Supplemental Appendix to "Alternative Assumptions to Identify LATE in Fuzzy Regression Discontinuity Designs"

Yingying Dong

University of California Irvine*

February 2018

Abstract

This document provides proofs of the theorems and details of the empirical application and illustration in "Alternative Assumptions to Identify LATE in Fuzzy Regression Discontinuity Designs."

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

1 Proofs

Proof that LS2 implies LS1: 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 $f_{\cdot|\cdot}(\cdot|\cdot)$ denote the unconditional and conditional probability density or mass functions, respectively. In particular, let $f_{\mathbf{w}|z}(\mathbf{w}|z)$ denote the mixed joint density of $\mathbf{w}_i$ conditional on $z_i = z$, i.e., $f_{\mathbf{w}|z}(\mathbf{w}|z) = f_{y_0,y_1,z|\psi}(y_0, y_1, z|\psi_i = \psi)\Pr(\psi_i = \psi)/f_z(z)$.

Assumption LS2 states that $f_{z|\mathbf{w}}(z|\mathbf{w})$ is continuous in z, and $f_z(z)$ is continuous and strictly positive at $z = z_0$. By Bayes' Rule, $f_{\mathbf{w}|z}(\mathbf{w}|z) = f_{z|\mathbf{w}}(z|\mathbf{w})\,f_{\mathbf{w}}(\mathbf{w})/f_z(z)$, so $f_{\mathbf{w}|z}(\mathbf{w}|z)$ is continuous

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

in z at $z = z_0$. By definition $\mathbf{w}_i := (y_{0i}, y_{1i}, \psi_i)$, then probability of each type of individual $\Pr(\psi_i = \psi|z_i = z) = \int_{\Omega_1}\int_{\Omega_0} f_{\mathbf{w}|z}(\mathbf{w}|z)dy_0dy_1$ 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, $f_{y_0,y_1|\psi,z}(y_0, y_1|\psi_i = \psi, z_i = z) = f_{\mathbf{w}|z}(\mathbf{w}|z)/Pr(\psi_i = \psi|z_i = z)$ for $\psi_i = \psi \in \{A, N, C\}$. Both $f_{\mathbf{w}|z}(\mathbf{w}|z)$ and $\Pr(\psi_i = \psi|z_i = z)$ are continuous in z at $z = z_0$, so $f_{y_0,y_1|\psi,z}(y_0, y_1|\psi_i = \psi, z_i = z)$ 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 $E(y_{ti}|\psi_i = \psi, z_i = z)$ for $t = 0, 1$,and $\psi_i = \psi \in \{A, N, C\}$ are continuous in z at $z = z_0$.

Proof of Theorem 1: Given LS1, monotonicity, and the definitions of individual types, we have

$$
\begin{aligned}
y^+ - y^- &= \lim_{z\downarrow z_0} E\left[\alpha_i + \beta_i x_i|z_i = z\right] - \lim_{z\uparrow z_0} E\left[\alpha_i + \beta_i x_i|z_i = z\right] \\
&= \lim_{z\downarrow z_0} E\left[\beta_i x_i|z_i = z\right] - \lim_{z\uparrow z_0} E\left[\beta_i x_i|z_i = z\right] \\
&= \lim_{z\downarrow z_0} \left\{E\left[\beta_i|z_i = z, x_{1i} = 1\right]\Pr\left[x_{1i} = 1|z_i = z\right]\right\} \\
&\quad - \lim_{z\uparrow z_0} E\left\{\left[\beta_i|z_i = z, x_{0i} = 1\right]\Pr\left[x_{0i} = 1|z_i = z\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= E\left[\alpha_i|z_i = z_0\right] + E\left[\beta_i|z_i = z_0, \psi_i = C\right]\Pr\left[\psi_i = C|z_i = z_0\right] \\
&\quad + E\left[\beta_i|z_i = z_0, \psi_i = A\right]\Pr\left[\psi_i = A|z_i = z_0\right] \\
&\quad - E\left[\alpha_i|z_i = z_0\right] + E\left[\beta_i|z_i = z_0, \psi_i = A\right]\Pr\left[\psi_i = A|z_i = z_0\right] \\
&= E\left[\beta_i|z_i = z_0, \psi_i = C\right]\Pr\left[\psi_i = C|z_i = z_0\right].
\end{aligned}
$$

In addition,

$$\begin{aligned} x^{+}-x^{-} &= \lim_{z\downarrow z_0} E\left[x_i|z_i = z\right] - \lim_{z\uparrow z_0} E\left[x_i|z_i = z\right] \\ &= E\left[x_{1i}|z_i = z_0\right] - E\left[x_{0i}|z_i = z_0\right] \\ &= E\left[x_{1i} > x_{0i}|z_i = z_0\right] \\ &= \Pr\left[\psi_i = C|z_i = z_0\right]. \end{aligned}$$

By A2, $x^{+} - x^{-} \neq 0$, then

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

Proof of Theorem 2: LI implies $\frac{\partial}{\partial z} E\left[y_{1i} - y_{0i}|z_i = z, \psi_i = C(z)\right]|_{z=z_0} = 0$. Define $y^{+\prime} \equiv \lim_{z\downarrow z_0} \frac{\partial}{\partial z} E\left[y_i|z_i = z\right]$ and $y^{-\prime} \equiv \lim_{z\uparrow z_0} \frac{\partial}{\partial z} E\left[y_i|z_i = z\right]$, and similarly define $x^{+\prime}$ and $x^{-\prime}$ by replacing y_i with x_i. The following shows that under continous differentiability of the conditional means and probabilities in Assumption LS1, one can identify $\frac{\partial}{\partial z} E\left[y_{1i} - y_{0i}|z_i = z, \psi_i = C(z)\right]|_{z=z_0}$ and hence test $\frac{\partial}{\partial z} E\left[y_{1i} - y_{0i}|z_i = z, \psi_i = C(z)\right]|_{z=z_0} = 0$.

$$
\begin{aligned}
& \frac{\partial}{\partial z} E\left[y_{1i} - y_{0i} | z_i = z, \psi_i = C(z)\right]|_{z=z_0} \\
= \; & \lim_{z \downarrow z_0} \frac{\partial}{\partial z} E\left[y_i | z_i = z, \psi_i = C(z)\right] - \lim_{z \uparrow z_0} \frac{\partial}{\partial z} E\left[y_i | z_i = z, \psi_i = C(z)\right] \\
= \; & \lim_{z \downarrow z_0} \frac{\partial}{\partial z} \left\{ E\left[y_i 1\left(\psi_i = C(z)\right) | z_i = z\right] / \Pr\left(\psi_i = C(z) | z_i = z\right) \right\} \\
& - \lim_{z \uparrow z_0} \frac{\partial}{\partial z} \left\{ E\left[y_i 1\left(\psi_i = C(z)\right) | z_i = z\right] / \Pr\left(\psi_i = C(z) | z_i = z\right) \right\} \\
= \; & \lim_{z \downarrow z_0} \frac{\partial}{\partial z} \left\{ E\left[y_i | z_i = z\right] / \Pr\left(\psi_i = C(z) | z_i = z\right) \right\} - \lim_{z \downarrow z_0} \frac{\partial}{\partial z} \left\{ E\left[y_i | z_i = z\right] / \Pr\left(\psi_i = C(z) | z_i = z\right) \right\} \\
= \; & \lim_{z \downarrow z_0} \left\{ \frac{\frac{\partial}{\partial z} E\left[y_i | z_i = z\right]}{\Pr\left(\psi_i = C(z) | z_i = z\right)} - E\left[y_i | z_i = z\right] \frac{\frac{\partial}{\partial z} \Pr\left(\psi_i = C(z) | z_i = z\right)}{\left(\Pr\left(\psi_i = C(z) | z_i = z\right)\right)^2} \right\} \\
& - \lim_{z \uparrow z_0} \left\{ \frac{\frac{\partial}{\partial z} E\left[y_i | z_i = z\right]}{\Pr\left(\psi_i = C(z) | z_i = z\right)} - E\left[y_i | z_i = z\right] \frac{\frac{\partial}{\partial z} \Pr\left(\psi_i = C(z) | z_i = z\right)}{\left(\Pr\left(\psi_i = C(z) | z_i = z\right)\right)^2} \right\} \\
= \; & \frac{\lim_{z \downarrow z_0} \frac{\partial}{\partial z} E\left[y_i | z_i = z\right] - \lim_{z \uparrow z_0} \frac{\partial}{\partial z} E\left[y_i | z_i = z\right]}{\Pr\left(\psi_i = C(z) | z_i = z\right)} \\
& - \frac{\lim_{z \downarrow z_0} E\left[y_i | z_i = z\right] - \lim_{z \uparrow z_0} E\left[y_i | z_i = z\right]}{\Pr\left(\psi_i = C(z) | z_i = z\right)} \left\{ \frac{\frac{\partial}{\partial z} \Pr\left(\psi_i = C(z) | z_i = z\right)}{\Pr\left(\psi_i = C(z) | z_i = z\right)} \right\} \\
= \; & \frac{1}{x^+ - x^-} \left[y^{+\prime} - y^{-\prime} + \frac{y^+ - y^-}{x^+ - x^-} \left(x^{+\prime} - x^{-\prime} \right) \right],
\end{aligned}
$$

where the first equality follows from the smoothness conditions and the fact that for compliers $y_i = y_{1i}$ for $z_i = z \geq z_0$ and $y_i = y_{0i}$ for $z_i = z < z_0$, the second equality follows from the law of total expectation, the third equality follows from the fact that by the smoothness conditions, $\frac{\partial}{\partial z} \frac{E\left[y_i 1\left(\psi_i = \psi(z)\right) | z_i = z\right]}{\Pr\left(\psi_i = C(z) | z_i = z\right)} - \frac{\partial}{\partial z} \frac{E\left[y_i 1\left(\psi_i = \psi(z)\right) | z_i = z\right]}{\Pr\left(\psi_i = C(z) | z_i = z\right)} = 0$ for $\psi = A$ and N, the forth equality follows from the chain rule, and the last equality follows from $\Pr\left(\psi_i = C(z) | z_i = z_0\right) = x^+ - x^-$ and by the smoothness conditions, $\frac{\partial}{\partial z} \Pr\left(\psi_i = C(z) | z_i = z\right)|_{z=z_0} = \frac{\partial}{\partial z} E\left[x_{1i} - x_{0i}\right]|_{z_i = z_0} = \lim_{z \downarrow z_0} E\left[x_i | z_i = z\right] - \lim_{z \downarrow z_0} E\left[x_i | z_i = z\right] = x^{+\prime} - x^{-\prime}$.

Relating τ_0 and τ_1 in $y_i = \alpha_0 + \alpha_1 (z_i - z_0) + \tau_0 x_i + \tau_1 (z_i - z_0) x_i + e_i$ with the corresponding parameters in the reduced-form outcome and treatment equation, where all equations are expressed as local linear regressions, it is easy to verify that $\frac{1}{x^+ - x^-} \left[y^{+\prime} - y^{-\prime} + \frac{y^+ - y^-}{x^+ - x^-} \left(x^{+\prime} - x^{-\prime} \right) \right]$ corresponds to the local two stage least squares (2SLS) estimand for τ_1, using $d_i := 1(z_i - z_0 \geq 0)$ and $d_i(z_i - z_0)$ as excluded IVs for x_i.

2 Empirical Applications

This section provides additional details for investigating treatment effect heterogeneity (TED) in the RD design of Lee (2008) and this paper's fuzzy RD design application.

2.1 Lee (2008)

The analysis draws on the same data as those used in Lee (2008) and Lee and Lemieux (2010).[1] The sample consists of 6,558 elections from 1946 to 1998. The treatment in this case is an indicator for the Democratic Party being the incumbent party. The running variable is the Democratic Party's winning margin in election t. The outcome is whether a democratic candidate won in election $t+1$.

Following Lee and Lemieux (2010), I use local linear regressions to estimate the local causal effect of being an incumbent party. Analogous to using local linear regressions to estimate means at a boundary point, local quadratic regressions may be appropriate for estimating slopes. See, e.g., discussion in Porter (2003) and Calonico, Cattaneo, and Titiunik (2014). I therefore use local quadratic regressions to estimate the derivative of the RD treatment effect (corresponding to the slope change) at the RD cutoff.

For kernel choices, I adopt both the boundary optimal triangular kernel (Fan and Gijbels, 1996), and 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.

[1] 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).

Table A1 Sharp RD Estimates of the Incumbency Advantage

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

Note: All RD LATE estimates are based on bias-corrected robust inference proposed by Calonico, Cattaneo and Titiunik (CCT, 2014), using local linear 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); 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.

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).

Table A2 The Derivative of the Incumbency Advantage

	(1)	(2)	(3)	(4)	(5)	(6)
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
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.

Table A1 and Table A2 report estimates of the treatment effect and the treatment effect derivative, respectively. The treatment effect derivative 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.

Consistent with estimates in Lee (2008) and Lee and Lemieux (2010), the average incumbency effect is estimated to be between 0.364 and 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 treatment effect derivative is between 1.143 and 1.349, so given 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 and their derivatives are all statistically significant. Therefore, LI is not likely to hold in this case. In particular, the incumbency advantage depends on the incumbent party's winning margin.

Table A3 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 A3 reports the estimated jumps or kinks at the RD cutoff in the conditional mean of an important covariate, the Democratic vote share from the previous election. Also reported are the estimated jumps and kinks in the empirical density of the running variable at the RD cutoff. I only report 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. None of the estimated jumps and kinks are statistically significant. These results support the plausibility of the smoothness assumption, and hence validity of the RD design, even though LI likely does not hold by the results in Table A2.

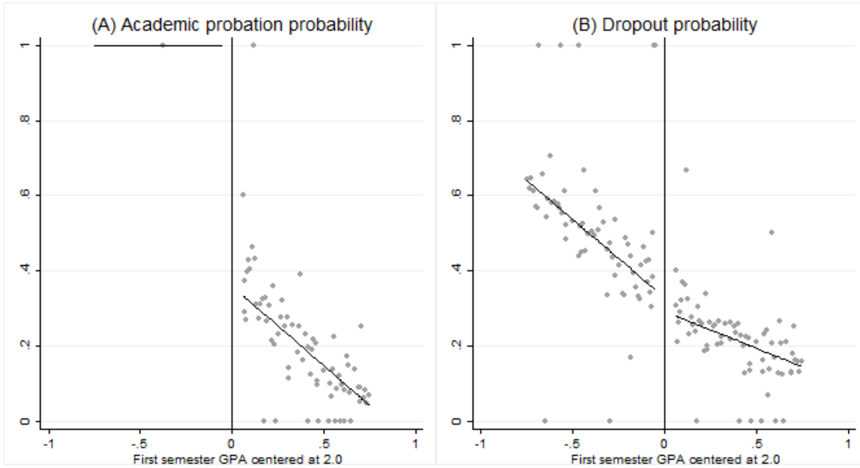

Figure 1: Academic Probation and Dropout Rates against First-semester GPA

2.2 Academic Probation and Dropout Rate

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.

I consider a fuzzy RD design based on the probation rule in college and evaluate the impact of academic probation on the dropout probability. I further evaluate how the discouragement effect of academic probation depends on the running variable, a student's pre-treatment GPA.

The data used here represent the entire population of the first-time freshmen cohorts between 1992 and 2002 from a large Texas university collected under the Texas Higher Education Opportunity Project (THEOP). The outcome of interest y_i is a binary indicator for dropout, which is 1 if a student drop out of college and 0 otherwise. The running variable z_i is the first semester GPA. The treatment x_i is the indicator for ever being placed on academic probation. [2] The total sample size is 64,310.

[2]The actual probation status is not observed in the data. Here I define treatment to be 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.' In

Figure 3 presents the probation probability and the dropout rate conditional on the first semester GPA. For those whose first semester GPAs fall below the probation threshold, the probation probability is 1 by construction; for those whose first semester GPAs fall marginally above, there is still an over 60% chance for them to be placed on probation later. Note that a dramatic slope change is present in the probation probability at the 2.0 cutoff, indicating that the fraction of 'compliers' largely depends on the running variable. In the mean time, the dropout rate also shows a discernible slope change the 2.0 cutoff, even though there is only a small jump right at the cutoff.

Table A1 Fuzzy RD Estimates of the Impact of Academic Probation on Dropout Rates

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

Note: All RD LATE estimates are based on bias-corrected robust inference proposed by Calonico, Cattaneo and Titiunik (CCT, 2014), using local linear 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); Standard errors are in parentheses; * significant at the 10% level, ** significant at the 5% level, ***significant at the 1% level.

Table A4 reports estimates of RD LATE. Placement on academic probation is shown to have a small yet insignificant impact on the college dropout rate right at the first-semester probation threshold. However, Table 1 in the main text shows that the estimated TED is large and negative, so the impacts of being placed on probation largely depend on a student's pre-treatment GPA. The further away a sudent falls below the required GPA threshold, the greater the impacts are on their dropout probabilities.

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.

Table A5 reports the estimated jumps and kinks in the density of the running variable and the conditional means of covariates at the probation threshold. 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.

Table A5 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 estimates are reported; Robust standard errors are in parentheses. The bin size used to generate the empirical density is .006.

References

[1] Calonico, S., M.D. Cattaneo and R.Titiunik, (2014), “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” Econometrica, 82(6), 2295–2326.

[2] 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.

[3] 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).

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

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

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

[7] 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.

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