Econometrics 2

Arieda Muço

Central European University

Spring 2020

Fuzzy Regression Discontinuity

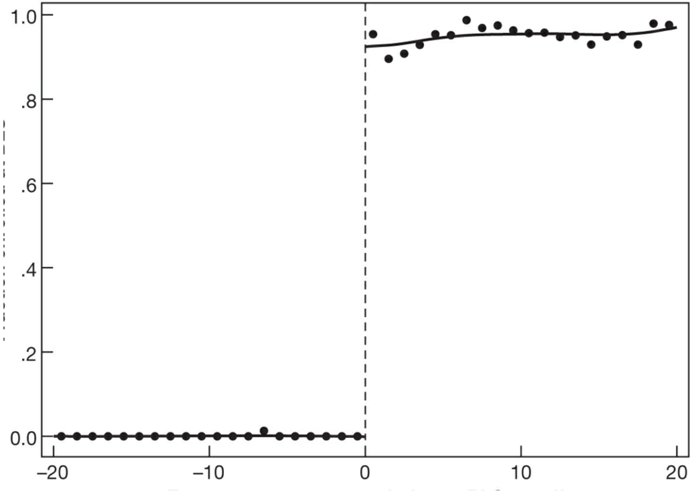
Probability of receiving treatment does not change from zero to one at the cutoff, but

$$\lim_{z \uparrow z_0} P(D_i = 1 | z_i = z) \neq \lim_{z \downarrow z_0} P(D_i = 1 | z_i = z)$$

other variables also determine treatment assignment, but incentives to participate change discontinuously at the threshold.

- Examples: Incentives to participate in some program may change discontinuously at the cutoff but are not powerful enough to move everyone from non participation to participation.

Example



Fuzzy Regression Discontinuity is IV

$$\hat{\rho}_f = \frac{\underset{z \downarrow z_0}{lim} E(Y_i|z_i = z) - \underset{z \uparrow z_0}{lim} E(Y_i|z_i = z)}{\underset{z \downarrow z_0}{lim} E(D_i|z_i = z) - \underset{z \uparrow z_0}{lim} E(D_i|z_i = z)}$$

Fuzzy RD is numerically equivalent and conceptually similar to instrumental variables

- Numerator: "jump" in the regression of the outcome on the running variable, z. **Reduced form**
- Denominator: "jump" in the regression of the treatment indicator on the running variable z. **First stage**

Idea: use the discontinuity created by this rule to identify the treatment effect

First stage relationship between Z and D

- In the sharp RD, D_i was determined by z_i and z_0; in the fuzzy RD, the conditional probability of treatment jumps at z_0
- The relationship between the probability of treatment and z_i can be written as:

$$P(D_i = 1|z_i) = E(D_i = 1|z_i) = g_0(z_i) + [g_0(z_i) - g_1(z_i)]T_i$$

- where $T_i = 1$ if $(z_i \geq z_0)$ and 0 otherwise; indicating the point of discontinuity in $E(D_i = 1|z_i)$

Fuzzy design

Fig 3: Assignment Probabilities (Fuzzy RD)

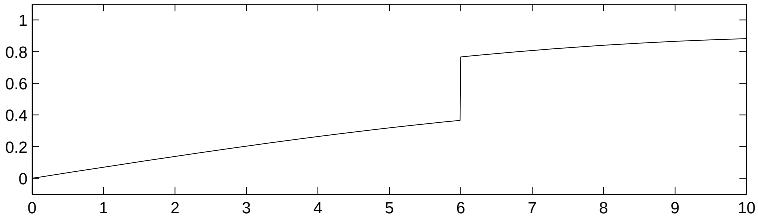
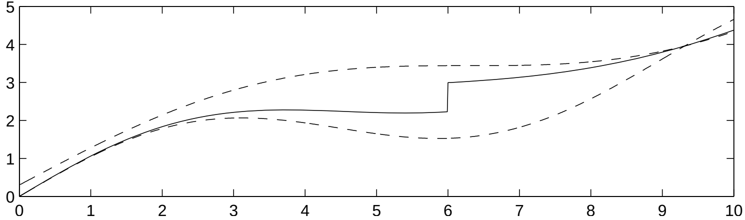


Fig 4: Potential and Observed Outcome Regression (Fuzzy RD)



First stage relationship between Z and D

- One can use both T_i as well as the interaction terms as instruments for D_i.
- The first stage would be:

$$D_i = \gamma_0 + \gamma_1 z_i + \gamma_2 z_i^2 + \gamma_3 z_i^3 + ... + \gamma_p z_i^p + \pi T_i + \zeta_{1i}$$

- where π is the causal effect of T_i on the conditional probability of treatment
- The fuzzy RD reduced form is:

$$Y_i = \mu + \kappa_1 z_i + \kappa_2 z_i^2 + \kappa_3 z_i^3 + ... + \kappa_p z_i^p + \rho\pi T_i + \zeta_{2i}$$

Fuzzy RD with varying Treatment Effects - Second Stage

- As in the sharp RD case one can allow the smooth function to be different on both sides of the discontinuity
- The second stage model with interaction terms would be the similar to before:

$$Y_i = \alpha + \beta_{01}\tilde{z}_i + \beta_{02}\tilde{z}_i{}^2 + \beta_{03}\tilde{z}_i{}^3 + ... + \beta_{0p}\tilde{z}_i{}^p$$
$$+\rho D_i + \beta_1^* D_i\tilde{z}_i + \beta_2^* D_i\tilde{z}_i{}^2 + \beta_3^* D_i\tilde{z}_i{}^3 + ... + \beta_p^* D_i\tilde{z}_i{}^p + \eta_i$$

Where $\tilde{z}_i$ are now not only normalized with respect to z_0 (also fitted values obtained from the first stage regression)

Fuzzy RD with varying Treatment Effects - First Stage

- Again one can use both T_i as well as the interaction terms as instruments for D_i
- Only using T the estimated first stages would be:

$$D_i = \gamma_{00} + \gamma_{01}\tilde{z}_i + \gamma_2\tilde{z}_i{}^2 + \gamma_3\tilde{z}_i{}^3 + ... + \gamma_p\tilde{z}_i{}^p +$$

$$+\pi T_i + \gamma_1^* T_i\tilde{z}_i + \gamma_2^* T_i\tilde{z}_i{}^2 + \gamma_3^* T_i\tilde{z}_i{}^3 + ... + \gamma_p^* T_i\tilde{z}_i{}^p + \eta_i$$

What does Fuzzy RD Estimate?

- As Hahn, Todd and van der Klaauw (2001) point out, one needs the same assumptions as in the standard IV framework
- As with other binary IVs, the fuzzy RD is estimating LATE: the average treatment effect for the compliers
- In RD, the compliers are those whose treatment status changed as we moved the value of z_i from just to the left of z_0 to just to the right of z_0

Challenges to RD (not only Fuzzy)

- Treatment is not as good as randomly assigned around the cutoff, z_0, when agents are able to manipulate their running variable scores. This happens when:
  1. The assignment rule is known in advance
  2. Agents are interested in adjusting
  3. Agents have time to adjust
- **Examples:** re-take an exam, self-reported income, etc.
- Some other unobservable characteristic changes at the threshold, and this has a direct effect on the outcome
- **Example:** Age thresholds used for policy (i.e., person turns 18, and faces more severe penalties for crime) is correlated with other variables that affect the outcome (i.e., graduation, voting rights, etc.)

Several formalized tests to evaluate the severity of these problems...

Manipulation of the running variable

Sorting on the running variable

- Assume a treatment, D, is assigned by rule $z \geq z_0$, if individuals chose z such that they sort into D then we say individuals are sorting on the running variable
- **Example:** Suppose a doctor plans to randomly assign heart patients to a medicine that lowers cholesterol and a placebo to study the effect of the medicine on heart attacks within 10 years. The doctor randomly assigns patients to two different waiting rooms, A and B, and plans to give those in A the medicine and those in B the placebo. If some of the patents learn of the planned treatment assignment mechanism, what would we expect to happen? And how would you check for it?

McCrary Density Test

Sorting on the running variable

- We would expect waiting room A to become crowded. In the RDD context, sorting on the running variable implies heaping on the "good side" of z_0 the running variable
- McCrary (2008) suggests a formal test. Under the null the density should be continuous at the cutoff point. Under the alternative hypothesis, the density should increase at the cutoff (where D is viewed as good)
    1. Partition the assignment variable into bins and calculate frequencies (i.e., number of observations) in each bin
    2. Treat those frequency counts as dependent variable in a local linear regression

McCrary Density Test

The McCrary Density Test has become mandatory for every analysis using RD

- If you can estimate the conditional expectations, with data on the running variable. In principle you can always do a density test
- For RD to be useful, you already need to know something about the mechanism generating the assignment variable and how susceptible it could be to manipulation. Note the rationality of economic actors that this test is built on
- A discontinuity in the density is "suspicious" it suggests manipulation is probably going on. This is a high-powered test. You need a lot of observations at z_0 to distinguish a discontinuity in the density from noise

McCrary Density Test

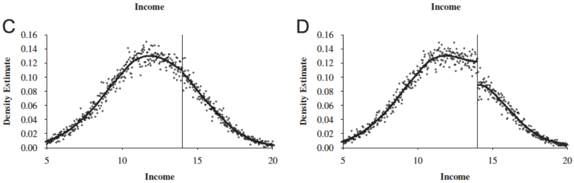


Figure: Panel C is density of income when there is no pre-announcement and no manipulation. Panel D is the density of income when there is pre-announcement and manipulation. From McCrary (2008).

Lee (2008) Incumbency Effect

- David Lee (2008) "Randomized Experiments from non-random selection in US House Elections", Journal of Econometrics analyzes the incumbency effect using Democratic incumbents for US congressional elections
- A large political science literature on the "incumbency advantage" having won an election once helps you win subsequent elections.
- Empirical challenge: how to separate incumbency advantage from selection? (i.e., candidates win multiple elections because they are better)

Lee (2008) Incumbency Effect

- Identification: Incumbency is assigned to a candidate discontinuously at 50% voteshare under two-party democracy with majority rule
- Lee (2008) analyzes the probability of winning the election in year $t+1$ by comparing candidates who just won to candidates who just lost the election in year t.

Density test

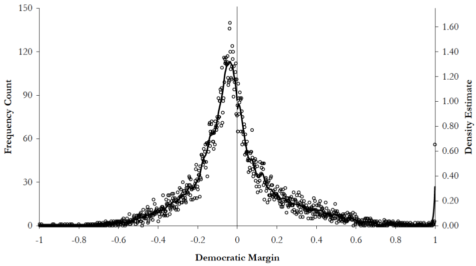


Figure: Democratic vote share relative to cutoff: popular elections to the House of Representatives, 1900-1990 (McCrary 2008).

More evidence of manipulation

- Contrast this with roll call voting in the US House of Representatives
- Coordination is expected because these are repeated games, votes are public records, and side payments are possible in the form of future votes
- Bills around the cutoff are more likely to be passed than not. Seems like a good candidate for RD
- Fails McCrary Density Test; cannot use RD because policy decisions are not quasi-randomly assigned at the cutoff

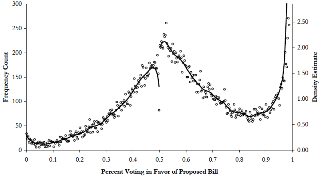


Figure: Percent voting Yeay: Roll Call Votes, US House of Representatives, 1857-2004 (McCrary 2008)

Balance test on covariates

- This is a type of placebo test. For RD to be valid in your study, you do not want to observe a discontinuity around the cutoff, z_0, for average values of covariates that should not be affected by the treatment (e.g., pretreatment characteristics)
- Question: What does a jump in the average values of pre-treatment characteristics have to do with the continuous (smoothness) assumption?
  - Choose pre-treatment covariates, X, and do a similar graphical plot as you did for Y
  - You **do not want** to see a jump around the cutoff, z_0
  - A formal balance test involves the same procedure used to estimate the treatment effect, only use X instead of Y as a LHS variable
  - Can combine test on multiple covariates into a single test

Visualizing Balanced Placebos

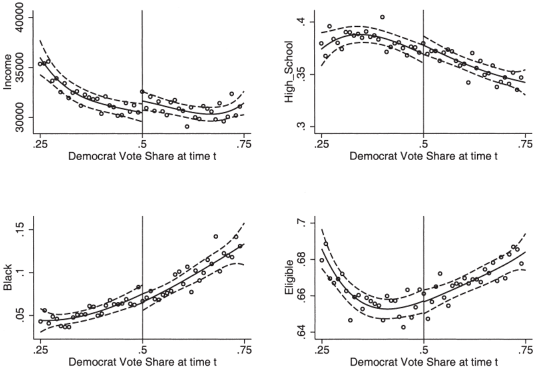


Figure: Figure 3 from Lee, Moretti and Butler (2004), "Do Voters Affect or Elect Policies?" *Quarterly Journal of Economcis*. Panels refer to (top left to bottom right) the following district characteristics: real income, percentage with high-school degree, percentage black, percentage eligible to vote. Circles represent the average characteristic within intervals of 0.01 in Democratic vote share. The continuous line represents the predicted values from a fourth-order polynomial in vote share fitted separately for points above and below the 50 percent threshold. The dotted line represents the 95 percent confidence interval.

Jumps at non-discontinuous points

- Imbens and Lemieux (2008) suggest to look at one side of the discontinuity (e.g., $z < z_0$), take the median value of the running variable in that section, and pretend it was a discontinuity
- Then test whether in reality there is a discontinuity at ${z_0}'$. You do not want to find anything
- Look for non-effects in non-treatment periods

Democracy, Redistribution, and Political Participation: Evidence From Sweden 1919 –1938 (Econometrica, 2014)

- Does the form of Democracy matter?
  - Test direct vs representative democracy
- The elite can capture democratic political process by exercising their de facto political power
  - Lack of (pro-poor) political parties in direct democracy made it harder for the citizens to solve their collective action problem
  - The chairman of the town meeting has agenda setting power
  - Elite relies on intimidation

Tyrefors Hinnerich and Pettersson-Lidbom(2014)

- Local Government Act (LGA) of 1863 granted local governments independent income taxation rights
- The bulk of local government revenues was still is raised through a local proportional income tax
- Focus on rural municipalities (Provide important public services such as education and social welfare)
- In 1918, local governments with a population of more than 1,500 people were required by national law to have representative those with population below could have 1,500 direct democracy
- If a local government had switched to a representative system, it could not switch back within a five-year period democracy
- Running variables: Population in 1918 and Population in the year t-1 for the period from 1919

Data

- Panel dataset for about 2,500 local governments for the period 1918-1938
- Published and unpublished material produced by Statistics Sweden
- National Archives of Sweden and was collected by hand. For the published material, we have digitized it by using data-entry services in India
- Swedish State church kept vital statistics until 1991
- Local governments could try to control individuals in and out
- No evidence of sorting around the threshold

Estimation

- Let be the running variable z_i, population in local government (either population size in year $t-1$ or population size in 1918)
- The local polynomial regression, equivalent to the OLS estimation using only observations around the cut-off

$$Y_i = \alpha + \beta D_i + f(z_i) + u_i$$

- where $D_i = 1$ if local government has direct democracy
- β measures the effect of having direct rather than representative democracy
- The eligibility rule $T = 1$ if $z_i \leq 1500$ as an instrument for treatment status

TABLE II

DESCRIPTIVE STATISTICS[a]

Variables	Mean	St. Dev.	Min	Max	Obs.
Panel A. Outcome variables 1919–1938					
Per capita social welfare spending	6.31	4.06	0	59.96	48,128
Per capita spending on indoor relief	2.07	2.72	0	52.24	45,724
Per capita spending on outdoor relief	4.16	2.83	0	29.76	45,728
Percentage of organized citizens	9.04	18.0	0	198	48,152
Panel B: Forcing variables					
Population size at time $t-1$	1,717	2,004	91	26,491	48,164
Population in 1918	1,715	1,988	110	21,648	2,400
Panel C: Baseline or pre-treatment characteristics as measured in 1917 or 1918					
Per capita social welfare spending, 1918	2.48	2.10	0	41.41	2,398
Per capita spending on indoor relief, 1918	1.25	2.05	0	40.35	2,398
Per capita spending on outdoor relief, 1918	1.22	0.81	0	6.41	2,400
Percentage of organized citizens, 1917	7.59	12.3	0	270	2,380
Number of total recipients including children, 1917	58	104	0	1,714	2,400
Number of adult recipients, 1917	38	59	0	1,090	2,370
Number of children directly supported, 1917	7	15	0	289	2,370
Number of children indirectly supported, 1917	14	38	0	581	2,370
Number of people receiving full support, 1917	21	28	0	295	2,400
Number of people boarded out, 1917	8	13	0	139	2,370
Number of people in public institutions, 1917	13	20	0	196	2,370
Number of public institutions, 1917	0.76	0.58	0	8	2,400
Number of slots available in public institutions, 1917	19	24	0	200	2,400
Total area (km^2), 1918	18,160	81,181	0	1.95e+06	2,371
Land area (km^2), 1918	17,025	75,530	15	1.81e+06	2,371
Arable land (km^2), 1918	1,566	1,213	0	13,524	2,400
Total income tax base, 1918	195,656	452,911	786	6.10e+06	2,400
Economic structure (% agriculture), 1918	49.5	22.1	0	98.5	2,370
Number of eligible male voters at the parliamentary elections, 1917	359	371	0	4,373	2,400
Number of voters at the parliamentary elections, 1917	229	233	0	3,003	2,387
Proportion of left-wing voters at the parliamentary elections, 1917	0.30	0.20	0	1.00	2,380

[a] All nominal values are in SEK and deflated with CPI with 1914 as the base year.

Reduced Form

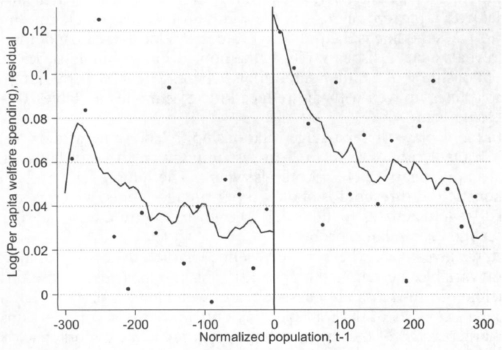


FIGURE 1.—Reduced-form relationship using population in year $t-1$ as the forcing variable. The dependent variable is the residual from a regression of per capita welfare spending on 21 covariates. Plotted points are conditional means with a binwidth of 20. The solid line is the predicted values of a local linear smoother with a rectangualar kernel and a bandwidth of 60.

First Stage

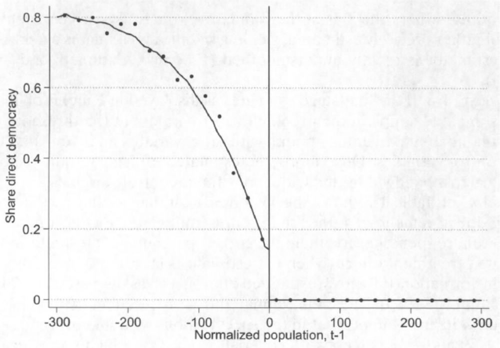


FIGURE 2.—First-stage relationship using population in year $t-1$ as the forcing variable. The dependent variable is an indicator variable for having direct democracy. Plotted points are conditional means with a binwidth of 20. The solid line is the predicted values of a local linear smoother with a rectangular kernel and a bandwidth of 60.

Wald Estimator

TABLE V

LOCAL LINEAR ESTIMATES FROM THE REGRESSION-DISCONTINUITY DESIGN WHEN THE FORCING VARIABLE IS POPULATION IN 1918[a]

Bandwidths:	20	30	40	50	60
Panel A: Reduced-form relationship					
Reduced-form effect	−0.536	−0.434*	−0.350	−0.287	−0.292
(no covariates)	(0.334)	(0.255)	(0.216)	(0.203)	(0.188)
Reduced-form effect	−0.461**	−0.412***	−0.422***	−0.379***	−0.272***
(including pre-treatment covariates)	(0.216)	(0.145)	(0.109)	(0.102)	(0.097)
Panel B: First-stage relationship					
First-stage effect	0.420***	0.319***	0.421***	0.392***	0.452***
(no covariates)	(0.129)	(0.114)	(0.116)	(0.108)	(0.106)
First-stage effect	0.453***	0.430***	0.422***	0.427***	0.472***
(including pre-treatment covariates)	(0.130)	(0.099)	(0.102)	(0.102)	(0.094)
Panel C: Wald or IV estimates					
Treatment effect	−1.274	−1.362*	−0.831	−0.732	−0.645
(no covariates)	(0.868)	(0.796)	(0.508)	(0.509)	(0.411)
Treatment effect	−1.017	−0.958**	−1.000***	−0.886***	−0.577**
(including pre-treatment covariates)	(0.630)	(0.453)	(0.370)	(0.315)	(0.233)
Number of municipalities	35	43	54	64	79
Number of observations	239	295	372	439	544

[a]Each entry is a separate local linear regression with a uniform kernel. All specifications allow for the RD slope to differ across the threshold and include a full set of time fixed effects. The dependent variable in Panels A and C is per capita welfare spending in logarithmic form. The dependent variable in Panel B is an indicator for having direct democracy rather than representative democracy. Panel C is the Wald estimator, the ratio between the reduced-form effect and the first-stage estimate. The forcing variable is population in year 1918. See the text for a description of included pre-treatment covariates. Standard errors, clustered at both the municipality level and the running variable, are within parentheses (Cameron, Gelbach, and Miller (2011)). Coefficients significantly different from zero are denoted by the following system: *10%, **5%, and ***1%.

McRary Test

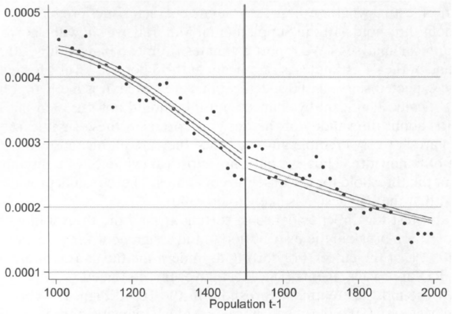


FIGURE 3.—The McCrary density test: population in year $t-1$ as the forcing variable. McCrary (2008) is a test of whether the density of the forcing variable, the population size in year $t-1$, is continuous at the population threshold. The point estimate is 0.0002 with a standard error of 0.0254.

Book References

- Angrist, Joshua D., and Jörn-Steffen Pischke. Mostly harmless econometrics: An empiricist's companion. Princeton university press, 2008.
- Angrist, Joshua D., and Jörn-Steffen Pischke. Mastering' metrics: The path from cause to effect. Princeton University Press, 2014.
- Cunningham, Scott. "Causal Inference: The Mixtape."