) &= \mathbb{E}(g(W_i) | X_i = c^-). \end{aligned}$$

By putting $g(w) = w^k$, for example, the corollary shows that the data identify the k th moment of the distribution of W_i among manipulators and non-manipulators just to the right of the cutoff, where $k \in \mathbb{N}$. If W_i was the age of an individual, for instance, this approach would make it possible to determine whether manipulators are older on average than non-manipulators. By putting $g(w) = \mathbb{I}\{w \leq r\}$, we obtain an identification result for the (conditional) CDF of the covariates at any level $r \in \mathbb{R}$, which means that we can identify

any feature of the (conditional) distribution of W_i .

4. ESTIMATION AND INFERENCE

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, and how one can construct valid confidence intervals. Our methods make use of many existing results from the literature on estimation and inference under partial identification, but also involve steps that are specific to our setting. The methods also differ somewhat conceptually depending on whether the definition of the bounds involves an optimization operator (as in the case of Theorems 2 and 3) or not (as in the case of Theorems 1 and 4).

In both cases, estimating the bounds and conducting inference 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 weights are decreasing to zero with the distance of an observation to the cutoff. The exact value of this weights is 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, that will later be combined to obtain estimates of the bounds. 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 $\boldsymbol{\pi}_p(x) = (1/0!, x/1!, x^2/2!, \dots, x^p/p!)'$ and $K_h(x) = K(x/h)/h$ for any $x \in \mathbb{R}$, and define the $(p+1)$ -vector $\mathbf{e}_1 = (1, 0, \dots, 0)'$. The data available to the econometrician is an independent and identically distributed sample $\{(Y_i, D_i, X_i), i = 1, \dots, n\}$ of size n .

4.1.1. Proportion of Manipulators.

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 $K^{p+}(x) = \mathbf{e}_1' S^{-1} (1, x, \dots, x^p)' K(x)$ 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 $j = 1, \dots, 2p$ being constants that depend 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, our estimate of the proportion of manipulators τ is then given by

$$\widehat{\tau} = 1 - \widehat{f}^- / \widehat{f}^+.$$

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

$$\begin{aligned}\hat{f}^+ &= \mathbf{e}'_1 \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \int_c^\infty (\tilde{f}_n(x) - \boldsymbol{\pi}_p(x-c)' \beta)^2 K_h(x-c) dx \quad \text{and} \\ \hat{f}^- &= \mathbf{e}'_1 \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \int_{-\infty}^c (\tilde{f}_n(x) - \boldsymbol{\pi}_p(x-c)' \beta)^2 K_h(x-c) dx.\end{aligned}$$

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 that will become apparent below, it will be useful to 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 vector that contains the right limit of the respective conditional quantile functions and their first p derivatives at the cutoff by

$$\begin{aligned}Q_{\nabla}^+(t) &= (Q_{Y|X}(t, c^+), \partial_x Q_{Y|X}(t, x)|_{x=c^+}, \dots, \partial_x^p Q_{Y|X}(t, x)|_{x=c^+})' \quad \text{and} \\ Q_{\nabla}^+(t, d) &= (Q_{Y|XD}(t, c^+, d), \partial_x Q_{Y|XD}(t, x, d)|_{x=c^+}, \dots, \partial_x^p Q_{Y|XD}(t, x, d)|_{x=c^+})',\end{aligned}$$

respectively. 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

$$\begin{aligned}\hat{Q}_{\nabla}^+(t) &= \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n \rho_t(Y_i - \boldsymbol{\pi}_p(X_i - c)' \beta) K_h(X_i - c) \mathbb{I}\{X_i \geq c\}, \quad \text{and} \\ \hat{Q}_{\nabla}^+(t, d) &= \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n \rho_t(Y_i - \boldsymbol{\pi}_p(X_i - c)' \beta) K_h(X_i - c) \mathbb{I}\{X_i \geq c, D_i = d\},\end{aligned}$$

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

$$\begin{aligned} m^{L+} &= \mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q^+(\tau)), \\ m^{U+} &= \mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q^+(1 - \tau)), \\ m^{L+}(t_1, t_0, d) &= \mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q^+(t_d, d), D_i = d), \\ m^{U+}(t_1, t_0, d) &= \mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q^+(1 - t_d, d), D_i = d), \end{aligned}$$

for the various truncated conditional expectation functions, and

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

for the “non-truncated” conditional expectations that we have to estimate.

We begin by discussing estimation of the truncated conditional expectations m^{L+} and m^{U+} , as this 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

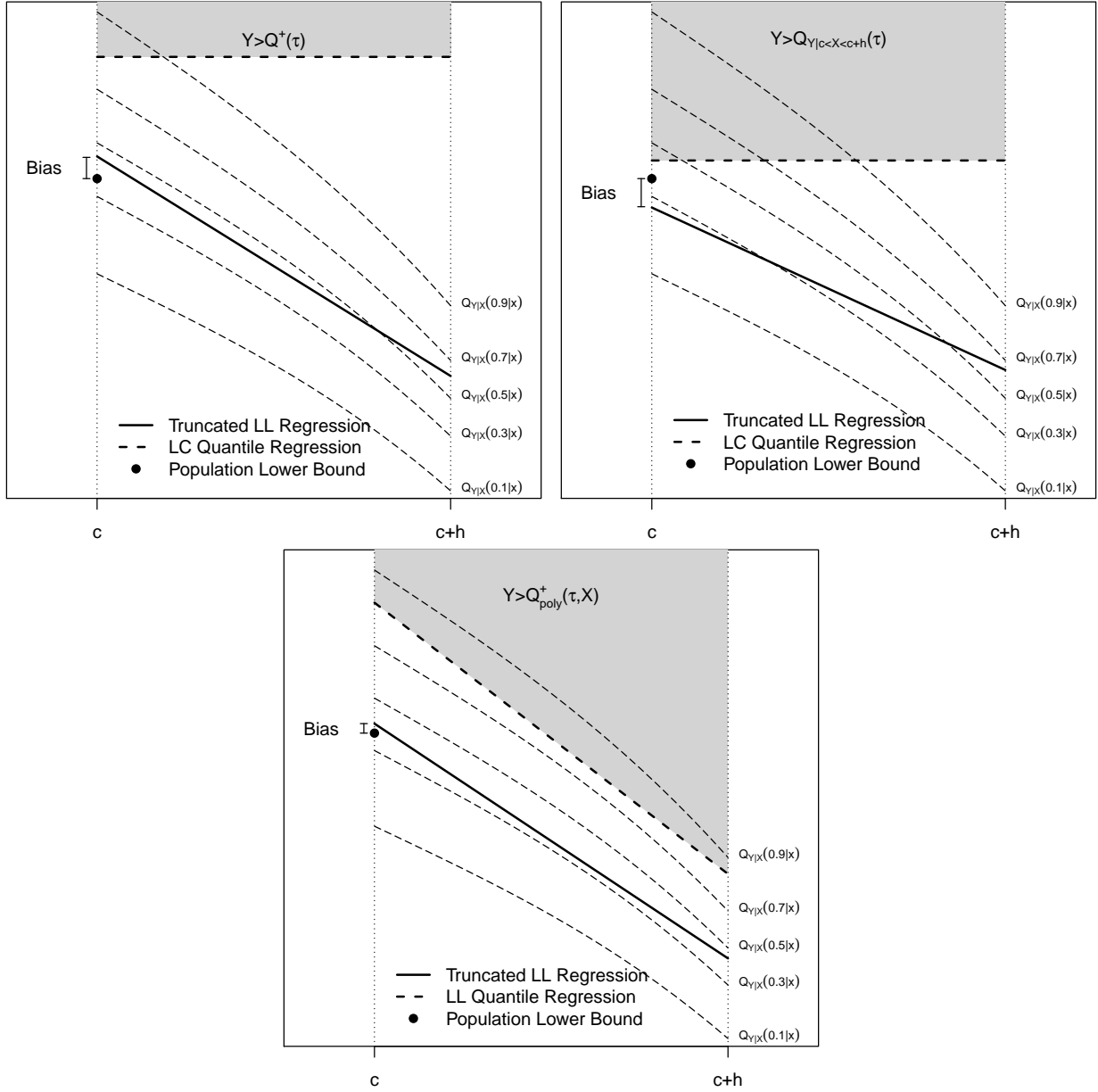


Figure 4.1: Illustration of the benefits of polynomial truncation.

has several undesirable properties, as illustrated in the top left panel of Figure 4.1 for the case of m^{L+} 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 going to be truncated

from this neighborhood will be substantially smaller than $\hat{\tau}$. And second, those units that are being truncated will 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+} and m^{U+} if a uniform kernel is used would be to run local polynomial regressions in the subsamples with $Y_i \leq \hat{Q}_{Y|c < X < c+h}(\hat{\tau})$ and $Y_i \geq \hat{Q}_{Y|c < X < c+h}(1 - \hat{\tau})$, respectively, where $\hat{Q}_{Y|c < X < c+h}$ 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 of the bounds. Stated more formally, our “polynomially truncated” local polynomial regression estimator of m^{L+} and m^{U+} can be described as follows. In a first step, we compute

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

which is the estimated p th order polynomial approximation of the conditional quantile function $x \mapsto Q_{Y|X}(t, x)$ in a local neighborhood to the right of the cutoff. In a second step, we then estimate m^{L+} and m^{U+} 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:

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

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

The same issue that complicates estimating m^{L+} and m^{U+} 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 p th 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

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

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

respectively, with

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

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

These estimates can easily be computed by any software package able to solve weighted least squares problems.

Estimating the remaining conditional expectations we mentioned above poses no partic-

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

$$\begin{aligned}\widehat{m}^- &= \mathbf{e}'_1 \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n (Y_i - \boldsymbol{\pi}_p(X_i - c)' \beta)^2 K_h(X_i - c) \mathbb{I}\{X_i < c\}, \\ \widehat{g}^+ &= \mathbf{e}'_1 \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n (D_i - \boldsymbol{\pi}_p(X_i - c)' \beta)^2 K_h(X_i - c) \mathbb{I}\{X_i \geq c\}, \\ \widehat{g}^- &= \mathbf{e}'_1 \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n (D_i - \boldsymbol{\pi}_p(X_i - c)' \beta)^2 K_h(X_i - c) \mathbb{I}\{X_i < c\},\end{aligned}$$

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

4.2. Estimation and Inference under the Conditions of Theorem 1

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

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

The asymptotic distribution of these estimators is derived formally in the Appendix. Under standard regularity conditions, in both cases the estimators of the respective upper and lower bound converge jointly in distribution to a bivariate Normal distribution at the same rate as a one-dimensional nonparametric regression.

The asymptotic variances of our estimators can be characterized explicitly in principle, but due their complicated form a “plug-in” sample analogue estimator would be difficult to implement. In our empirical application below, we therefore compute the standard error \widehat{s}^j of $\widehat{\Gamma}_{0,SRD}^j$ for $j \in \{U, L\}$ through the usual bootstrap procedure. By following the approach suggested in Imbens and Manski (2004), one can then construct confidence intervals $CI_{1-\alpha}$ that cover the parameter Γ_0 with at least probability $1 - \alpha$. Such confidence intervals are

given by

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

with r_α 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 $\Phi(\cdot)$ the CDF of the standard normal distribution. Note that for $0 < \alpha < .5$ 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 construction ensures that the confidence interval has good finite-sample coverage properties irrespective of the length of the identified set.

4.3. Estimation and Inference under the Conditions of Theorem 2

Estimating the lower and upper bounds on the treatment effect Γ_0 under the conditions of Theorem 2 is a more difficult problem due to the more complicated structure of the estimands. Our bounds take the form of a non-smooth functional of the population distribution of the data. Hirano and Porter (2012) show that for such parameters there generally exists no locally asymptotically unbiased estimator. They also show for any estimator of such a parameter a bias correction procedure that reduces bias too much will eventually cause the variance of the estimator to explode. Chernozhukov, Lee, and Rosen (2013) argue that a naive sample analogue estimator of parameters defined by applying an optimization operator to an unknown function will typically have a large bias, and propose an alternative estimator that achieves a reasonable balance between bias reduction and accuracy. The techniques they develop can also be used for the construction of valid confidence intervals.

To estimate our bounds on Γ_0 under the conditions of Theorem 2, we therefore consider a minor modification of the approach proposed by Chernozhukov, Lee, and Rosen (2013).

Recall the representation of the set \mathcal{T} given after Lemma 2, and note that the lower and upper bound on Γ_0 can be written as

$$\Gamma_{0,FRD}^L = \inf_{t \in [0,1]} \Gamma_0^L(t) \quad \text{and} \quad \Gamma_{0,FRD}^U = \sup_{t \in [0,1]} \Gamma_0^U(t),$$

respectively, where

$$\Gamma_0^j(t) = \frac{\Delta_0^j(\eta_1(t), \eta_0(t))}{\Psi_0(\eta_1(t))}$$

for $j \in \{L, U\}$. This notation, which integrates 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 depends on nuisance parameters. We then define the corresponding sample analogue

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

for $j \in \{L, U\}$, where

$$\begin{aligned} \widehat{\Delta}_0^j(t_1, t_0) &= \widehat{m}^{j+}(t_1, t_0, 1) \cdot \frac{\widehat{g}^+(1-t_1)}{1-\widehat{\tau}} + \widehat{m}^{j+}(t_1, t_0, 0) \cdot \frac{(1-\widehat{g}^+)(1-t_0)}{1-\widehat{\tau}} - \widehat{m}^-, \\ \widehat{\Psi}_0(t_1) &= \frac{\widehat{g}^+ \cdot (1-t_1)}{1-\widehat{\tau}} - \widehat{g}^-, \quad \text{and} \quad \widehat{\eta}_d(t) = \widehat{\tau}_d^L + t \cdot (\widehat{\tau}_d^U - \widehat{\tau}_d^L); \end{aligned}$$

and estimates of the endpoints of the line that constitutes our estimates of the set \mathcal{T} are given by

$$\begin{aligned} \widehat{\tau}_1^L &= \max \left\{ 0, 1 - \frac{1-\widehat{\tau}}{\widehat{g}^+} \right\}, & \widehat{\tau}_1^U &= \min \left\{ 1 - \frac{(1-\widehat{\tau}) \cdot \widehat{g}^-}{\widehat{g}^+}, \frac{\widehat{\tau}}{\widehat{g}^+} \right\}, \\ \widehat{\tau}_0^L &= \min \left\{ 1, \frac{\widehat{\tau}}{1-\widehat{g}^+} \right\}, & \widehat{\tau}_0^U &= \max \left\{ 0, \widehat{\tau} - \frac{(1-\widehat{\tau}) \cdot (\widehat{g}^+ - \widehat{g}^-)}{1-\widehat{g}^+} \right\}. \end{aligned}$$

Simply minimizing $\widehat{\Gamma}_0^L(t)$ over $t \in [0, 1]$ would yield an estimator of $\Gamma_{0,FRD}^L$ that is downward-biased, and maximizing $\widehat{\Gamma}_0^U(t)$ over $t \in [0, 1]$ would yield an estimator of $\Gamma_{0,FRD}^U$ that is upward-biased. These naive estimators would thus tend to overstate the degree of partial identification, and produce conservative estimates of the identified set that are generally too

wide. The strategy proposed by Chernozhukov, Lee, and Rosen (2013) to address this issue involves adding a precision-correction term to an estimate of the bounding function before minimizing or maximizing it. In our context, the resulting estimates are⁶

$$\widehat{\Gamma}_{0,FRD}^L = \inf_{t \in [0,1]} \left(\widehat{\Gamma}_0^L(t) + \widehat{k}^L \cdot \widehat{s}^L(t) \right) \quad \text{and} \quad \widehat{\Gamma}_{0,FRD}^U = \sup_{t \in [0,1]} \left(\widehat{\Gamma}_0^U(t) - \widehat{k}^U \cdot \widehat{s}^U(t) \right),$$

respectively, where for $j \in \{U, L\}$ the term $\widehat{s}^j(t)$ is the pointwise standard error of $\widehat{\Gamma}_0^j(t)$, and \widehat{k}^j is an estimate of the median of the distribution of the random variable $\sup_{t \in [0,1]} (\widehat{\Gamma}_0^j(t) - \Gamma_0^j(t))/s^j(t)$, with $s^j(t)$ the population version of $\widehat{s}^j(t)$. A similar approach leads to a confidence interval for Γ_0 with nominal level $1 - \alpha$:

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

Here \widehat{k}_α^j is an estimate of the α -quantile of the distribution of $\sup_{t \in [0,1]} (\widehat{\Gamma}_0^j(t) - \Gamma_0^j(t))/s^j(t)$.

Constructing the estimates $\widehat{s}^j(t)$, \widehat{k}^j , and \widehat{k}_α^j for $j \in \{U, L\}$ faces some additional complications in our case. While one can in principle use the bootstrap to estimate the distribution of the various intermediate quantities, there are two difficulties when these estimators are transformed into the function $\widehat{\Gamma}_0^j(t)$. The first occurs if the denominator $\Psi_0(\eta_1(t))$ is close to zero (relative to the sampling uncertainty of the corresponding plug-in estimator), and the second shows up if one of the various max and min operators in the definition of the function $\eta_a(\cdot)$ is close to binding. Fortunately, in many empirical applications, including the one that we study below, none of these two issues appears, and thus the bootstrap can be used. More formally, the procedure that we propose below is going to be valid under the condition that the proportion τ of manipulators just to the right of the cutoff is relatively

⁶To simplify the exposition, we omit an analogue of the “inequality selection” step in Chernozhukov, Lee, and Rosen (2013). Adding such an additional layer to our estimation procedure would lead to a slightly more narrow estimate of the identified set.

small, in the sense that

$$\tau < \min \left\{ 1 - \mathbb{E}(D_i|X_i = c^+), \frac{\mathbb{E}(D_i|X_i = c^+) - \mathbb{E}(D_i|X_i = c^-)}{1 - \mathbb{E}(D_i|X_i = c^-)} \right\}. \quad (4.1)$$

This condition appears reasonable for many applications, including the one we study below, but of course does not hold universally. However, its plausibility can be investigated empirically by checking whether $\hat{\tau}$ is substantially smaller than estimates of the two quantities on the right-hand side of (4.1), accounting for sampling uncertainty. Condition (4.1) implies that

$$\tau_1^U = \tau/\mathbb{E}(D_i|X_i = c^+), \quad \tau_0^L = \tau/(1 - \mathbb{E}(D_i|X_i = c^+)), \quad \text{and} \quad \tau_1^L = \tau_0^U = 0$$

In turn, this means that all the max and min operators in the definition of the bounding functions $\Gamma_0^j(t)$ are redundant, and can thus be ignored when studying the asymptotic properties of $\hat{\Gamma}_0^j(t)$. Under some standard regularity conditions, one can then show validity of the bootstrap-based algorithm for estimating $\hat{s}^j(t)$, \hat{k}^j , and \hat{k}_α^j for $j \in \{U, L\}$:

1. Generate bootstrap samples $\{Y_{i,b}^*, D_{i,b}^*, X_{i,b}^*\}_{i=1}^n$, $b = 1, \dots, B$ by sampling with replacement from the original data $\{Y_i, D_i, X_i\}_{i=1}^n$ for some large integer B .
2. Calculate bootstrap estimates $\hat{\Gamma}_{0,b}^{j*}(t)$, $b = 1, \dots, B$ of the bounding functions from the bootstrap data for $j \in \{U, L\}$.
3. Put $\hat{s}^j(t)$ as the sample standard deviation of $\{\hat{\Gamma}_{0,b}^{j*}(t)\}_{b=1}^B$ for $j \in \{U, L\}$.
4. Put \hat{k}^j as the median and \hat{k}_α^j as the α quantile of $\{\max_{t \in \mathcal{U}}(\hat{\Gamma}_{0,b}^{j*}(t) - \hat{\Gamma}_0^j(t))/\hat{s}^j(t)\}_{b=1}^B$ for $j \in \{U, L\}$, where \mathcal{U} is some fine grid over the unit interval.

Here validity of the algorithm means that the resulting estimators $\hat{\Gamma}_{0,FRD}^L$ and $\hat{\Gamma}_{0,FRD}^U$ of the bounds possess the half-median-unbiasedness property described in Chernozhukov, Lee, and Rosen (2013), and that $CI_{1-\alpha}$ is an asymptotically valid $1 - \alpha$ confidence interval for the parameter Γ_0 .

4.4. Estimation and Inference under the Conditions of Theorems 3 and 4

Estimation and inference 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 the empirical implementation of our refined bounds in these cases rather brief. Following the remark after Theorem 3, estimates $\widehat{\Gamma}_{0,FRD(a)}^L$ and $\widehat{\Gamma}_{0,FRD(a)}^U$ of $\Gamma_{0,FRD(a)}^L$ and $\Gamma_{0,FRD(a)}^U$ can be obtained by proceeding exactly as in Section 4.3, after redefining $\widehat{\tau}_1^L = \widehat{\tau}_0^L = \widehat{\tau}$. The construction of confidence intervals is also analogous. Estimates of the bounds $\Gamma_{0,FRD(b)}^L$ and $\Gamma_{0,FRD(b)}^U$ from Theorem 4 can be constructed as

$$\widehat{\Gamma}_{0,FRD(b)}^L = \frac{\widehat{\Delta}_0^L(\widehat{\tau}/\widehat{g}^+, 0)}{\widehat{\Psi}_0(\widehat{\tau}/\widehat{g}^+)} \quad \text{and} \quad \widehat{\Gamma}_{0,FRD(b)}^U = \frac{\widehat{\Delta}_0^L(\widehat{\tau}/\widehat{g}^+, 0)}{\widehat{\Psi}_0(\widehat{\tau}/\widehat{g}^+)},$$

respectively, using the notation introduced in Section 4.3. The theoretical properties of these estimators can be analyzed using the same arguments employed for the bounds in the sharp RD design, and Imbens-Manski-style confidence intervals can be obtained in the same fashion as well.

Remark 5. While by construction $\Gamma_{0,FRD}^L \leq \Gamma_{0,FRD(a)}^L \leq \Gamma_{0,FRD(b)}^L$, it is not necessarily the case that $\widehat{\Gamma}_{0,FRD}^L \leq \widehat{\Gamma}_{0,FRD(a)}^L \leq \widehat{\Gamma}_{0,FRD(b)}^L$ in finite samples; and similarly for the respective upper bounds. One reason why the order of the estimated bounds might not correspond to the order of their population values is that these three objects can have quite different finite sample biases. These differences in biases could potentially be larger than the differences in the population values of the bounds. We see this phenomenon actually occur in our empirical application below.

5. EMPIRICAL APPLICATION

5.1. Motivation

Our empirical application considers the unemployment insurance (UI) program in Brazil. UI is a relevant program to consider. It typically varies benefits based on some discontinuous

rules. For instance, eligibility often requires a minimum number of months of employment prior to layoff. Workers have thus incentives to avoid being laid off prior to the cutoff, possibly postponing the timing of their layoff. Workers and firms have also higher incentives to terminate an employment relationship passed the eligibility cutoff, possibly increasing layoffs or bringing forward their timing. These incentives could lead to manipulation of the running variable, breaking down the point identification of the effects of UI eligibility at the cutoff. Such manipulation is particularly likely when firms do not internalize the externality that they impose on the UI budget when laying off a worker at a given time. This is the case for instance when UI benefits are not perfectly experienced-rated, i.e., when the level of the UI tax paid by the firm does not increase one-to-one with the benefits drawn by its workers. UI benefits are rarely perfectly experienced-rated in practice, even though theory suggests that it should (Blanchard and Tirole, 2006), and this is thought to significantly increase layoffs (e.g., Feldstein, 1976; Anderson and Meyer, 2000).

Brazil is also a relevant country to study. UI programs have been adopted or considered in a growing number of developing countries. A distinctive feature of their labor markets is the prevalence of both formal (registered) and informal (unregistered) workers. It is widely believed that the usual moral hazard problem with UI –the fact that UI distorts incentives to return to a formal job– may be more severe in a context of high informality because workers can work informally while drawing UI benefits. Yet, the existing evidence remains limited (Gerard and Gonzaga, 2014). Moreover, providing new evidence may be challenging because manipulation of UI eligibility may be more likely in a context of high informality.⁷ The utility costs of being (reported as) laid off after the cutoff 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.

⁷Another example of manipulation for eligibility for a social program in a developing country is Camacho and Conover (2011).

We use our new estimators below to bound the effects of UI eligibility on non-formal-employment duration (i.e., the time between two formal jobs) for non-manipulators at a specific eligibility cutoff. Our main purpose is to illustrate the applicability of our approach. However, the results are relevant for their own sake. Such a treatment effect is an important input to evaluate the optimal design of UI programs.⁸ Moreover, non-manipulators would be those still laid off at the cutoff if firms (and workers) were not increasing layoffs or bringing forward their timing in response to UI eligibility, for instance if UI benefits were fully experience-rated (as they should). Finally, the bounds have an economic interpretation. The lower bound for the impact on non-formal-employment duration captures the treatment effect under adverse selection: manipulators are assumed to stay longer without a formal job than non-manipulators. Conversely, the upper bound assumes advantageous selection: manipulators are assumed to be better able to find a new formal rapidly.

5.2. *Institutional Details and Data*

In Brazil, a worker who is reported as involuntarily laid off from a private formal job is eligible for UI under two conditions. She must have at least six months of continuous job tenure at layoff. Additionally, 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 this second condition for our application.⁹ Eligible workers can draw three, four, or five months of UI benefits if they have accumulated more than 6, 12, or 24 months of formal employment in the 36 months prior to layoff. Unfortunately, our data do not allow us to precisely construct a worker's accumulated tenure as measured by the UI agency around the 16-month cutoff (see Gerard and Gonzaga, 2014, for details). The benefit level depends

⁸See for instance Baily (1978) and Chetty (2006). The literature typically refers to the effect on non-employment duration because it considers countries where all jobs are assumed to be formal.

⁹There is also evidence of manipulation around the six-month cutoff (Gerard and Gonzaga, 2014). However, the 16-month cutoff appears more arbitrary and thus less likely to coincide with other policy discontinuities. In contrast, six months of job tenure may be a salient milestone for evaluating employees' performance.

nonlinearly on the average wage in the three months prior to layoff. The replacement rate is 100% at the bottom of the wage distribution but is already down to 60% for a worker who earned three times the minimum wage (the full benefit schedule is provided in the Appendix). Finally, UI benefits are not experienced-rated in Brazil.

We exploit two administrative datasets. The first one is a longitudinal matched employee-employer dataset covering by law the universe of formal employees. Importantly, the data include hiring and separation dates, and the reason for separation.¹⁰ The second dataset is the registry of all UI payments. Individuals in both datasets are identified through the same ID number. We use data from 2002 to 2010. In 2009, there were about 40,300,000 formal employees and 625,650 new UI beneficiaries in each month. 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.

5.3. *Sample Selection*

Our sample of analysis is constructed as follows. First, we consider all workers, 18 to 55 years old, who have lost a private full-time formal job and drawn UI benefits between 2002 and 2010. Workers who find a new formal job before exhausting their benefits are entitled to draw the remaining benefits following a new layoff in Brazil, 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 cutoff is sharp. To do so, we focus on workers who drew the maximum number of UI benefits (five). Second, 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 (other eligibility condition) and we focus on workers whose new layoff date falls within 120 days of the 16-month eligibility

¹⁰Every year, firms must report workers formally employed during the previous calendar year. The data also include information on wage, tenure, age, race, gender, education, sector of activity, and establishment size and location.

cutoff. Third, we drop workers whose first layoff date fell after the 29th of a month. The policy rules create bunching at the 16-month cutoff even in absence of manipulation for workers laid off in the last few days of a month. This is because the relevant month, 16 months later, may not have more than 28, 29, or 30 days. For instance, all workers laid off between October 29th and October 31st in 2007 became eligible on February 28th, 2009. Finally, we follow workers for at least one year after the new layoff date. As a result, we drop workers laid off in 2010.

Our resulting sample of analysis 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.¹¹ Those are not the characteristics of the average formal employee or the average UI beneficiary in Brazil, but characteristics of workers for whom the 16-month cutoff may be binding.

5.4. *Graphical Evidence*

Figure 5.2 displays some patterns in our data graphically. Observations are aggregated by day between the layoff date and the eligibility cutoff date according to the 16-month rule. Panel A provides evidence of manipulation. The share of observations located to the right of the cutoff is higher than the share located to the left of the cutoff by about 10%. 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 on the left of the cutoff, but it discontinuously jumps from 40% to 80% at the cutoff. Panel C shows that the eligibility rule was enforced. The share of workers drawing some UI benefits is close to zero on the left of the cutoff, but it increases to 80%

¹¹They 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 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.

at the cutoff. Workers on the right of the cutoff 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 remain without a formal job for at least one year after layoff. It increases discontinuously from 25% to 30% at the cutoff. Similarly, panel F shows a clear discontinuity in non-formal-employment duration (censored at one year), which increases from 170 days to 220 days at the cutoff.

The discontinuities in panels E and F could be due to a treatment effect. They could also be due to a selection effect. Workers on each side of the cutoff may have different potential outcomes in the presence of manipulation. Our new estimators allow us to bound treatment effects for non-manipulators, despite the possibility of sample selection.

5.5. *Bounds*

We implement our estimators for the outcome in panel F, non-formal-employment duration. We censor the outcome at one year such that we can measure it for all workers in our sample. The censoring also serves an illustrative purpose. The share of right censored observation is always larger than the estimated share of manipulators. Our lower bound thus always truncates a discrete outcome (whether an observation is censored or not), while our upper bound essentially truncates a continuous outcome (uncensored durations).¹² Results are presented in Table 1 for an edge kernel (Cheng, Fan, and Marron, 1997) and a bandwidth of ten days around the cutoff.¹³ For bounds in the FRD case that involve numerical optimization, we use a grid search to look for the infimum and supremum using 51 values for $t \in [0, 1]$. 95% confidence intervals are based on 500 bootstrap samples.

Panel A of Table 1 shows that UI take-up increases by 75.5 percentage points at the

¹²We present the distribution of our outcome variable on the left and on the right of the cutoff in the Appendix.

¹³We do not have theoretical results on the optimal bandwidth for the estimation of our bounds. The optimal bandwidth for a naive RD design that does not take manipulation into account is 8.1 days using Imbens and Kalyanaraman (2012). Our estimates are similar if we use bandwidths of 30 or 50 days.

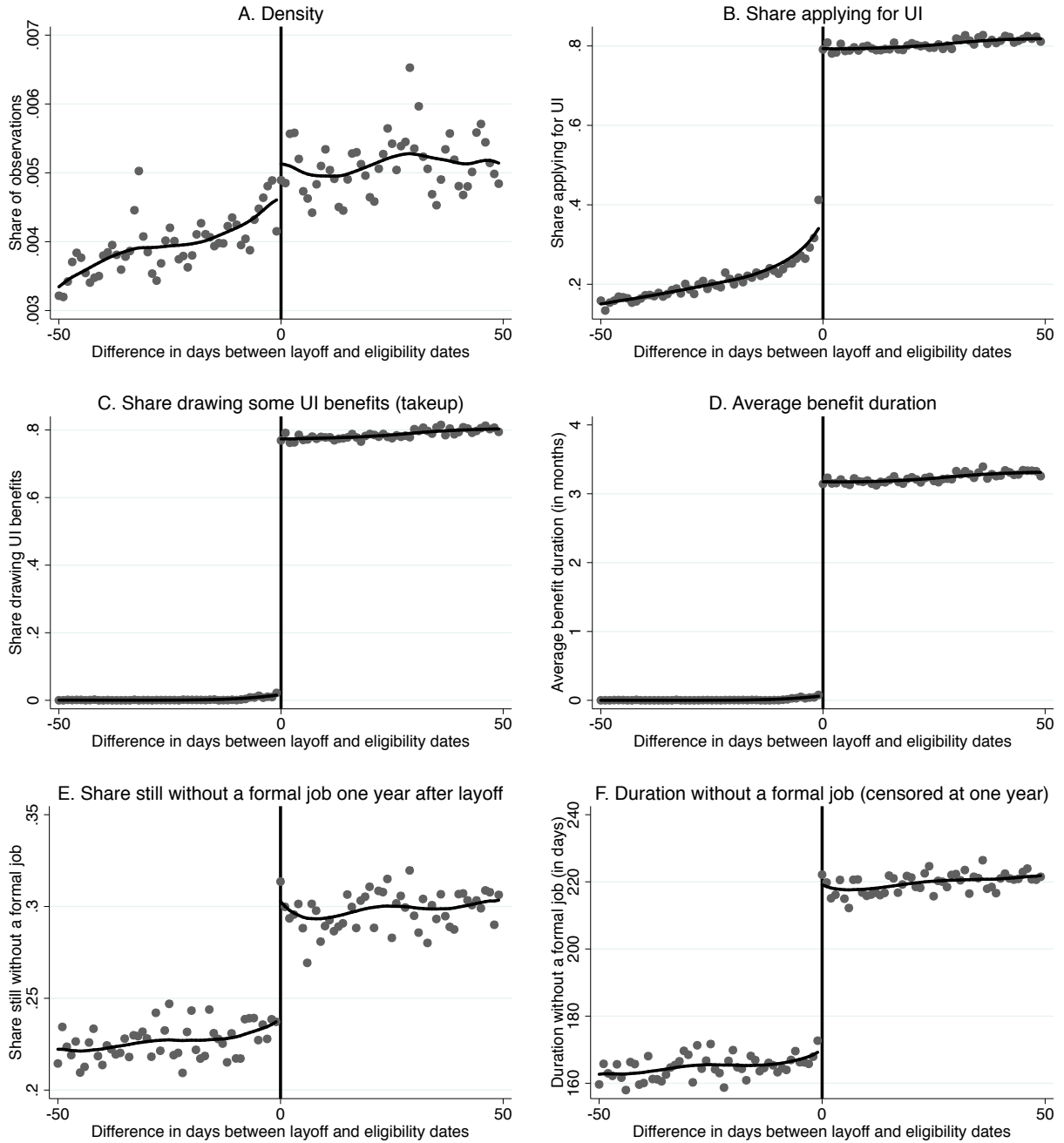


Figure 5.2: Graphical evidence for our empirical application

cutoff, but that manipulators are estimated to account for 10.1% of observations on the right of the cutoff. We then report results from two types of exercises. First, we consider a

SRD design in which *eligibility* for UI benefits is defined as the treatment of interest. 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 uptake of UI benefits being the treatment. Naive RD estimates that assume no manipulation yield a 49.1-day increase in non-formal-employment duration (ITT/SRD) and a 65.1-day increase for UI takers (LATE/FRD). Those estimates, however, confound treatment and selection effects. Table 1 thus also provides estimates of our bounds for the treatment effects, using SRD formulas (resp. FRD formulas) for the ITT estimates (resp. LATE estimates). The bounds are relatively large, which is not surprising given the extent of manipulation. Point estimates for the standard upper bounds are 72.6 days and 88.3 days for the ITT and LATE estimates, respectively. Point estimates for the standard lower bounds are 32.8 days (ITT) and 45.8 days (LATE). Our two refinements seem to have no meaningful additional identifying power in this case, as they yield estimates that are qualitatively very similar to the standard bounds.¹⁴

In sum, a naive RD design that ignores manipulation estimates an increase in censored duration of 28.8% at the eligibility cutoff. Our bounds imply that the magnitude of this estimate may be heavily affected by selection, but that the treatment effect for non-manipulators is an increase in censored duration of at least 15.5% (lower limit of the confidence interval). Therefore, granting UI eligibility at the existing cutoff would entail significant behavioral responses, and increase UI costs, even if UI benefits were fully experience-rated. Of course, this estimate is particular to the sample of workers for whom the cutoff may be binding. It must also be put in perspective with the welfare gains from providing these workers with additional insurance. For example, the welfare gain is arguably lower for workers with high turnover rates, for whom layoff is an expected event.

¹⁴Note that our estimation procedures are not constructed such that the estimates of the refined bounds are strictly more narrow than those of the standard bounds. The fact that estimate of the lower bound under refinement B is actually slightly smaller than the estimate of the standard lower bound, for example, is most likely due to differences in small sample bias between these two estimation procedures.

Table 1: Results from our empirical application

	Point/Interval Estimate	95% Conf. Interval
A. Basic Inputs		
UI take-up at cutoff	.7553	[.7435, .767]
Share of manipulators (τ)	.1014	[.0704, .1324]
B. ITT/SRD estimates		
ITT - Ignoring Manipulation	49.14	[44.07, 54.22]
ITT - Standard Bounds (Th.1)	[32.81, 72.63]	[26.35, 80.46]
C. LATE/FRD estimates		
LATE - Ignoring Manipulation	65.06	[58.54, 71.59]
LATE - Standard Bounds (Th.2)	[45.75, 88.32]	[38.85, 96.68]
LATE - Refinement A (Th.3)	[43.79, 87.82]	[34.53, 96.75]
LATE - Refinement B (Th.4)	[45.21, 88.28]	[35.95, 96.13]

Total number of observations within our bandwidth of 10 days around the cutoff: 42037. The counterfactual value of the outcome at the cutoff for non-manipulators (\hat{m}^-) was estimated at 170.4.

6. CONCLUSIONS

In this paper, we propose a partial identification approach to dealing with 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. We recommend the use of such confidence intervals in applications irrespective of the outcome of McCrary’s (2008) test for manipulation. The approach is illustrated with an application to the Brazilian unemployment insurance (UI) system.

A. PROOFS OF IDENTIFICATION RESULTS

A.1. Proof of Lemma 1

Since the density of X is continuous around c among non-manipulators 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. \square

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

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

and

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

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 “manipulator” is randomly assigned

with probability τ_d to units with treatment status d and running variable above 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 . \square

A.3. Proof of Theorem 1

By Assumption 3, we have that $\mathbb{E}(Y_i | X_i = c^-) = \mathbb{E}(Y_i | X_i = c^-, M_i = 0)$. Thus, we only need to show that $\mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q_{Y|X}(1 - \tau | c^+))$ and $\mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q_{Y|X}(\tau | c^+))$ are sharp lower and upper bounds for $\mathbb{E}(Y_i | X_i = c^+, M_i = 0)$. 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. \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 $\Delta_0^U(\tau_1, \tau_0)$ are sharp lower and upper bounds for Δ_0 . By Assumption 3, $\mathbb{E}(Y_i | X_i = c^-) = \mathbb{E}(Y_i | X_i = c^-, M_i = 0)$. Following the main text's argument, we thus only need to show that

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

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 $\mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_1 | c^+, 1), D_i = 1)$ and $\mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q_{Y|XD}(1 - \tau_0 | c^+, 0), D_i = 0)$ 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 $\mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_1 | c^+, 1), D_i = 1)$ and $\mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q_{Y|XD}(\tau_0 | c^+, 0), D_i = 0)$ 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. \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 $\mathbb{P}(D_i = 1 | X_i = c^+, M_i = 1) \geq \mathbb{P}(D_i = 1 | X_i = c^+, M_i = 0)$ 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 $\mathbb{P}(D_i = 1 | X_i = c^+, M_i = 1) = 1$ implies $\tau_1 = \tau/\mathbb{E}(D_i | X_i^* = c^-)$. In addition, $\mathbb{P}(D_i = 1 | X_i = c^+, M_i = 1) = 1$ implies that $\tau_0 = 0$. \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) $\mathbb{E}(Y_i | X_i = x, Y_i \leq Q^+(\tau))$, $\mathbb{E}(Y_i | X_i = x, Y_i \geq Q^+(1 - \tau))$, $Q_{Y|X}(\tau, x)$ 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) $\mathbb{V}(Y_i | X_i = x, Y_i \leq Q^+(\tau))$, $\mathbb{V}(Y_i | X_i = x, Y_i \geq Q^+(1 - \tau))$ 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 $nh^{2p+1} \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$.

Under these conditions, we have the following result.

Theorem B.1. Suppose that Assumptions B.1–B.4 hold. Then

$$\sqrt{nh} \begin{pmatrix} \widehat{\Gamma}_{0,SRD}^L - \Gamma_{0,SRD}^L \\ \widehat{\Gamma}_{0,SRD}^U - \Gamma_{0,SRD}^U \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

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

Joint normality of the estimators of the upper and lower bound is an important condition for the results in Imbens and Manski (2004) to apply. To prove the theorem, we begin by defining the following shorthand notation for various derivatives:

$$\begin{aligned} m_Q^{L+} &= \partial_t \mathbb{E}(Y_i | X_i = c^+, Y_i \leq t) |_{t=Q^+(\tau)}, & m_\tau^{L+} &= \partial_t \mathbb{E}(Y_i | X_i = c^+, Y_i \leq Q^+(t)) |_{t=\tau}, \\ m_Q^{U+} &= \partial_t \mathbb{E}(Y_i | X_i = c^+, Y_i \geq t) |_{t=Q^+(1-\tau)}, & m_\tau^{U+} &= \partial_t \mathbb{E}(Y_i | X_i = c^+, Y_i \geq Q^+(1-t)) |_{t=\tau}. \end{aligned}$$

It will also be useful to 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:

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

Finally, with $\mathbf{0} = (0, 0, \dots, 0)'$, $\mathbf{1} = (1, 1, \dots, 1)'$ and $\mathbf{h}_p = (1, h, h^2, \dots, h^p)'$, 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 (scaled) intermediate estimators:

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

The statement of the theorem then follows from Lemma B.1–B.2 below, with $\Sigma = AVA'$ for A as defined above and V as given in Lemma B.1.

Lemma B.1. *Suppose that Assumptions B.1–B.4 hold. Then $\sqrt{nh} \cdot \mathbf{v} \xrightarrow{d} N(0, V)$ for some positive definite matrix V .*

Proof. (Sketch; the be completed) It follows from standard arguments 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 that is $o_P((nh)^{-1/2}) + h^{p+1}$. 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. \square

Lemma B.2. *Suppose that Assumptions B.1–B.4 hold. Then the estimators \widehat{m}^{L+} and \widehat{m}^{U+} satisfy the following stochastic decomposition:*

$$\begin{aligned} \widehat{m}^{L+} &= \tilde{m}^{L+} + m_{Q(\tau)}^{L+} \cdot \mathbf{h}'_p (\widehat{Q}_{\nabla}^+(\tau) - Q_{\nabla}^+(\tau)) + m_{\tau}^{L+} \cdot (\widehat{\tau} - \tau) + o_P((nh)^{-1/2}) + O(h^{p+1}) \quad \text{and} \\ \widehat{m}^{U+} &= \tilde{m}^{U+} + m_{Q(1-\tau)}^{U+} \cdot \mathbf{h}'_p (\widehat{Q}_{\nabla}^+(1-\tau) - Q_{\nabla}^+(1-\tau)) + m_{\tau}^{U+} \cdot (\widehat{\tau} - \tau) + o_P((nh)^{-1/2}) + O(h^{p+1}). \end{aligned}$$

Proof. (Sketch; the be completed) Follows from stochastic equicontinuity arguments. \square

C. APPENDIX TABLES AND GRAPHS

We present here some supporting graphs. Figure C.3 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 (10-day window around the cutoff). Figure C.4 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.

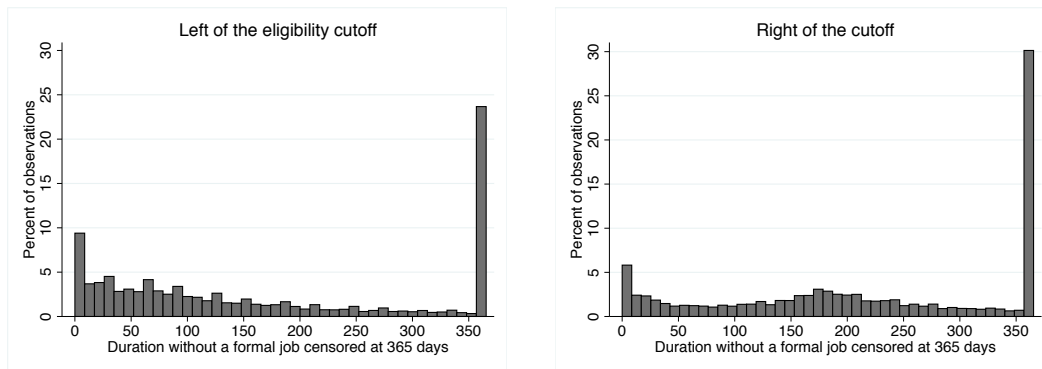


Figure C.3: Distribution of our outcome variable on each side of the cutoff

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 (10-day window on each side of the cutoff).

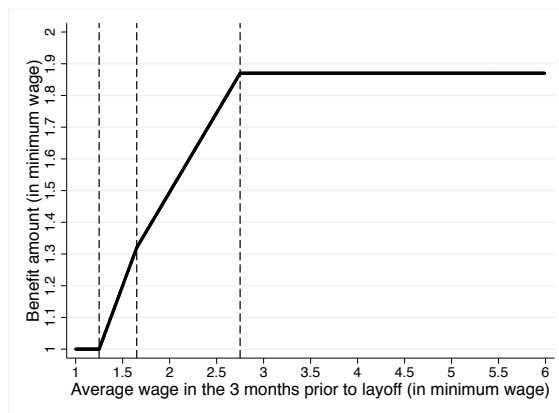


Figure C.4: Monthly UI benefit amount

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.

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., AND X. SHI (2013): “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 81(2), 609–666.
- 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.
- BLANCHARD, O., AND J. TIROLE (2006): “The Joint Design of Unemployment Insurance and Employment Protection: A First Pass,” *Journal of the European Economic Association*, 6(1), 45–77.
- BLUNDELL, R., A. GOSLING, H. ICHIMURA, AND C. MEGHIR (2007): “Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds,” *Econometrica*, 75(2), 323–363.

- 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, 41–64.
- 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. (2006): “A General Formula for the Optimal Level of Social Insurance,” *Journal of Public Economics*, 90, 1879–1901.
- 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.
- FRANSEN, 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 KLAUW (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.
- 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.
- ROKKANEN, M. (2015): “Exam Schools, Ability, and the Effects of Affirmative Action: Latent Factor Extrapolation in the Regression Discontinuity Design,” Unpublished manuscript.
- ROMANO, J., AND A. SHAIKH (2010): “Inference for the Identified Set in Partially Identified Econometric Models,” *American Economic Journal: Economic Policy*, 78(1), 169–211.
- RUBIN, D. (1974): “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies,” *Journal of Educational Psychology*, 66(5), 688–701.
- SAEZ, E. (2010): “Do Taxpayers Bunch at Kink Points?,” *American Economic Journal: Economic Policy*, 2(3), 180–212.
- 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.
- 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.