Regression discontinuity design

Matthieu Stigler
Matthieu.Stigler@gmail.com

NIPFP, 3 June 2014

Outline

1. Introduction
2. The econometrics of RDD
3. The statistics of RDD
4. Sensitivity tests
5. Conclusion

The idea

Introduction

The econometrics of RDD

The statistics of RDD

Sensitivity tests

Conclusion

The goal: Estimate a treatment effect in a case where there might be self-selection/endogeneity.

The framework: Selection is based on a simple known (administrative) rule, based on a cutoff

The idea: treated and untreated are in general different, but those just around the cutoff should be similar.

The solution: make a local comparison, comparing just above with just below

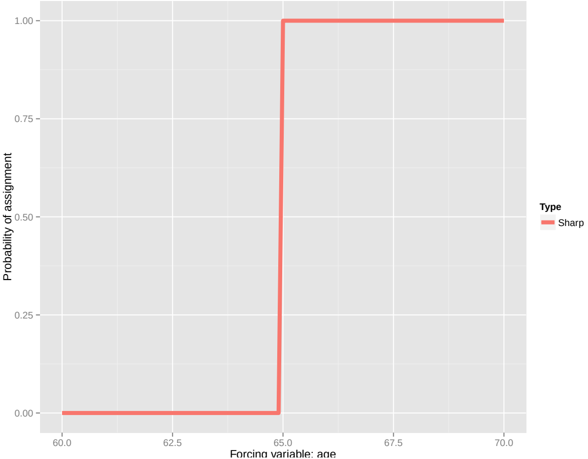
Two cases

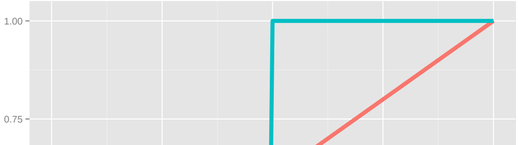
Introduction
The econometrics of RDD
The statistics of RDD
Sensitivity tests
Conclusion

Two (three) cases:

- **Sharp RDD:** Assignment is strict (i.e. all above are treated, all below untreated)
- **Fuzzy RDD:** Assignment does not lead to full compliance
- **Mixed:** One sided-compliance

Figure 1: Assignment probabilities in sharp versus fuzzy





Examples

Introduction
The econometrics of RDD
The statistics of RDD
Sensitivity tests
Conclusion

Age: Retirement and Medicare (Card 2004). Alcohol in US and mortality (Carpenter et al 2007)

School score: Angrist et al. 2011, 2012...

Class size: If above 40 students, split 1 class into 2. (Angrist Lavy 1999)

Geo. location: Geo-wise treatment (for state, country): compare individuals at the border

Majority rule: For elections, Lee (2008)

Height: Military service ?

Poverty rate: Indian BPL? Super fuzzy...

The history

Majour contributions in eco:

- Thistlewaite and Campbell (1960)
- Angrist (1999) Fuzzy RDD, early application in economics
- Hahn et al (2001) Identification of ATE by RDD
- Porter (2003): Estimation of RDD
- Imbens and Lemieux (2008) State of the art in RDD
- Lee and Lemieux (2010) State of the art in RDD

cf Cook (2008) *Waiting for Life to Arrive: A History of the RDD*

Causality in econometrics

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Granger: X causes Y if X helps predict Y

Counterfactual: X causes Y if there is a difference between the Y observed when X occurred and the un-observed counterfactual Y *if X had not occurred*.

Refs: Granger (1969), Holland (1980)

Causality in econometrics

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Definitions:

Def: treatment variable

Individual i has treatment status:

$$D_i = \begin{cases} 1 & \text{if treated} \\ 0 & \text{if not treated} \end{cases}$$

Def: potential outcome

Individual i has outcome:

$$\begin{aligned} Y_i(D_i): &\quad \text{oberved outcome (treated or not)} \\ Y_i(1-D_i): &\quad \text{countefactual/hypothetical outcome} \end{aligned}$$

Term counter-factual

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Beware that counterfactual is used in two ways:

- **Hypothetical:** The unobserved outcome: $Y_i(0)|D_i = 1$ or $Y_i(1)|D_i = 0$
- **Approximation:** The closest factual: $Y_i(0)|D_i = 0$ for $Y_i(0)|D_i = 1$

Term has changed from: *we use **as** counter-factual* to *we use **the** counter-factual*

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Term counter-factual II

An example of the similarity of counter-factual notions:

Everyday life

Do you feel warmer with the jacket?

- Than you were before? $\Rightarrow$ closest factual
- Than you would be without? $\Rightarrow$ unobserved outcome

Counterfactual thinking is not new:

- Used every day, studied extensively in psychology
- Lewis (1973) "Causation", *Journal of Philosophy*
- Used in law, forensic medicine...

Term counter-factual III

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Another confusion in economics about Counterfactual analysis

CGE: What if policy A (country A reduces tarif by 3%)

Impact evaluation: Answers factual questions by asking counter-factual questions.

But IE answers counter-factual questions: treatment effect on the untreated.

Causal inference

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Def: Effects of the causes

The cause D causes the effect $Y_i(D=1) - Y_i(D=0)$

Fundamental Problem of Causal Inference (Holland 1980)

It is impossible to observe simultaneously the value of $Y_i(D=1)$ and $Y_i(D=0)$ on the same unit and, therefore, it is impossible to observe the effect of D for i.

Causal inference: the statistical approach

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

In statistics, given the fundamental problem, we do inference for population, not individuals:

Definition: average treatment effect

$$ATE \equiv E[Y_i(1) - Y_i(0)]$$

Back to causalities

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

We see now more fundamental differences between Granger and Counterfactual causality:

- Granger: Defines causes of effects: Among $X_1, X_2, \ldots, X_k$ which cause(s) Y?
- Counterfactual: Measures effects of causes: what is effect of X_1 on Y?

But is is really effect**s** of cause**s**? Most usually, is single effect of single cause.

Causal inference: the statistical approach

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Two more definitions

Definition

$$\text{ATE on treated: } ATT \equiv E[Y_i(1) - Y_i(0)|D = 1]$$
$$\text{ATE on untreated: } ATU \equiv E[Y_i(1) - Y_i(0)|D = 0]$$

Estimating the ATE

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

One could think that ATE is estimated simply:

$$\begin{aligned} ATE &\equiv E[Y_i(1) - Y_i(0)] \\ &= E[Y_i(1)] - E[Y_i(0)] \end{aligned}$$

Unfortunately, we do oberve $E[Y_i(1)]$ only for the $\{j|D_j = 1\}$ i.e. we observe $E[Y_i(1)|D_i = 1]$!

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

The frameworks in impact evaluation:

- **Sample selection:** Old-fashioned
- **RCT:** allocate D randomly, so $E[Y_i(1)|D_i = 1] = E[Y_i(1)]$ and hence just compare means!
- **Matching:** Assume D is based on observables: ATT and ATU, ATE
- **Diff-diff** Assume D is based only on time-invariant unobservables: ATT
- **IV:** Find exogeneous moves in D (estimating LATE/CATE)
- **RDD:** Assume D is random close to threshold: Local ATE

Identification in the sharp case

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

In the sharp RDD, treatment is assigned based on whether the forcing variable x is below/above a threshold c.

$$D_i = 1(x_i \geq c) = \begin{cases} 1 & \text{if } x_i \geq c \\ 0 & \text{if } x_i < c \end{cases}$$

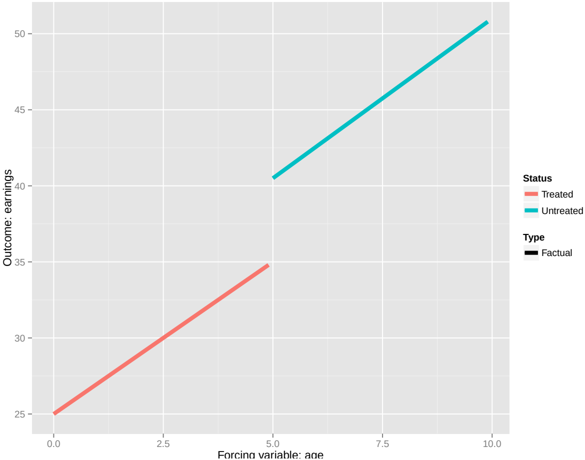
We will identify the ATE at a point:

Definition: point ATE

$$ATE(c) \equiv E[Y_i(1) - Y_i(0)|x_i = c]$$

Note here we have a local ATE in a well-defined sense (local in space of x)!

Figure 2: Assignment probabilities in sharp versus fuzzy



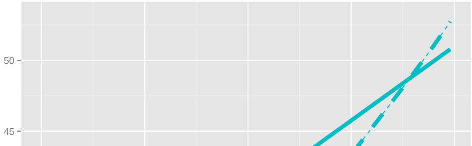
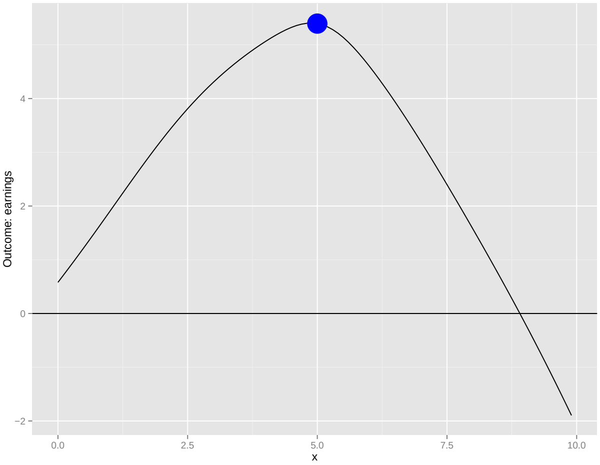


Figure 3: Assignment probabilities in sharp versus fuzzy



The RDD estimand

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

To estimate:

$$ATE(c) \equiv E[Y_i(1) - Y_i(0)|x_i = c] = \overbrace{E[Y_i(1)|c]}^{\text{Factual}} - \overbrace{E[Y_i(0)|c]}^{\text{Counter-factual}}$$

we approximate the *counter-factual* untreated at c by the *factual* untreated just below c

$$\lim_{x \uparrow c} E[Y_i(0)|x_i = c] \approx E[Y_i(0)|x_i = c]$$

So end up using:

$$\overbrace{\lim_{x \downarrow c} E[Y_i(1)|x_i = c]}^{\text{Just treated}} - \overbrace{\lim_{x \uparrow c} E[Y_i(0)|x_i = c]}^{\text{Just untreated}}$$

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Assumptions

Main assumptions

- Identifying assumption: $E[Y(1)|c]$ and $E[Y(0)|c]$ are continuous in c.
- No manipulation of x around c

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

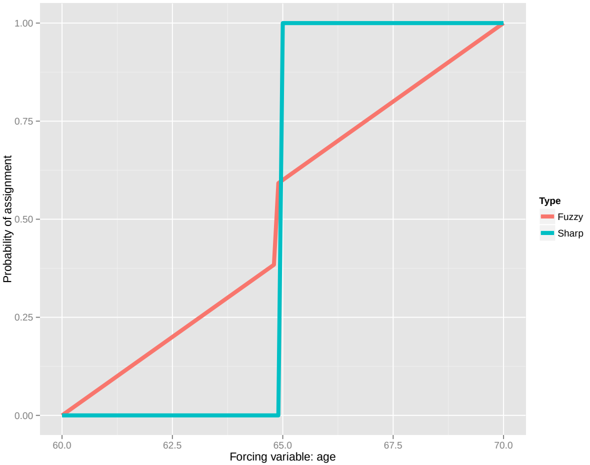
Assumption 2

Test 2: Manipulation of forcing variable

Check if individuals can manipulate their value around x?

- Age? (if *declared* age...)
- Declared income?
- Number of employees?
- Test score: can retake exam?

Figure 4: Assignment probabilities in sharp versus fuzzy



Cases

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Sharp case: $D_i = 1(x_i \geq c)$

Fuzzy case: $\lim_{x \downarrow c} Pr(D_i = 1|x) \neq \lim_{x \uparrow c} Pr(D_i = 1|x)$

Sharp new: $\overbrace{\lim_{x \downarrow c} Pr(D_i = 1|x)}^{=1} - \overbrace{\lim_{x \uparrow c} Pr(D_i = 1|x)}^{=0} = 1$

Mixed case: One-sided compliance only

Cases

In the fuzzy RDD, we use:

$$ATE \equiv \frac{\lim_{x\downarrow c} E[Y|X=x] - \lim_{x\uparrow c} E[Y|X=x]}{\lim_{x\downarrow c} E[D|X=x] - \lim_{x\uparrow c} E[D|X=x] \underbrace{\lim_{x\downarrow c} E[D|X=x] - \lim_{x\uparrow c}}_{=1 \text{ for sharp!}}}$$

This is just the form the IV takes for binary D and binary Z: Wald estimator! (see MHE 4.1.2)

$$IV_{\text{Wald}} \equiv \frac{E[Y|z_i=1] - E[Y|z_i=0]}{E[D|z_i=1] - E[D|z_i=0]}$$

Identification in the fuzzy case

Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

But in the fuzzy case, like with standard IV, we estimate only the complier-average effect (CATE/LATE)

And here, we only get a local version of that! $\Rightarrow$ local-local-average!

On the IV interpretation of RDD

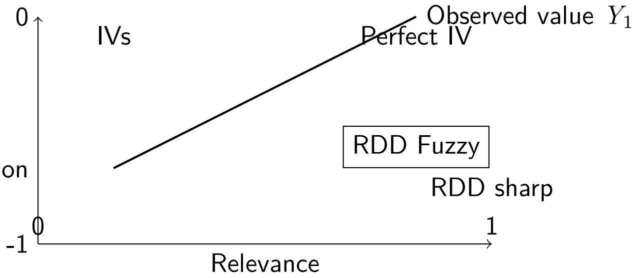
Introduction
The econometrics of RDD
Causal inference with potential outcomes
Sharp RDD
Fuzzy RDD
The statistics of RDD
Sensitivity tests
Conclusion

Remember the two conditions for IV with Y, X, Z:

Relevant: Should be relevant in first stage $(\beta \neq 0)$:
$X_i = \alpha + \beta Z_i + \varepsilon_i^X$

Exclusion: Should be excluded of structural form: $(\theta = 0)$:
$Y_i = \gamma + \delta X_i + \theta Z_i + \varepsilon_i^Y$

RDD recognises $\theta = 0$ only locally!



Parametric model

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

We want to estimate: $\lim_{x \downarrow c} E[Y_i(1)|c] - \lim_{x \uparrow c} E[Y_i(0)|c]$

So can estimate separately on each side of each side of c:

Left: $(1-D)Y = \alpha_l + \beta_l(1-D)(X-c) + \epsilon_l$

Right: $DY = \alpha_r + \beta_r D(X-c) + \epsilon_r$

with $D = I(X \geq c)$, and test $H_0 : \quad \alpha_r = \alpha_l$

Assuming homoskedasticity ($\sigma_l = \sigma_r$, can estimate instead:

$$Y = \alpha + \tau D + \beta_1(X-c) + \beta_2 D(X-c) + \epsilon$$

Now we have $\tau = \alpha_r - \alpha_l$

Parametric model: extension II

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Eventually, we could assume the same slope on each side:

$$Y = \alpha + \tau D + \beta(X - c) + \epsilon$$

Parametric model: extension II

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Eventually, add higher-order polynomials:

$$Y = \alpha + \tau D + \beta_1^1 (X-c) + \beta_2^1 D(X-c) + \ldots + \beta_1^p (X-c)^p + \beta_2^p D(X-c)^p + \epsilon$$

What if use polynomial of order zero? **Group means test (=RCT!)**

What is parametric?

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

What is parametric?

- Functional form
- Distributional assumptions (cf MLE)

What is non-parametric here?

- Local/pointwise estimation **This is exactly what we want!**

Local/pointwise estimation

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

With kernel regression/smoother, want to estimate:

$$\mathrm{E}(Y|X) = m(X)$$

The idea is to estimate $\hat{y}(x_i)$ averaging over local neighbours.

$$\widehat{m}_h(x) = \sum_{i=1}^{n} w_i Y_i$$

Two approaches for w_i:

- use k nearest neighbours.
- use all neighbours within a window h.

Local/pointwise estimation

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

The Nadaraya-Watson (1964) estimator use a window with weights:

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^{n} \mathcal{K}_h(x - X_i)\, Y_i}{\sum_{i=1}^{n} \mathcal{K}_h(x - X_i)}$$

where $\mathcal{K}$ is a kernel with a bandwidth h.

Kernel and bandwidth choice

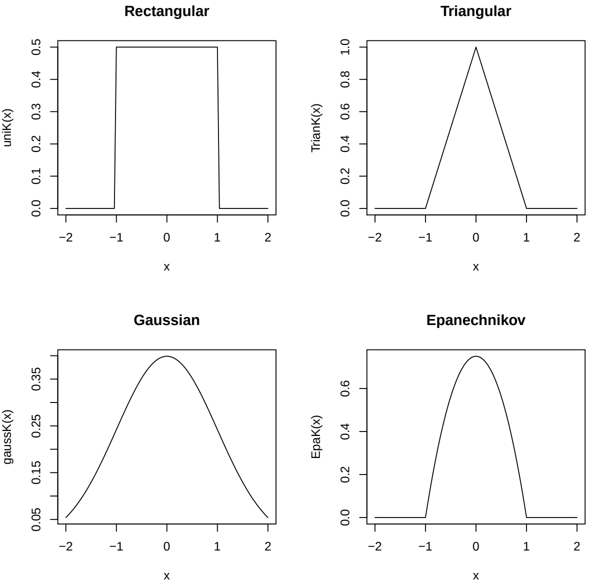
Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Parameters in non-parametric estimation:

Kernel $\mathcal{K}$: Rectangular, Gaussian...

Bandwidth h:
- $h \searrow$: bias $\searrow$, variance $\nearrow$
- $h \nearrow$: bias $\nearrow$, variance $\searrow$

Figure 5: various kernels



Problems

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Problems with Nadaraya-Watson:

- Biased
- Higher bias at boundary point **This is however what we want :-(**

Solution: Use local-linear regression

Local-linear regression

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Local-linear regression (LLR): fit a local weighted OLS at every point!
So we get $\hat{m}(x_i)$, and also for every point get a new $\hat{\alpha}(x), \hat{\beta}(x)$

It is a simple weighted regression, with weights obtained from kernel around x.

$$\hat{\alpha}(c), \hat{\beta}(c), \hat{\tau}(c) = \arg\min_{\alpha,\beta,\tau} \sum_{i=1}^{n} \left(Y_i - \alpha - \tau D - \beta(X_i - c)\right)^2 \mathcal{K}\left(\frac{X_i - c}{h}\right)$$

Local-linear regression

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Links with LLR:

OLS: If use LLR with $h \to \infty$ and rectangular $\mathcal{K}$: LLR $\to$ OLS!

Nadaraya-Watson: NW is a LLR with only a constant!

Choice of bandwidth

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Two procedures to choose bandwidth:

Plug-in: minimise (asymptotic) $\mathrm{MSE}(h) = \mathrm{bias}(h)^2 + \mathrm{var}(h)$
pluging-in estimates of bias and variance

Cross-validation: Minimise empirical MSE: $\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{\mu}(X_i)\right)^2$

Inference

Introduction
The econometrics of RDD
The statistics of RDD
Parametric approach
Nonparametric approaches
Sensitivity tests
Conclusion

Distribution: Normal!

Two types of variance estimators for non-parametric:

- Parametric: $\sigma \left(X' X \right)^{-1}$
- Non-parametric: Messy!!

Rate of convergence: slower than parametric!

Sensitivity tests

Introduction
The econometrics of RDD
The statistics of RDD
Sensitivity tests
Conclusion

- Bandwidth sensitivity
- Manipulation of forcing variable
- Placebo tests
- Discontinuity also for covariates?

Sensitivity tests: sensitivity to bandwidth

Test 1: sensitivity to bandwidth

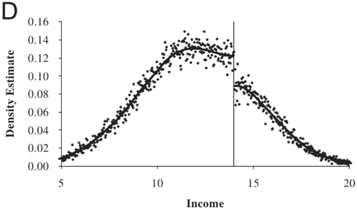
Compute estimate at various bandwidths, compare

Sensitivity tests: Manipulation of forcing variable 1

Introduction
The econometrics of RDD
The statistics of RDD
Sensitivity tests
Conclusion

Test: look for discontinuity of density around threshold

McCrary (2008): look for discontinuity of the density of X at point c:



Sensitivity tests: Manipulation of forcing variable 2

Introduction
The econometrics of RDD
The statistics of RDD
Sensitivity tests
Conclusion

Actually, this approach helps also to check how a constraint is binding, in general.
Cf paper of Card and DellaVigna (2012) *How Do Authors Respond to Page Limits*

Figure 6: AER limits (Card DellaVigna 2012)

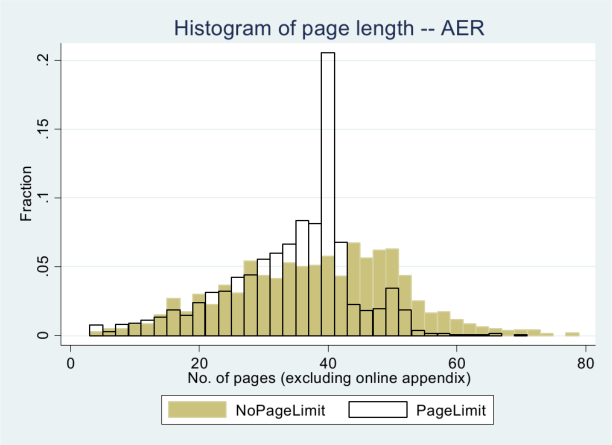
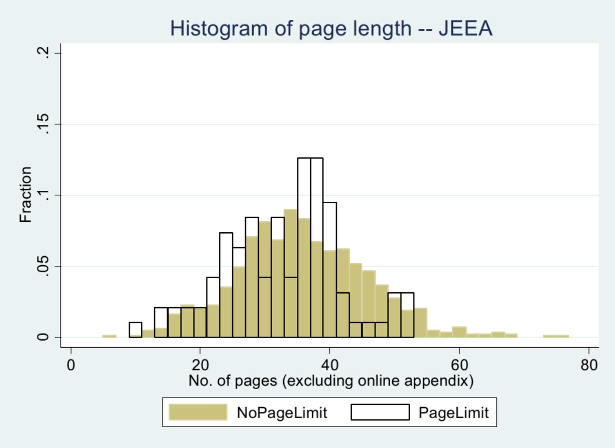


Figure 7: JEEA limits (Card DellaVigna 2012)



Sensitivity tests: Placebo tests

Introduction
The econometrics of RDD
The statistics of RDD
Sensitivity tests
Conclusion

Test 3: Placebo tests

Test for effect at a point known for known effect: estimate RDD at non-cutoff point.

Test 4: Discontinuity also for covariates

Like placebo, run RDD using as new y a covariate.

But passing test is not necessary! Just reduces credibility.

RDD

Introduction
The econometrics of RDD
The statistics of RDD
Sensitivity tests
Conclusion

RDD in a nutshell

- Estimating ATE just around cutoff
- So identify effect only around the cutoff, low external validity
- Step 1: Visual representation important
- Step 2: Estimation by parametric or non-parametric
- Step 3: Few sensitivity tests