An Alternative Assumption to Identify LATE in Regression Discontinuity Designs

Yingying Dong
University of California Irvine*

September 2014

Abstract

One key assumption Imbens and Angrist (1994) use to identify a Local Average Treatment Effect (LATE) is independence of the instrument from potential treatment and potential outcomes. Hahn, Todd and van der Klaauw (2001) employ a local version of this assumption to identify a LATE in Regression Discontinuity (RD) models. This paper shows that this local independence assumption may not hold in many empirical applications, and that this assumption can be replaced with an empirically plausible and partially testable weak behavioral assumption. Given no defiers and a discontinuity in the treatment probability, LATE in both sharp and fuzzy designs is identified under just smoothness conditions. The required smoothness can be satisfied when there is no sorting near the RD threshold. The independence assumption is partially testable given smoothness. An empirical application is provided.

JEL Codes: C21, C25
Keywords: Regression discontinuity, Sharp design, Fuzzy design, LATE, Treatment effect heterogeneity, Treatment effect derivative, Selection into compliers, Sorting, RD density test

1 Introduction

Regression discontinuity (RD) designs have been widely used in many areas of empirical research. In a seminal paper, Hahn, Todd and van der Klaauw (2001, hereafter HTV) provide a set of formal assumptions for identifying and estimating a local average treatment effect (LATE) using RD designs. A key assumption used by HTV is that the treatment effect and potential treatment status are jointly

*I would like to thank Josh Angrist, Arthur Lewbel, Yaming Yu, Damon Clark, Marianne Bitler, Kevin Roth, and David Powell for their insightful comments. I am also grateful to Thomas Lemieux for help with the election data. All errors are my own.

Yingying Dong, Department of Economics, 3151 Social Science Plaza, University of California Irvine, CA 92697-5100, USA. Email: yyd@uci.edu. http://yingyingdong.com/

independent of the running variable in a neighborhood of the RD cutoff.[1] This assumption is based on the independence assumption in the original LATE paper by Imbens and Angrist (1994).

Most of the empirical studies using RD designs do not cite any research that shows that their estimates can be given a causal interpretation, while those do cite HTV (2001).[2] However, this paper shows that the HTV type independence assumption may not hold in empirical applications and that this assumption can be replaced with a weaker, empirically more plausible and partially testable behavioral assumption. The running variable is frequently one of the key determinants of (or at least is correlated with) outcomes, restricting the treatment effect to be independent of the running variable therefore places an undesirable restriction on treatment effect heterogeneity.[3,4] As I show later, this type of heterogeneity would arise naturally in many RD applications.

This paper provides formal smoothness conditions that suffice to identify both sharp and fuzzy design RD without the HTV independence assumption. I also provides a weak behavioral assumption, in the spirit of Lee (2008), that leads to the required smoothness to hold in both sharp and fuzzy RD designs. Results in this paper provide formal support for the popular practice of performing McCrary's (2008) density test to assess validity of fuzzy design RD. Given minimal further smoothness, I show that one may partially test the HTV independence assumption. These results are applied to the data of Lee (2008) showing that the identifying assumption in this paper plausibly holds, but the HTV independence assumption likely does not.

Lee (2008) discusses continuity of the conditional density (conditional on an individual's 'identity') of the running variable to establish local randomization and hence causal inference. The discussion focuses on sharp design RD. Lee further shows that this assumption generates strong testable implications, i.e., individuals cannot precisely manipulate the running variable to be just above or

[1]The alternative assumption HTV impose is either that the treatment effect is constant across individuals or that the treatment be independent of the treatment effect conditional on the running variable near the RD threshold. As HTV note, they rule out self-selection into treatment based on idiosyncratic gains, and so is often not be realistic.

[2]E.g., looking at the 150 published empirical RD studies in the top general-interest or field journals in the past 10 years, 62 of them cited HTV (2001), while the other either just mentioned the survey papers by Imbens and Lemieux (2008) or Lee and Lemieux (2010) or did not cite any.

[3]See, e..g., Chapter 6 of Angrist and Pischke (2008) for discussion of cases where the treatment effect is allowed to depend on the running variable.

[4]The RD treatment effect can potentially be a function of relevant observed and unobserved covariates, so it may either directly depend on the running variable or is correlated with it through other covariates. A dependency means a non-zero average derivative of the RD treatment effect with respective to the running variable.

just below the cutoff so that covariate means should be balanced on either side of the cutoff. This paper formally establishes RD identification based on smoothness conditions and then similarly relate smoothness to a lack of manipulation over the running variable. The discussion instead focuses on fuzzy design RD, with sharp design following as a special case. This requires dealing with (via smoothness assumptions) the probabilities with which individuals may self-select into types such as compliers or always takers. The discussion leads to precisely the same identification results as those by HTV, but under a weaker assumption.

The rest of the paper is organized as follows: The next section provides some motivating empirical examples; Section 3 formally shows RD identification under alternative assumptions; Section 4 discusses testing the independence assumption by HTV; Section 5 provides an empirical application of the results. Short concluding remarks are provided in Section 6.

2 Motivating Examples

To motivate the discussion in this paper, consider the standard sharp design RD model estimating the electoral advantage of incumbency in the US house of representatives in Lee (2008). The treatment is an indicator that the Democratic Party was incumbent. The running variable is the Democratic Party's winning margin, or the difference between the Democratic Party's vote share and its strongest opponent's. The outcome is whether a Democrat won the next election. In this case, the independence assumption by HTV would require that the incumbent party's electoral advantage does not depend on its winning margin. Figure 1, which is reproduced from Figure 5-(a) in Lee (2008), provides clear evidence against this assumption.[5]

Figure 1 shows how the probability of a Democrat winning in election $t+1$ depends on its winning margin in election t. The slope gets steeper right above the threshold, implying that the larger is the incumbent party's share in the previous election, the greater is their chance of winning the next election, i.e., the incumbency advantage depends on the winning margin.

Consider another RD model estimating the effect of the Adams Scholarship program on college choices (Goodman, 2008). The Adams Scholarship program provides qualified students tuition wa-

[5]Intuitively, the incumbent party's electoral advantage may depend on its winning margin either directly or indirectly because, e.g., winning margin is positively correlated with a party's strength. In either case, the independence assumption fails.

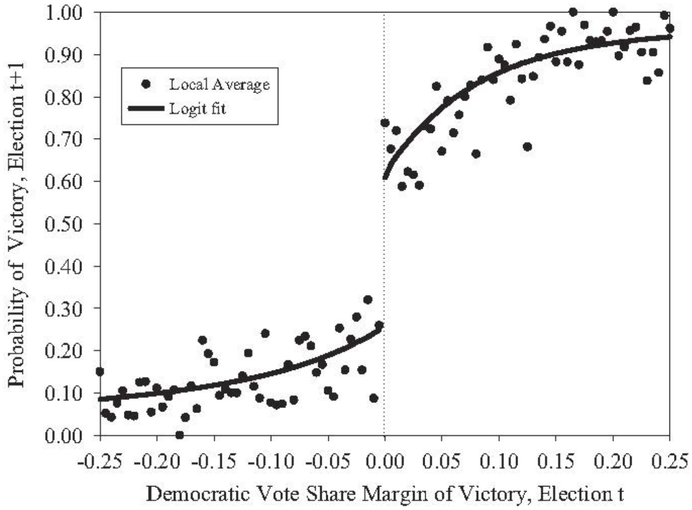

Figure 1: Probability of the Democratic Party winning election t+1 against its winning margin in election t

vers at in-state public colleges in Massachusetts of the US, with the goal of attracting talented students to the state's public colleges. The treatment is qualifying for the Adams Scholarship. It is determined by whether a student's test score from the Massachusetts Comprehensive Assessment System (MCAS) exceeds a certain threshold, so the running variable is the MCAS test score. Figure 2 below shows the probability of choosing a four-year public college against the number of grade points to the eligibility threshold.

As is clear from Figure 2, the probability of choosing a four-year public college jumps at the eligibility threshold, but then declines quickly once further above the threshold. The dramatic downward slope change at the threshold suggests that a student's response to an Adams Scholarship likely depends on her test score, which if true would invalidate the HTV independence assumption. Indeed, Dong and Lewbel (2014) provides nonparametric estimates of the derivative of the RD treatment effect with respect to the running variable at the cutoff (corresponds to the slope change at the cutoff in a sharp design like this) and show that they are negative and strongly significant. The estimates are also insensitive to covariates. Similarly using a differences in differences (DID) analysis, Goodman (2008) shows that qualified students with test scores near the eligibility threshold react much more strongly to the price change than students with test scores further above the threshold. This is likely because students trade college quality with prices. Better qualified students may be admitted to private colleges of much higher quality, and hence face a large quality drop if they instead accept the Adams Scholarship and attend a Massachusetts public college. In contrast, for marginal winners (those with

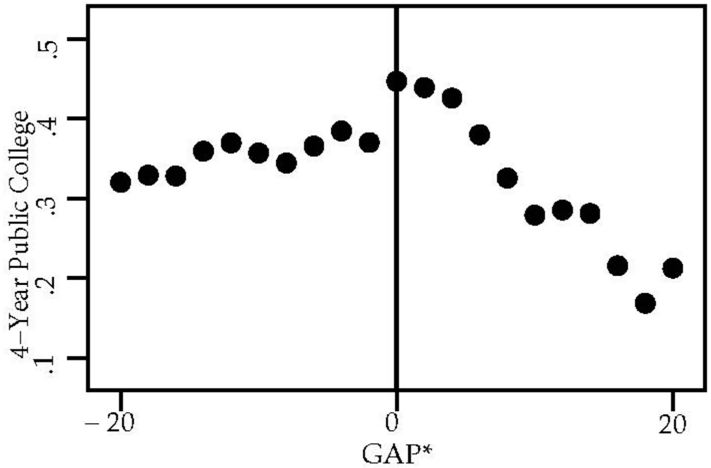

Figure 2: Probability of choosing a 4-year public college against the grade points from the eligibility threshold

test scores right above the threshold) the quality difference is smaller or non-existent, making the choice of a public college with a scholarship relatively more worthwhile given its lowered price (See Goodman 2008 for more discussion).

A third example is the RD model used to evaluate the impact of remedial education on students' outcomes (see, e.g., Jacob and Lefgren 2004 and Matsudaira 2008). The treatment is receiving remedial education, such as attending summer school, if a student's test score falls below some threshold failing grade, and the outcome is later academic performance. The HTV independence assumption requires that the effectiveness of remedial education for marginal students does not depend on one's pre-treatment test score. In contrast, the smoothness assumption imposed in this paper only requires that no students have precise manipulation of their test scores and no other changes at the cutoff have an impact on students' later academic performance.

3 Identification and Discussion

I focus on fuzzy design, treating sharp design as a special case. The discussion uses the same notation as in HTV (2001). Let y_{1i} and y_{0i} be the potential outcomes when an individual i is treated or not treated, respectively (Neyman 1923, Fisher 1935, Rubin 1974, 1990). Let x_i be a binary treatment indicator, so $x_i = 1$ if treated and 0 otherwise. The observed outcome can then be written as $y_i = \alpha_i + \beta_i x_i$, where $\alpha_i \equiv y_{0i}$, and $\beta_i \equiv y_{1i} - y_{0i}$. For a given covariate z_i, define the potential treatment status as $x_i(z)$ for a given value z that z_i could take on. When z_i is an instrument, one of the key assumptions for identifying a LATE in Imbens and Angrist (1994) is that the triplet $(y_{0i}, y_{1i}, x_i(z))$ is jointly independent of z_i (See Condition 1 of their Theorem 1). This independence assumption subsumes random assignment of the instrument z_i and an exclusion restriction asserting that z_i affects the outcomes only through its effect on the treatment x_i (see discussion in Angrist, Imbens and Rubin, 1996).

In the RD framework, z_i is the running variable, where z_0 is the RD cutoff. In discussing the fuzzy design RD with a variable treatment effect, HTV analogously assume that $(\beta_i, x_i(z))$ is jointly independent of z_i in a neighborhood of z_0 (See their conditions in Theorem 2). The other assumptions required to identify the RD LATE include continuity of the conditional mean of y_{0i}, a discontinuity in the treatment probability at z_0 and monotonicity.

In the following, I show that continuity of the density of the running variable for every 'individual' (defined later) is a fundamental identifying assumption for RD models. In sharp design RD, this assumption leads to continuity of the conditional means of potential outcomes; in fuzzy design RD, this assumption analogously leads to continuity of the conditional means of potential outcomes for different types of individuals (i.e., always takers, never takers, compliers and defiers) as well as continuity of probabilities of different types.

The resulted smoothness, along with two other standard LATE assumptions, monotonicity (no defiers) and a discontinuity in the treatment probability (a positive fraction of compliers), can identify the RD LATE even when the treatment effect is arbitrarily heterogenous and is correlated with treatment, which allows for self-selection into treatment based on idiosyncratic gains and selection into different types. Importantly, there can be endogenous selection into compliers, as long as the probability of being a complier is smooth at the cutoff.

For any $z \in (z_0 - \varepsilon, z_0 + \varepsilon)$ that z_i can take on and some small $\varepsilon > 0$, define an individual's potential treatment below the cutoff as $x_{0i}(z) \equiv x_i(z)$ if the observed $z < z_0$, so the potential treatment below the cutoff is the same as the observed treatment status in this case, and $x_{0i}(z) \equiv \lim_{\varepsilon \to 0} x_i(z_0 - \varepsilon)$ if the observed $z \geq z_0$ and the limit exists. Similarly, define an individual's potential treatment above the cutoff as $x_{1i}(z) \equiv x_i(z)$ if the observed $z \geqslant z_0$ and $x_{1i}(z) \equiv \lim_{\varepsilon \to 0} x_i(z_0 + \varepsilon)$ if the observed $z < z_0$ and the limit exists.[6]

Given the above definitions, for an individual with $z_i = z < z_0$, her treatment status below the

[6]Note that defining these unobserved counterfactuals this way is without loss of generality, which just simplifies the derivation, since what matters are only those at the limit when z goes to z_0.

cutoff is observed and is given by $x_{0i}(z) = x_i(z)$, and her counterfactual treatment just above the cutoff is $x_{1i}(z) \equiv \lim_{\varepsilon \to 0} x_i(z_0 + \varepsilon)$. Similarly for an individual with $z \geq z_0$, her treatment status above the cutoff is observed and is given by $x_{1i}(z) = x_i(z)$, and her counterfactual treatment just below the cutoff is $x_{0i}(z) \equiv \lim_{\varepsilon \to 0} x_i(z_0 - \varepsilon)$.

Given $z_i = z$, one can then define the following four types of individuals as events in a common probability space $(\Omega, \mathcal{F}, P)$: (Angrist, Imbens and Rubin, 1996):

Always taker: $A = \{x_{1i}(z) = x_{0i}(z) = 1\}$

Never taker: $N = \{x_{1i}(z) = x_{0i}(z) = 0\}$

Complier: $C = \{x_{1i}(z) = 0, x_{0i}(z) = 1\}$

Defier: $D = \{x_{1i}(z) = 1, x_{0i}(z) = 0\}$

Individual types are therefore allowed to depend on the value of the running variable. The standard RD models identify a LATE for compliers at $z = z_0$, i.e., individuals having $x_{1i}(z_0) = 0$ and $x_{0i}(z_0) = 1$. For notational convenience, I will suppress the argument to simply use x_{0i} and x_{1i} whenever there is no confusion. Let $d_i = 1\{z_i \geq z_0\}$, where $1\{\cdot\}$ is an indicator function equal to 1 if the expression in the bracket is true, and 0 otherwise. The observed treatment can then be written as $x_i = x_{0i} + d_i(x_{1i} - x_{0i})$.

ASSUMPTION A1a (Smoothness): Define the random vector $\mathbf{w}_i \equiv (y_{0i}, y_{1i}, x_{0i}, x_{1i})$. The conditional density of the running variable z_i, $f_{z|\mathbf{w}}(z \mid \mathbf{w})$ is continuous in a neighborhood of $z = z_0$ for all $\mathbf{w} \in supp(\mathbf{w}_i)$, and the density of the running variable $f_z(z)$ is continuous and strictly positive in a neighborhood of $z = z_0$.

An individual i can be seen as defined by the vector $\mathbf{w}_i$. Given her draw of the running variable, $\mathbf{w}_i$ would completely determine her treatment status x_i and outcome y_i. A1a is a statement asserting that for each individual defined by $\mathbf{w}_i$ the density of the running variable is continuous.

Given A1a, any particular individual's probability of being just above z_0 is bounded away from 0 and 1, implying that they do not have precise control over the running variable. Note that continuity of $f_{z|\mathbf{w}}(z \mid \mathbf{w})$ also rules out other discrete changes at the cutoff that would affect potential outcomes, e.g., there shouldn't be other policies or programs using the same cutoff.

Note that $\mathbf{w}_i \equiv (y_{0i}, y_{1i}, x_{i0}, x_{i1})$ puts no restrictions on treatment effect heterogeneity and on selection into different types of individuals. So, e.g., individuals can still self-select to be compliers, as long as the probability of being a complier is smooth at the RD cutoff, which in turn is guaranteed by Assumption A1a. Assumption A1a is similar to Condition 2b in Lee (2008), except that $\mathbf{w}_i$ in Lee (2008) is a one-dimensional random variable representing an individual's 'identity,' and that the discussion in Lee (2008) focuses on sharp design RD.

The following lemma shows that given A1a, the conditional means of potential outcomes y_{0i} and y_{1i} for each type of individuals and the probabilities of selection into different types are continuous at the cutoff z_0, which is expressed in Assumption A1b below.

ASSUMPTION A1b (Smoothness): $E\left[y_{0i} \mid x_{0i}=t_0, x_{1i}=t_1, z_i=z\right]$, $E\left[y_{1i} \mid x_{0i}=t_0, x_{1i}=t_1, z_i=z\right]$ and $Pr\left[x_{0i}=t_0, x_{1i}=t_1 \mid z_i=z\right]$, for $t_0=0,1$ and $t_1=0,1$ are continuous in z at $z=z_0$.

A1b nests the sharp design assumption by HTV as a special case. For sharp design, everyone is a complier, so $E\left[y_{0i} \mid x_{0i}=t_0, x_{1i}=t_1, z_i=z\right]=E\left[y_{0i} \mid z_i=z\right]$ and $E\left[y_{1i} \mid x_{0i}=t_0, x_{1i}=t_1, z_i=z\right]=E\left[y_{1i} \mid z_i=z\right]$. Therefore, for sharp design, A1b reduces to the assumption that $E\left[y_{0i} \mid z_i=z\right]$ and $E\left[y_{1i} \mid z_i=z\right]$ are continuous at $z=z_0$.

Given the smoothness in A1b, $E\left[x_i \mid z_i=z\right]=E\left[x_{0i} \mid z_i=z\right]$ for $z<z_0$ and $E\left[x_i \mid z_i=z\right]=E\left[x_{1i} \mid z_i=z\right]$ for $z \geqslant z_0$ are continuous at $z=z_0$. Note that at least one-sided continuity of the treatment probability below or above the cutoff is implicitly assumed in the first-stage treatment equation, which is typically estimated by local linear or parametric polynomial regressions.

LEMMA: If A1a holds, then A1b holds.

Proofs are in the Appendix. Define $x^+ \equiv \lim_{\varepsilon\to 0} E\left[x_i \mid z_i=z+\varepsilon\right]$, $x^- \equiv \lim_{\varepsilon\to 0} E\left[x_i \mid z_i=z-\varepsilon\right]$, $y^+ \equiv \lim_{\varepsilon\to 0} E\left[y_i \mid z_i=z+\varepsilon\right]$ and $y^- \equiv \lim_{\varepsilon\to 0} E\left[y_i \mid z_i=z-\varepsilon\right]$. A1b guarantees that these limits exist, as shown in the proof of the theorem below.

ASSUMPTION A2 (RD): $x^+ \neq x^-$.

ASSUMPTION A3 (Monotonicity): There exists $\varepsilon>0$ such that $x_{1i}(z) \geq x_{0i}(z)$ for all $z \in (z_0-\varepsilon, z_0+\varepsilon)$.

A2 assumes a discontinuity in the treatment probability, which means existence of compliers and A3 rules out defiers. Both are also assumed by HTV, following from Imbens and Angrist (1994). However, I use A1b, which is guaranteed by A1a, to replace the HTV independence or alternative assumptions. So the above assume neither that x_i is independent of β_i, nor that $(\beta_i, x_i(z))$ is jointly independent of z_i for z_i near z_0.[7]

THEOREM: Given assumptions A1b, A2 and A3, the local average treatment effect for compliers at $z_i = z_0$ is identified and is given by $E\left[y_{1i} - y_{0i} \mid z_i = z_0, C\right] = \frac{y^+ - y^-}{x^+ - x^-}$.

This theorem shows that assumptions A1b, A2 and A3 suffice to obtain the standard RD identification result established in HTV (2001).

Compared with A1b, A1a is stronger than necessary.[8] In fact, given A1a, the joint distribution of $\mathbf{w}_i \equiv (y_{0i}, y_{1i}, x_{i0}, x_{i1})$ is continuous at the RD cutoff, so one may identify any distributional effects in addition to mean effects (see, e.g., Frandsen, Frolich and Melly, 2012). A1a is more appealing considering its plausible behavioral interpretation and testable implications. In particular, A1a provides a formal support for using the density test to assess the validity of fuzzy design RD (see Lee 2008 for discussion in sharp design and related discussion in McCrary 2008).

Intuitively, given no precise manipulation listed in A1a, d_i is a valid instrument. Then by Theorem 1 of Imbens and Angrist (1994), the ratio of the two intention-to-treat causal estimands $\frac{y^+ - y^-}{x^+ - x^-}$ identifies a LATE for compliers who change treatment status when the instrument d_i changes value from 0 to 1.

By relaxing the independence assumption in HTV (2001), these alternative assumptions lead to the opportunity of identifying the derivative of the RD treatment effect with respect to the running variable. Dong and Lewbel (2014) provides nonparametric identification and estimation of this derivative. They show that this derivative can be used to investigate external validity of RD LATE, to extrapolate the RD LATE away from the RD cutoff, and to identify how the RD LATE would change when the threshold is marginally changed.

Further, Dong (2014) assumes continuous differentiability of the conditional density in A1a, a stronger version of no-manipulation, and shows that when there is no discontinuity in the treatment probability, one may still identify a causal effect of the treatment using a slope change (a kink) in the

[7] Given A1a, the joint distribution for $(y_{0i}, y_{1i}, x_{i0}, x_{i1})$ given $z_i = z$ is continuous at $z = z_0$. The independence assumption imposed by HTV may be seen to hold in the limit as z approaches to z_0, which is what is required for identification.

[8] For example, one may randomize separately right above and right below the cutoff (with different sample sizes) so that A1b holds, but the density of the running variable is not necessarily continuous.

treatment probability. The identified effect is a limit form of the RD LATE, or a marginal treatment effect (MTE), as proposed by Heckman and Vytlacil (2005, 2007).

There exist several RD studies that do not impose independence. Except for the previously mentioned Frandsen, Frolich and Melly (2012), Battistin and Rettore (2008) relax the HTV independence assumption by looking at the one-sided fuzzy RD design without always takers. They show that in this case, as in a fuzzy design RD, continuity of the conditional mean of y_{0i} is sufficient for identification. Neither of them discuss more generally weaker assumptions for RD identification.

4 Testing Independence

In this section, I discuss testing the implication of the independence assumption by HTV, utilizing the results of Dong and Lewbel (2014). The test is based on minimal further smoothness (assuming continuous differentiability instead of just continuity).[9]

Dong and Lewbel (2014) shows that given continuous differentiability instead of just continuity of the conditional means and probabilities in A1b, for both sharp and fuzzy RD designs, one can nonparametrically identify and estimate the derivative of the RD treatment effect with respect to the running variable at the cutoff, i.e., $\partial E(y_{1i} - y_{0i} \mid z_i = z, C)/\partial z \mid_{z=z_0}$. It is referred to as the treatment effect derivative or TED. TED measures how the treatment effect depends on the running variable near the cutoff z_0. In sharp design, the treatment $x_i = d_i$, so TED corresponds to the slope change at the RD cutoff.

The independence assumption by HTV requires that the treatment effect does not to depend on the running variable near the cutoff and hence requires TED to be zero. One can then test partially his independence assumption by testing whether the estimated TED is significant or not. In case of a significant TED, the independence assumption by HTV can be rejected; however, the identified LATE is still valid as long as the required smoothness holds.

Standard RD validity tests can be used to test the alternative continuity assumption imposed in this paper, which I do not discuss in detail. One can similarly test the further smoothness required for identifying TED by testing additionally continuity of the slope of the density of the running variable

[9]Virtually all empirical implementations of RD models satisfy this slightly stronger assumption. In particular, parametric models generally assume polynomials or other differentiable functions, while most nonparametric estimators, including local linear regressions, assume (for establishing asymptotic theory) at least continuous differentiability.

and the slopes of the conditional means of covariates at the RD cutoff.

5 Empirical Application

Following Lee (2008) and Lee and Lemieux (2010), this section estimates the RD model of incumbency advantage in the US house election. I show that in this case the smoothness assumption plausibly holds, but the independence assumption does not, given smoothness.

Recall that the treatment in this case is an indicator of the Democratic Party being the incumbent party, the running variable is the Democratic Party's vote share margin in election t, and the outcome is whether a democrat won in election $t + 1$.

I use the same data used in Lee (2008) and estimate similar local linear and local polynomial regressions. Following Lee and Lemieux (2010) I adopt a uniform kernel but explore the sensitivity of the estimates to greatly varying bandwidths. The sample consists of 6,558 elections over the 1946 - 98 period (see Lee 2008 for more detail).[10] I report estimates of the treatment effect as well as TED in the top panel of Table 1. TED in this case measures how the incumbency advantage depends on the incumbent party's winning margin, corresponding to the slope change at the cutoff in Figure 1.

The last row of Table 1 presents estimates from a local linear regression based on the optimal bandwidth (0.172) used in Lee and Lemieux (2010), which was chosen by a cross-validation procedure described in Ludwig and Miller (2007). Slightly smaller (0.15) and larger (0.20) bandwidths are also used in additional local linear regressions. Further, various low-order polynomials are estimated for equally spaced bandwidths ranging from 0.25 to 1, where bandwidth equal to 1 corresponds to the full sample. The order of polynomial for each bandwidth is chosen using the Akaike information criterion (AIC). Robust standard errors are calculated as proposed by Imbens and Lemieux (2008).

The estimated incumbency effects and their derivatives are robust to different bandwidths and orders of polynomials. Consistent with estimates in Lee (2008) and Lee and Lemieux (2010), the average incumbency effect is estimated to be 0.36-0.42 across the 10 specifications, meaning that when the Democratic Party is the incumbent party, it increases their probability of winning the next election by 36% - 42%. The estimated TED based on the optimal bandwidth is 1.134 and based on the full range of data is 1.293. Both are statistically significant. Estimates based on other bandwidths

[10] The difference between Lee's (2008) sample and that in Caughey and Sekhon (2011) is discussed in Lee and Lemieux (2013).

Table 1 RD estimates of the treatment effect and treatment effect derivative (TED)

Bandwidth	Treatment effect		TED		Optimal order of polynomial	Observations
	Dependent variable: Winning in election t+1					
1.00	0.385	(0.039)***	1.293	(0.514)**	4	6,558
0.50	0.370	(0.043)***	1.542	(0.678)**	3	4,900
0.45	0.363	(0.046)***	1.574	(0.801)**	3	4,560
0.40	0.407	(0.036)***	1.186	(0.368)***	2	4,169
0.35	0.393	(0.038)***	1.150	(0.451)**	2	3,772
0.30	0.381	(0.042)***	1.446	(0.577)**	2	3,283
0.25	0.375	(0.047)***	1.664	(0.791)**	2	2,763
0.20	0.423	(0.033)***	0.988	(0.263)***	1	2,265
0.15	0.409	(0.040)***	0.998	(0.424)**	1	1,765
[0.172]	0.417	(0.037)***	1.134	(0.339)***	1	1,993
	Dependent variable: Density of winning margin in election t					
0.50	0.125	(0.131)	-1.290	(1.064)	2	4,900
	Dependent variable: Vote share in election t-1					
0.50	-0.010	(0.014)	0.177	(0.252)	3	4,900

Note: The optimal bandwidth for the local linear regression (the last row of the top panel) is obtained using the cross-validation procedure described in Lee and Lemieux (2010). The optimal order of the polynomial for each bandwidth is chosen based on the Akaike's information criterion. The density of the running variable is calculated using 200 bins over the range -0.5 to 0.5. Robust standard errors are in parentheses; *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level.

are of similar magnitude, so given a 1 percentage point increase in the Democrats' winning margin, the probability for them to win the next election roughly increases by 1%.

The estimated derivatives are all significant, suggesting that the incumbency effect significantly depends on the incumbent party's winning margin, so the HTV independence assumption can be rejected practically. However, the bottom two panels of Table 1 report the estimated discontinuities or slope changes in the density of the running variable and the conditional mean of an important covariate, the Democratic vote share from the previous election. None of these estimates are statistically significant, so the smoothness assumption plausibly holds in this case. Therefore, the RD design is still valid.

6 Conclusions

This paper shows that given a discontinuity in the treatment probability and monotonicity, identification of the RD LATE in both sharp and fuzzy designs can be established under just smooth conditions, instead of the independence assumption by Hahn, Todd and Van der Klaauw (2001). This paper also provides an empirically plausible and partially testable weak behavioral assumption that leads to the required smoothness to hold.

Relaxing the HTV independence assumption is necessary to identify the derivative of the RD treatment effect with respect to the running variable. Such a derivative is useful in evaluating treatment heterogeneity and hence external validity of the RD LATE in the neighborhood of the RD threshold, in extrapolating the RD LATE away from the cutoff, and in evaluating how the RD LATE would change if the threshold is marginally changed. This alternative framework also allows to generalize standard RD identification to incorporate identification based on kinks (or slope changes) as in Dong (2014).

Results in this paper provide a formal support for the popular practice of performing McCrary's (2008) density test to assess the validity of fuzzy design RD. A simple test is proposed to evaluate the independence assumption by HTV. I apply these results to the data of Lee (2008) and show that the smoothes assumption required plausibly holds while the independence assumption by HTV does not, given smoothness.

7 Appendix

Proof of Lemma: Let $f_{.}$ and $f_{.|.}$ denote the unconditional and conditional probability density or mass functions, respectively. For simplicity, the following analysis assumes that y_{0i} and y_{1i} are continuous, though essentially the argument holds when y_{0i} and y_{1i} are discrete. So for example, $f_{\mathbf{w}|z}(\mathbf{w} \mid z)$ denotes the mixed joint density of $\mathbf{w}_i$ conditional on $z_i = z$, i.e., $f_{\mathbf{w}|z}(\mathbf{w} \mid z) = f_{y_0,y_1,z|x_0,x_1}(y_0, y_1, z \mid x_{0i} = t_0, x_{1i} = t_1) \Pr(x_{0i} = t_0, x_{1i} = t_1)/f_z(z)$. All the analysis applies to $z_i \in (z_0 - \varepsilon, z_0 + \varepsilon)$ for some small $\varepsilon > 0$.

Assumption A1a states that $f_{z|\mathbf{w}}(z \mid \mathbf{w})$ is continuous in z, and $f_z(z)$ is continuous and strictly positive at $z = z_0$. By Bayes' Rule, $f_{\mathbf{w}|z}(\mathbf{w} \mid z) = f_{z|\mathbf{w}}(z \mid \mathbf{w}) f_{\mathbf{w}}(\mathbf{w})/f_z(z)$, so $f_{\mathbf{w}|z}(\mathbf{w} \mid z)$ is continuous in z at $z = z_0$. By definition $\mathbf{w}_i \equiv (y_{0i}, y_{1i}, x_{i0}, x_{i1})$, then probability of each type of individual $\Pr(x_{0i} = t_0, x_{1i} = t_1 \mid z_i = z) = \int_{\Omega_1}\int_{\Omega_0} f_{\mathbf{w}|z}(\mathbf{w} \mid z)dy_0dy_1$ for $t_0 = 0, 1$ and $t_1 = 0, 1$ is continuous in z at $z = z_0$, where $\mathbf{\Omega}_t$ is the conditional support of y_{ti} for $t = 0, 1$ conditional on $z_i = z$.

Again by Bayes' Rule, $f_{y_0,y_1|x_0,x_1,z}(y_0, y_1 \mid x_{0i} = t_0, x_{1i} = t_1, z_i = z) = f_{\mathbf{w}|z}(\mathbf{w} \mid z)/Pr(x_{0i} = t_0, x_{1i} = t_1 \mid z_i = z)$ for $t_0 = 0, 1$ and $t_1 = 0, 1$. Both $f_{\mathbf{w}|z}(\mathbf{w} \mid z)$ and $Pr(x_{0i} = t_0, x_{1i} = t_1 \mid z_i = z)$ are continuous in z at $z = z_0$, so $f_{y_0,y_1|x_0,x_1,z}(y_0, y_1 \mid x_{0i} = t_0, x_{1i} = t_1, z_i = z)$ for $t_0 = 0, 1$ and $t_1 = 0, 1$ is continuous in z at $z = z_0$. It follows that type-specific conditional means of potential outcome $E(y_{ti} \mid x_{0i} = t_0, x_{1i} = t_1, z_i = z)$ for $t = 0, 1$, $t_0 = 0, 1$ and $t_1 = 0, 1$ are continuous in z at $z = z_0$.

Proof of Theorem: Given assumptions A1b and A3 as well as the definitions of individual types, we have

$$\begin{aligned}
y^{+} &\equiv \lim_{e\to 0} E\left[y_i \mid z_i = z_0+\varepsilon\right] = \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i x_i \mid z_i = z_0+\varepsilon\right] \\
&= \lim_{\varepsilon\to 0} E\left[\alpha_i \mid z_i = z_0+\varepsilon, x_i = 0\right]\Pr\left[x_i = 0 \mid z_i = z_0+\varepsilon\right] \\
&+ \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i \mid z_i = z_0+\varepsilon, x_i = 1\right]\Pr\left[x_i = 1 \mid z_i = z_0+\varepsilon\right] \\
&= \lim_{\varepsilon\to 0} E\left[\alpha_i \mid z_i = z_0+\varepsilon, x_{1i} = 0\right]\Pr\left[x_{1i} = 0 \mid z_i = z_0+\varepsilon\right] \\
&+ \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i \mid z_i = z_0+\varepsilon, x_{1i} = 1\right]\Pr\left[x_{1i} = 1 \mid z_i = z_0+\varepsilon\right] \\
&= \lim_{\varepsilon\to 0} E\left[\alpha_i \mid z_i = z_0 + \varepsilon, x_{1i} = 0, x_{0i} = 0\right]\Pr\left[x_{1i} = 0, x_{10i} = 0 \mid z_i = z_0+\varepsilon\right] \\
&+ \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i \mid z_i = z_0 + \varepsilon, x_{1i} = 1, x_{0i} = 0\right]\Pr\left[x_{1i} = 1, x_{0i} = 0 \mid z_i = z_0+\varepsilon\right] \\
&+ \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i \mid z_i = z_0 + \varepsilon, x_{1i} = 1, x_{0i} = 1\right]\Pr\left[x_{1i} = 1, x_{0i} = 1 \mid z_i = z_0+\varepsilon\right] \\
&= \lim_{\varepsilon\to 0} E\left[\alpha_i \mid z_i = z_0 + \varepsilon, N\right]\Pr\left[N \mid z_i = z_0+\varepsilon\right] \\
&+ \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i \mid z_i = z_0 + \varepsilon, C\right]\Pr\left[C \mid z_i = z_0+\varepsilon\right] \\
&+ \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i \mid z_i = z_0 + \varepsilon, A\right]\Pr\left[A \mid z_i = z_0+\varepsilon\right] \\
&= E\left[\alpha_i \mid z_i = z_0, N\right]\Pr\left[N \mid z_i = z_0\right] \\
&+ E\left[\alpha_i + \beta_i \mid z_i = z_0, C\right]\Pr\left[C \mid z_i = z_0\right] \\
&+ E\left[\alpha_i + \beta_i \mid z_i = z_0, A\right]\Pr\left[A \mid z_i = z_0\right],
\end{aligned}$$

where the fourth equality follows from definition of potential treatment status x_{0i} and x_{1i}, which are implicit functions of the running variable, the fifth equality follows from monotonicity, the six equality follows from definitions of types, and the last equality follows from continuity of conditional means of potential outcomes for each type of individuals and continuity of probabilities of individual types at $z = z_0$.

Similarly we have

$$
\begin{aligned}
& y^- \equiv \lim_{e\to 0} E\left[y_i \mid z_i = z_0 - \varepsilon\right] = \lim_{\varepsilon\to 0} E\left[\alpha_i + \beta_i x_i \mid z_i = z_0 - \varepsilon\right] \\
& = E\left[\alpha_i \mid z_i = z_0, N\right] \Pr\left[N \mid z_i = z_0\right] \\
& + E\left[\alpha_i \mid z_i = z_0, C\right] \Pr\left[C \mid z_i = z_0\right] \\
& + E\left[\alpha_i + \beta_i \mid z_i = z_0, A\right] \Pr\left[A \mid z_i = z_0\right].
\end{aligned}
$$

Therefore,

$$
y^+ - y^- = E\left[\beta_i \mid z_i = z_0, C\right] \Pr\left[C \mid z_i = z_0\right].
$$

In addition,

$$
\begin{aligned}
x^+ - x^- &\equiv \lim_{\varepsilon\to 0} E\left[x_i \mid z_i = z_0 + \varepsilon\right] - \lim_{\varepsilon\to 0} E\left[x_i \mid z_i = z_0 - \varepsilon\right] \\
&= \lim_{\varepsilon\to 0} E\left[x_{1i} \mid z_i = z_0 + \varepsilon\right] - \lim_{\varepsilon\to 0} E\left[x_{0i} \mid z_i = z_0 - \varepsilon\right] \\
&= \lim_{\varepsilon\to 0} \Pr\left[x_{1i} = 1, x_{0i} = 0 \mid z_i = z_0 + \varepsilon\right] + \lim_{\varepsilon\to 0} \Pr\left[x_{1i} = 1, x_{0i} = 1 \mid z_i = z_0 + \varepsilon\right] \\
&\quad - \lim_{\varepsilon\to 0} \Pr\left[x_{1i} = 1, x_{0i} = 1 \mid z_i = z_0 - \varepsilon\right] \\
&= \Pr\left[C \mid z_i = z_0\right] + \Pr\left[A \mid z_i = z_0\right] - \Pr\left[A \mid z_i = z_0 - \varepsilon\right] \\
&= \Pr\left[C \mid z_i = z_0\right],
\end{aligned}
$$

where the third equality follows from monotonicity, the fourth equality follows from definitions of types and continuity of probability of each type at $z_i = z_0$.

By A2, $x^+ - x^- \neq 0$, so putting the above equations together gives

$$
E\left[y_{1i} - y_{0i} \mid z_i = z_0, C\right] = \frac{y^+ - y^-}{x^+ - x^-}.
$$

References

[1] Angrist, J., G. Imbens and D. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," Journal of the American Statistical Association, 91(434), 444-455.

[2] Angrist, J. and J. S. Pischke (2008) Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press.

[3] Battistin E. and E. Rettore (2008): "Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs," Journal of Econometircs, 142, 715-730.

[4] Caughey, D. and J. S. Sekhon (2011): “Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942-2008,” Political Analysis, 19, 385-408.

[5] Dinardo, J. and D. S. Lee (2011), "Program Evaluation and Research Designs," in Handbook of Labor Economics, Ashenfelter and Card (eds.), vol. 4a, Chap. 5, 463-536.

[6] Dong, Y. (2014): "Jump or Kink? Identification of the Binary Treatment Regression Discontinuity Design without the Discontinuity," Unpublished manuscript.

[7] Dong Y. and A. Lewbel (2014): "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models," Boston College working paper.

[8] Fisher, R. (1935): "Design of Experiments," Oliver and Boyd, Edinburgh.

[9] Frandsen B., M. Frolich, and B. Melly (2012): "Quantile treatment effects in the regression discontinuity design," Journal of Econometrics, 168, 382-395.

[10] Goodman, J. (2008): "Who Merits Financial Aid?: Massachusetts’ Adams Scholarship," Journal of Public Economics 92(10) 2121-2131.

[11] Hahn, J., P. Todd, W. van der Klaauw (2001): "Identification and estimation of treatment effects with a regression-discontinuity design," Econometrica 69(1), 201-209.

[12] Heckman, J. J. and E. Vytlacil (2005),“Structural Equations, Treatment Effects, and Econometric Policy Evaluation," Econometrica, 73(3), 669-738.

[13] Heckman, J. J. and E. J. Vytlacil (2007), “Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation," Handbook of Econometrics 6, in: J.J. Heckman & E.E. Leamer (ed.), 4779-4874.

[14] Imbens, G. and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," Econometrica, 62(2), 467-475.

[15] Imbens, G. and T. Lemieux (2008): "Regression discontinuity designs: A guide to practice," Journal of Econometrics, 142, 615-635.

[16] Jacob, B. A., and L. Lefgren, (2004) “Remedial Education and Student Achievement: A Regression-Discontinuity Analysis,” Review of Economics and Statistics, 86, 226-244.

[17] Lee, D.S. (2008): "Randomized Experiments from Non-random Selection in U.S. House Elections," Journal of Econometrics, 142(2), 675-97.

[18] Lee, D.S., T. Lemieux (2010): "Regression Discontinuity Designs in Economics," Journal of Economic Literature, 48, 281-355.

[19] Lee, D.S., T. Lemieux (2014): "Regression Discontinuity Designs in Social Sciences," in Regression Analysis and Causal Inference, H. Best and C. Wolf (eds.), Sage.

[20] Ludwig J., and D. L. Miller (2007), "Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design," The Quarterly Journal of Economics, 122, 159-208.

[21] Matsudaira J.D. (2008): "Mandatory summer school and student achievement," Journal of Econometrics, 142, 829-850.

[22] McCrary, J. (2008): “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” Journal of Econometrics, 142(2), 698-714.

[23] Neyman, J. 1923: “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9,” Translated in Statistical Science, 5(4), 465-480, 1990.

[24] Rubin, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," Journal of Educational Psychology, 66, 688-701.

[25] Rubin, D. (1990): "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," Statistical Science, 5, 472-480.