An Alternative Assumption to Identify LATE in Regression Discontinuity Design

Yingying Dong
University of California Irvine*

May 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) methods. This paper shows that this local independence assumption may not hold in empirical applications, and that this assumption can be replaced with an empirically plausible and partially testable weak behavioral assumption. This paper extends the results of Lee (2008) to fuzzy RD designs, showing that, given no defiers and a discontinuity in the treatment probability, the LATE is identified in fuzzy or sharp RD designs if no individuals (always takers, never takers, or compliers) can precisely manipulate the running variable. Empirical examples are provided.

JEL Codes: C21, C25

Keywords: Regression discontinuity, Sharp design, Fuzzy design, LATE, Treatment effect heterogeneity, Treatment effect derivative, Manipulation, Sorting, Density test

1 Introduction

Regression discontinuity (RD) design has 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 models. A key assumption used by HTV is that the treatment effect and potential treatment status are jointly

*I am grateful to Josh Angrist, Arthur Lewbel and Yaming Yu for their insightful comments and 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.\textsuperscript{1} This assumption is based on the independence assumption in the original LATE paper by Imbens and Angrist (1994).

In this paper I show that the HTV type independence assumption may not hold in empirical applications of RD models and that this assumption can be replaced with an empirically plausible and partially testable weak behavioral assumption. Intuitively, independence may be violated because 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 place restrictions on the form of treatment effect heterogeneity.\textsuperscript{2}

This paper lists formal smoothness assumptions that suffice to identify both sharp and fuzzy design RD without the HTV independence assumption. This paper also provides a weak behavioral assumption, similar to those in Lee (2008), that is sufficient to make the required smoothness conditions hold both sharp and fuzzy design RD.\textsuperscript{3} 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 a minimal further smoothness assumption, HTV’s independence assumption is shown to be testable. I apply these results to an empirical application showing that the identifying assumption in this paper plausibly holds, but the HTV independence assumption likely does not.

In the context of sharp design RD, 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. Lee further shows that this assumption generates strong testable implications, i.e., individuals cannot precisely manipulate the running variable to sort themselves to just above or just below the cutoff so that covariate means should be balanced on either side of the cutoff. Results in this paper are consistent with those in Lee (2008) and similarly relate smoothness to a lack of ma-

\textsuperscript{1}An alternative assumption HTV impose is that the treatment be independent of the treatment effect conditional on the running variable near the RD threshold. As HTV note, this assumption rules out self-selection into treatment based on idiosyncratic gains, and so is often not be realistic. Another proposed alternative is that the treatment effect is constant across individuals.

\textsuperscript{2}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.

\textsuperscript{3}Cattaneo, Frandsen and Titiunik (2013) propose doing inference in RD design as local randomized experiments under stronger assumptions than those imposed by HTV or here. In particular, they assume that there exists a neighborhood around the cutoff where randomization holds. Their assumptions require that, e.g., related conditional mean functions are constant as a function of the score, whereas here I only assume continuity.
nipulation 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.\(^4\) The discussion leads to precisely the same identification results as those by HTV, but under an alternative smoothness assumption.

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

\section{Motivating Examples}

To motivate the discussion in this paper, consider the standard RD model estimating the electoral advantage of incumbency in the US house of representatives in Lee (2008). The treatment is the Democratic Party being an incumbent party, which is determined by the party’s vote share exceeding its strongest opponent’s.\(^5\) The running variable is the Democrats’ winning margin, and the outcome is the probability of a Democrat winning the next election. In this case, the independence assumption by HTV would require that the incumbent’s electoral advantage not depend on its winning margin. Figure 1, which is reproduced from Figure 5-(a) in Lee (2008), provides clear evidence against this assumption.

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 in the next election, so the incumbency advantage depends on the winning margin.

Consider another RD model estimating the effect of the Adams Scholarship program on college

---

\(^4\)Cattaneo, Frandsen and Titiunik (2013) also consider alternative assumptions. However, their assumptions are stronger than those imposed by HTV or here, and are made to analyze RD design as local randomized experiments. In particular, they assume that there exists a neighborhood around the cutoff where randomization holds. Their assumptions require that, e.g., conditional mean outcomes are constant as a function of the score, whereas I only assume continuity of functions like these. See also discussion in DiNardo and Lee (2008).

\(^5\)Due to the largely two-party system, the strongest opponent is almost always the Republican Party, and the outcome for the Republican Party is therefore a mirror image of that for the Democratic Party (Lee 2008).
Figure 1: Probability of the Democratic Party winning election t+1 against its winning margin in election t

choices (Goodman, 2008). The Adams Scholarship program provides qualified students tuition waivers 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 $x$ is qualifying for an Adams Scholarship. A student qualifies for the scholarship if her test score from the Massachusetts Comprehensive Assessment System (MCAS) exceeds certain thresholds. So the running variable $z$ is the MCAS test score. Figure 2 below presents the probability of choosing a four-year public college as a function of 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 cutoff, but then declines quickly with test scores above the eligibility threshold. The dramatic downward slope change at the cutoff suggests that a student’s response to an Adams Scholarship likely depends on how far she is from the eligibility threshold, which if true would invalidate the HTV independence assumption. Indeed, Dong and Lewbel (2014) formally estimate the derivative of the RD treatment effect with respect to the running variable (which equals the slope change at the cutoff) and show in this case that it is negative and strongly significant. This indicates that, at least in the neighborhood of the threshold, the effect of an Adams Scholarship on college choices decreases with test scores. 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 farther above the threshold. This is likely because students trade off between college quality with price. 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 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 evaluating the impact of remedial education on students’ outcomes (see, e.g., Jacob and Lefgren 2004). The treatment is receiving remedial education if a student’s test score falls below some threshold failing grade, and the outcome is later academic performance. The independence assumption in HTV requires that the effectiveness of remedial education not depend on one’s pre-treatment test score near the threshold failing grade. 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

All the discussion in this section applies to $z_i$ in a neighborhood of the RD cutoff, i.e., $z_i \in [z_0 - \varepsilon, z_0 + \varepsilon]$ for some small $\varepsilon > 0$. The discussion focuses on fuzzy design, treating sharp design as a special case.

Using the same notation as in HTV and based on the counterfactual framework (Neyman 1923, Fisher 1935, Rubin 1974, 1990), let $y_{1i}$ and $y_{0i}$ be the potential outcomes when an individual $i$ is treated or not treated, respectively. 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 \( \alpha_i = y_{0i} \), and \( \beta_i = y_{1i} - y_{0i} \). 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). With the stable unit treatment value assumption (SUTVA), this independence assumption corresponds to an exclusion restriction asserting that the instrument \( z_i \) 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. 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), where \( z_0 \) is the RD cutoff. The other assumptions required to identify the RD LATE are continuity of the conditional mean of \( y_{0i} \), the presence of 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.

This smoothness assumption, along with with monotonicity (ruling out defiers) and existence of a discontinuity in the treatment probability (requiring a positive fraction of compliers) at the cutoff, can identify the RD LATE when the treatment effect varies across individuals and is correlated with the treatment, which allows for self-selection into treatment based on idiosyncratic gains, and hence into different types. For example, this paper’s assumptions allow for endogenous selection into compliers, as long as the probability of being a complier is smooth at the cutoff.

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 \( z \geq z_0 \) and \( x_{1i}(z) \equiv \lim_{\varepsilon \to 0} x_i(z_0 + \varepsilon) \) if the observed \( z < z_0 \) and the limit exists. So \( x_{0i}(z) \) and \( x_{1i}(z) \) are well

\[ \text{Note that defining these unobserved counterfactuals this way is without loss of generality, which just simplifies the} \]
defined for any $z \in [z_0 - \varepsilon, z_0 + \varepsilon]$.

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

For a value $z_i$ can take on $z$, we can then define the following four types of individuals in a common probability space (Angrist, Imbens and Rubin, 1996):

- **Always takers**: $A = \{i: x_{1i}(z) = x_{0i}(z) = 1\}$
- **Never takers**: $N = \{i: x_{1i}(z) = x_{0i}(z) = 0\}$
- **Compliers**: $C = \{i: x_{1i}(z) = 0, x_{0i}(z) = 1\}$
- **Defiers**: $D = \{i: x_{1i}(z) = 1, x_{0i}(z) = 0\}$

Individual types are therefore allowed to depend on $z$. 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 $x_i = x_{0i} + d_i(x_{1i} - x_{0i})$. Since $x_{0i}$ and $x_{1i}$ are functions of $z$, $x_i$ is also a function of $z$.

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

Assumption A1a is similar to Condition 2b in Lee (2008), except that $w_i$ 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. A1a is a statement asserting that for each individual the density of the running variable is continuous. To see this, one can think of an individual $i$ as being defined by the vector $w_i$, since $w_i$ consists of information on her potential treatment and potential outcomes, which would completely determine her observed outcome $y_i$ given her draw of the running variable.

The counterfactuals $y_{0i}, y_{1i}, x_{0i},$ and $x_{1i}$ can themselves be functions of observed and unobserved individual and environmental characteristics, so an ‘individual’ here is defined along with the related derivation, since what matters are only those at the limit when $z$ goes to $z_0$. 

---

7
policy environment. Given continuity and strict positiveness of the conditional density $f_{z|w}(z | w)$, the conditional cdf of $z_i$ is such that $0 < 1 - F_{z|w}(z_0 | w) < 1$, i.e., any particular individual’s probability of being just above the threshold $z_0$ is between 0 and 1, so they do not have precise control over the running variable. Note that continuity of $f_{z|w}(z | w)$ 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.

Importantly, $w_i = (y_{0i}, y_{1i}, x_{i0}, x_{i1})$ puts no restrictions on the heterogeneity of treatment effect and on selection into different types of individuals. 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):** $E(y_{0i} | x_{0i} = t_0, x_{1i} = t_1, z_i = z)$, $E(y_{1i} | x_{0i} = t_0, x_{1i} = t_1, z_i = z)$ and $Pr(x_{0i} = t_0, x_{1i} = t_1 | z_i = z)$, 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, in particular, for sharp design, everyone is a complier $Pr(x_{0i} = t_0, x_{1i} = t_1 | z_i = z) = 0$ for $t_0 = t_1$ i.e., for always takers and never takers, and $Pr(x_{0i} = t_0, x_{1i} = t_1 | z_i = z) = 1$ for $t_0 < t_1$, i.e., for compliers. $E(y_{0i} | x_{0i} = t_0, x_{1i} = t_1, z_i = z) = E(y_{0i} | z_i = z)$ and $E(y_{1i} | x_{0i} = t_0, x_{1i} = t_1, z_i = z) = E(y_{1i} | z_i = z)$. Therefore for sharp design RD A1b states that $E(y_{0i} | z_i = z)$ and $E(y_{1i} | z_i = z)$ are continuous at $z = z_0$, which is sufficient for identification of a causal effect in sharp design. For fuzzy design, analogously it requires that the conditional means of the potential outcomes are smooth for each type of individuals and that the probabilities of different ‘types’ are smooth.

**LEMMA:** If A1a holds, then A1b holds.

Proofs are in the Appendix. Given A1b, the conditional means of potential outcomes for each type and the probabilities of different types are all continuous at the cutoff. It follows that the mean outcome difference at the cutoff then equals the mean outcome change for compliers and defiers. Further assuming no defiers and a positive fraction of compliers leads to the standard RD identification result, i.e., the ratio of the mean outcome discontinuity to the treatment probability discontinuity identifies a LATE at the RD cutoff.
Let $x^+ \equiv \lim_{\epsilon \to 0} E x_i \mid z_i = z + \epsilon, x^- \equiv \lim_{\epsilon \to 0} E x_i \mid z_i = z - \epsilon, y^+ \equiv \lim_{\epsilon \to 0} E y_i \mid z_i = z + \epsilon$ and $y^- \equiv \lim_{\epsilon \to 0} E y_i \mid z_i = z - \epsilon$. A1b guarantees that these limits exist, as shown below in the proof of the theorem.

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

ASSUMPTION A3 (Monotonicity): There exists $\epsilon > 0$ such that $x_i(z_0 + \epsilon) \geq x_i(z_0 - \epsilon)$ for all $0 < \epsilon \leq \epsilon$.

A2 is the assumption that a discontinuity in the treatment probability exists. A3 assumes away defiers. Both are assumed by HTV, and similarly assumed in Imbens and Angrist (1994). The difference between the above assumptions and those by HTV is that the above does not assume independence – neither that $x_i$ is independent of $\beta_i$, nor that $(\beta_i, x_i(z))$ is jointly independent of $z_i$ for $z_i$ near $z_0$; instead, the above assumes smoothness of conditional means and probabilities, which is further guaranteed by continuity of the conditional density of the running variable.

Given monotonicity in A3 and smoothness in A1b, $E (x_i \mid z_i = z) = E (x_{0i} \mid z_i = z) = \Pr [x_{1i} = 1, x_{0i} = 1 \mid z_i = z]$ for $z < z_0$ and $E (x_i \mid z_i = z) = E (x_{1i} \mid z_i = z) = \Pr [x_{1i} = 1, x_{0i} = 1 \mid z_i = z] + \Pr [x_{1i} = 1, x_{0i} = 0 \mid z_i = z]$ for $z \geq z_0$ 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 local linear regressions, which are typically used to estimate RD models. With a local linear regression based on a uniform kernel, $x_i = a_0 + a_1 (z_i - z_0) + b_0 d_i + b_1 d_i (z_i - z_0) + u_i$ where $u_i$ is an error term, $E [x_i \mid z_i = z] = a_0 + a_1 (z - z_0)$ for $z < z_0$ and $E [x_i \mid z_i = z] = (a_0 + b_0) + (a_1 + b_1) (z - z_0)$ for $z \geq z_0$. Both are continuous by specification.

Given $x_i = x_{0i} + d_i (x_{1i} - x_{0i})$ and monotonicity in A3, $x_i = d_i$ for compliers, i.e., for compliers treatment is entirely determined by whether one is above or below the cutoff. Continuity of the density of the running variable for compliers then requires that compliers do not have precise control over the running variable and hence treatment is randomly assigned among compliers. Note that 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 continuity assumptions in A1a.

---

7Given A1a, the joint distribution for $(y_{0i}, y_{1i}, x_{0i}, x_{1i})$ 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.
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 $E \left[ y_{1i} - y_{0i} \mid z_i = z_0, C \right] = \frac{y^+ - y^-}{x^+ - x^-}$.

The above theorem shows that assumptions A1b, A2 and A3 suffice to obtain the standard RD identification result as established by HTV, i.e., the ratio of the discontinuity in the mean outcome to the discontinuity in the treatment probability is the LATE for compliers at the RD cutoff.

Note that compared with A1b, A1a is stronger than necessary for identification of LATE in fuzzy design RD. However, A1a is more appealing for its plausible behavioral interpretation, and for its testable implications. Continuity of $f_{z|w}(z \mid w)$ implies no precise manipulation of the running variable and hence no sorting around the RD cutoff (see Lee 2008 for discussion in sharp design and related discussion in McCrary 2008). Standard tests for manipulation include testing continuity of the conditional means of pre-determined covariates (Lee 2008 and Lee and Lemieux 2010) and testing continuity of the unconditional density of the running variable (McCrary, 2008). The above results then ensure that these tests are also valid for fuzzy design RD.

The above result can be more intuitively interpreted using A1a and the result of Imbens and Angrist (1994). By the proof of the Lemma, continuity of $f_{z|w}(z \mid w)$ implies continuity of $f_{w|z}(w \mid z)$, i.e., the mixed joint density of $(y_{0i}, y_{1i}, x_{0i}, x_{1i})$ conditional on $z_i = z$ is continuous in $z$ at $z = z_0$. It follows that any functions of $(y_{0i}, y_{1i}, x_{0i}, x_{1i})$ are continuous in $z$ at $z = z_0$, so the left limits always equal the right limits; therefore, $(y_{0i}, y_{1i}, x_{0i}, x_{1i})$ is jointly independent of the crossing threshold dummy $d_i$ in the limit when $z$ approaches $z_0$. By Theorem 1 of Imbens and Angrist (1994), the ratio of 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.

Note that this paper’s discussion focuses on identifying mean treatment effects; however, given assumption A1a, the conditional distribution of $(y_{0i}, y_{1i}, x_{0i}, x_{1i})$ is continuous at the RD cutoff, so one can more generally identify any distributional effects. One example is Frandsen, Frolich and Melly (2012), who identify the distribution of $y_{1i}$ and $y_{0i}$ and hence quantile treatment effects in RD models. One can show that given A1a, the assumptions in Frandsen, Frolich and Melly (2012) hold.

A few existing studies of RD identification do not rely on the assumption that treatment is independent of the treatment effect. For example, Dong and Lewbel (2014) impose further smoothness,
in particular, continuous differentiability of similar conditional mean functions to identify the treat-
ment effect derivative (TED) and the marginal threshold treatment effect (MTTE) in RD models. The
former is the derivative of the treatment effect with respect to the running variable, corresponding to
the slope change at the cutoff, while the latter refers to the change in the treatment effect resulting
from a marginal change in the cutoff threshold. Dong (2013) similarly utilizes smoothness of condi-
tional means of potential outcomes and smoothness of probabilities of types to identify the standard
RD LATE in a context where the probability of treatment has a kink instead of (or in addition to) a
discontinuity. Frandsen, Frolich and Melly (2012) utilize continuity of the conditional distribution
function of potential outcomes to identify quantile treatment effects in RD models. In addition, Bat-
tistin 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.

4 Testing Independence

In order to investigate whether the independence assumption by HTV holds in an empirical appli-
cation, additional assumptions are needed. Here I assume minimal further smoothness (assuming
continuous differentiability instead of just continuity) and test validity of this independence assump-
tion based on the results of Dong and Lewbel (2014).9

Dong and Lewbel (2014) shows that assuming continuous differentiability instead of just continu-
ity, one can nonparametrically identify and estimate the derivative of the treatment effect with respect
to the running variable at the RD cutoff $\tau_1 E(y_{1i} - y_{0i} | z_i = z_0, C)$, referred to as the treatment effect
derivative or TED. Intuitively, one can think of TED as the coefficient of the interaction term between
the treatment $x_i$ and $z_i - z_0$ in a (local) linear regression of $y_i$ on a constant, $z_i - z_0$, $x_i$ and $x_i(z_i - z_0)$.
Assuming a uniform kernel, $y_i = \gamma_0 + \gamma_1 (z_i - z_0) + \tau_0 x_i + \tau_1 x_i (z_i - z_0) + \nu_i$, and so TED is
$\tau_1$, which measures how the treatment effect depends on the running variable near the cutoff $z_0$, and

9This is a minimal further smoothness assumption in that 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. Also similar to testing the continuity assumption, one can partially test this
further smoothness assumption.
corresponds to the slope change at the cutoff in sharp design RD.

The independence assumption by HTV requires that treatment effect does not depend on the running variable in the neighborhood of the cutoff and hence that TED is zero. One can then test this 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. Such tests include testing continuity of covariates means and continuity of the density of the running variable at the cutoff. One can similarly test the further smoothness required for identifying TED by testing continuity of the slope of the density and the slopes of the conditional means at the RD cutoff.

5 Empirical Application

To illustrate the results in the previous sections, this section estimates the RD model of incumbency advantage in the US house election following Lee (2008) and Lee and Lemieux (2010). I show that in this case the smoothness assumption plausibly holds, since they are partially testable, 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 difference between the Democratic Party’s vote share and its strongest opponent’s in election $t$, and the outcome is a democrat winning in election $t + 1$.

I use the same sample used in Lee (2008) and estimate similar local linear and local polynomial regressions. The data consist of 6,558 elections over the 1946 - 98 period (see Lee 2008 for more detail). I report not only the RD estimates of the treatment effect of being an incumbent party on the probability of winning the next election, but also the derivative of the treatment effect with respect to the running variable. This derivative measures how the incumbency advantage depends on the incumbent party’s winning margin in the neighborhood of the cutoff, corresponding to the slope change at the cutoff in Figure 1.

The estimates are reported in the top panel of Table 1 below. The last row gives estimates from a local linear regression using the optimal bandwidth chosen by a cross-validation procedure described in Lee and Lemieux (2010). The rest are estimates based on low-order polynomials with various
Table 1 RD estimates of the treatment effect and treatment effect derivative (TED)

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

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.

The estimated incumbency effects and the derivatives are robust to different bandwidths and different orders of polynomials. Consistent with those reported 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 an incumbent party, it increases their probability of winning the next election by 36% - 42%. The estimated derivative based on the optimal bandwidth (0.172) is 1.134 and based on the full range of data is 1.293. Both are statistically significant. Estimates based on other bandwidths ranging from 0.15 to 1, the full sample. The order of polynomial for each bandwidth is chosen using the Akaike information criterion (AIC). Following Lee and Lemieux (2010) I adopt a uniform kernel but explore the sensitivity of the estimates to greatly varying bandwidths. Robust standard errors are calculated as proposed by Imbens and Lemieux (2008).

10The range of bandwidth is restricted to be 0.15 or above, because the optimal cross-validation bandwidth is 0.172 and for bandwidths less than 0.10, the optimal order of polynomial is zero.
bandwidths are of similar magnitude, so if the Democrats’ winning margin increases by 1 percentage point, 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’s winning margin, and so the HTV independence assumption can be rejected. 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 and the RD estimates are still valid.

6 Conclusions

This paper shows that RD models can identify LATE under a weak behavioral assumption, instead of the assumptions by Hahn, Todd and Van der Klaauw (2001). In particular, continuity of the conditional density of the running variable, along with existence of a discontinuity in the treatment probability and monotonicity, can establish identification of a causal effect not only in sharp design RD but also in fuzzy design RD. This continuity assumption requires that no individuals (always takers, never takers or compliers) have precise control over the running variable to be just above or below the RD threshold, and that no other programs or treatments use the same threshold, which can be readily tested.

I briefly discuss how to test the independence assumption by HTV based on a minimal smoothness assumption and provide an empirical application showing that the smoothness assumption required plausibly holds while the HTV independence assumption does not, given smoothness.

This paper extends the results of Lee (2008) to fuzzy design RD (with sharp design following as a special case) and formally establishes RD identification based on a similar smoothness assumption. Results in this paper provides formal support for the popular practice of performing McCrary’s (2008) density test to assess validity of fuzzy design RD.
7 Appendix

Proof of Lemma 1: Let $f_z(z)$ be the unconditional density of $z_i$. Continuity of $f_z(w | z)$ implies continuity of $f_z(z)$ in $z = z_0$. Let $f_{w|z}(w | z)$ denote the mixed joint density of $w_i$ conditional on $z_i = z$. That is, $f_{w|z}(w | z) = f_{y_0,y_1|x_0,x_1,y}(y_0,y_1,z | x_0 = t_0, x_1 = t_1) Pr(x_0 = t_0, x_1 = t_1)/f_z(z)$. By Bayes’ Rule, $f_{w|z}(w | z) = f_z(w | z) f(w)/f(z)$, so continuity of $f_z(w | z)$ and $f_z(z)$ implies continuity of $f_{w|z}(w | z)$ in $z = z_0$. It follows immediately that $Pr(x_0 = t_0, x_1 = t_1 | z_i = z)$ for $t_0 = 0, 1$ and $t_1 = 0, 1$ is continuous in $z$ at $z = z_0$, since $Pr(x_0 = t_0, x_1 = t_1 | z_i = z) = \int y_1 \int_{y_0} f_{w|z}(w | z) dy_0 dy_1$.

Let $f_{y_0,y_1|x_0,x_1}(y_0,y_1 | x_0 = t_0, x_1 = t_1, z_i = z)$ be the conditional density of $y_0$, $y_1$ conditional on $x_0$, $x_1$, and $z_i$. Again by Bayes’ Rule, $f_{y_0,y_1|x_0,x_1}(y_0,y_1 | x_0 = t_0, x_1 = t_1, z_i = z) = f_{w|z}(w | z)/Pr(x_0 = t_0, x_1 = t_1 | z_i = z)$. Both $f_{w|z}(w | z)$ and $Pr(x_0 = t_0, x_1 = t_1 | z_i = z)$ are continuous in $z$ at $z = z_0$, so $f_{y_0,y_1,x_0,x_1}(y_0,y_1 | x_0 = t_0, x_1 = t_1, z_i = z)$ for $t_0 = 0, 1$ and $t_1 = 0, 1$ is continuous in $z$ at $z = z_0$. It follows that $E(y_0 | x_0 = t_0, x_1 = t_1, z_i = z)$, $E(y_1 | x_0 = t_0, x_1 = t_1, z_i = z)$, and $E(\beta_i | x_0 = t_0, x_1 = t_1, z_i = z)$ for $t_0 = 0, 1$ and $t_1 = 0, 1$ are continuous in $z$ at $z = z_0$.

Proof of Theorem 1: For convenience, let $E^+$ and $E^-$ be the right or left limit of an expectation function, respectively. Given assumptions A1b and A3 as well as the definitions of individual types, we have

\[
y^+ - y^- \equiv \lim_{\varepsilon \to 0} E[y_i | z_i = z_0 + \varepsilon] - \lim_{\varepsilon \to 0} E[y_i | z_i = z_0 - \varepsilon] \\
= \lim_{\varepsilon \to 0} E[a_i + \beta_i x_i | z_i = z_0 + \varepsilon] - \lim_{\varepsilon \to 0} E[a_i + \beta_i x_i | z_i = z_0 - \varepsilon] \\
= \lim_{\varepsilon \to 0} \left\{ E[a_i | z_i = z_0 + \varepsilon, x_i = 0] Pr[x_i = 0 | z_i = z_0 + \varepsilon] \right\} \\
+ \lim_{\varepsilon \to 0} \left\{ E[a_i + \beta_i | z_i = z_0 + \varepsilon, x_i = 1] Pr[x_i = 1 | z_i = z_0 + \varepsilon] \right\} \\
- \lim_{\varepsilon \to 0} \left\{ E[a_i | z_i = z_0 - \varepsilon, x_i = 0] Pr[x_i = 0 | z_i = z_0 - \varepsilon] \right\} \\
- \lim_{\varepsilon \to 0} \left\{ E[a_i + \beta_i | z_i = z_0 - \varepsilon, x_i = 1] Pr[x_i = 1 | z_i = z_0 - \varepsilon] \right\} \\
= \lim_{\varepsilon \to 0} \left\{ E[a_i | z_i = z_0 + \varepsilon, x_{0i} = 0] Pr[x_{0i} = 0 | z_i = z_0 + \varepsilon] \right\} \\
+ \lim_{\varepsilon \to 0} \left\{ E[a_i + \beta_i | z_i = z_0 + \varepsilon, x_{0i} = 1] Pr[x_{0i} = 1 | z_i = z_0 + \varepsilon] \right\} \\
- \lim_{\varepsilon \to 0} \left\{ E[a_i | z_i = z_0 - \varepsilon, x_{0i} = 0] Pr[x_{0i} = 0 | z_i = z_0 - \varepsilon] \right\} \\
- \lim_{\varepsilon \to 0} \left\{ E[a_i + \beta_i | z_i = z_0 - \varepsilon, x_{0i} = 1] Pr[x_{0i} = 1 | z_i = z_0 - \varepsilon] \right\}
\]
where the fourth equality follows from continuity of the density of \( z_i \), and continuity of the conditional means of the potential treatment \( E \left[ x_{1i} \mid z_i = z \right] \) and \( E \left[ x_{0i} \mid z_i = z \right] \) at \( z = z_0 \), the fifth equality follows from monotonicity, and the last equality follows from continuity of the conditional means of potential outcomes for each type of individuals at \( z = z_0 \).
where the third equality follows from continuity of the density of $z_i$ and continuity of $E [x_{1i} \mid z_i = z]$ and $E [x_0i \mid z_i = z]$ at $z = z_0$, and the fourth equality follows from the monotonicity assumption.

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


