Alternative Assumptions to Identify LATE in Fuzzy Regression Discontinuity Designs

Yingying Dong

University of California Irvine*

This version: February 2018

Abstract

There exist two alternative assumptions to identify local average treatment effects (LATE) in fuzzy regression discontinuity (RD) designs: local independence (LI) and local smoothness (LS). Together with the usual LATE assumptions requiring existence of a first-stage and treatment monotonicity, either of these two assumptions is sufficient to identify RD LATE. I discuss the practical (and testable) implications of these alternative assumptions, and show that weakening LI by LS might be empirically relevant. However, when LI does hold, there are some practical implications one may explore. Numerical and empirical examples are briefly presented. (Word Count: 2,907)

JEL codes: C21, C25
Keywords: Regression discontinuity design, Local independence, Local smoothness

1 Introduction

Regression discontinuity (RD) designs have been widely used in many areas of empirical research. There exist two alternative assumptions to identify local average treatment effects (LATE) in fuzzy RD designs: local independence (LI) and local smoothness (LS). Together with the usual LATE assumptions requiring existence of a first-stage and treatment monotonicity, either of these two assumptions is sufficient to identify RD LATE.[1] This paper discusses these alternative assumptions, and show that weakening LI by LS might be empirically relevant. However, when LI does hold, there are some practical implications one may explore.

*Yingying Dong, Department of Economics, University of California Irvine, CA 92697, USA. Email: yyd@uci.edu. http://yingyingdong.com/

[1] This paper focuses on fuzzy RD designs, with sharp design following as a special case.

LI requires that individual treatment effects and potential treatment status are jointly independent of the running variable in the neighborhood of the RD cutoff (Hahn, Todd and van der Klaauw, 2001, hereafter HTK). It is a local version of the independence assumption proposed in the LATE framework (Imbens and Angrist, 1994 and Angrist, Imbens, and Rubin 1996). In contrast, LS only requires that the conditional means (or distributions) of potential outcomes and potential treatment status are smooth near the RD cutoff (see, e.g., Frandsen, Frolich and Melly, 2012, Dong (2015) and Dong and Lewbel, 2015), which can be seen as a smooth parallel of the LATE independence assumption.

In the next Section 2, I formally present LI and LS, and briefly discuss identification of RD LATE under LS. In Section 3, I discuss their practical (and testable) implications, and present an empirical application. Section 4 concludes. All proofs and details of the empirical application are provided in the supplemental online Appendix.

2 LI, LS and RD LATE

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). 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 $\beta_i := y_{1i} - y_{0i}$. Define the potential treatment status as $x_i(z)$ for a given value z that z_i could take on. When z_i is a binary instrument, one of the key assumptions for identifying LATE in Imbens and Angrist (1994, Condition 1 of their Theorem 1) is that $(y_{0i}, y_{1i}, x_i(z))$ is jointly independent of z_i.

In the RD framework, z_i is the running variable, and z_0 is the RD cutoff. The following discussion applies to $z \in (z_0 - \varepsilon, z_0 + \varepsilon)$ for some small $\varepsilon > 0$. In discussing the fuzzy RD design with a variable treatment effect, HTK (Assumption A3 (i)) analogously assume the following LI assumption.

ASSUMPTION LI (HTK, 2001): $(\beta_i, x_i(z))$ is jointly independent of z_i near z_0.

The above assumption implies that the individual treatment effect β_i is independent of the running variable z_i locally near z_0. In practice, potential outcomes can depend on the running variable directly or indirectly through omitted covariates, so LI places a restriction on treatment effect heterogeneity. In the next section, I provide empirical scenarios where this type of heterogeneity arises naturally and should be taken into account for policy evaluation.

In addition, assume $x_i = h(z_i, u_i)$, where u_i is a vector of (un)observables other than the running variable z_i. Then $x_i(z) := h(z, u_i)$. LI requires $x_i(z)$ to be independent of z_i near z_0, which further implies that u_i is independent of z_i near z_0.

LS relaxes the above restrictions. To formally introduce the LS assumption, I extend the definitions of individual types of the LATE framework (Imbens and Angrist, 1994 and Angrist, Imbens, and Rubin, 1996) to the RD setup. For an individual with $z_i = z$, let $x_{1i}(z)$ and $x_{0i}(z)$ be her potential treatment status if she is above or below the cutoff, respectively.[2] Given $z_i = z$, an individual's type $\psi_i(z) = A(z)$ if $x_{1i}(z) = x_{0i}(z) = 1$ (always takers), $\psi_i(z) = N(z)$ if $x_{1i}(z) = x_{0i}(z) = 0$ (never takers), $\psi_i(z) = C(z)$ if $x_{1i}(z) > x_{0i}(z)$ (compliers), and $\psi_i(z) = D(z)$ if $x_{1i}(z) < x_{0i}(z)$ (defiers). For notational convenience, whenever there is no confusion, I will suppress the argument to use x_{0i} (x_{1i}) to denote $x_{0i}(z)$ ($x_{1i}(z)$) and similarly use A, N, C, and D to denote individual types.

ASSUMPTION LS1: $E\left[y_{xi}|\psi_i = \psi, z_i = z\right]$ for $x \in \{0, 1\}$ and $\Pr\left(\psi_i = \psi|z_i = z\right)$ for $\psi \in \{A, N, C, D\}$ are continuous in z at $z = z_0$.

LS1 assumes that the conditional means of potential outcomes for each type of individuals and the probabilities of different types are continuous at the RD cutoff. For sharp designs, everyone is a complier. LS1 then reduces to the assumption that $E\left[y_{xi}|z_i = z\right]$, $x = 0, 1$, are continuous at $z = z_0$.[3]

[2] Let $d_i = 1(z_i \geq z_0)$. d_i is binary and is a deterministic function of z_i, then one can re-write $x_i = h_1(z_i, u_i)\, d_i + h_0(z_i, u_i)(1 - d_i)$, where $h_1(z_i, u_i)$ and $h_0(z_i, u_i)$ describe the treatment assignment above and below the cutoff, respectively. Then define $x_{di}(z) := h_{di}(z, u_i)$. For an individual i, If $z \geq z_0$, $x_{1i}(z)$ is her observed treatment status above the cutoff, while $x_{0i}(z)$ would be her counterfactual treatment status if she were below the cutoff. The converse holds for $z < z_0$.

[3] To identify quantile treatment effects (QTE) in RD designs, Frandsen, Frolich and Melly (2012) assume that the conditional distributions of potential outcomes, conditional on individual types are smooth.

LS1 can be justified by a weak and empirically plausible behavioral 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, which further implies local randomization and hence causal inference. For fuzzy designs, one needs to take into account the probabilities with which individuals may self-select into types such as compliers.

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

ASSUMPTION LS2: $f_{z|\mathbf{w}}(z|\mathbf{w})$ is continuous in a neighborhood of $z = z_0$ for all $\omega \in \mathcal{W}$. $f_z(\cdot)$ is continuous and strictly positive in a neighborhood of $z = z_0$.

LS2 imposes smoothness on the conditional and unconditional densities of the running variable. It asserts that for each 'individual' defined by ω_i, the conditional density of the running variable z_i conditional on ω_i is smooth. By Bayes' Rule, $f_{\mathbf{w}|z}(\mathbf{w}|z) = f_{z|\mathbf{w}}(z|\mathbf{w}) f_{\mathbf{w}}(\mathbf{w})/f_z(z)$, so LS2 implies that $f_{\mathbf{w}|z}(\mathbf{w}|z)$ is continuous in z at $z = z_0$, where $f_{\mathbf{w}|z}(\mathbf{w}|z)$ denotes the (possibly mixed) joint density of $\mathbf{w}_i$ conditional on $z_i = z$. Other than smoothness, LS2 imposes no restrictions on $y_{1i} - y_{0i}$, so treatment effects can be arbitrarily heterogeneous and individuals can self-select into treatment and different types. For example, there can be endogenous selection into compliers, as long as the probability of being a complier is smooth at the cutoff.

LS2 implies LS1, which provides theoretical ground for performing McCrary's (2008) density test in fuzzy RD designs. LS2 is stronger than necessary. The density of z_i can have a discontinuity, but LS1 can still hold. LS2 is appealing considering its plausible behavioral interpretation and testable implications. Note that cannot directly test continuity of the above conditional density, but only the unconditional density of the running variable. Therefore, passing the RD density test is neither necessary nor sufficient for validity of RD designs.

Define $y^+ := \lim_{\varepsilon \to 0} E\left[y_i | z_i = z + \varepsilon\right]$, $y^- := \lim_{\varepsilon \to 0} E\left[y_i | z_i = z - \varepsilon\right]$, and similarly x^+ and x^-.

THEOREM 1: Assume that $\Pr\left(\psi_i = D\right) = 0$ in the neighborhood of z_0, and $x^+ \neq x^-$. Then under LS1, $E\left[y_{1i} - y_{0i}|z_i = z_0, \psi_i = C\right] = \frac{y^+ - y^-}{x^+ - x^-}$.

Assuming no defiers and existence of a discontinuity, LS1 yields the standard RD identification result that was first established by HTK.

3 Discussion and Empirical Illustration

Either LI or LS can be used to identify RD LATE. They also have readily testable implications.

Assume $y_i = g\left(x_i, z_i, \upsilon_i\right)$, where υ_i is a vector of other (un)observable covariates other than the running variable z_i. Consider a local linear approximation of it above and below the cutoff (assuming a uniform kernel): $y_i = \alpha_0 + \alpha_1\left(z_i - z_0\right) + \tau_0 x_i + \tau_1\left(z_i - z_0\right) x_i + e_i$. $\tau_0 + \tau_1 z_i$ is then the treatment effect. LI requires the treatment effect to be independent of z_i near z_0, and hence $\tau_1 = 0$. It is clearest to illustrate LI or lack of treatment effect heterogeneity in the running variable in sharp RD designs: in sharp designs $\tau_1 = 0$ means no slope change at the cutoff; further having treatment effects to be independent of the running variable implies no slope or any higher order derivative changes right at the RD cutoff.

Consider the classic RD design of Lee (2008). x_i is an indicator for the Democratic Party being the incumbent party. z_i is the Democratic Party's winning margin. y_i is whether a Democrat won the next election. LI requires that the incumbent party's electoral advantage does not depend on its winning margin. However, such a dependency may exist directly or indirectly, due to, e.g., omitted (un)observables.

Figure 1 (reproduced from Figure 5-(a) of Lee 2008) shows the probability of a Democrat winning election $t + 1$ given 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 depend on the winning margin.[4]

[4] τ_1 in the RD design of Lee (2008) is estimated to be between 1.143 and 1.349, statistically significant at the 5% or 1% level (provided in the online supplemental Appendix). That is, given a 1 percentage point increase in the Democrats' winning margin, the probability for a Democrat to win the next election increases by 1.143% to 1.349%.

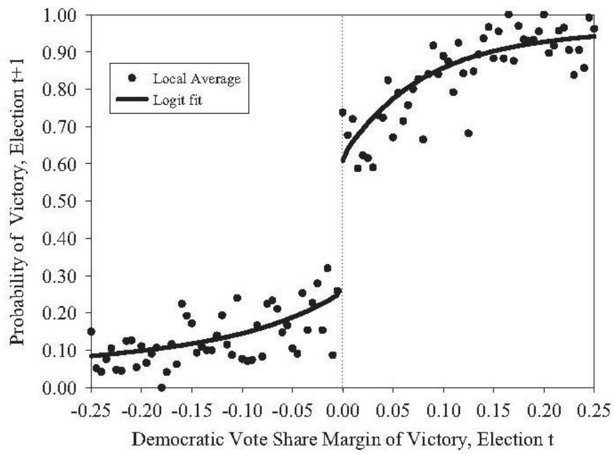

Figure 1: Probability of the Democratic Party winning election t+1 against its winning margin in election t

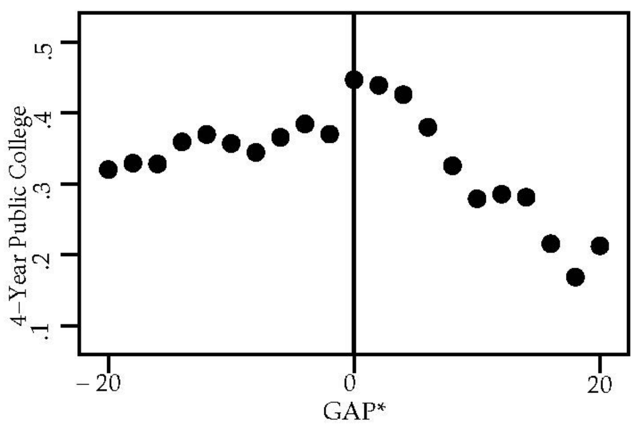

Figure 2: Probability of choosing a 4-year public college against the grade points from the eligibility threshold

Consider another RD design of Goodman (2008) estimating the effect of the Adams Scholarship program on college choices. The treatment x_i is eligibility for the Adams Scholarship. It is determined by a student's standard test score z_i exceeding a certain threshold. The Adams Scholarship program provides qualified students tuition wavers at in-state public colleges in Massachusetts, United States.

Figure 2 shows the probability of choosing a four-year public college conditional on the number of grade points to the eligibility threshold. The dramatic downward slope change at the threshold suggests that a student's response to an Adams Scholarship likely depends on her test score, and thereby invalidates LI. Goodman (2008) indeed shows that marginally qualified react much more strongly to the price change than students with test scores further above the threshold. Students trade college quality with prices.

Better qualified students may be admitted to private colleges of much higher quality, and hence would face a large quality drop had they accepted the Adams Scholarship. In contrast, the quality difference is smaller or non-existent for marginal winners.

One may formally assess validity of LI, given smoothness. LI implies that any derivatives of the RD treatment effects evaluated at the RD cutoff are zero. I focus on the first derivative $\frac{\partial}{\partial z} E\left[\beta_i | z_i = z, \psi_i = C(z)\right]|_{z=z_0}$. Dong and Lewbel (2015) show that $E\left[\beta_i | z_i = z, \psi_i = C\right]|_{z=z_0}$, referred to as treatment effect derivative (TED), can be nonparametrically identified and estimated. One can then test significance of the estimated TED to evaluate LI. The following theorem summarizes this observation.

THEOREM 2 Assume that the conditional means and probabilities in Assumption LS1 are continuously differentiable.[5] LI implies a testable restriction $\frac{\partial}{\partial z} E\left[\beta_i | z_i = z, \psi_i = C(z)\right]|_{z=z_0} = 0$.

Consider again the local linear approximation $y_i = \alpha_0 + \alpha_1 (z_i - z_0) + \tau_0 x_i + \tau_1 (z_i - z_0) x_i + e_i$. TED is captured by τ_1. In fuzzy designs, one can make inference on τ_1 by local two stage least squares (2SLS) estimation for a given bandwidth, using $d_i := 1(z_i - z_0 \geq 0)$ and $d_i (z_i - z_0)$ as excluded IVs.[6,7]

The following briefly presents an empirical application. Details of the data and further discussions are provided in the supplemental online Appendix. I consider how placement on academic probation affects the dropout probability in college, focusing on treatment effect heterogeneity in the pre-treatment GPA, the running variable.

I use confidential data on the entire first-time freshmen cohorts between 1992 and 2002 from a large Texas university. The outcome of interest y_i is a binary indicator for dropout, which is 1 if a student dropped out of college, and 0 otherwise. The running variable z_i is the first semester GPA. The treatment x_i is an indicator for ever being placed on academic probation, which is set to be 1 if a student's cumulative or semester GPA was below the school-wide cutoff 2.0, and 0 otherwise.

[5]As noted in Dong and Lewbel (2015), although identification of RD LATE requires only continuity conditions, the standard estimation (via, e.g., local linear regressions) and inference of RD LATE requires continuous differentiability conditions.

[6]For sharp designs, $x_i = d_i$, this reduces to estimating the slope change at the RD cutoff.

[7]Instead of using a uniform kernel, one can apply other kernel functions to estimate kernel weighted local 2SLS.

Placement on academic probation is shown to have a small, positive, yet insignificant impact on the college dropout rate right at the first-semester probation threshold (see Table A4 in the online Appendix). However, estimates of TED in Table 1 below reveal that the impacts increase significantly as a student's GPA moves marginally below the cutoff. For example, the estimated TED is -0.568 using a triangular kernel and the optimal bandwidth proposed by Calonico, Cattaneo and Titiunik (2014). The estimate is statistically significant at the 1% level. That is, when the pre-treatment GPA decreases by 0.1, the probability of dropping out of college for a student on probation increases by 5.68%. The estimated TED is large in magnitude, compared with the estimated RD LATE, which is 6.8% with a standard error 0.053. Estimates using alternative bandwidths and kernel choices show similar results. These results strongly suggest that the probation effect depends on how far one is from the probation threshold and hence LI is violated in this case.

Table 1 The derivative of the Impact of Academic Probation on Dropout Rates

	(1)	(2)	(3)	(4)	(5)	(6)
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
Kernel	Triangular	Triangular	Triangular	Uniform	Uniform	Uniform

Note: All estimates are based on local quadratic regressions; Columns (1) and (3) use the CCT optimal bandwidth, Columns (2) and (4) use the optimal bandwidth proposed by Imbens and Kalyanaraman (2014), and Columns (3) and (6) use the cross validation bandwidth proposed by Ludwig and Miller (2007); 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 check for smoothness, I estimate the jumps and kinks in the density of the running variable and the conditional means of covariates at the probation threshold. Covariates investigated include dummies for male, Black, and Hispanic, as well as an indicator for being ranked among the top 25% of the high school class. These results are reported in Table A5 in the online Appendix. 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 LI does not hold in this case.

Additionally, LI implies that (un)observable covariates that determine potential treatment status, u_i, are independent of the running variable near z_0, so one can test independence of covariate means from the running variable near z_0 to check for LI. This is similar to testing smoothness of the conditional means of covariates to evaluate the smoothness conditions. This kind of tests are essentially falsification tests.

LI might be restrictive in some empirical scenarios. However, when LI holds, there are practical implications one may explore. First, the estimated RD local effects have stronger external validity. Second, for sharp designs LI implies no slope changes at the cutoff, so with the standard local polynomial estimation, a wider range of bandwidth choices can be applied. In particular, one does not need to under-smooth in order to shrink the bias to zero to have correct inference. In practice, the robust bias-corrected inference of Calonico, Cattaneo, and Tituinik (2014) is proposed by taking into account the slope change so that inference remains valid even for “too large” bandwidth choices. That is, one can enjoy robust estimation that is less sensitive to bandwidth choice and bias correction.

4 Conclusion

Either local independence (LI) or a smooth parallel of it, local smoothness (LS), can be used to identify RD LATE. Both have readily testable implications. This paper discusses these two alternative assumptions and note that LS is closely related to a weak and empirically plausible behaviorial assumption, in the spirit of Lee (2008). However, when LI holds, there are some practical implications one may explore. The discussion also provides theoretical ground for McCrary’s (2008) density test in fuzzy RD designs.

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] 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.

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

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

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

[6] 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.

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

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

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

[10] 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.

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