

Econometrics 2 – Part 1

Arieda Muço

Central European University

Winter 2023

Fuzzy Regression Discontinuity

Probability of receiving treatment does not change from 0 to one at the cutoff, but

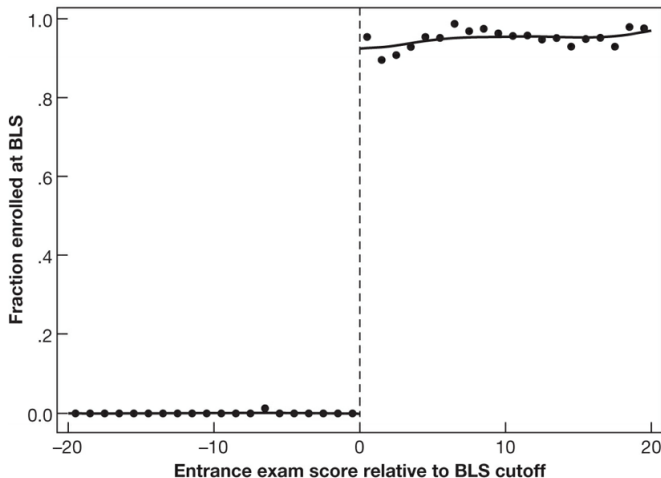
$$\lim_{z \uparrow z_0} P(D_i = 1 | z_i = z) \neq \lim_{z \downarrow z_0} P(D_i = 1 | z_i = z)$$

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

Enrollment at BLS



Fuzzy Regression Discontinuity is IV

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

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 $E(D_i = 1|z_i)$

Fig 3: Assignment Probabilities (Fuzzy RD)

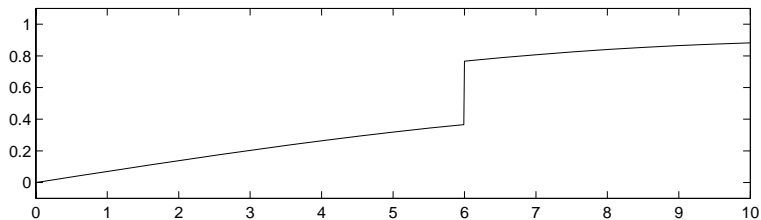
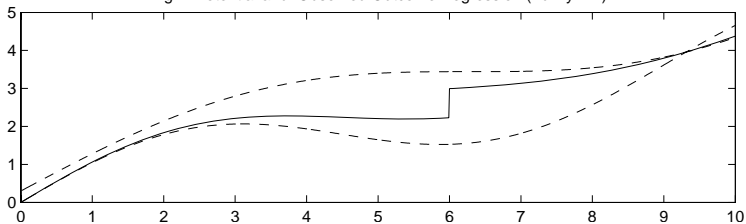


Fig 4: Potential and Observed Outcome Regression (Fuzzy RD)



Fuzzy RD as IV

- Use T_i as instrument for D_i .
- The first stage would be:

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

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

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

Fuzzy RD as IV - Second Stage

- As in the sharp RD case one can allow the smooth function to be different on both sides of the discontinuity
- Use both T_i as well as the interaction terms as instruments for D_i .
- The second stage model with interaction terms would be the similar to before.

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

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 as IV - First Stage

- Again one can use both T_i as well as the interaction terms as instruments for D_i

$$D_i = \gamma_{00} + \gamma_{01}\tilde{z}_i + \gamma_2\tilde{z}_i^2 + \gamma_3\tilde{z}_i^3 + \dots + \gamma_p\tilde{z}_i^p + \\ + \pi T_i + \gamma_1^* T_i \tilde{z}_i + \gamma_2^* T_i \tilde{z}_i^2 + \gamma_3^* T_i \tilde{z}_i^3 + \dots + \gamma_p^* T_i \tilde{z}_i^p + \eta_i$$

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.
- In other words, the cutoff is endogenous
- **Example:** Age thresholds used for policy (i.e., a person turns 18, and faces more severe penalties for a 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

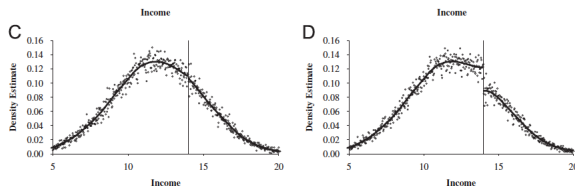


Figure: Panel C is density of income when there is no pre-announcement and no manipulation. Panel D is the density of income when there is pre-announcement and manipulation. From McCrary (2008).

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

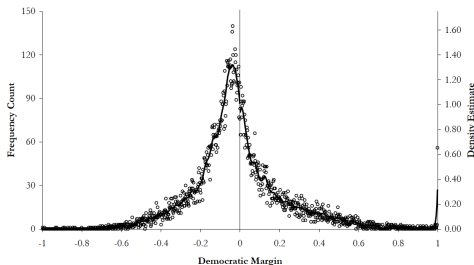


Figure: Democratic vote share relative to cutoff: popular elections to the House of Representatives, 1900-1990 (McCrary 2008).

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

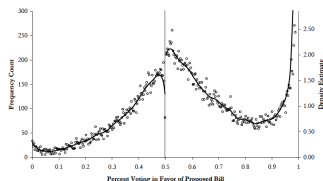
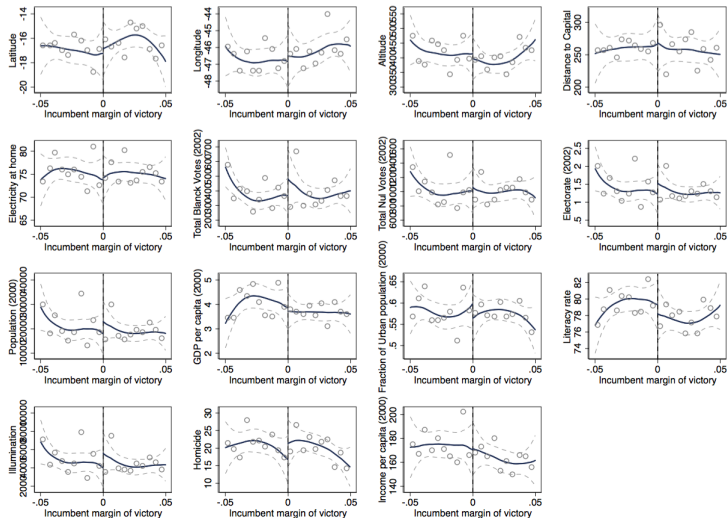


Figure: Percent voting Yeay: Roll Call Votes, US House of Representatives, 1857-2004 (McCrary 2008)

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 Covariate Balance



Jumps at non-discontinuous points

- Imbens and Lemieux (2008) suggest looking 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

Examples Fuzzy RD

- Tyrefors Hinnerich and Pettersson-Lidbom, Econometrica (2014)
- Does the form of Democracy matter?
- 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 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 democracy*; those with a population below 1,500 *direct democracy*
- If a local government had switched to a representative system, it could not switch back within a five-year period
- Running variables: Population in 1918 and Population in the year $t-1$ for the period from 1919

- 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

First Stage

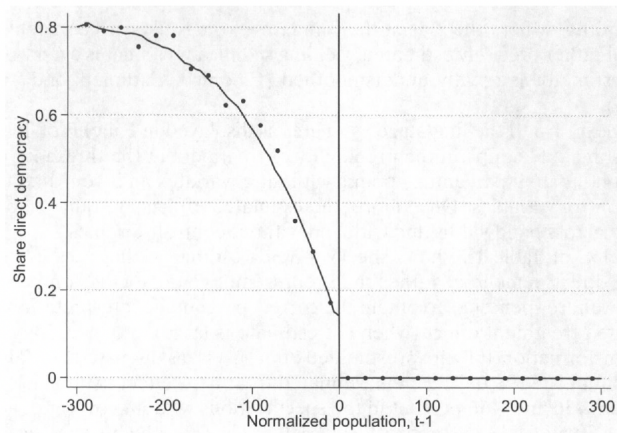


FIGURE 2.—First-stage relationship using population in year $t - 1$ as the forcing variable. The dependent variable is an indicator variable for having direct democracy. Plotted points are conditional means with a binwidth of 20. The solid line is the predicted values of a local linear smoother with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

Reduced Form

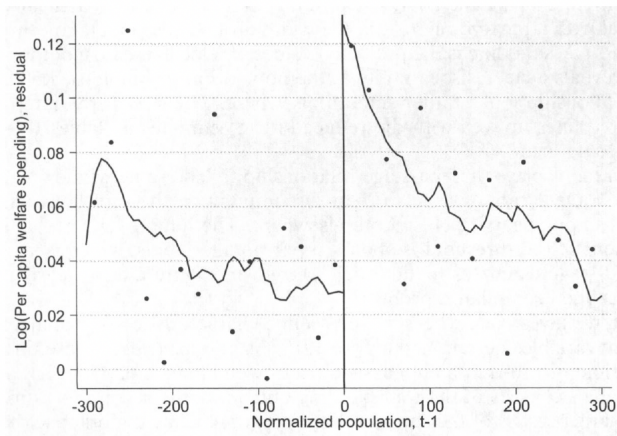


FIGURE 1.—Reduced-form relationship using population in year $t - 1$ as the forcing variable. The dependent variable is the residual from a regression of per capita welfare spending on 21 covariates. Plotted points are conditional means with a binwidth of 20. The solid line is the predicted values of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

Wald Estimates

LOCAL LINEAR ESTIMATES FROM THE REGRESSION-DISCONTINUITY DESIGN WHEN THE FORCING VARIABLE IS POPULATION IN YEAR $t - 1$ ^a

Bandwidths:	20	40	60	80	100	120
Panel A: Reduced-form relationship						
Reduced-form effect (no covariates)	-0.107* (0.058)	-0.075 (0.046)	-0.093** (0.042)	-0.117*** (0.037)	-0.084** (0.034)	-0.078** (0.034)
Reduced-form effect (including pre-treatment covariates)	-0.092** (0.036)	-0.093*** (0.028)	-0.101*** (0.031)	-0.114*** (0.029)	-0.089*** (0.027)	-0.083*** (0.025)
Panel B: First-stage relationship						
First-stage effect (no covariates)	0.140*** (0.030)	0.168*** (0.036)	0.165*** (0.037)	0.143*** (0.034)	0.154*** (0.034)	0.169*** (0.036)
First-stage effect (including pre-treatment covariates)	0.161*** (0.038)	0.183*** (0.039)	0.167*** (0.038)	0.148*** (0.034)	0.155*** (0.034)	0.168*** (0.035)
Panel C: Wald or IV estimates						
Treatment effect (no covariates)	-0.768* (0.446)	-0.445 (0.284)	-0.565** (0.260)	-0.819*** (0.291)	-0.549** (0.229)	-0.461** (0.209)
Treatment effect (including pre-treatment covariates)	-0.574** (0.230)	-0.511*** (0.173)	-0.604*** (0.207)	-0.771*** (0.241)	-0.572*** (0.191)	-0.492*** (0.165)
Number of local governments	158	193	232	252	274	296
Number of observations	520	1,021	1,535	2,074	2,608	3,113

^aEach 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 $t - 1$. 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

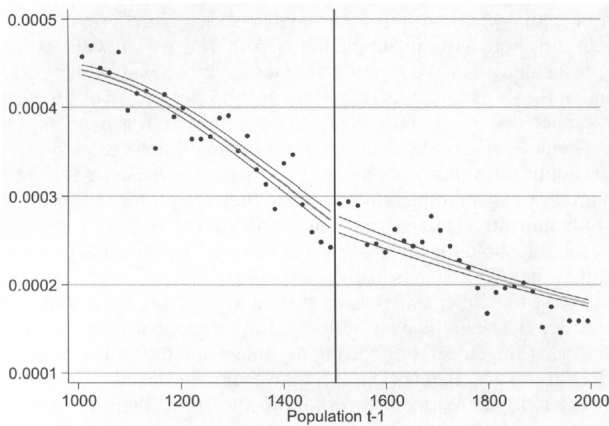


FIGURE 3.—The McCrary density test: population in year $t - 1$ as the forcing variable. McCrary (2008) is a test of whether the density of the forcing variable, the population size in year $t - 1$, is continuous at the population threshold. The point estimate is 0.0002 with a standard error of 0.0254.