



HAL
open science

Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs

Erich Battistin, Enrico Rettore

► **To cite this version:**

Erich Battistin, Enrico Rettore. Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs. *Econometrics*, 2007, 142 (2), pp.715. 10.1016/j.jeconom.2007.05.006 . hal-00501793

HAL Id: hal-00501793

<https://hal.science/hal-00501793>

Submitted on 12 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs

Erich Battistin, Enrico Rettore

PII: S0304-4076(07)00114-5
DOI: doi:10.1016/j.jeconom.2007.05.006
Reference: ECONOM 2943

To appear in: *Journal of Econometrics*

Cite this article as: Erich Battistin and Enrico Rettore, Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs, *Journal of Econometrics* (2007), doi:10.1016/j.jeconom.2007.05.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jeconom

Ineligibles and Eligible Non-Participants as a Double Comparison Group in Regression-Discontinuity Designs

Erich Battistin

Department of Statistics, University of Padova
Institute for Fiscal Studies, London

Enrico Rettore

Department of Statistics, University of Padova

May 19, 2007

Abstract

In a sharp Regression-Discontinuity Design (RDD) the participation status deterministically depends on whether a pre-programme characteristic is above or below a specified threshold. The attractiveness of such a design rests on close similarities with a formal experiment. Nevertheless, it is of limited applicability since participation into a programme is seldom determined according to this rule. Besides, in the presence of heterogeneous effects a sharp RDD only allows identification of mean effects for individuals around the threshold for participation. Two results are presented in this paper, and they both partially overcome the two limitations described above. We 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. We also show that, in this set-up, the regularity conditions required for the identification of the mean counterfactual outcome for participants marginally eligible for the programme are essentially the same as in a sharp RDD.

Keywords: programme evaluation; second comparison group; specification tests; treatment effects.

JEL Classification: C4; C8

1 Introduction

In this paper we show how discontinuities in the probability of participation induced by the eligibility criteria for being enrolled in a programme can be used to test the performance of alternative non-experimental estimators of the programme effects.

The central issue in the evaluation of the impact of an intervention is to separate its causal effect from the confounding effect of other factors influencing the outcome of interest. Random assignment of individuals to the intervention defines treatment and control groups that are equivalent in all respects, except for their treatment status. Thus, if a randomized experiment is well implemented, any post-intervention difference between treatment and control individuals can safely be attributed to the intervention itself. The main advantage of experiments is that the assumptions they rest upon are generally more plausible than those made in an observational setting (see Heckman and Smith, 1995). In the latter case, the identification of causal relationships rests on assumptions about individuals' behavior whose plausibility is often controversial.

In the absence of random assignment a fairly favourable situation to the researcher arises when the assignment mechanism leading individuals to participate into the programme is fully specified. This situation applies to those instances in which participation follows a *sharp* Regression Discontinuity Design (RDD; see Thistlethwaite and Campbell, 1960, Rubin, 1977, Trochim, 1984). In this design, assignment to the programme solely depends on whether one or more observable pre-intervention variables satisfy a set of conditions *known* to the analyst. As an example, we can think of situations in which individuals willing to participate are divided into two groups according to whether or

not a pre-intervention measure exceeds a known threshold, but only individuals scoring above that threshold are assigned to the programme.

This design features both advantages and disadvantages with respect to its non-experimental competitors. On the one hand, in a neighborhood of the threshold for selection a sharp RDD presents some features of a pure experiment. In this sense, it is certainly more attractive than other non-experimental designs. Since individuals in the treatment and comparison group solely differ with respect to the variable determining the participation status (and with respect to any other variable correlated to it), one can control for confounding factors by contrasting marginal participants to marginal non-participants. In this context, the term ‘marginal’ refers to those units *not too far* away from the threshold for selection.

The comparison of mean outcomes for participants and non-participants at the margin identifies the mean impact of the intervention *locally* at the threshold for selection. Intuitively, for identification at the cut-off point to hold it must be the case that any discontinuity in the relationship between the outcome of interest and the variable determining the treatment status is fully attributable to the treatment itself. This requires some regularity conditions at the threshold for selection discussed by Hahn *et al.* (2001; HTV in the following).

On the other hand, the sharp RDD features two main limitations. First, its feasibility is confined to those instances in which assignment takes place *only* on observable pre-intervention variables; as a matter of fact, this is not often the case. Second, even when such a design applies, it only allows identification of the mean impact of the intervention at the threshold for selection. In the common situation of heterogeneous returns to participation, the local effect might be very

different from the effect for individuals away from the threshold for selection. To identify the mean impact on a broader population of participants one has to rely on non-experimental estimators, whose consistency for the intended impact rests on assumptions about the behaviour of individuals.

Throughout this paper we consider the case in which an eligibility rule splits the relevant population into eligible and ineligible individuals, and participation of the former group is determined according to rules potentially *unknown* to the researcher. Examples of such a design are labour market programmes for which participation is voluntary for individuals satisfying a condition on age, or means tested programmes. College enrollment, for which only a subset of applicants is enrolled amongst those passing an achievement test, also fits this design.

The plan of the paper is as follows. First, we show that in the set-up described above the mean impact for participants around the threshold for eligibility is identified essentially under the same regularity conditions required in a sharp RDD, no matter how eligible individuals are selected into the programme (see Section 3). Although references have been made in the literature to the potential of using eligibility rules to identify mean impacts (see Heckman, 1992, Angrist, 1998, Heckman *et al.*, 1999, and van der Klaauw, 2002), to the best of our knowledge the regularity conditions required for identification have not been discussed so far.

Second, we show that eligibility rules for participation into a programme can be used to assess the validity of non-experimental estimators for the programme effects (see Section 4). In particular, we show that the selection bias arising from the non-random selection of eligible individuals into the programme is identified at the threshold for eligibility, so that one can formally test whether any of the

long array of existing non-experimental estimators can correct for this bias. If this hypothesis is not rejected at the threshold for eligibility, one may feel more confident to use that non-experimental estimator to identify the causal effect on a broader population (typically, the one represented by all participants).

Several links to the literature are established. In particular, we show that our first result is closely related to what discussed by Bloom (1984), Heckman (1990) and Angrist and Imbens (1991). We also stress the relationship with the idea in Rosenbaum (1987) of using two comparison groups for the identification of causal effects. Finally, we point out similarities between our specification test and the set of specification tests derived by Heckman and Hotz (1989), as well as the link to the characterization of the selection bias provided by Heckman *et al.* (1998).

The remainder of this paper is organized as follows. Section 2 discusses similarities between a fully randomized experiment and a RDD. Section 3 describes the set-up in which participation is determined by self-selection amongst eligible individuals and derives the regularity conditions required for identification of the average effect on marginal participants. Section 4 shows how to use this set-up to validate non-experimental estimators for the treatment effect. Section 5 presents an empirical application of the identification strategies discussed in the paper and Section 6 concludes.

2 Identification of treatment effects in Regression Discontinuity Designs

This section presents the basic features of a RDD and highlights the similarities with a randomized experiment. The discussion of identification issues arising in

a RDD is based on HTV, to which the interested reader is referred for further details.

Following the notation of the potential outcome approach to causal inference, let (Y_1, Y_0) be the two potential outcomes one would experience by participating and not participating into the programme, respectively.¹ The causal effect of the treatment on *that* specific subject is then defined as the difference between these outcomes, $\beta = Y_1 - Y_0$, which is not observable since being exposed to (denied) the programme reveals Y_1 (Y_0) but conceals the other potential outcome.²

Let I be the binary variable for the treatment status, with $I = 1$ for participants and $I = 0$ for non-participants. If the assignment is determined by randomization and subjects comply with the assignment, the following condition holds true by construction

$$(Y_1, Y_0) \perp I.$$

The attractiveness of randomization is that the difference between mean outcomes for participants and non-participants identifies the mean impact of the programme

$$E\{\beta\} = E\{Y_1|I = 1\} - E\{Y_0|I = 0\}, \quad (1)$$

since conditioning on I in the right-hand side of (1) is irrelevant by construction. In other words, randomization allows the researcher to use information on non-participants to identify the mean counterfactual outcome for participants,

¹For reviews of the evaluation problem see Heckman, LaLonde and Smith (1999) and Imbens (2004). For the potential outcome framework, the main references are Fisher (1935), Neyman (1935), Quandt (1972), Roy (1951) and Rubin (1974).

²In what follows we will discuss the case in which the programme impact β varies across individuals, which in most instances is the relevant case. HTV also derive the regularity conditions required for the identification of treatment effects in the case of constant impact across individuals.

namely what participants would have experienced had they not participated into the programme.

A RDD arises when the treatment status depends on an *observable* individual characteristic S and there exist a *known* point in the support of S where the probability of participation changes discontinuously. Throughout this paper, S is assumed to be continuous on the real line. Formally, if \bar{s} is the discontinuity point, then a RDD is defined if

$$Pr\{I = 1|\bar{s}^+\} \neq Pr\{I = 1|\bar{s}^-\}. \quad (2)$$

Here and in the following \bar{s}^+ and \bar{s}^- refer to those individuals *marginally* above and below \bar{s} , respectively. Moreover, to ease the exposition and without any loss of generality, we will deal with the case in which the following inequality holds

$$Pr\{I = 1|\bar{s}^+\} - Pr\{I = 1|\bar{s}^-\} > 0.$$

Following Trochim (1984), the distinction between *sharp* and *fuzzy* RDD depends on the size of the discontinuity in (2). The former design occurs when the probability of participating conditional on S steps from zero to one as S crosses the threshold \bar{s} . That is, the treatment status deterministically depends on whether individuals' values of S are above \bar{s}

$$I = \mathbb{1}(S \geq \bar{s}). \quad (3)$$

A fuzzy RDD occurs when the size of the discontinuity at \bar{s} is smaller than one. For example, a fuzzy RDD can be thought as an instance in which I is a deterministic function of S for all subjects but this function is different across individuals (see HTV).

Although a RDD lacks random assignment of individuals to the treatment group, it shares an attractive feature with experimental designs. We will address this issue by considering the case of a sharp RDD; we will discuss the fuzzy case further below in this section. Let

$$Y = Y_0 + I(s)\beta$$

be the observed outcome as it results from taking or not taking part into the programme. The dependence of the treatment status I on the variable S is stressed by writing $I(s)$. The difference of mean outcomes for individuals marginally above and below the threshold \bar{s}

$$E\{Y|\bar{s}^+\} - E\{Y|\bar{s}^-\} \quad (4)$$

can be written as

$$E\{Y_0|\bar{s}^+\} - E\{Y_0|\bar{s}^-\} + E\{I(s)\beta|\bar{s}^+\} - E\{I(s)\beta|\bar{s}^-\}, \quad (5)$$

which simplifies to

$$E\{Y_0|\bar{s}^+\} - E\{Y_0|\bar{s}^-\} + E\{\beta|\bar{s}^+\}$$

because of (3). The following condition is then sufficient for the mean impact of the treatment at \bar{s}^+ to be identified in a sharp RDD (it corresponds to Assumption (A1) in HTV).

Condition 1. The mean value of Y_0 conditional on S is a continuous function of S at \bar{s} .

Accordingly, Condition 1 requires that in the counterfactual world no discontinuity would take place at the threshold for selection. The attractiveness of the RDD is apparent here. By contrasting mean outcomes for participants

and non-participants at the margin, one can identify the average impact of the programme for individuals in a right-neighborhood of \bar{s} , thus a local version of the parameter in (1)

$$E\{\beta|\bar{s}^+\} = E\{Y|\bar{s}^+\} - E\{Y|\bar{s}^-\}.$$

The identification of $E\{\beta|\bar{s}^-\}$, that is of the mean impact from extending the programme to marginally excluded individuals, requires an additional continuity condition on the conditional mean $E\{Y_1|S\}$.³ This additional assumption together with Condition 1 are equivalent to Assumptions (A1) and (A2) in HTV. Note that, by exploiting the relationship in (3) and by assuming that the distribution of (Y_0, Y_1) as a function of S is continuous at the discontinuity point, the following condition holds true

$$(Y_1, Y_0) \perp I | S = \bar{s}. \quad (6)$$

Because of this property, perhaps a bit rethorically, a sharp RDD is often referred to as a quasi-experimental design (Cook and Campbell, 1979).^{4 5}

When individuals do not comply with the mandated status resulting from a sharp assignment, drop out of the programme or seek alternative treatments if denied it (see, for example, Battistin and Rettore, 2002), a fuzzy RDD arises.

In this case, the continuity of Y_0 and Y_1 at \bar{s} is no longer enough to ensure

³Note that, in practice, it is difficult to think of cases where Condition 1 is satisfied and the same condition does not hold for Y_1 .

⁴It is worth stressing again that to meaningfully define marginal units (with respect to \bar{s}) the selection variable S has to be continuous. Estimation of the right-hand side (left-hand side) of (4) makes use of data only in a neighborhood on the right (left) side of the discontinuity point. Unless one is willing to make some parametric assumptions about the regression curve away from \bar{s} , only data local to the discontinuity point help to estimate the jump. Asymptotically the neighborhood needs to shrink as with usual non-parametric estimation, implying a non-standard asymptotic theory for the resulting estimator of the mean impact (see HTV and Porter, 2002).

⁵Lee (2004) suggests simple tools to test for the validity of (6) based on the idea of comparing individuals marginally above and below the threshold with respect to variables which *cannot* be affected by the programme. Finding that the two groups of individuals present systematic differences in the values of these variables would cast serious doubts on the validity of (6).

the orthogonality condition in (6). Accordingly, the mean impact at \bar{s} cannot be identified by simply comparing the mean outcome for marginal participants to the mean outcome for marginal non-participants, and additional conditions are required to recover meaningful causal parameters from (4). The following condition corresponds to Assumption (A3)(i) in HTV.⁶

Condition 2. The triple $(Y_0, Y_1, I(s))$ is stochastically independent of S in a neighborhood of \bar{s} .

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

If Condition 2 is satisfied, then expression in (4) can be written as

$$E\{\beta|I(\bar{s}^+) > I(\bar{s}^-)\}Pr\{I(\bar{s}^+) > I(\bar{s}^-)\} - E\{\beta|I(\bar{s}^+) < I(\bar{s}^-)\}Pr\{I(\bar{s}^+) < I(\bar{s}^-)\}.$$

Then, under the additional

Condition 3. Participation into the programme is *monotone* around \bar{s} , that is

it is either the case that $I(\bar{s}^+) \geq I(\bar{s}^-)$ for all subjects or the case that

$I(\bar{s}^+) \leq I(\bar{s}^-)$ for all subjects.

⁶As an alternative to Condition 2, HTV also consider the following local orthogonality condition

$$(Y_1, Y_0) \perp I | S = \bar{s},$$

which rules out non-random selection based either on (Y_1, Y_0) or on any variable stochastically related to (Y_1, Y_0) .

the mean impact

$$E\{\beta|I(\bar{s}^+) \neq I(\bar{s}^-)\} = \frac{E\{Y|\bar{s}^+\} - E\{Y|\bar{s}^-\}}{E\{I|\bar{s}^+\} - E\{I|\bar{s}^-\}}, \quad (7)$$

is identified (Condition 3 corresponds to Assumption (A3)(ii) in HTV). The parameter in (7) represents the mean impact of the programme on those individuals in a neighborhood of \bar{s} who would switch their treatment status if the threshold for participation switched from just above their score to just below it (see Imbens and Angrist, 1994, and Angrist *et al.*, 1996).

It is worth noting that Condition 3 is an assumption on individuals' behavior which is not testable. It corresponds to assuming that the individual specific function $I(s)$ is monotone the same way for all subjects in a neighborhood of \bar{s} . Moreover, if Condition 3 holds, the denominator in the right-hand side of the previous expression identifies the proportion of complying individuals at \bar{s} , that is the subpopulation the effect in (7) refers to. Nevertheless, it is not observable which individuals the group of compliers consists of. Whether or not the resulting mean impact is a policy relevant parameter depends on the specific case (see Heckman, 1997, for a discussion).

Apparently, a sharp RDD allows the identification of the mean impact on a broader population than the one identified in a fuzzy RDD. Moreover, stronger regularity conditions are needed in the fuzzy case. While in the sharp case Condition 1 is sufficient to ensure the identification of the mean impact for marginal participants, in the fuzzy case Condition 2 and Condition 3 together imply that the impact on the subpopulation of compliers in a neighborhood of \bar{s} is identified. Heckman *et al.* (1999) emphasize this point by saying that much of the simplicity of the design is lost moving from a sharp RDD to a fuzzy RDD.

Two major drawbacks hamper the applicability of the RDD. Firstly, in an

observational study it is very often the case that units self-select into the treatment rather than being exogenously selected on a pre-programme measure. If this is the case, the RDD set-up as introduced so far no longer applies. Secondly, even in those instances in which the RDD applies, if the impact is heterogeneous across individuals such a design is not informative about the impact on individuals away from \bar{s} . These two issues will be dealt with in the next sections.

3 A partially fuzzy design

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 labour 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.

To fix ideas, let S be a continuous pre-programme characteristic and let the *eligibility status* be established according to the deterministic rule $\mathbb{1}(S \geq \bar{s})$. That is, subjects are eligible for the programme if and only if they present a value of the variable S above a known threshold \bar{s} . Throughout our discussion it will be assumed that S is observable for all individuals. If all eligibles (and only them) participated into the programme, a sharp RDD would arise. For example, if participation were mandatory for all eligible individuals, the effect of the programme at the threshold for eligibility would be identified by (4) provided that Condition 1 holds.

As a matter of fact, it is often the case that some eligible individuals self-select into the programme while some others do not (typically, when participation is on a voluntary basis). Individuals' heterogeneity about information on

the availability of the programme, preferences and opportunity costs are factors likely to influence the participation decision in several instances. Accordingly, the population turns out to be divided into three subgroups: *ineligibles*, *eligible non-participants* and *participants*. Our analysis develops with reference to the general case in which the researcher has no knowledge of the rule leading eligible individuals to self-select into the programme (that is, selection is on *unobservables*). In what follows, it will be assumed that information on the three groups of individuals is available to the researcher.

As a result of the eligibility rule and of self-selection, the probability of participation for those individuals scoring a value of S below the threshold \bar{s} is zero by definition, since they are not eligible for the programme. The probability of participation for those scoring above \bar{s} is smaller than one because participation is not mandatory. As a result, the probability of participation is discontinuous at the threshold for eligibility and the size of the discontinuity is less than one (i.e. according to the terminology introduced in the previous section, a fuzzy RDD is defined).

Van der Klaauw (2002, p.1284) explicitly mentions the potential for using the RDD arising from the eligibility criteria for a social programme. Heckman *et al.* (1999, pp.1971-1972) recognize the fuzzy RDD nature of this set-up and point out that in this case the estimand in (7) identifies the mean impact on participants at \bar{s} .

3.1 Identification of programme effects

To recover the regularity conditions required for identification consider again the difference in (4). Since participation is precluded to marginally ineligible

($I(\bar{s}^-) = 0$), the expression in (5) becomes

$$E\{Y_0|\bar{s}^+\} - E\{Y_0|\bar{s}^-\} + E\{I(s)\beta|\bar{s}^+\}.$$

If Condition 1 holds, by using the law of iterated expectations and by noting that $E\{I(s)\beta|I = 0, \bar{s}^+\} = 0$ the previous expression equals

$$E\{I(s)\beta|\bar{s}^+\} = E\{\beta|I = 1, \bar{s}^+\}Pr\{I = 1|\bar{s}^+\},$$

so that the mean impact on participants in a right-neighborhood of \bar{s} is identified by

$$E\{\beta|I = 1, \bar{s}^+\} = \frac{E\{Y|\bar{s}^+\} - E\{Y|\bar{s}^-\}}{E\{I|\bar{s}^+\}}. \quad (8)$$

In other words, Condition 1 is sufficient for the average effect of the treatment on the treated to be identified locally at the threshold for eligibility \bar{s} .

It turns out that, despite the *prima facie* fuzzy RDD nature of this set-up, the LATE (Imbens and Angrist, 1994) at the discontinuity point is identified under the same condition used to estimate the average treatment effect in the sharp design. The result rests on the fact that the probability of participation on the left-hand side of \bar{s} is zero by design, and this simplifies the expression in (5) without further assumptions on individuals' behavior.⁷

An alternative way of deriving the result exploits a close analogy between the set-up discussed in this section and the one in Bloom (1984). In a fully experimental setting, Bloom (1984) notes that even if some individuals randomly assigned to the treatment eventually do not show-up, the identification of the mean impact on participants is still secured by the experiment. By analogy, exploiting the fact that the eligibility rule defines a randomized experiment in

⁷Results by HTV and Porter (2002) on non-parametric inference in a RDD straightforwardly apply to the estimation of (8).

a neighborhood of \bar{s} and the fact that eligible non-participants play the role of Bloom's (1984) no-shows, the intuition suggests that the mean impact on participants in a neighborhood of \bar{s} is also identified.

This is exactly what we have derived so far. The key relationship identification rests upon is the equality stated by Condition 1, that is

$$E\{Y_0|\bar{s}^+\} = E\{Y_0|\bar{s}^-\}. \quad (9)$$

The left-hand-side of (9) can be written as the weighted average of the mean outcome for eligible participants and for eligible non-participants, respectively

$$E\{Y_0|I = 1, \bar{s}^+\}\phi + E\{Y_0|I = 0, \bar{s}^+\}(1 - \phi),$$

where $\phi = E\{I|\bar{s}^+\}$ is the probability of self-selection into the programme conditional on marginal eligibility. The last expression combined with (9) yields

$$E\{Y_0|I = 1, \bar{s}^+\} = \frac{E\{Y_0|\bar{s}^-\}}{\phi} - E\{Y_0|I = 0, \bar{s}^+\}\frac{1 - \phi}{\phi}. \quad (10)$$

Namely, the *counterfactual* mean outcome for marginal participants is identified by a linear combination of *factual* mean outcomes for marginal ineligible and for marginal eligibles not participating into the programme. The coefficients of this combination add up to one and are a function of ϕ , which is identified from observed data.

Hence, equation (10) implies that the mean impact on participants is identified, since by definition

$$E\{\beta|I = 1, \bar{s}^+\} = E(Y_1|I = 1, \bar{s}^+) - E(Y_0|I = 1, \bar{s}^+).$$

The right-hand-side of the previous expression can be rearranged using (10) to obtain the expression in (8). The result that the counterfactual mean outcome for marginal participants is identified will play a crucial role in the next section.

The identification of $E\{\beta|I = 1, \bar{s}^-\}$, namely the impact on individuals who would participate into the programme if the threshold for eligibility were marginally reduced, requires regularity conditions stronger than Condition 1. More precisely, if Condition 2 holds then the following two equalities are jointly satisfied

$$E\{I(s)\beta|\bar{s}^-\} = E\{I(s)\beta|\bar{s}^+\},$$

$$E\{I|\bar{s}^-\} = E\{I|\bar{s}^+\}.$$

As the former equality implies

$$E\{\beta|I = 1, \bar{s}^-\}E\{I|\bar{s}^-\} = E\{\beta|I = 1, \bar{s}^+\}E\{I|\bar{s}^+\},$$

we have

$$E\{\beta|I = 1, \bar{s}^-\} = E\{\beta|I = 1, \bar{s}^+\}.$$

3.2 Related results

Five general comments follow from the results presented so far.

First, the comparison of participants to ineligible individuals or the comparison of participants to eligible non-participants do not allow identification of any causal parameter. It follows from Bloom's (1984) key result in (10) adapted to the RDD case that identification of the mean counterfactual outcome for participants around \bar{s} is achieved by using information on *both* groups of non-participants. In this sense, the availability of information separately on *three* groups of subjects - participants, eligible non-participants and ineligibles - is crucial for identification.

Second, as pointed out by Little and Yau (1998) following the discussion in Angrist *et al.* (1996), the validity of the result in Bloom (1984) rests upon

an exclusion restriction stating that, net of the actual participation status, the random assignment is irrelevant for the outcome of interest (see also Heckman *et al.*, 1999). By analogy, an exclusion restriction must be satisfied in our set-up locally at \bar{s} . Note that such a restriction is implicit in Condition 1: if, by fixing the actual participation status, Y_0 were affected by the status defined by the eligibility rule, then $E(Y_0|S)$ would be discontinuous at \bar{s} .

Third, to achieve identification we do not need to model how eligible individuals self-select (or are selected) into the programme. Thus, identification of the mean impact does not need any behavioral assumptions on the selection process.

Fourth, our result (as well as Bloom's, 1984, one) can also be derived as a special case of Heckman (1990) and Angrist and Imbens (1991). The authors prove that even if participation into the programme takes place as a result of self-selection, the mean impact on participants is identifiable provided that (i) there exists a random variable Z affecting the participation into the programme and orthogonal to the potential outcomes and (ii) the probability of participation conditional on Z is zero in at least one point of the support of Z . Condition (i) qualifies Z as an instrumental variable for the problem.

In our case, since the eligibility status is orthogonal to the potential outcomes in a neighborhood of \bar{s} and $Pr(I = 1|\bar{s}^-) = 0$, the eligibility status meets the conditions stated by Angrist and Imbens (1991) in a neighborhood of \bar{s} . Identification of the mean impact on participants at \bar{s} follows (see the discussion in HTV on the property of the IV estimator in this instance). Accordingly, in the set-up considered in this paper S is exploited both as a control variable and

to define an instrumental variable.⁸

Finally, the case in which the partially fuzzy design is characterised by the conditions $Pr(I = 1|\bar{s}^+) = 1$ and $Pr(I = 1|\bar{s}^-) > 0$ straightforwardly fits our framework. Moving from the continuity restriction $E\{Y_1|\bar{s}^+\} = E\{Y_1|\bar{s}^-\}$ one recovers $E\{Y_1|I = 0, \bar{s}^-\}$, the counterfactual mean outcome for non-participants at \bar{s}^- . Then, both the mean impact for non-participants at \bar{s}^- and the selection bias with respect to Y_1 at \bar{s} are identifiable.

4 Validating non-experimental estimators of the mean impact on participants

4.1 Specification tests

This section shows that if data are available on the three groups of individuals resulting from the set-up discussed in Section 3, one can use the information around the threshold for eligibility to test for the validity of non-experimental estimators of the programme effect for participants *away* from the threshold.

As pointed out in the previous section, Condition 1 is sufficient to identify the mean impact of the programme on participants *marginally* eligible for it, even if they are a non-random sub-group from the set of eligible individuals. However, if the gain from being exposed to the programme is heterogeneous with respect to S , the mean impact for individuals in a neighborhood of the threshold for eligibility is *not* informative on the impact of the programme for individuals away from this point, thus precluding identification of the effect for *all* participants.⁹

⁸In a fully parametric set-up (and under the additional linearity assumption) the following regression could be estimated

$$Y = \alpha_0 + \alpha_1 S + \alpha_2 I + \varepsilon,$$

using the eligibility status $Z = \mathbb{1}(S \geq \bar{s})$ as an instrumental variable for I with S entering the equation to control for the selection induced by the eligibility rule.

⁹Note that local identification provided by the RDD can be used to test for the hetero-

In order to identify the mean impact on participants *away from* \bar{s}

$$E\{\beta|I = 1, s\} = E\{Y_1|I = 1, s\} - E\{Y_0|I = 1, s\}, \quad s \geq \bar{s}$$

one must resort to one of the long array of non-experimental estimators available in the literature which adjust for selection bias under alternative assumptions (see Heckman *et al.*, 1999, Blundell and Costa Dias, 2000, and Imbens, 2004, for a review). The main problem with non-experimental identification strategies is that these assumptions are intrinsically not testable.

To fix the ideas, in what follows we will focus on the class of non-experimental estimators based on the assumption of selection on observables, but the same line of reasoning applies to other estimators. If the restriction of selection on observables holds true, all the variables driving self-selection of individuals *and* stochastically related to the potential outcomes are observable to the researcher. Formally, identification of the mean impact on participants rests on the existence of an observable vector of individual characteristics X such that the following conditions are jointly satisfied

$$Y_0 \perp I | S, X, \quad S \geq \bar{s} \quad (12.a)$$

$$\text{var}\{I = 1 | S, X\} > 0, \quad S \geq \bar{s}. \quad (12.b)$$

Then, it is as if eligible individuals were randomly assigned to the treatment with a probability of assignment depending on S and X , provided that such probability is non-degenerate at each value of these variables. If the set of variables X is rich enough for this assumption to be credible, the counterfactual

generality of the effect. Under a smoothness condition, a constant (with respect to S) treatment impact implies that the derivatives of the regression curve are the same on both sides of \bar{s} . One could test for this by directly modelling the regression of Y on S and I . Alternatively, a non-parametric implementation of this idea could compare the partially linear (Battistin and Rettore, 2002; Porter, 2002) and the local polynomial (HTV, 2002) estimators of the discontinuity, which have the same convergence rate under constant treatment effects (Porter, 2002)

outcome for participants presenting characteristics (S, X) can be approximated by the actual outcome of non-participants presenting the same characteristics.¹⁰ It follows that the effect on participants away from the threshold \bar{s} could be identified after controlling for X at each given value of S .

Let

$$sb(s) = E\{Y_0|I = 1, s\} - E\{Y_0|I = 0, s\}, \quad s \geq \bar{s} \quad (13)$$

be the *selection bias* that affects the raw comparison of participants and eligible non-participants scoring $S = s$, with $S \geq \bar{s}$. The first term on the right-hand-side of (13) is the mean counterfactual outcome for participants. The second term is the mean factual outcome for eligible non-participants. This quantity summarises pre-programme differences between eligible individuals self-selected in and out of the programme at each level of S , with $S \geq \bar{s}$.

Using the result stated in (10), the mean counterfactual outcome for participants on the right hand side of (13) is identified in a neighborhood of \bar{s} provided that Condition 1 holds. Accordingly $sb(\bar{s}^+)$, the selection bias for individuals marginally eligible for the programme, is also identified. Clearly, identification is precluded as S moves away from \bar{s} .

Let

$$sb(s, x) = E\{Y_0|I = 1, s, x\} - E\{Y_0|I = 0, s, x\}, \quad s \geq \bar{s}$$

be the bias term for a specific subpopulation of eligibles indexed by x and s , where X are the variables claimed to properly account for the selection bias.

If the orthogonality condition (12.a) holds, then $sb(s, x) = 0$ uniformly with

¹⁰The two conditions stated in (12) are stronger than is required, as conditional mean independence, instead of full independence, would suffice to identify the impact of the treatment on participants and the second condition need not to hold for values of X with no participants. Note that (12.b) implies that both participants and non-participants can be found at each value of (S, X) , $S \geq \bar{s}$, thus ruling out any common support problems.

respect to x and s . In particular, a necessary condition for the validity of the estimator based on the ignorability condition (12.a) is

$$sb(\bar{s}^+, x) = 0, \quad (14)$$

which is directly testable since $sb(s, x)$ is identified in a right-neighborhood of \bar{s} (provided Condition 1 holds). It follows that in a neighborhood of \bar{s} any test of the equality of the mean outcomes for ineligible and eligible non-participants conditional on X is a test for the ignorability of participation into the programme, thus a test on the validity of the estimator that corrects for X .

The rejection of the null hypothesis is sufficient to conclude that condition (12.a) does not hold. On the other hand, by not rejecting the null hypothesis one might feel more confident in controlling for X to estimate the effect away from \bar{s} . However, the acceptance of the null hypothesis is not conclusive about the validity of the estimator for the whole population of participants: in fact, the test is not informative on whether the ignorability condition holds away from \bar{s} .

The testing procedure can be non-parametrically implemented along the lines of what discussed by HTV and Porter (2002). Once the different composition with respect to the X 's of ineligible and eligible non-participants around the threshold is accounted for, an estimate for the discontinuity of the regression of Y on S at \bar{s} is required. Porter's (2002) approach yields more precise estimates of such quantity, hence more powerful test of the no discontinuity hypothesis, at the cost of imposing extra regularity conditions. Battistin and Rettore (2002) point out that under the no discontinuity hypothesis it is hard to imagine cases in which such extra regularity conditions are not met. Accordingly, the strategy

suggested by Porter (2002) seems to be preferred.¹¹

4.2 Related results

The results presented in the previous section are easily summarized as follows. The ignorability condition (12.a) represents an identifying restriction for the effect of the programme on participants. On the other hand, for the set-up considered in this paper the condition in (14) represents an over-identifying restriction for the same parameter in a neighborhood of \bar{s} . Hence, the parameter of interest is *locally* over-identified, and condition (12.a) can be tested at \bar{s} . Exploiting restrictions to test the assumptions non-experimental estimators rest upon is also the idea in Rosenbaum (1984) and Heckman and Hotz (1989).

Strong similarities with other results presented in the literature can be established. First, since a RDD can be interpreted as an experiment at \bar{s} , the specification test presented above develops along the same lines of what LaLonde (1986) and Heckman *et al.* (1998) develop in a fully experimental set-up. In both cases there is a benchmark estimate of the mean impact - the RDD estimate in the former case, the experimental estimate in the latter case - to which the researcher is ready to give credibility. Then, the researcher compares non-experimental estimates to “benchmark” estimates and interprets any discrepancy thereof as a violation of the identifying restrictions the non-experimental estimator rests upon¹².

¹¹The core of the testing procedure can easily be implemented by exploiting the parametric set-up presented at the end of Section 3. By considering information for ineligible and eligible non-participants (and assuming linearity throughout), the following regression can be estimated

$$Y = \alpha_0 + \alpha_1 S + \alpha_2 X + \alpha_3 Z + \varepsilon,$$

where $Z = \mathbb{1}(S \geq \bar{s})$ denotes the eligibility status, and then test for $\alpha_3 = 0$.

¹²Interpreting the specification test as a comparison of a non-experimental estimate of the mean impact on participants to the RDD benchmark clarifies why the specification test is unfeasible in the standard fuzzy RDD. As in a fuzzy set-up the RDD estimand corresponds to the mean impact on compliers at \bar{s} , non-experimental estimates of this parameter cannot be recovered being the set of compliers unobservable to the researcher.

Experimental data allow to characterize the selection bias and to test the validity of non-experimental estimators for the effect on participants (see Heckman *et al.*, 1998). In the current set-up, this result is attained only with reference to participants around \bar{s} . However, the availability of experimental data is rarely encountered in the evaluation of social policies, especially in European countries, while it is very often the case that a policy is targeted to a population of eligible individuals whose participation into the programme is on a voluntary basis. In this situation the information required to implement the specification test described in the previous section is in principle available. This provides researchers with a tool to validate non-experimental estimators of the mean impact on participants.

Finally, while discussing the role of a second comparison group in observational studies Rosenbaum (1987, see example 2) provides an example which resembles, albeit loosely, the set-up we refer to. The Advanced Placement programme (APP) provides students with the opportunity to earn college credits for work done in high school. Some schools do not offer APP, and in those that do only a small minority of students participate. Two comparison groups naturally arise in this context: students enrolled in high schools not offering APP and students who do not participate although enrolled in schools offering APP.

Rosenbaum (1987) argues that the two comparison groups can be exploited to test the ignorability condition on which the matching estimator for the effect of APP on participants relies. This can be achieved by testing whether the two groups present the same mean outcomes after their composition is balanced by using a matching procedure. Clearly Rosenbaum's (1987) comparison groups resemble ineligible and eligible non-participants of our set-up. The crucial dif-

ference is that in the former case the rule according to which high schools decide whether to offer APP is unknown to the researcher while in the current set-up the eligibility rule is fully specified. As a consequence, the null hypothesis tested by Rosenbaum (1987) is whether the matching procedure is able to compensate for the selection bias induced by the *two-stage* (schools and students) selection process. Our null hypothesis (14) is whether the matching procedure is able to compensate only for the selection bias induced by self-selection amongst the eligibles, locally at \bar{s} . Otherwise stated, Rosenbaum (1987) puts a heavier burden on the matching procedure than we do by assuming that it solves for a more complex selection process.

5 An empirical application

In this section we use survey data collected for the evaluation of PROGRESA (*programa de Educación, Salud y Alimentación*) to apply the estimation and specification strategies discussed above.

The PROGRESA programme aimed at encouraging investments in education, health and nutrition through relatively large monetary transfers and in-kind benefits given to poor households in rural Mexico. The programme had three main components: an education component, for which transfers were contingent upon children's regular attendance at school; a health component, which consisted of a vaccination programme, growth and development check ups for children as well as courses for mothers; and finally, a component consisting of monetary and nutritional supplements supplied to infants and lactating mothers.¹³

¹³The evaluation of the impact of PROGRESA was done by IFPRI in a series of reports that are summarized in Skoufias (2001).

Participation of households into the programme was originally designed to be the result of a two step procedure (see Behrman and Todd, 1999). In the first step, around 500 localities sharing a high degree of marginality were randomized into a ‘pilot’ group and a ‘control’ one, respectively. In the second step, only *poor* households living in ‘pilot’ localities were considered eligible to receive the programme, the eligibility status being determined on the basis of a *poverty index* at the household level obtained from a discriminant analysis applied separately on each of the seven geographical regions involved in the experiment.

As discussed by Buddelmeyer and Skoufias (2003; see Table 1), a simple descriptive analysis confirms that eligibility was well approximated by a *sharp* design for five out of seven regions, with eligible households scoring values of the discriminant score *below* a region-specific threshold.¹⁴ The programme was rolled out in ‘pilot’ localities in July 1998.

All households in the evaluation sample were surveyed for the first time in October/November 1997, thus before the programme was introduced, and were re-interviewed six times over the three following years. The criterion described above resulted in just above 50% of the households in the total sample being eligible for receiving the programme. However, this number increased to nearly 80% of the evaluation sample during 1998 as a result of a major revision to the eligibility rule (the *densificaciòn*, which is the Spanish name that was given to this process). Such a revision was undertaken to increase the number of households with certain characteristics that were felt to be under-represented when the eligibility status was first determined, although not much is known

¹⁴In fact, as discussed by Buddelmeyer and Skoufias (2003), these thresholds fell roughly at the mode of the distribution of the poverty index and were approximately the same across regions.

about the criteria actually followed to boost the eligible population (see, for example, the discussion in Buddelmeyer and Skoufias, 2003).

Despite this revision process, administrative errors prevented the majority of the boosted sample from receiving the programme. In particular, according to official payment records around 60% of this group had not received any benefits by March 2000, mostly because they were never incorporated into the programme due to some administrative errors and, only to a minor extent, because of non-compliance (see Hoddinott *et al.*, 2000).

In our analysis we exclusively focus on ‘pilot’ localities, in which four groups of households can be defined on the basis of their participation status: *eligibles*, i.e. participating households who were eligible for the programme before the revision was undertaken; *densificados*, i.e. participating households who became eligible for the programme after the *densificaciòn*; *ineligibles*, i.e. households who were not eligible for the programme after the *densificaciòn*; *forgotten densificados*, i.e. households who did not participate in the programme though they were eligible for it after the *densificaciòn*.

Because of the randomization, the comparison of the mean outcome for *eligibles* in the ‘pilot’ areas to the mean outcome for *eligibles* in the ‘control’ areas identifies the mean impact on *eligibles*. Note however that the experimental design fails to identify the mean impact on participants, namely *eligibles* and *densificados*, since the selection of *densificados* from the pool of households who were ineligible before the *densificaciòn* took place according to a non-random and at least partly unknown rule.

**** TABLE 1 AROUND HERE ****

Table 1 presents a breakdown of households by the aforementioned groups

using data from the second follow-up (November 1998).¹⁵ For simplicity, we keep only households living in the five regions where the eligibility status before the *densificaciòn* is well approximated by a *sharp* rule, dropping the very few households whose index was not consistent with the eligibility status.¹⁶ Separately for each group of households, we report the minimum value, the maximum value and the mean value of the poverty index, whose values have been standardized by within-region standard deviations after calculating the differences from the region-specific thresholds.¹⁷

In what follows we will work with *two* groups of participant households (i.e. those labeled *eligibles* and those labeled *densificados*) and *one* group of non-participant households (obtained by pooling those labeled *ineligibles* and *forgotten*). We will focus on educational outcomes of children in the aforementioned groups using individual level data from the baseline survey collected in October/November 1997, as well as from two follow-up surveys collected in October/November 1998 and October/November 1999. In particular, we will look at the effect of PROGRESA on the proportion of children enrolled at school separately for four groups defined by gender and age. In sum, there are almost 11,000 children aged 6-16 belonging to the *eligibles* group, 1,127 belonging to the *densificados* group and 3,600 belonging to the group of *non-participants*. The full report on the estimation of the impact on educational outcomes can be found in Behrman *et al.* (2001).

By defining these three groups of children we are replicating the design

¹⁵The breakdown by November 1999 is informationally equivalent and therefore is not reported.

¹⁶A further set of immigrant households was added to the original sampling frame starting from the first follow-up survey (see Angelucci, 2004), but they are not considered in the analysis presented below.

¹⁷We used a robust estimator of the standard deviation to control for the effect of outliers in the score.

described in Section 3, where the labels for participation and eligibility have been symmetrically reversed: all subjects below the threshold comply with the original sharp RDD assignment while some of the subjects above the threshold violate the sharp RDD assignment according to an unknown rule. It is therefore straightforward to show that the average effect of the treatment on non-participants (ETNP) around the threshold $\bar{s} = 0$ is identified

$$E\{Y_1 - Y_0 | I = 0, \bar{s}^+\} = \frac{E\{Y | \bar{s}^-\} - E\{Y | \bar{s}^+\}}{Pr\{I = 0 | \bar{s}^+\}},$$

provided that the continuity condition on the mean of Y_1 conditional on S holds at \bar{s} . Accordingly, selection bias is defined with respect to the outcome Y_1 .

The plan of our analysis is to replicate (for the ETNP) results presented in Figure 2a, Figure 2b and Figure 2c in Behrman *et al.* (2001) using an identification strategy based on the regression discontinuity. First, we estimate the ETNP around $\bar{s} = 0$, separately for boys and girls and for the age groups 6 – 11 and 12 – 16 (to distinguish between primary and secondary school levels). Second, we test for the presence of selection bias at $\bar{s} = 0$ by comparing school enrollment amongst marginal *eligibles* to school enrollment amongst marginal *densificados*. Third, in those cases where the selection bias is not zero at $\bar{s} = 0$, we look for a set of pre-programme characteristics x such that the bias is zero once they are controlled for in the analysis. The effects on school enrollment are evaluated one year (1998) and two years (1999) after the programme started.

Note that in this context the test for the presence of selection bias is a tool to check whether selection of *densificados* from the pool of households who were not eligible before the *densificaciòn* is locally ignorable with respect to Y_1 . Otherwise stated, we test whether the pool of *densificados* on the one hand and the pool of *ineligibles* and *forgotten densificados* on the other hand share, on

average, the same outcome Y_1 locally at $\bar{s} = 0$. On failing to reject the null, that is on finding that there is no systematic difference between these two groups, one would be more confident to compare *densificados* to non-participants (either in 'pilot' or in 'control' areas) to identify the mean impact on *densificados*. By averaging the mean impacts on *eligibles* and on *densificados*, respectively, using the relative size of the two groups as weights one gets the mean impact on participants.

Discontinuities in the outcome of interest are estimated by contrasting the mean outcome for individuals whose distance from $\bar{s} = 0$ is within a fixed bandwidth, and a sensitivity analysis with respect to different values of this bandwidth is presented. To test for the presence of selection bias at $\bar{s} = 0$, whether conditional or unconditional on x , we use information for *eligibles* and for *densificados* only. We control for x using three different methods: a linear probability model, where school enrollment is regressed on a dummy variable for whether or not individuals have $s \leq 0$ and the additional controls x ; propensity score matching, where individuals with $s \leq 0$ are matched to individuals with $s > 0$ presenting similar values of the propensity score; and propensity score weighting, where mean outcomes for participants are compared to weighed outcomes for non-participants to balance the composition of the two groups with respect to the propensity score (see Imbens, 2004). As the three sets of results were informationally equivalent, we decided to report only those from matching.

**** TABLE 2 AROUND HERE ****

Estimation results for the parameters of interest are reported in Table 2, together with the sample size associated to the values of the bandwidth h considered (h is expressed in unit of standard deviation). The significance level has

been derived by considering bootstrap confidence intervals, obtained by clustering children outcomes at the household level. In line with the results discussed by Behrman *et al.* (2001), the programme effect on non-participants is larger in the second year, for the age group 12 – 16 (for which the enrollment rate is much smaller than for the younger group) and for girls.

As for the bias at \bar{s} our results suggest that it is of non-negligible size and statistically significant for girls in the age group 12 – 16. In correcting for this bias, we consider a number of individual specific and village level variables known from previous studies to be good predictors of participation at school, such as family characteristics and family composition as well as a rich list of village level characteristics. After matching children around the discontinuity point on this set of variables separately for the four groups considered, the bias turns out to be statistically zero in all cases, though not negligible in some of them. In particular, for $h = 0.1$ the sample size might be too small resulting in large sample variability for the estimates.

6 Conclusions and caveats

It is well known from the literature on the Regression-Discontinuity Design (RDD; see HTV) that when participation into a programme deterministically depends on whether an observable pre-programme characteristic lies on either side of a specified threshold (*sharp* design), the identification of mean effects for marginal participants is attained under fairly weak conditions.

On the other hand, when the probability of participation change discontinuously at a specified point of the support of the pre-programme characteristic but the size of the discontinuity is less than one not a *fuzzy* design arises, which

requires stronger regularity conditions and allows identification of mean effects for a sub-set of marginal participants.

In this paper we have shown that when an intervention is targeted to a population of eligible individuals but is actually administered only to self-selected eligibles, it is worth collecting information separately on *three* groups of subjects: ineligibles, eligible non-participants and participants. A “partially fuzzy” design is then defined, since individuals ineligible for the programme are denied participation but the participation status varies amongst eligible individuals (typically, if participation is on a voluntary basis). We have derived the regularity conditions required to identify the mean impact on participants marginally eligible for the programme by *jointly* exploiting both groups. Despite the *prima facie* fuzzy nature of the design we deal with, we have shown that identification requires the same regularity conditions as in a sharp design.

Second, we have shown that the selection bias for subjects at the margin between eligibility and ineligibility is identifiable. Local identification provided by the RDD can then be used to test the validity of alternative identifying restrictions on which non-experimental estimators for participants rely. By design, such a test is informative only at the threshold for eligibility, thus results cannot be generalized to the whole population (unless one is willing to impose additional restrictions). The value of the specification test is that if it rejects the non-experimental estimator locally then this is enough to reject it altogether.

The results presented rest on the *Stable Unit Treatment Value Assumption* (SUTVA; Rubin, 1977). According to this assumption, the outcome experienced by each individual is not affected by assignment and receipt of treatment by other individuals. In particular SUTVA rules out substitution effects (see the

Table 1: Households living in pilot localities by eligibility status in November 1998

		poverty index		
		mean	min	max
Participants				
<i>eligibles</i>	4,000	-0.6891	-2.8811	0.0000
<i>densificados</i>	496	0.3068	0.0008	2.1676
Non-participants				
<i>ineligibles</i>	1,046	0.8564	0.0030	3.3530
<i>forgotten</i>	731	0.4560	0.0029	2.3256

discussion in Battistin and Rettore, 2003).

A further threat to the validity of our results occurs when subjects ineligible for the programme on the basis of a pre-programme observable characteristic purposively modify it to become eligibles. For example, mean-tested policies targeted to improve family income might induce some subjects to deliberately reduce their labor income to become eligible. If this is the case, the design loses its fundamental feature, namely to provide a nearly randomised assignment rule in a neighborhood of the threshold for eligibility. Simple tests to detect entry effects are presented in Battistin and Rettore (2003) and in Lee (2004).

Acknowledgements

This paper, previously circulated as “Another Look at the Regression Discontinuity Design” (first version January 2002), benefited from helpful comments by two anonymous referees and the editors, from discussion with Alberto Abadie, Manuela Angelucci, Orazio Attanasio, Richard Blundell, David Card, Vincenzo Di Maro, Hide Ichimura, Andrea Ichino and Barbara Sianesi and from comments by audiences at ESEM 2002, CEPR/IZA Conference “Improving Labor Market Performance: The Need for Evaluation” (Bonn, October 2002), Statistics Canada Symposium 2002, LABORatorio Conference “New perspectives in

public policy evaluation” (Turin, November 2002), Brucchi Luchino workshop (Padova, December 2002), Dalarna University, Cemmap, “The Regression Discontinuity Method in Economics: Theory and Applications” (Banff, May 2003) and Ente Einaudi. Financial support from MIUR to the project “Dynamics and inertia in the Italian labour market and policies evaluation (data-bases, measurement issues, substantive analyses)” is gratefully acknowledged.

References

- [1] Angelucci, M., 2004, Border enforcement, aid and migration, Ph.D. Thesis in Economics, University College London.
- [2] Angrist, J.D., 1998, Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants, *Econometrica*, 66, 2, 249-288.
- [3] Angrist, J.D. and G.W. Imbens, 1991, Sources of Identifying Information in Evaluation Models, NBER Technical Working Paper 117.
- [4] Angrist, J.D, G.W. Imbens and D.B. Rubin, 1996, Identification of Causal Effects Using Instrumental Variables (with discussion), *Journal of the American Statistical Association*, 91, 434, 444-472.
- [5] Battistin, E. and E. Rettore, 2002, Testing for programme effects in a regression discontinuity design with imperfect compliance, *Journal of the Royal Statistical Society A*, 165, 1, 1-19.
- [6] Battistin, E. and E. Rettore, 2003, Another look at the Regression Discontinuity Design, Working Paper 01/03, Cemmap, London.

- [7] Behrman, J., P. Sengupta and P.E. Todd, 2001, Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment, University of Pennsylvania and the International Food Policy Research Institute, Washington, D.C.
- [8] Behrman, J. and P.E. Todd, 1999, Randomness in the Experimental Samples of PROGRESA (Education, Health, and Nutrition programme), International Food Policy Research Institute, Washington, D.C.
- [9] Bloom, H.S., 1984, Accounting for No-Shows in Experimental Evaluation Designs, *Evaluation Review*, 8, 225-246.
- [10] Blundell, R. and M. Costa Dias, 2000, Evaluation methods for non-experimental data, *Fiscal Studies*, 21, 4, 427-468.
- [11] Buddelmeyer, H. and E. Skoufias, 2003, An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA, IZA Discussion Paper 827.
- [12] Cook, T.D. and D.T. Campbell, 1979, *Quasi-Experimentation. Design and Analysis Issues for Field Settings* (Houghton Mifflin Company, Boston).
- [13] Fisher, R.A., 1935, *The Design of Experiments* (Oliver&Boyd, Edinburgh).
- [14] Hahn, J., P.E. Todd and W. van der Klaauw, 2001, Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design, *Econometrica*, 69, 1, 201-209.
- [15] Heckman, J.J., 1990, Varieties of Selection Bias, *The American Economic Review*, 80, 2, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association, 313-318.

- [16] Heckman, J.J., 1992, Randomization and Social Policy Evaluation, in: C. Manski and I. Garfinkel, eds., *Evaluating Welfare and Training Programs*, (Harvard University Press, Cambridge) 201-230.
- [17] Heckman, J.J., 1997, Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making programme Evaluations, *Journal of Human Resources*, XXXII, 441-462.
- [18] Heckman, J.J. and V.J. Hotz, 1989, Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social programmes: The Case of Manpower Training, *Journal of the American Statistical Association*, 84, 862-874.
- [19] Heckman, J.J. and J. Smith, 1995, Assessing the Case for Social Experiments, *The Journal of Economic Perspectives*, 9, 2, 85-110.
- [20] Heckman, J.J., H. Ichimura, J. Smith and P.E. Todd, 1998, Characterizing Selection Bias Using Experimental Data, *Econometrica*, 66, 1017-1098.
- [21] Heckman, J.J. R. Lalonde and J. Smith, 1999, The Economics and Econometrics of Active Labor Market programmes, in: A. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3 (Elsevier Science, Amsterdam).
- [22] Hoddinott, J., E. Skoufias and R. Washburn, 2000, *The Impact of PROGRESA on Consumption: A Final Report*, International Food Policy Research Institute, Washington, D.C.
- [23] Imbens, G.W., 2004, Semiparametric Estimation of Average Treatment Effects under Exogeneity: a Review, *The Review of Economics and Statistics*, 86, 1, 4-29.

- [24] Imbens, G.W. and J.D. Angrist, 1994, Identification and Estimation of Local Average Treatment Effects, *Econometrica*, 62, 467-476.
- [25] van der Klaauw, W., 2002, Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach, *International Economic Review*, 43, 4, 1249-1287.
- [26] Lee, D., 2004, Randomized Experiments from Non-random Selection in U.S. House Elections, revision of the NBER Working Paper 8441.
- [27] Little, R.J.A. and L. Yau, 1998, Statistical Techniques for Analyzing Data from Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model, *Psychological Methods*, 3, 2, 147-159.
- [28] LaLonde, R., 1986, Evaluating the econometric evaluations of training programmes with experimental data, *American Economic Review*, 76, 604-20.
- [29] Neyman, J. (with co-operation by Iwazskiewicz, K. and S. Kolodziejczyk), 1935, Statistical Problems in Agricultural Experimentation, Supplement of the *Journal of the Royal Statistical Society*, 2, 107-180.
- [30] Porter, J., 2002, Asymptotic bias and optimal convergence rates for semi-parametric kernel estimators in the regression discontinuity model, Discussion Paper 1989, Harvard Institute of Economic Research.
- [31] Quandt, R., 1972, Methods for Estimating Switching Regressions, *Journal of the American Statistical Association*, 67, 306-310.
- [32] 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, 385, 41-48.

- [33] Rosenbaum, P.R., 1987, The Role of a Second Control Group in an Observational Study, *Statistical Science*, 2, 3, 292-306.
- [34] Roy, A., 1951, Some Thoughts on the Distribution of Earnings, *Oxford Economic Papers*, 3, 135-146.
- [35] Rubin, D.B., 1974, Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, 66, 688-701.
- [36] Rubin, D.B., 1977, Assignment to Treatment Group on the Basis of a Covariate, *Journal of Educational Statistics*, 2, 4-58.
- [37] Skoufias, E., 2001, PROGRESA and its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI, International Food Policy Research Institute, Washington, D.C.
- [38] Thistlethwaite, D.L. and D.T. Campbell, 1960, Regression discontinuity analysis: an alternative to the ex post facto experiment, *Journal of Educational Psychology*, 51, 6, 309-317.
- [39] Trochim, W., 1984, *Research Design for Program Evaluation: the Regression-Discontinuity Approach* (Sage Publications, Beverly Hills).

Table 2: Estimation and specification testing results

sample size 1998		$h = 0.10$			$h = 0.15$			$h = 0.20$		
		boys	girls	all	boys	girls	all	boys	girls	all
6-11	eligibles	158	176	334	273	281	554	380	399	779
	densificados	80	67	147	119	108	227	141	127	268
	non-participants	80	71	151	146	123	269	183	173	356
12-16	eligibles	141	125	266	237	216	453	328	296	624
	densificados	66	63	129	100	97	197	118	115	233
	non-participants	92	82	174	145	140	285	199	188	387
sample size 1999		$h = 0.10$			$h = 0.15$			$h = 0.20$		
		boys	girls	all	boys	girls	all	boys	girls	all
6-11	eligibles	138	158	296	238	264	502	344	370	714
	densificados	74	61	135	106	100	206	127	117	244
	non-participants	71	62	133	133	106	239	163	153	316
12-16	eligibles	105	101	206	167	160	327	227	219	446
	densificados	42	49	91	64	72	136	73	86	159
	non-participants	67	56	123	101	94	195	133	126	259
average effect for non-participants at $\bar{s} = 0$		$h = 0.10$			$h = 0.15$			$h = 0.20$		
		boys	girls	all	boys	girls	all	boys	girls	all
6-11	1998	3.67	0.61	2.19	2.06	-0.23	0.93	2.25**	0.14	1.17
	12-16	0.19	7.24	3.60	-1.71	7.53	2.83	0.13	11.73*	5.80
6-11	1999	5.42*	7.43*	6.06**	2.25	4.62	3.21*	1.50	4.69*	3.03**
	12-16	-6.99	9.48	0.90	-2.38	15.60*	6.55	-0.19	18.45**	9.16*
bias at $\bar{s} = 0$		$h = 0.10$			$h = 0.15$			$h = 0.20$		
		boys	girls	all	boys	girls	all	boys	girls	all
6-11	1998	1.87	-1.14	0.46	1.31	-0.50	0.42	1.15	-0.47	0.35
	12-16	3.45	10.03	6.80	0.19	11.82**	5.93	-1.60	12.54**	5.39
6-11	1999	0.00	3.02	1.21	0.10	1.11	0.55	-0.08	1.21	0.52
	12-16	-6.67	15.44*	5.35	-3.48	16.74**	7.11	-3.40	19.15**	8.68*
bias at $\bar{s} = 0$ after controlling for x		$h = 0.10$			$h = 0.15$			$h = 0.20$		
		boys	girls	all	boys	girls	all	boys	girls	all
6-11	1998	1.43	-1.05	-0.14	0.39	-0.50	-0.24	0.06	0.04	0.26
	12-16	17.88	25.39	8.55	0.04	8.01	7.97	-1.61	0.86	1.16
6-11	1999	0.00	7.46	3.02	0.24	0.54	5.82	-0.13	2.00	1.24
	12-16	-12.30	13.21	7.26	-5.09	8.41	9.89	0.56	-4.95	1.49

Significance based on bootstrap confidence intervals, obtained by 500 replications and by clustering at the household level. **: 95% *: 90%