Evaluating the effects of university grants using regression discontinuity designs

Fabrizia Mealli   Carla Rampichini

Dipartimento di Statistica "Giuseppe Parenti"
Università di Firenze

Bristol June 29th, 2011

Outline

1. Evaluation of Italian University grants
2. Causal inference and potential outcomes
3. Eligibility, grant assignment and causal effect
   - Sharp RD designs
   - Identification of treatment effects via RD designs
   - Implementing RD in practice
   - Estimation
4. Estimation of the grant effect
5. Extending the result: Difference-In-Difference assumptions

Evaluation of Italian University grants

- The Italian State universities offer some grants every year to eligible freshmen (HS grade ≥ 70, economic indicator $S \leq \widetilde{s}$)
- The main objective of this intervention is to give equal opportunity to achieve higher education to motivated students irrespective of their income.
- We investigate whether the grant is an effective tool to **prevent students from low-income families from dropping out of higher education**.
- Data come from a larger research project, conducted for the Department of Education (Biggeri and Catalano, 2006).

Causal inference I

- The research questions that motivate most studies in statistics-based sciences are causal in nature.
- The aim of standard analysis is to infer associations among variables, estimate the likelihood of past and future events, as well as update the likelihood of events in light of new evidence or new measurements, using samples drawn of that population. These tasks are managed well by statistical analysis so long as experimental conditions remain the same.
- Causal analysis goes one step further; its aim is to infer aspects of the data generating process.
- With the help of such aspects, one can deduce not only the likelihood of events under static conditions, but also the dynamics of events under changing conditions.

Causal inference II

- The additional information needed for making such predictions is provided by causal assumptions.
- The role of this information is to identify those aspects of the world that remain invariant when external conditions change, say due to treatments or policy decisions.
- Correlation does not imply causation: one cannot substantiate causal claims from associations alone, even at the population level
- behind every causal conclusion there must lie some causal assumption that is not testable.

The potetial outcomes approach

- A statistical framework for causal inference that has received increasing attention in recent years is the one based on potential outcomes.
- This approach is rooted in the statistical work on randomized experiments by Fisher (1918, 1925) and Neyman (1923), as extended by Rubin (1974, 1976, 1977, 1978, 1990) and subsequently by others to apply to nonrandomized studies and other forms of inference.
- This perspective was called Rubin's Causal Model (RCM, Holland, 1986) by because of it viewed causal inference as a problem of missing data with explicit mathematical modeling of the assignment mechanism as a process for revealing the observed data.

The Fundamental Problem of Causal Inference

- Y outcome: drop-out or not
  - Let Y_0 denote the outcome given the control treatment
  - and Y_1 the outcome given the active treatment (i.e. grant assignment)
- *We can observe at most one of the potential outcomes for each unit*
- the individual causal effect

$$\beta = Y_1 - Y_0$$

is *not observable*

The average treatment effect

- Because individual causal effects can never be observed, inference usually focuses on average causal effects, e.g.

$$ATE = E(Y_1 - Y_0)$$

- Under randomisation, ATE is estimated by comparing the group of treated subjects with a group of untreated subjects, usually called the control group.
- By itself, however, the presence of multiple units does not solve the problem of causal inference

The Stable Unit Treatment Value Assumption

- When using multiple units, we need the potential outcomes to remain a function of treatment assignment for a given unit, irrespective of assignment of other units.
- This assumption is the so called SUTVA (Stable Unit Treatment Value Assumption), and will be assumed to hold throughout the paper.
- SUTVA states that potential outcomes for an individual are unaffected by the treatment assignments of other individuals (Rubin, 1980).

Randomisation I

- The comparison of treated and untreated subjects leads to unbiased estimation of the average treatment effects if, in expectation, the only difference between the two groups is the treatment assignment.
- This is true if treatment assignment is randomised.
  - Let T denote the *treatment status*: 1 treated, 0 untreated
  - Formally, if assignment to treatment is randomised and treatment assignment coincides with treatment received, i.e. subjects comply with assignment

$$\Downarrow$$

we can assume that (Y_1, Y_0) are independent of T, i.e. $(Y_1, Y_0) \perp\!\!\!\perp T$

Randomisation II

- Under randomisation ATE can be estimated by examining the difference between the sample averages of the observed values for treated and untreated subjects

$$Y^{obs} = Y_1 T + Y_0(1 - T)$$

- given that

$$ATE = E(Y_1 - Y_0) = E(Y_1 \mid T = 1) - E(Y_0 \mid T = 0)$$

- That is, randomisation allows one to use information on untreated subject to identify the mean potential outcome for treated subjects, and viceversa, because, under randomisation, conditioning on T is irrelevant.

Eligibility, grant assignment and causal effect

- The **evaluation** of the grant effect is **quite complicate** ...
  - only some of the eligible students apply for a grant (i.e. application is not mandatory but voluntary) $\Rightarrow$ **self-selection**
  - due to **budget constraints**, only some of the applicants receive the treatment, i.e. applicants with S below an *ex-post* threshold $\hat{s}$, $\hat{s} < \tilde{s}$.
- ... we can exploit the **assignment rule** to identify the grant effect
  - the applicants cannot precisely manipulate S around $\widehat{s} \Rightarrow$ treatment is **locally randomized**
  - **RD Designs**: the discontinuity gap at the ex-post threshold $\widehat{s}$ identifies the treatment effect of interest (Lee and Lemieux, 2009).

Data I

- In our application, the probability of receiving the grant depends on a family income indicator S.
- The Table reports the *ex-ante* and *ex-post* eligibility thresholds (in Euros) along with the minimum and maximum grant amounts.

Table: Economic *ex-ante* eligibility thresholds for a reference family of three members and grant amount range (Euros). Academic year 1999.

University	*Ex-ante thresholds*		*Ex-post*	*Grant amount*	
	annual income	*assets*	*threshold* $\hat{s}$	*min*	*max*
Catania	23240.56	61974.83	12912.00	1394.43	2065.83
Milano	26734.55	69509.82	30374.00	1343.00	3460.00
Padova	23034.22	61515.18	15812.51	1551.44	3602.30
Salerno	25882.84	67139.40	12214.00	2995.45	3460.26

Data II

- We consider freshmen enrolled in 1999, who live out of town and cannot commute, and meeting the ex-ante eligibility criteria
- We define the *ex-post* eligibility status of a student by considering the *ex-post* thresholds reported in the previous Table: $Z = \mathbb{I}[S < \hat{s}]$.

Application status	*ex-post eligibility status.*					
	Catania		*Milano*		*Padova*	
	$Z = 0$	$Z = 1$	$Z = 0$	$Z = 1$	$Z = 0$	$Z = 1$
$A = 0$	740	640	308	249	130	46
$A = 1$	425	536	27	272	246	306
Total	1165	1176	335	521	376	352

- In general, to evaluate the grants causal effect, we would like to **compare** students who received a grant with students who did not, all other things being equal.

Which comparison group?

- The **voluntary** nature of the **application** and the **assignment rule** divide the population of *ex-ante* eligible students into four subgroups

Application status	*Eligibility* status	
	$Z = 0$ $(S \geq \hat{s})$	$Z = 1$ $(S < \hat{s})$
$A = 0$	ineligible non-applicants	eligible non-applicants
$A = 1$	ineligible applicants	*participants* $(\Rightarrow T = 1)$

1. we could compare non-applicants who meet the ex-ante and ex-post eligibility criteria (those in the upper right corners) with applicants who receive the grant (*participants*).
   - However, this comparison may be *confounded by the voluntary application process* and by the consequent possible differences in unobservable characteristics that relate to dropout.
2. or we could compare applicants who receive a grant (*participants*) with applicants who do not receive a grant (more plausible comparison).

Comparison between applicants who receive a grant and applicants who do not receive a grant

The grant assignment rule makes these two groups very different and not directly comparable, because they have different and *non-overlapping values of the economic indicator S*.

Eligibility rule

Students who apply for a grant receive the grant, provided that they have a value of the economic indicator S below the cutoff $\hat{s}$

- The participation status changes deterministically with S, i.e., the probability of being treated for applicants with $S < \hat{s}$ is 1, while the probability of being treated for applicants with $S > \hat{s}$ is 0.
- the **probability of participation is discontinuous at the threshold** $\widehat{s} \Rightarrow$ a **sharp RDD** is defined for the applicants, even if students have self-selected into the programme.

Regression discontinuity (RD) designs

RD methods aim to evaluating causal effects of interventions where assignment to a treatment is determined (at least partly) by the value of an *observed covariate* lying on either side of a cutoff point

- First introduced by Thistlewaite and Campbell (1960)
- RD methods did not attract much attention in the economics literature until recently
- From late 1990s there has been a growing number of studies in economics applying and extending RD methods (Angrist and Lavy, 1999; Black, 1999; Hahn et al., 2001; Van der Klaauw, 2002; Lee, 2007).

Basic features of RDDs I

The treatment status T is *not randomized*, but depends on an observable continuous random variable S.

Moreover, there exists a known cutoff point $\hat{s}$ in the support of S where the probability of treatment received changes discontinuously, i.e.,

$$P(T = 1 \mid \hat{s}^{+}) \neq P(T = 1 \mid \hat{s}^{-})$$

where $Pr(T = 1 \mid \hat{s}^{+}) = \lim_{s \to \hat{s}^{+}} = P(T = 1 \mid S = \hat{s})$, and $Pr(T = 1 \mid \hat{s}^{-}) = \lim_{s \to \hat{s}^{-}} = P(T = 1 \mid S = \hat{s})$.

This assignment rule determines a RDD.

The basic idea underlying the RDD is that subjects below and above the threshold c are similar, so that a **quasi-randomization** can be assumed **around the threshold**.

Basic features of RDDs II

- The basic idea underlying the RDD is that subjects just below and just above the threshold are similar, so that a quasi-randomisation can be assumed around the threshold.

RRDs caveats I

- The existence of a *treatment* being a *discontinuous function* of an assignment variable is *not sufficient* to justify the validity of an RD design: RD designs can be invalid if individuals can *precisely manipulate* the assignment variable.
- If individuals even while having some influence are **unable** to *precisely manipulate* the assignment variable⇒ the variation in treatment *near the threshold* is **randomized** as though from a randomized experiment
- RD designs can be analyzed and tested like **randomized experiments**.
- It follows that all baseline characteristics (i.e. determined prior to the realization of the assignment variable) should have the *same distribution* just above and just below the cutoff

RRDs caveats II

- If there is a discontinuity in these baseline covariates, then at a minimum, the underlying identifying assumption of individuals inability to precisely manipulate the assignment variable is unwarranted.
- Thus, the *baseline covariates* are used to *test the validity* of the RD design.

Two Cases

Usually two special cases of RDD are considered

- Sharp RDD: all units with values on one side of the cutoff value are exposed to the program, and all units on the other side of the cutoff value are not exposed.
- Fuzzy RDD: around the cutoff value the participation (exposure) rate changes sharply, but not from zero to one.

Sharp RDD

- Eligibility rule: individuals who apply for a grant receive the grant, provided that they have an economic index equal to a given *cutoff* or below
  $\Rightarrow$ the participation status changes deterministically with the economic index
- Policy question: does participation affect the dropout probability?

Sharp Regression Discontinuity for the Grants case

Y_i is dropout status for the i-th student.

S_i is the economic indicator for the i-th student.

$\hat{s}$ is cutoff value

$$T_i = \begin{cases} 0 & \text{if } S_i > \hat{s} \\ 1 & \text{if } S_i \leq \hat{s} \end{cases}$$

- A student receives the grant if her economic indicator is less than $\hat{s}$.
- The economic index is likely to be correlated with the outcome of interest and, by design, is the only variable that determines the treatment status.

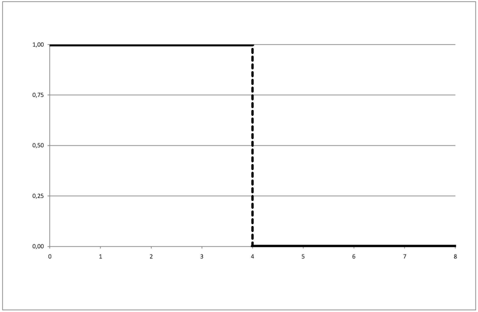
The Sharp Regression Discontinuity Design

The treatment is assigned as a *deterministic function* of an observed covariate which is also related to the outcome of interest

$$T_i = 1\{S_i \leq s\}.$$

In our setting, all the students above the threshold are denied participation, while all individuals below the threshold receive the grant, i.e. are *treated*.

$\Rightarrow$ participation probability jumps from zero to one as the selection variable crosses the threshold



Sharp RDD I

- It seems a good idea to match participants to non-participants similar with respect to values of this variable (selection on observables)
- It follows that the effect of participation on the outcome of interest can be estimated only for individuals in the common support, that is *around the cutoff*
- We consider the **applicants** to estimate the grant effect at the threshold $\hat{s}$:

$$E(Y_1 - Y_0 \mid A = 1, \hat{s}) = E(Y_1 \mid A = 1, \hat{s}^-) - E(Y_0 \mid A = 1, \hat{s}^+).$$

Y_1: outcome when $Z = 1$ (treated), Y_0: outcome when $Z = 0$.
▶ The probability of being treated for applicants with $S < \hat{s}$ is 1, while the probability of being treated for applicants with $S > \hat{s}$ is 0.

$$Pr(T = 1 \mid A = 1, Z = 1) - Pr(T = 1 \mid A = 1, Z = 0) = 1$$

Sharp RDD II

- the **probability of participation is discontinuous at the threshold** $\widehat{s}$

$\Rightarrow$a **sharp RDD** is defined for the applicants, even if students have self-selected into the programme.

Causal effect identification

The average treatment effect at $\hat{s}$, is identified as:

$$\tau = \mathbb{E}(Y_1 - Y_0 \mid S = \hat{s}) = \mathbb{E}(Y_1 \mid \hat{s}^+) - \mathbb{E}(Y_0 \mid \hat{s}^-)$$

where $\mathbb{E}(Y_1 \mid \hat{s}^+) = \lim_{s \to \hat{s}^+} \mu_1(S)$, and $\mathbb{E}(Y_0 \mid \hat{s}^-) = \lim_{s \to \hat{s}^-} \mu_0(S)$,

★In the SRD design we look at the discontinuity in the conditional expectation of the outcome given the forcing variable

$$\lim_{s \to \hat{s}^+} \mathbb{E}[Y_i | S_i = s] - \lim_{s \to \hat{s}^-} \mathbb{E}[Y_i | S_i = s]$$

$$= \lim_{s \to \hat{s}^+} \mu_1(s) - \lim_{s \to \hat{s}^-} \mu_0(S)$$

which is interpreted as *average causal effect* τ of the treatment *at* the discontinuity point $\hat{s}$.

Smoothness

- A sufficient condition to identify the causal effect around the threshold is that *the expected values of Y_0 and Y_1 conditionally on S are continuous functions of S in $\hat{s}$.*
- This condition can be tested indirectly comparing eligible and ineligible non-applicants around the cutoff point $\hat{s}$.
- If fact, for the non-applicants, Y_0 can be observed above and below $\hat{s}$, i.e. $Y^{obs} = Y_0$ for all values of S, so that the continuity of the conditional mean value of Y_0 given S at $\hat{s}$, i.e.

$$\lim_{s \to \tilde{s}^+} E(Y_0 \mid A = 0, S = s) = \lim_{s \to \tilde{s}^-} E(Y_0 \mid A = 0, S = s)$$

can be tested.

Implementing RD designs in practice

- Estimation
- Graphical identification of the effect
- Choice of the regression model
- Tests to assess the validity of the RD designs
  - balanced covariates on the two sides of the threshold
  - density of the assignment variable continuous at the threshold

Estimation of causal effects in a sharp RDD

Estimation of causal effects in a sharp RDD can be formalized as a problem of estimating the **two conditional means** of the outcome variable at each side of the *cutoff* point $\hat{s}$.

Statistical Question: How to estimate objects like

$$\lim_{s \to \hat{s}^+} \mathbb{E}[Y_i | S_i = s]$$

- Graphical analysis
- Differences between average outcome for $T_i = 1$ and $T_i = 0$ groups
- Local linear regression

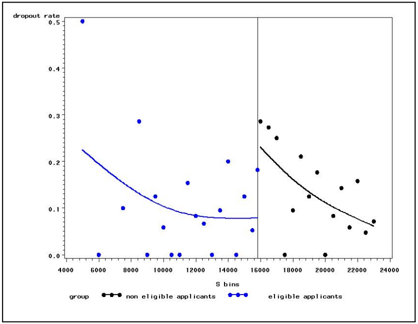
Graphical analysis

A first indication of the treatment effect can be obtained by means of a graphical representation (Lee &Lemieux,2009)

- Put observations in bins.
- Calculate average outcomes within bins, and see whether there is a jump around the cutoff value.

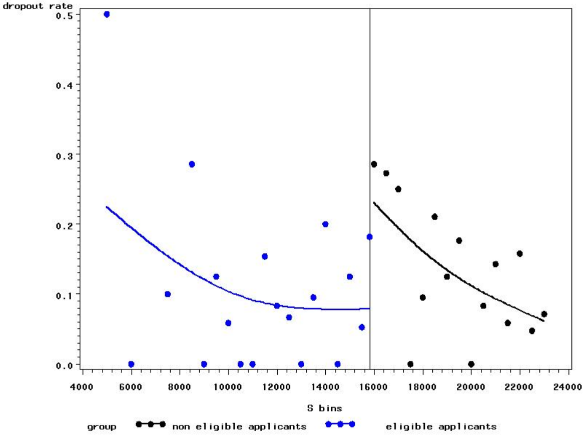
In our analysis we have:

- created bins for S_i (economic indicator).
- computed the proportion of dropouts for students where the economic indicator is in a particular bin.
- Plot the dropout rate by bin.



Dropout proportion by application and eligibility status

Comparing the dropout rate just to the left and right sides of the threshold for applicants we obtain an indication of the grant effects.



Indentifiability of causal effect: sharp RDD

- In presence of heterogeneous effects, a RDD allows identification of an *average treatment effect* for (a subset of) applicants *around the threshold* $\widehat{s}$.
- A sufficient condition to identify the causal effect around the threshold is that *the expected values of Y_0 and Y_1 conditionally on S are continuous functions of S in $\hat{s}$.*
- This condition can be tested indirectly for non applicants.

Continuity analysis for non-applicants I

- We can compare eligible and ineligible non-applicants around the cutoff point $\hat{s}$.
- If fact, for the non-applicants, Y_0 can be observed above and below $\hat{s}$, i.e. $Y^{obs} = Y_0$ for all values of S, so that the continuity of the conditional mean value of Y_0 given S at $\hat{s}$, i.e.

$$\lim_{s \to \tilde{s}^+} E(Y_0 \mid A = 0, S = s) = \lim_{s \to \tilde{s}^-} E(Y_0 \mid A = 0, S = s)$$

  can be tested.
    - If the null hypothesis of continuity on non-applicants is rejected, then it will be hard to assume that the continuity assumption holds for applicants.
    - On the other hand, by not rejecting the null hypothesis for non-applicants, one might feel more confident in assuming continuity for the applicants and in trusting the estimates obtained with the sharp RDD.

Continuity analysis for non-applicants II

- In a neighborhood of $\hat{s}$, any test of the equality of the mean outcomes of non-applicants below and above $\hat{s}$ is a test for the continuity of $E(Y_0 \mid A = 0, S)$ at $\hat{s}$.
- We used a semi-parametric estimator of the discontinuity of the conditional mean value of Y^{obs} at $\hat{s}$ (see paper for details).
- The jump is never significant, thus the continuity assumption is not rejected.

Table: Continuity analysis for non-applicants ($A = 0$) - polynomial regression of dropout.

University	h_{opt}	$n\ obs$	K	$\widehat{\tau}$	$s.e.$
Catania	912	110	1	0.006	0.203
Milano	1319	51	1	0.075	0.246
Padova	1477	15	1	0.049	0.661

Continuity analysis for non-applicants III

- In order to support the plausibility of the assumptions underlying the sharp RDD, we also test the continuity assumption at $\hat{s}$ of the conditional means of all the observed covariates for applicants.
- No discontinuity of the conditional means of any of the covariates was found.

Estimation of the grant effect via sharp RDD

- Estimation of the causal effect in the RD designs can be formalized as a problem of estimating the **two conditional means** of the outcome at each side of the cutoff point $\hat{s}$.
- This can be viewed as a standard non parametric regression problem with the peculiarity that the interest is on the *evaluation of the regression function at a single point* and that this point is a *boundary point* $\Rightarrow$ the simple kernel estimator is biased (Porter, 2003).
- To reduce the bias, usually local linear or polynomial regressions with an optimal choice of the bandwidth are used (Imbens and Kalyanaraman, 2009; Lee and Lemieux, 2009).

Non parametric estimation of the grant effect

Local polynomial regression (Hahn et al., 2001; Porter, 2003) calls for three important choices (Lee and Lemieux, 2009):

- the **kernel**: we consider a rectangular kernel, i.e. a standard regression over an interval of width h on both sides of the threshold $\hat{s}$;
- the **bandwidth** h: we try a range of bandwidth going from the whole range, i.e. 16000 on each side, to a small bandwidth of 1000 euros;
- the **polynomial terms** to be included in the regression model

$$E(Y \mid S) = \alpha_r + \tau \cdot T + \sum_{k=1}^{K} \beta_{rk} \cdot (S - \hat{s})^k + \sum_{k=1}^{K} \delta_k \cdot T \cdot (S - \hat{s})^k,$$

with $\hat{s} - h \leq S \leq \hat{s} + h$, $\delta_k = \beta_{lk} - \beta_{rk}$, and βs are the parameters of the polynomial terms on the left and right sides of $\widehat{s}$.

Simple comparison

The simple model with zero order polynomial and a large bandwidth is equivalent to the difference of the averages of the outcome computed using observation at each side of the threshold

Simple Comparison: Calculate difference between average outcome for $T_i = 1$ and $T_i = 0$ groups
This amounts to estimating a regression function

$$Y_i = \alpha + \tau \cdot T_i + \varepsilon_i$$

assuming random assignment of grants.

Non parametric estimation of the grant effect 2

To choose the *bandwidth* and the *polynomial*, we follow the two step procedure suggested by Lee and Lemieux (2009)

1. estimate a Rule-Of-Thumb bandwidth over the whole range of S.
2. use the ROT bandwidth to estimate the optimal bandwidth h_{ROT} right at the threshold.

- To choose among h_{ROT}s, we considered the h_{OPT} bandwidth found following the *optimal bandwidth choice* procedure suggested by Imbens and Kalyanaraman (2009)
- the polynomial is chosen comparing models *via AIC*

Sharp RDD on applicants

- We estimated the grant effect at $\hat{s}$, so that τ can be interpreted as the effect of the treatment on the dropout rate at the cutoff point.
- For all the universities but Padova, the effect of the grant at the threshold is not significant, irrespective of the choice of the bandwidth and polynomial order.
- The following Table shows the estimates corresponding to a common bandwidth of 4000 Euros.

Sharp RDD results I

Pooled regression (robust s.e) for varying bandwidth and polynomial degree.

Table: Sharp RDD on applicants: pooled regression (robust s.e) with bandwidth $h = 4000$ for varying polynomial degree (*Value corresponding to optimal polynomial degree following AIC in bold*)

University	*polynomial order*						N
	0		1		2		
	$\widehat{\tau}$	*s.e.*	$\widehat{\tau}$	*s.e.*	$\widehat{\tau}$	*s.e.*	
Milano	**0.029**	0.039	-0.004	0.072	-0.016	0.098	54
Catania	**-0.018**	0.042	-0.001	0.101	-0.171	0.170	311
Padova	-0.066	0.041	**-0.132**	0.078	-0.186	0.112	268

Sharp RDD results II: University of Padua I

- To evaluate the robustness of the results for the University of Padova, we replicate the estimates for
  - a range of bandwidths from the optimal bandwidth of 793 euros to a bandwidth of 10000 Euros
  - and for orders of the polynomial up to two (higher orders are never chosen by the AIC).

h	*polynomial order* 0		1		2		*n obs*
	$\widehat{\tau}$	*s.e.*	$\widehat{\tau}$	*s.e.*	$\widehat{\tau}$	*s.e.*	
793	**-.200**	0.102	-.071	0.209	-.153	0.358	62
1000	**-.151**	0.092	-.201	0.187	0.023	0.296	79
2000	-.095	0.060	**-.184**	0.111	-.231	0.171	142
3000	**-.074**	0.048	-.141	0.093	-.248	0.139	214
4000	-.066	0.041	**-.132**	0.078	-.186	0.112	268
6000	-.050	0.034	**-.101**	0.062	-.184	0.090	369
8000	-.039	0.030	**-.105**	0.055	-.145	0.080	442
10000	-.042	0.029	**-.102**	0.052	-.159	0.076	462

Sharp RDD results II: University of Padua II

- As expected, the precision of the estimates decreases with the bandwidth.
- The estimates are rather stable in terms of significance, even if their values range from -7% to -20 % (when we consider the optimal polynomial order for each bandwidth).
- The effect at $h_{OPT} = 793$ Euros is about -0.20.

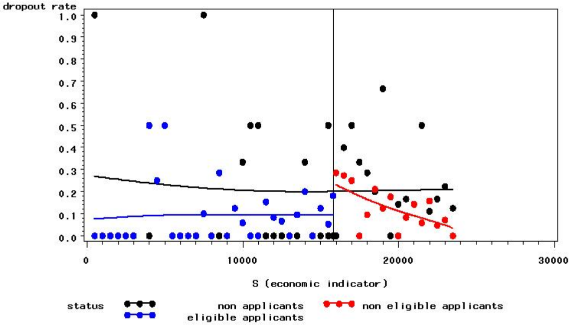
The analysis *at the threshold* shows that the grant **reduces the drop out rate** of about **20%** points.

Using non applicants to model selection bias I

How can we estimate the effect of the grant **below the threshold**, i.e. for poorer students?

- Applicants and non applicants are different, also controlling for observable covariates, so it is not possible to use matching techniques under the unconfoundedness assumption of application.

Using non applicants to model selection bias II



- The *selection bias* arising from non random selection of applicants is *identified above the threshold* for non eligible students.
- We can formally test whether any of the existing non-experimental estimator can correct for this bias.

Using non applicants to model selection bias III

- If the hypothesis is not rejected for the non eligible, one may *feel more confident* to use non-experimental estimators to identify causal effects for the *eligible applicants*.
- We correct the *difference* between applicants and non applicants *below the threshold*, using the observed *difference* in Y_0, i.e. between applicants and non-applicants *above the threshold* $\hat{s}$ *(This difference is modelled as a function of* S*, see paper)*.

The basic assumption of this approach, with the obvious modifications to conform with the RDD notation and setting, is that, in the absence of treatment, the following condition holds:

Using non applicants to model selection bias IV

Condition

The difference in the average outcome between applicants and non-applicants in the absence of the treatment is constant

$$E(Y_0 \mid Z, A = 1) - E(Y_0 \mid Z, A = 0) = c$$

i.e., it does not depend on the eligibility status, Z.

Using non applicants to model selection bias V

Under *Condition 1*, the causal effect is identified as follows:

$$
\begin{aligned}
E(Y_1 \mid T=1) \;-\; & E(Y_0 \mid T=1) \\
= \;& E(Y_1 \mid Z=1, A=1) - E(Y_0 \mid Z=1, A=1) \\
= \;& E(Y_1 \mid Z=1, A=1) - E(Y_0 \mid Z=1, A=1) \\
& -E(Y_0 \mid Z=1, A=0) + E(Y_0 \mid Z=1, A=0) = \\
= \;& [E(Y_1 \mid Z=1, A=1) - E(Y_0 \mid Z=1, A=0)] + \\
& -[E(Y_0 \mid Z=1, A=1) - E(Y_0 \mid Z=1, A=0)] = \\
= \;& [E(Y_1 \mid Z=1, A=1) - E(Y_0 \mid Z=1, A=0)] + \\
& -[E(Y_0 \mid Z=0, A=1) - E(Y_0 \mid Z=0, A=0)]
\end{aligned}
$$

Condition 1 can be weakened by assuming the following alternative Condition 2:

Using non applicants to model selection bias VI

Condition

The difference between applicants and non-applicants in the absence of the treatment is a function of S*, i.e.:*

$$E(Y_0 \mid S, A = 1) - E(Y_0 \mid S, A = 0) = f(S)$$

with a functional form such that:

$$f(S) = g(f_1(S))\mathbb{I}(S < \hat{s}) + f_1(S)\mathbb{I}(S \geq \hat{s})$$

where $g(.)$ *is the identity function.*

Grant effect estimation under DID assumptions

1. logit model for Y on (ex-post) eligibles conditionally on A, S, and $A * S$.
2. logit model for Y_0 on non eligibles conditionally on A, S, and $A * S$.
3. the coefficients of model 2 are used to *correct* the parameters of model 1

S	p_1	$p_1 - p_0$	$s.e.(p_1 - p_0)$
6000	0.16514	0.05966	0.16470
8000	0.13235	0.02690	0.14003
10000	0.10525	-0.00018	0.11839
12000	0.08317	-0.02223	0.10125
14000	0.06538	-0.03999	0.09091
16000	0.05119	-0.05416	0.08966

$p_k = \widehat{P}(Y_k = 1 | Z = 1, A = 1, S),\ k = 0, 1$

▶ the grant effect decreases with S.

Conclusions I

- Results show that,*at the threshold*, the grant is an *effective tool* to prevent students from low income families from dropping out of higher education.
- Under some relatively weak nonlinear *DID* type of assumptions, results show that moving *below the threshold*, i.e. for poorer students, the impact of the grant becomes smaller and *not significant*
  ▶ grants do **not** seem to be **effective** in changing the decision of the **poorest students** to abandon their university studies.

- The applicants not receiving the grant and the non applicants allowed to characterize the *selection bias* and to exploit DID assumptions to estimate the grant effect for a larger group of eligibles.
- In quasi-experimental designs it is crucial to explore the information embedded in the data.

Conclusions II

- In this application we exploited the partition of students induced by the application status and the assignment rule.
- The proposed tools could be applied in similar frameworks, whenever the participation threshold is different from the eligibility threshold and the forcing variable S is observable for applicants and non applicants.

References I

Battistin E. and Rettore E. (2008). Ineligibles and Eligible Non-Participants as a Double Comparison Group in Regression Discontinuity Designs, *Journal of Econometrics*, 142, 715-730.

Biggeri L. and Catalano G. EDS (2006). *L'efficacia delle politiche di sostegno agli studenti universitari*, Bologna: Il Mulino.

Hahn, J., Todd, P.E., van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69, 201-209.

Imbens G. W. and Kalyanaraman K. (2009). Optimal Bandwidth Choice for the Regression Discontinuity estimator, *IZA*, Discussion Paper, n. 3995.

Imbens G. W. and Lemieux T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615-635.

Lee D.S. and Lemieux T. (2009). Regression Discontinuity Designs in Economics *NBER*, Working paper n. 14723.

Porter J.R. (2003). Reducing bias in observational Estimation in the Regression Discontinuity Model. Harvard Institute of Economic Research, *Research Paper Series*.

References II

Rubin, D.B. (1980). Discussion of "Randomization analysis of experimental data: the Fisher randomization test" by D. Basu. *Journal of the American Statistical Association*, 75, 591-593.