An Alternative Assumption to Identify LATE in Regression Discontinuity Designs

Yingying Dong

University of California Irvine*

This version: March 2016

Abstract

One of the key identifying assumptions for regression discontinuity (RD) designs is the local independence assumption (LIA). This paper shows that LIA puts a restriction on treatment effect heterogeneity and hence may not hold in many empirical applications. This paper then shows that LATE in both sharp and fuzzy designs can be identified under alternative smoothness conditions, and that the required smoothness can be satisfied given a weak and empirically plausibly behavioral assumption, in the spirit of Lee (2008). A sufficient (but stronger than necessary condition) is smoothness of the conditional density of the running variable, which provides formal justification for McCrary's (2008) density test in fuzzy RD designs. Theoretical and empirical relevance of the discussion is illustrated in two empirical applications.

JEL codes: C21, C25

Keywords: Regression discontinuity, Sharp design, Fuzzy design, LATE, Treatment effect heterogeneity, Treatment effect derivative, 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. One of the key assumptions used by HTV is that the treatment effect and potential treatment status

*I would like to thank Josh Angrist, Arthur Lewbel, Yaming Yu, Damon Clark, Marianne Bitler, Kevin Roth, David Powell and Brigham Frandsen for their helpful 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/

are jointly independent of the running variable in the neighborhood of the RD cutoff.[1] This local independence assumption (LIA) is a local version of the independence assumption proposed in the original LATE paper by Imbens and Angrist (1994).

This paper shows that it is both theoretically and empirically useful to relax LIA in RD designs. First of all, restricting the treatment effect to be independent of the running variable places an undesirable restriction on treatment effect heterogeneity, since in RD designs the running variable is frequently one of the key determinants of (or at least is correlated with) outcomes.[2,3] As I show later, this type of heterogeneity would arise naturally in many empirical scenarios. Second of all, by not allowing the treatment effect to depend on the running variable (implying no slope or higher order derivative changes right at the RD cutoff in the sharp RD design), such an independence assumption prevents the opportunity to explore any discrete slope or derivative changes at the RD cutoff. In practice, taking into account the slope change serves as the basis for recent papers such as Calonico, Cattaneo and Titiunik (2014), Dong and Lewbel (2015) and Bertanha and Imbens (2014) among others.

This paper provides formal smoothness conditions that suffice to identify LATE in both sharp and fuzzy RD designs without LIA. This paper further relates the required smoothness to a weak behav-

[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, these alternative assumptions rule out self-selection into treatment based on idiosyncratic gains, and so is often not be realistic.

[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.

[3] 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.

ioral assumption, in the spirit of Lee (2008). For the special case of sharp RD designs, Lee (2008) provides behavioral assumptions that lead to continuity of the conditional density (conditional on an individual's 'identity') of the running variable, and hence local randomization and causal inference. In contrast, this paper discusses identification of LATE in RD designs based on minimal smoothness conditions and then similarly relate smoothness to a lack of manipulation over the running variable. The analysis here 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 theorem provides precisely the same identification results as those by HTV, but under a smoothness assumption instead of LIA.

This paper also discusses a testable implication of LIA given smoothness and evaluates the two alternative assumptions empirically. In particular, LIA implies locally constant treatment effect. When LIA is plausible, one does not need to under-smooth in order to shrink the bias to zero and hence to have correct inference. The robust biased-corrected inference proposed by Calonico, Cattaneo and Titiunik (2014) will be asymptotically equivalent to the inference without bias-correction.

Results in this paper provide formal support for performing McCrary's (2008) density test in fuzzy design RD. Many empirical applications implement McCrary's (2008) density test to assess the plausibility of RD assumptions. Despite its popularity, no justification for this test exists for fuzzy designs. Hitherto, the only rationale was Lee's sharp design analysis.

There exist several RD studies that do not impose independence. 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, continuity of the conditional mean of potential outcome without treatment is sufficient for identification, though they do not discuss more generally weaker assumptions for RD identification. Frandsen, Frolich and Melly (2012) provide identification and estimation of quantile RD treatment effects without imposing the independence assumption. In contrast, this paper discusses theoretically and empirically identification of the standard RD LATE under two alternative assumptions, i.e., smoothness vs. local independence, and relates the identifying assumption to behavioral interpretations. More recently, Bertanha and Imbens (2014) adopt a similar smoothness assumption to evaluate the external validity in fuzzy RD designs. Yanagi (2015) adopts precisely this paper's framework to provide identification of a local weighted average treatment effect (LWATE) in the presence of measurement error in RD designs.

The rest of the paper is organized as follows: The next section shows a few motivating empirical examples. Section 3 provides RD identification under the alternative assumptions. Section 4 discusses the smoothness condition vs. LIA. Section 5 presents two empirical applications evaluating both assumptions. Brief concluding remarks are provided in Section 6.

2 Motivating Examples

To motivate the discussion in this paper, consider an outcome model (or a local linear approximation of it) $y_i = a + b(z_i - z_0) + \tau x_i + \tau_1 (z_i - z_0) x_i + e_i$, where y_i is the observed outcome, x_i is a binary treatment indicator, so $x_i = 1$ if treated and 0 otherwise, and z_i is the running variable with z_0 being the RD threshold. The treatment effect is then $\tau + \tau_1 z_i$. LIA requires the treatment effect to be independent of z_i near z_0, so $\tau_1 = 0$. In sharp design, $\tau_1 = 0$ means no slope change at the cutoff.

In the following, I briefly discuss a few sharp and fuzzy RD examples to motivate the discussion in this paper. First consider the standard sharp RD design estimating the electoral advantage of incumbency in the US house of representatives in Lee (2008). The treatment x_i is the Democratic Party being the incumbent party. The running variable z_i is the Democratic Party's winning margin, or the

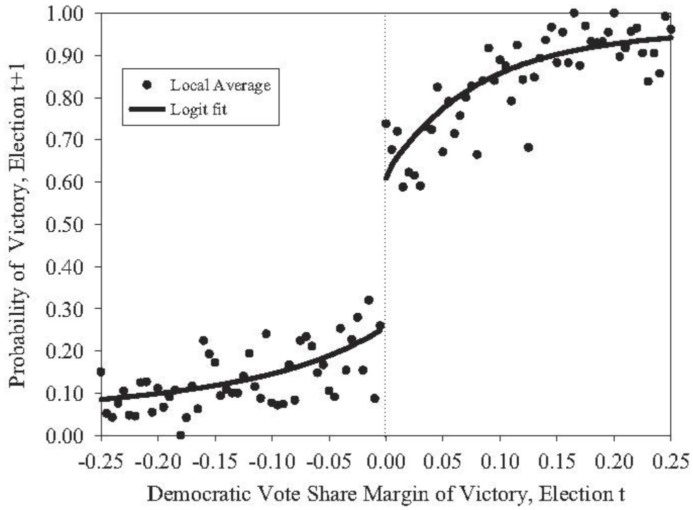

Figure 1: Probability of the Democratic Party winning election t+1 against its winning margin in election t

Democratic Party's vote share minus its strongest opponent's, so $z_0 = 0$. 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. Intuitively, the incumbent party's electoral advantage may depend on its winning margin either directly or indirectly.

Figure 1, which is reproduced from Figure 5-(a) in Lee (2008), 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 the incumbent party's share is in the previous election, the greater their chance of winning the next election, i.e., the incumbency advantage may well depend 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 wavers at in-state public colleges in Massachusetts, United States, with the goal of attracting talented students to the state's public colleges. The treatment x_i is eligibility for the Adams Scholarship. It is determined by whether z_i, 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, and hence the treatment effect, likely depends on her test score, which if true would invalidate LIA. Note that as will be shown later, the dramatic downward slope change is induced neither by manipulation or by missing covariates in this case. Using a differences in differences (DID) approach, 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 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.

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. LIA 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

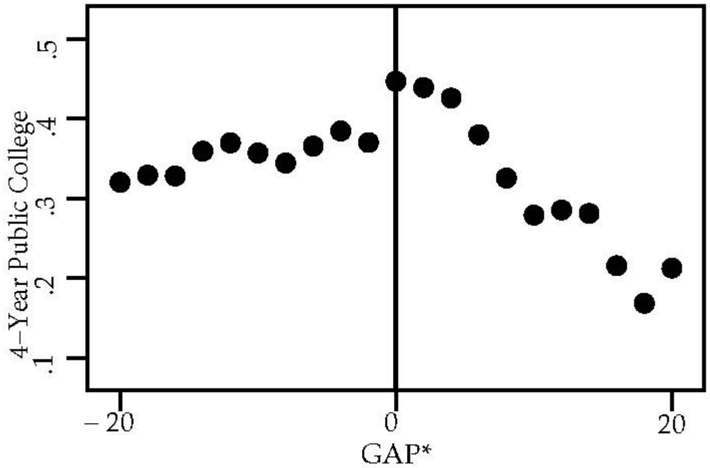

Figure 2: Probability of choosing a 4-year public college against the grade points from the eligibility threshold

academic performance.

3 Identification Given Smoothness

This section discusses formal identification of the LATE in RD designs with smoothness conditions. The discussion focuses on the fuzzy design, treating the sharp design as a special case. To facilitate the discussion, I use the same notation as in HTV (2001). Let y_{1i} and y_{0i} be the potential outcomes for an individual i under treatment or no treatment, respectively (Neyman 1923, Fisher 1935, Rubin 1974, 1990). Recall that x_i is the treatment indicator. The observed outcome can then be written as $y_i = \alpha_i + \beta_i x_i$, where $\alpha_i := y_{0i}$, and $\beta_i := y_{1i} - y_{0i}$. 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 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, and z_0 is the RD cutoff. In discussing the fuzzy RD design 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).

In the following I show that what is required to identify RD LATE is continuity of the conditional probability of different types of individuals (i.e., always takers, never takers, compliers and defiers) and continuity of type-specific means of potential outcomes, conditional on the running variable in the neighborhood of the RD cutoff. In the sharp design, everyone is a complier, the required smoothness reduces to continuity of the conditional means of potential outcomes. Further I show that the required smoothness is readily satisfied given continuity of the density of the running variable for every 'individual' characterized by potential outcomes and types, extending Lee (2008)'s sharp design results to handle fuzzy designs. However, note that smoothness of such conditional density is sufficient but stronger than necessary to identify average treatment effects.

Let an individual's treatment status below the RD cutoff be generated by the random function $x_{0i}(z)$ for any $z \in (z_0 - \varepsilon, z_0)$ and that above the RD cutoff be generate by $x_{1i}(z)$ for $z \in [z_0, z_0 - \varepsilon)$ for some small $\varepsilon > 0$. Further define an individual's counterfactual status just above or just below the cutoff as $x_{1i} := \lim_{\varepsilon \to 0} x_{1i}(z_0 + \varepsilon)$ and $x_{0i} := \lim_{\varepsilon \to 0} x_{0i}(z_0 - \varepsilon)$, respectively, if these limits exist.[4]

For the known fixed threshold $z = z_0$, one can then define four types of individuals following Angrist, Imbens, and Rubin (1996): $\psi_i = A$ if $x_{1i} = x_{0i} = 1$ (always takers), $\psi_i = N$ if $x_{1i} = x_{0i} = 0$ (never takers), $\psi_i = C$ if $x_{1i} > x_{0i}$ (compliers), and $\psi_i = D$ if $x_{1i} < x_{0i}$ (defiers).

ASSUMPTION A1a (Smoothness): $E\left[y_{ti} \mid \psi_i = \psi, z_i = z\right]$ and $\Pr\left[\psi_i = \psi \mid z_i = z\right]$, for $t \in \{0, 1\}$ and $\psi \in \{A, N, C\}$ are continuous in z at $z = z_0$.

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

A1a assumes smoothness of conditional means (probabilities), which replaces LIA. A1a states that the conditional means of potential outcomes y_{0i} and y_{1i} for each type of individuals $\psi \in \{A, N, C\}$ and the probabilities of different types are continuous at the cutoff z_0.

A1a nests the sharp design assumption by HTV as a special case. For sharp design, everyone is a complier, so A1a 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 $z = z_0$.

Define the random vector $\mathbf{w}_i := (y_{0i}, y_{1i}, \psi_i)$ with support $\mathcal{W}$. Denote the conditional density of the running variable z_i conditional on $\mathbf{w}_i$ as $f_{z|\mathbf{w}}(\cdot)$ and the density of the running variable as $f_z(\cdot)$.

ASSUMPTION A1b (Stronger Smoothness): $f_{z|\mathbf{w}}(\cdot)$ is continuous in a neighborhood of $z = z_0$ for all $\mathbf{w} \in \mathcal{W}$. $f_z(\cdot)$ is continuous and strictly positive in a neighborhood of $z = z_0$.

Assumption A1b is a statement asserting that for each individual defined by $\mathbf{w}_i$ the density of the running variable z_i is continuous. For an individual i, given her draw of the running variable, $\mathbf{w}_i$ completely determines her treatment status x_i and outcome y_i. A1b 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.

Given A1b, 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 $f_{z|\mathbf{w}}(\cdot)$ 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.

$\mathbf{w}_i := (y_{0i}, y_{1i}, \psi_i)$ puts no restrictions on treatment effect heterogeneity. Therefore, the alternative assumptions justify identification of the RD LATE even when the treatment effect is arbitrarily heterogenous (and particularly is correlated with the running variable). These assumptions allow for self-selection into treatment based on idiosyncratic gains and selection into different types. So, e.g., there can be endogenous selection into compliers, as long as the probability of being a complier is smooth at the cutoff.

LEMMA: Given A1b, A1a holds.

Proofs are in the Appendix. Compared with A1a, A1b is stronger than necessary. For example, one may have missing observations above or below the cutoff so that the density has a discontinuity. As long as at both sides the observations are missing at random, A1a may still hold. A1b is more appealing considering its plausible behavioral interpretation and testable implications. For example, A1a provides a formal support for using the McCrary (2008)'s density test to assess the validity of fuzzy design RD (see Lee 2008 for discussion in sharp design). Note that one can not test the continuity of the conditional density, but only the unconditional density of the running variable. It then follows that what is tested is neither sufficient nor necessary for the validity of an RD design.

Define $x^+ := \lim_{\varepsilon\to 0} E\left[x_i \mid z_i = z+\varepsilon\right]$, $x^- := \lim_{\varepsilon\to 0} E\left[x_i \mid z_i = z-\varepsilon\right]$, $y^+ := \lim_{\varepsilon\to 0} E\left[y_i \mid z_i = z+\varepsilon\right]$ and $y^- := \lim_{\varepsilon\to 0} E\left[y_i \mid z_i = z-\varepsilon\right]$. These limits exist, given our assumption of smoothness (see the proof of the theorem).

ASSUMPTION A2 (Monotonicity): $\Pr\left(\psi_i = D\right) = 0$.

ASSUMPTION A3 (RD): $x^+ \neq x^-$.

A2 rules out defiers. A2 can be weakened by the assumption requiring conditionally more compliers than defiers, conditional on potential outcomes (de Chaisemartin, 2015). A3 assumes a first-stage exists, i.e., there is a positive fraction of compliers. Both A2 and A3 are also assumed by HTV, following from Imbens and Angrist (1994). The only difference is that the above assumes 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. Instead I assume A1a, which is guaranteed by A1b. [5]

THEOREM: Given assumptions A1a, A2 and A3, the local average treatment effect for compliers at $z_i = z_0$ is identified and is given by $E\left[y_{1i} - y_{0i} \mid z_i = z_0, \psi_i = C\right] = \frac{y^+ - y^-}{x^+ - x^-}$.

This theorem shows that assumptions A1a, A2 and A3 suffice to obtain the standard RD identification results established in HTV (2001).

The above theorem shows identification of mean treatment effects in RD models. Given the stronger smoothness A1b, the distribution of $\mathbf{w}_i := (y_{0i}, y_{1i}, \psi_i)$ 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).

Intuitively, no precise manipulation implied by A1b means that $d_i := 1(z_i \geq z_0)$ is a valid instrument, where $1(\cdot)$ is an indicator function equal to 1 if the expression in the bracket is true, and 0 otherwise. Then by Theorem 1 of Imbens and Angrist (1994), the ratio of $y^+ - y^-$ to $x^+ - x^-$ (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.

4 Smoothness vs. Local Independence

In the following I discuss the theoretical and empirical importance of relaxing LIA. LIA implies that treatment effects are locally constant, while the alternative smoothness assumption requires only that treatment effects be smooth near the RD cutoff. Therefore the alternative assumption can be useful

[5] Given A1b, the joint distribution for (y_{0i}, y_{1i}, ψ_i) 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.

whenever it is necessary to incorporate such treatment effect heterogeneity and/or smoothness.

One example is Calonico, Cattaneo and Tituinik (2014). They propose robust bias-corrected confidence intervals that are not sensitive to "too large" bandwidth choices. Under LIA, such a bias correction and hence robust inference would not be necessary. Take for an example a sharp design RD model, the bias from the local linear regression above or below depends on the derivative of the conditional mean of the outcome right above and below the RD cutoff. Under the local independence assumption, the two derivatives will be the same and hence the incurred biases from above and below will be cancelled out.[6]

Another example is Dong and Lewbel (2014). They investigate external validity of RD LATE, and propose a non-parametric approach of extrapolating the RD LATE away from the RD cutoff, and identifying how the RD LATE would change when the threshold is marginally changed with additional assumptions. Under LIA, the external validity away from the cutoff is guaranteed and so one can directly apply the standard RD LATE to points near but not at the RD cutoff.

A third example is Bertanha and Imbens (2014), who also discuss evaluating the external validity of the RD LATE at the current threshold. In particular, they propose simple tests to evaluate whether one can generalize the RD LATE to subpopulations other than compliers, and to subpopulations other than those with forcing variable equal to the threshold. They employ smoothness conditions (citing this paper) since under LIA, external validity of the RD LATE to points other than the RD cutoff holds automatically.

One may assess the validity of LIA, given smoothness. LIA requires that individuals' treatment effects and types do not depend on the running variable near the RD cutoff. Therefore, given LIA,

[6] For fuzzy design RD models, the bias depends on the derivative differences in both the conditional means of treatment and the conditional means of the outcome from above and blow the RD cutoff.

$\partial E(y_{1i} - y_{0i} \mid z_i = z, \psi_i = C)/\partial z \mid_{z=z_0} = 0$. As discussed, if the true outcome model or a local linear approximation of it (with a uniform kernel) is $y_i = a + b(z_i - z_0) + \tau x_i + \tau_1 (z_i - z_0) x_i + e_i$ for observations near the RD cutoff, then LIA implies that $\tau_1 = 0$. In sharp design RD models, the treatment is $x_i = d_i := 1(z_i \geq z_0)$, so $\tau_1 = 0$ means no slope change at the RD cutoff.

Under minimal further smoothness assumptions, i.e., assuming continuous differentiability instead of just continuity of the conditional means and probabilities in A1b, one can nonparametrically identify and estimate $\partial E(y_{1i} - y_{0i} \mid z_i = z, \psi_i = C)/\partial z \mid_{z=z_0}$ for both sharp and fuzzy RD designs (see, discussion in Dong and Lewbel, 2014). Note that virtually all empirical implementations of RD models satisfy this slightly stronger assumption.[7] For sharp design, this amounts to estimating the slope change at the RD cutoff by local polynomial regressions. For fuzzy design, with a uniform kernel, one can estimate this derivative by the coefficient of the interaction term between the treatment x_i and the (re-centered) running variable $(z_i - z_0)$ in the outcome equation $y_i = a + b(z_i - z_0) + \tau x_i + \tau_1 (z_i - z_0) x_i + e_i$, with d_i and $d_i (z_i - z_0)$ be the excluded IVs. More complicated kernels can be accommodated by estimating weighted two stage least squares (2SLS) instead of 2SLS, with the chosen kernel function as weights. One can then evaluate LIA by testing the significance of this estimated derivative. If it is significant, then LIA likely does not hold.

The alternative smoothness assumption proposed in this paper can be evaluated by testing the continuity of the empirical density of the running variable and the continuity of the pre-determined covariate means. I do not discuss these tests in detail, since they are the standard practice currently. One can similarly test the further smoothness required for identifying the derivative of the treatment effect by testing continuity at the RD cutoff of both the level and slope (the intercept and the first-

[7] 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.

derivative of local polynomial regressions) of the density of the running variable as well as those of the conditional mean functions of pre-determined covariates.

5 Empirical Applications

This section provides two empirical applications. One is for the sharp design, and the other is for the fuzzy design. The goal is to evaluate the plausibility of the two alternative assumptions, i.e., independence vs. smoothness. I provide evidence that the smoothness conditions plausibly hold in both cases, while the independence assumption does not, given smoothness.

5.1 Sharp Design

This section estimates the sharp RD model of incumbency advantage in the US house election, using the same data as those used in Lee (2008) and Lee and Lemieux (2010).[8] Recall that the treatment in this case is an indicator that the Democratic Party was the incumbent party, the running variable is the Democratic Party's winning margin in election t, and the outcome is whether a democratic candidate won in election $t+1$.

The sample consists of 6,558 elections over the 1946 - 98 period (see Lee 2008 for more detail). Following Lee and Lemieux (2010) and Porter (2003), I use local linear regressions to estimate the local causal effect of being an incumbent party. Analogous to using local linear regressions to estimate the means at a boundary point, local quadratic regressions may be appropriate for estimating slopes (see, e.g., discussion in Calonico, Cattaneo, and Titiunik, 2014). I therefore adopt local quadratic regressions to estimate the derivative of the RD treatment effect (corresponding to the slope change)

[8]Caughey and Sekhon (2011) show possible manipulation in this case. However, Lee and Lemieux (2014) notice that this can be explained by the sampling differences between Caughey and Sekhon (2011) and Lee (2008).

at the RD cutoff.

Triangular kernel is recommended for estimating conditional means at a boundary point (Fan and Gijbels, 1996). I adopt the triangular kernel, but also report results using the uniform kernel, which is frequently used for convenience. Three different bandwidth estimators are used to choose the optimal bandwidth for the local linear or local quadratic regressions. These are the plug-in estimator proposed by Calonico, Cattaneo and Titiunik (2014), the plug-in estimator proposed by Imbens and Kalyaranaman (2014), and the cross-validation estimator proposed by Ludwig and Miller (2007). Estimates of the treatment effect as well as the derivative of the treatment effect at the RD cutoff are reported Table 1 and Table 2, respectively. The derivative of the treatment effect 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.

Table 1 Sharp RD Estimates of the Incumbency Advantage

	CCT	IK	CV	CCT_u	IK_u	CV_u
RD LATE	0.387	0.402	0.411	0.364	0.386	0.414
	(0.050)***	(0.046)***	(0.039)***	(0.051)***	(0.046)***	(0.040)***
Bandwidth	0.160	0.147	0.202	0.118	0.116	0.153
N	1,850	1,725	2,291	1,405	1,375	1,784
Polynomial order	1	1	1	1	1	1

Note: This table uses data from Lee (2008); All RD LATE estimates are based on bias-corrected robust inference proposed by Calonico, Cattaneo and Titiunik (CCT, 2014), using local linear regressions; using local linear regressions; CCT and IK refer to the optimal bandwidths proposed by CCT and Imbens and Kalyanaraman (2014), respectively; CV refers to the cross validation bandwidth proposed by Ludwig and Miller (2007); CCT, IK, and CV use the triangular kernel, and CCT_u, IK_u and CV_u use the Uniform kernel; Standard errors are in parentheses; Standard errors are in parentheses; * significant at the 10% level, ** significant at the 5% level, ***significant at the 1% level.

As one can see from Table 1 and Table 2, the estimated incumbency effects and their derivatives are largely robust to different bandwidth choices. Consistent with estimates in Lee (2008) and Lee and Lemieux (2010), the average incumbency effect is estimated to be 0.364 - 0.414, meaning that when the Democratic Party is the incumbent party, it increases their probability of winning the next election by 36.4% to 41.4%. The estimated derivative of the treatment effect is 1.143 - 1.349, so given

Table 2 The Derivative of the Incumbency Advantage

	CCT	IK	CV	CCT_u	IK_u	CV_u
TED	1.349	1.240	1.209	1.391	1.143	1.169
	(0.599)**	(0.442)***	(0.411)***	(0.560)**	(0.434)***	(0.450)***
Bandwidth	0.353	0.432	0.452	0.304	0.361	0.352
N	3,796	4,410	4,570	3,289	3,866	3,785
Polynomial order	2	2	2	2	2	2

Note: This table uses data from Lee (2008); All estimates are based on local quadratic regressions; CCT and IK refer to the optimal bandwidths proposed by Calonico, Cattaneo, and Titiunik (2014) and Imbens and Kalyanaraman (2014), respectively; CV refers to the cross validation bandwidth proposed by Ludwig and Miller (2007); CCT, IK, and CV use the triangular kernel, and CCT_u, IK_u, and CV_u uses the uniform kernel; Bandwidth and sample size N refer to those of the outcome equation; Bootstrapped Standard errors based on 500 simulations are in parentheses; * significant at the 10% level, ** significant at the 5% level, ***significant at the 1% level.

a 1 percentage point increase in the Democrats' winning margin, the probability for their candidates to win the next election increases by 1.143% to 1.349%. The estimated incumbency effects as well as their derivatives are all statistically significant. Significance of the estimated derivative of the treatment effect suggests that the incumbency advantage indeed depends on the incumbent party's winning margin. That is, LIA can be rejected in this case.

Table 3 Smoothness of the Covariate Mean and Density of the Running Variable

	Estimates		Bandwidth	No of obs	Polynomial order
	Previous Election Vote Share				
Jump	-0.001	(0.015)	0.190	2,170	1
Kink	-0.150	(0.337)	0.239	2,663	2
	Density of the Winner Margin in Election t				
Jump	0.138	(0.161)	0.205	82	1
Kink	2.400	(3.629)	0.212	84	2

Note: Standard errors are in parentheses; All estimates use the CCT optimal bandwidth and the triangular kernel.

Table 3 reports at the RD cutoff the estimated jumps or kinks in the conditional mean of an important covariate, the Democratic vote share from the previous election. Also reported is the estimated jump and kink in the empirical density of the running variable at the RD cutoff. I present only estimates by the local linear or quadratic regressions with triangular kernels and bandwidths chosen by the Clonico, Cattaneo and Tituinik's (2014) plug-in estimator. Estimates using uniform kernels and other bandwidths are similar and are therefore suppressed to save space. Estimates in both tables show that none of the estimated jumps and kinks are statistically significant, so the smoothness assumption plausibly holds. These results support that the RD design in this case is valid, even though LIA likely does not hold.

5.2 Fuzzy Design

This section estimates the impact of academic probation on the sbsequent dropout probability, using a fuzzy RD design based on the probation rule in colleges. Nearly all colleges and universities in the US adopt academic probation to motivate students to stay above a certain performance standard. Typically students are placed on academic probation if their GPAs fall below a pre-determined threshold. Students on academic probation face the real threat of being suspended if their performance continues to fall below.

Despite of the prevalence of this academic policy, limited empirical research exists investigating its impacts on students' outcomes. In a seminal study, Lindo et al. (2010) examine the causal effects of academic probation on students' performance using data from a large Canadian university. Fletcher and Tokmouline (2010) perform similar analysis using the US data. They both show that placement on academic probation discourages students from enrolling in school. Both studies adopt the standard sharp RD design to evaluate the effects of the first year (or first-term) probation.

This paper adopts a fuzzy RD design to evaluate the impact of ever placement on probation on the overall college dropout rate. More importantly, this paper evaluates how the discouragement effect of academic probation depends on the running variable, a student's pre-treatment GPA, which would not be allowed for under LIA. Let Y be the binary indicator for dropout, which is 1 if a student drop out of college and 0 otherwise. The running variable R is the first semester GPA. The treatment T is

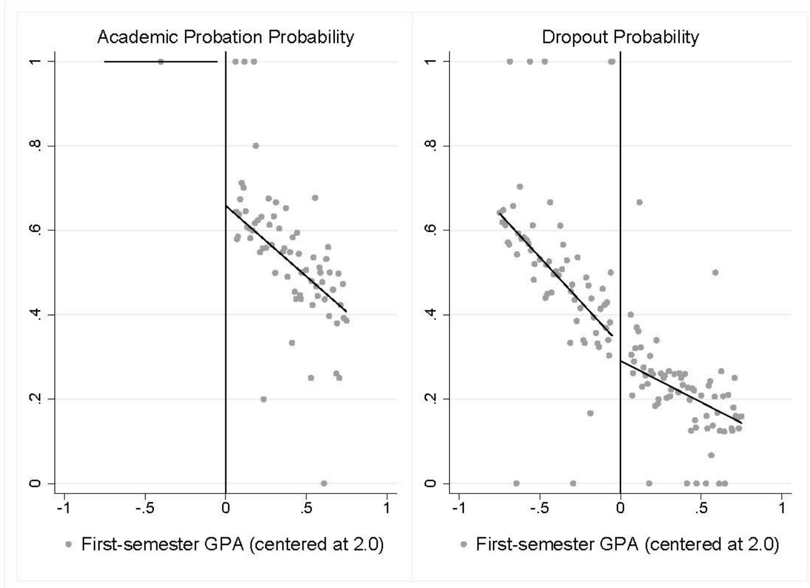

Figure 3: Academic Probation and Dropout Rates against First-semester GPA

the indicator for ever being placed on academic probation in college. I use the confidential data from a large public university in Texas, which are collected under the Texas Higher Education Opportunity Project (THEOP). An undergraduate at this university is considered as 'scholastically deficient' if his or her semester or cumulative GPA falls below 2.0. The actual probation status is not observed in the data. In this paper's analysis, treatment is 1 as long as a student's cumulative or semester GPA is below the school-wide cutoff 2.0, i.e., when a student is considered as 'scholastically deficient.'[9] The data used here represent the entire population of the first-time freshmen cohorts between 1992 and 2002. The total sample size is 64,310.

Figure 3 presents the probability of ever being placed on academic probation and the dropout probability as functions of the first semester GPA. In the left graph, for those whose first semester GPA falls below the probation threshold, the probability of being on probation is 1 by construction;

[9]In practice, when a student is considered as scholastically deficient, he or she may only be given an academic warning. However, a quick survey administered to the relevant academic deans shows that students are generally placed on probation in this case.

for those whose first semester GPA is marginally above, there is still an over 60% chance for them to be placed on academic probation later. This graph shows a dramatic slope change at the RD cutoff, indicating that the probability of compliers depends dramatically on the running variable. In the right graph, the dropout probability also shows a discernible slope change the RD cutoff, in addition to a small level change.

Table 4 Fuzzy RD Estimates of the Impact of Academic Probation on Dropout Rates

	CCT	IK	CV	CCT_u	IK_u	CV_u
1st-stage discontinuity	-0.343	-0.345	-0.352	-0.336	-0.341	-0.378
	(0.010)***	(0.010)***	(0.016)***	(0.010)***	(0.010)***	(0.023)***
RD-LATE	0.068	0.108	0.108	0.073	0.128	0.120
	(0.053)	(0.084)	(0.085)	(0.049)	(0.084)	(0.106)
Bandwidth	0.869	0.723	0.681	0.762	0.568	0.487
N	31,396	25,149	23,623	26,780	19,413	15,763
Polynomial order	1	1	1	1	1	1

Note: This table uses the Administration and Transcript Data from a Public University in Texas; All RD LATE estimates are based on bias-corrected robust inference proposed by Calonico, Cattaneo and Titiunik (CCT, 2014), using local linear regressions; CCT and IK refer to the optimal bandwidths proposed by CCT and Imbens and Kalyanaraman (2014), respectively; CV refers to the cross validation bandwidth proposed by Ludwig and Miller (2007); CCT, IK, and CV use the triangular kernel, and CCT_u, IK_u and CV_u use the Uniform kernel; Standard errors are in parentheses; * significant at the 10% level, ** significant at the 5% level, ***significant at the 1% level.

The estimated change in the probation probability (shown in Table 4) at the cutoff is largely insensitive to the bandwidths and kernel functions used. The estimates range from -33.6% to -37.8%, all of which are statistically significant at the 1% level. Table 4 also reports the estimated RD LATE. Placement on academic probation is shown to have a small, positive, yet insignificant impact on the college dropout rate right at the RD cutoff. However, in Table 5, I report the estimated TED along with the first-stage slope change. The latter indicates how the probability of compliers depend on the running variable at the RD cutoff in this fuzzy RD model. Both sets of derivative estimates are largely robust to the choices of optimal bandwidths and kernel functions and are statistically significant at the 1% level. The estimated TED is -0.568 by the CCT bias-corrected estimation with a triangular kernel. This implies that the discouragement effect of placement on academic probation increases significantly as a student's GPA moves marginally below the cutoff. In particular, when a student's first-semester GPA decreases by 0.1, the probability of dropping out of college once on probation increases by 5.68%. This is large in magnitude, compared with the change of 7-13% in the dropout rate at the cutoff. These results provide strong evidence that both individual types and the RD treatment effects depend on the running variable at the RD cutoff and hence LIA is violated.

Table 5 The derivative of the Impact of Academic Probation on Dropout Rates

	CCT	IK	CV	CCT_u	IK_u	CV_u
1st-stage derivative	-0.371	-0.269	-0.336	-0.400	-0.298	-0.342
	(0.022)***	(0.077)***	(0.033)***	(0.018)***	(0.090)***	(0.045)***
TED	-0.568	-0.584	-0.555	-0.754	-0.639	-0.591
	(0.150)***	(0.376)	(0.144)***	(0.103)***	(0.422)	(0.147)***
Bandwidth	1.604	0.868	1.648	1.807	0.726	1.358
N	54,151	31,396	54,595	59,306	25,185	47,846
Polynomial order	2	2	2	2	2	2

Note: This table uses the Administration and Transcript Data from a Public University in Texas; All estimates are based on local quadratic regressions; CCT and IK refer to the optimal bandwidths proposed by Calonico, Cattaneo and Titiunik (2014) and Imbens and Kalyanaraman (2014), respectively; CV refers to the cross validation bandwidth proposed by Ludwig and Miller (2007); CCT, IK, and CV use the triangular kernel, and CCT_u, IK_u, and CV_u uses the uniform kernel; Bandwidth and sample size N refer to those of the outcome equation; Bootstrapped Standard errors based on 500 simulations are in parentheses; * significant at the 10% level, ** significant at the 5% level, ***significant at the 1% level.

To test the smoothness conditions in this case, I estimate the jumps and kinks in the conditional means of covariates and those in the density of the running variable at the RD cutoff. Covariates investigated include male, Black, and Hispanic dummies, as well as an indicator for being in the top 25% of the high school class. The results are reported in Table 6. None of the estimated jumps and kinks are statistically significant, indicating that the smoothness conditions are plausible and hence the RD design is still valid, even though LIA is violated.

6 Conclusions

This paper shows that given a discontinuity in the treatment probability and monotonicity, identification of LATE in RD designs can be established under just smoothness conditions, in particular

Table 6 Smoothness of Covariate Means and Density of 1st-semester GPA

	Jump		Kink	
Male	-0.007	(0.019)	-0.084	(0.220)
Black	-0.004	(0.010)	0.035	(0.075)
Hispanic	0.010	(0.013)	0.102	(0.172)
Top 25% of HS Class	0.013	(0.018)	0.085	(0.175)
Density of 1st-semester GPA	0.381	(0.369)	-0.266	(1.646)

Note: CCT bias-corrected intercept and slope change estimates are reported; Robust standard errors are in parentheses. The bin size used to generate the empirical density is .006.

smoothness of conditional means of potential outcomes for each type of individuals (always takers, never takers and compliers) and smoothness of individual types. This paper also provides an empirically plausible weak behavioral assumption, in the spirit of Lee (2008), that leads to the required smoothness. Results in this paper formally justify using McCrary's (2008) density test to evaluate the validity of fuzzy design RD models.

Except for its empirical relevance, the smoothness condition can be theoretically important, particularly in cases where one needs to incorporate fully treatment effect heterogeneity or how the treatment effect changes with respect to the running variable (which underlies, e.g., the bias correction and robust inference in Calonico, Cattaneo and Titiunik, 2014). Lastly using the data from Lee (2008 and the confidential THEOP data, I show that in both cases the RD designs are valid, since the smoothness assumptions plausibly hold, even though the local independence assumption may not, given smoothness.

7 Appendix

Proof of Lemma: For simplicity, assume that y_{0i} and y_{1i} are continuous, though analogous analysis can be done when y_{0i} and y_{1i} are discrete. The following discussion applies to $z_i = z \in (z_0 - \varepsilon, z_0 + \varepsilon)$ for some small $\varepsilon > 0$. Let $f_{\cdot}(\cdot)$ and $f_{\cdot|\cdot}(\cdot)$ denote the unconditional and conditional probability density or mass functions, respectively. In particular, let $f_{\mathbf{w}|z}(\mathbf{w} \mid z)$ denote the mixed joint density of $\mathbf{w}_i$ conditional on $z_i = z$, i.e., $f_{\mathbf{w}|z}(\mathbf{w} \mid z) = f_{y_0,y_1,z|\psi}(y_0, y_1, z \mid \psi_i = \psi)\Pr(\psi_i = \psi)/f_z(z)$.

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 $\mathbf{w}_i := (y_{0i}, y_{1i}, \psi_i)$, then probability of each type of individual $\Pr(\psi_i = \psi \mid z_i = z) = \int_{\Omega_1}\int_{\Omega_0} f_{\mathbf{w}|z}(\mathbf{w} \mid z)dy_0dy_1$ for $\psi_i = \psi \in \{A, N, C\}$ is continuous in z at $z = z_0$, where Ω_t is the conditional support of y_{ti} for $t = 0, 1$ conditional on $z_i = z$.

By Bayes' Rule, $f_{y_0,y_1|\psi,z}(y_0, y_1 \mid \psi_i = \psi, z_i = z) = f_{\mathbf{w}|z}(\mathbf{w} \mid z)/Pr(\psi_i = \psi \mid z_i = z)$ for $\psi_i = \psi \in \{A, N, C\}$. Both $f_{\mathbf{w}|z}(\mathbf{w} \mid z)$ and $\Pr(\psi_i = \psi \mid z_i = z)$ are continuous in z at $z = z_0$, so $f_{y_0,y_1|\psi,z}(y_0, y_1 \mid \psi_i = \psi, z_i = z)$ for $\psi_i = \psi \in \{A, N, C\}$ is continuous in z at $z = z_0$. It follows that type-specific conditional means of potential outcome $E(y_{ti} \mid \psi_i = \psi, z_i = z)$ for $t = 0, 1$,and $\psi_i = \psi \in \{A, N, C\}$ are continuous in z at $z = z_0$.

Proof of Theorem: Given assumptions A1a and A3 as well as the definitions of individual types, we have

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

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

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

Therefore,

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

In addition,

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

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

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

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] Bertanha, M. and G. W. Imbens (2014): "External Validity in Fuzzy Regression Discontinuity Designs," NBER Working Paper No. 20773.

[5] 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.

[6] de Chaisemartin, C. (2015): "Tolerating Defiance? Identification of treatment effects without monotonicity," working paper, Department of Economics, University of Warwick.

[7] 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.

[8] Dong Y. and A. Lewbel (2015): "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models," The Review of Economics and Statistics, 97(5), 1081-1092.

[9] Fan, J. and Gijbels, I. (1996), Local Polynomial Modeling and its Implications, Monographs on Statistics and Applied Probability no. 66 (Boca Raton: Chapman and Hall/CRC).

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

[11] Fletcher, J. M. and M. Tokmouline (2010): "The Effects of Academic Probation on College Success: Lending Students a Hand or Kicking Them While They Are Down," Working Paper.

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

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

[14] 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.

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

[16] Imbens W. G. and Kalyanaraman, K. (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," Review of Economic Studies,79 (3): 933-959.

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

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

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

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

[21] 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.

[22] Lindo M. J., N. J. Sanders and P. Oreopoulos (2010): "Ability, Gender, and Performance Standards: Evidence from Academic Probation," American Economic Journal: Applied Economics 2: 95-117.

[23] 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.

[24] Porter, J. (2003): "Estimation in the Regression Discontinuity Model," unpublished manuscript, Department of Economics, University of Wisconsin, Madison.

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

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

[27] 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.

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

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

[30] Yanagi (2015): "Regression Discontinuity Designs with Nonclassical Measurement Error," working paper, Department of Economics, Hitotsubashi University.