Dipartimento di Statistica "Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

WORKING PAPER 2009/04

Exploiting Nonlinear Difference-in-Difference Assumptions in a Regression Discontinuity Design

Fabrizia Mealli,
Carla Rampichini

Università degli Studi di Firenze

Exploiting Nonlinear Difference-in-Difference Assumptions in a Regression Discontinuity Design [(⋆)]

Fabrizia Mealli and Carla Rampichini

Dipartimento di Statistica "Giuseppe Parenti", Università di Firenze

e-mail: mealli@ds.unifi.it, rampichini@ds.unifi.it

Summary: The paper tackles the problem of identification of treatment effects in a regression-discontinuity-design (RDD) in the presence of heterogeneous effects. A RDD allows identification of average treatment effects only for a subset of individuals around the threshold for the participation status. The paper shows how a sharp RDD may provide: a) additional ways to define specification tests for the continuity assumptions at the discontinuity point on which identification usually rests; b) additional ways to test the performance of alternative non-experimental estimators of programme effects away from the threshold; c) alternative identification assumptions, similar to those on which nonlinear difference-in-difference estimators rest, which can be partially tested, and allow to extend estimation results away from the threshold. The considered set-up is one where a budget-constraint induced threshold splits the relevant population into two groups, the *ex-post* eligible and ineligible individuals, and application in both groups is determined according to rules potentially unknown to the researcher, so application (or participation) is not mandatory but voluntary. The proposed tools are applied to the evaluation of Italian university grants. Applicants meeting some ex-ante eligibility criteria receive a grant if their family economic indicator S is below a threshold $\hat{s}$. 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. However, under some relatively weak nonlinear difference-in-difference type of assumptions, results show that moving below the threshold, thus for less well-off (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.

Keywords: evaluation of university grants, nonlinear DID, sharp RDD, specification tests.

1. Introduction

Since the late 1990s, regression discontinuity applications have increased their popularity: the attractiveness of such designs probably rests on its similarity with a formal randomized experiment and the consequent perception that the identifying assumptions are relatively weak and plausibly hold in many circumstances. In a regression-discontinuity design (RDD) the participation status depends, either completely or partly, by the value of an observed preprogramme characteristic S being above or below a specified threshold. A recent volume of the Journal of Econometrics (Imbens and Lemieux, 2008) contains a collection of papers reviewing new theoretical developments as well as practical issues in implementation of regression-discontinuity methods.

[(⋆)] Financial suppor for this research was provided by MIUR COFIN 2005 grant.

In presence of heterogeneous effects, a RDD allows identification of an average treatment effect for (a subset of) individuals around the threshold for the participation status. Battistin and Rettore (2008) propose some specification tests to overcome this limitation: in particular they show that, when individuals self-select into participation conditional on some eligibility criteria, a sharp RDD provides a natural framework to define a specification test for the non-experimental estimation of programme effects for participants away from the threshold. Along the same lines and within a similar, though not identical, context, we show how a sharp RDD may provide: a) additional ways to define specification tests for the continuity assumptions at the discontinuity point on which identification usually rests; b) additional ways to test the performance of alternative non-experimental estimators of programme effects away from the threshold; c) alternative identification assumptions, similar to those on which nonlinear difference-in-difference (DID) estimators rest(Athey and Imbens, 2006), which can be partially tested, and allow to extend estimation results away from the threshold.

The particular setting we consider is one where the threshold for participation is different from the threshold of initial eligibility; such situation may arise in the presence of budget constraints characterizing some active labor market policies. In Italy, for example, vocational training is organized in such a way so that, for some training courses, only individuals having a high school grade above a certain threshold are initially admitted to the program; ranking of applicants is also based on high school grade and, due to limited resources, only part of the applicants scoring above a *new* threshold are actually admitted to the program. So, participation is assigned to some of the initially eligible applicants as in a sharp RDD. However, high school grade is usually observed also for non applicants, who can, as a consequence, also be divided into two groups depending on their score being above or under the budget-constraint induced threshold. Similar labor market programmes include all those in which participation is voluntary for individuals satisfying a condition on, say, age, or means tested programmes, where we assume that those characteristics determining admission to the programme are also observed for the non applicants.

Throughout the paper, we consider the case in which a budget-constraint induced threshold splits the relevant population into two groups, the ex-post eligible and ineligible individuals, and application in both groups is determined according to rules potentially unknown to the researcher, so we assume that application or participation is not mandatory but voluntary.

The plan of the paper is as follows. We first introduce sharp and fuzzy RDD and state the identifying conditions required to estimate causal estimands around the threshold (Section 2). In Section 3 we describe the particular RDD setting considered in the paper, characterized by the presence of non-eligible applicants. A continuity test is presented in Section 4, as an additional tool complementing the long arrays of specification tests and sensitivity analyses discussed in Imbens and Lemieux (2008). In Section 5.1, we show additional ways to test the performance of alternative non-experimental estimators of programme effects away from the threshold; in particular we show that the selection bias arising from non random selection of applicants is identified above the threshold for the non eligible, so that one can formally test whether any of the existing non-experimental estimator can correct for this bias. If the hypothesis is not rejected for the non eligible, one may feel more confident to use non-experimental estimators to identify causal effects on a broader population (usually that of the eligible applicants). Finally, in Sections 5.2

and 5.3we show how to exploit difference-in-difference designs, based on the availability of non eligible applicants, to characterize the bias, and estimate the effect away from the threshold by non linear difference-in-difference type estimation strategies. Links with the existing literature are established, in particular with **?**, Heckman and Hotz (1989), Heckman *et al.* (1998).

The proposed tools are applied to the evaluation of Italian university grants in Section 6, where we try to assess whether grants are an effective tool to prevent students from low income families from dropping out of higher education. Section 7 discusses the results and concludes.

2. The Regression Discontinuity Design

In order to introduce the basic features of a RDD, consider a binary treatment, and define the treatment variable T, with $T = 1$ if a subject receives the treatment of interest, e.g,. she participates in a programme, and $T = 0$ if a subject does not receive the treatment. For every subject, it is possible to define two *potential outcomes* $(Y_1; Y_0)$ that a subject would experience by being treated and not treated, respectively (Rubin, 1974). The *causal effect* of the treatment for a specific subject is in general a non-observable comparison between Y_1 and Y_0, e.g., $Y_1 - Y_0$, since only one of the potential outcomes can be observed, the other one being concealed by exposure or non exposure to the treatment.

If assignment to treatment is randomized and treatment assignment coincides with treatment received, i.e., subjects comply with assignment, then we can assume that $(Y_1, Y_0) \perp\!\!\!\perp T$, so the average causal effect, ATE, can be estimated by comparing the sample averages of Y for treated and not treated subjects, given that

$$ATE = E(Y_1 - Y_0) = E(Y_1 \mid T = 1) - E(Y_0 \mid T = 0), \tag{1}$$

that is, randomization allows to use information on non-participants to identify the mean counterfactual outcome for participants, and viceversa, because, under randomization, conditioning on T is irrelevant.

Suppose that the treatment status T is not randomized, but depends on an observable continuous random variable S; moreover, there exists a known cutoff point $\tilde{s}$ in the support of S where the probability of treatment received changes discontinuously, i.e.,

$$Pr(T = 1 \mid \tilde{s}^+) \neq Pr(T = 1 \mid \tilde{s}^-)$$

where $Pr(T = 1 \mid \tilde{s}^+) = \lim_{s \to \tilde{s}^+} Pr(T = 1 \mid S = s)$, while $Pr(T = 1 \mid \tilde{s}^-) = \lim_{s \to \tilde{s}^-} Pr(T = 1 \mid S = s)$. This rule determines a Regression Discontinuity Design (RDD). The basic idea underlying the RDD is that subjects below and above the threshold $\tilde{s}$ are similar, so that a quasi-randomization can be assumed around the threshold. Usually two special cases of RDD are considered: sharp and fuzzy RDD.

2.1. Sharp RDD

In a sharp RDD, the treatment variable T depends deterministically on the value of S, typically $T \equiv \mathbb{I}(S < \tilde{s})$, so that $Pr(T = 1 \mid \tilde{s}^-) - Pr(T = 1 \mid \tilde{s}^+) = 1$. In this case, the average impact of the treatment (1) at $\tilde{s}$:

$$E(Y_1 - Y_0 \mid \tilde{s}) = E(Y_1 \mid \tilde{s}^-) - E(Y_0 \mid \tilde{s}^+) \tag{2}$$

where $E(Y_1 \mid \tilde{s}^-) = \lim_{s \to \tilde{s}^-} E(Y_1 \mid S = s)$, and $E(Y_0 \mid \tilde{s}^+) = \lim_{s \to \tilde{s}^+} E(Y_0 \mid S = s)$, is identified if the following sufficient continuity Condition 1 holds (Battistin and Rettore, 2008; Hahn *et al.*, 2001).

Condition 1 *The mean values of Y_0 and of Y_1 conditional on S are continuous functions of S at $\tilde{s}$.*

2.2. Fuzzy RDD

A fuzzy RDD (Trochim, 1984) may arise when individuals do not comply with the mandated status resulting from a sharp assignment rule, drop out of the programme or seek alternative treatments if denied it. In a fuzzy RDD the probability of participation decreases (increases) as S crosses $\tilde{s}$, so the following inequality holds:

$$Pr(T = 1 \mid \tilde{s}^-) - Pr(T = 1 \mid \tilde{s}^+) > 0. \tag{3}$$

In this case, the continuity of Y_0 and Y_1 at $\tilde{s}$ expressed by Condition 1 is no longer enough to ensure identifiability and the following two additional Conditions 2 and 3 are needed (Battistin and Rettore, 2008; Hahn *et al.*, 2001):

Condition 2 *The triple ($Y_0; Y_1; T$) is stochastically independent of S in a neighborhood of $\tilde{s}$.*

The independence between T and S in a neighborhood of $\tilde{s}$ corresponds to imposing the restriction that assignment at $\tilde{s}$ takes place as if it were randomized. On the other hand, the independence between $Y_0; Y_1$ and S at $\tilde{s}$ corresponds to an exclusion restriction asserting that, in a neighborhood of $\tilde{s}$, S affects the outcomes only through its effect on T (see the discussion on the role of the exclusion restriction in Angrist *et al.*, 1996).

Condition 3 *Participation into the programme is monotone around $\tilde{s}$.*

Under conditions 1, 2 and 3, it is possible to identify the average impact:

$$E\left[Y_1 - Y_0 \mid T(Z = 1) = 1, T(Z = 0) = 0, \tilde{s}\right] \tag{4}$$

where $Z = \mathbb{I}(S < \tilde{s})$ is the mandated status variable and $T(Z = j)$ is the treatment status, $j = 0, 1$. The expected value (4) represents the mean impact of the programme on those individuals in a neighborhood of $\tilde{s}$ who would switch their treatment status if the threshold for participation switched from just below their score to just above it, i.e., the mean impact of the programme on the *compliers* (Imbens and Angrist, 1994; Angrist *et al.*, 1996).

3. A RDD with non eligible applicants

Social interventions are often targeted to specific groups of individuals meeting a fully specified set of conditions for eligibility. Means tested programmes (such as food stamp programmes) or labor market programmes, whose eligibility criteria depend on the duration of unemployment or on the age of individuals, are frequently encountered examples of such a scheme. Italian university grants are another example.

To fix ideas, let S be a continuous pre-programme characteristic (e.g., family income) and let the eligibility status be established according to the deterministic rule $S < \tilde{s}$. That is, subjects are eligible for the programme ($Z = 1$) if and only if they have a value of the variable S below a known threshold $\tilde{s}$, so $Z = \mathbb{I}(S < \tilde{s})$. Throughout our discussion, it will be assumed that S is observable for all individuals. If all eligibles (and only them) apply to the programme and there are no budget or other constraints, so that all the applicants are assigned to the treatment, a sharp RDD would arise. For example, if participation to the programme is mandatory for all eligible individuals, the treatment effect at the threshold for eligibility would be identified provided that Condition 1 holds. However, it is often the case that some eligible individuals self-select into the programme while some others do not, typically when application is on a voluntary basis. Individuals' heterogeneity about information on the availability of the programme and individuals' preferences are among the factors likely to influence the application decision in several instances. Accordingly, the population turns out to be divided into three subgroups: ineligibles, eligible non-applicants and eligible applicants, i.e., participants. This situation is represented in Table 1, where A is the application indicator, i.e., $A = 1$ for eligible individuals applying to receive the treatment, and $A = 0$ otherwise. In a typical RDD, the following equality obviously holds: $Pr(A = 0 \mid Z = 0) = 1$. Note that the treatment indicator T is equal to $A \times Z$. This situation with eligible non-applicants has been widely discussed and analyzed in Battistin and Rettore (2008), who show that it provides a natural framework to define specification tests for non-experimental estimation of causal effects for participants away from the threshold $\tilde{s}$.

Table 1: *Application and eligibility status in a typical fuzzy RDD*

Application status	*Eligibility* status	
	$Z = 0 (S \geq \tilde{s})$	$Z = 1 (S < \tilde{s})$
$A = 0$	ineligibles	eligible non-applicants
$A = 1$	///////	applicants ($\equiv T = 1$)

Suppose now that, not only people self-select into the programme, but, due to budget constraints, only some of the eligible applicants receive the treatment, i.e., those applicants with S below an *ex-post* threshold $\hat{s}$, where $\hat{s} < \tilde{s}$. As before, define A the application indicator (i.e., $A = 1$ if a subject applies to receive the treatment and 0 otherwise), while $Z = \mathbb{I}(S < \hat{s})$ is now the (ex-post) eligibility indicator, so the treatment indicator is again $T = A \times Z$. In the previous example (Table 1) $Z = 0$ implies $A = 0$, while now this is not necessarily the case, because the threshold $\hat{s}$ is defined *after* the application deadline, and it is presumably unknown to the applicants.

Accordingly, in this case the population turns out to be divided into four subgroups:

ineligible applicants, eligible applicants (participants), ineligible non-applicants, and eligible non-applicants. It will be assumed that information on the four subgroups of individuals is available to the researcher. This situation is described in Table 2.

Table 2: *Application and eligibility status in case of ex-post eligibility*

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* $(\Rightarrow T = 1)$

The first two groups, i.e., the applicants, can be used to estimate the treatment effect at $\hat{s}$ for those who apply:

$$E(Y_1 - Y_0 \mid A = 1, \hat{s}) = E(Y_1 \mid A = 1, \hat{s}^-) - E(Y_0 \mid A = 1, \hat{s}^+). \quad (5)$$

Indeed, conditionally on application ($A = 1$), T is equivalent to Z and the probability of being treated for those individuals scoring a value of S below the threshold $\hat{s}$ is one, while the probability of being treated for those scoring above $\hat{s}$ is zero. As a result, the probability of participation is discontinuous at the threshold for (ex-post) eligibility, i.e.,

$$Pr(T = 1 \mid A = 1, Z = 1) - Pr(T = 1 \mid A = 1, Z = 0) = 1 \quad (6)$$

so that a sharp fuzzy RDD is defined for the applicants, even if people have self-selected into the programme.

Note that, because $\hat{s}$ is unknown to the applicants, the situation appears to be ideal to apply a sharp RDD. Presumably, in fact, we do not expect any discontinuity in the distribution of the forcing variable S at the threshold induced by manipulation by individual agents, because of their ignorance about $\hat{s}$.

The other two groups, i.e., the non-applicants are not involved directly in the estimation of the programme effect; however they may play a role in testing and finding alternative identifying assumptions, which have, by the best of our knowledge, sofar not been exploited. The non-applicants can in fact be used: (*i*) to define and implement specification tests for the continuity assumptions on which identification in the sharp RDD rests, (*ii*) to find additional ways to test the performance of alternative non-experimental estimators of programme effects away from the threshold, and (*iii*) to find alternative nonlinear difference-in-difference type of assumptions which would allow to extend estimation results away from the threshold.

4. Another continuity test in the presence of non applicants

The presence of eligible and non-eligible applicants can be exploited in order to test the continuity assumption (Condition 1), needed for the *sharp* RDD, comparing eligible and non-eligible *non applicants* around the cutoff point $\hat{s}$.

Note that, for the non-applicants, Y_0 can be observed above and below $\hat{s}$, so that the continuity of the conditional mean value of Y_0 given S at $\hat{s}$ can be tested, i.e.

$$\lim_{s \to \tilde{s}^+} E(Y_0 \mid S, A = 0) = \lim_{s \to \tilde{s}^-} E(Y_0 \mid S, A = 0). \quad (7)$$

The test can be done on the conditional expected value or on the density function of Y_0, depending on the causal estimand of interest. If the null hypothesis of continuity is rejected, then it will be hard to assume that the continuity assumption holds for the applicants, which is required for identification. On the other hand, by not rejecting the null hypothesis, one might feel more confident in assuming continuity for the applicants and in trusting the estimates obtained with the sharp RDD. 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 $E(Y_0 \mid S, A = 0)$ at $\hat{s}$. For example, one could test for continuity by directly modelling the regression of Y_0 on S and the eligibility status $Z = \mathbb{I}(S < \hat{s})$: any test on the eligibility status coefficient is a continuity test. Alternatively, one can use some semi-parametric estimator of the discontinuity of Y_0 at $\hat{s}$ (Porter, 2003). This test should be viewed as an additional tool, complementing the long arrays of specification tests and sensitivity analyses based on tests for discontinuities in the average values of the covariates, discontinuities in the average outcome at different values of S and, discontinuities in the conditional density of S (McCrary, 2008), summarized and discussed in Imbens and Lemieux (2008).

5. Extending the RDD

In the following we will show how to use the information on non-applicants to perform specification tests of alternative non-experimental estimators of programme effects away from the threshold and alternative identification assumptions, which can be partially tested, and allow to extend estimation results away from the threshold. Indeed, if the effect of being treated is heterogeneous with respect to S, the mean impact of individuals in a neighborhood of $\hat{s}$ is not informative at all on the impact for individuals away from $\hat{s}$.

5.1. Ignorability of application

In order to estimate the effect of the treatment on a larger subset of treated units, we usually need to rely on an assumption of unconfoundedness (**?**), i.e., $Y_0 \perp\!\!\!\perp T \mid X, S$, with $S < \widehat{s}$, where X is a vector of observed pre-treatment covariates. Unconfoundedness cannot be tested because Y_0 is observed only for $T = 0$. However, note that when $S < \widehat{s}$, $T = 0$ if $A = 0$, i.e., if a subject does not apply to participate into the programme, so that $Y_0 \perp\!\!\!\perp T \mid X, S$ is equivalent to $Y_0 \perp\!\!\!\perp A \mid X, S$ when $S < \widehat{s}$. The condition $Y_0 \perp\!\!\!\perp A \mid X, S$ can be tested for $S \geq \widehat{s}$, by comparing ineligible non-applicants with the ineligible applicants. Any procedure allowing to test $H_0 : Y_0 \perp\!\!\!\perp A \mid X, S$ for $S \geq \widehat{s}$ can be used. For example, given that the conditional mean independence would suffice to identify the average treatment effect, one could model the conditional mean of Y_0 given A, X, and S with a flexible regression function. A test on the coefficient of A is a test on the conditional mean independence of Y_0. Note that this test is similar to the one proposed by Battistin and Rettore (2008), but is more powerful, because it uses the information for all the values $S \geq \hat{s}$ and not just the information at the threshold $\hat{s}$. The idea to use different comparison groups to test ignorability conditions is not novel, and can be found in works by Rosenbaum (1984, 1987), LaLonde (1986), and Heckman *et al.* (1998). Some examples in Rosenbaum (1987) resembles our set-up closely. If the null hypothesis is not rejected, one can be more confident in using a matching type estimator controlling for X and S away from the cut-off point $\widehat{s}$ and the non-applicants with $S < \widehat{s}$ as a control group. On the contrary, the rejection of the null hypothesis implies that the unconfoundedness assumption does not hold for $S \geq \widehat{s}$, so it plausibly does not hold also for $S < \widehat{s}$ and it is not possible to estimate the mean impact away from the cutoff point using non applicants as a control group. In this case, i.e., in the presence of a selection bias, one can exploit the presence of non-eligible applicants to test or assume alternative conditions, characterizing the bias (Heckman et al., 1998), which allow to extend the estimation away from the cut-off point, as proposed in the next Section.

5.2. Difference-in-difference designs

The presence of non eligible applicants ($A = 1, Z = 0$) allows to exploit the data to estimate the average effect on all the eligible applicants by means of a difference-in-difference approach (Angrist and Krueger, 2000; Card and Krueger, 1993; Abadie, 2005; Athey and Imbens, 2006). The basic assumption of this approach, with the obvious modifications to conform with the RDD notation, is that, in the absence of treatment, the following condition holds:

Condition 4 *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.

To see how Condition 4 allows the identification of a causal effect, assume that the causal effect of interest is

$$E(Y_1 \mid T = 1) - E(Y_0 \mid T = 1) \tag{8}$$

where the second term of the difference is unobservable. Given that $T = A \times Z$ the difference in (8) can also be written as:

$$E(Y_1 \mid Z = 1, A = 1) - E(Y_0 \mid Z = 1, A = 1). \tag{9}$$

Under Condition 4, the estimand (9) can be identified as follows:

$$\begin{aligned} E(Y_1 \mid Z = 1, A = 1) &- E(Y_0 \mid Z = 1, A = 1) = \\ &= E(Y_1 \mid Z = 1, A = 1) - E(Y_0 \mid Z = 1, A = 1) \\ &\quad -E(Y_0 \mid Z = 1, A = 0) + E(Y_0 \mid Z = 1, A = 0) = \\ &= [E(Y_1 \mid Z = 1, A = 1) - E(Y_0 \mid Z = 1, A = 0)] + \\ &\quad - [E(Y_0 \mid Z = 1, A = 1) - E(Y_0 \mid Z = 1, A = 0)] = \\ &= [E(Y_1 \mid Z = 1, A = 1) - E(Y_0 \mid Z = 1, A = 0)] + \\ &\quad - [E(Y_0 \mid Z = 0, A = 1) - E(Y_0 \mid Z = 0, A = 0)] \end{aligned} \tag{10}$$

where the last equality holds due to Condition 4. All the terms in (10) are observable. Note that Z is a deterministic (not one-to-one) function of S, thus a sufficient Condition for Condition 4 to hold is the following:

Condition 5

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

does not depend on S.

Condition 5 can be tested for $S \geq \hat{s}$; if it holds for $S \geq \hat{s}$, we can be more confident to assume it also holds for $S < \hat{s}$. Note that a constant bias c does not imply that the treatment effect is constant with respect to S. Moreover Condition 5 allows to remove the bias without involving covariates, but only the observation of S. As a consequence, such identifying assumption appears to be particularly attractive when few covariates are observed, making the unconfoundedness assumption not plausible, but S is observed on both applicants and non applicants.

A simple way to test Condition 5 is by means of a regression model of Y_0 on S, A, and the interaction between S and A for non-eligible units ($Z = 0$): a test on the interaction coefficient is equivalent to test Condition 5. In fact, if the null hypothesis on the interaction coefficients is not rejected, the difference in Condition 5 does not depend on S and its value is estimated by the coefficient of A. The linearity assumption can be relaxed using semiparametric regression models.

If Condition 5 holds, so does Condition 4, and equation (10) can be used to derive an estimator for the estimand in (8) away from the threshold $\widehat{s}$, i.e, $S < \widehat{s}$. Note that the identification strategy does not entail parametric restrictions. However, both non parametric or parsimonious parametric approximations can be used for testing and estimation. Condition 5 can be weakened by assuming the following alternative Condition 6:

Condition 6 *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.*

Given that $f_1(S)$ is identifiable for $S \geq \hat{s}$ and $E(Y_0 \mid S < \hat{s}, A = 0)$ is observable (by the presence of eligible non applicants), Condition 6 can be exploited in order to estimate $E(Y_0 \mid S < \hat{s}, A = 1)$. This expected value can then be contrasted with an estimate of $E(Y_1 \mid S < \hat{s}, A = 1)$ in order to estimate the causal effect for participants over $S < \hat{s}$. The expected values can be estimated via parametric or non parametric approaches. This strategy is analogous to that proposed in Athey and Imbens (2006) in a nonlinear difference-in-difference setting. In this paper we consider differences in regression functions instead of considering differences in distributions.

5.3. DID assumptions with categorical response

As it is well known, difference-in-difference assumptions are not invariant to non-linear transformations of the response variable Y. Parametric models for categorical response variables usually involve some non-linear transformation of the expected value of Y, so the DID assumptions stated in the previous section should be modified accordingly. When using a parametric model, a possible strategy is to identify a causal estimand as a function of the model parameters, by exploiting a modified DID assumption, and then use the estimated model to recover the estimand of main interest.

For example, consider a logit model for the binary response Y: in this case the logit of the expected value of Y is modeled as a linear function of the covariates. In our setting, we can specify a logit model for Y_0 as a function of S and A, when $S \geq \hat{s}$:

$$E(Y_0 \mid S, A) = P\{Y_0 = 1 \mid S, A\} = \frac{\exp(\alpha + \beta S + \gamma A + \delta A \times S)}{1 + \exp(\alpha + \beta S + \gamma A + \delta A \times S)} \qquad (11)$$

A test on δ corresponds to test that the difference between the logits:

$$logit P\{Y_0 = 1 \mid S, A = 1\} - logit P\{Y_0 = 1 \mid S, A = 0\} = \gamma + \delta S \qquad (12)$$

does not depend on S, where $logit P\{Y_0 = 1 \mid S, A\} = \alpha + \beta S + \gamma A + \delta A \times S$ and the difference (12) is the log of the following odds ratio:

$$\frac{P(Y_0 = 1 \mid S, A = 1)}{1 - P(Y_0 = 1 \mid S, A = 1)} \Big/ \frac{P(Y_0 = 1 \mid S, A = 0)}{1 - P(Y_0 = 1 \mid S, A = 0)}. \qquad (13)$$

If $\delta = 0$, the odds ratio (13) does not depend on S. If the null hypothesis $H_0 : \delta = 0$ is not rejected, it is possible to estimate the following causal estimand, by exploiting a strategy similar to that in (10):

$$\frac{P(Y_1 = 1 \mid T = 1)}{1 - P(Y_1 = 1 \mid T = 1)} \Big/ \frac{P(Y_0 = 1 \mid T = 1)}{1 - P(Y_0 = 1 \mid T = 1)}. \qquad (14)$$

Note that (14) is a different causal estimand than (8). The parametrization of the logit model, however, allows to recover (8), so the following condition is sufficient for identifying (8) in a logit model setting:

Condition 7 *The difference between the logits*

$$logit P\{Y_0 = 1 \mid S, A = 1\} - logit P\{Y_0 = 1 \mid S, A = 0\}$$

does not depend on S*.*

In order to estimate (14) under Condition 7, one can specify a logit model for Y on applicants and non applicants, and S, Z and A as covariates. The coefficient of Z in such a model represents the logit of the odds- ratio (14), *corrected* for the differences between applicants and non applicants. The model can be extended in order to allow for heterogeneous effects with respect to S, by including interaction terms between S and Z.

Even when Condition 7 does not hold, we can use a weaker condition by assuming that:

Condition 8 *The difference between the logit of applicants and non applicants in the absence of the treatment is a function of* S*, i.e.:*

$$logit P\{Y_0 = 1 \mid S, A = 1\}) - logit P\{Y_0 = 1 \mid S, A = 0\}) = f(S)$$

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.*

Note that this condition corresponds to Condition 6 on a logit scale.

6. Evaluation of Italian University grants

The Italian State Universities offer some grants every year to eligible freshmen: eligibility is based on merit (high school grade ≥ 70) and economic needs. The number of grants awarded is usually constrained by a fixed budget, so that not all the students applying for a grant receive one, even if they are eligible. Applicants are ranked on the basis of an economic indicator, that depends in a deterministic way on family income, property and personal assets, and family structure. Only a varying percentage (could be also 100% in some universities) of high-ranking eligible applicants receive a grant. The amount of aid received may vary with the value of students' economic indicator. The main objective of this intervention is to give equal opportunity to achieve higher education to motivated students irrespective of their income.

Dropout from university is a relevant phenomenon in Italy: indeed, the low rate of graduates among Italian youths is mainly due to the high dropout rate rather than to a low enrollment rate. Moreover, the majority of dropouts happen at the end of the first year of the university studies. The high dropout rate may be due to several circumstances, some of which depend on the socio-economic status of the students (Mealli, Rampichini, 2002; Biggeri and Catalano, 2006). We investigate whether the grant is an effective tool to prevent students from low-income families from dropping out of higher education. The response variable of interest is a binary variable Y that equals 1 if a student drops out at the end of the first year of study and 0 otherwise.

We show which causal parameters can be identified from available information: under some of the conditions stated in previous Sections, the grant's assignment rule can be exploited to estimate the grant's effects at different values of the economic indicator S.

We concentrate the analysis on the freshmen enrolled in 1999 at the University of Padua[(1)], considering only the students who live out of Padua who cannot commute: in addition to the grant, these students are also awarded accommodation in students' houses or, alternatively, accommodation expenses. The 75% of the beneficiaries received an amount of 3190 euros per year, the remaining *richer* beneficiaries received an amount of 2050 euros.

Data include information on: the economic indicator S, high school grade, grant status and grant amount, student status in subsequent year (still enrolled or dropped out), and other student characteristics. Table 3 reports the average characteristics of students meeting the ex-ante eligibility criteria, i.e., with a high school grade ≥ 70 and an income indicator $S < 23395.5$ euros.

The economic indicator is available for students who apply for university fee reduction and for students who apply for a grant. The students can apply for a grant if their economic indicator S is less than $\tilde{s} = 23395.5$ euro. We define *non eligible applicants* those applicants not receiving the grant. Non applicants will be classified as *eligible*

(1) Data come from a larger research project, conducted for the Department of Education by a research group which included the authors, studying the effects of students' aids in 11 Italian State Universities (Biggeri and Catalano, 2006).

Table 3: *Summary statistics (averages) for the 1999 cohort of freshmen resident out of Padua (excluding commuters) meeting the ex-ante eligibility criteria (high school grade* ≥ 70 *and income indicator* $S < 23395.5$ *euros)*

Variable	*Application and eligibility status*			Total
	non applicants	applicants		
		non eligible	eligible	
Dropout	0.20	0.13	0.13	0.14
Economic Indicator	17203	19231	9533	14784
Female	0.59	0.60	0.64	0.61
High school grade	81.6	85.7	85.7	84.8
High school type				
Humanities	0.13	0.13	0.14	0.13
Scientific	0.36	0.41	0.34	0.37
Late enrollment	0.36	0.18	0.19	0.22
University school subject				
Science	0.14	0.23	0.18	0.19
Medicine	0.12	0.10	0.10	0.11
Humanities	0.44	0.39	0.47	0.43
Eng+Arch	0.13	0.16	0.15	0.15

non applicants and *non eligible non applicants* depending on their eligibility status. The eligibility status is defined *ex-post* on the economic indicator value $\hat{s} = 15812.5$ euros. In fact, in 1999 only the 306 applicants with $S < 15812.5$, out of the 577 applicants, received the grant, as shown in Table 4. Among the ex-ante eligible students, 166 did not apply for a grant.

Table 4: *1999 cohort of freshmen resident out of Padua (excluding commuters) meeting the ex-ante eligibility criteria (high school grade* ≥ 70 *and income indicator S* < 23395.5 *euros)*

Application status	*ex-post eligibility status*		Total
	$Z = 0$	$Z = 1$	
$A = 0$	118	48	166
$A = 1$	271	306	577
Total	389	354	743

Figure 1 reports the dropout rate by application and ex-post eligibility status and intervals of width 500 euros of the economic indicator S. Dropout rates are interpolated using the SM method implemented in the GPLOT procedure of SAS (SAS , 2008). The method produces a cubic spline that minimizes a linear combination of the sum of squares of the residuals of fit and the integral of the square of the second derivative. Applicants appear to be more motivated and with a lower dropout rate than non applicants, even in the absence of the grant. Among applicants, those receiving the grant show a lower dropout rate, with an evident jump around the ex-post threshold $\hat{s}$. The non eligible applicants show a larger elasticity of dropping out with respect to S, consistently with their higher

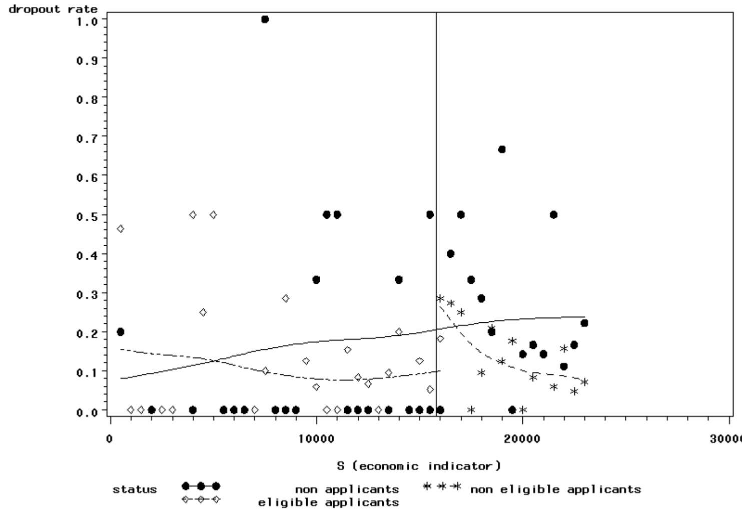

Figure 1: *Dropout rate by application and eligibility status*

motivation to receive a monetary aid.

We will now show how to estimate the grant effect at the discontinuity point $\hat{s}$ via a sharp RDD, assuming Condition 1 holds. We will then show how to estimate the grant effect for all the eligible applicants, i.e., with S below the threshold $\hat{s}$, using data on non applicants, under either an ignorability or a DID type of assumption. For both testing and estimation, we employ simple parametric models, which are used here for simplicity and efficiency reasons due to small sample sizes. As explained in previous Sections, however, identification is not driven by these parametric assumptions. In general, one can be more flexible by employing alternative semi-parametric estimation strategies.

6.1. Continuity test on non applicants

In order to test Condition 1, we fitted a logit model for $E(Y_0 \mid S, Z)$ on non applicants ($A = 0$). As illustrated in Section 4, the test on the coefficient of Z corresponds to a continuity test. We estimate models with different orders of polynomials for S, finding that the coefficient for Z is always non significant (see Table 5), thus Condition 1 could not be rejected. Moreover, the continuity assumption at $\hat{s}$ of the conditional means for the covariates is not rejected for all the covariates reported in Table 3.

6.2. Sharp RDD on applicants

Given that the continuity Condition 1 is not rejected, we estimated the grant effect at $\hat{s}$. To this end, we fitted a logit model for $E(Y \mid S, Z)$, on applicants *around* the threshold $\hat{s}$, i.e., for $5000 < S \leq 23395.5$ euros.

Using the results reported in Table 6, the estimated probabilities of dropping out at the threshold value $\widehat{s} = 15812.51$ for eligible and non-eligible applicants are $P(Y = 1 \mid Z = 1, S = \widehat{s}) = 0.100$ and $P(Y = 1 \mid Z = 0, S = \widehat{s}) = 0.239$ respectively. The

Table 5: *Continuty test on non applicants: logit model for dropout (where $A = 0$)*

Coefficient	*Estimate*	$s.e.$	*Wald chi-square*	$p - value$
Intercept	-1.3465	1.5456	0.7589	0.3837
S	-0.00014	0.000423	0.1029	0.7483
Z	-0.4835	1.1337	0.1819	0.6697
S^2	2.241E-8	4.214E-8	0.2827	0.5950
S^3	-751E-15	1.08E-12	0.4879	0.4849

Table 6: *Sharp RDD on applicants: logit model for dropout on applicants ($A = 1$, $5000 < S \leq 23395.5$)*

Coefficient	*Estimate*	$s.e.$	*Wald chi-square*	$p - value$
Intercept	-4.8597	2.1569	5.0764	0.0243
Z	-0.8741	0.5227	2.7964	0.0945
S	0.000508	0.000258	3.8794	0.0489
S^2	-1.84E-8	8.222E-9	5.0327	0.0249

effect of the grant around the threshold is significant: the grant reduces the dropout rate of 13.9% points, with a 95% confidence interval $[-23.3\%, -4.5\%]$. The confidence interval is calculated as the estimated difference $\pm 2s.e.$, where the $s.e.$ of the estimated difference is obtained via the delta method (Agresti, 2002; Greene, 2000).

6.3. Ignorability test

In order to estimate the effect of the grant away from the threshold, in particular for *poorer* students with $S < \hat{s}$, one could use matching techniques, under the unconfoundedness assumption of application. Before doing so, we can assess the plausibility of this identifying assumption with the ignorability test of Section 5.1. To this end, we estimated a logit model for the dropout rate on non eligible students, including student covariates (gender, high school grade and type, university school subject, late enrollment), the economic indicator S, and the application indicator A.

The model results are reported in Table 7. Given that the coefficient of A is significant, the ignorability assumption is rejected: applicants appear to be different from non applicants even conditional on covariates; in particular they show a lower dropout rate, even in the absence of the grant.

The logit model used to test the unconfoundeness assumption contains parametric assumptions that can be relaxed, for example using a propensity score matching on non eligible students, with the application status A as the treatment.

6.4. Using DID assumptions

Given that the ignorability assumption is not plausible, in order to estimate the grant effects away from the threshold $\hat{s}$, we investigate on the selection bias induced by the difference between applicants and non applicants, by testing difference-in-difference types of assumption.

Table 7: *Ignorability test: logit model for dropout on non eligible students ($Z = 0$)*

Coefficient	*Estimate*	$s.e.$	Wald chi-square	$p-value$
Intercept	5.2768	2.1226	6.1805	0.0129
Application status A	-0.7453	0.3446	4.6786	0.0305
Economic indicator S	-0.00017	0.000073	5.5190	0.0188
Female	1.1913	0.4124	8.3467	0.0039
High school grade	-0.0410	0.0190	4.6547	0.0310
High school subject				
Humanities	-2.0379	0.7807	6.8150	0.0090
Scientific	-1.1022	0.3669	9.0241	0.0027
Late enrollment	0.9617	0.3531	7.4165	0.0065
University school subject				
Science	-0.2190	0.5553	0.1556	0.6933
Medicine	-1.9178	0.8680	4.8821	0.0271
Humanities	-0.5999	0.4826	1.5452	0.2138
Eng+Arch	0.8474	0.5888	2.0716	0.1501

Condition 5, or Condition 8 on the logit scale, is required in order to apply DID methods. Figure 1 shows, descriptively, that the difference in the dropout rate among non eligible applicants and non applicants may depend on S. We therefore apply a nonlinear DID estimation strategy assuming Condition 8 holds.

The estimation strategy is as follows: we estimated a logit model for Y on eligible students conditional on A, S, and the interaction between A and S. Table 8 reports the estimated coefficients, $\boldsymbol{\beta}' = [\beta_0, \beta_A, \beta_S, \beta_{AS}]$. We then estimated a logit model for Y_0 on non eligible students conditional on A, S, and the interaction between A and S. The estimated coefficients are reported in Table 9, $\boldsymbol{\gamma}' = [\gamma_0, \gamma_A, \gamma_S, \gamma_{AS}]$. An estimate of $f(S)$ in Condition 8 is obtained considering $\gamma_A + \gamma_{AS}S$. An estimate of the unobservable $logitP(Y_0 = 1 \mid S < \widehat{s}, A = 1)$ is obtained by adding $(\gamma_A + \gamma_{AS}S)$ to $logitP(Y_0 = 1 \mid S < \widehat{s}, A = 0) = \beta_0 + \beta_S S$. So the coefficients of logit model on the eligible students are *corrected* according to the the strategy explained in Sections 5.2 and 5.3 provided that Condition 8 holds. Table 10 and Figure 2 report the *corrected* estimates of the grant effects on the treated at different values of the economic indicator S, where the $s.e.$ of the effects are obtained via the delta method.

Table 8: *Logit model for dropout on eligible ($Z = 1$)*

Coefficient	*Estimate*	*s.e.*	*Wald chi-square*	*p-value*
Intercept	-1.4282	0.6592	4.6937	0.0303
A	-1.2980	0.7305	3.1568	0.0756
s	0.000047	0.000097	0.2342	0.6284
A*s	-0.00018	0.000103	2.9432	0.0862

It is worth to note that the grant effect decreases with S. Moreover, consistently

Table 9: *Logit model for dropout on non eligible ($Z = 0$)*

Coefficient	*Estimate*	*s.e.*	*Wald chi-square*	*p-value*
Intercept	-0.4778	0.5561	0.7383	0.3902
A	-0.5930	0.6809	0.7584	0.3838
s	-0.00014	0.000096	2.2543	0.1332
A*s	-0.00006	0.000133	0.2240	0.6360

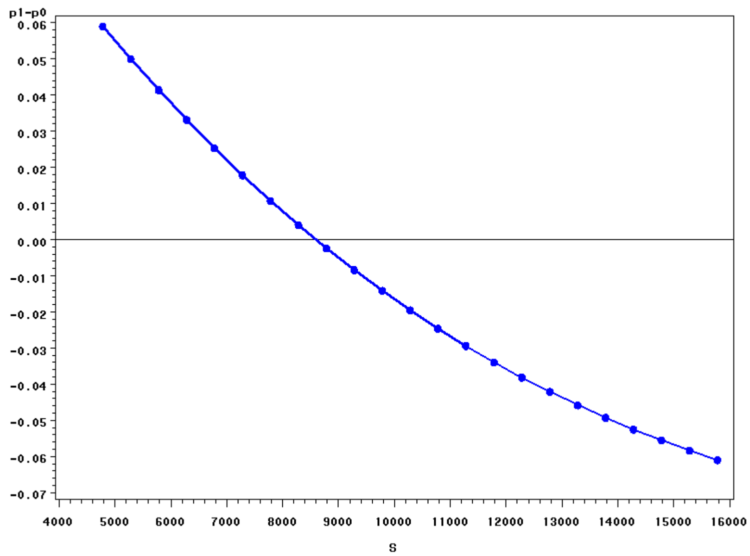

Figure 2: *Grant effect on the treated under non linear DID assumptions*

with the results obtained with the sharp RDD, there is some evidence that the grant is effective in reducing the dropout rate only for richer students, although the effects are not significant. The result suggests that the grant, while effective at the threshold, is on the contrary insufficient to change the decision of the poorest student to drop out of their university studies. This result fills a gap of other studies where nothing was said about the effect of the grant for the students in real needs.

7. Final remarks

In the paper, we have explored different ways of using the groups of eligible non applicants and non-eligible applicants to provide tools for specification testing and identification in a RDD. In particular, the existence of a group of applicants who were denied participation allows to test for the presence of selection bias arising from non random selection of applicants and to exploit (nonlinear) difference-in-difference assumptions to estimate the effect away from the threshold. The evaluation of Italian university grants was used as an illustrative example. Results show that, *at the threshold*,

Table 10: *Grant effect on eligibles under non linear DID assumptions*

S	p_1	$p_1 - p_0$	$s.e.(p_1 - p_0)$
4783.92	0.19358	0.059025	0.24111
5283.92	0.18364	0.050013	0.23104
5783.92	0.17410	0.041396	0.22121
6283.92	0.16496	0.033170	0.21161
6783.92	0.15620	0.025327	0.20225
7283.92	0.14783	0.017862	0.19314
7783.92	0.13983	0.010765	0.18429
8283.92	0.13220	0.004029	0.17572
8783.92	0.12492	-0.002357	0.16742
9283.92	0.11799	-0.008402	0.15943
9783.92	0.11140	-0.014116	0.15175
10283.92	0.10513	-0.019511	0.14442
10783.92	0.09917	-0.024598	0.13746
11283.92	0.09352	-0.029388	0.13090
11783.92	0.08815	-0.033893	0.12478
12283.92	0.08307	-0.038123	0.11914
12783.92	0.07825	-0.042090	0.11401
13283.92	0.07370	-0.045805	0.10945
13783.92	0.06938	-0.049280	0.10550
14283.92	0.06530	-0.052526	0.10220
14783.92	0.06145	-0.055552	0.09958
15283.92	0.05781	-0.058370	0.09768
15783.92	0.05437	-0.060990	0.09650
16283.92	0.05112	-0.063422	0.09605
16783.92	0.04806	-0.065674	0.09632
$p_k = \widehat{P}(Y_k = 1 \mid Z = 1, A = 1, S)$, $k = 0, 1$			

the grant is an effective tool to prevent students from low income families from dropping out of higher education. However, under some relatively weak nonlinear difference-in-difference type of assumptions, they also show that moving away from the threshold 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 proposed tools can be applied to other settings, whenever the threshold for participation is different from the threshold of initial eligibility.

References

Abadie A. (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies*, 72, 1-19.

Agresti, A. (2002). *Categorical Data Analysis*, 2nd Edition, New York: Wiley.

Athey S. and Imbens G. W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, 74 (2), 431-497.

Angrist, J.D., Imbens, G.W., Rubin, D.B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, 91 (434), 444-472.

Angrist J. and Krueger A. (2000). "Empirical Strategies in Labor Economics", in *Handbook of Labor Economics*, Ashenfelter O. and Card D. , eds. North Holland: Elsevier, 1277-1366.

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.

Card D. and Krueger A. (1993). Minimum Wages and Employment: A Case Study of the Fast-food Industry in New Jersey and Pennsylvania, *American Economic Review*, 84 (4), 772-784.

Greene, W. H. (2000). *Econometric Analysis*, 4th Ed., New York: Prentice Hall.

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.

Heckman, J. J. and Hotz, V. J. (1989). Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs. The Case of Manpower Training. *Journal of the American Statistical Association*, 84, 862-880.

Heckman, J. J. and Ichimura, H. and Smith, J. and Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66 (5), 1017-1098.

Imbens, G.W., Angrist, J.D. (1994). Identification and estimation of local average treatment effects. *Econometrica* 62, 467-476.

Imbens G. W. and Lemieux T. Eds (2008). The regression discontinuity design: Theory and applications, 142, Issue 2.

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

LaLonde R. (1986). Evaluating the econometric evaluations of training programmes with experimental data *America Economic Review*, 76, 604-620.

McCrary J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142, 698-714.

Mealli F. and Rampichini C. (2002). Impact Evaluation of University Grants , 2002 Proceedings of the American Statistical Association, Statistical Computing Section [CD-ROM], Alexandria, VA: American Statistical Association.

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

Rosenbaum P.R. (1984). From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment. *Journal of the American Statistical Association*, 79, 41-48.

Rosenbaum P.R. (1987). The role of a second control group in an observational study. *Statistical Science*, 2, 292-306.

Rosenbaum P.R., Rubin D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70, 41-55.

Rubin D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psycology*, Vol. 66, 688-701.

SAS Institute Inc. (2008). *SAS/GRAPH® 9.2 Reference.* Cary, NC: SAS Institute Inc.

Trochim, W. (1984). *Research Design for Program Evaluation: the Regression-Discontinuity Approach*, Beverly Hills: Sage Publications. 18