Regression–Discontinuity Analysis: A Survey of Recent Developments in Economics

Wilbert van der Klaauw

Abstract. This paper provides a discussion of recent developments related to the applicability of the regression discontinuity design in economics. It reviews econometric issues, such as identification and estimation methods, as well as a number of sensitivity and validity tests of importance in empirical application.

1. Introduction

The regression–discontinuity (RD) data design is an evaluation design that was first introduced by Thistlethwaite and Campbell (1960) as an alternative to the randomized experiment for evaluating social programs and interventions. From its inception the design has been termed *quasi-experimental*, reflecting its intuitive connection to purely randomized experimental designs. Unlike experimental data, where randomized assignment guarantees comparability between persons in the treated group and in the control group, assignment in an RD design like that in observational data is not random and persons who receive treatment will differ systematically from those who do not. However, data from an RD design differ from observational data in that we have specific knowledge about the assignment rule that influences how persons are assigned to or selected into treatment. More specifically, the design requires that there is a known *cut-off* point in treatment assignment or in the probability of treatment receipt as a function of one or more

---

Wilbert van der Klaauw, Microeconomics and Regional Studies Function, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045, USA. E-mail: Wilbert.VanderKlaauw@ny.frb.org.

The views and opinions offered in this paper do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System as a whole.

LABOUR 22 (2) 219–245 (2008) JEL C14, C21

continuous assignment variables, generating a discontinuity in the treatment recipiency rate at that point.

Thistlethwaite and Campbell's original article concerned the evaluation of the impact of receiving a National Merit Award on students' success in obtaining additional college scholarships and on their career aspirations. In their study an award was given to all students who obtained a minimum score on a scholarship examination. Thistlethwaite and Campbell's insight was that one could take advantage of knowing the cut-off score to learn about the impact of award receipt for persons near the cut-off. Assuming that students who scored just below the minimum score represent a valid no-treatment comparison group for those who scored just above the minimum, one could evaluate its impact by relating average outcomes for award recipients just above the cut-off with those of non-recipients just below it. That is, under certain comparability conditions, the assignment near the cut-off can be seen as behaving almost as if random.

Although the design has been of continued, though limited, interest to evaluation research methodologists (Cook and Campbell, 1979; Trochim, 1984), it has recently experienced considerable gains in popularity among econometricians and empirical economists (Angrist and Krueger, 1999; Hahn *et al.*, 2001; Porter, 2003; Imbens and Lemieux, 2008a, 2008b). This movement has generated several key econometric advances, including the formal derivation of identification conditions for causal inference and the introduction of new semi-parametric estimation procedures for the design. The concurrent development of a large number of empirical applications is extending insights into the design's use and has led to the emergence of several sensitivity and validity tests.

The emerging popularity of the RD design in applied economic research is related to several factors. First, the assignment rules in many existing programs and procedures for allocating social resources, even though usually not specifically developed for evaluation purposes, nonetheless frequently lend themselves to RD evaluations because of their structure. Often, program resources are allocated based on a formula with a cut-off structure that determines the allocation of resources to recipients. The use of cut-offs is conceptually compatible with the political and social goal of allocating scarce resources to those persons who need or deserve them most. Eligibility cut-offs applied to a continuous quantified measure of need, such as family income, are often used to screen potential recipients of a program or in allocating funds. A second positive feature of the design is that it is intuitive and its results can be easily conveyed, often with a visual depiction of sharp changes in both treatment assignment and average outcomes around the cut-off value of the assignment variable (McEwan and Urquiola, 2005). Third, it is possible for a researcher to choose from among several different estimation methods to estimate effects that have credible causal interpretations (Hahn *et al*., 2001).

One area of economic research where the design has been particularly useful in recent years has been the evaluation of educational interventions. In fact, the first two empirical applications in the economics literature fall within this area. van der Klaauw (1997, 2002) exploited discontinuities in the amount of financial aid offered as a function of a student's ability score to evaluate its effect on student enrollment decisions. Angrist and Lavy (1999) used a rule that generated discontinuities in the relationship between average class size and beginning-of-the-year grade enrollment, to estimate the effect of class size on student test scores. Education programs are frequently prescribed or offered to schools or students who score below a cut-off on some scale (student performance, poverty), and school and program funding decisions are often based on allocation formulas containing discontinuities. Among others, the approach has recently been used to evaluate the effectiveness of a remedial education program (Jacob and Lefgren, 2004a), a mandatory summer school program (Matsudaira, 2008), state financial aid programs (Kane, 2003), federal Title I funding for compensatory education (van der Klaauw, 2008), funding for the Head Start program in the USA, which provides pre-school, health, and other services to low-income children aged 3–5 (Ludwig and Miller, 2007), classroom peer effects (Angrist and Lang, 2004), teacher training programs (Jacob and Lefgren, 2004b), performance-related incentive pay to teachers (Lavy, 2004), school voucher programs (Chakrabarti, 2008a, 2008b), the impact of school characteristics on housing prices (Black, 1999; Kane *et al*., 2006), and the design is being considered for evaluating Reading First, a federal program in the USA that provides funding for reading programs in the early grades (Bloom *et al*., 2005).

The design has also proven useful in evaluating the socio-economic impacts of a diverse set of government programs and laws. The RD approach was adopted to evaluate the impacts of a US anti-discrimination law (Hahn *et al*., 1999), an anti-poverty program in Mexico (Buddelmeyer and Skoufias, 2004), expansions in government-provided health insurance to low-income households (Card and Shore-Sheppard, 2004), a federal disability insurance program (Chen and van der Klaauw, 2008), legislation aimed at reducing air pollution (Chay and Greenstone, 2005), state foreclosure laws (Pence, 2006), a social assistance program in Quebec (Lemieux and Milligan, 2008), compulsory military service in Germany (Bauer *et al.*, 2004), the parental notification act in Texas requiring parents to be notified of a minor's request for an abortion (Joyce *et al.*, 2006), a federal inmate classification system determining prison conditions (Chen and Shapiro, 2007), the incumbency advantage in elections (Lee, 2008), and the impact of unionization on establishment closure (DiNardo and Lee, 2004). In all these applications, the treatment variable or the probability of receiving treatment changes discontinuously as a function of one or more underlying variables, which is the defining characteristic of RD data designs.

In the next two sections I will review the conditions under which the existence of cut-offs in the treatment selection mechanism constitutes a valuable source of identifying information. The exposition will be closely related to that in Hahn, Todd and van der Klaauw (2001, HTV in what follows) and van der Klaauw (2002) and will follow Trochim (1984) in distinguishing between two different forms of the design, depending on whether the treatment assignment is related to the assignment variable by a deterministic function (sharp design) or a stochastic one (fuzzy design). Sections 4 and 5 consider parametric and semi-parametric estimation methods for estimating treatment effects. Sensitivity and validity tests are discussed in Section 6, and Section 7 concludes with a discussion of several extensions to the basic RD model.

2. The sharp RD design

Consider the general problem of evaluating the causal effect of a binary treatment on an outcome variable, using a random sample of individuals where for each individual i we observe an outcome measure y_i (e.g. a student's performance on an academic test at the end of given year) and a binary treatment indicator t_i, equal to one if treatment was received and zero otherwise (e.g. receipt at the beginning of the year of a fixed scholarship amount). The evaluation problem that arises in determining the effect of t on y is due to the fact that each individual either receives or does not receive treatment and is never simultaneously observed in both states. Let

$y_i(1)$ be the outcome given treatment, and $y_i(0)$ the outcome in the absence of treatment. Then the actual outcome we observe equals $y_i = t_i y_i(1) + (1 - t_i) y_i(0)$. A common regression model representation for the observed outcome can then be written as

$$y_i = \alpha + \beta_i t_i + u_i, \quad [1]$$

where $\alpha = E[y_i(0)]$, $\beta_i = y_i(1) - y_i(0)$, and $y_i(0) = \alpha + u_i$.

Comparing average outcomes of participants (treatment recipients) with non-participants (non-recipients) would generally not provide us with an estimate of the average treatment effect $E[\beta_i]$ because

$$\begin{aligned} E[y_i(1)|t_i = 1] - E[y_i(0)|t_i = 0] = E[\beta_i] &+ (E[u_i|t_i = 1] - E[u_i|t_i = 0]) \\ &+ \Pr(t_i = 0)(E[\beta_i|t_i = 1] - E[\beta_i|t_i = 0]), \end{aligned} \quad [2]$$

where we used the fact that $E[\beta_i] = E[\beta_i|t_i = 1] - \Pr(t_i = 0) \times (E[\beta_i|t_i = 1] - E[\beta_i|t_i = 0])$.

If average outcomes for recipients and non-recipients differed even in the absence of treatment, or if average outcome gains resulting from treatment were different for both groups of individuals, one or both of the last two parenthesized terms in the equation will not equal zero. Although randomized assignment of treatment would guarantee both to be zero, in most observational studies there would usually be reasons to doubt this to be the case. In an ordinary least squares regression of the outcome variable on the treatment indicator, this would imply that the estimated coefficient for the treatment indicator will generally not have a causal interpretation, whereas in the case of randomized assignment it would estimate $E[\beta_i]$, the average treatment effect in the population.

In a sharp RD design, individuals are assigned to or selected for treatment solely on the basis of a cut-off score on an observed continuous variable x. This variable, alternatively called the *assignment*, *selection*, *running*, or *ratings* variable, may represent a single characteristic or a composite variable constructed using multiple characteristics. Those who fall below some distinct cut-off point $\bar{x}$ are placed in the control group ($t_i = 0$), whereas those on or above that point are placed in the treatment group ($t_i = 1$) (or vice versa). Thus, assignment occurs through a known and measured deterministic decision rule: $t_i = t(x_i) = 1\{x_i \geq \bar{x}\}$ where $1\{\cdot\}$ is the indicator function.

As the assignment variable itself may be correlated with the outcome variable, the assignment mechanism may be non-random. Consequently when comparing average outcomes for persons who received and did not receive treatment, the effect of t on y will be confounded with that of x, implying that the two bias terms in [2] generally will not equal zero. Moreover, although the sharp design represents a special case of selection on observables (Heckman and Robb, 1985), common solutions to such selection problems, such as propensity score-matching methods, are not applicable here as the design violates the *strong ignorability condition* of Rosenbaum and Rubin (1983), which, in addition to requiring u to be independent of t conditional on x, requires $0 < \Pr(t = 1|x) < 1$ for all x whereas here $\Pr(t = 1|x) \in \{0, 1\}$. In the terminology of Heckman *et al*. (1997), there exists no region of common support required for matching.

However, if it is reasonable to assume that persons close to the threshold with very similar x values are comparable, then we may view the design as almost experimental near $\bar{x}$, suggesting that we could evaluate the causal impact of treatment by comparing the average outcomes for individuals with ratings just above and below the cut-off. More formally, consider the following *local continuity assumption*:

$E[u_i|x]$ and $E[\beta_i|x]$ are continuous in x at $\bar{x}$, or equivalently, $E[y(1)|x]$ and $E[y(0)|x]$ are continuous at $\bar{x}$,

then assuming that the density of x is positive in a neighborhood containing $\bar{x}$,

$$\begin{aligned}\lim_{x\downarrow\bar{x}} E[y_i|x] - \lim_{x\uparrow\bar{x}} E[y_i|x] &= \lim_{x\downarrow\bar{x}} E[\beta_i t_i|x] - \lim_{x\uparrow\bar{x}} E[\beta_i t_i|x] \\ &\quad + \lim_{x\downarrow\bar{x}} E[u_i|x] - \lim_{x\uparrow\bar{x}} E[u_i|x] \\ &= E[\beta_i|\bar{x}],\end{aligned}$$

where we used $E[y_i|x]$ as shorthand notation for $E[y_i|x_i = x]$, with $E[\beta_i|\bar{x}]$ similarly representing $E[\beta_i|x_i = \bar{x}]$. The RD approach of comparing average outcomes just right and left of the cut-off therefore identifies the average treatment effect for individuals close to the discontinuity point. It is important to recognize that the continuity assumption formalizes the condition discussed earlier that individuals just above and below the cut-off are 'comparable', requiring them to have similar average potential outcomes when receiving treatment and when not. Thus identification is achieved assuming only smoothness in expected potential outcomes at the discontinuity without any parametric functional form restrictions. The smoothness assumption represents an unusual functional form restriction, which is required here to take advantage of the known discontinuity in the treatment rule.

It is a limitation of the RD design that without further assumptions, such as a 'common effect' assumption, we might learn about treatment effects only for a subpopulation of persons with values of x near the discontinuity point. In case of heterogeneous impacts, the local effect may be very different from the effect at values further away from the threshold. On the other hand, as pointed out by HTV this local effect is highly relevant to policy-makers who are contemplating either expanding or limiting eligibility or participation via a small change in the cut-off.

The continuity assumption required for identification fulfils an important function. Even if treatment receipt is determined solely on the basis of a cut-off score on the assignment variable, this is not a sufficient condition for the identification of a meaningful causal effect. The continuity assumption rules out coincidental functional discontinuities in the x–y relationship such as those caused by other programs that use assignment mechanisms based on the exact same assignment variable and cut-off. In addition, as will be discussed in greater detail in Section 6, the continuity restriction tends to rule out certain types of behavior both on the part of potential treatment recipients who exercise control over their value of x, and on the part of program administrators in choosing the assignment variable and cut-off point.

3. The fuzzy RD design

In the second type of the RD design, referred to in the literature as the *fuzzy RD design* (Campbell, 1969), instead of a deterministic assignment rule, treatment assignment depends on x in a stochastic manner, but one in which the propensity score function $\Pr(t = 1|x)$ is again known to have a discontinuity at $\bar{x}$. Instead of a 0–1 step function, the treatment probability as a function of x could now contain a jump at the cut-off that is less than 1.

The fuzzy design is akin to a case of mis-assignment relative to the cut-off value in a sharp design, with values of x near the cut-off appearing in both treatment and control groups. This situation is analogous to having no-shows (treatment group members who do not receive treatment) and/or crossovers (control group member who do receive the treatment) in a randomized experiment. The mis-assignment may occur if, in addition to the position of the individual's score relative to the cut-off value, assignment is based on additional variables observed by the administrator, but unobserved by the evaluator.

For example, a decision to offer a scholarship may be based in part on whether a continuous measure of academic ability exceeds a given cut-off value, but also on recommendation letters that may be unobserved by the evaluator. In this case, in evaluating its impact on subsequent academic achievement, scholarship recipients with a given ability score could be expected to have higher average outcomes than the non-recipients with the same ability score, even in the absence of any genuine award effect. As a consequence scholarship receipt provides information beyond that given by the student's ability score. That is, there is some characteristic of the treatment or control group (teachers' evaluations of a student's potential to succeed as expressed in recommendation letters) that is both associated with receipt of the treatment and associated with the outcome of interest so as to lead to a false attribution of causality regarding treatment and outcome. Thus a comparison of average outcomes of recipients and non-recipients, even if near the cut-off, would not generally lead to correct inferences regarding an average treatment effect.

On the other hand, a comparison of average outcomes for *all* cases, irrespective of recipiency status, with scores just right and left of the cut-off point could still be informative. To see how it is still possible to exploit the discontinuity in the selection rule to identify an average treatment effect of interest in the fuzzy RD case, notice that

$$\lim_{x\downarrow\bar{x}} E[y_i|x] - \lim_{x\uparrow\bar{x}} E[y_i|x] = \left(\lim_{x\downarrow\bar{x}} E[\beta_i t_i|x] - \lim_{x\uparrow\bar{x}} E[\beta_i t_i|x]\right) + \left(\lim_{x\downarrow\bar{x}} E[u_i|x] - \lim_{x\uparrow\bar{x}} E[u_i|x]\right). \quad [3]$$

Considering first the case where the treatment effect is locally constant ($\beta_i = \beta$ in a neighborhood around $\bar{x}$), then under the same local continuity assumption as in the sharp design, the first term on the right-hand side in this equation equals $\beta[\lim_{x\downarrow\bar{x}} E[t_i|x] - \lim_{x\uparrow\bar{x}} E[t_i|x]]$ and the common treatment effect β is thus identified by

$$\frac{\lim_{x \downarrow \bar{x}} E[y|x] - \lim_{x \uparrow \bar{x}} E[y|x]}{\lim_{x \downarrow \bar{x}} E[t|x] - \lim_{x \uparrow \bar{x}} E[t|x]}, \qquad [4]$$

where the denominator is always non-zero because of the known discontinuity of $E[t|x]$ at $\bar{x}$.

In case of heterogenous treatment effects, as shown by HTV, under the local continuity assumption, and a local conditional independence assumption requiring t_i to be independent of β_i conditional on x near $\bar{x}$, the first parenthesized term on the right-hand side of equation [3] equals $\lim_{x \downarrow \bar{x}} E[\beta_i|x] \lim_{x \downarrow \bar{x}} E[t_i|x] - \lim_{x \uparrow \bar{x}} E[\beta_i|x] \lim_{x \uparrow \bar{x}} E[t_i|x]$, implying that the ratio in [4] identifies $E[\beta_i|x = \bar{x}]$, the average treatment effect for cases with values of x close to $\bar{x}$.

If individuals self-select into treatment, or are selected for treatment on the basis of expected gains from treatment, then the conditional independence assumption may be violated. HTV also consider the case where $t_i(x)$, individual ith treatment assignment given any x, is a deterministic function that varies across individuals. They show that in this case, under a weaker local monotonicity assumption similar to that made in Imbens and Angrist (1994), the ratio defined in [4] will instead identify a local average treatment effect (LATE) at the cut-off point, defined as: $\lim_{e \downarrow 0} E[\beta_i | t_i(\bar{x} + e) - t_i(\bar{x} - e) = 1, x = \bar{x}]$. This causal effect represents the average treatment effect of the 'compliers', i.e. the subgroup of individuals whose treatment status would switch from non-recipient to recipient if their score x crossed the cut-off. The share of this group in the population in the neighborhood of the cut-off is equal to the denominator of [4].

As illustration, consider again the scholarship example. Assume that awards are based not only on a student's academic ability score relative to a threshold but also on the student's minority status, in such a way that although all minority students receive the scholarship, only those non-minority students with ability scores above the cut-off do. In this case, if applicants' minority status was unobserved by the evaluator, the scholarship assignment rule would correspond to that of a fuzzy design. The LATE would then apply to the subgroup of students with academic ability scores close to the cut-off for whom scholarship receipt depends on the position of their ability score relative to the cut-off, which in this case would be the subsample of non-minority students.

Although it is often the case that the identity of the so-called compliers associated with an estimated LATE is unknown, in many

RD applications the treatment assignment or selection rule is sufficiently well known that the group can be precisely characterized. In fact, in many situations the actual rule is deterministic and based on predetermined variables and therefore representing a sharp RD design, but needs to be treated as a fuzzy design because one or more of the determinants are unobserved. The subgroup to which the LATE applies can then usually be characterized in terms of particular values or regions for the omitted variables [see the applications by van der Klaauw (2008) and Chen and van der Klaauw (2008)].

Another case in which the ratio in [4] identifies an average treatment effect for a well-defined subgroup of the population is one where an eligibility rule divides the population into eligibles and non-eligibles according to a sharp RD design, and where eligible individuals self-select into treatment. As pointed out by Battistin and Rettore (2008), under the local continuity assumption the first parenthesized term on the right-hand side of [3] equals $\lim_{x\downarrow\bar{x}} E[\beta_i|t_i = 1, x] \cdot \lim_{x\downarrow\bar{x}} E[t_i|x]$, whereas the second term is zero implying that the local continuity assumption alone is sufficient for the ratio in [4] to identify $E[\beta_i|t_i = 1, x = \bar{x}]$, the average treatment effect on the treated, for those near the cut-off. This result requires no restrictions on the self-selection behavior of eligibles, and is similar to results by Bloom (1984) and Angrist and Imbens (1991).

4. Parametric estimation

The identification results in the previous section indicate that estimation of treatment effects in the case of an RD design requires estimating boundary points of conditional expectation functions. With a sufficiently large number of observations one could in principle focus on units within a very small interval around the cut-off point and simply compare average outcomes for units just left and right of the discontinuity point. Increasing the interval around the cut-off point is likely to produce a bias in the effect estimate, especially if the assignment variable was itself related to the outcome variable conditional on treatment status. However, if one is willing to make additional assumptions about this relationship, one could use more observations and extrapolate from above and below the cut-off point to what a tie-breaking randomized experiment would have shown. This double extrapolation combined with the exploitation of the 'randomized experiment' around the cut-off point represents the two main features by which RD analysis has been characterized in the evaluation literature (e.g. Cook and Campbell, 1979).

In line with this idea of imposing additional assumptions, the most common empirical strategy in the literature has been to adopt parametric specifications for the conditional expectations functions. To understand the implementation of this approach, it is useful to consider the following alternative specification of the outcome regression equation [1] in case of a sharp RD design:

$$y_i = m(x_i) + \delta t_i + e_i, \qquad [5]$$

where $e_i = y_i - E[y_i|t_i, x_i]$, $t_i = 1\{x_i \geq \bar{x}\}$, and $m(x) = \alpha + E[u_i|x] + (E[\beta_i|x] - E[\beta_i|\bar{x}])1\{x \geq \bar{x}\}$. Note that under the local continuity assumption $m(x)$ will be a continuous function of x at $\bar{x}$, with δ, the discontinuity in the average outcome at the cut-off, representing $E[\beta_i|\bar{x}]$, the average treatment effect at $\bar{x}$. This specification therefore suggests that if the correct specification of $m(x)$ were known, and was included in the regression, we could consistently estimate the treatment effect for the sharp RD design.

The concept of including a specification of $m(x)$ in the regression of y on t in order to correct for selection bias due to selection on observables is known in the econometrics literature as the *control function approach* (Heckman and Robb, 1985). Empirical researchers have tended to favor global polynomials or splines (piecewise polynomials) where, even though continuous at the cut-off, $m(x)$ is specified as a different polynomial function of x on either side of the cut-off (Trochim, 1984; van der Klaauw, 2002; McCrary, 2008). Although traditionally in the RD evaluation literature this was usually done using linear controls, allowing for non-linearities in $m(x)$ can be important, especially in cases where we suspect x and y to be non-linearly related, such as when the outcome variable is bounded, or when we have reason to expect this relationship to change as a result of the program. In the scholarship award example, the link between student performance and academic ability may be enhanced or reduced after receiving a scholarship that pays for all college expenses. In that scenario, we would expect $m(x)$ (even though continuous at $\bar{x}$) to be non-differentiable at the cut-off.

Note that although the control function approach allows us to expand the estimation sample beyond the subset of observations close to cut-off, it clearly still would require a much larger sample of observations to provide the same precision as a randomized experiment. This is due to the collinearity between the terms in $m(x)$ and t in the regression equation, which reduces the independent variation in treatment status across observations, and thus the precision of program impact estimates. Goldberger (1972) and Bloom *et al.* (2005) consider two different versions of a constant treatment effect model with a linear $m(x)$ specification and compute that the sample for an RD analysis must respectively be 2.75 and 4 times that for a corresponding experiment to achieve the same degree of precision.

Generally, in case of mis-assignment relative to the cut-off score, the inclusion of $m(x)$ in the regression equation [5] is no longer adequate for avoiding biases due to group non-equivalence. A single exception to this is the case of random mis-assignment considered by Cain (1975), where the assignment error is independent of e given x. Not only would inclusion of the correct control function in this case produce a consistent estimate of the average treatment effect δ near the cut-off, it would actually be possible to non-parametrically identify the average treatment effect over the region of common support, by simply comparing average outcomes of the treated and non-treated at any given value of the assignment variable x.

In other fuzzy RD design cases, estimation of a control-function-augmented outcome equation [5] will generally not provide an estimate of a meaningful treatment effect. As discussed by Barnow *et al.* (1980), in case of a constant treatment effect, δ will be estimated with bias, where the bias will depend on the covariance of t and e conditional on x and may be positive or negative. It is, however, possible to solve this selection bias problem, by estimating the same control-function-augmented outcome equation, but where t_i is now replaced by an estimate of the propensity score $E[t_i|x_i]$. Assuming local independence of t_i and β_i conditional on x, then in a neighborhood of $\bar{x}$,

$$y_i = m(x_i) + \delta E[t_i|x_i] + w_i, \qquad [6]$$

where $w_i = y_i - E[y_i|x_i]$ and $m(x) = \alpha + E[u_i|x] + (E[\beta_i|x] - E[\beta_i|\bar{x}])E[t|x]$. With the local continuity assumption again implying that $m(x)$ will be continuous at the cut-off, and with $E[t_i|x_i]$ being discontinuous at $\bar{x}$, δ in this regression will measure the ratio in [4], which in this case equals the average local treatment effect $\delta = E[\beta_i|\bar{x}]$. Similarly, δ can be interpreted as a LATE if we replaced the local independence assumption with the local monotonicity condition of Imbens and Angrist (1994).

This characterization logically leads to the two-stage procedure adopted by van der Klaauw (2002), where in the first stage we specify the treatment or selection rule in the fuzzy RD design as

$$t_i = E[t_i|x_i] + v_i = f(x_i) + \gamma 1\{x_i \geq \bar{x}\} + v_i, \qquad [7]$$

where $f(\cdot)$ is some function of x, which is continuous at $\bar{x}$. By specifying the functional form of f (or by estimating f semi- or non-parametrically) we can estimate γ, the discontinuity in the propensity score function at $\bar{x}$. In the second stage the control-function-augmented outcome equation [6] is then estimated with t_i replaced by the first-stage estimate of $E[t_i|x_i] = \Pr[t_i = 1|x_i]$ as in Maddala and Lee (1976). With correctly specified $f(x)$ and $m(x)$ functions, this two-stage procedure yields a consistent estimate of the treatment effect. The approach is similar in spirit to those proposed earlier in the RD evaluation literature by Spiegelman (1979) and Trochim and Spiegelman (1980).

It is worth pointing out that in case of a parametric approach, if we assume the same functional form for $m(x)$ and $f(x)$, then the two-stage estimation procedure described here will be equivalent to two-stage least squares (in case of linear-in-parameter specifications) with $1\{x_i \geq \bar{x}\}$ and the terms in $m(x)$ serving as instruments. Because of the popularity of this particular parametrization, the RD approach is often interpreted as being equivalent to an Instrumental Variable approach, as it implicitly imposes an exclusion restriction by excluding $1\{x_i \geq \bar{x}\}$ as a variable in the outcome equation.

5. Semi-parametric estimation

In case of the RD design, valid parametric inference requires a correct specification both of the control function $m(x)$ and of $f(x)$ in the treatment equation. Although our identification results relied on a local continuity assumption, the parametric estimation approach imposes global continuity and frequently also global differentiability (except perhaps at the discontinuity point) of the conditional expectation functions. Even though this permits us to use all data points including those far from the cut-off, the choice of functional form and of the order of the polynomial in polynomial specifications is delicate.

One way to reduce the potential for mis-specification bias is to continue assuming global continuity and differentiability but to estimate the functions $m(x)$ and $f(x)$ semi-parametrically. van der Klaauw (2002) proposed the use of a power series approximation for estimating these functions as $\sum_{j=1}^{J}\eta_j \cdot x^j$, where the number of power functions, J, is estimated from the data by generalized cross-validation as in Newey *et al*. (1990). In this case, the correct standard errors would generally be larger than conventional standard errors reflecting the fact that the chosen polynomial specification is an approximation.

More recently proposed semi-parametric estimation methods rely on even less restrictive smoothness conditions away from the discontinuity to estimate the size of the discontinuities in the conditional expectation functions $E[y|x]$ and $E[t|x]$. Estimation of the limits $\lim_{x\downarrow\bar{x}}E[z|x]$ and $\lim_{x\uparrow\bar{x}}E[z|x]$ in [4] in these cases is based mainly on data in a neighborhood on either side of the cut-off point. Asymptotically this neighborhood needs to shrink, as with usual non-parametric estimation, implying that we should expect a slower than parametric rate of convergence in estimating treatment impacts.

HTV considered the use of kernel methods, where the left- and right-hand side limits of the conditional expectations $E[z|x]$ that appear in [4] are estimated using Nadaraya–Watson estimators defined as $\Sigma_i k_h(\bar{x} - x_i)z_i w_i / \Sigma_i k_h(\bar{x} - x_i)w_i$ and $\Sigma_i k_h(\bar{x} - x_i)z_i(1 - w_i)/\Sigma_i k_h(\bar{x} - x_i)(1 - w_i)$ where $w_i = 1\{x \geq \bar{x}\}$, $k_h(\cdot) = \frac{1}{h}k(\cdot/h)$, $k(\cdot)$ is a kernel function, and h denotes a bandwidth that controls the size of the local neighborhood to average over. Each term represents a weighted average of z for data contained in a small neighborhood left or right of the cut-off, with the weights depending on the distance from the cut-off $(\bar{x} - x)$ as well as the bandwidth and kernel function. In case of a one-sided uniform kernel this leads to the 'local Wald' estimator of HTV:

$$\frac{\sum_{i\in S} y_i w_i \Big/ \sum_{i\in S} w_i - \sum_{i\in S} y_i(1-w_i) \Big/ \sum_{i\in S}(1-w_i)}{\sum_{i\in S} t_i w_i \Big/ \sum_{i\in S} w_i - \sum_{i\in S} t_i(1-w_i) \Big/ \sum_{i\in S}(1-w_i)} \qquad [8]$$

where S denotes the subsample around the cut-off point defined by $\bar{x} - h \leq x_i \leq \bar{x} + h$. This estimator is numerically equivalent to a Wald estimator applied to a discontinuity sample S, where t_i is instrumented by w_i.

This kernel-based estimator, though consistent, suffers from the same poor asymptotic bias behavior that many non-parametric conditional expectation estimators have at boundary points. In case of a positive slope of $m(x)$ near $\bar{x}$, the average outcome for observations just right of the cut-off will generally provide an upward biased estimate of $\lim_{x\downarrow\bar{x}}E[y|x]$ whereas the average outcome of observations just to the left of the cut-off would provide a downward biased estimate of $\lim_{x\uparrow\bar{x}}E[y|x]$. Therefore, in the case of a sharp design, where the treatment effect is estimated by taking the difference between these two averages, this generates a positive finite sample bias, implying that the bandwidth needs to shrink at a very fast rate, leading to a slow rate of convergence.

An alternative approach considered by HTV and Porter (2003) is to estimate the limits in [4] using local polynomial regression (Fan, 1992). Such estimators are known for their superior boundary behavior (Fan and Gijbels, 1996). Although HTV focused on the linear case, Porter also considered higher-order local polynomial estimators. The order p local polynomial estimator of the right-hand side limit of a conditional expectation $E[z|x]$ is defined as the value of $\hat{\delta}$, which solves the minimization problem

$$\min_{\delta,b_1,\cdots,b_p} \frac{1}{n}\sum_{i=1}^{n} k_h(x_i-\bar{x})w_i\left[z_i-\delta-b_1(x_i-\bar{x})-\cdots-b_p(x_i-\bar{x})^p\right]^2.$$

The estimator of the left-hand side limit represents the value of δ, which minimizes the same criterion but with w_i replaced by $1-w_i$. HTV and Porter show that by accounting for the polynomial type behavior of the conditional expectation near the boundary, the bias in the boundary estimate becomes of much lower order compared with that in kernel estimation. Moreover, as shown by Porter, RD estimators based on local polynomial regression achieve the optimal rate of convergence.

Another estimator that, under certain conditions, attains the optimal convergence rate is based on partially linear model estimation. This estimator, which was proposed by Porter (2003) and based on Robinson's (1988) partially linear model estimator, imposes additional smoothness on the conditional expectation functions by assuming that $m(x)$ in equation [5] is not only continuous, but also continuously differentiable at $\bar{x}$. The partially linear model estimator for estimating δ in equation [5] can be found by minimizing the average squared deviation between $(y-\delta w)$ and the non-parametric estimate of $m(x)$. That is, it is the value of δ that solves

$$\min_{\delta}\sum_{i=1}^{n}\left[y_i-\delta w_i-\sum_{j=1}^{n} r_j^i(y_j-\delta w_j)\right]^2$$

where $r_j^i = k_h(x_i - x_j) \Big/ \sum_{l=1}^{n} k_h(x_i - x_l)$. The RD estimate in case of a fuzzy design can then be computed by using the same approach to estimate the discontinuity in $E[t|x]$, followed by taking the ratio of the two estimated discontinuities.

Unlike the local polynomial and kernel estimators, which are based on weighted averages calculated with data belonging to only one side of the discontinuity, the partial linear estimator at each value of x uses data from both sides of the cut-off. Thus in estimating $m(x)$ near $\bar{x}$, positive weight is given to data points from both sides of the threshold. This is in fact the reason for why this estimator does not suffer the same poor bias behavior as the estimator based on one-sided kernel estimation, because it exploits the fact that the biases at either side of the discontinuity are comparable and thus in a sense can be cancelled out. Porter shows that the partially linear estimator achieves the same order of bias reduction as the local polynomial estimator, and similarly achieves the optimal convergence rate, as long as the additional smoothness condition on the derivatives at the discontinuity point hold. The later condition, however, may not be desirable for analysing heterogeneous treatment effect cases in which we suspect an interaction effect between treatment receipt and the assignment variable that would produce a jump in the derivative of $m(x)$ at $\bar{x}$ (as illustrated in the earlier scholarship example). For this reason, the local or piecewise polynomial estimators can be expected to be more robust.

Finally, the results derived by HTV and Porter regarding the asymptotic properties of the local polynomial regression and partially linear model estimators rely on a known degree of smoothness of the conditional expectations functions $m(x)$ and $f(x)$. As pointed out by Sun (2005), if the degree of smoothness is unknown and incorrectly chosen, the estimates obtained with these methods may actually inflate, rather than reduce the bias due to the boundary problem. To avoid this problem, Sun proposes the use of an adaptive estimator, which first estimates the degree of smoothness in the data prior to implementing either estimator.

6. Sensitivity analysis and validity tests

Given the common practice of researchers to estimate RD treatment effects parametrically, and in light of the minimal assumptions required for identification, it is important to supplement any parametric analysis with several robustness checks. First, it is informative to start an RD analysis with a graphical portrayal of the data to show the presence of a discontinuity in the probability of assignment. An effective way of doing this is to plot averages for equally sized non-overlapping bins, on either side of the cut-off such that no histogram bin includes points to both the left and right of the point of discontinuity (see, for example, DiNardo and Lee, 2004; Lee *et al*., 2004; van der Klaauw, 2008; Lemieux and Milligan, 2008). A similar analysis for the outcome variable would then provide a first indication of the existence of a non-zero treatment effect, if the plot reveals a similar discontinuity at the same cut-off in the average outcome measure.

Second, it is important to investigate how sensitive the parametric estimates are to alternative and more flexible parametric assumptions, for example, by adding higher-order terms in polynomial specifications, and by exploring polynomial splines (with separate polynomials on both sides of the discontinuity). Third, and close in spirit to the idea of local identification, is to examine the robustness of the parametric results by restricting the sample to a subset of observations more closely clustered around the cut-off. A linear control function is likely to provide a reasonable approximation of the true functional form within a small neighborhood of the cut-off. Alternatively one could report easy-to-calculate 'local Wald' estimates for varying bandwidth sizes (van der Klaauw, 2002, 2008). By taking increasingly narrow windows around the discontinuity point, the influence of data points further away from the discontinuity is reduced and will generally lead to a reduction in the risk of a mis-specification bias, but at the obvious cost of a loss in efficiency.

The internal validity of the RD approach relies on the local continuity of conditional expectations functions around the discontinuity point. Continuity implies that the average *potential* outcomes for individuals marginally below the threshold will equal those for the group just above the threshold. For each specific application of the design, it is important to consider the potential for economic behavior to invalidate the local continuity assumption. If agents exercise control over their values of the assignment variable, or if administrators can strategically choose what assignment variable to use or which cut-off point to pick then comparability near the cut-off may be violated. Both types of behavior may lead to sorting of individuals around the cut-off point, where those below the cut-off may differ on average from those just above the cut-off. Continuity will also be violated in the presence of other programs that use a discontinuous assignment rule with the exact same assignment variable and cut-off score.

To judge the potential for behavioral effects that could lead to violation of the continuity assumption, it is important to analyse agents' ability and incentive to affect their values of the assignment variable. If the assignment rule and selection mechanism are well understood, and treatment is considered beneficial, then it would be plausible that agents would engage in manipulation of the assignment variable in order to obtain desirable treatment assignments. Such sorting is less likely if the existence of the assignment rule is not known and hard to uncover, if the location of the cut-off is unknown or uncertain, if the assignment variable cannot be manipulated or if there is insufficient time for agents to do so.

Similarly, it is important to analyse the extent to which administrators can define the assignment variable and cut-off score, as well as the information that is available at the time these variables are chosen. If the choice is made based on the realized distribution of the assignment variable, it may in certain cases be optimal for administrators to set the cut-off by picking a randomly realized 'break point' in the data where by chance the average characteristics of individuals change. Existing rules or laws often predetermine the location of the cut-off score, restricting the administrator's scope for strategic behavior. For example, a rule may specify that the cut-off be computed as the average of the assignment variable in a well-defined population.

Although in many cases it will be likely that agents or administrators (or both) will exercise some control over the value of the assignment variable or position of the cut-off, this does not necessarily imply that the continuity assumptions will be violated. Lee (2008) analyses the conditions under which an ability to manipulate the assignment variable may invalidate the RD identification assumptions. He shows in the context of a sharp RD design that as long as individuals do not have *perfec*t control over the position of the assignment variable relative to the cut-off score, the continuity assumption will be satisfied. More precisely, as long as the score on which treatment is based contains an independent random chance element, so that conditional on the individual's choices and characteristics the density of the realized score for each individual is continuous, it will imply local continuity of average potential outcomes at the discontinuity point. Moreover, this model generates variation in the treatment status that resembles that of randomized treatment assignment within a neighborhood of the cut-off. That is, the model implies that the local conditional independence assumption will be satisfied, such that close to this threshold, all variables determined prior to assignment (both observable and unobservable) will be independent given treatment status.

Although an ability to manipulate the value of the assignment variable will therefore not necessarily invalidate the RD approach, clearly one would expect it to affect the composition of the sample around the discontinuity point: the program may cause different subgroups to locate near the cut-off than would have been the case if the program rules were unknown or if agents had no control over their value of the assignment variable. This in turn affects the interpretation of the estimated treatment effect. As pointed out by Lee (2008) the RD approach in such cases would identify a weighted average treatment effect, with individuals whose underlying characteristics or actions that make them more likely to obtain a draw of the score near the cut-off receiving more weight than those who are unlikely to obtain such a draw. Furthermore, the counterfactual would not correspond to a world without the program, but rather one with a program but where the individual does not receive treatment due to having an assignment score just below the cut-off. Thus the RD approach would estimate a causal effect for a subgroup of the population that may be program dependent, an effect that would nevertheless continue to inform the policy-maker on the consequences of small changes in the cut-off score.

These results indicate that for sorting to undermine the causal interpretation of the RD approach, agents need to be able to sort precisely around the assignment cut-off. Although in many cases the extent of agents' control (and thus the local continuity assumption) is fundamentally untestable, a number of additional validity tests have been developed to bolster the credibility of the RD design. A first validity test is to look for direct evidence of precise sorting around the assignment cut-off by searching for a sharp break in the distribution of the assignment variable at the cut-off. If individuals cannot exercise precise control over the assignment score or do not know the cut-off, then the distribution should be smooth close to the cut-off value (Chen and van der Klaauw, 2008; Lee, 2008; Lemieux and Milligan, 2008). If they can, we would expect to find a jump in this distribution at the discontinuity point. Evidence of such a jump would then be suggestive of a violation of the RD assumptions.

McCrary (2008) developed a local linear density estimator to test for a discontinuity in the density at a given point. More informally, one could plot the frequency of the assignment variable for equally sized non-overlapping bins on either side of the cut-off. This approach was used by McEwan and Urquiola (2005) to investigate the applicability of the RD approach for evaluating the impact of class size on student outcomes. The discontinuity in their analysis arises due to a law in Chile mandating a maximum class size of 45 students, which generates drops in the average class size level as a function of grade-specific enrollment at values 45, 90, 135, etc. A similar maximum class size rule was exploited by Angrist and Lavy (1999). McEwan and Urquiola present histograms that reveal spikes in the distribution of school enrollment, with sharply higher numbers of schools at or just below these enrollment cut-off levels, which is consistent with precise sorting of schools around these levels. This seems plausible in that schools appear to be able to informally discourage some students from attending.

The absence of a discontinuity in the density of the assignment variable is neither a necessary nor a sufficient condition for valid inference. As pointed out by McCrary (2008) a density test will only be informative if manipulation is monotonic, i.e. if the program induces agents to change the assignment variable in one direction only. Second, such manipulation or gaming may be for reasons that are unrelated to potential outcomes, such as a desire to avoid administrative hassle involved with treatment receipt, while leaving continuity of expected potential outcome functions intact. However in most cases where the program rules are well known and individuals have direct control over the assignment score, and where one would not expect to find a discontinuity in the density if the program had no effect, finding such a discontinuity would compromise the validity of the RD approach.

A second validity test, which has been frequently applied to analyse the credibility of the RD approach, is to search for evidence that individuals on either side of the cut-off are observationally similar. If continuity assumptions are correct then we would expect both groups of individuals to have similar average observed and unobserved characteristics. To visually test for an imbalance of observed covariates at the cut-off we could for each observable characteristic plot its average for non-overlapping bins on either side of the cut-off. Alternatively, we can test whether the relationship between the assignment variable and any baseline covariates is smooth in the vicinity of the discontinuity point by repeating the RD analysis treating the characteristics as outcome variables (van der Klaauw, 2008). In their analysis of class size effects, McEwan and Urquiola (2005) found sharp breaks in the distribution of student characteristics and family income at the discontinuity points. For example, private school students in schools just to the right of these cut-offs (those in small classes) tend to have higher average family income than those just to the left (those in large ones). Thus, the treatment of smaller classes is confounded with a relative improvement in observed student socio-economic status that is associated with the earlier finding of sorting close to the cut-offs. In a similar vein, one can test the local independence result implied by the behavioral 'imperfect control' model of Lee. This condition implies that in a neighborhood around the cut-off all variables determined prior to assignment should be independent of treatment status. Thus characteristics of recipients and non-recipients should also be comparable in this case.

Although a lack of significant jumps in observable characteristics along the discontinuity provides evidence consistent with the validity of the RD design, this does not necessarily imply that no such discontinuity exists in unobservables. At the same time, finding evidence of a discontinuity does not necessarily imply that the RD identification conditions are violated. Finding such a discontinuity would only be relevant if the observed characteristic is actually related to the outcome of interest. This suggests a third test, in which we assess whether RD estimates are sensitive to inclusion of observed characteristics as controls. This test can be interpreted as a test for an imbalance in *relevan*t characteristics (van der Klaauw, 2002, 2008). If the local continuity assumptions are satisfied then the only possible gain from controlling for observables in parametric or semi-parametric estimation of treatment effects in an RD design would come from a reduction in sampling variability (assuming they have explanatory power). However, if including these controls lead to significant changes in estimates, this would suggest that these continuity assumptions may be violated. An alternative way to test the sensitivity of RD estimates to the inclusion of individual covariates, proposed by Lee (2008), is to first regress the outcome variable on a vector of individual characteristics and then to repeat the RD analysis using the residuals as outcome variable, instead of the outcome variable itself.

In some applications data are available from a baseline period in which the program did not yet exist, or for a group of individuals that was not eligible for treatment. In such a case the credibility of the design can be significantly enhanced by repeating the RD analysis with such data. Finding a zero treatment effect in such a falsification test would suggest that a non-zero post-program effect was not an artifact of the specific RD model specification or estimation approach chosen. It also could be used to rule out the possibility that the estimated effects are caused by another program using the same cut-off and assignment variable.

An additional specification test for a spurious relationship between treatment and outcome at a cut-off point was proposed by Kane (2003). His test analyses whether the actual cut-off fits the data better than other nearby cut-offs. To do so, Kane estimated his model for a set of alternative cut-off values, and plotted the log likelihood value associated with each. A clear spike in the log likelihood at the actual relative to alternative cut-off values could help allay concerns that the found (local) relationship was spurious.

Finally, it is worth pointing out that although evidence of sorting would often imply that the RD approach would not estimate a meaningful treatment effect, in many cases the finding of a non-zero effect estimate can nevertheless be seen as evidence that a treatment matters. For example, in analysing the impact in the USA of an anti-discrimination or equal employment opportunity law affecting firms with 15 or more workers, Hahn *et al.* (1999) found a discontinuity in the percentage of minorities employed in small firms, with firms with 13 and 14 workers having a significantly lower share of minority workers than firms with 15 or 16 workers. As firms directly choose their number of employees, sorting represents a legitimate concern to this RD application. Although the results in Hahn *et al.* did not find evidence of heaping at total employment values just below 15, it is difficult to rule out the possibility that the significant jump in the share of minority workers at 15 was the result of strategic employment decisions in order to avoid the risk of sanctions for violating the law. Although this would render the RD estimate of the law's impact biased, the presence of a non-zero jump nevertheless provides credible evidence that the law actually affected firm hiring and lay-off decisions.

7. Extensions

The analysis of the RD design in this paper has focused on the binary treatment case with a selection rule containing a single discontinuity at a known cut-off in a continuously distributed variable. The case of a binary treatment and single known cut-off can be readily extended to one where there are multiple treatment dose levels and multiple cut-offs or 'cut-off ranges' within which the treatment dose varies continuously.

With multiple treatment dose levels for t, equation [1] can be interpreted as describing the average potential outcomes across individuals under alternative treatment dose assignments. In case of a sharp design, the impact defined in [4] at a discontinuity point can then be interpreted as the average impact of a change in treatment dose equal to the jump at the discontinuity point, for individuals near the cut-off. With a fuzzy design, t_i is again a random variable given x, but the conditional mean $E[t_i|x]$ is known to be discontinuous at a given cut-off. Considering the more general case where treatment effects may be non-linear, with β_i in [1] replaced by $\beta_i(t_i)$, the results of Angrist and Imbens (1995) can be extended to show that under the same continuity and local monotonicity assumptions discussed earlier, [4] identifies a weighted average causal response across the different treatment dose levels observed in the data, for those whose treatment dose changes discontinuously at the cut-off value (see van der Klaauw, 2002).

When there are several discontinuity points, one can identify treatment effects over a wider range of the support of x. As the number of points approaches infinity, a discontinuity design approximates the conditions of a randomized experiment. If treatment effects were locally constant, say within quintiles of x, then an RD design with a fixed number of discontinuity points would allow identification of the range of treatment impacts. Multiple discontinuity points also allow a test of the common effect assumption (see HTV).

Porter (2003) considered the case where the treatment assignment or selection mechanism is known to have a single discontinuity, but where the exact location of the discontinuity point is unknown. He proposes the use of methods developed for the estimation of structural breaks in time-series data to estimate the point in the distribution where the discontinuity was most likely to have taken place. One can then proceed with the RD analysis as before, treating the estimated cut-off as the known true one. Importantly, Porter points out that estimators of the discontinuity point have a faster convergence rate than the estimators of the discontinuity size and so do not affect the limiting distribution results derived for the known cut-off case.

Finally, a potentially important practical issue that arises in empirical applications of the RD design is that rather than being continuous, the assignment or selection variable is often discrete or only reported in coarse intervals. As pointed out by Lee and Card (2008), in such a case it is no longer possible to compare observations just above or just below the cut-off because of the inevitable gap between discrete values. Even if the number of observations goes to infinity, we will never have any observations *jus*t below the cut-off to estimate the left-side limits in [4]. Therefore, the conditions presented earlier are no longer sufficient for the identification of a causal effect. Additional assumptions are required to allow interpolation and extrapolation between discrete values. Lee and Card discuss such parametric assumptions as well as inference in this case.

References

Angrist J. D. and Imbens G. W. (1991) 'Sources of Identifying Information in Evaluation Models', NBER Technical Working Paper 0117.

Angrist J. D. and Imbens G. W. (1995) 'Two-stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity', *Journal of American Statistical Association* 90(430): 431–442.

Angrist J. D. and Krueger A. B. (1999) 'Empirical Strategies in Labor Economics' in Ashenfelter O. and Card D. (eds.) *Handbook of Labor Economics*, Vol. 3, Amsterdam: North Holland: 1277–1366.

Angrist J. D. and Lang K. (2004) 'Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program', *American Economic Review* 94: 1613–1634.

Angrist J. D. and Lavy V. (1999) 'Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement', *Quarterly Journal of Economics* 114: 533–575.

Barnow B. S., Cain G. G. and Golberger A. S. (1980) 'Issues in the Analysis of Selectivity Bias' in Stormsdorfer E. and Farkas G. (eds.) *Evaluation Studies Review Annual*, Vol. 5, Beverly Hills, CA: Sage: 43–59.

Battistin E. and Rettore E. (2008) 'Ineligibles and Eligible Non-participants as a Double Comparison Group in Regression Discontinuity Designs', *Journal of Econometrics* 142(2): 715–730.

Bauer T. K., Bender S. and Schmidt C. M. (2004) 'Evaluating the Labor Market Effect of Compulsory Military Service: A Regression–Discontinuity Approach', Unpublished manuscript, IZA.

Black S. E. (1999) 'Do Better Schools Matter? Parental Valuation of Elementary Education', *Quarterly Journal of Economics* 114(2): 577–599.

Bloom H. S. (1984) 'Accounting for No-shows in Experimental Evaluation Designs', *Evaluation Review* 8: 225–246.

Bloom H. S., Kemple J., Gamse B. and Jacob R. (2005) 'Using Regression Discontinuity Analysis to Measure the Impacts of Reading First', Paper presented at the annual conference of the American Educational Research Association, Montreal, Canada.

Buddelmeyer H. and Skoufias E. (2004) 'An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA', World Bank Policy Research Working Paper No. 3386.

Cain G. (1975) 'Regression and Selection Models to Improve Nonexperimental Comparisons' in Bennett C. and Lumsdaine A. (eds.) *Evaluation and Experiment*, New York: Academic Press: 297–317.

Campbell D. T. (1969) 'Reforms as Experiments', *American Psychologist* 24: 409–429.

Card D. and Shore-Sheppard L. D. (2004) 'Using Discontinuous Eligibility Rules to Identify the Effects of Federal Medicaid Expansions on Low-income Children', *Review of Economics and Statistics* 86(3): 752–766.

Chakrabarti R. (2008a) 'Can Increasing Private School Participation and Monetary Loss in a Voucher Program Affect Public School Performance? Evidence from Milwaukee', Federal Reserve Bank of New York, Staff Report 300, *Journal of Public Economics* 92: 1371–1393.

Chakrabarti R. (2008b) 'Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs', Federal Reserve Bank of New York, Staff Report 315.

Chay K. Y. and Greenstone M. (2005) 'Does Air Quality Matter? Evidence from the Housing Market', *Journal of Political Economy* 113: 376–424.

Chen K. and Shapiro J. M. (2007) 'Does Harsher Prison Conditions Reduce Recidivism? A Discontinuity-based Approach', *American Law and Economics Review* 9(1): 1–29.

Chen S. and van der Klaauw W. (2008) 'The Work Disincentive Effects of the Disability Insurance Program in the 1990s', *Journal of Econometrics* 142(2): 757–784.

Cook T. D. and Campbell D. T. (1979) *Quasi-experimentation: Design and Analysis Issues for Field Settings*, Boston, MA: Houghton-Miffin.

DiNardo J. and Lee D. S. (2004) 'Economic Impacts of New Unionization on U.S. Private Sector Employers: 1984–2001', *Quarterly Journal of Economics* 119(4): 1383–1442.

Fan J. (1992) 'Design-adaptive Nonparametric Regression', *Journal of the American Statistical Association* 87: 998–1004.

Fan J. and Gijbels I. (1996) *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.

Goldberger A. S. (1972) 'Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations', Discussion Paper 123-172, Madison, IRP.

Hahn J., Todd P. and van der Klaauw W. (1999) 'Evaluating the Effect of an Antidiscrimination Law using a Regression–Discontinuity Design', NBER Working Paper 7131, Cambridge, MA.

Hahn J., Todd P. E. and van der Klaauw W. (2001) 'Identification and Estimation of Treatment Effects with a Regression–Discontinuity Design', *Econometrica* 69: 201–209.

Heckman J. J. and Robb R. (1985) 'Alternative Methods for Evaluating the Impact of Interventions' in Heckman J. and Singer B. (eds.) *Longitudinal Analysis of Labor Market Data*, New York: Cambridge University Press: 156–245.

Heckman J. J., Ichimura H. and Todd P. E. (1997) 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme', *Review of Economic Studies* 64: 605–654.

Imbens G. W. and Angrist J. D. (1994) 'Identification and Estimation of Local Average Treatment Effects', *Econometrica* 62(2): 467–476.

Imbens G. W. and Lemieux T. (2008a) 'Special Issue Editors' Introduction: The Regression Discontinuity Designs — Theory and Applications', *Journal of Econometrics* 142(2): 611–614.

Imbens G. W. and Lemieux T. (2008b) 'Regression Discontinuity Designs: A Guide to Practice', *Journal of Econometrics* 142(2): 615–635.

Jacob B. A. and Lefgren L. (2004a) 'The Impact of Teacher Training on Student Achievement: Quasi-experimental Evidence from School Reform Efforts in Chicago', *Journal of Human Resources* 39(1): 50–79.

Jacob B. A. and Lefgren L. (2004b) 'Remedial Education and Student Achievement: A Regression–Discontinuity Analysis', *Review of Economics and Statistics* 86(1): 226–244.

Joyce T., Kaestner R. and Colman S. (2006) 'Changes in Abortions and Births and the Texas Parental Notification Law', *The New England Journal of Medicine* 354(10): 1031–1038.

Kane T. J. (2003) 'A Quasi-experimental Estimate of the Impact of Financial Aid on College-going', NBER Working Paper No. W9703.

Kane T. J., Riegg S. and Staiger D. (2006) 'School Quality, Neighborhoods, and Housing Prices', *American Law and Economics Review* 8(2): 183–212.

van der Klaauw W. (1997) 'A Regression–Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment', C. V. Starr Center for Applied Economics, New York University, Working Paper 97-10.

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.

van der Klaauw W. (2008) 'Breaking the Link between Poverty and Low Student Achievement: An Evaluation of Title I', *Journal of Econometrics* 142(2): 731–756.

Lavy V. (2004) 'Performance Pay and Teachers' Effort, Productivity and Grading Ethics', NBER Working Paper 10622.

Lee D. S. (2008) 'Randomized Experiments from Non-random Selection in U.S. House Elections', *Journal of Econometrics* 142(2): 675–697.

Lee D. S. and Card D. (2008) 'Regression Discontinuity Inference with Specification Error', *Journal of Econometrics* 142(2): 655–674.

Lee D. S., Moretti E. and Butler M. J. (2004) 'Do Voters Affect or Elect Policies? Evidence from the U.S. House', *Quarterly Journal of Economics* 119(3): 807–860.

Lemieux T. and Milligan K. (2008) 'Incentive Effects of Social Assistance: A Regression Discontinuity Approach', *Journal of Econometrics* 142(2): 807–828.

Ludwig J. and Miller D. L. (2007) 'Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design', *Quarterly Journal of Economics* 122(1): 159–208.

Maddala G. S. and Lee L. (1976) 'Recursive Models with Qualitative Endogenous Variables', *Annals of Economic and Social Measurement* 5(4): 525–545.

Matsudaira J. (2008) 'The Effects of Mandatory Summer School, a Regression–Discontinuity Analysis', *Journal of Econometrics* 142(2): 829–850.

McCrary J. (2008) 'Testing for Manipulation of the Running Variable in the Regression Discontinuity Design', *Journal of Econometrics* 142(2): 698–714.

McEwan P. J. and Urquiola M. (2005) 'Economic Behavior and the Regression–Discontinuity Design: Evidence from Class Size Reduction', Unpublished manuscript, Columbia University.

Newey W. K., Powell J. L. and Walker J. R. (1990) 'Semiparametric Estimation of Selection Models: Some Empirical Results', *American Economic Review Papers and Proceedings* 80: 324–328.

Pence K. M. (2006) 'Foreclosing on Opportunity: State Laws and Mortgage Credit', *Review of Economics and Statistics* 88(1): 177–182.

Porter J. (2003) 'Estimation in the Regression Discontinuity Model', Unpublished manuscript, Harvard University.

Robinson P. (1988) 'Root-n-consistent Semiparametric Regression', *Econometrica* 56: 931–954.

Rosenbaum P. R. and Rubin D. B. (1983) 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika* 40: 41–55.

Spiegelman C. (1979) 'Estimating the Effect of a Large Scale Pretest Posttest Social Program' in *Proceedings of the Social Statistics Section*, American Statistical Association: 370–373.

Sun Y. (2005) 'Adaptive Estimation of the Regression Discontinuity Model', Unpublished manuscript, University of California, San Diego, CA.

Thistlethwaite D. and Campbell D. (1960) 'Regression–Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment', *Journal of Educational Psychology* 51: 309–317.

Trochim W. K. (1984) *Research Design for Program Evaluation: The Regression–Discontinuity Approach*, Beverly Hills, CA: Sage.

Trochim W. K. and Spiegelman C. (1980) 'The Relative Assignment Variable Approach to Selection Bias in Pretest-posttest Group Designs' in *Proceedings of the Survey Research Methods Section*, American Statistical Association: 376–380.