Applying Regression Discontinuity Designs to Measure Treatment Effects

Petra E. Todd[1]

[1]University of Pennsylvania, Dept of Economics

March, 2012

Overview of RD design

- Goal is to evaluate causal impacts of an intervention
- Assignment to treatment is determined in part by the value of an observed covariate lying on either side of a fixed threshold (cut-off)
- Design first introduced by Thistlewaite and Campbell (1960) to evaluate the effect of National Merit awards on career aspirations of award recipients.
- Analyzed by Goldberger (1972) in the context of evaluating education interventions and Berk and Rauma (1983) in analzying effect of an unemployment benefit program on recidivism rates.

- Many studies implicitly rely on nonlinearities or discontinuities in the assignment rule (Black (1996) and Angrist and Krueger (1991)).
- Since 1990's, there has been a large number of studies in economics and other fields applying and extending RD methods (will discuss many examples later)
- New theoretical advances in interpretation and estimation

Potential Outcomes Framework

- Potential outcomes associated with treated and untreated states

$$Y_i(0), Y_i(1)$$

- Framework laid out in Fisher (1951), Roy (1951), Quandt (1972), Rubin (1978)
- Interest usually focuses on

$$Y_i(1) - Y_i(0)$$

- Let $W_i = 1$ if unit i exposed to treatment, else $W_i = 0$.
- Observed outcome

$$\begin{aligned} Y_i &= (1 - W_i) Y_i(0) + W_i Y_i(1) \\ &= Y_i(0) + W_i (Y_i(1) - Y_i(0)) \end{aligned}$$

Assignment to Treatment

- Let (X_i, Z_i) be a vector of covariates or pretreatment variables known not to be affected by treatment (e.g. pre-test score, age)
- Assignment to treatment is determined either completely or partly by the value of X_i being on either side of a fixed threshold.
- $Y_i(0)$ or $Y_i(1)$ may also be associated with X_i, but the dependence is assumed to be smooth
- Discontinuities in the conditional distribution of Y_i (or in its conditional expectation) are attributed to a causal effect of treatment.

Two Types of Designs

- *Sharp Regression Discontinuity (SRD) Design*

$$W_i = 1\{X_i \geq c\} \tag{1}$$

  All individuals with covariates of c or greater are assigned to treatment.

- We can use the discontinuity in the conditional expectation of the outcome given the covariate to uncover an average causal effect of treatment

$$\tau_{SRD} = lim_{x \downarrow c} E(Y_i | X_i = x) - lim_{x \uparrow c} E(Y_i | X_i = x) \tag{2}$$

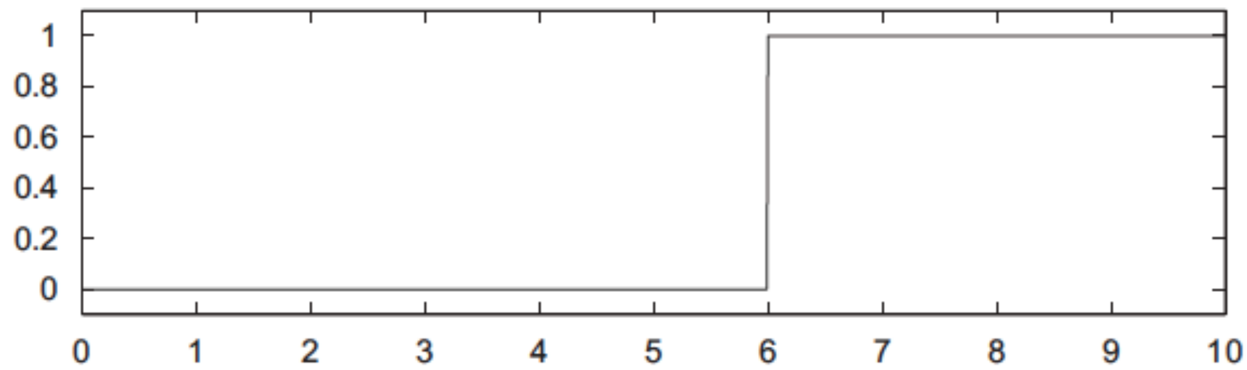


Fig. 1. Assignment probabilities (SRD).

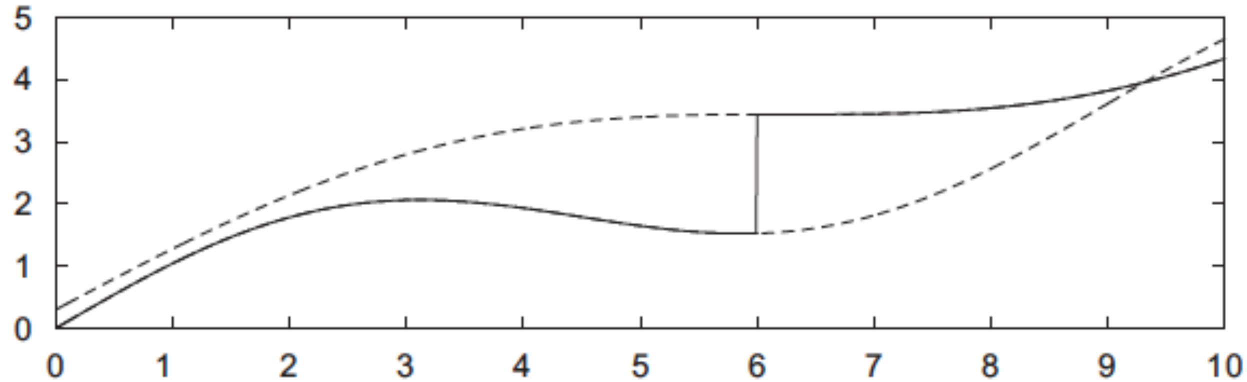


Fig. 2. Potential and observed outcome regression functions.

Two Types of Designs

- *Fuzzy Regression-Discontinuity (FRD) Design*
- Prob of receiving treatment need not change from 0 to 1 at the threshold, but there there is a discontinuous jump in the probability, so that

$$lim_{x \downarrow c} E(W_i | X_i = x) - lim_{x \uparrow c} E(W_i | X_i = x) \neq 0$$

$$or, equivalently,$$

$$lim_{x \downarrow c} Pr(W_i = 1 | X_i = x) - lim_{x \uparrow c} Pr(W_i = 1 | X_i = x) \neq 0$$

- Treatment effect can be obtained by the ratio

$$\tau_{FRD} = \frac{lim_{x \downarrow c} E(Y_i | X_i = x) - lim_{x \uparrow c} E(Y_i | X_i = x)}{lim_{x \downarrow c} E(W_i | X_i = x) - lim_{x \uparrow c} E(W_i | X_i = x)} \qquad (3)$$

Why?

Assume constant treatment effect τ_{FRD}.

$$\mathrm{Y}_i = Y_i(0) + W_i[Y_i(1) - Y_i(0)]$$

$$\mathrm{Y}_i = Y_i(0) + W_i \tau_{FRD}$$

$$\lim_{x \downarrow c} E(Y_i|X_i = x) =$$

$$\lim_{x \downarrow c} E(Y_i(0)|X_i = x) + lim_{x \downarrow c} E(W_i|X_i = x)\tau_{FRD}$$

$$\lim_{x \uparrow c} E(Y_i|X_i = x) =$$

$$\lim_{x \downarrow c} E(Y_i(0)|X_i = x) + lim_{x \uparrow c} E(W_i|X_i = x)\tau_{FRD}$$

Take the difference, use the fact that $E(Y_i(0)|X_i = x)$ is continuous at $X_i = c$ and solve for τ_{FRD}.

$$\lim_{x\downarrow c} E(Y_i|X_i = x) - lim_{x\uparrow c} E(Y_i|X_i = x) =$$
$$\lim_{x\downarrow c} E(W_i|X_i = x) - lim_{x\uparrow c} E(W_i|X_i = x)\tau_{FRD}$$

$$\tau_{FRD} = \frac{lim_{x\downarrow c} E(Y_i|X_i = x) - lim_{x\uparrow c} E(Y_i|X_i = x)}{lim_{x\downarrow c} E(W_i|X_i = x) - lim_{x\uparrow c} E(W_i|X_i = x)}$$

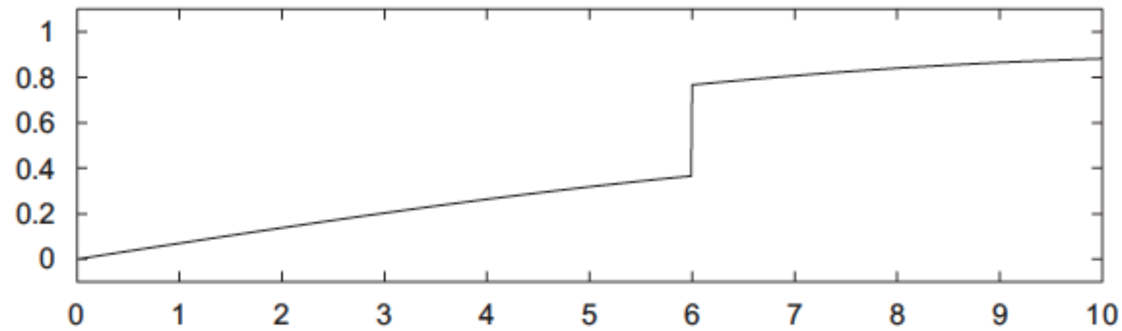


Fig. 3. Assignment probabilities (FRD).

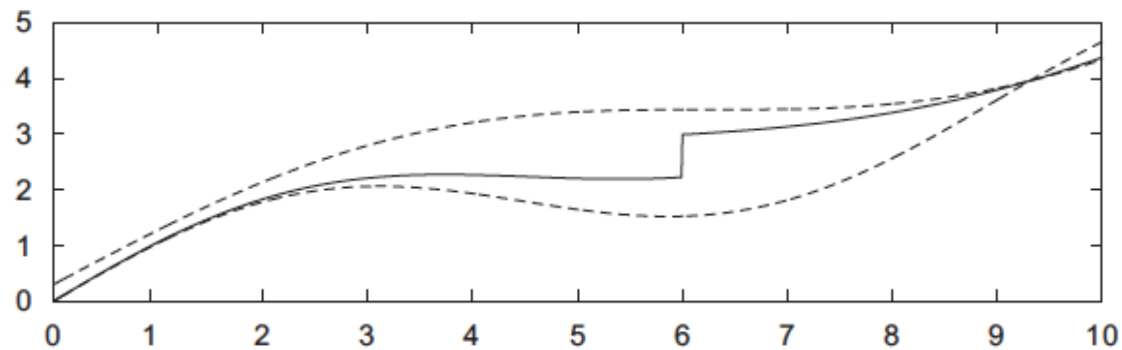


Fig. 4. Potential and observed outcome regression (FRD).

Interpretation of FRD when treatment response is heterogeneous

- Assume that treatment effect varies by unit τ_{FRD} random
- Let $W_i(x)$ be the potential treatment status given cut-off point x, for x in a neighborhood of c.
- $W_i(x) = 1$ if unit i would take treatment if cut-off equals x
- Assume monotonicity: $W_i(x)$ is nonincreasing in x at $x = c$

Define compliance status

- *Compliers:* have

$$lim_{x\downarrow X_i} W_i(X_i) = 0, lim_{x\uparrow X_i} W_i(X_i) = 1, \tag{4}$$

would get treatment if cut-off X_i or below, would not get treatment otherwise

- *Nevertakers:* do not get treatment either way

$$lim_{x\downarrow X_i} W_i(X_i) = 0, lim_{x\uparrow X_i} W_i(X_i) = 0 \tag{5}$$

- *Always takers:* get treatment either way

$$lim_{x\downarrow X_i} W_i(X_i) = 1, lim_{x\uparrow X_i} W_i(X_i) = 1 \tag{6}$$

For example, consider a program that assigns children with a pre-test score below a threshold to some remedial intervention (e.g. a summer reading program).

- *Compliers* are children who participate in the program only if their test score is below the threshold and not otherwise. They comply with their assigned treatment status.
- *Always-takers* are children who manage to receive the intervention regardless (e.g. parents request that they attend the program)
- *Never-takers* are children who do not attend the program even if assigned to it.

Interpretation of τ_{FRD}

In that case, τ_{FRD} gives the average treatment effect for compliers. (shown in Hahn, Todd and Van der Klaauw, 2001, building on insights of Angrist and Imbens, 1994, about LATE estimators).

Another example of FRD Design

- Van Der Klaauw (2002)
- Studies effect of financial aid on college admissions
- Association is ambiguous. More generous financial aid offers make students more likely to attend, but those students are also likely to have more generous offers from other places.
- x_i - numerical score assigned to college application based on the objective part of the application (SAT scores, grades)

$$\begin{aligned} G_i &= 1 \text{ if } 0 \leq X_i < c_1 \\ G_i &= 2 \text{ if } c_1 \leq X_i < c_2 \\ &\vdots \\ G_i &= L \text{ if } c_{L-1} \leq X_i \end{aligned}$$

- Fuzzy design - numerical score not the only factor. essays, rec letters also important.

Comparison of RD Approach with a Matching Approach

- Matching assumes

$$Y(0), Y(1) \perp\!\!\!\perp W | X \tag{8}$$

- In that case, treatment effect can be obtained by comparing people with same x values who did and did not receive treatment

$$E(Y(1) - Y(0) | X = x) = E(Y | W = 1, X = c) - E(Y | W = 0, X = c)$$

- This approach would not exploit the jump in the probability of assignment at the discontinuity point
- It could not be implemented with a sharp design, where there is no overlap.
- Treated units with $x_i = c$ include both compliers and alwaystakers.
- Unconfoundedness is based on units being comparable if covariates are similar.

External and Internal Validity of RD Designs

- When treatment response is heterogeneous, RD approach provides estimates for subpopulation with x_i=c.
- If FRD and treatment effect heterogeneous, then effect is further restricted to the effect on compliers only (and compliers cannnot be identified in the data)
- The RD design has high internal validity (valid with the population studied), but potentially limited external validity (limited application to outside populations)

Estimation

- For sharp design, need estimators of two limits

$$\tau_{SRD} = lim_{x \downarrow c} E(Y_i | X_i = x) - lim_{x \uparrow c} E(Y_i | X_i = x) \tag{10}$$

- Could estimate each limit by kernel regression

$$\hat{\mu}_l(x) = \frac{\sum_{x_i < c} Y_i K(\frac{x_i - x}{h})}{\sum_{x_i < c} K(\frac{x_i - x}{h})} \tag{11}$$

$$\hat{\mu}_r(x) = \frac{\sum_{x_i \geq c} Y_i K(\frac{x_i - x}{h})}{\sum_{x_i \geq c} K(\frac{x_i - x}{h})} \tag{12}$$

With rectangular (uniform) kernel

- $K(u) = 1/2$ for $-1 \leq u \leq 1$, $= 0$ elsewhere

$$\tau_{SRD} = \frac{\sum_{i=1}^{n} Y_i 1(c \leq X_i \leq c+h)}{\sum_{i=1}^{n} 1(c \leq X_i \leq c+h)} - \frac{\sum_{i=1}^{n} Y_i 1(c-h \leq X_i < c)}{\sum_{i=1}^{n} 1(c-h \leq X_i < c)} \quad (13)$$

- Simple kernel regression suffers from boundary bias problem - slower rate of convergence at boundary points than in interior points.

Boundary bias

$$plim\hat{\mu}_r(c) = \frac{\int_c^{c+h} \mu(x) f(x) dx}{\int_c^{c+h} f(x) dx} = \mu_{r(c)} + lim_{x \downarrow c} \frac{\partial}{\partial x} \mu(x) \frac{h}{2} + O(h^2)$$

- bias is linear in the bandwidth, h.
- At interior points, bias is usually of order h^2. Convergence of bias to 0 is slower at boundary points.

Recommended alternative: Local linear regression

- Fan and Gijbels (1996) discuss local linear regression methods that have the same order of convergence at boundary points as in interior points.
- These methods fit a regression to observations within a distance h on either side of the discontinuity point.

$$min_{\alpha_r,\beta_r} \sum_{i=1,x_i \geq c}^{n} (Y_i - \alpha_r - \beta_r(X_i - c))^2 K(\frac{X_i - c}{h}) \qquad (14)$$

- $\hat{\alpha}_r$ provides an estimator of μ_r at the point x=c
- Obtain $\hat{\alpha}_l$ similarly and then obtain treatment effect as $\hat{\alpha}_r - \hat{\alpha}_l$.
- Local linear regression has same variance as kernel regression, but faster rate of convergence of bias at boundary points.(Fan and Gijbels, 1996).

Estimation under the FRD design

Again, we need to estimate the expected value of the outcome on both sides of the discontinuity point

$$min_{\alpha_r^y,\beta_r^y} \sum_{i=1,x_i \geq c}^{n} (Y_i - \alpha_r - \beta_r (X_i - c))^2 K(\frac{X_i - c}{h}) \tag{15}$$

$$min_{\alpha_l^y,\beta_l^y} \sum_{i=1,x_i < c}^{n} (Y_i - \alpha_l - \beta_l (X_i - c))^2 K(\frac{X_i - c}{h}) \tag{16}$$

Estimation under the FRD design cont...

- In addition, estimate the expected value of the treatment indicator on both sides of the discontinuity point

$$min_{\alpha_r^w,\beta_r^w} \sum_{i=1,x_i\geq c}^{n} (W_i - \alpha_r - \beta_r(X_i - c))^2 K(\frac{X_i - c}{h}) \tag{17}$$

$$min_{\alpha_l^w,\beta_l^w} \sum_{i=1,x_i< c}^{n} (W_i - \alpha_l - \beta_l(X_i - c))^2 K(\frac{X_i - c}{h}) \tag{18}$$

- The RD treatment effect estimate under the FRD is

$$\hat{\tau} = \frac{\hat{\alpha}_r^y - \hat{\alpha}_l^y}{\hat{\alpha}_r^w - \hat{\alpha}_l^w}$$

Smoothing parameter selection

- For a given bandwidth,h, let the regression function at x be

$$\begin{aligned}\hat{\mu}(x) &= \hat{\alpha}_l(x) \text{ if } x < c \\ &= \hat{\alpha}_r(x) \text{ if } x \geq c\end{aligned}$$

- Define the cross-validation criterion as

$$CV_y(h) = \frac{1}{N}\sum_{i=1}^{n}(Y_i - \hat{\mu}_{-i}(X_i))^2$$

- where $\hat{\mu}_{-i}(X_i)$ is the so-called *leave-one-out* estimator, that leaves out the ith datapoint in calculating the estimate at X_i.

Smoothing parameter selection

- Choose h to minimize $CV_y(h)$ over a grid of possible bandwidths.

$$\mathrm{h}_{CV}^{opt} = argmin_h CV_y(h)$$

- Typically, get a cross-validation "check function"
- It is also possible to choose the bandwidth locally, focussing only on data points within close distance to the cut-off point c.
- Can choose a separate bandwidth for estimating the regression function of W_i given X_i

Assessing the variance of the estimator

- Can obtain standard errors using bootstrap methods
- Bootstrap methods are useful when it is cumbersome to obtain asymptotic standard errors.

  (i) Generate B bootstrap subsamples from the original data (can use 100% sampling with replacement.)
  (ii) Estimate treatment effect within each bootstrap sample
  (iii) The estimate of the treatment effect is based on the original data. The empirical variation across bootstrap estimates provides an estimator of the variance.

$$var(\widehat{\mu(x_i)}) = \frac{1}{B}\sum_{i=1}^{B}(\hat{\mu}_b(x_i) - \bar{\hat{\mu}}_b(x_i))^2$$

Should analysis condition on other covariates?

- There may be other covariates (Z) that are observed and that determine outcomes
- Presence of these covariates rarely changes the identification strategy. The distribution of outcomes is usually continuous in other covariates.
- Do not necessarily need to condition on other covariates.
- In practice, conditioning on Z may be helpful if we use observations on X that are not too close to c.

Graphical analysis

- Integral part of RD analysis
- RD -> treatment impact measured by a discontinuity in expected value of outcome at a particular point
- Inspect histogram estimate of avg value of the outcome around the threshold - is there evidence of a jump?
- Calculate averages that are not smoothed over the cut-off
- Also verify that there is a jump in the probability of treatment at the cut-off point
- It is also useful to inspect graphs for covariates and density of the “forcing" variable to assess credibility
- Plot average values of other covariates
- Plot the density of the forcing variable to look for evidence of manipulation (e.g. individuals know the threshold and can manipulate their value of x_i, for example, by retaking a test.)

RD Examples: Card, D., Dobkin, C.,(2008,AER)

Effect of health insurance coverage on health care utilization

- Medicare eligibility at age 65 leads to sharp changes in the health insurance coverage of the U.S. population and health care utilization increases after age 65.
- Paper compares health-related outcomes (such as different kinds of doctor visits and proceures) among people just before and just after 65, also examining results disaggregated according to group characteristics.
- It follows DiNardo and Lee (2004) and assumes the age profiles in equations (1), (2a) and (2b) are continuous polynomials with potential discontinuities in the derivatives at age 65.
- They also fit many of the models using local linear regression (as suggested by Hahn, Todd and van der Klaauw, 2001) and find results to be relatively robust.

RD Examples: Lalive (2007, J of Econometrics)

Examines whether extended benefits affect unemployment duration

- Analyzes effect of a targeted program that extends the max duration of unemployment benefits from 30 weeks to 209 weeks in Austria for individuals 50 and older living in certain geographic regions.
- There are sharp discontinuities in treatment assignment at age 50 and at the geographical border between eligible and ineligible regions.
- Uses social security data and data on unemployed.
- Two identification strategies: (i) compare individuals around the age cut-off, (ii) compare individuals across geographic borders
- Finds that job search is prolonged by 0.09 weeks per additional week of benefits for women and unemployment duration increase by 0.32 weeks per additional week of benefits for women.

RD Examples: Lalive (2007, J of Econometrics)

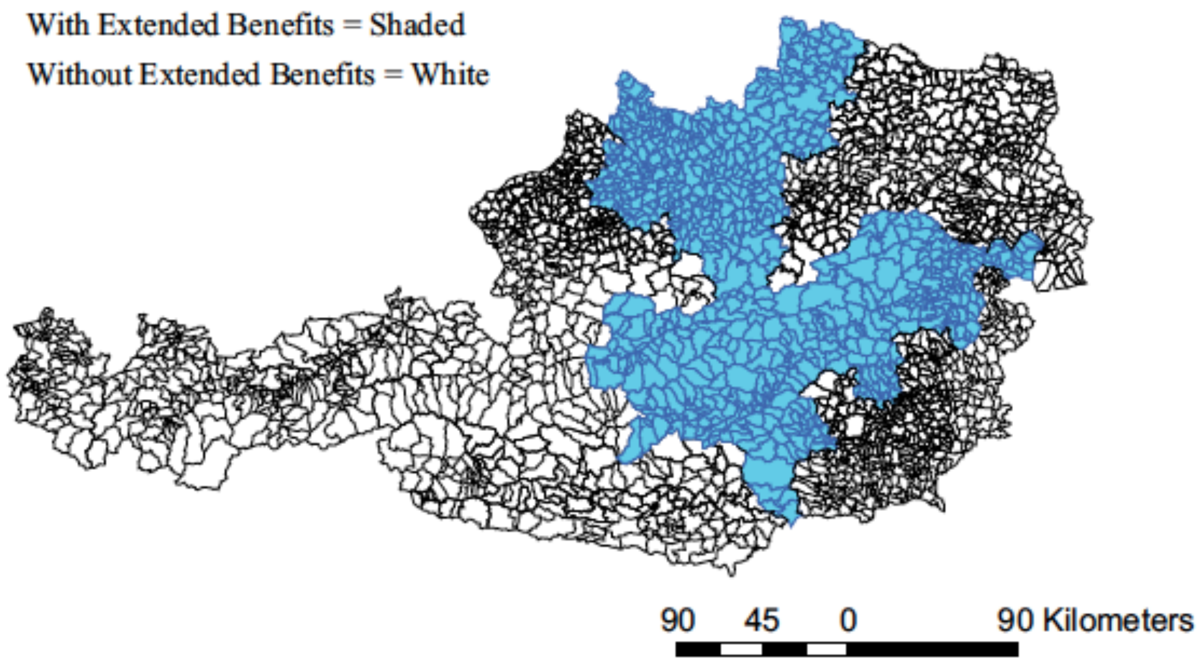


Fig. 1. Regional distribution of REBP.

RD Examples: Lalive (2007, J of Econometrics)

Table 1
Selected descriptive statistics (means)

	Column		
Living in Age bracket	(1) Treated region 50–53 years	(2) Treated region 46–49 years	(3) Control region 50–53 years
A. Men			
Age (years)	51.7	48.0	51.7
Distance to border (minutes)	28.2	27.2	−39.2
Married (share)	0.828	0.785	0.821
Construction (share)	0.481	0.492	0.600
Number of spells	4,759	4,975	8,537
B. Women			
Age (years)	51.5	48.1	51.9
Distance to border (minutes)	27.1	26.6	−37.1
Married (share)	0.780	0.696	0.721
Construction (share)	0.030	0.027	0.034
Number of spells	3,466	2,193	3,625

Source: Own calculations, based on ASSD.

RD Examples: Lalive (2007, J of Econometrics)

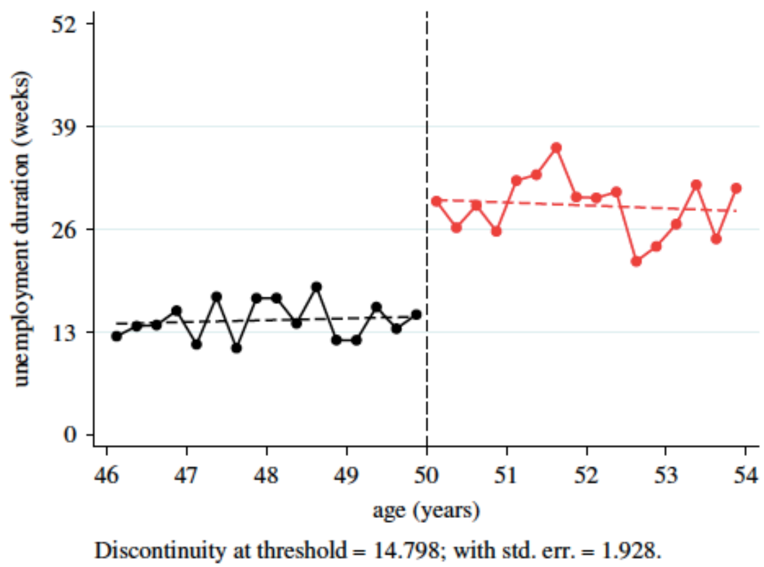


Fig. 2. The effect of REBP on unemployment duration for men: age threshold. Sample restricted to inflow into unemployment the period 8/1989 until 7/1991 (during REBP) and to individuals living in treated region.

RD Examples: Lalive (2007, J of Econometrics)

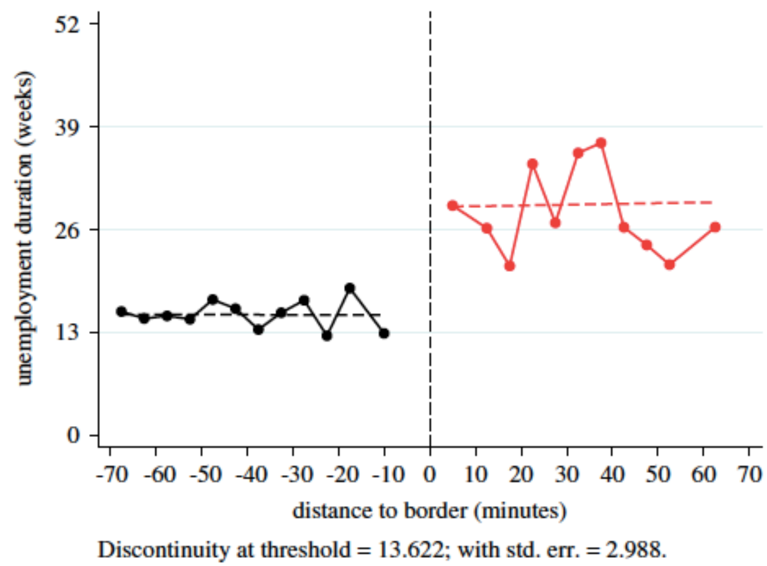


Fig. 3. The effect of REBP on unemployment duration for men: border threshold. Sample restricted to inflow into unemployment the period 8/1989 until 7/1991 (during REBP) and to individuals aged 50 years or older.

RD Examples: Lalive (2007, J of Econometrics)

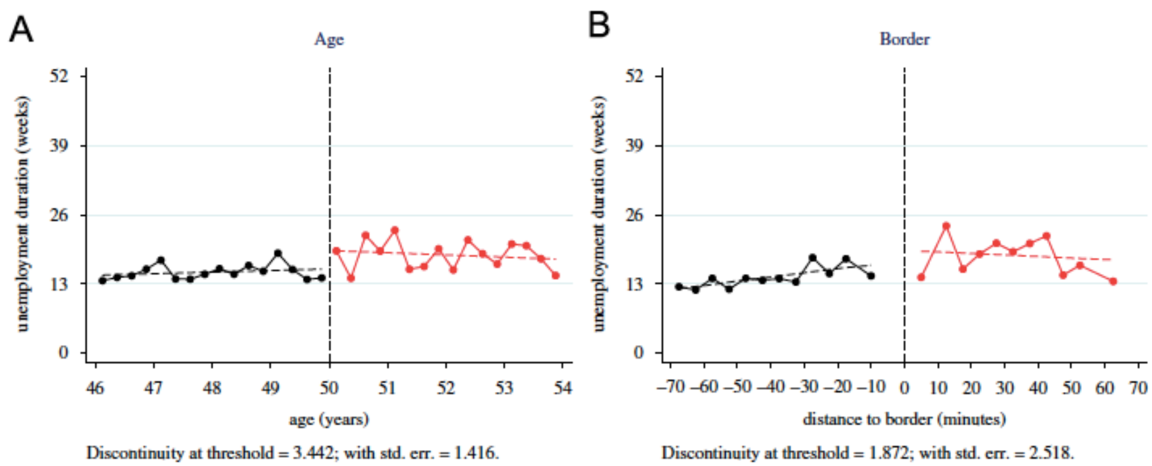


Fig. 4. The effects of age and distance before REBP: men. Sample restricted to inflow into unemployment in the period 1/1986 until 12/1987 (before REBP). Sample for age identification is restricted to treated region. Sample for border identification is restricted to individuals aged 50 years or older.

RD Examples: Jacob, B.A., Lefgren, L., (2004, Restat)

Effect of summer school and grade retention on student performance

- Analyzes the effectiveness of remedial education programs on test scores.
- In 1996, Chicago public schools instituted an accountability policy that tied summer school and grade retention to performance on standardized tests.
- Finds that summer school increased academic achievement in reading and mathematics and that these positive effects remain in the two years following the summer school program.
- Grade retention did not have negative consequences for third graders and increased short run performance.
- Retention had no impact on math performance of older students (sixth graders) and a negative impact on reading.

RD Examples: Jacob, B.A., Lefgren, L., (2004, Restat)

Effect of summer school and grade retention on student performance

- Uses administrative data from the Chicago Public School System
- 40% of third-graders and 30% of sixth graders failed to meet promotional standards.
- 3% of students who scored below the cut-off received waivers from summer school, so design was fuzzy.

Figure 1.—Student Progress Under the Chicago Accountability Policy

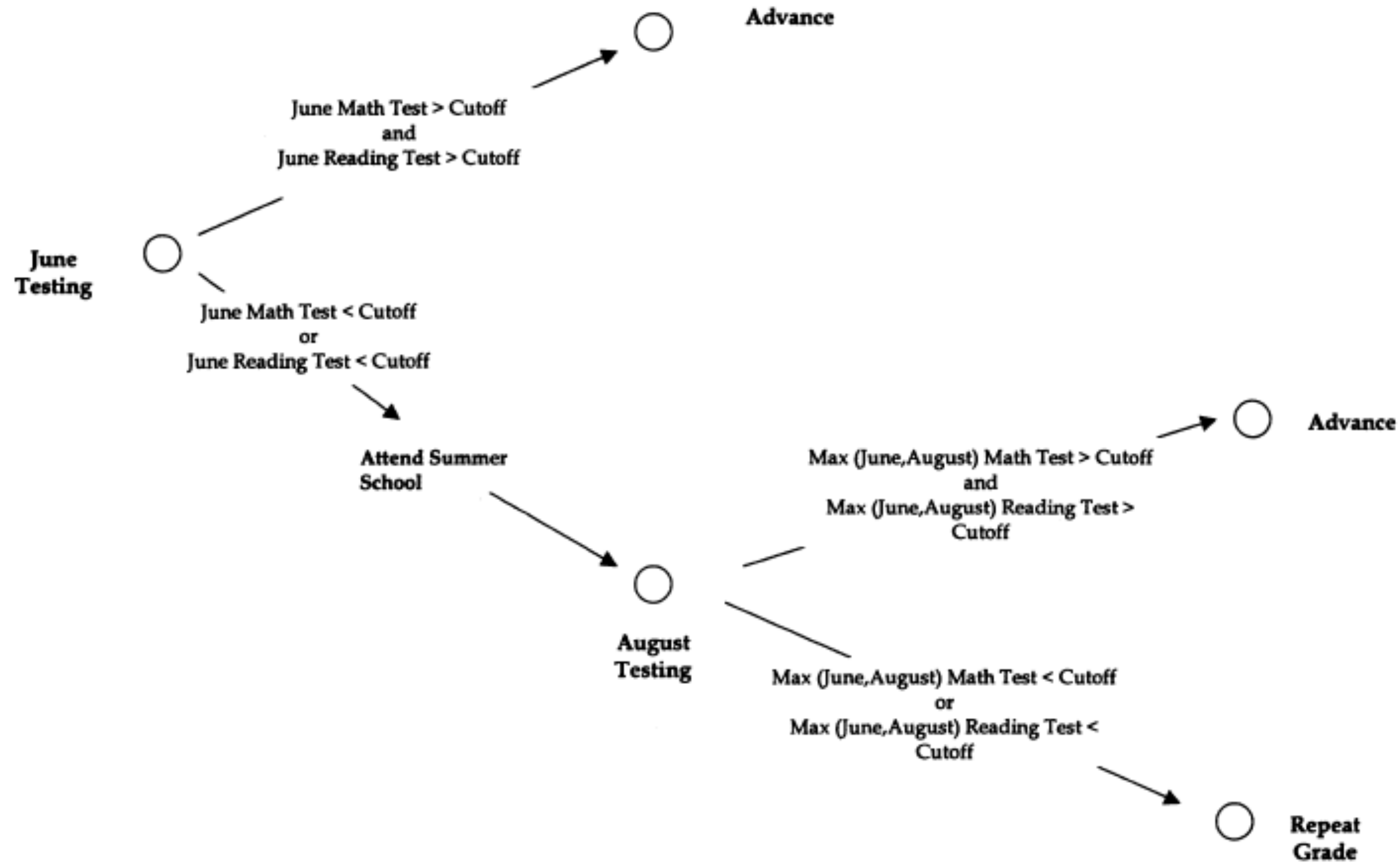
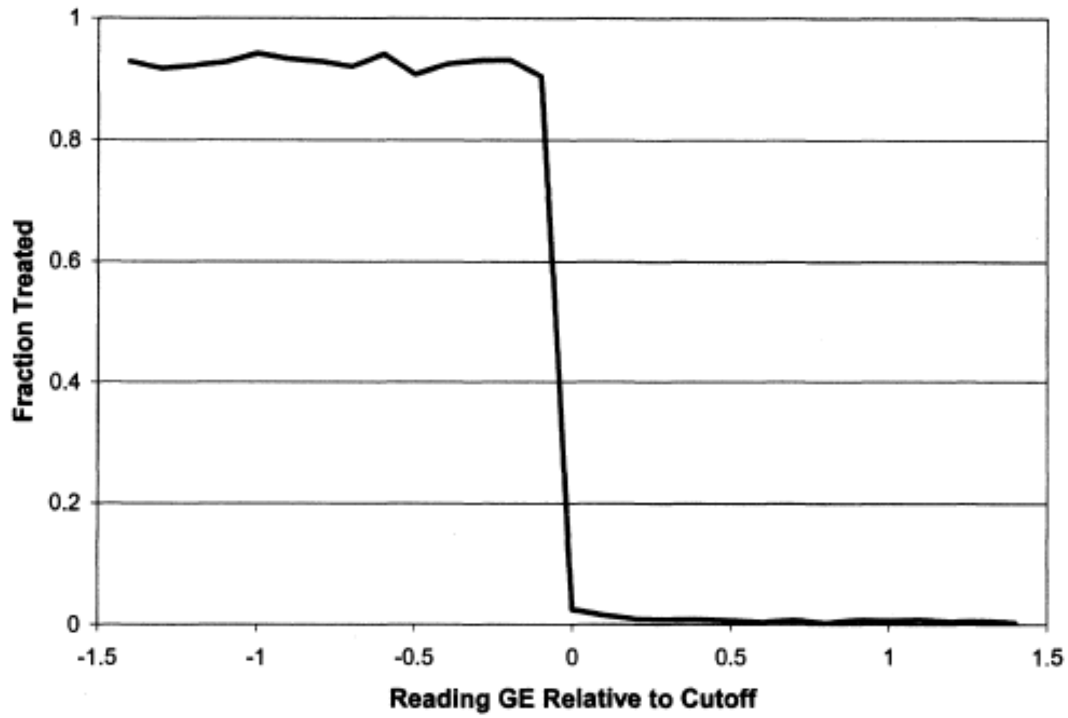


Figure 2.—The Relationship between June Reading Scores and the Probability of Attending Summer School or Being Retained

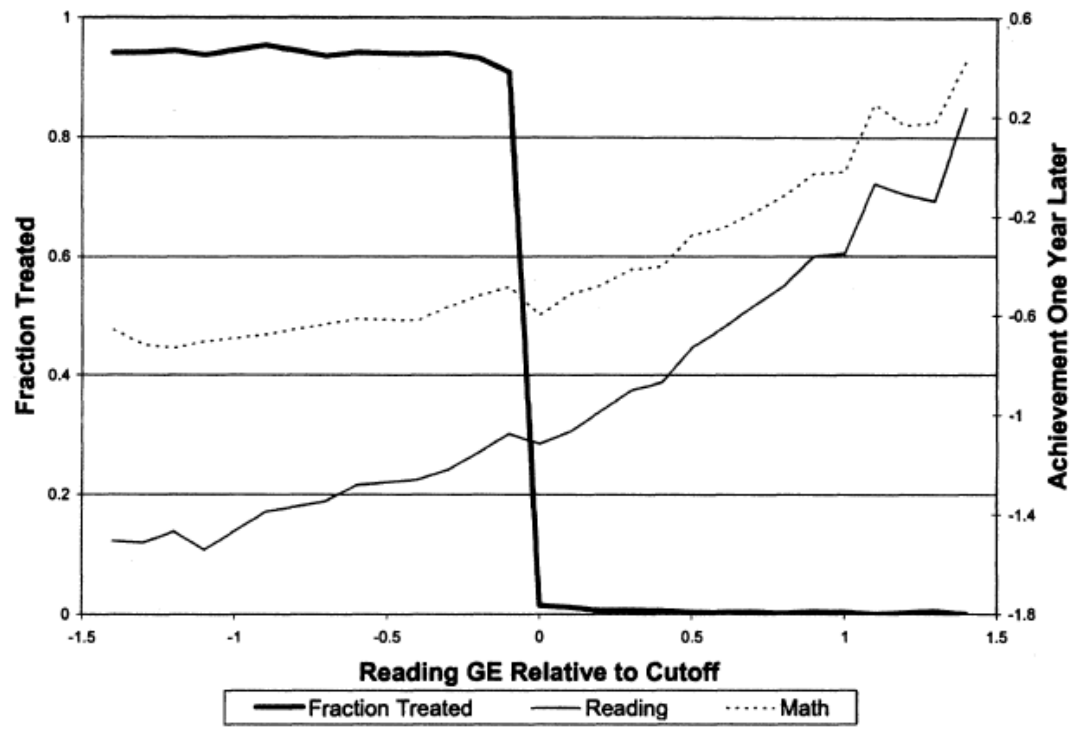


Sample of third- and sixth-grade students from 1997 to 1999 whose June math score exceeded the promotional cutoff but whose June reading score did not.

TABLE 3.—THE NET EFFECT OF SUMMER SCHOOL AND GRADE RETENTION ON STUDENT ACHIEVEMENT

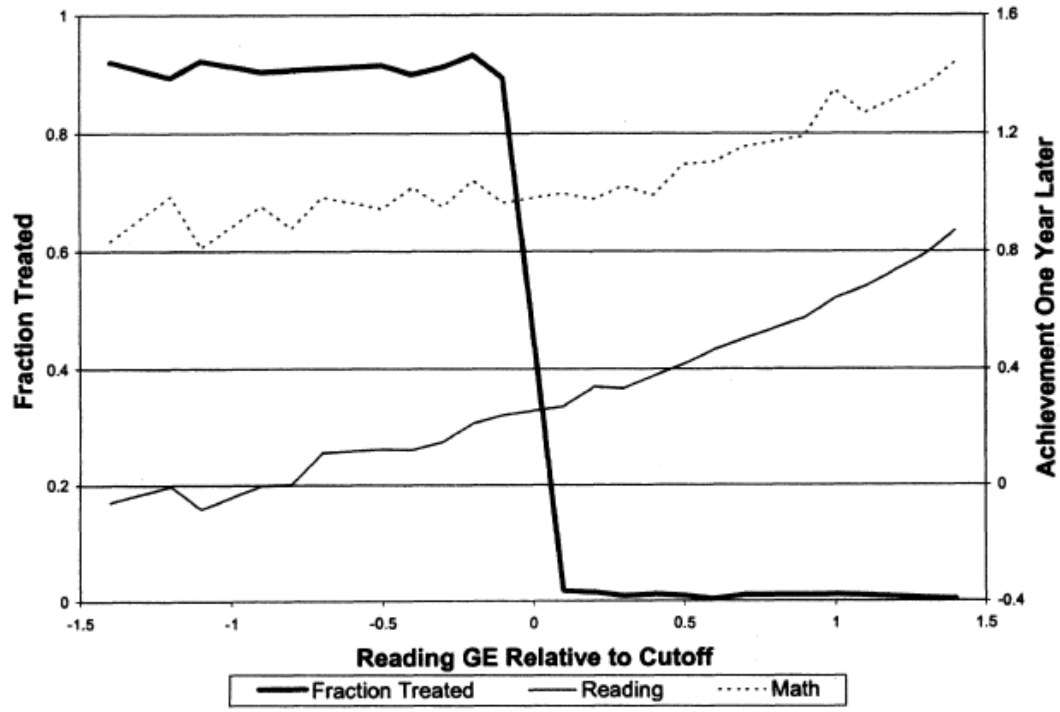
	Specification		
Dependent Variable	OLS (1)	IV (2)	IV (3)
Third grade:			
Reading:			
1 year (n = 13,687)	0.082	0.112	0.104
	(0.019)	(0.026)	(0.025)
2 years (n = 12,806)	0.032	0.064	0.062
	(0.020)	(0.027)	(0.026)
Math:			
1 year (n = 13,664)	0.155	0.132	0.136
	(0.019)	(0.026)	(0.024)
2 years (n = 12,802)	0.066	0.087	0.095
	(0.021)	(0.027)	(0.026)
Sixth grade:			
Reading:			
1 year (n = 7,920)	−0.013	0.012	0.024
	(0.022)	(0.029)	(0.027)
2 years (n = 7,262)	−0.027	−0.015	0.000
	(0.024)	(0.032)	(0.030)
Math:			
1 year (n = 7,904)	0.056	0.077	0.077
	(0.016)	(0.021)	(0.021)
2 years (n = 7,249)	0.007	0.018	0.019
	(0.019)	(0.025)	(0.023)
Additional performance and demographic covariates	No	No	Yes

FIGURE 4.—THE RELATIONSHIP BETWEEN READING AND MATH PERFORMANCE AND JUNE READING PERFORMANCE FOR THIRD-GRADE STUDENTS



Sample of third-grade students from 1997 to 1999 whose June math score exceeded the promotional cutoff but whose June reading score did not.

Figure 5.—The Relationship between Reading and Math Performance and June Reading Performance for Sixth-Grade Students



Sample of third-grade students from 1997 to 1999 whose June reading score exceeded the promotional cutoff but whose June math score did not.

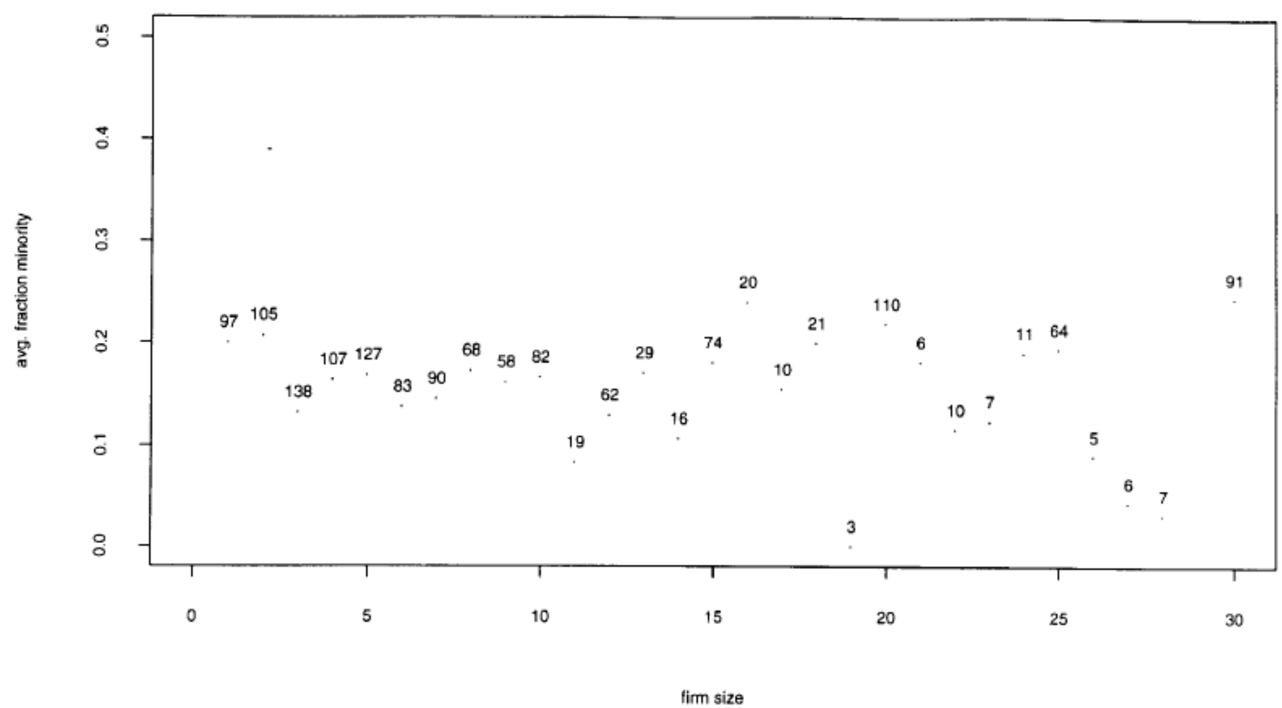
Examples: Hahn, J., Todd, P., Van Der Klaauw, W., (1999).

Evaluating the effect of an anti discrimination law

- Assesses the impact of an anti-discrimination law on minority hiring that mandates that firms with 15 or more employees make reports to the government about the ethnic/racial/gender composition of their work.
- Firms with at least 15 employees are covered by Title VII of the Civil Rights Act (1972 Amendment extended coverage from firms with 25 or more to firms with 15 or more employees).
- Uses a sharp RD design.
- Finds that law led to modest increase in minority hiring.

Examples: Hahn, J., Todd, P., Van Der Klaauw, W., (1999).

Figure 1: Avg. Fraction Minority by Firm Size in 1987 (Number of Observations shown at each point)



Examples: Hahn, J., Todd, P., Van Der Klaauw, W., (1999).

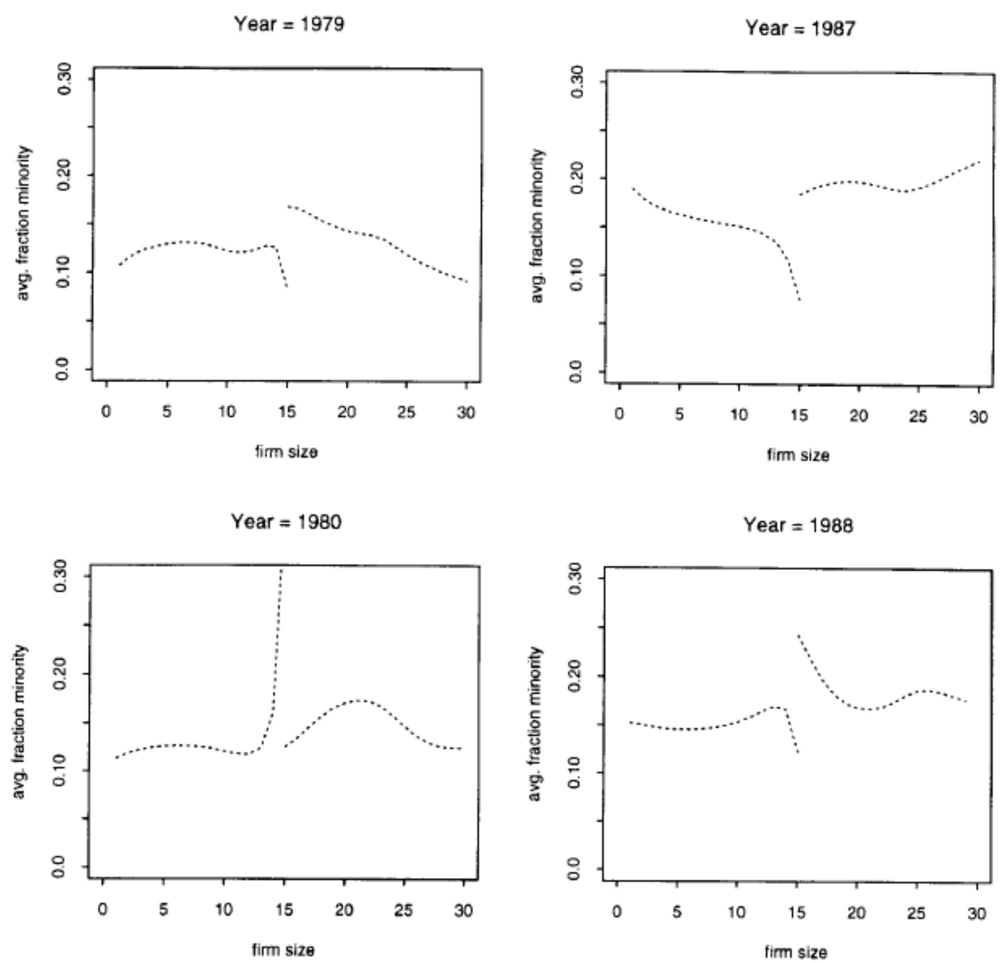
Table 2
Estimated Effects of EEOC-reporting on Percentage Minority in the Firm[a]
(asymptotic standard errors reported in parentheses)

	Year									
Bandwidth	**1979**	**1980**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**
8	-8.7 (9.1)	-26.8 (11.3)	9.8 (6.7)	10.9 (5.3)	12.2 (8.9)	9.8 (11.1)	4.7 (5.9)	11.1 (4.1)	3.7 (10.4)	7.6 (9.1)
10	-2.1 (5.9)	-7.5 (9.0)	5.1 (6.9)	10.1 (4.6)	3.7 (7.5)	1.5 (6.0)	4.3 (7.3)	9.4 (4.6)	2.7 (10.0)	6.1 (8.8)
12	3.3 (4.5)	1.9 (13.2)	3.3 (8.7)	10.5 (4.7)	-1.3 (8.3)	5.5 (6.7)	7.4 (9.0)	9.6 (4.8)	2.6 (12.0)	4.7 (12.0)
14	8.2 (5.4)	8.8 (18.6)	1.9 (11.2)	10.6 (5.1)	-2.7 (8.0)	8.0 (8.4)	10.3 (8.8)	11.5 (4.7)	5.3 (13.8)	-1.5 (14.0)
global IV	3.3 (0.3)	7.1 (0.4)	8.4 (0.3)	6.7 (0.3)	8.3 (0.3)	8.2 (0.3)	7.2 (0.3)	7.9 (0.4)	11.0 (0.5)	7.4 (0.4)
local IV with bw=12	0.92 (0.3)	3.03 (0.3)	2.8 (0.2)	4.0 (0.1)	3.7 (0.3)	3.8 (0.2)	2.6 (0.2)	3.0 (0.3)	3.5 (0.3)	-1.0 (0.3)

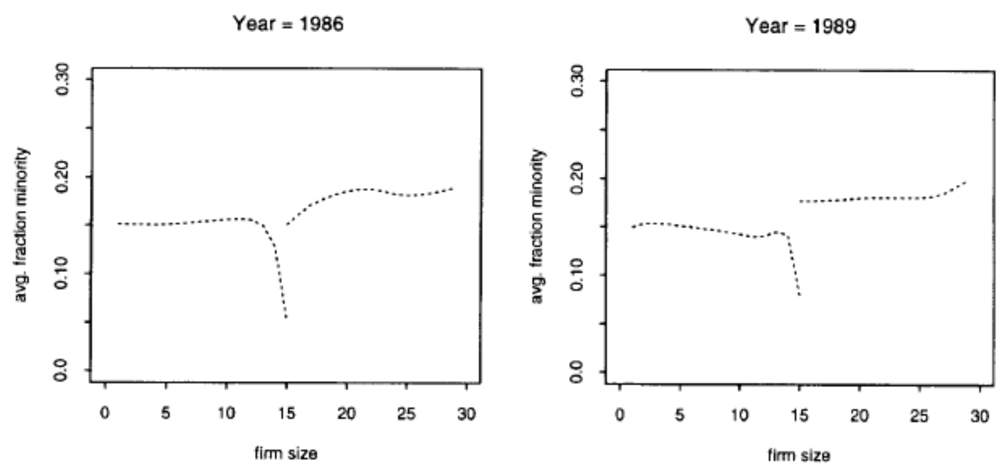
(a) The first four rows report results from using the estimator based on local linear regression described in section 4.2 of the text. The last two rows report results using a global Wald estimator (i.e. one that uses all the data) and a local Wald estimator (one that only uses data within a bandwidth).

Examples: Hahn, J., Todd, P., Van Der Klaauw, W., (1999).

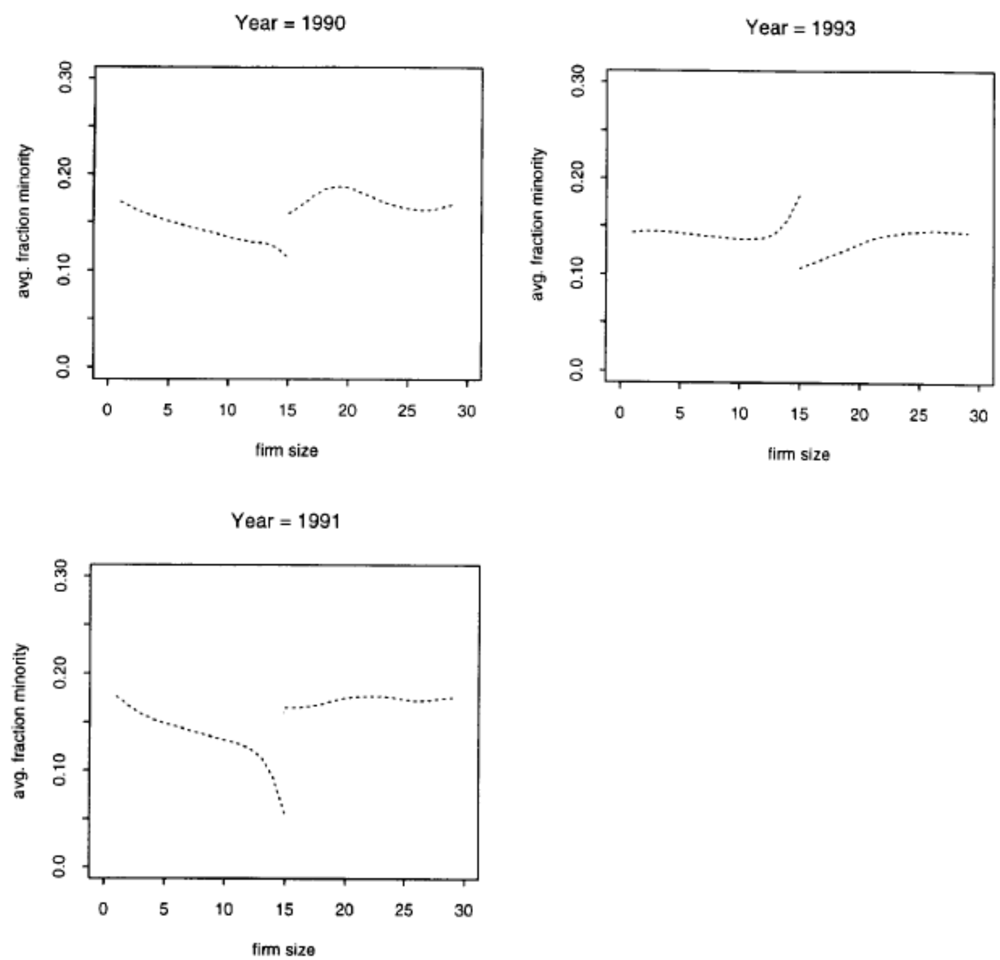
Figure 2: Estimated Percentage Minority Conditional on Firm Size (bw=12)



Examples: Hahn, J., Todd, P., Van Der Klaauw, W., (1999).



Examples: Hahn, J., Todd, P., Van Der Klaauw, W., (1999).



RD Examples

Effect of class size on scholastic achievement

Angrist, J.D., Lavy, V. (1999, QJE)

- Analyzes the effect of class size on student test scores using data from Israel and exploiting a discontinuity created by , which states that a class be added whenever average class size reaches 40 students.
- Finds that reducing class size induces a significant and substantial increase in test scores for fourth and fifth graders, although not for third graders.

RD Examples

Estimating the value parents place on school quality

Black, S., (1999,QJE)

- Uses house prices to infer the value parents place on school quality.
- Compares, within school districts, the prices of houses located on attendance district boundaries - houses that differ only by the elementary school the child attends.
- This comparison removes the variation in neighborhoods, taxes, and school spending.
- Finds that parents are willing to pay 2.5 percent more for a 5 percent increase in test scores.
- Possible that parents on either side are different, so that estimate is a lower bound on valuation.

RD Examples:
Chay, K., McEwan, P., Urquiola, M., (AER, 2005)
Effects of a school incentive program on test score performance

- Evaluates the effect of a school-incentive program in Chile (the Chile-900 program) in which resources were allocated based on cutoffs in schools' mean test scores.
- Shows how a regression discontinuity design that exploits the discrete nature of the selection rule can be used to evaluate the program.
- Finds that the P-900 program had significant but modest size effects on test score gains.

RD Examples: DiNardo, J., Lee, D.S., 2004, QJE)

Effect of unionization on labor market outcomes

- Using US establishment-level data on establishments that faced union organizing drives during 1984-1999, this paper uses a sharp RD design to estimate the impact of unionization on business survival, employment, output, productivity, and wages.
- Compares outcomes for employers where unions won the election by a close margin with those where the unions lost by a close margin (e.g. 49% compared to 51%).
- Impacts on all outcomes are small and impacts on wages are close to zero. Concludes that mandates for employers to bargain with unions had little effect.

RD Examples: Card, D., Mas, A., Rothstein, J., (2006, QJE)

Tests for discontinuities in the dynamics of neighborhood racial composition

- Theoretical models of social interactions (Schelling (1971)) predict tipping behavior in neighborhoods - e.g. once the minority share in a neighborhood exceeds a so-called tipping point,all the whites leave.
- This paper uses regression discontinuity methods and Census tract data from 1970 through 2000 to test for discontinuities in the dynamics of neighborhood racial composition.
- Finds evidence for tipping-like behavior in most cities, with a distribution of tipping points ranging from 5% to 20% minority share, but evidence for tipping points in on other outcomes, like house prices.

RD Examples: Lee (2007, J of Econometrics)

Effect of incumbancy advantage

- Uses data on US Congressional election returns from 1946 to 1998.
- Analyzes the effect of the incumbancy advantage at the level of the party at the district level, without regard to the identify of the nominee for the party.
- For example, analyzes the prob of winning the election in t+1 given that democrats won the election in t, coming districts where they won by a close margin to districts where they lost by a close margin.
- Paper recommends checking the density of observables to test for systematic selection around the cut-off point.
- Finds that democrats who just barely win the election are much more likely to run for office and succeed in the next election compared to democrats who barely lose, which implies a large incumbency advantage. (also see Moretti and Butler, 2004, QJE)

Recommended RD Practices

(Imbens and Limieux, 2007)

Sharp RD Designs

1. Graph the data by computing the average value of the outcome variable over a set of bins. The bin width should be large enough to have a sufficient amount of precision so that the plots looks smooth on either side of the cut-off value, but also small enough to be able to see the jump around the cut-off value.

2. Estimate the treatment effect by running linear regressions on both sides of the cut-off points using only data within a bin width h of the cut-off point. These are kernel regressions using a rectangular kernel.

-standard errors can be computed using standard least squares methods (using robust standard errors).

-optimal bandwidth can be chosen using cross-validation

3. Examine robustness of the results by

(a) looking at possible jumps in the value of other covariates around the cut-off point.

(b) using various values of the bandwidth, with and without controlling for other covariates in the regression.

4. The performance of the estimator can be improved by using nonparametric local linear regression and computing the standard errors either using a plug-in approach or by bootstrapping.

Fuzzy Regression Discontinuity Designs

1. Graph the average outcomes over a set of bins as in the case of SRD, but also graph the probability of treatment.

2. Estimate the treatment effect using TSLS applied only to data within h of the cut-off (above and below), which is numerically equivalent to computing the ratio in the estimate of the jump (at the cutoff point) in the outcome variable over the jump in the treatment variable.

3. Standard errors can be computed using robust TSLS estimates or using a plug-in estimator.

4. Robustness can be examined using similar approaches as in SRD.

5. The performance of the estimator can again be improved by using nonparametric local linear regression instead and computing the standard errors either using a plug-in approach or by bootstrapping.