Evaluating the Effect of University Grants on Student Dropout: Evidence from a Regression Discontinuity Design Using Bayesian Principal Stratification Analysis

Fan Li
fli@stat.duke.edu
Department of Statistics
Duke University

Joint work with Fabrizia Mealli and Alessandra Mattei

February 6, 2015
Motivation: Evaluation of Italian University Grants

- Italian state universities offer some grants every year to eligible students.
- **Objective:** give equal opportunity to achieve higher education to motivated students irrespective of their financial background.
- **Grant allocation rule:** A student must (1) meet the eligibility criteria; and (2) apply for the grant.
- **Eligibility:** an economic measurement of the student’s family income and assets (forcing variable) falling below or above a pre-determined threshold.
- Application is voluntary, and also ineligible students may apply.
Challenges

- Application is not mandatory but voluntary ⇒ self-selection: Applicants and non-applicants may be different in important but unmeasured ways
- Due to budget constraints, only some of the applicants receive the grant: Poor students and rich students may be different in important but unmeasured ways
Regression discontinuity design (RDD)

- Our strategy: exploit the assignment rule to identify the grant effect
- The assignment rule: only applicants with the economic measure $S$ below a fixed threshold $s_0$ get the grant
- The threshold is set by the administration, not the applicants $\Rightarrow$ eligibility can be viewed as locally randomized at $s_0$ and the students are comparable around $s_0$
- There is a discontinuity in the treatment probability at $s_0$
- The discontinuity gap in the outcome at $s_0$ can be interpreted as the treatment effect at $s_0$
- Regression discontinuity design (RDD)
Regression discontinuity design (RDD)

- First introduced by Thistlewaite and Campbell (1960) in psychology
- Regain popularity since the 90s, especially in economics (e.g., Imbens and Lemieux, 2008, Lee and Lemieux, 2010)
- RDD: the probability of receiving the treatment changes discontinuously at a cutoff point $s_0$ of a covariate $S$ (the forcing variable):

$$\Pr(W = 1 \mid s_0^+) \neq \Pr(W = 1 \mid s_0^-),$$

where $\Pr(W = 1 \mid s_0^+) = \lim_{s \to s_0^+} \Pr(W = 1 \mid S = s)$, and

$\Pr(W = 1 \mid s_0^-) = \lim_{s \to s_0^-} \Pr(W = 1 \mid S = s)$

- Two types: Sharp RDD (SRD) and Fuzzy RDD (FRD)
- Hahn, Todd and van der Klaauw (2001) first related fuzzy RDDs to instrumental variables
Graphical illustration: SRD

Fig. 1. Assignment probabilities (SRD).

Fig. 2. Potential and observed outcome regression functions.
Graphical illustration: FRD

Fig. 3. Assignment probabilities (FRD).

Fig. 4. Potential and observed outcome regression (FRD).
Graphical Illustration: Italian University Grants

Application Rate

Dropout Rate

□ Non-Applicants

• Applicants
Goals

- **Goal 1**: Evaluate if the grant is effective to prevent students from low-income families from dropping out of university.

- **Goal 2**: Provide a probabilistic formulation of the assignment mechanism underlying RDDs using the *principal stratification* framework *(Frangakis and Rubin, 2002)*.

- **Goal 3**: Show how to capitalize on the application data to draw extra causal information.

- **Goal 4**: Develop a Bayesian approach to conduct exact causal inference for fuzzy RDDs.
  - Bayesian paradigm is particularly useful to draw inference in complex settings like RDDs.
  - Bayesian methods are attractive in dealing with small samples in RDDs.
  - Inference about finite-sample and super-population estimands can be drawn using the same inferential procedures.
The Potential Outcome Approach to Causal Inference
(Neyman, 1923; Fisher, 1925; Rubin, 1974, 1978)

- $Z_i =$ Eligibility status: Binary treatment
  \[ Z_i = z \in \{0, 1\} = \{\text{Ineligible, Eligible}\} \]

- $S_i =$ Economic measure: The forcing variable
  \[ Z_i = 1\{S_i \leq s_0\} \quad s_0 = 15000 \text{ euros (threshold)} \]

- Under SUTVA, each unit $i$ has two potential outcomes for each post-treatment variable
- $Y_i(z):$ Potential outcome (dropout) given assignment to eligibility status $z$
- Potential grant application status given assignment to eligibility status $z$
  \[ A_i(z) = \begin{cases} 
  1 & \text{If student } i \text{ applies} \\
  0 & \text{If student } i \text{ does not apply} 
\end{cases} \]

- Potential grant receipt status given assignment to eligibility status $z$
  \[ W_i(z) = \begin{cases} 
  1 & \text{If student } i \text{ receives a grant} \\
  0 & \text{If student } i \text{ does not receive a grant} 
\end{cases} \]

- The Italian University grant allocation rule implies:
  \[ W_i(0) = 0 \text{ (as ineligible units have no access to grant)} \quad W_i(z) = z \times A_i(z) \]
Principal Strata of Application: \( G_i = (A_i(0), A_i(1)) \)

(Frangakis and Rubin, 2002)

- **Never-applicants**: Subjects who would not apply irrespective of their eligibility status
  \[ n = (0, 0) \equiv \{ i : A_i(0) = 0, A_i(1) = 0 \} \]

- **Always-applicants**: Subjects who would apply irrespective of their eligibility status
  \[ a = (1, 1) \equiv \{ i : A_i(0) = 1, A_i(1) = 1 \} \]

- **Compliant-applicants**: Subjects who would not apply if ineligible and would apply if eligible
  \[ a = (0, 1) \equiv \{ i : A_i(0) = 0, A_i(1) = 1 \} \]

- **Defiant-applicants**: Subjects who would apply if ineligible and would not apply if eligible
  \[ d = (1, 0) \equiv \{ i : A_i(0) = 1, A_i(1) = 0 \} \]

- **Key property of \( G \)**: \( G \) is not affected by assignment
- **Treatment comparisons within a principal stratum are always causal effects**
Causal Estimands

Super-Population estimands

\[ \tau_{a}^{SP} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|G_i = a] \quad \tau_{c}^{SP} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|G_i = c] \]

\[ \tau^{SP} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|G_i = \{a, c\}] = \mathbb{E}[Y_i(1) - Y_i(0)|W_i(1) = 1] \]

\[ = \frac{\tau_{a}^{SP} Pr(G_i = a) + \tau_{c}^{SP} Pr(G_i = c)}{Pr(G_i = a) + Pr(G_i = c)} \]

- \( \tau^{SP} \) = average principal causal effects for the compliers (\( W_i(0) = 0, W_i(1) = 1 \)): Standard IV estimand
- Compliers are union of always-applicants and compliant-applicants

Corresponding finite-sample estimands can be defined
Probabilistic Treatment Assignment Mechanism

Key: reconstruct the hypothetical experiment underlying a RDD by viewing the forcing variable as a random variable with a probability distribution:

**Assumption 1.** *(Local Overlap).* There exists a subset of units, $\mathcal{U}_{s_0}$, such that for each $i \in \mathcal{U}_{s_0}$, $\Pr(S_i \leq s_0) > \epsilon$ and $\Pr(S_i > s_0) > \epsilon$ for some sufficiently large $\epsilon > 0$.

**Assumption 2.** *(Shape of the Overlap Set).* There exists $h > 0$ such that for each $\epsilon > 0$, $\Pr(s_0 - h \leq S_i \leq s_0 + h) > 1 - \epsilon$, for each $i \in \mathcal{U}_{s_0}$.

**Assumption 3.** *(Local Stable Unit Treatment Value Assumption).* For each $i \in \mathcal{U}_{s_0}$, consider two eligibility statuses $Z_i' = 1(S_i' \leq s_0)$ and $Z_i'' = 1(S_i'' \leq s_0)$, with $S_i' \neq S_i''$.

If $Z_i' = Z_i''$ then $A_i(Z') = A_i(Z'')$, $W_i(Z') = W_i(Z'')$, $Y_i(Z') = Y_i(Z'')$.

**Assumption 4.** *(Local Randomization).* For each $i \in \mathcal{U}_{s_0}$,

$$\Pr(S_i|A_i(0), A_i(1), W_i(0), W_i(1), Y_i(0), Y_i(1), X_i) = \Pr(S_i)$$

Local Causal Estimands: “Local” version of the causal estimands within $\mathcal{U}_{s_0}$: $\tau_{a,s_0}, \tau_{c,s_0}, \tau_{s_0}$.
Bayesian inference considers the observed values to be realizations of random variables and the unobserved values to be unobserved random variables.

Given $\mathcal{U}_{s_0}$, under Assumptions 1-4 and exchangeability, Bayesian inference involves three sets of models:

1. Model for principal stratum membership: $\Pr(G_i|X_i; \theta)$;
2. Model for potential outcomes: $\Pr(Y_i(0), Y_i(1)|G_i, X_i; \theta)$;
3. Priors for the parameters $p(\theta)$.

$G_i = (A_i(0), A_i(1)) = (A_i^{obs}, A_i^{mis})$ is unknown, and the likelihood involves mixtures.

Posterior computation via a Gibbs sampler with data augmentation (to sample $A_i^{mis}$ for each unit).
Two additional assumptions

- **Assumption 5.** (Monotonicity):
  \[ A_i(1) \geq A_i(0), \quad \text{for all } i \in U_{s_0} \]

- **Assumption 6.** (Exclusion Restriction (ER) for never-applicants):
  \[ \Pr(Y_i(1)|G_i = n; i \in U_{s_0}) = \Pr(Y_i(0)|G_i = n; i \in U_{s_0}) \]
Application to Italian University Grants

- Administrative data of 1st year students in Universities of Florence and Pisa in year 2004 – 2006

- Forcing variable: A combined economic measure of family income and asset. Threshold: 15,000 euros

- The forcing variable is hard to manipulate

- Covariates: Sex, high school type, high school grade, enrollment year, university, major in university
Selection of the subpopulation $\mathcal{U}_{s_0}$

- Selection of the subpopulation $\mathcal{U}_{s_0}$: Choice of the bandwidth $h$ defining an interval around the threshold, $s_0$, where our RD assumptions (Assumptions 1 through 4) hold.

- Conventional RD approaches using local polynomial regression also involve bandwidth selection, but for a very different objective, namely finding an optimal balance between precision and bias at the threshold for local polynomials.

- Imbens-Kalyanaraman optimal bandwidth $= 3295.7$ (Imbens and Kalyanaraman, 2012)
The role of the local randomization assumption in the selection of the subpopulation $\mathcal{U}_{s_0}$

- Under Assumption 4, in the subpopulation $\mathcal{U}_{s_0}$, pre-treatment variables should be well balanced in the two subsamples defined by assignment.

- Test of the sharp null hypothesis of no effect of assignment on pre-treatment covariates: Each $h$ such that the sharp null hypothesis of no effect cannot be rejected defines a potential subpopulation, $\mathcal{U}_{s_0}$.

- Rejection of the sharp null hypothesis can be interpreted as evidence against the local randomization assumption (Assumption 4).

- Choice of the pre-treatment covariates: It is important to consider all variables known (believed) to be related to both assignment and the outcome.
Assessing the balance of the pre-treatment variables

- We start with a small bandwidth $h$ around the threshold and then increase $h$ until the sharp null hypothesis is rejected.
- Randomization tests (Cattaneo et al., 2013)
- Multiple testing problem:

<table>
<thead>
<tr>
<th></th>
<th>Not reject</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null True</td>
<td>$A_0$</td>
<td>$R_0$</td>
<td>$M_0$</td>
</tr>
<tr>
<td>Null False</td>
<td>$A_1$</td>
<td>$R_1$</td>
<td>$M_1$</td>
</tr>
<tr>
<td>Total</td>
<td>$A$</td>
<td>$R$</td>
<td>$M$</td>
</tr>
</tbody>
</table>

Need to control for *compound error measures*

- Bonferroni method (Family wise error rate $= \Pr(R_0 \geq 1)$):
  $$\alpha^B = \frac{\alpha}{M}$$
- Benjamini- Hochberg method
  (False discovery rate $= \mathbb{E}[R_0/R|R > 0] \Pr(R > 0)$)
  $$\alpha^{BH} = \frac{\alpha \cdot K^*}{M} \quad K^* = \max_{k=1,\ldots,M} \left\{ p(k) \leq \frac{\alpha \cdot k}{M} \right\}$$
  $$p(1) \leq \ldots \leq p(M)$$ (ordered $p$–values)

- Bayesian methods: The role of the parameter $h$
- Bayesian approach to multiple testing (e.g., Scott and Berger, 2005)
Randomization test: P-values of difference-in-means

*(Cattaneo et al., 2013)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>bandwidth (h)</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(n=1 042)</td>
<td>(n=2 108)</td>
<td>(n=3 166)</td>
<td>(n=4 197)</td>
<td>(n=16 361)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td>.574</td>
<td>.629</td>
<td>.741</td>
<td>1.000</td>
<td>.003</td>
</tr>
<tr>
<td>HS Type (Other)</td>
<td></td>
<td>.563</td>
<td>.387</td>
<td>.632</td>
<td>.779</td>
<td>.308</td>
</tr>
<tr>
<td></td>
<td>Humanity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
<td>1.000</td>
<td>.436</td>
<td>.794</td>
<td>.143</td>
</tr>
<tr>
<td></td>
<td>Tech</td>
<td>.846</td>
<td>.587</td>
<td>.974</td>
<td>.750</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>HS Grade</td>
<td>.249</td>
<td>.879</td>
<td>.658</td>
<td>.686</td>
<td>.000</td>
</tr>
<tr>
<td>Year (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>.943</td>
<td>.174</td>
<td>.469</td>
<td>.847</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>.833</td>
<td>.517</td>
<td>.904</td>
<td>1.000</td>
<td>.000</td>
</tr>
<tr>
<td>University</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PI versus FI)</td>
<td></td>
<td>.325</td>
<td>.623</td>
<td>.007</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td>Major in University (Other)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Humanity</td>
<td>.202</td>
<td>.023</td>
<td>.034</td>
<td>.146</td>
<td>.214</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>1.000</td>
<td>.855</td>
<td>.376</td>
<td>.344</td>
<td>.429</td>
</tr>
<tr>
<td></td>
<td>Social Science</td>
<td>.942</td>
<td>.792</td>
<td>.933</td>
<td>1.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>BioMed</td>
<td>.328</td>
<td>.074</td>
<td>.351</td>
<td>.389</td>
<td>.487</td>
</tr>
<tr>
<td></td>
<td>Tech</td>
<td>.196</td>
<td>.242</td>
<td>.455</td>
<td>.874</td>
<td>.000</td>
</tr>
</tbody>
</table>

\[
\alpha = 0.05 \implies \alpha^B = \frac{0.05}{13} = 0.004 \quad \text{and} \quad \frac{h}{\alpha^{BH}} = 0.000 \quad 0.000 \quad 0.000 \quad 0.004 \quad 0.035
\]
<table>
<thead>
<tr>
<th>Variable</th>
<th>bandwidth (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 (n=1,042)</td>
</tr>
<tr>
<td>Sex</td>
<td>0.95</td>
</tr>
<tr>
<td>HS Type (Other)</td>
<td></td>
</tr>
<tr>
<td>Humanity</td>
<td>0.95</td>
</tr>
<tr>
<td>Science</td>
<td>0.91</td>
</tr>
<tr>
<td>Tech</td>
<td>0.81</td>
</tr>
<tr>
<td>HS Grade</td>
<td>0.96</td>
</tr>
<tr>
<td>Year (2004)</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>0.96</td>
</tr>
<tr>
<td>2006</td>
<td>0.92</td>
</tr>
<tr>
<td>University</td>
<td>0.92</td>
</tr>
<tr>
<td>(PI versus FI)</td>
<td></td>
</tr>
<tr>
<td>Major in University (Other)</td>
<td></td>
</tr>
<tr>
<td>Humanity</td>
<td>0.90</td>
</tr>
<tr>
<td>Science</td>
<td>0.86</td>
</tr>
<tr>
<td>Social Science</td>
<td>0.82</td>
</tr>
<tr>
<td>BioMed</td>
<td>0.78</td>
</tr>
<tr>
<td>Tech</td>
<td>0.63</td>
</tr>
</tbody>
</table>
Model Specification

- Model for principal strata:
  \[
  \begin{align*}
  \Pr(G_i = a) &= \Pr(G_i^*(a) \leq 0), \\
  \Pr(G_i = n) &= \Pr(G_i^*(a) > 0 \text{ and } G_i^*(n) \leq 0), \\
  \Pr(G_i = c) &= 1 - \Pr(G_i = n) - \Pr(G_i = a),
  \end{align*}
  \]

  where
  \[
  G_i^*(a) = \alpha_{a0} + S_i \alpha_a^{(S)} + X_i' \alpha_a^{(X)} + \epsilon_{ai}, \quad \epsilon_{ai} \sim N(0, 1)
  \]
  \[
  G_i^*(n) = \alpha_{n0} + S_i \alpha_a^{(S)} + X_i' \alpha_n^{(X)} + \epsilon_{ni}, \quad \epsilon_{ni} \sim N(0, 1)
  \]

- Model for outcome (probit link):
  \[
  \Pr(Y_i(z) = 1 | G_i = g, S_i, X_i) = \Phi \left( \beta_{0,g,z} + S_i \beta_{g,z}^{(S)} + X_i' \beta_{g,z}^{(X)} \right)
  \]

- We assume that parameters are a priori independent and use multivariate normal priors for the coefficients with mean 0 and large variances

- To calculate the sample-average estimates, we assume the association between \( Y_i(0) \) and \( Y_i(1) \) to be 0
### Results

<table>
<thead>
<tr>
<th>Estimand</th>
<th>Population-average</th>
<th>Sample-average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median 2.5% 97.5%</td>
<td>Median 2.5% 97.5%</td>
</tr>
<tr>
<td>$h = 500$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Pr(G_i = a)$</td>
<td>.323 .294 .355</td>
<td>.322 .309 .336</td>
</tr>
<tr>
<td>$\Pr(G_i = c)$</td>
<td>.060 .031 .105</td>
<td>.041 .021 .090</td>
</tr>
<tr>
<td>$\Pr(G_i = n)$</td>
<td>.616 .570 .650</td>
<td>.637 .590 .651</td>
</tr>
<tr>
<td>$\tau_{a,s_0}$</td>
<td>$-.153 -.313 -.030$</td>
<td>$-.152 -.307 -.038$</td>
</tr>
<tr>
<td>$\tau_{c,s_0}$</td>
<td>.045 $-.170 .497$</td>
<td>.074 $-.256 .545$</td>
</tr>
<tr>
<td>$\tau_{s_0}$</td>
<td>$-.116 -.253 -.005$</td>
<td>$-.120 -.265 -.009$</td>
</tr>
<tr>
<td>$h = 1000$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Pr(G_i = a)$</td>
<td>.336 .312 .365</td>
<td>.333 .318 .354</td>
</tr>
<tr>
<td>$\Pr(G_i = c)$</td>
<td>.043 .002 .086</td>
<td>.027 .002 .075</td>
</tr>
<tr>
<td>$\Pr(G_i = n)$</td>
<td>.623 .584 .652</td>
<td>.640 .599 .645</td>
</tr>
<tr>
<td>$\tau_{a,s_0}$</td>
<td>$-.161 -.273 -.052$</td>
<td>$-.161 -.270 -.057$</td>
</tr>
<tr>
<td>$\tau_{c,s_0}$</td>
<td>.028 $-.745 .828$</td>
<td>.031 $-.778 .871$</td>
</tr>
<tr>
<td>$\tau_{s_0}$</td>
<td>$-.132 -.242 -.021$</td>
<td>$-.139 -.247 -.034$</td>
</tr>
</tbody>
</table>
Posterior predictive model checking

(e.g., Gelman et al., 1996)

<table>
<thead>
<tr>
<th>$h$</th>
<th>$\mathcal{G}$</th>
<th>Signal</th>
<th>Noise</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>${a}$</td>
<td>.095</td>
<td>.630</td>
<td>.094</td>
</tr>
<tr>
<td></td>
<td>${c}$</td>
<td>.254</td>
<td>.325</td>
<td>.254</td>
</tr>
<tr>
<td></td>
<td>${a, c}$</td>
<td>.338</td>
<td>.273</td>
<td>.370</td>
</tr>
<tr>
<td>1000</td>
<td>${a}$</td>
<td>.411</td>
<td>.425</td>
<td>.419</td>
</tr>
<tr>
<td></td>
<td>${c}$</td>
<td>.400</td>
<td>.444</td>
<td>.473</td>
</tr>
<tr>
<td></td>
<td>${a, c}$</td>
<td>.493</td>
<td>.335</td>
<td>.518</td>
</tr>
</tbody>
</table>
Bivariate Analysis

- Besides dropout, a secondary outcome, “credit taken”, is also of interest.
- Another reason to conduct bivariate analysis is to sharpen the analysis:
  - Recent results (Mealli and Pacini, 2012; Li, Mattei and Mealli, 2012) show using multiple outcomes sharpens the inference for causal effects in principal stratification.
  - Mercatanti, Li, and Mealli (2015): under correct specification, in normal mixture models, bivariate models generally lead to larger observed information of the parameters than corresponding marginal univariate models.
- **Intuition**: Even if two outcomes are conditionally independent within each cluster, the marginal dependence still help predict cluster membership.
- Compliance types are essentially latent mixtures. Multiple outcomes help to better disentangle the mixture.
The current Italian university grants seem to be effective in reducing dropout from universities among students from families with annual economic measure around 15,000 euros.

Always-applicants and compliant-applicants are found to be heterogeneous with respect to the effect of the grants.

It appears more beneficial for education administrations to lower the eligibility criteria to allow more applicants to get the grant, than to increase the amount of the grant to awarders.

Directions for future research: (1) Sequential RDDs using multiple year data; (2) bivariate analysis.
References


