DISCUSSION PAPER SERIES

IZA DP No. 3638

Quantile Treatment Effects in the Regression Discontinuity Design

Markus Frölich
Blaise Melly

August 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Quantile Treatment Effects in the Regression Discontinuity Design

Markus Frölich
University of Mannheim
and IZA

Blaise Melly
Brown University

Discussion Paper No. 3638
August 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3638
August 2008

ABSTRACT

Quantile Treatment Effects in the Regression Discontinuity Design

This paper shows nonparametric identification of quantile treatment effects (QTE) in the regression discontinuity design (RDD) and proposes simple estimators. Quantile treatment effects are a very helpful tool to characterize the effects of certain interventions on the outcome distribution. The distributional impacts of social programs such as welfare, education, training programs and unemployment insurance are of large interest to economists.

JEL Classification: C13, C14, C21

Keywords: quantile treatment effect, causal effect, endogeneity, regression discontinuity

Corresponding author:

Markus Frölich
Universität Mannheim
Abteilung Volkswirtschaftslehre
L7, 3-5
68131 Mannheim
Germany
E-mail: froelich@uni-mannheim.de

1 Introduction

In recent years, the regression discontinuity design (RDD) has received tremendous attention in applied economic research.[1] All these applications focus on the estimation of average treatment effect. In many research areas, one is not only interested in mean impacts, but also in the *distributional* consequences of treatment interventions. In the field of education (e.g. Angrist and Lavy (1999), Puhani and Weber (2007)), educational inequality e.g. in cognitive achievement is of large public interest. When examining the effects of training (e.g. Black, Galdo, and Smith (2005)), policy makers are often more interested in the effects at the lower quantiles than at the upper quantiles. When analyzing the effects of unemployment insurance on unemployment durations (e.g. Lalive (2008)), the distribution of the unemployment durations is of interest, e.g. the risk of becoming long-term unemployed.[2]

Quantile treatment effects (QTE) are a convenient tool to characterize the potentially heterogenous impacts of variables on different points of an outcome distribution. In this paper we show how QTE can be identified *nonparametrically* in the regression discontinuity design and propose nonparametric estimators.[3] We also discuss the identification of the potential outcome distributions.

2 Identification of QTE in the RDD

Following the setup of Hahn, Todd, and van der Klaauw (2001), let $D_i \in \{0,1\}$ be a binary treatment variable, let Y_i^0, Y_i^1 be the individual potential outcomes. The potential outcomes as well as the treatment effect are permitted to vary freely across individuals, i.e. no constant treatment effect is assumed. In the examples mentioned, D may represent school quality, class

[1] For an incomplete list see e.g. Angrist and Lavy (1999), Battistin and Rettore (2002), Battistin and Rettore (2008), Black (1999), Black, Galdo, and Smith (2005), Black, Jang, and Kim (2006), Buddelmeyer and Skoufias (2003), Brügger, Lalive, and Zweimüller (2008), Chay and Greenstone (2005), Chay, McEwan, and Urquiola (2005), DiNardo and Lee (2004), Fredriksson and Öckert (2006), Forslund and NordströmSkans (2006), Imbens and Lemieux (2008), Jacob and Lefgren (2004a), Jacob and Lefgren (2004b), Gormley and Phillips (2005), Guryan (2001), Lalive (2008), Lalive, Wüllrich, and Zweimüller (2008), Leuven, Lindahl, Oosterbeek, and Webbink (2007), Matsudaira (2008), NordströmSkans and Lindqvist (2005), Öckert (2008), Puhani and Weber (2007), van der Klaauw (2002), van der Klaauw (2008) and the special issue of the Journal of Econometrics 2008.

[2] Note that this distribution also identifies the hazard rates.

[3] In future work we are going to derive the asymptotic properties of these estimators.

size, participation in training etc. Let Z_i be a variable that influences the treatment variable in a discontinuous way, e.g. total school enrollment, profiling risk score etc.

In the literature, often two different designs are examined: the *sharp* design where D_i changes for everyone at a known threshold z_0, e.g.

$$D_i = 1(Z_i \geq z_0). \tag{1}$$

In this sharp design, *all* individuals change programme participation status exactly at z_0. The *fuzzy* design, on the other hand, permits D to also depend on other (unobserved) factors but assumes that the treatment probability changes discontinuously at z_0:

$$\lim_{\varepsilon \to 0} E\left[D|Z = z_0 + \varepsilon\right] - \lim_{\varepsilon \to 0} E\left[D|Z = z_0 - \varepsilon\right] \neq 0. \tag{2}$$

This fuzzy design includes the sharp design as a special case when the left hand side of (2) is equal to one. Therefore the following discussion focusses on the more general fuzzy design.[4]

Identification of treatment effects requires two assumptions. First, the conditional distribution of Y^0 has to be continuous at z_0. Furthermore, an assumption on the treatment effect is required. Hahn, Todd, and van der Klaauw (2001) consider two different versions: In the first, they assume that the treatment effect is independent of D conditional on Z being near z_0. This is some kind of selection on observables assumption. As an alternative, they consider an instrumental variable type assumption, which assumes for the local *compliers* that the potential outcomes and the potential treatment status are independent of Z, near z_0. In the sharp design, both assumptions are equivalent. In the fuzzy design, the IV type assumption is much more frequently used in applications. We therefore focus on the IV type approach.

For stating the identification results, it is helpful to introduce more precise notation first. Let $\mathcal{N}_\varepsilon$ be an ε neighbourhood about z_0 and partition $\mathcal{N}_\varepsilon$ into $\mathcal{N}_\varepsilon^+ = \{z : z \geq z_0, z \in \mathcal{N}_\varepsilon\}$ and $\mathcal{N}_\varepsilon^- = \{z : z < z_0, z \in \mathcal{N}_\varepsilon\}$. According to their reaction to the instrument z over $\mathcal{N}_\varepsilon$ we can

[4]The fuzzy design may apply when the treatment decision contains some element of discretion. Case workers may have some discretion about whom they offer a programme, or they may base their decision also on criteria that are unobserved to the econometrician.

partition the population into five subpopulations:

$$\begin{aligned}
\mathcal{T}_{i,\varepsilon} &= a \quad \text{if} \quad D_i(z)=1 \quad \forall z \in \mathcal{N}_\varepsilon^- \quad \text{and} \quad D_i(z)=1 \quad \forall z \in \mathcal{N}_\varepsilon^+ \\
\mathcal{T}_{i,\varepsilon} &= n \quad \text{if} \quad D_i(z)=0 \quad \forall z \in \mathcal{N}_\varepsilon^- \quad \text{and} \quad D_i(z)=0 \quad \forall z \in \mathcal{N}_\varepsilon^+ \\
\mathcal{T}_{i,\varepsilon} &= c \quad \text{if} \quad D_i(z)=0 \quad \forall z \in \mathcal{N}_\varepsilon^- \quad \text{and} \quad D_i(z)=1 \quad \forall z \in \mathcal{N}_\varepsilon^+ \\
\mathcal{T}_{i,\varepsilon} &= d \quad \text{if} \quad D_i(z)=1 \quad \forall z \in \mathcal{N}_\varepsilon^- \quad \text{and} \quad D_i(z)=0 \quad \forall z \in \mathcal{N}_\varepsilon^+ \\
\mathcal{T}_{i,\varepsilon} &= i \quad \text{if} \quad D_i(z) \text{ is nonconstant over } \mathcal{N}_\varepsilon^- \text{ or over } \mathcal{N}_\varepsilon^+.
\end{aligned}$$

These subpopulations are a straightforward extension of the LATE concept of Imbens and Angrist (1994). The first group contains those units that will *always* be treated, the second contains those that will *never* be treated, and the third and fourth group contains the units that are treated only on one side of z_0. The fifth group (labelled indefinite) contains all units that switch at other values than z_0 and that react non-monotonously, e.g. they may first switch from $D=0$ to 1 and then back for increasing values of z. We will assume that in the limit only the first three groups exist, and that the fraction of compliers is positive. Note that in the sharp design, everyone is a complier for $\varepsilon \to 0$.

Assumption 1:

i) Existence of compliers	$\lim_{\varepsilon\to 0} \Pr(\mathcal{T}_\varepsilon = c \mid Z = z_0) > 0$
ii) Monotonicity	$\lim_{\varepsilon\to 0} \Pr(\mathcal{T}_\varepsilon = t \mid Z \in \mathcal{N}_\varepsilon) = 0 \qquad \text{for } t \in \{d, i\}$
iii) Independent IV	$\lim_{\varepsilon\to 0} \Pr(\mathcal{T}_\varepsilon = t \mid Z \in \mathcal{N}_\varepsilon^+) - \Pr(\mathcal{T}_\varepsilon = t \mid Z \in \mathcal{N}_\varepsilon^-) = 0 \qquad \text{for } t \in \{a, n, c\}$
iv) IV Exclusion	$\lim_{\varepsilon\to 0} F_{Y^1 \mid Z\in\mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = t}(u) - F_{Y^1 \mid Z\in\mathcal{N}_\varepsilon^-, \mathcal{T}_\varepsilon = t}(u) = 0 \qquad \text{for } t \in \{a, c\}$
	$\lim_{\varepsilon\to 0} F_{Y^0 \mid Z\in\mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = t}(u) - F_{Y^0 \mid Z\in\mathcal{N}_\varepsilon^-, \mathcal{T}_\varepsilon = t}(u) = 0 \qquad \text{for } t \in \{n, c\}$
v) Density at threshold	$F_Z(z)$ is differentiable at z_0 and $f_Z(z_0) > 0$

In words, Assumption 1 requires the existence of some compliers and the absence of defiers near z_0. In addition, the potential outcomes and the type are jointly independent of Z near z_0, that is $(Y^0, Y^1, \mathcal{T}) \perp\!\!\!\perp Z$ in a neighbourhood of z_0.

Define the τ-th quantile of Y as $Q_Y^\tau = \inf\{y : F_Y(y) \geq \tau\}$. Define $Q_{Y^d|c}^\tau = \lim_{\varepsilon\to 0} Q_{Y^d|Z\in\mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c}^\tau$ as the limit for the local compliers. The quantile treatment effect (QTE) for the compliers is then defined as

$$\Delta_{QTE}^\tau = Q_{Y^1|c}^\tau - Q_{Y^0|c}^\tau.$$

The following theorem shows that $Q^{\tau}_{Y^1|c}$ and $Q^{\tau}_{Y^0|c}$ are identified under Assumption 1 by a very simple weighted quantile regression with weights 1 and -1. Define $I^+ = 1(Z \geq z_0)$ and $I^- = 1 - I^+$.

Theorem 1 (Quantiles of potential outcomes) *Under Assumption 1, the quantiles of the potential outcomes for the local compliers are identified as the solution of the following optimization problem*

$$\begin{aligned} Q^{\tau}_{Y^1|c} &= \lim_{\varepsilon \to 0} \arg\min_q E\left[\rho_\tau(Y-q)\left(2I^+ - 1\right) | Z \in \mathcal{N}_\varepsilon, D = 1\right] \\ Q^{\tau}_{Y^0|c} &= \lim_{\varepsilon \to 0} \arg\min_q E\left[\rho_\tau(Y-q)\left(2I^- - 1\right) | Z \in \mathcal{N}_\varepsilon, D = 0\right] \end{aligned}$$

where $\rho_\tau(u) = u \cdot \{\tau - 1(u < 0)\}$. *(All proofs are given in the appendix.)*

Based on this representation, a straightforward estimator of the quantiles is obtained as

$$\begin{aligned} \hat{Q}^{\tau}_{Y^1|c} &= \arg\min_q \sum_{i:D_i=1} \rho_\tau(Y_i - q)\left(2 \cdot 1(Z_i \geq z_0) - 1\right) K\left(\frac{Z_i - z_0}{h}\right) \\ \hat{Q}^{\tau}_{Y^0|c} &= \arg\min_q \sum_{i:D_i=0} \rho_\tau(Y_i - q)\left(2 \cdot 1(Z_i \geq z_0) - 1\right) K\left(\frac{Z_i - z_0}{h}\right), \end{aligned}$$

where K is a kernel function. These are simple univariate quantile regression with weights 1 and -1, multiplied with kernel weights. Note that $\hat{Q}^{\tau}_{Y^1|c}$ and $\hat{Q}^{\tau}_{Y^0|c}$ are estimated from independent observations, since the former uses only the $D_i = 1$ observations and the latter only the $D_i = 0$ observations.

Despite its simplicity one should note that the objective function of the weighted quantile regression estimator is not convex since some of the weights are negative. This complicates the optimization problem a little because local optima could exist and conventional linear programming algorithms cannot be used. The problem is, however, not very serious because we have two one-dimensional estimation problems in the treated and non-treated populations. In addition, the objective function can change only at the values of Y_i observed in the sample such that only n weighted means need to be computed to find the global minimum.

We explore now an alternative approach that may be fruitful when we want to estimate many (or all) QTE. To this end we estimate the cumulative distribution function. Define

$$F_{Y^d|c}(u) = \lim_{\varepsilon \to 0} F_{Y^d|Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c}(u).$$

Theorem 2 shows that the distribution functions of the potential outcomes for compliers are identified by the ratio of two weighted means.

Theorem 2 (Distribution of potential outcomes) *Under Assumption 1, the distribution of the potential outcomes for the local compliers are identified as*

$$\begin{aligned} F_{Y^1|c}(u) &= \lim_{\varepsilon\to 0}\frac{E\left[1\left(Y\leq u\right)\cdot\left(2I^+-1\right)|Z\in\mathcal{N}_\varepsilon, D=1\right]}{E\left[2I^+-1|Z\in\mathcal{N}_\varepsilon, D=1\right]} \\ F_{Y^0|c}(u) &= \lim_{\varepsilon\to 0}\frac{E\left[1\left(Y\leq u\right)\cdot\left(2I^--1\right)|Z\in\mathcal{N}_\varepsilon, D=0\right]}{E\left[2I^--1|Z\in\mathcal{N}_\varepsilon, D=0\right]}. \end{aligned}$$

Note that in the *sharp* design, everyone is a complier at z_0, such that the cdf of the potential outcomes in the population is identified in this case as

$$\begin{aligned} \lim_{\varepsilon\to 0}F_{Y^1|Z\in\mathcal{N}_\varepsilon}(u) &= \lim_{\varepsilon\to 0}E\left[1\left(Y\leq u\right)|Z\in\mathcal{N}_\varepsilon, D=1\right] \\ \lim_{\varepsilon\to 0}F_{Y^0|Z\in\mathcal{N}_\varepsilon}(u) &= \lim_{\varepsilon\to 0}E\left[1\left(Y\leq u\right)|Z\in\mathcal{N}_\varepsilon, D=0\right]. \end{aligned}$$

Based on Theorem 2, straightforward estimators of the distribution functions are obtained as

$$\begin{aligned} \hat{F}_{Y^1|c}\left(u\right) &= \frac{\sum_{i:D_i=1}1\left(Y_i\leq u\right)\left(2\cdot 1\left(Z_i\geq z_0\right)-1\right)K\left(\frac{Z_i-z_0}{h}\right)}{\sum_{i:D_i=1}\left(2\cdot 1\left(Z_i\geq z_0\right)-1\right)K\left(\frac{Z_i-z_0}{h}\right)}, \\ \hat{F}_{Y^0|c}\left(u\right) &= \frac{\sum_{i:D_i=0}1\left(Y_i\leq u\right)\left(2\cdot 1\left(Z_i\geq z_0\right)-1\right)K\left(\frac{Z_i-z_0}{h}\right)}{\sum_{i:D_i=0}\left(2\cdot 1\left(Z_i\geq z_0\right)-1\right)K\left(\frac{Z_i-z_0}{h}\right)}. \end{aligned}$$

The estimated distribution function is well-behaved for all types of outcome variables while the quantiles identified in Theorem 1 will be well-behaved only when Y is continuously distributed. Therefore, Theorem 2 is interesting for discrete or mixed outcome variables. Furthermore, since we have a closed-form solution for the distribution function, its estimation may be a first step towards the estimation of the QTE. The negativity of some of the weights, however, implies that the estimated distribution function will not be monotonously increasing in finite samples. This problem can be solved by monotonizing the estimated distribution function using the method of Chernozhukov, Fernandez-Val, and Galichon (2007) and finally inverting it to obtain the quantiles.

3 QTE with estimated threshold probability

In the preceding section we have used the fact that $\lim_{\varepsilon \to 0} \Pr\left(Z \geq z_0 | Z \in \mathcal{N}_\varepsilon\right) = \frac{1}{2}$. This result follows by differentiability of F_Z at z_0 and led to the very simple formulae of Theorems 1 and 2. In small samples, however, we may not have very many data points available at z_0 and therefore have to rely on a larger smoothing window. In this case, the number of data points could be asymmetric around z_0, and we could obtain more precise estimates by estimating the probability of being above the threshold (within the smoothing area). Define $p_\varepsilon = \Pr\left(Z \geq z_0 | Z \in \mathcal{N}_\varepsilon\right)$ for a given ε.

Theorem 3 (Quantiles of potential outcomes with estimated threshold probability) *Under Assumption 1, the quantiles of the potential outcomes for the local compliers are identified as*

$$\begin{aligned} Q^\tau_{Y^1|c} &= \lim_{\varepsilon \to 0} \arg\min_q E\left[\rho_\tau\left(Y - q\right) \frac{I^+ - p_\varepsilon}{p_\varepsilon\left(1 - p_\varepsilon\right)} | Z \in \mathcal{N}_\varepsilon, D = 1\right] \\ Q^\tau_{Y^0|c} &= \lim_{\varepsilon \to 0} \arg\min_q E\left[\rho_\tau\left(Y - q\right) \frac{I^+ - p_\varepsilon}{p_\varepsilon\left(1 - p_\varepsilon\right)} | Z \in \mathcal{N}_\varepsilon, D = 0\right]. \end{aligned}$$

This result simplifies to Theorem 1 by using that $\lim_{\varepsilon \to 0} p_\varepsilon = \frac{1}{2}$. The representation of Theorem 3, nevertheless, suggests a different estimation strategy where one plugs in an estimate of $\Pr\left(Z \geq z_0 | Z \in \mathcal{N}_\varepsilon\right)$ instead of the value $\frac{1}{2}$. Using the estimated p_ε often performed better in Monte Carlo simulations in small samples than when using $\lim_{\varepsilon \to 0} p_\varepsilon = \frac{1}{2}$. In some sense this result appears to be related to the well-known result in the propensity score matching literature that estimators which use the estimated propensity score are more efficient than estimators that use the true propensity score. This result might not be directly transferable here, though, since we are in a nonparametric context.

For completeness, we give the identification results for the distribution function.

Theorem 4 (Distribution of potential outcomes with estimated threshold probability) *Under Assumption 1, the distribution of the potential outcomes for the local compliers are*

identified as

$$
\begin{aligned}
F_{Y^1|c}(u) &= \lim_{\varepsilon \to 0} \frac{E\left[1\left(Y \leq u\right) \frac{I^+ - p_\varepsilon}{p_\varepsilon(1-p_\varepsilon)} | Z \in \mathcal{N}_\varepsilon, D = 1\right]}{E\left[\frac{I^+ - p_\varepsilon}{p_\varepsilon(1-p_\varepsilon)} | Z \in \mathcal{N}_\varepsilon, D = 1\right]} \\
F_{Y^0|c}(u) &= \lim_{\varepsilon \to 0} \frac{E\left[1\left(Y \leq u\right) \frac{I^+ - p_\varepsilon}{p_\varepsilon(1-p_\varepsilon)} | Z \in \mathcal{N}_\varepsilon, D = 0\right]}{E\left[\frac{I^+ - p_\varepsilon}{p_\varepsilon(1-p_\varepsilon)} | Z \in \mathcal{N}_\varepsilon, D = 0\right]}.
\end{aligned}
$$

4 QTE in RDD with covariates

In this section, we extend the regression discontinuity design to incorporate additional covariates X in a fully nonparametric way, and suppose that Assumption 1 holds conditionally on X.

There are several reasons why one might want to control for X. To mention a few: Covariates can help to eliminate small sample biases, especially if the number of observations close to the threshold z_0 is rather small such that one also has to include observations in the estimation process that are further apart. This point is emphasized particularly in Black, Galdo, and Smith (2005). We also permit that the density $f_{X|Z}$ is discontinuous at z_0. This can occur when the variable Z itself is confounded, e.g. in a situation of dynamic treatment assignment as in van der Klaauw (2008). It can also occur when different data collection schemes have been used for individuals above the threshold z_0 versus those below z_0, e.g. if those above z_0 have been hospitalized while those below z_0 received outpatient care with restricted follow-up data collection. Another reason for incorporating covariates applies when the threshold crossing at z_0 itself affects various X variables that one would like to control for. For example, Z may represent proximity to a state or regional border and crossing the border is associated with certain changes in laws or regulations that one is interested in. At the same time, a few other covariates may change in distribution as well, which one would like to control for. For a recent example see Brügger, Lalive, and Zweimüller (2008). As a final example for a discontinuity in $f_{X|Z}$ we consider the decomposition between direct and indirect effects of the treatment effect. X is here a post-treatment variable, and a change in treatment status D may have an effect on Y via X as well as a direct effect on Y. While RDD estimation without covariates estimates the total effect, in various situations one is interested in disentangling the direct from the indirect effect, which under certain conditions

can be done by controlling for X.[5]

We assume in the following that Assumption 1 holds conditionally on X. Theorems 1 to 4 now apply immediately to the treatment effect conditionally on X. In many situations we are however more interested in the unconditional effect, i.e. the effect on all local compliers irrespective of their value of X. There are at least three reasons why unconditional effects are interesting. First, for the purpose of evidence-based policy making a small number of summary measures can be more easily conveyed to the policy makers and the public than a large number of estimated effects for each and every value of X. Second, unconditional effects can be estimated more precisely than conditional effects. Third, the definition of the unconditional effects does not depend on the variables included in X.[6] One can therefore consider different sets of control variables X and still estimate the same object, which is useful for examining robustness of the results to the set of control variables.

The following results identify the unconditional effects, which are obtained by first conditioning on X and thereafter integrating with respect to X. For identification we need a common support restriction with respect to X and we also assume the existence of a density. (At the expense of more complex notation we could also easily permit discrete X.)

Assumption 2: Assume Assumption 1(i), (ii), (v) and Assumption 1(iii) and (iv) conditionally on X. Further assume:

- Common support $\lim_{\varepsilon \to 0} Supp(X|Z \in \mathcal{N}_\varepsilon^+) = \lim_{\varepsilon \to 0} Supp(X|Z \in \mathcal{N}_\varepsilon^-)$
- Density at threshold $\lim_{\varepsilon \to 0} F_{X|Z \in \mathcal{N}_\varepsilon^+}(x)$ and $\lim_{\varepsilon \to 0} F_{X|Z \in \mathcal{N}_\varepsilon^-}(x)$ exist and are differentiable in x at z_0 with pdf $f^+(x|z_0)$ and $f^-(x|z_0)$, respectively.

(Regarding notation: $f^+(x, z_0) = f^+(x|z_0)f(z_0)$ refers to the joint distribution of X and Z whereas $f^+(x|z_0)$ refers to the conditional distribution of X. Analogously for the limit from below.)

Note that we permit $f(x, z)$ to be continuous at z_0, i.e. $f^+(x|z_0) = f^-(x|z_0)$, or to be discontinuous, i.e. $f^+(x|z_0) \neq f^-(x|z_0)$.

With these additional assumptions, we can identify the quantile and cumulative distribution functions of the potential outcomes. The formulae, however, are not so neat as in Section 2.

[5] Even if $f_{X|Z}$ is not discontinuous at z_0, we conjecture that there may be efficiency gains by incorporating X. We will analyze this issue in more detail in future work.

[6] This, of course, is only true if X contains only pre-treatment variables.

Define $p_\varepsilon(x) = \Pr(Z \geq z_0 | X = x, Z \in \mathcal{N}_\varepsilon)$.

Theorem 5 (Quantiles of the potential outcomes) *Under Assumption 2, $Q^\tau_{Y^1|c}$ and $Q^\tau_{Y^0|c}$ are the solutions of the following optimization problem*

$$\begin{aligned} Q^\tau_{Y^1|c} &= \lim_{\varepsilon \to 0} \arg\min_q E\left[\rho_\tau(Y-q)\frac{I^+ - p_\varepsilon(X)}{p_\varepsilon(X)(1-p_\varepsilon(X))} | Z \in \mathcal{N}_\varepsilon, D=1\right] \\ Q^\tau_{Y^0|c} &= \lim_{\varepsilon \to 0} \arg\min_q E\left[\rho_\tau(Y-q)\frac{I^+ - p_\varepsilon(X)}{p_\varepsilon(X)(1-p_\varepsilon(X))} | Z \in \mathcal{N}_\varepsilon, D=0\right]. \end{aligned}$$

An analogous result is obtained for the distribution function.

Theorem 6 (Distribution of potential outcomes) *Under Assumption 2, the distribution of the potential outcomes for the local compliers are identified as*

$$\begin{aligned} F_{Y^1|c}(u) &= \lim_{\varepsilon \to 0} \frac{E\left[1(Y \leq u)\frac{I^+ - p_\varepsilon(X)}{p_\varepsilon(X)(1-p_\varepsilon(X))} | Z \in \mathcal{N}_\varepsilon, D=1\right]}{E\left[\frac{I^+ - p_\varepsilon(X)}{p_\varepsilon(X)(1-p_\varepsilon(X))} | Z \in \mathcal{N}_\varepsilon, D=1\right]} \\ F_{Y^0|c}(u) &= \lim_{\varepsilon \to 0} \frac{E\left[1(Y \leq u)\frac{I^+ - p_\varepsilon(X)}{p_\varepsilon(X)(1-p_\varepsilon(X))} | Z \in \mathcal{N}_\varepsilon, D=0\right]}{E\left[\frac{I^+ - p_\varepsilon(X)}{p_\varepsilon(X)(1-p_\varepsilon(X))} | Z \in \mathcal{N}_\varepsilon, D=0\right]}. \end{aligned}$$

References

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

Battistin, E., and E. Rettore (2002): "Testing for programme effects in a regression discontinuity design with imperfect compliance," *Journal of Royal Statistical Society Series A*, 165, 39–57.

——— (2008): "Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs," *Journal of Econometrics*, 142, 715–730.

Black, B., H. Jang, and W. Kim (2006): "Does Corporate Governance Predict Firms Market Values? Evidence from Korea," *Journal of Law, Economics and Organization*, 22, 366–413.

Black, D., J. Galdo, and J. Smith (2005): "Evaluating the regression discontinuity design using experimental data," *mimeo*, University of Michigan, USA.

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

Brügger, B., R. Lalive, and J. Zweimüller (2008): "Does culture affect unemployment? Evidence from the Barriere des Roestis," *mimeo, University of Zürich*, ., .

Buddelmeyer, H., and E. Skoufias (2003): "An evaluation of the performance of regression discontinuity design on PROGRESA," *IZA discussion paper*, 827.

Chay, K., and M. Greenstone (2005): "Does air quality matter? Evidence from the housing market," *Journal of Political Economy*, 113, 376–424.

Chay, K., P. McEwan, and M. Urquiola (2005): "The central role of noise in evaluating interventions that use test scores to rank schools," *American Economic Review*.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2007): "Quantile and probability curves without crossing," MIT working paper.

DiNardo, J., and D. Lee (2004): "Economic impacts of new unionization on private sector employers: 1984-2001," *Quarterly Journal of Economics*, 119, 1383–1441.

Forslund, A., and O. NordströmSkans (2006): "Swedish youth labour market policies revisited," *Viertelsjahreshefte zur Wirtschaftsforschung*, 75, 168–185.

Fredriksson, P., and B. Öckert (2006): "Is early learning really more productive? The effect of school starting age on school and labor market performance," *IFAU Discussion Paper 2006:12*.

Gormley, W., and D. Phillips (2005): "The effects of universal pre-k in Oklahoma: research highlights and policy implications," *The Policy Studies Journal*, 33, 65Ű81.

Guryan, J. (2001): "Does money matter? Regression-discontinuity estimates from education finance reform in Massachusetts," *NBER*, Working paper, 8269.

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

Imbens, G., and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

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

Jacob, B., and L. Lefgren (2004a): "The impact of teacher training on student achievement: quasi-experimental evidence from school reform efforts in Chicago," *Journal of Human Resources*, 39, 50–79.

——— (2004b): "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," *Review of Economics and Statistics*, 86, 226–244.

Lalive, R. (2008): "How do extended benefits affect unemployment duration? A regression discontinuity approach," *Journal of Econometrics*, 142, 785–806.

Lalive, R., J. Wüllrich, and J. Zweimüller (2008): "Do financial incentives for firms promote employment of disabled workers: a regression discontinuity approach," *mimeo, University of Zürich*, x, x.

Leuven, E., M. Lindahl, H. Oosterbeek, and D. Webbink (2007): "The effect of extra funding for disadvantaged pupils on achievement," *Review of Economics and Statistics*, 89, 721–736.

Matsudaira, J. (2008): "Mandatory summer school and student achievement," *Journal of Econometrics*, 142, 829–850.

NordströmSkans, O., and L. Lindqvist (2005): "Causal effects of subsidized carrer breaks," *IFAU Working paper*, 2005, 17.

Puhani, P., and A. Weber (2007): "Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany," *Empirical Economics*, 32, 359–386.

van der Klaauw, W. (2002): "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach," *International Economic Review*, 43, 1249–1287.

——— (2008): "Breaking the link between poverty and low student achievement: An evaluation of Title I," *Journal of Econometrics*, 142, 731–756.

Öckert, B. (2008): "What's the value of an acceptance letter? Using admissions data to estimate the return to college," mimeo, IFAU, Uppsala.