

Quantile treatment effects in the regression discontinuity design

Markus Frölich

Universität Mannheim and IZA (Bonn), Germany

Blaise Melly

Brown University, Providence, RI, USA

First version: December 2007

Last changes: March 12, 2010

Abstract:

This paper shows nonparametric identification of quantile treatment effects (QTE) in the regression discontinuity design. The distributional impacts of social programs such as welfare, education, training programs and unemployment insurance are of large interest to economists. QTE are an intuitive tool to characterize the effects of these interventions on the outcome distribution. We propose uniformly consistent estimators for both potential outcome distributions (treated and non-treated) for the population of interest as well as other function-valued effects of the policy including in particular the QTE process. The estimators are straightforward to implement and attain the optimal rate of convergence for one-dimensional nonparametric regression. We apply the proposed estimators to estimate the effects of summer school on the distribution of school grades, complementing the results of Jacob and Lefgren (2004).

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

JEL classification: C13, C14, C21

This paper completes and replaces the earlier IZA working paper 3638 that was circulated under the same title. We would like to thank Josh Angrist, Marine Carrasco, Ulrich Müller, Bernd Fitzenberger and seminar participants at Bonn IZA, Brown University, Freiburg, Georgetown, IFPRI Washington, Maryland, Montréal, Ohio State, Paris INSEE-Crest Malinvaud, Princeton, St.Gallen, Uppsala IFAU, Zürich, ESEM 2009 in Barcelona, the Cemmap Workshop on Quantile Regression 2009 and the Labor Economics Workshop at Engelberg for very useful comments that helped improve the paper. Markus Frölich, Universität Mannheim, IZA Bonn and ZEW Mannheim, froelich@uni-mannheim.de; Blaise Melly, Brown University, Department of Economics, Providence, RI, USA, blaise_melly@brown.edu.

1 Introduction

The regression discontinuity design (RDD) was first introduced by Thistlethwaite and Campbell (1960) as a quasi-experimental design for evaluating social programs and interventions. The prediction made by Campbell and Stanley (1963) that the RDD is “very limited in its range of applications (that are) mainly educational” has been proven wrong by the recent literature. The RDD has received tremendous attention in many fields, e.g. labor markets, political economy health, criminology, environment, development.¹

The RDD has attracted a more modest attention in theoretical econometrics. Hahn, Todd, and van der Klaauw (2001) describe this identification strategy using the treatment effect framework and formalize the assumptions required to identify causal effects. They also provide local linear estimators to non-parametrically estimate the mean treatment effect. Porter (2003) complements their work by considering alternative estimators. Lee and Card (2008) consider the case when the forcing variable is discrete. They interpret deviations of the true regression function from a given approximating function as specification errors and propose appropriately corrected standard errors. McCrary (2008) develops a test of the manipulation of the running variable related to the continuity of its density function. Frölich (2007) incorporates covariates in a fully nonparametric way and shows that efficiency gains are obtained and that the rate of convergence does not depend on the number of covariates. Imbens and Lemieux (2008), van der Klaauw (2008) and Lee and Lemieux (2009) have surveyed both the applied and theoretical literature on the RDD.

Despite this growing number of studies, the RDD has been used so far only to estimate average treatment effects. Yet, in many research areas, one is often not only interested in average impacts, but also in the distributional consequences of treatment interventions. In the field of education (e.g. Jacob and Lefgren (2004), Leuven, Lindahl, Oosterbeek, and Webbink (2007)), educational *inequality* e.g. in cognitive achievement is of large public interest. When examining the effects

¹For an incomplete list see e.g. Angrist and Lavy (1999), Battistin and Rettore (2002), Battistin and Rettore (2008), Black (1999), Black, Galdo, and Smith (2007), Black, Jang, and Kim (2006), Black, Smith, Berger, and Noel (2003), 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 Skans (2006), Imbens and Lemieux (2008), Gormley and Phillips (2005), Guryan (2001), Jepsen, Mueser, and Troske (2009), Lalive (2008), Lalive, Würlrich, 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.

of training (e.g. Black, Galdo, and Smith (2007)), 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.

In this paper we identify and estimate nonparametrically the treatment effects on the distribution of the outcome variable in the RDD.² We obtain uniformly consistent estimates for the distribution functions of the potential outcomes. If the dependent variable is continuous, quantile treatment effects (QTE) are a convenient way to report heterogeneous impacts of treatments on different points of an outcome distribution. Therefore, we show how quantile treatment effects can also be identified and estimated. More generally, we obtain uniformly consistent estimates for functionals of the distribution functions, e.g. for the Gini coefficient, the Lorenz curve or distribution treatment effects.³ Our estimators are based on the local-linear estimation of the distribution function. If we choose optimally the bandwidth, our estimators are consistent at the $n^{-\frac{2}{5}}$ rate, which is the optimal convergence rate for one-dimensional nonparametric estimation.

Even if one is not primarily interested in the distributional impacts or the impact on inequality, one may still use the method proposed to reduce susceptibility to outliers. Compared to the widely used mean RDD estimator, a *median* RDD estimator can provide more stable estimates when the outcome variable is noisy, e.g. wages or earnings. The quantiles are well-defined even if the outcome variable does not have finite moments due to fat tails. This is akin to the discussion on mean versus median regression, see e.g. Koenker and Bassett (1978) and Koenker (2005), who stress the robustness of median regression to outliers. This may be particularly relevant for the

²The identification of the distribution of the potential outcomes was first shown in section 4.3 of Frölich (2007). These results were extended in the IZA working paper 3638 that was released in August 2008. The present paper supplants this working paper and our previous results. Frandsen (2009) has proposed an estimator based on our identification strategy using local linear quantile regression. Frandsen's estimator is different from ours in many ways. However, it is inefficient compared to our estimator. It does not achieve the optimal convergence rate of $n^{-\frac{2}{5}}$. (See his Assumption A7.) Guiteras (2008) suggests an interesting alternative identification strategy based on Chernozhukov and Hansen (2005). His assumptions are neither more nor less general than ours. However, we view his assumptions as less natural in the RDD. His identification approach relies on a monotonicity assumption in the outcome equation, whereas we rely on a monotonicity assumption in the selection equation, which appears natural in the RDD. His approach is only applicable for a continuous outcome variable Y , whereas we can allow for Y to be discrete, continuous or mixed discrete-continuous, e.g. earnings with a mass point at zero.

³We proceed in this respect as in Chernozhukov, Fernández-Val, and Melly (2007). For a further discussion on inequality measures see also Firpo (2008).

RDD since the number of observations close to the discontinuity threshold is often relatively small. In many applications, estimated effects on higher-education or employment are often significant whereas effects on earnings or wages are insignificant, because of the large variance of the latter estimates, see e.g. Jepsen, Mueser, and Troske (2009).

The results in this paper are related to previous efforts to estimate quantile treatment effects. Koenker and Bassett (1978) propose a parametric estimator for conditional quantile functions while Chaudhuri (1991) suggests a nonparametric estimator. In the RDD, the QTEs could be estimated via nonparametrically weighted quantile regressions. However, we prefer to estimate first the distribution function because the weights can be negative in the RDD, which causes the objective function to be non-convex. In the fuzzy discontinuity design the proposed estimator can be interpreted as an instrumental variable estimator where the discontinuity is used as a binary instrument for the binary treatment. This is similar to the framework of Abadie, Angrist, and Imbens (2002) and Abadie (2003) with the added complication that we must control nonparametrically for the continuous running variable.

We illustrate how our estimators work in practice by applying the methodology to the data used by Jacob and Lefgren (2004). They exploit an administrative rule introduced by Chicago public schools in 1996 that tied summer school attendance to accomplishment on tests. Using the discontinuity implied by this rule they find positive mean effects of the summer school on later educational achievement. We complement their results by showing that the effects were clearly larger at the upper quantile than at the lower quantiles of the test score distribution. This shows that summer school is particularly effective for motivated students who have presently difficulties.

The rest of the paper is organized as follows. In section 2 identification is considered. Section 3 proposes estimators and derives their limiting distributions. We extend the previous results to the case where covariates are present in section 4. Section 5 applies the estimation and inference procedures to estimate distributional effects of summer school on educational attainment. Section 6 concludes with a summary of the results.

2 Identification of QTE in the RDD

We define causal effects using the potential-outcome notation in the framework known as the Neyman-Fisher-Rubin causal model.⁴ We are interested in the effect of a binary treatment D on an outcome variable Y . We observe n units, indexed by $i = 1, \dots, n$, which are drawn randomly and independently from a large population. Let Y_i^1 and Y_i^0 be the potential outcomes of individual i . Hence, Y_i^1 would be realized if individual i were to receive treatment and Y_i^0 would be realized otherwise. The potential outcomes as well as the treatment effects $Y_i^1 - Y_i^0$ are permitted to vary freely across individuals, i.e. no constant treatment effect is assumed.

We observe each unit only once and, therefore, observe the realized outcome $Y_i = Y_i^0 (1 - D_i) + Y_i^1 D_i$ but not both potential outcomes. The identification strategy of the causal treatment effects will exploit the presence of Z_i , a variable that influences D_i in a discontinuous way. In the example in Section 5, Z_i will be performance on a test before the summer break. Note that Z is not an instrument because it is allowed to have a direct effect on Y . However, this direct impact is assumed to be 'smooth' while the effect on D is assumed to be discontinuous at a known threshold z_0 .

In the literature, two different designs are often examined. In the *sharp* design D_i changes for everyone at z_0 ,

$$D_i = 1(Z_i \geq z_0). \quad (1)$$

In this sharp design, *all* individuals change programme participation status exactly at z_0 . In many applications, however, the treatment decision contains some element of discretion. Caseworkers 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. In this case, known as the *fuzzy* design, D is permitted to also depend on other (unobserved) factors but we assume that the treatment probability changes discontinuously at z_0 :

$$\lim_{\varepsilon \rightarrow 0} E[D|Z = z_0 + \varepsilon] - \lim_{\varepsilon \rightarrow 0} E[D|Z = z_0 - \varepsilon] \neq 0. \quad (2)$$

The 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.⁵

⁴See Neyman (1935), Fisher (1935) and Rubin (1978).

⁵Battistin and Rettore (2008) introduce the mixed sharp fuzzy design as a special case of the fuzzy design.

In addition to the discontinuity (2), which is in fact a testable assumption, for identification it is further required that the conditional means of $E[Y^0|Z]$ and $E[Y^1|Z]$ are continuous at z_0 . With these two assumptions, Z can act as an instrumental variable locally in a neighbourhood about z_0 .⁶ Since we allow for heterogenous treatment effects, identification requires further a monotonicity condition similar to that of Imbens and Angrist (1994). We identify the causal effects of D on the distribution of Y for the local compliers. Note that in the sharp design the monotonicity assumption is automatically satisfied and everyone is a complier.

For stating the identification results, it is helpful to introduce more precise notation first. Let $\mathcal{N}_\varepsilon$ be a symmetric ε neighborhood 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 partition the population into five subpopulations:

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

The definition of these subpopulations is a direct extension of the concept of Imbens and Angrist (1994). The first group contains those units that will *always* be treated (if $Z \in \mathcal{N}_\varepsilon$), the second contains those that will *never* be treated (if $Z \in \mathcal{N}_\varepsilon$), 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 react non-monotonously over the $\mathcal{N}_\varepsilon$ neighbourhood, e.g. they may first switch from $D = 0$ to 1 and then back for increasing values of z . We will assume that the last two groups have measure zero for ε sufficiently small. Note that in the sharp design, everyone is a complier for any $\varepsilon > 0$.

Assumption I: There exists some (arbitrary) positive $\bar{\varepsilon}$ such that for every positive $\varepsilon \leq \bar{\varepsilon}$

⁶Hahn, Todd, and van der Klaauw (2001) consider also as an alternative a kind of selection on observables assumption. They assume that the treatment effect is independent of D conditional on Z being near z_0 . We focus on the IV type approach because it seems to dominate the literature. In the sharp design, both assumptions are equivalent. In any case, in both designs the same estimator is obtained whichever of the two assumptions is invoked.

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

This assumption requires that for every small-enough neighbourhood, the threshold acts like a local instrumental variable. Assumption I (i) requires that $E[D|Z]$ is in fact discontinuous at z_0 , i.e. we assume that some units change their treatment status exactly at z_0 . Assumption I (ii) requires that, in a very small neighborhood of z_0 , the instrument has a weakly monotonous impact on $D_i(z)$. Assumptions I (iii) and (iv) impose the continuity of the type and the distribution of the potential outcomes as a function of Z at z_0 .⁷ Assumption I (v) requires that observations close to z_0 exist.

Under this assumption the distribution functions of the potential outcomes for local compliers are identified. Define $F_{Y^d|c}(u) = \lim_{\varepsilon \rightarrow 0} F_{Y^d|Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon=c}(u)$. Define also $I^+ = 1(Z \geq z_0)$ and $I^- = 1 - I^+$.

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

$$F_{Y^1|c}(u) = \lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq u)(I^+ - p_\varepsilon) | Z \in \mathcal{N}_\varepsilon, D = 1]}{E[I^+ - p_\varepsilon | Z \in \mathcal{N}_\varepsilon, D = 1]}$$

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

where $p_\varepsilon = \Pr(Z \geq z_0 | Z \in \mathcal{N}_\varepsilon)$ for $\varepsilon > 0$.

Hence, the distribution functions can be estimated by local regression in a neighbourhood of z_0 . Details of the estimator will be discussed in the next section.

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

$$F_{Y^1}(u) = \lim_{\varepsilon \rightarrow 0} E[1(Y \leq u) | Z \in \mathcal{N}_\varepsilon, D = 1]$$

$$F_{Y^0}(u) = \lim_{\varepsilon \rightarrow 0} E[1(Y \leq u) | Z \in \mathcal{N}_\varepsilon, D = 0].$$

⁷This is slightly weaker than assuming $(Y^1, Y^0, \mathcal{T}_\varepsilon) \perp\!\!\!\perp 1(Z \geq z_0) | Z \in \mathcal{N}_\varepsilon$ for $\varepsilon \rightarrow 0$.

The previous theorem showed that the distribution functions are identified irrespective of whether the outcome variable Y is continuous or discrete. Identification of the distribution function also implies the identification of all functionals of the marginal distribution function of the outcome with and without the treatment. We provide detailed results for a popular functional, the quantile treatment effect (QTE) process, but similar results may be obtained for other functionals such as the variance, the Gini coefficient, the Lorenz curve, the Theil index or the interquartile or interdecile range. These results can also be used to develop tests of (first or second order) stochastic dominance.

When Y is continuously distributed, we are often interested in the quantiles of the potential outcomes and in particular the quantile treatment effect. Suppose that Y^1, Y^0 are continuously distributed which implies that the distribution functions are monotonously increasing and invertible. Let $Q_Y^\tau = \inf\{u : F_Y(u) \geq \tau\}$ be the quantile operator. Define $Q_{Y^d|c}^\tau = \lim_{\varepsilon \rightarrow 0} Q_{Y^d|Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c}^\tau$ as the τ^{th} quantile of Y^d for the local compliers. The quantile treatment effect (QTE) for the compliers is then defined as

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

We could identify the quantiles by first estimating the entire distribution functions, using the results of Theorem 1, and thereafter inverting them. A more direct approach to identify and estimate QTE consists in using the representation of conditional quantiles as the solution to a minimization problem, as in Koenker and Bassett (1978). The following result gives a direct representation of the quantiles and of the Δ_{QTE}^τ . The expressions are obtained via the first order conditions of a minimization/maximization problem.

Theorem 2 (Quantiles of the potential outcomes) *Under Assumption I and assuming that Y^0 and Y^1 are continuously distributed, the quantiles of the potential outcomes for the local compliers are identified as*

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

where $\rho_\tau(u) = u \cdot \{\tau - 1(u < 0)\}$ is the check function and $p_\varepsilon = \Pr(Z \geq z_0 | Z \in \mathcal{N}_\varepsilon)$ for $\varepsilon > 0$.

Alternatively, we could identify the QTE Δ_{QTE}^τ directly as

$$\left(Q_{Y^0|c}^\tau, \Delta_{QTE}^\tau\right) = \lim_{\varepsilon \rightarrow 0} \arg \min_{a,b} E \left[\rho_\tau(Y - a - bD) (I^+ - p_\varepsilon) (2D - 1) | Z \in \mathcal{N}_\varepsilon \right], \quad (3)$$

which corresponds to a local weighted bivariate quantile regression.

Hence, the quantiles can be obtained by simple univariate weighted quantile regressions. 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. Conventional linear programming algorithms therefore will not work. Implementation via estimation of the cdf, as identified by Theorem 1, with subsequent inversion to obtain the quantiles is more convenient in practice. This is the way how we implement the estimator in Section 3.

Before we discuss estimation of the cdf and the QTE in more detail, we note that the previous expressions can be simplified considerably. By Assumption I (v) and the symmetry of $\mathcal{N}_\varepsilon$ it follows by l'Hospital that $\lim_{\varepsilon \rightarrow 0} p_\varepsilon = \lim_{\varepsilon \rightarrow 0} \Pr(Z \geq z_0 | Z \in \mathcal{N}_\varepsilon) = \frac{1}{2}$. Therefore we can identify the distribution and the quantiles of the potential outcomes as

Corollary 3 (Distribution of potential outcomes) *Under Assumption I, the distribution of the potential outcomes for the local compliers are identified as*

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

Corollary 4 (Quantiles of potential outcomes) *Under Assumption I and assuming that Y^0 and Y^1 are continuously distributed, the quantiles of the potential outcomes for the local compliers are identified as the solution of the following optimization problem*

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

In the following sections, however, we will base our estimators on Theorems 1 and 2 instead of Corollaries 3 and 4 because using an estimated p_ε often performed better in Monte Carlo simulations in small samples than using the fact that $\lim_{\varepsilon \rightarrow 0} p_\varepsilon = \frac{1}{2}$.⁸

⁸In small samples, we may not have very many data points close to z_0 available and therefore have to rely on a

3 Estimators and asymptotic properties

In this section we suggest estimators for the distribution functions of the potential outcomes $F_{Y^1|c}$ and $F_{Y^0|c}$ and for the quantile treatment effect process Δ_{QTE}^τ . We also derive asymptotic results for these processes.

We first examine the estimator of the distribution function of the potential outcomes, which has the advantage that it applies naturally irrespective of whether Y is continuous or discrete. For example, when one considers earnings, there is usually a substantial masspoint at zero, which does not pose any problems when estimating the cdf.

A natural estimator of $F_{Y^1|c}$ following from Theorem 1 can be motivated as follows. Let K_i be some kernel weights depending on the distance between Z_i and z_0 and a bandwidth converging to zero. We permit the weights to be local linear regression weights and will be more specific further below. A natural estimator of $F_{Y^1|c}(u)$ is thus

$$\hat{F}_{Y^1|c}(u) = \frac{\sum 1(Y_i \leq u) D_i (I_i^+ - \hat{p}_\varepsilon) K_i}{\sum D_i (I_i^+ - \hat{p}_\varepsilon) K_i}$$

where a natural estimator of p_ε is $\sum I_i^+ K_i / \sum K_i$. We thus obtain after a few calculations

$$\begin{aligned} & \frac{\sum_{i:I_i^+=1} 1(Y_i \leq u) D_i K_i}{\sum_{i:I_i^+=1} K_i} - \frac{\sum_{i:I_i^+=0} 1(Y_i \leq u) D_i K_i}{\sum_{i:I_i^+=0} K_i} \\ &= \frac{\frac{\sum_{i:I_i^+=1} D_i K_i}{\sum_{i:I_i^+=1} K_i} - \frac{\sum_{i:I_i^+=0} D_i K_i}{\sum_{i:I_i^+=0} K_i}}{\frac{\sum_{i:I_i^+=1} K_i}{\sum_{i:I_i^+=1} K_i} - \frac{\sum_{i:I_i^+=0} K_i}{\sum_{i:I_i^+=0} K_i}}. \end{aligned}$$

Define for a general random variable W the following right limit function $m_W^+ = \lim_{\varepsilon \rightarrow 0} E[W|Z = z_0 + \varepsilon]$ and the corresponding left limit function $m_W^- = \lim_{\varepsilon \rightarrow 0} E[W|Z = z_0 - \varepsilon]$. The variable W will represent at different places $1(Y \leq u) \cdot D$ or $1(Y \leq u) \cdot (1 - D)$ or $(1 - D)$ or D . Note that in every case W has bounded support and the previously defined limit functions are therefore bounded. We can write the suggested estimator as

$$\hat{F}_{Y^1|c}(u) = \frac{\hat{m}_{1(Y \leq u)D}^+ - \hat{m}_{1(Y \leq u)D}^-}{\hat{m}_D^+ - \hat{m}_D^-}$$

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 p_ε for a given ε . 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. Those results, however, are not directly transferable here, since we are in a nonparametric context. We plan to discuss this issue in future work and will consider only estimators that use an estimated p_ε in this paper.

and similarly for the non-treated outcome

$$\hat{F}_{Y^0|c}(u) = \frac{\hat{m}_{1(Y \leq u)(1-D)}^+ - \hat{m}_{1(Y \leq u)(1-D)}^-}{\hat{m}_{1-D}^+ - \hat{m}_{1-D}^-}.$$

If we use local linear weights, which appears appropriate here since we are effectively estimating conditional means at boundary points (from the left and right of z_0), m_W^+ is estimated as the value of a that solves

$$\arg \min_{a,b} \sum_{i=1}^n (W_i - a - b(Z_i - z_0))^2 \cdot I_i^+ K\left(\frac{Z_i - z_0}{h}\right).$$

Analogously m_W^- is estimated by using only observations to the left of z_0 .

Before we can state the main asymptotic results, we have to define precisely the regularity conditions that we assume.

Assumption R

- (i) IID sampling: The data $\{(Y_i, D_i, Z_i)\}$ are iid
- (ii) Smoothness and limit conditions: The left and right limits of the functions $E[1(Y \leq u) | Z, D = 0]$, $E[1(Y \leq u) | Z, D = 1]$ and $E[D | Z]$ exist at z_0 and these functions are twice continuously differentiable with respect to Z at z_0 with second derivative Hölder continuous in a left and a right ε -neighborhood of z_0 , respectively, and uniformly on $y \in \mathcal{Y}$, where $\mathcal{Y}$ is a compact subset of $\mathbb{R}$.
- (iii) Density: The density f_Z is bounded away from zero and is twice continuously differentiable at z_0 with second derivative Hölder continuous in a ε -neighborhood of z_0 .
- (iv) Compliers: The fraction of compliers $P_c = m_D^+ - m_D^-$ is bounded away from zero.
- (v) Bandwidth conditions: $nh \rightarrow \infty$ and $\sqrt{nh}h^2 \rightarrow \Xi < \infty$.
- (vi) Kernel: K is symmetric, bounded, zero outside a compact set and integrates to one.

Assumption R (i) is made mainly for simplicity and because it is satisfied in most applications of the RDD. Assumption R (ii) assures sufficient smoothness on both sides of the threshold. It guarantees that the distribution functions are equicontinuous in a neighborhood of z_0 . Note that assumption R (ii) implies also the existence of the left and right limits of all covariance functions since the random variables $1(Y \leq u)$ and D are all binary. Condition R (iv) is equivalent to assuming that we have a strong instrument in an IV problem. Condition R (v) balances bias and variance of the estimator. For $\Xi > 0$, squared bias and variance are of the same order. If we choose a bandwidth such that $\Xi = 0$, the bias vanishes asymptotically. To simplify the notation,

the same bandwidth is used for all functions on both sides of the threshold. The results can be adapted to allow for different bandwidths in a straightforward way as long as the convergence rates of the bandwidths are the same. A kernel function with compact support is assumed for convenience in the technical derivations.

Our main asymptotic result shows the joint weak convergence of $\sqrt{nh_n}(\hat{F}_{Y^1|c}(u) - F_{Y^1|c}(u))$ and $\sqrt{nh_n}(\hat{F}_{Y^0|c}(u) - F_{Y^0|c}(u))$ in the space $\ell^\infty(\mathcal{Y})$ of all bounded functions on $\mathcal{Y}$ equipped with the supremum norm.

Theorem 5 (Limit distribution for distribution functions) *If assumptions I and R are satisfied, the estimators $\hat{F}_{Y^0|c}(u)$ and $\hat{F}_{Y^1|c}(u)$ of the distribution functions for the compliers $F_{Y^0|c}(u)$ and $F_{Y^1|c}(u)$ jointly converge in law to the following Gaussian processes:*

$$\sqrt{nh_n} \left(\hat{F}_{Y^j|c}(u) - F_{Y^j|c}(u) \right) \Longrightarrow Z_j(u), \quad j \in \{0, 1\},$$

in $\ell^\infty(\mathcal{Y})$, where $u \mapsto Z_j(u)$ have mean functions

$$b_j(u) = \frac{\Xi}{P_c} C'_K \left\{ \frac{\partial^2 m_{1(Y \leq u)}^{+}(D+j-1)}{\partial z^2} - F_{Y^j|c}(u) \frac{\partial^2 m_D^{+}}{\partial z^2} - \frac{\partial^2 m_{1(Y \leq u)}^{-}(D+j-1)}{\partial z^2} + F_{Y^j|c}(u) \frac{\partial^2 m_D^{-}}{\partial z^2} \right\}$$

where C'_K is a constant that depends on the kernel function,⁹ $\frac{\partial^2 m_W^{+}}{\partial z^2} = \lim_{\varepsilon \rightarrow 0} \frac{\partial^2 E[W|Z=z_0+\varepsilon]}{\partial z^2}$ for a general random variable W and $\frac{\partial^2 m_W^{-}}{\partial z^2}$ is the similar left limit function. The covariance functions are, for $j, k \in \{0, 1\}$,

$$v_{j,k}(u, \tilde{u}) = C_K \frac{1}{P_c^2 f_Z(z_0)} \left(\omega_{j,k}^{+}(u, \tilde{u}) + \omega_{j,k}^{-}(u, \tilde{u}) \right)$$

where C_K is a constant that depends on the kernel function,¹⁰

$$\omega_{j,k}^{+}(u, \tilde{u}) = \lim_{\varepsilon \rightarrow 0} \text{Cov} \left\{ (D+j-1) (1(Y \leq u) - F_{Y^j|c}(u)), (D+k-1) (1(Y \leq \tilde{u}) - F_{Y^k|c}(\tilde{u})) \mid Z \in \mathcal{N}_\varepsilon^{+} \right\}$$

and $\omega_{j,k}^{-}(y, \tilde{y})$ is the analogous left limit.

A simple corollary of Theorem 5 is that the estimators of the distribution functions evaluated at a particular value $u \in \mathcal{Y}$ are asymptotically jointly normally distributed

$$\sqrt{nh_n} \left(\hat{F}_{Y^j|c}(u) - F_{Y^j|c}(u) \right) \sim N(b_j(u), v_{j,j}(u, u)), \quad j \in \{0, 1\}.$$

⁹The exact formula can be found in the appendix. $C'_K = \frac{11}{190}$ for the Epanechnikov kernel function.

¹⁰The exact formula can be found in the appendix. $C_K = \frac{56832}{12635}$ for the Epanechnikov kernel function.

This result can be extended to any finite collection of $u_k \in \mathcal{Y}$, $k = 1, \dots, K$.

Note that the bias functions $b_j(u)$ disappear if we choose $\Xi = 0$. This choice of the bandwidth implies that we undersmooth the functions to be estimated. This has the obvious advantage of simplifying the asymptotic inference, but may provide less accurate inference in finite samples. The asymptotic covariances are the sum of the covariances of the estimated functions rescaled by $P_c^2 f_Z(z_0)$.

A possible way to characterize the effect of the treatment on the outcome Y consists in estimating the distribution treatment effect (DTE) defined as $\Delta_{DTE}^u = F_{Y^1|c}(u) - F_{Y^0|c}(u)$. We naturally estimate this function by $\hat{\Delta}_{DTE}^u = \hat{F}_{Y^1|c}(u) - \hat{F}_{Y^0|c}(u)$. Corollary 6 gives the limiting distribution of $\hat{\Delta}_{DTE}^u$.

Corollary 6 (Limit distribution for distribution treatment effects) *If assumptions I and R are satisfied, the estimator $\hat{\Delta}_{DTE}^u = \hat{F}_{Y^1|c}(u) - \hat{F}_{Y^0|c}(u)$ of the distribution treatment effects for the compliers Δ_{DTE}^u converge to the following Gaussian process*

$$\sqrt{nh_n} \left(\hat{\Delta}_{DTE}^u - \Delta_{DTE}^u \right) \Rightarrow Z^1(u) - Z^0(u)$$

in $\ell^\infty(\mathcal{Y})$ with mean function $b_1(u) - b_0(u)$ and covariance function $v_{1,1}(u, \tilde{u}) + v_{0,0}(u, \tilde{u}) - 2v_{0,1}(u, \tilde{u})$.

As indicated by the title of this paper, the main focus is not the distribution treatment effects but the quantile treatment effects, which we find more intuitive. A disadvantage of considering quantiles is that they have a well-defined asymptotic distribution only if the outcome is continuous. We therefore make the additional Assumption Q from now on.

Assumption Q: $F_{Y^0|c}(u)$ and $F_{Y^1|c}(u)$ are both continuously differentiable with continuous density functions $f_{Y^0|c}(u)$ and $f_{Y^1|c}(u)$ that are bounded above and away from zero on $\mathcal{Y}$.

We could estimate the quantile treatment effects by the sample analog of Theorem 2. This minimization problem is, however, a non-convex optimization problem because some weights are positive while others are negative. This requires grid searches or algorithms for nonconvex problems that do not guarantee a global optimum.¹¹ Therefore, we follow a more direct strategy

¹¹Abadie, Angrist, and Imbens (2002) encountered a similar problem and they proposed to convexify the problem by using the projection of the weights on the space spanned by D , X and Y . This requires an additional nonparametric regression and it is unclear if this additional step will preserve the asymptotic distribution of the estimator. Our setup is simpler as we would not need to project on X .

here and invert the estimated distribution function. Another consequence of having negative and positive weights is that the estimated distribution function is non-monotone: $\hat{F}_{Y^j|c}(u)$ may decrease when we increase u . Of course, this is only a small sample problem because the assumed monotonicity ensures that the estimated distribution function is asymptotically strictly increasing. Even being only a small sample problem, this is nevertheless disadvantageous if one wants to invert $\hat{F}_{Y^j|c}(u)$. We follow here the suggestion of Chernozhukov, Fernandez-Val, and Galichon (2007) and monotone the estimated distribution functions by re-arrangements. This does not affect the asymptotic properties of the estimator but allows us to invert it. This procedure consists of a sequence of closed-form steps and is therefore very quick.

We now derive the limiting distribution of the quantile functions and of other functionals of the distribution functions via the functional delta method (see chapter 3.9 in van der Vaart and Wellner (1996)). This requires establishing the Hadamard differentiability of the functionals, which is well known in the case of the quantile functions.¹²

Theorem 7 (Limit distribution for quantile functions) *If assumptions I, R and Q are satisfied, the estimators $\hat{Q}_{Y^0|c}(\tau)$ and $\hat{Q}_{Y^1|c}(\tau)$ of the quantile functions for the compliers $Q_{Y^0|c}(\tau)$ and $Q_{Y^1|c}(\tau)$ jointly converge to the following Gaussian processes:*

$$\sqrt{nh_n} \left(\hat{Q}_{Y^j|c}(\tau) - Q_{Y^j|c}(\tau) \right) \Rightarrow -f_{Y^j|c}(Q_{Y^j|c}(\tau))^{-1} Z_j(Q_{Y^j|c}(\tau)) := W^j(\tau), \quad j \in \{0, 1\}$$

in $\ell^\infty((0, 1))$ with mean function $b_j^q(\tau) = -f_{Y^j|c}(Q_{Y^j|c}(\tau))^{-1} b_j(Q_{Y^j|c}(\tau))$ and covariance function $v_{j,k}^q(\tau, \tilde{\tau}) = f_{Y^j|c}(Q_{Y^j|c}(\tau))^{-1} f_{Y^k|c}(Q_{Y^k|c}(\tilde{\tau}))^{-1} v_{j,k}(Q_{Y^j|c}(\tau), Q_{Y^k|c}(\tilde{\tau}))$.

The straightforward Corollary 8 gives the limiting distribution of the quantile treatment effect process. As explained above, this implies that $\sqrt{nh_n} \left(\hat{\Delta}_{QTE}^{\tau_k} - \Delta_{QTE}^{\tau_k} \right)$ is jointly normally distributed for any finite sequence $\tau_k \in (0, 1)$, $k = 1, \dots, K$.

Corollary 8 (Limit distribution for quantile treatment effects) *If assumptions I, R and Q are satisfied, the estimator $\hat{\Delta}_{QTE}^\tau$ of the QTE for the compliers Δ_{QTE}^τ converges to the following Gaussian process*

$$\sqrt{nh_n} \left(\hat{\Delta}_{QTE}^\tau - \Delta_{QTE}^\tau \right) \Rightarrow W^1(\tau) - W^0(\tau)$$

¹²We follow the same strategy as Chernozhukov, Fernández-Val, and Melly (2007).

in $\ell^\infty((0, 1))$ with mean function $b_1^q(\tau) - b_0^q(\tau)$ and covariance function $v_{1,1}^q(\tau, \tilde{\tau}) + v_{0,0}^q(\tau, \tilde{\tau}) - 2v_{0,1}^q(\tau, \tilde{\tau})$.

Our last main result shows that various smooth functionals of both distribution functions satisfy a functional central limit theorem.

Theorem 9 (Limit distribution for smooth functionals) *Let $\phi(u, F_{Y^0|c}, F_{Y^1|c})$ be a functional taking value in $\ell^\infty(\mathcal{Y})$ that is Hadamard differentiable in $(F_{Y^0|c}, F_{Y^1|c})$ tangentially to the set of continuous functions with derivative (ϕ'_0, ϕ'_1) . If assumptions I and R are satisfied, the plug-in estimator $\hat{\phi}(u) \equiv \phi(u, \hat{F}_{Y^0|c}, \hat{F}_{Y^1|c})$ converges to the following Gaussian process:*

$$\sqrt{nh_n} \left(\hat{\phi}(u) - \phi(u) \right) \Rightarrow \phi'_0(u) Z_0(u) + \phi'_1(u) Z_1(u)$$

in $\ell^\infty((0, 1))$.

To conclude this section, we apply this powerful result to derive the limiting distribution of the Lorenz curve and the Gini coefficient, that will be examined in our application. The Lorenz curves and their estimates are defined for $j \in \{0, 1\}$ as

$$L^j(\tau) = \frac{\int_0^\tau Q_{Y^j|c}(t) dt}{\int_0^1 Q_{Y^j|c}(t) dt}, \quad \hat{L}^j(\tau) = \frac{\int_0^\tau \hat{Q}_{Y^j|c}(t) dt}{\int_0^1 \hat{Q}_{Y^j|c}(t) dt}.$$

The Hadamard derivative of the map from the distribution function to the Lorenz curve can be found e.g. in Barrett and Donald (2000). Using their result we obtain the following limiting distribution for the simple plug-in estimator:

$$\sqrt{nh_n} \left(\hat{L}^j(\tau) - L^j(\tau) \right) \Rightarrow \frac{\int_0^\tau W^1(t) dt}{\int_0^1 Q_{Y^1|c}(t) dt} - \frac{L^1(\tau)}{\int_0^1 Q_{Y^1|c}(t) dt} \int_0^1 W^1(t) dt \equiv \mathcal{L}(\tau)$$

in $\ell^\infty((0, 1))$ with mean function

$$b_j^l(\tau) = \frac{\int_0^\tau b_j^q(t) dt}{\int_0^1 Q_{Y^j|c}(t) dt} - \frac{L^j(\tau)}{\int_0^1 Q_{Y^j|c}(t) dt} \int_0^1 b_j^q(t) dt$$

and covariance function

$$\begin{aligned} v_{j,k}^l(\tau, \tilde{\tau}) &= \frac{1}{\int_0^1 Q_{Y^j|c}(t) dt \int_0^1 Q_{Y^k|c}(t) dt} \\ &\times \left(\int_0^\tau \int_0^{\tilde{\tau}} v_{j,k}^q(t, \tilde{t}) d\tilde{t} dt + L^j(\tau) L^k(\tilde{\tau}) \int_0^1 \int_0^1 v_{j,k}^q(t, \tilde{t}) d\tilde{t} dt \right. \\ &\quad \left. - L^j(\tau) \int_0^1 \int_0^{\tilde{\tau}} v_{j,k}^q(t, \tilde{t}) d\tilde{t} dt - L^k(\tilde{\tau}) \int_0^\tau \int_0^1 v_{j,k}^q(t, \tilde{t}) d\tilde{t} dt \right). \end{aligned}$$

The Gini coefficient is defined by

$$G^j = 1 - 2 \int_0^1 L^j(t) dt, \quad \hat{G}^j = 1 - 2 \int_0^1 \hat{L}^j(t) dt$$

Our simple plug-in estimator is asymptotically normally distributed with bias $-2 \int_0^1 b_j^l(t) dt$ and variance $4 \int_0^1 \int_0^1 v_{j,k}^l(t, \tilde{t}) d\tilde{t} dt$:

$$\sqrt{nh_n} \left(\hat{L}^j(\tau) - L^j(\tau) \right) \rightarrow 2 \int_0^1 \mathcal{L}(t) dt.$$

4 Extension: QTE in the 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 I holds conditionally on X . Frölich (2007) discusses in detail why one might want to control for X . Even if we believe that the RDD is valid without conditioning, we might want to check for the robustness of the results when we include covariates. In addition, Frölich (2007) shows that including covariates will increase the precision of the estimates.¹³ In other cases, the variable Z itself may be confounded, e.g. in a situation of dynamic treatment assignment as in van der Klaauw (2008). 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. Under certain conditions we can disentangle the direct from the indirect effects by controlling for X . See e.g. Brügger, Lalive, and Zweimüller (2008).

Whatever reasoning is used to justify the inclusion of covariates, we assume in the following that Assumption I holds *conditionally* on X . (Note that we permit that, when not conditioning on X , Assumption I may or may not be valid.) The identification results stated above 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 for all 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.¹⁴ Third, the definition of the unconditional effects does not depend on the

¹³This has been shown only for average effects. We will analyze this issue in more detail in future work for QTE.

¹⁴Unless one imposes parametric assumptions.

variables included in X .¹⁵ 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 .

Assumption C: Suppose Assumption I (i), (ii) and (v). Suppose further that Assumption I (iii) and (iv) are true *conditionally* on X . Further assume:

(vi) Common support $\lim_{\varepsilon \rightarrow 0} \text{Supp}(X|Z \in \mathcal{N}_\varepsilon^+) = \lim_{\varepsilon \rightarrow 0} \text{Supp}(X|Z \in \mathcal{N}_\varepsilon^-)$

With this assumption, we can identify the quantile and cumulative distribution functions of the potential outcomes for all compliers. Similar expressions as in Theorems 1 and 2 are obtained but the weights are now a function of $p_\varepsilon(x) = \Pr(Z \geq z_0|X = x, Z \in \mathcal{N}_\varepsilon)$. The following Theorem 10 shows the results for the QTEs.

Theorem 10 (Identification of QTEs in the presence of covariates) *Under Assumption C,*

$$\left(Q_{Y^0|C}^\tau, \Delta_{QTE}^\tau\right) = \lim_{\varepsilon \rightarrow 0} \arg \min_{a,b} E \left[\rho_\tau(Y - a - bD) \frac{I^+ - p_\varepsilon(X)}{p_\varepsilon(X)(1 - p_\varepsilon(X))} (2D - 1) | Z \in \mathcal{N}_\varepsilon \right].$$

This result shows that the unconditional QTE can be estimated via a simple weighted quantile regression where the covariates X only enter in the weights via $p_\varepsilon(x)$.¹⁶¹⁷ Again, the weights in the previous expression are sometimes positive and sometimes negative such that conventional linear programming algorithms fail because of the non-convexity. In the application we will therefore again proceed by first estimating the entire distribution functions and thereafter obtain the quantiles via inversion. The results for the distribution functions are given in the appendix.

¹⁵This, of course, is only true if X contains only pre-treatment variables.

¹⁶Note that these weights are similar to the weights in Theorems 3.1 b and c of Abadie (2003). The weights used here, however, are localized with respect to z_0 . The nonparametric setup of the RDD requires smoothing with respect to Z . Therefore none of the conditions of Abadie (2003) is satisfied in our local nonparametric framework. This also implies that we cannot obtain a parametric rate of convergence as discussed before.

¹⁷Hirano, Imbens, and Ridder (2003) also use a weighting approach in a setup where treatment is unconfounded given X .

5 Application: effects of summer school attendance

In this section, we apply the methods proposed earlier to estimate the effects of a summer school program for weak students on academic achievement. We use administrative data from the Chicago Public School system, taken from Jacob and Lefgren (2004).¹⁸ We use the same identification strategy as in this influential article but consider the effects of the summer school on the entire distribution of test scores. While Jacob and Lefgren (2004) acknowledge the interest in quantile treatment effects (p. 233), they estimate only average effects due to the absence, at that time, of suitable methods for quantiles and distributions.

The regression discontinuity design follows from an administrative rule that tied summer school attendance to the performance on a standardized test. In 1996, the Chicago Public Schools required students in third, sixth and eighth grade to take a mathematics and a reading exam in June. Students who did not meet the standards at both tests were required to attend a six-weeks summer school. This discontinuity in the rule can be used to identify the effect of the summer school attendance on similar tests one and two years after the summer school. More details can be found in the original article Jacob and Lefgren (2004).

We focus on students for whom the reading cutoff was binding (that is, students who passed math) because many more students failed in reading than in mathematics. For this sample, the reading test score is the unique running variable and this situation corresponds exactly to our framework. With two tests (math and reading), two grades (third and sixth) and two periods (one and two years after the summer school), we could consider eight outcome variables. Due to space limitations, we provide detailed results for the math test score one year after the summer school for the third graders. We focus on this outcome for two reasons. First, the treatment effects are estimated most precisely for this outcome variable in Jacob and Lefgren (2004). Second, the support of this outcome variable comprises so many values that it can reasonably be treated as a continuous variable. The results for other outcomes will be briefly sketched at the end of this section.

The administrative rule suggests a sharp discontinuity. However, some course waivers were given and a small percentage of students who scored above the threshold were observed participating. This fuzziness of the selection rule does not preclude the identification of the treatment effects for the subpopulation of compliers as discussed in Section 2. Figure 1 shows the relationship

¹⁸We thank Brian Jacob and Lars Lefgren for supporting us in accessing their data.

between the reading test score (the running variable Z) and the treatment probability. The discontinuity is extremely clear and compliers represent about 90% of the population at the threshold.

Figure 1 shows also three quantiles of the outcome variable as a function of the running variable Z . If the summer school attendance had no effect, we would not observe any discontinuity in the relationship between the running variable Z and the outcome at the threshold. Yet, we can recognize a drop in the conditional 95th percentile (-3.5), median (-2.8) and 5th percentile (-1.3) exactly at the threshold. While these drops suggest that the treatment had a positive effect, the increasing size of the drops suggests that the effects were stronger at the top of the distribution. Note, however, that the differences between the quantiles above and below the threshold do not estimate the quantile treatment effects in the fuzzy discontinuity design. Therefore, we present now the results obtained using the consistent estimators suggested in this paper.

The implementation of the estimators requires choosing the bandwidth and the kernel function. The choice of the kernel function is typically considered as being not very crucial. We have used the Epanechnikov kernel but the robustness checks with the uniform kernel (as in Jacob and Lefgren (2004)) do not reveal any sensitivity. For the bandwidth, we use the same bandwidth as in the original article.¹⁹ This can be considered as a conservative choice since one-sided cross validation gives larger bandwidths. When using the larger bandwidths, the estimated quantile treatment effects are similar but estimated with a smaller variance. From now on, we will use as outcome variable the math test *achievement gains* instead of the math tests themselves in order to be more in line with the analysis of Jacob and Lefgren (2004). This does not affect the results very much since we estimate the effects only for the population with reading score in a neighbourhood of z_0 , where also the pre-treatment math score is balanced between left and right.

Figure 2 plots the estimated distribution functions of the potential outcomes $\hat{F}_{Y^1|c}$ and $\hat{F}_{Y^0|c}$ together with pointwise 95% confidence intervals based on Theorem 5. At the lower end of the distribution, $\hat{F}_{Y^1|c}$ and $\hat{F}_{Y^0|c}$ are rather similar, but they diverge at the upper end of the distribution. The two confidence intervals do not overlap over a large part of the distribution.

Figure 2 also reveals that the distribution of the outcome variable can be considered as reasonably smooth. This allows inverting the distribution functions and estimating the quantile treat-

¹⁹Actually, Jacob and Lefgren (2004) use a different bandwidth on the left (2) and on the right (4). We use the same bandwidth (2) on both sides.

ment effect process. Figure 3 plots the estimated quantile treatment effects with pointwise 95% confidence intervals. This confirms the visual impression obtained from the distribution functions of Figure 2: the effect of summer school attendance is small or perhaps even negative at the bottom and increases monotonically when we move up the distribution. Note that the QTEs estimated at different quantiles are positively correlated. This means that the differences between two QTEs may be significant even when the confidence intervals overlap. In fact, we reject at the 3% level the equality between the 0.9 and 0.1 QTEs or between the 0.1, 0.25, 0.75 and 0.9 QTEs, for instance.

Some caution must be exercised when interpreting these results, because of the population for which they are identified. We cannot simply interpret the quantile as unobserved ability since all students had the same test score before the summer school. Conditionally on being relatively mediocre at the first test, the best students benefited the most from attending the summer school. We interpret the unobserved heterogeneity more like a measure of motivation: motivated students benefited from being helped during the summer while unmotivated students did not gain anything from additional school hours. Students who do not like going to school will not progress when more school is being imposed. On the other hand, interested students who suffered from a lack of resources and support at home could benefit from this additional support. In light of these results, the summer school cannot be considered as a panacea, but it was effective for the majority of the students.²⁰

Figure 4 provides an alternative visualization of the results which might facilitate interpretation. It shows the estimated density functions of the potential outcomes, $\hat{f}_{Y^1|c}$ and $\hat{f}_{Y^0|c}$, obtained from the distribution functions of Figure 2. This graph shows very clearly that, for the population of compliers who scored close to the threshold z_0 , summer school leads not only to a rightward shift but also to a widening of the test distribution. Many students gain from summer school, but clearly not everyone, and some may even be harmed by attending summer school.²¹

²⁰Note that the students who attended the summer school have also taken a test in August. If their score was good enough, they were allowed to advance. If their score was still too low, they were retained. We do not try to identify separately the effect of grade retention. Our results represent the total effect of summer school participation and potential grade retention. We do not believe that the heterogeneity of the effects is driven by grade retention because the weakest students were retained and Jacob and Lefgren (2004) find a positive effect for grade retention.

²¹We do not require a rank invariance assumption as in Chernozhukov and Hansen (2005) or Guiteras (2008). We therefore identify only the marginal distributions. Figures 3 and 4 only show that most quantiles move upwards, yet individual treatment effects could still be negative for many students.

In our analysis so far, we have not included any control variables. Jacob and Lefgren (2004) in fact include two year dummies in their main specification and 23 other additional control variables for their robustness checks. The main goal of this exercise is to check indirectly the validity of the identifying assumptions. The arguments for the validity of the RD design in this application do not rest on the presence of any conditioning variable. Therefore, conditioning on any pre-determined covariates should not significantly change the results. Of course, this procedure will detect lack of local randomization only if the local self-selection is related to the observed characteristics.

Theorem 10 allows us to include additional control variables without changing the estimand because we integrate them out to obtain the unconditional effects. The large number of control variables X included prevents to use a completely nonparametric approach. We estimate the "propensity score" $p_\varepsilon(X)$ by a parametric linear logistic regression and then implement our estimator in the same way as we did without covariates. The QTEs estimated this way are plotted in Figure 3, which show that they are very close to those estimated without covariates. As in the original article, we find no reason to reject the validity of the design.

We have concentrated on one outcome variable (math test one year after the summer school) so far. In Figure 5 we show the QTE estimates for several other outcomes variables: math and reading test scores one and two years after the summer school, respectively.²² In the figure, we have smoothed the results because the reading test score is less smooth and the estimated distribution functions were quite jumpy. The general pattern is the same for all outcomes, obtaining larger QTEs for larger quantiles.

Table 1 pins down the results presented visually in the Figures 1 to 5 by giving the treatment effects on six summary statistics of the outcome distributions. Standard errors were estimated using the sample analogs of the asymptotic variances derived in Theorem 9. The mean and the median effects show that the summer school attendance had a positive effect on the location of the outcome distribution. These effects are significantly different from zero at levels ranging from 7.5% to 0.01%, depending on the outcome and on the statistic. The standard deviation, interdecile difference ($Q_{Y|c}^{0.9} - Q_{Y|c}^{0.1}$), interquartile difference ($Q_{Y|c}^{0.75} - Q_{Y|c}^{0.25}$) and Gini coefficient measure the treatment effect on the dispersion of the outcomes. With one exception (reading score two years after the summer school) the point estimates indicate that the treatment increased the dispersion of the scores. Some of these effects, most notably the interdecile ranges, are significant and confirm

²²The confidence intervals for these estimates are similar to those shown in Figure 3.

the visual impression discussed above.²³

6 Conclusion

In this paper, we described how the regression discontinuity design can be used to evaluate the impact of endogenous treatments on the entire distribution of outcome variables. We showed that both potential outcome distributions are identified for the population affected by the discontinuity. We introduced estimators for these two distribution functions and showed their joint convergence to continuous Gaussian processes. We also obtained uniformly consistent estimates for functionals of the distribution functions and we considered in detail the quantile treatment effect process when the dependent variable is continuous. By appropriate bandwidth choice, our estimators are consistent at the $n^{-2/5}$ rate, which is the optimal convergence rate for one-dimensional nonparametric estimation.

The approach was illustrated through estimation of the quantile treatment effects of summer school attendance on later educational performance. We used the same data and identification strategy as in Jacob and Lefgren (2004). Our results showed the heterogeneity of the treatment effects, with no effect at the bottom of the distribution and significantly positive effects at the top of the distribution.

We believe that the estimators suggested in this paper have many interesting applications in economics, statistics and other social sciences. Since the late 1990s, a growing number of studies have exploited threshold rules to estimate program effects. Additional insights could be gained in all applications where the outcome is not binary by estimating the distributional effects of the treatment. We make the estimators available in Stata to make them conveniently accessible and usable.

²³The Gini coefficient makes sense only for variables with a positive support. Therefore, we transformed the dependent variables to satisfy this condition. This is debatable since this measure is not invariant to location changes. These results do not contradict the results using other measures of inequality.

A Proofs

A.1 Proof of Theorem 1: Distribution functions of potential outcomes

We have to show that

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

where $p_\varepsilon = \Pr(Z \geq z_0 | Z \in \mathcal{N}_\varepsilon) = E[I^+ | Z \in \mathcal{N}_\varepsilon]$.

In the following, we will prove the first equation and mention that the derivations for the second equation are analogous.

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq u)(I^+ - p_\varepsilon) | Z \in \mathcal{N}_\varepsilon, D = 1]}{E[I^+ - p_\varepsilon | Z \in \mathcal{N}_\varepsilon, D = 1]} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq u)(I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon]}{E[(I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon]} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq u)(I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon^+] p_\varepsilon + E[1(Y \leq u)(I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon^-] (1 - p_\varepsilon)}{E[(I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon^+] p_\varepsilon + E[(I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon^-] (1 - p_\varepsilon)} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq u) D | Z \in \mathcal{N}_\varepsilon^+] - E[1(Y \leq u) D | Z \in \mathcal{N}_\varepsilon^-]}{E[D | Z \in \mathcal{N}_\varepsilon^+] - E[D | Z \in \mathcal{N}_\varepsilon^-]}. \end{aligned} \quad (4)$$

As next step, we examine the first term of the numerator of (4). Define $g(Y)$ to be a one-dimensional measurable function such that $\lim_{\varepsilon \rightarrow 0} E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+]$ and $\lim_{\varepsilon \rightarrow 0} E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^-]$ exist. We will later use the two functions $g(Y) = 1(Y \leq u)$ and $g(Y) = 1$. By Assumption I (v), $\lim_{\varepsilon \rightarrow 0} E[g(Y) D | Z \in \mathcal{N}_\varepsilon^+]$ and $\lim_{\varepsilon \rightarrow 0} E[g(Y) D | Z \in \mathcal{N}_\varepsilon^-]$ exist and are observable. By the law of total probability,

$$\begin{aligned} E[g(Y) D | Z \in \mathcal{N}_\varepsilon^+] &= E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = a] \Pr(\mathcal{T}_\varepsilon = a | Z \in \mathcal{N}_\varepsilon^+) \\ &\quad + E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = n] \Pr(\mathcal{T}_\varepsilon = n | Z \in \mathcal{N}_\varepsilon^+) \\ &\quad + E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = c] \Pr(\mathcal{T}_\varepsilon = c | Z \in \mathcal{N}_\varepsilon^+) \\ &\quad + E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = d] \Pr(\mathcal{T}_\varepsilon = d | Z \in \mathcal{N}_\varepsilon^+) \\ &\quad + E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = i] \Pr(\mathcal{T}_\varepsilon = i | Z \in \mathcal{N}_\varepsilon^+). \end{aligned}$$

By Assumption I (ii) $\lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = d | Z \in \mathcal{N}_\varepsilon^+) = \lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = i | Z \in \mathcal{N}_\varepsilon^+) = 0$. By definition of the types, $E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = n] = 0$, $E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = c] = E[g(Y^1) | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = c]$ and $E[g(Y^1) D | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = a] = E[g(Y^1) | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = a]$.

Therefore, we obtain

$$\begin{aligned}
\lim_{\varepsilon \rightarrow 0} E[g(Y) D | Z \in \mathcal{N}_\varepsilon^+] &= \lim_{\varepsilon \rightarrow 0} E[g(Y^1) | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = a] \Pr(\mathcal{T}_\varepsilon = a | Z \in \mathcal{N}_\varepsilon^+) \\
&\quad + \lim_{\varepsilon \rightarrow 0} E[g(Y^1) | Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = c] \Pr(\mathcal{T}_\varepsilon = c | Z \in \mathcal{N}_\varepsilon^+) \\
&= \lim_{\varepsilon \rightarrow 0} E[g(Y^1) | Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = a] \Pr(\mathcal{T}_\varepsilon = a | Z \in \mathcal{N}_\varepsilon) \\
&\quad + \lim_{\varepsilon \rightarrow 0} E[g(Y^1) | Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c] \Pr(\mathcal{T}_\varepsilon = c | Z \in \mathcal{N}_\varepsilon)
\end{aligned}$$

where the second equality follows from Assumption I (iii) and (iv).

By similar calculations, we obtain

$$\lim_{\varepsilon \rightarrow 0} E[g(Y) D | Z \in \mathcal{N}_\varepsilon^-] = \lim_{\varepsilon \rightarrow 0} E[g(Y^1) | Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = a] \Pr(\mathcal{T}_\varepsilon = a | Z \in \mathcal{N}_\varepsilon).$$

When we apply this intermediate result with respect to $g(Y) = 1(Y \leq u)$ and $g(Y) = 1$, respectively, we can write expression (4) as

$$\begin{aligned}
&\lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq u) D | Z \in \mathcal{N}_\varepsilon^+] - E[1(Y \leq u) D | Z \in \mathcal{N}_\varepsilon^-]}{E[D | Z \in \mathcal{N}_\varepsilon^+] - E[D | Z \in \mathcal{N}_\varepsilon^-]} \\
&= \frac{\lim_{\varepsilon \rightarrow 0} E[1(Y^1 \leq u) | Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c] \Pr(\mathcal{T}_\varepsilon = c | Z \in \mathcal{N}_\varepsilon)}{\lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = c | Z \in \mathcal{N}_\varepsilon)} \\
&= \lim_{\varepsilon \rightarrow 0} E[1(Y^1 \leq u) | Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c] \\
&= \lim_{\varepsilon \rightarrow 0} F_{Y^1 | Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c}(u) \\
&= F_{Y^1 | c}(u),
\end{aligned}$$

because $\lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = c | Z \in \mathcal{N}_\varepsilon)$ is strictly positive by Assumption I (i). The identification of $F_{Y^0 | c}(u)$ is similar, with $1 - D$ replacing D , and is therefore omitted. Note that this result is similar to Lemma 2.1 in Abadie (2002). The difference is that all derivations are localized with respect to z_0 , which gives the nonparametric features of this result.

A.2 Proof of Theorem 2: Quantile functions of potential outcomes

Starting from the results of Theorem 1, $Q_{Y^1 | c}^\tau$ and $Q_{Y^0 | c}^\tau$ are the solutions of the following two moment conditions

$$\begin{aligned}
\lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq Q_{Y^1 | c}^\tau) (I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon]}{E[(I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon]} &= \tau \\
\lim_{\varepsilon \rightarrow 0} \frac{E[1(Y \leq Q_{Y^0 | c}^\tau) (I^+ - p_\varepsilon) (1 - D) | Z \in \mathcal{N}_\varepsilon]}{E[(I^+ - p_\varepsilon) (1 - D) | Z \in \mathcal{N}_\varepsilon]} &= \tau
\end{aligned}$$

or equivalently

$$\begin{aligned}
\lim_{\varepsilon \rightarrow 0} E\left[\left\{1(Y \leq Q_{Y^1 | c}^\tau) - \tau\right\} (I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon\right] &= 0 \\
\lim_{\varepsilon \rightarrow 0} E\left[\left\{1(Y \leq Q_{Y^0 | c}^\tau) - \tau\right\} (I^+ - p_\varepsilon) (1 - D) | Z \in \mathcal{N}_\varepsilon\right] &= 0
\end{aligned}$$

$Q_{Y^1|c}^\tau$ and $Q_{Y^0|c}^\tau$ are identified by these moment conditions. Renaming $Q_{Y^1|c}^\tau$ with $Q_{Y^0|c}^\tau + \Delta_{QTE}^\tau$ and adding the first equation into the second one, we obtain

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} E \left[\left\{ 1 \left(Y \leq Q_{Y^0|c}^\tau + D\Delta_{QTE}^\tau \right) - \tau \right\} (I^+ - p_\varepsilon) D | Z \in \mathcal{N}_\varepsilon \right] &= 0 \\ \lim_{\varepsilon \rightarrow 0} E \left[\left\{ 1 \left(Y \leq Q_{Y^0|c}^\tau + D\Delta_{QTE}^\tau \right) - \tau \right\} (I^+ - p_\varepsilon) | Z \in \mathcal{N}_\varepsilon \right] &= 0 \end{aligned}$$

which are the first order conditions of the weighted quantile regression stated in equation (3).

A.3 Proof of Theorem 5: Limit distribution for distribution functions

We will prove this theorem in several steps. Lemma 11 gives the limiting distribution of $\hat{m}_W^+$ for a generic random variable W that is bounded between 0 and 1. Lemma 13 gives the limiting distribution of $\hat{F}_{Y^1|c}(u)$ and $\hat{F}_{Y^0|c}(u)$ evaluated at a finite number of points. Finally, we show that the process is asymptotically tight, which concludes the proof.

Define the kernel constants $\mu_l = \int u^l K(u) du$ and $\bar{\mu}_l = \int_0^\infty u^l K(u) du$. Define also $\tilde{\mu} = \bar{\mu}_2 \bar{\mu}_0 - \bar{\mu}_1^2$.

Furthermore define $\ddot{\mu}_l = \int_0^\infty u^l K^2(u) du$.²⁴

Lemma 11 (Linear representation of the local linear estimator) *For a generic random variable W , which is bounded between 0 and 1, the estimator $\hat{m}_W^+$ is defined as the value of a that solves*

$$\arg \min_{a,b} \sum_{i=1}^n (W_i - a - b(Z_i - z_0))^2 \cdot I_i^+ K\left(\frac{Z_i - z_0}{h}\right).$$

Its limiting distribution is given by

$$\sqrt{nh} (\hat{m}_W^+ - m_W^+) \xrightarrow{d} N \left(\Xi \frac{\partial^2 m_W^+}{\partial z^2} \frac{\bar{\mu}_2^2 - \bar{\mu}_1 \bar{\mu}_3}{2\tilde{\mu}}, \sigma_W^{2+} \frac{\bar{\mu}_2^2 \ddot{\mu}_0 - 2\bar{\mu}_2 \bar{\mu}_1 \ddot{\mu}_1 + \bar{\mu}_1^2 \ddot{\mu}_2}{f(z_0) \tilde{\mu}^2} \right).$$

Proof. $\hat{m}_W^+$ can also be written as

$$\min_{\beta} \sum_{i=1}^n (W_i - \mathbb{Z}_i' \beta)^2 \cdot K_i I_i^+$$

where

$$\mathbb{Z}_i = \left(1, \frac{Z_i - z_0}{h} \right)',$$

²⁴For the Epanechnikov kernel with support $[-1, 1]$, i.e. $K(u) = \frac{3}{4} (1 - u^2) 1(|u| < 1)$ the kernel constants are $\mu_0 = 1$, $\mu_1 = \mu_3 = \mu_5 = 0$, $\mu_2 = 0.2$, $\mu_4 = \frac{6}{70}$, $\bar{\mu}_0 = 0.5$, $\bar{\mu}_1 = \frac{3}{16}$, $\bar{\mu}_2 = 0.1$, $\bar{\mu}_3 = \frac{1}{16}$, $\bar{\mu}_4 = \frac{3}{70}$, and $\tilde{\mu} = \frac{19}{1280}$ and $\ddot{\mu}_0 = 0.3$, $\ddot{\mu}_1 = \frac{3}{32}$ and $\ddot{\mu}_2 = \frac{3}{70}$. The explicit calculations for the kernel dependent constants of bias and variance in Theorem 5 refer to this Epanechnikov kernel.

$I_i^+ = 1 (Z_i \geq z_0)$ and $\beta = (a, hb)'$. We thus obtain

$$\hat{m}_W^+ = e_1' \left(\sum_{i=1}^n \mathbb{Z}_i \mathbb{Z}_i' K_i I_i^+ \right)^{-1} \sum_{i=1}^n \mathbb{Z}_i W_i K_i I_i^+,$$

where e_1 is a column vector of zeros with first element being one. Define $\hat{A}_+ = \sum_{i=1}^n \mathbb{Z}_i \mathbb{Z}_i' K_i I_i^+$ we obtain

$$= e_1' \hat{A}_+^{-1} \sum_{i=1}^n \mathbb{Z}_i K_i I_i^+ (W_i - E[W_i|Z = Z_i] + E[W_i|Z = Z_i])$$

and with Taylor expansion

$$= e_1' \hat{A}_+^{-1} \sum_{i=1}^n \mathbb{Z}_i K_i I_i^+ \left(W_i - E[W_i|Z = Z_i] + \left\{ m_W^+ + (Z_i - z_0) \frac{\partial m_W^+}{\partial z} + \frac{(Z_i - z_0)^2}{2} \frac{\partial^2 m_W^+}{\partial z^2} + o_p((Z_i - z_0)^2) \right\} \right).$$

After a few calculations the terms with $\frac{\partial m_W^+}{\partial z}$ drop and we obtain

$$\begin{aligned} \hat{m}_W^+ - m_W^+ &= e_1' \hat{A}_+^{-1} \sum_{i=1}^n \mathbb{Z}_i K_i I_i^+ (W_i - E[W_i|Z = Z_i]) \\ &\quad + e_1' \hat{A}_+^{-1} \sum_{i=1}^n \mathbb{Z}_i K_i I_i^+ \left\{ h^2 \frac{(Z_i - z_0)^2}{2h^2} \frac{\partial^2 m_W^+}{\partial z^2} + o_p(h^2) \right\}, \end{aligned}$$

where the remainder term is of order $o_p(h^2)$ because it is premultiplied with the kernel function K which is zero outside a compact set. Using the result of Lemma 12 we obtain

$$= e_1' A_+^{-1} \frac{1}{nh} \sum_{i=1}^n \mathbb{Z}_i K_i I_i^+ \left(W_i - E[W_i|Z = Z_i] + \frac{(Z_i - z_0)^2}{2} \frac{\partial^2 m_W^+}{\partial z^2} \right) (1 + o_p(1))$$

and finally

$$\hat{m}_W^+ - m_W^+ = \frac{1}{nh} \sum_{i=1}^n \frac{\bar{\mu}_2 - \bar{\mu}_1 \frac{Z_i - z_0}{h}}{\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2} \frac{K_i I_i^+}{f(z_0)} \left(W_i - E[W_i|Z = Z_i] + \frac{(Z_i - z_0)^2}{2} \frac{\partial^2 m_W^+}{\partial z^2} \right) (1 + o_p(1)). \quad (5)$$

Next, we calculate bias and variance. The bias is thus given as

$$\begin{aligned} E[\hat{m}_W^+ - m_W^+] &= \frac{1}{h} E \left[\frac{\bar{\mu}_2 - \bar{\mu}_1 \frac{Z_i - z_0}{h}}{\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2} \frac{K_i I_i^+}{f(z_0)} \frac{(Z_i - z_0)^2}{2} \frac{\partial^2 m_W^+}{\partial z^2} (1 + o_p(1)) \right] \\ &= \frac{\partial^2 m_W^+}{\partial z^2} \frac{h^2}{2} \int_0^\infty \frac{\bar{\mu}_2 - \bar{\mu}_1 u}{\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2} \frac{K(u)}{f(z_0)} u^2 f(z_0 + uh) du (1 + o(1)) \\ &= \frac{\partial^2 m_W^+}{\partial z^2} \frac{h^2}{2} \frac{\bar{\mu}_2^2 - \bar{\mu}_1 \bar{\mu}_3}{\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2} + o(h^2) \end{aligned}$$

by a change in variables and dominated convergence.

Next we derive the variance

$$\begin{aligned} & \text{Var}(\hat{m}_W^+ - m_W^+) \\ &= \frac{1}{nh^2} \text{Var} \left(\frac{\bar{\mu}_2 - \bar{\mu}_1 \frac{Z_i - z_0}{h}}{\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2} \frac{K_i I_i^+}{f(z_0)} \left(W_i - E[W_i|Z = Z_i] + h^2 \frac{(Z_i - z_0)^2}{2h^2} \frac{\partial^2 m_W^+}{\partial z^2} \right) (1 + o_p(1)) \right) \end{aligned}$$

which is clearly dominated by the first term

$$\begin{aligned} &= \frac{1}{nh^2} E \left[\left(\frac{\bar{\mu}_2 - \bar{\mu}_1 \frac{Z_i - z_0}{h}}{\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2} \frac{K_i I_i^+}{f(z_0)} (W_i - E[W_i|Z = Z_i]) \right)^2 \right] (1 + o(1)) \\ &= \frac{1}{nh} \int_0^\infty \frac{(\bar{\mu}_2 - \bar{\mu}_1 u)^2}{\left(\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2\right)^2} \frac{K^2(u)}{f^2(z_0)} \text{Var}(W_i|Z = z_0 + uh) f(z_0 + uh) du (1 + o(1)) \\ &= \frac{\sigma_W^{2+}}{nh f(z_0)} \frac{\bar{\mu}_2^2 \ddot{\mu}_0 - 2\bar{\mu}_2 \bar{\mu}_1 \ddot{\mu}_1 + \bar{\mu}_1^2 \ddot{\mu}_2}{\left(\frac{\bar{\mu}_2}{2} - \bar{\mu}_1^2\right)^2} (1 + o(1)) \end{aligned}$$

by dominated convergence.

With these results and the linear representation of (5) and because all higher order moments of W exist (because W has been assumed to be bounded between 0 and 1), asymptotic normality of $\hat{m}_W^+$ follows straightforwardly by a CLT for independent observations

$$\sqrt{nh}(\hat{m}_W^+ - m_W^+) \xrightarrow{d} N \left(\Xi \frac{\partial^2 m_W^+}{\partial z^2} \frac{\bar{\mu}_2^2 - \bar{\mu}_1 \bar{\mu}_3}{2\bar{\mu}}, \sigma_W^{2+} \frac{\bar{\mu}_2^2 \ddot{\mu}_0 - 2\bar{\mu}_2 \bar{\mu}_1 \ddot{\mu}_1 + \bar{\mu}_1^2 \ddot{\mu}_2}{f(z_0) \tilde{\mu}^2} \right).$$

■

Lemma 12 (Denominator of the local linear estimator) *Under the assumption that $nh \rightarrow \infty$,*

$$\frac{1}{nh} \sum_{i=1}^n \mathbb{Z}_i \mathbb{Z}_i' K_i I_i^+ = A_+ + o_p(h^2) \quad (6)$$

where the symmetric matrix A_+ is

$$\begin{bmatrix} \frac{f(z_0)}{2} + hf'(z_0)\bar{\mu}_1 + \frac{h^2}{2}f''(z_0)\bar{\mu}_2 & f(z_0)\bar{\mu}_1 + hf'(z_0)\bar{\mu}_2 + \frac{h^2}{2}f''(z_0)\bar{\mu}_3 \\ f(z_0)\bar{\mu}_1 + hf'(z_0)\bar{\mu}_2 + \frac{h^2}{2}f''(z_0)\bar{\mu}_3 & f(z_0)\bar{\mu}_2 + hf'(z_0)\bar{\mu}_3 + \frac{h^2}{2}f''(z_0)\bar{\mu}_4 \end{bmatrix}$$

plus lower order terms.²⁵

²⁵The matrix A_- is analogously

$$\begin{bmatrix} \frac{f(z_0)}{2} - hf'(z_0)\bar{\mu}_1 + \frac{h^2}{2}f''(z_0)\bar{\mu}_2 & -f(z_0)\bar{\mu}_1 + hf'(z_0)\bar{\mu}_2 - \frac{h^2}{2}f''(z_0)\bar{\mu}_3 \\ -f(z_0)\bar{\mu}_1 + hf'(z_0)\bar{\mu}_2 - \frac{h^2}{2}f''(z_0)\bar{\mu}_3 & f(z_0)\bar{\mu}_2 - hf'(z_0)\bar{\mu}_3 + \frac{h^2}{2}f''(z_0)\bar{\mu}_4 \end{bmatrix}.$$

Proof. The relationship (6) is shown via mean square convergence for each element of A_+ . Only the derivations for the (1, 2) element are shown here, with the derivations for the other elements being analogous.

Consider the (1, 2) element of $\frac{1}{nh} \sum_{i=1}^n \mathbb{Z}_i \mathbb{Z}_i' K_i I_i^+$ and denote it by ξ

$$\xi = \frac{1}{nh} \sum_{i=1}^n \frac{Z_i - z_0}{h} K_i I_i^+$$

which has the expected value:

$$\begin{aligned} E[\xi] &= \frac{1}{h} E \left[\frac{Z_i - z_0}{h} K_i I_i^+ \right] \\ &= \frac{1}{h} \int \frac{z - z_0}{h} K \left(\frac{z - z_0}{h} \right) 1(z \geq z_0) f(z) dz. \end{aligned}$$

With a change in variables: $u = \frac{z - z_0}{h}$ and an expansion about the point z_0 , considering only points to the right of z_0 , we obtain

$$\begin{aligned} &= \int u K(u) 1(u \geq 0) f(z_0 + uh) du \\ &= \int_0^\infty u K(u) \left\{ f(z_0) + uh f'(z_0) + \frac{u^2 h^2}{2} f''(z_0) \right\} du \\ &= f(z_0) \bar{\mu}_1 + h f'(z_0) \bar{\mu}_2 + \frac{h^2}{2} f''(z_0) \bar{\mu}_3 \end{aligned}$$

by bounded convergence.

To show convergence in mean square, it also needs to be shown that $Var(\xi)$ converges to zero:

$$\begin{aligned} &Var(\xi) \\ &= \frac{1}{n^2 h^2} Var \left(\sum \frac{Z_i - z_0}{h} K_i I_i^+ \right) = \frac{1}{n h^2} Var \left(\frac{Z_i - z_0}{h} K_i I_i^+ \right) \\ &= \frac{1}{n h^2} \left\{ E \left[\left(\frac{Z_i - z_0}{h} K_i I_i^+ \right)^2 \right] - E \left[\frac{Z_i - z_0}{h} K_i I_i^+ \right]^2 \right\} \\ &= \frac{1}{n h^2} \int_0^\infty \left(\frac{z - z_0}{h} \right)^2 K^2 \left(\frac{z - z_0}{h} \right) f(z) dz - O \left(\frac{1}{n} \right) \\ &= \frac{1}{n h} \int_0^\infty u^2 K^2(u) f(z_0 + uh) du - O \left(\frac{1}{n} \right) \end{aligned}$$

where a change in variables: $u = \frac{z - z_0}{h}$ and a Taylor expansion about the point z_0 has been used. As it has been assumed that $nh \rightarrow \infty$, the variance of ξ converges to zero by dominated convergence. Hence, mean square convergence has been shown, which implies convergence in probability by Chebyshev's inequality. ■

Lemma 13 *If assumptions I and R are satisfied, the estimators $\hat{F}_{Y^1|c}(u)$, $\hat{F}_{Y^0|c}(u)$, $\hat{F}_{Y^1|c}(\tilde{u})$, $\hat{F}_{Y^0|c}(\tilde{u})$ for any u and $\tilde{u} \in \mathcal{Y}$ jointly converge in law to the following normal distribution:*

$$\sqrt{nh} \begin{pmatrix} \hat{F}_{Y^1|c}(u) - F_{Y^1|c}(u) \\ \hat{F}_{Y^0|c}(u) - F_{Y^0|c}(u) \\ \hat{F}_{Y^1|c}(\tilde{u}) - F_{Y^1|c}(\tilde{u}) \\ \hat{F}_{Y^0|c}(\tilde{u}) - F_{Y^0|c}(\tilde{u}) \end{pmatrix} \xrightarrow{d} N(\mathcal{B}, \mathcal{V})$$

where $\mathcal{B}$ and $\mathcal{V}$ are defined below in the proof.

Proof. By Lemma 11 and 12 we have a linear representation of the local linear estimator (5) of a generic random variable W that is bounded between zero and one. Now we consider the joint normality of $\hat{F}_{Y^1|c}(u)$, $\hat{F}_{Y^0|c}(u)$, $\hat{F}_{Y^1|c}(\tilde{u})$ and $\hat{F}_{Y^0|c}(\tilde{u})$, i.e. of the estimators of the cdf at two different values u and $\tilde{u}$. For convenience we restate the definition of the estimator here:

$$\hat{F}_{Y^1|c}(u) = \frac{\hat{m}_{1(Y \leq u)D}^+ - \hat{m}_{1(Y \leq u)D}^-}{\hat{m}_D^+ - \hat{m}_D^-}.$$

Note that this estimator is a continuous function of estimators $\hat{m}_W^+$ and $\hat{m}_W^-$ for different generic variables W , once for $W = 1 (Y \leq u) D$ and once for $W = D$. In any case W is bounded between zero and one and all higher order moments exist.

Define the vector $\hat{\theta}$ as a function of the sample size (and of u and $\tilde{u}$) as $\hat{\theta} =$

$$(\hat{m}_D^+, \hat{m}_D^-, \hat{m}_{1(Y \leq u)D}^+, \hat{m}_{1(Y \leq u)D}^-, \hat{m}_{1(Y \leq u)(D-1)}^+, \hat{m}_{1(Y \leq u)(D-1)}^-, \hat{m}_{1(Y \leq \tilde{u})D}^+, \hat{m}_{1(Y \leq \tilde{u})D}^-, \hat{m}_{1(Y \leq \tilde{u})(D-1)}^+, \hat{m}_{1(Y \leq \tilde{u})(D-1)}^-)$$

By the Cramer-Wold device it is trivial to show that $\hat{\theta}$ converges to a jointly normal distribution: If $\lambda' \hat{\theta}$ converges to a normal random variable for every conformable constant nonzero vector λ , then $\hat{\theta}$ converges to a multivariate normal random variable. Because every element in $\hat{\theta}$ can be represented in linear form as in (5), $\lambda' \hat{\theta}$ is a sample average of random variables, for whom higher order moments exist. Applying a CLT to $\lambda' \hat{\theta}$, premultiplied with $\sqrt{nh}$, thus gives asymptotic normality, which thus implies joint normality of the elements in $\hat{\theta}$.

With this notation, we can restate the estimators of the cdf as

$$\begin{aligned} \hat{F}_{Y^1|c}(u) &= \frac{\hat{\theta}_3 - \hat{\theta}_4}{\hat{\theta}_1 - \hat{\theta}_2} & \text{and} & & \hat{F}_{Y^0|c}(u) &= \frac{\hat{\theta}_5 - \hat{\theta}_6}{\hat{\theta}_1 - \hat{\theta}_2} \\ \hat{F}_{Y^1|c}(\tilde{u}) &= \frac{\hat{\theta}_7 - \hat{\theta}_8}{\hat{\theta}_1 - \hat{\theta}_2} & \text{and} & & \hat{F}_{Y^0|c}(\tilde{u}) &= \frac{\hat{\theta}_9 - \hat{\theta}_{10}}{\hat{\theta}_1 - \hat{\theta}_2} \end{aligned}$$

where $\hat{\theta}_k$ refers to the k -th element of $\hat{\theta}$. Consider $\hat{F}_{Y^1|c}(u)$, $\hat{F}_{Y^0|c}(u)$, $\hat{F}_{Y^1|c}(\tilde{u})$ and $\hat{F}_{Y^0|c}(\tilde{u})$ as a vector of a continuous function of the asymptotically normal $\hat{\theta}$. It thereby follows by the continuous mapping theorem that the estimators of the cdf are jointly asymptotically normal

$$\sqrt{nh} \begin{pmatrix} \hat{F}_{Y^1|c}(u) - F_{Y^1|c}(u) \\ \hat{F}_{Y^0|c}(u) - F_{Y^0|c}(u) \\ \hat{F}_{Y^1|c}(\tilde{u}) - F_{Y^1|c}(\tilde{u}) \\ \hat{F}_{Y^0|c}(\tilde{u}) - F_{Y^0|c}(\tilde{u}) \end{pmatrix} \xrightarrow{d} N(\mathcal{B}, \mathcal{V}).$$

It remains to calculate the bias and covariance matrix, which can be obtained by the delta method. The first element of the bias $\mathcal{B}$ is

$$\frac{\Xi}{P_c} \underbrace{\frac{\bar{\mu}_2^2 - \bar{\mu}_1 \bar{\mu}_3}{2\bar{\mu}}}_{=-\frac{11}{190} \text{ for Epanechnikov}} \left\{ \frac{\partial^2 m_{1(Y \leq u)D}^+}{\partial z^2} - F_{Y^1|c}(u) \frac{\partial^2 m_D^+}{\partial z^2} - \frac{\partial^2 m_{1(Y \leq u)D}^-}{\partial z^2} + F_{Y^1|c}(u) \frac{\partial^2 m_D^-}{\partial z^2} \right\}.$$

The second element of the bias is

$$\frac{\Xi}{P_c} \frac{\bar{\mu}_2^2 - \bar{\mu}_1 \bar{\mu}_3}{2\bar{\mu}} \left\{ \frac{\partial^2 m_{1(Y \leq u)(D-1)}^+}{\partial z^2} - F_{Y^0|c}(u) \frac{\partial^2 m_D^+}{\partial z^2} - \frac{\partial^2 m_{1(Y \leq u)(D-1)}^-}{\partial z^2} + F_{Y^0|c}(u) \frac{\partial^2 m_D^-}{\partial z^2} \right\}$$

and the third and fourth elements of $\mathcal{B}$ are analogous to the first two elements with u replaced by $\tilde{u}$.

After tedious calculations and application of the dominated convergence theorem we obtain for the covariance matrix

$$\mathcal{V} = \underbrace{\frac{\bar{\mu}_2^2 \ddot{\mu}_0 - 2\bar{\mu}_2 \bar{\mu}_1 \ddot{\mu}_1 + \bar{\mu}_1^2 \ddot{\mu}_2}{\ddot{\mu}^2}}_{=\frac{56832}{12635} \text{ for the Epanechnikov kernel}} \frac{1}{P_c^2 f(z_0)} \begin{pmatrix} \omega_{Y^1(u)Y^1(u)}^{+-} & \omega_{Y^1(u)Y^0(u)}^{+-} & \omega_{Y^1(u)Y^1(\tilde{u})}^{+-} & \omega_{Y^1(u)Y^0(\tilde{u})}^{+-} \\ & \omega_{Y^0(u)Y^0(u)}^{+-} & \omega_{Y^0(u)Y^1(\tilde{u})}^{+-} & \omega_{Y^0(u)Y^0(\tilde{u})}^{+-} \\ & & \omega_{Y^1(\tilde{u})Y^1(\tilde{u})}^{+-} & \omega_{Y^1(\tilde{u})Y^0(\tilde{u})}^{+-} \\ & & & \omega_{Y^0(\tilde{u})Y^0(\tilde{u})}^{+-} \end{pmatrix},$$

where

$$\omega_{Y^1(u)Y^1(\tilde{u})}^{+-} = \lim_{\varepsilon \rightarrow 0} Cov \{ 1(Y \leq u)D - F_{Y^1|c}(u)D, 1(Y \leq \tilde{u})D - F_{Y^1|c}(\tilde{u})D \mid Z = z_0 + \varepsilon \}$$

is the right limit and define $\omega_{Y^1(u)Y^1(\tilde{u})}^{-}$ analogously as the left limit and

$$\omega_{Y^1(u)Y^1(\tilde{u})}^{+-} = \omega_{Y^1(u)Y^1(\tilde{u})}^{+} + \omega_{Y^1(u)Y^1(\tilde{u})}^{-}$$

is the sum of left and right limit. Similarly, for the covariance element between $\hat{F}_{Y^1|c}(u)$ and $\hat{F}_{Y^0|c}(\tilde{u})$

$$\omega_{Y^1(u)Y^0(\tilde{u})}^{+-} = \lim_{\varepsilon \rightarrow 0} Cov \{ 1(Y \leq u)D - F_{Y^1|c}(u)D, 1(Y \leq \tilde{u})(D-1) - F_{Y^0|c}(\tilde{u})D \mid Z = z_0 + \varepsilon \}.$$

The modifications for the definition of the other elements of $\mathcal{V}$ are obvious. ■

Lemma 13 shows the convergence of the finite dimensional distributions. The last step to prove Theorem 5 consists in verifying the asymptotic tightness. Starting from the asymptotic representation in Theorem 13, we have to show that the process $W_h(u)$ is asymptotically tight, where $W_h(u) = \sum_{i=1}^n Z_{ni}(u)$ and

$$Z_{ni}(u) = (nh_n)^{-1/2} P_c^{-1} K_i D_i \{ I_i^+ (1(Y_i \leq u) - F_{Y^1|c}(u)) - (1 - I_i^+) (1(Y_i \leq u) - F_{Y^1|c}(u)) \}.$$

We can consider separately the first and the second term in the curly brackets. We will show the result only for the first term. The second term is similar.

We will prove that the three conditions displayed in Theorem 2.11.9 in van der Vaart and Wellner (1996) are satisfied for $W_h(u)$.²⁶ Endow $\mathcal{Y}$ with the semimetric ρ defined by

$$\rho(u, \tilde{u}) = P_c^{-1} \sqrt{|F_{Y^1|c}(u) - F_{Y^1|c}(\tilde{u})|}.$$

Let us divide $\mathcal{Y}$ into sub-intervals $t_0 \leq t_1 \leq \dots \leq t_q$ where $\rho(u, \tilde{u}) \leq C\varepsilon$ for all $u, \tilde{u} \in [t_{j-1}, t_j], j = 1, \dots, q$ with C some constant which we will determine further on. For the partition $\mathcal{Y} = [t_0, t_1] \cup \dots \cup [t_{q-1}, t_q]$, we find that

$$|Z_{ni}(u) - Z_{ni}(\tilde{u})| \leq (nh_n)^{-1/2} P_c^{-1} K_i D_i (|1(Y_i \leq u) - 1(Y_i \leq \tilde{u})| + |F_{Y^1|c}(u) - F_{Y^1|c}(\tilde{u})|).$$

So

$$\sup_{u, \tilde{u} \in \mathcal{Y}_{\varepsilon j}} |Z_{ni}(u) - Z_{ni}(\tilde{u})|^2 \leq (nh_n)^{-1} \bar{K}^2 (C^2 \varepsilon^2 + |1(Y_i \leq u) - 1(Y_i \leq \tilde{u})|^2).$$

For the appropriate choice of C , this leads to

$$\sum_{i=1}^n E \left[\sup_{u, \tilde{u} \in \mathcal{Y}_{\varepsilon j}} |Z_{ni}(u) - Z_{ni}(\tilde{u})|^2 \right] \leq \varepsilon^2$$

Hence the bracketing number $N_{[]}(\varepsilon, \mathcal{Y}, L_2^n)$ is equal to $O(\varepsilon^{-1})$ and we get

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\varepsilon, \mathcal{Y}, L_2^n)} d\varepsilon = \int_0^{\delta_n} \sqrt{O(\varepsilon^{-1})} d\varepsilon \rightarrow 0$$

when $\delta_n \rightarrow 0$. This verifies the third condition of Theorem 2.11.9 in van der Vaart and Wellner (1996). We do not need to verify the second condition, since our partition of $\mathcal{Y}$ is independent of n .

Finally, we have to check whether for all $\eta > 0$

$$\sum_{i=1}^n E \left[\sup_{u \in \mathcal{Y}} |Z_{ni}(u)| \cdot 1 \left(\sup_{u \in \mathcal{Y}} |Z_{ni}(u)| > \eta \right) \right] \rightarrow 0 \text{ as } n \rightarrow \infty$$

Since $P_c^{-1} K_i I_i^+ (D_i \cdot 1(Y_i \leq u) - F_{Y^1|c}(u) D)$ is bounded uniformly, we get that $\sup_{u \in \mathcal{Y}} |Z_{ni}(u)| = O((nh_n)^{-1/2})$ which is always smaller than η for n sufficiently large. So the first condition is also satisfied.

This verifies that the process is asymptotically tight, which, in combination with Lemma 13, implies the result of Theorem 5 by Theorem 1.5.4 in van der Vaart and Wellner (1996).

A.4 Proof of Corollary 6: Limit distribution for distribution treatment effects

This result follows from Theorem 5 by the continuous mapping theorem (see for instance Theorem 1.9.5 in van der Vaart and Wellner (1996)).

²⁶Similar applications of this result can be found in van Keilegom and Akritas (1999), Braekers and Veraverbeke (2003), Gaddah and Braekers (2009),

A.5 Proof of Theorem 7: Limit distribution for quantile functions

This result follows from Theorem 5 by the functional delta method, since the quantile operator is Hadamard differentiable for absolutely continuous functions, which is assumed in Assumption Q (see for instance section 2.2.4 in Kosorok (2008) for a definition of the functional delta method and an application to the quantile operator).

A.6 Proof of Corollary 8: Limit distribution for quantile treatment effects

This result follows from Theorem 7 by the continuous mapping theorem (see for instance Theorem 1.9.5 in van der Vaart and Wellner (1996)).

A.7 Proof of Theorem 9: Limit distribution for smooth functionals

This result follows from Theorem 5 by the functional delta method.

A.8 Proof of Theorem 10: Identification of QTEs in the presence of covariates

We will proceed in three steps. We first show a regression representation of the distribution functions. This result is very similar to Theorem 5 in Frölich (2007).

Lemma 14 (Distributions with covariates) *Under Assumption C, the potential outcome distributions for the local compliers are identified as*

$$\begin{aligned} F_{Y^1|c}(u) &= \lim_{\varepsilon \rightarrow 0} \frac{\int (m^+(1(Y \leq u)D|x) - m^-(1(Y \leq u)D|x)) dF_{X|Z \in \mathcal{N}_\varepsilon}(x)}{\int (m^+(D|x) - m^-(D|x)) dF_{X|Z \in \mathcal{N}_\varepsilon}(x)} \\ F_{Y^0|c}(u) &= \lim_{\varepsilon \rightarrow 0} \frac{\int (m^+(1(Y \leq u)(D-1)|x) - m^-(1(Y \leq u)(D-1)|x)) dF_{X|Z \in \mathcal{N}_\varepsilon}(x)}{\int (m^+(D|x) - m^-(D|x)) dF_{X|Z \in \mathcal{N}_\varepsilon}(x)} \end{aligned}$$

where $m^+(W|x) = E[W|X=x, Z \in \mathcal{N}_\varepsilon^+]$ and $m^-(W|x) = E[W|X=x, Z \in \mathcal{N}_\varepsilon^-]$ for any random variable W .

Proof. Let $g(y)$ be a real measurable and absolutely integrable function. Consider the interior terms

$$\begin{aligned} & m^+(g(Y)D|x) - m^-(g(Y)D|x) \\ &= m^+(g(Y^1)|x, \mathcal{T}_\varepsilon = a) \Pr(\mathcal{T}_\varepsilon = a|X, Z \in \mathcal{N}_\varepsilon^+) - m^-(g(Y^1)|x, \mathcal{T}_\varepsilon = a) \Pr(\mathcal{T}_\varepsilon = a|X, Z \in \mathcal{N}_\varepsilon^-) \\ & \quad + m^+(g(Y^1) \cdot 0|x, \mathcal{T}_\varepsilon = n) \Pr(\mathcal{T}_\varepsilon = n|X, Z \in \mathcal{N}_\varepsilon^+) - m^-(g(Y^1) \cdot 0|x, \mathcal{T}_\varepsilon = n) \Pr(\mathcal{T}_\varepsilon = n|X, Z \in \mathcal{N}_\varepsilon^-) \\ & \quad + m^+(g(Y^1)|x, \mathcal{T}_\varepsilon = c) \Pr(\mathcal{T}_\varepsilon = c|X, Z \in \mathcal{N}_\varepsilon^+) - m^-(g(Y^1) \cdot 0|x, \mathcal{T}_\varepsilon = c) \Pr(\mathcal{T}_\varepsilon = c|X, Z \in \mathcal{N}_\varepsilon^-) \\ & \quad + m^+(g(Y^1) \cdot 0|x, \mathcal{T}_\varepsilon = d) \Pr(\mathcal{T}_\varepsilon = d|X, Z \in \mathcal{N}_\varepsilon^+) - m^-(g(Y^1)|x, \mathcal{T}_\varepsilon = d) \Pr(\mathcal{T}_\varepsilon = d|X, Z \in \mathcal{N}_\varepsilon^-) \\ & \quad + m^+(g(Y^1) \cdot D|x, \mathcal{T}_\varepsilon = i) \Pr(\mathcal{T}_\varepsilon = i|X, Z \in \mathcal{N}_\varepsilon^+) - m^-(g(Y^1) \cdot D|x, \mathcal{T}_\varepsilon = i) \Pr(\mathcal{T}_\varepsilon = i|X, Z \in \mathcal{N}_\varepsilon^-). \end{aligned}$$

Now consider the integral

$$\lim_{\varepsilon \rightarrow 0} \int (E[g(Y)D|X, Z \in \mathcal{N}_\varepsilon^+] - E[g(Y)D|X, Z \in \mathcal{N}_\varepsilon^-]) dF_{X|Z \in \mathcal{N}_\varepsilon}$$

and insert the previous expression. All terms have well defined limits, mostly zero. The limits of the terms for the $\mathcal{T}_\varepsilon = a$ and $\mathcal{T}_\varepsilon = n$ populations are zero by Assumptions C (iii) and (iv). Since $g(Y)$ is absolutely integrable and since conditional probabilities are bounded and since $\lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = d|X, Z \in \mathcal{N}_\varepsilon^-)$ and $\lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = i|X, Z \in \mathcal{N}_\varepsilon)$ are zero by Assumption C (ii), also the limits of the terms for the $\mathcal{T}_\varepsilon = d$ and $\mathcal{T}_\varepsilon = i$ subpopulations are zero. Hence, it remains

$$= \lim_{\varepsilon \rightarrow 0} \int E[g(Y^1)|X, Z \in \mathcal{N}_\varepsilon^+, \mathcal{T}_\varepsilon = c] \Pr(\mathcal{T}_\varepsilon = c|X, Z \in \mathcal{N}_\varepsilon^+) dF_{X|Z \in \mathcal{N}_\varepsilon}.$$

By Assumptions C (iii) and C (iv), this is equal to

$$= \lim_{\varepsilon \rightarrow 0} \int E[g(Y^1)|X, Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c] \Pr(\mathcal{T}_\varepsilon = c|X, Z \in \mathcal{N}_\varepsilon) dF_{X|Z \in \mathcal{N}_\varepsilon}.$$

By making use of Bayes' theorem, we obtain

$$\begin{aligned} &= \lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = c|Z \in \mathcal{N}_\varepsilon) \int E[g(Y^1)|X, Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c] dF(X|\mathcal{T}_\varepsilon = c, Z \in \mathcal{N}_\varepsilon) \\ &= \lim_{\varepsilon \rightarrow 0} \Pr(\mathcal{T}_\varepsilon = c|Z \in \mathcal{N}_\varepsilon) \cdot E[g(Y^1)|Z \in \mathcal{N}_\varepsilon, \mathcal{T}_\varepsilon = c]. \end{aligned}$$

Applying this result with $g(y) = 1(Y \leq u)$ and $g(y) = 1$, for the numerator and denominator respectively, we obtain the result of lemma 14 for Y^1 . The result for Y^0 can be obtained similarly with D replaced by $1 - D$ in all expressions. ■

The distribution function has also a weighting representation:

Lemma 15 (Weighted distributions with covariates) *Under Assumption C, the potential outcome distributions for the local compliers are identified as*

$$\begin{aligned} F_{Y^1|c}(u) &= \lim_{\varepsilon \rightarrow 0} \frac{E\left[1(Y \leq u) \frac{1(Z \geq z_0) - p_\varepsilon(X)}{p_\varepsilon(X)(1 - p_\varepsilon(X))} (2D - 1) | Z \in \mathcal{N}_\varepsilon, D = 1\right]}{E\left[\frac{1(Z \geq z_0) - p_\varepsilon(X)}{p_\varepsilon(X)(1 - p_\varepsilon(X))} (2D - 1) | Z \in \mathcal{N}_\varepsilon, D = 1\right]} \\ F_{Y^0|c}(u) &= \lim_{\varepsilon \rightarrow 0} \frac{E\left[1(Y \leq u) \frac{1(Z \geq z_0) - p_\varepsilon(X)}{p_\varepsilon(X)(1 - p_\varepsilon(X))} (2D - 1) | Z \in \mathcal{N}_\varepsilon, D = 0\right]}{E\left[\frac{1(Z \geq z_0) - p_\varepsilon(X)}{p_\varepsilon(X)(1 - p_\varepsilon(X))} (2D - 1) | Z \in \mathcal{N}_\varepsilon, D = 0\right]} \end{aligned}$$

where $p_\varepsilon(x) = \Pr(Z \geq z_0|X = x, Z \in \mathcal{N}_\varepsilon)$ for $\varepsilon > 0$.

Proof. This representation can be obtained by using Bayes' theorem starting from the results of Lemma 14. For any real measurable and absolutely integrable function $g(y)$

$$\begin{aligned} &\int (E[g(Y)D|X, Z \in \mathcal{N}_\varepsilon^+] - E[g(Y)D|X, Z \in \mathcal{N}_\varepsilon^-]) dF_{X|Z \in \mathcal{N}_\varepsilon} \\ &= \int E\left[\frac{g(Y)D \cdot 1(Z \geq z_0)}{p_\varepsilon(x)} - \frac{g(Y)D(1 - 1(Z \geq z_0))}{1 - p_\varepsilon(x)} | X, Z \in \mathcal{N}_\varepsilon\right] dF_{X|Z \in \mathcal{N}_\varepsilon} \\ &= E\left[g(Y)D \frac{I^+ - p_\varepsilon(x)}{p_\varepsilon(x)(1 - p_\varepsilon(x))} | Z \in \mathcal{N}_\varepsilon\right] \\ &= \Pr(D = 1|Z \in \mathcal{N}_\varepsilon) E\left[g(Y) \frac{1(Z \geq z_0) - p_\varepsilon(X)}{p_\varepsilon(X)(1 - p_\varepsilon(X))} (2D - 1) | Z \in \mathcal{N}_\varepsilon, D = 1\right] \end{aligned}$$

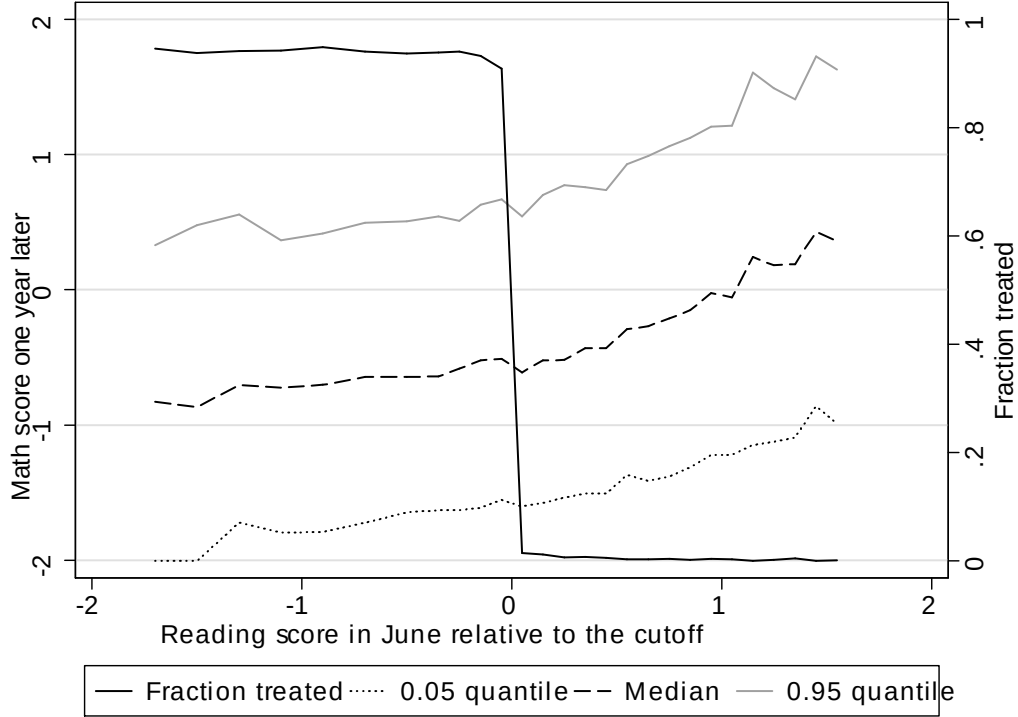
We obtain Lemma 15 by applying this result to the numerators and denominators of Lemma 14.

■

Proof of Theorem 10. The proof follows from Lemma 15 using the same line of arguments as in the proof of Theorem 2. ■

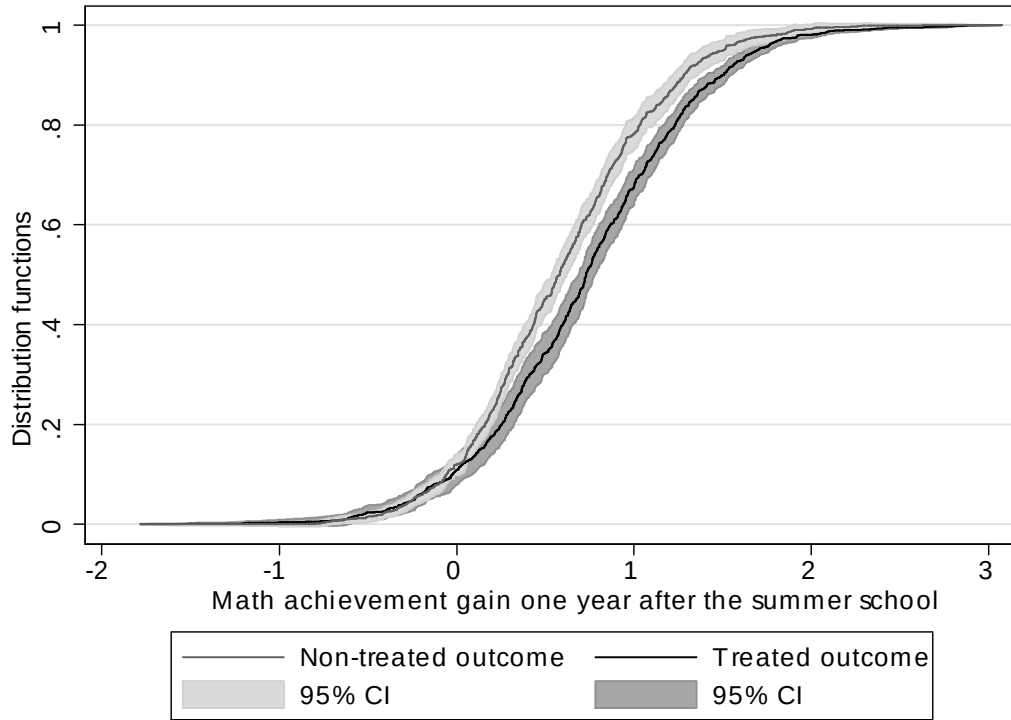
B Figures and Table

Figure 1: The relationship between the forcing variable Z , the probability of attending the summer school and three quantiles of the observed outcome



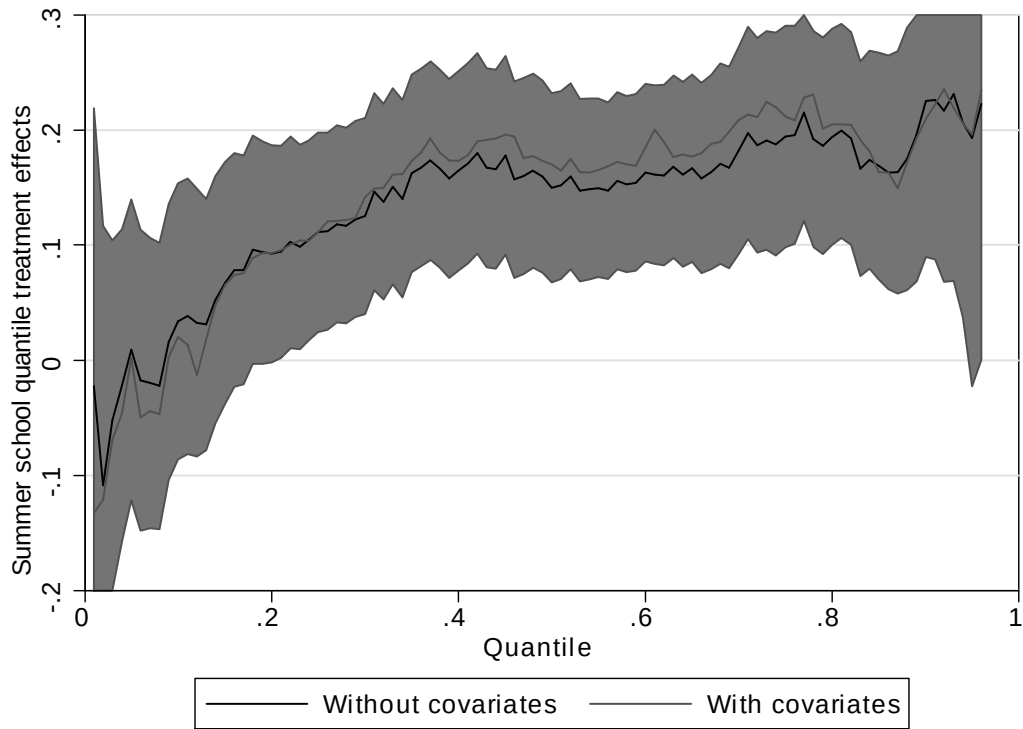
Note: June reading grade equivalents relative to threshold (Z variable, with $z_0 = 0$). Bold line: Probability of treatment as a function of Z , i.e. fraction of students with same value of Z who attended the summer school. Dashed, dotted and gray lines: median, 0.05 and 0.95 quantile of the math grade equivalents one year after the summer school as a function of Z . Population: Third-grade students from 1997 to 1999 whose math score exceeded the promotional cutoff.

Figure 2: Distribution functions of the potential outcomes



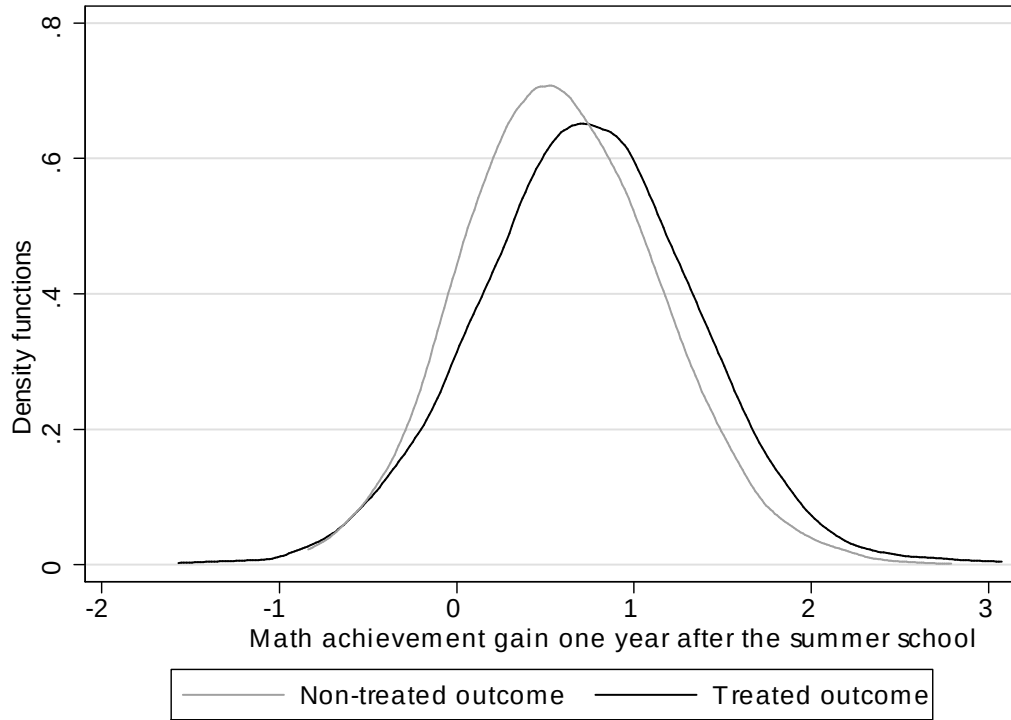
Note: Estimated distribution functions of the potential outcomes $\hat{F}_{Y^1|c}$ and $\hat{F}_{Y^0|c}$, and pointwise 95% confidence intervals based on Theorem 5. Outcome variable is the achievement gain in math test score one year after the summer school. Population: Third-grade students from 1997 to 1999 whose math score exceeded the promotional cutoff and whose reading test score was just below or just above the threshold.

Figure 3: Quantile treatment effects of summer school attendance on math test score



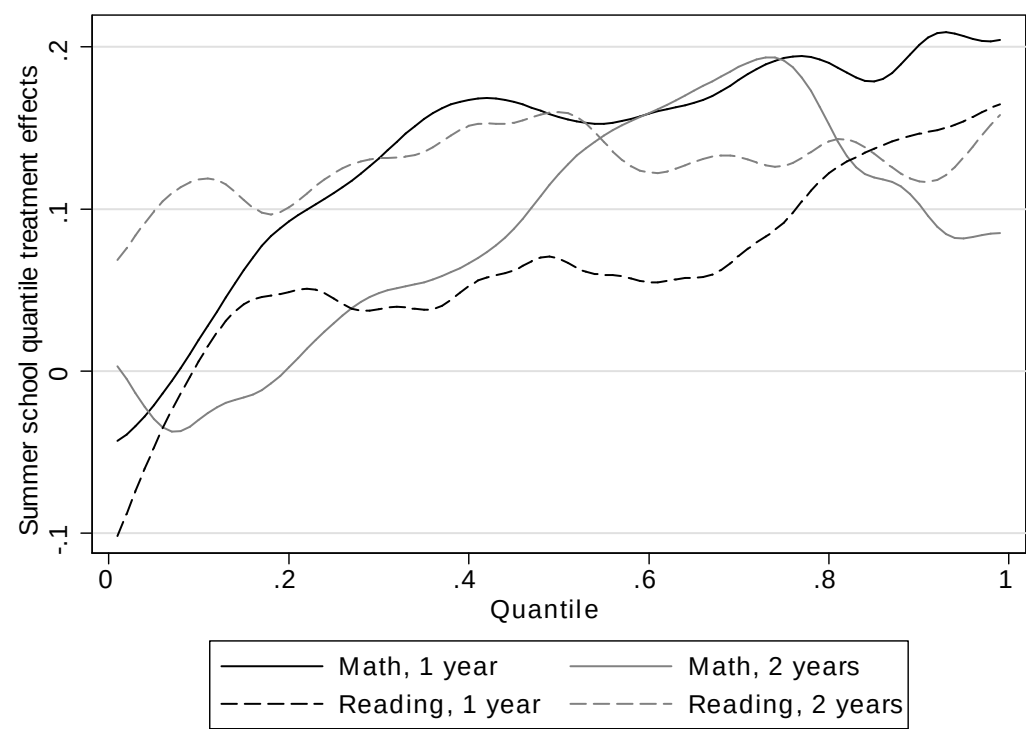
Note: Estimated quantile treatment effects, with and without covariates. Pointwise 95% confidence intervals are shown for the QTE estimated without covariates. Outcome variable is the achievement gain in math test score one year after the summer school. Population: Third-grade students from 1997 to 1999 whose math score exceeded the promotional cutoff and whose reading test score was just below or just above the threshold.

Figure 4: Density functions of the potential outcomes



Note: Estimated density functions of the potential outcomes $\hat{f}_{Y1|c}$ and $\hat{f}_{Y0|c}$. Density functions only shown for illustration. No confidence intervals given. Outcome variable is the achievement gain in math test score one year after the summer school. Population: Third-grade students from 1997 to 1999 whose math score exceeded the promotional cutoff and whose reading test score was just below or just above the threshold.

Figure 5: Quantile treatment effects of summer school attendance on other outcomes



Note: Quantile treatment effects, estimated without covariates, for four outcome variables: achievement gain in math and reading test scores one and two years after the summer school. Population: Third-grade students from 1997 to 1999 whose math score exceeded the promotional cutoff and whose reading test score was just below or just above the threshold.

Table 1: Summer school treatment effects on summary statistics of the outcome distribution

Statistic	Outcome			
	Math gain		Reading gain	
	1 Year	2 Years	1 Year	2 Years
Mean	0.140 (3.78)	0.085 (2.34)	0.063 (1.78)	0.127 (3.53)
Median	0.163 (3.82)	0.125 (2.77)	0.112 (2.74)	0.151 (3.15)
Standard deviation	0.070 (1.43)	0.045 (1.47)	0.061 (1.40)	0.012 (0.37)
Interdecile range	0.175 (2.02)	0.162 (1.94)	0.200 (2.40)	-0.024 (-0.29)
Interquartile range	0.083 (0.97)	0.171 (2.05)	0.000 (0.00)	-0.023 (-0.27)
Gini coefficient	0.007 (1.36)	0.006 (1.56)	0.005 (1.63)	-0.000 (-0.02)

Note: Treatment effects on six summary statistics of the outcome distribution. Four outcomes are considered: achievement gains in math and reading scores one and two years after the summer school. The outcomes have been re-localized to avoid negative outcomes. This allows calculating the Gini coefficient but does not affect the other statistics. t-values are given in parentheses. The standard errors were estimated using the sample analog (plug-in principle) of the asymptotic formulas derived in Theorem 9. Population: Third-grade students from 1997 to 1999 whose math score exceeded the promotional cutoff and whose reading test score was just below or just above this threshold.

References

- ABADIE, A. (2002): “Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models,” *Journal of the American Statistical Association*, 97, 284–292.
- ABADIE, A. (2003): “Semiparametric instrumental Variable estimation of treatment response models,” *Journal of Econometrics*, 113, 231–263.
- ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica*, 70, 91–117.
- 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.
- BARRETT, G. F., AND S. G. DONALD (2000): “Statistical Inference with Generalized Gini Indices of Inequality and Poverty,” Working Paper, University of New South Wales.
- 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. SMITH, M. BERGER, AND B. NOEL (2003): “Is the Threat of Reemployment Services More Effective Than the Services Themselves? - Evidence from Random Assignment in the UI System,” *American Economic Review*, 93, 1313–1327.
- BLACK, D. A., J. GALDO, AND J. A. SMITH (2007): “Evaluating the regression discontinuity using experimental data,” Mimeo, University of Michigan.
- BLACK, S. (1999): “Do ‘Better’ Schools Matter? Parental Valuation of Elementary Education,” *Quarterly Journal of Economics*, 114, 577–599.
- BRAEKERS, R., AND N. VERAVERBEKE (2003): “Testing for the partial Koziol-Green model with covariates,” *Journal of Statistical Planning and Inference*, 115, 181–192.
- BRÜGGER, B., R. LALIVE, AND J. ZWEIMÜLLER (2008): “Does culture affect unemployment? Evidence from the Barriere des Roestis,” *mimeo, University of Zürich*.
- BUDELMEYER, H., AND E. SKOUFIAS (2003): “An evaluation of the performance of regression discontinuity design on PROGRESA,” *IZA discussion paper*, 827.
- CAMPBELL, D., AND J. STANLEY (1963): “Experimental and quasi-experimental designs for research on teaching,” in *Handbook of Research on Teaching*, ed. by N. Gage. Rand McNally, Chicago.
- CHAUDHURI, P. (1991): “Global nonparametric estimation of conditional quantile functions and their derivatives,” *Journal of Multivariate Analysis*, 39, 246–269.

- 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*, pp. 1237–1258.
- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, AND A. GALICHON (2007): “Quantile and probability curves without crossing,” MIT working paper.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2007): “Inference on Counterfactual Distributions,” *mimeo*.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “An IV model of quantile treatment effects,” *Econometrica*, 73, 245–261.
- DINARDO, J., AND D. LEE (2004): “Economic impacts of new unionization on private sectors employers: 1984-2001,” *Quarterly Journal of Economics*, 119, 1383–1441.
- FIRPO, S. (2008): “Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures,” *mimeo*.
- FISHER, R. (1935): *Design of Experiments*. Oliver and Boyd, Edinburgh.
- FORSLUND, A., AND O. N. SKANS (2006): “Swedish youth labour market policies revisited,” *Vierteljahreshefte zur Wirtschaftsforschung*, 75, 168–185.
- FRANDSEN, B. R. (2009): “Nonparametric identification and estimation of regression discontinuity quantile treatment effects,” *Mimeo*, MIT.
- 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*.
- FRÖLICH, M. (2007): “Regression discontinuity design with covariates,” IZA working paper No. 3024.
- GADDAH, A., AND R. BRAEKERS (2009): “Weak convergence for the conditional distribution function in a Koziol-Green model under dependent censoring,” *Journal of Statistical Planning and Inference*, 139, 930–943.
- GORMLEY, W., AND D. PHILLIPS (2005): “The effects of universal pre-k in Oklahoma: research highlights and policy implications,” *The Policy Studies Journal*, 33, 65–81.
- GUITERAS, R. (2008): “Estimating quantile treatment effects in a regression discontinuity design,” *Mimeo*, MIT.
- 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 KLAUW (2001): “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 69, 201–209.

- HIRANO, K., G. IMBENS, AND G. RIDDER (2003): “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, 71, 1161–1189.
- 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. A., AND L. LEFGREN (2004): “Remedial education and student achievement: a regression-discontinuity analysis,” *The Review of Economics and Statistics*, 86, 226–244.
- JEPSEN, C., P. MUESER, AND K. TROSKE (2009): “Labor-market returns to the GED using regression discontinuity analysis,” *mimeo*, x(x).
- KOENKER, R. (2005): *Quantile Regression*. Cambridge University Press, Cambridge.
- KOENKER, R., AND G. BASSETT (1978): “Regression quantiles,” *Econometrica*, 46, 33–50.
- KOSOROK, M. R. (2008): *Introduction to empirical processes and semiparametric inference*. Springer Verlag.
- 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*.
- LEE, D., AND D. CARD (2008): “Regression discontinuity inference with specification error,” *Journal of Econometrics*, 142, 655–674.
- LEE, D. S., AND T. LEMIEUX (2009): “Regression discontinuity designs in economics,” NBER working paper No. 14723.
- 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.
- MCCRARY, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142, 698–714.
- NEYMAN, J. (1935): “Statistical problems in agricultural experiments,” *Supplement to the Journal of the Royal Statistical Society*, II, 107–180.
- NORDSTRÖMSKANS, O., AND L. LINDQVIST (2005): “Causal effects of subsidized career breaks,” *IFAU Working paper*, p. 17.

- PORTER, J. (2003): “Estimation in the regression discontinuity model,” mimeo, Department of Economics, University of Wisconsin.
- 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.
- RUBIN, D. (1978): “Bayesian inference for causal effects: the role of randomization,” *Annals of Statistics*, pp. 34–58.
- THISTLETHWAITE, D., AND D. CAMPBELL (1960): “Regression-discontinuity analysis: An alternative to the ex post facto experiment,” *Journal of Educational Psychology*, 51, 309–317.
- VAN DER KLAUW, W. (2002): “Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach,” *International Economic Review*, 43, 1249–1287.
- VAN DER KLAUW, W. (2008): “Regression-discontinuity analysis: a survey of recent developments in economics,” *Labour*, 22, 219–245.
- VAN DER VAART, A., AND J. WELLNER (1996): *Weak convergence and empirical processes: with applications in statistics*. Springer Verlag, New York.
- VAN KEILEGOM, I., AND M. G. AKRITAS (1999): “Transfer of tail information in censored regression models,” *The Annals of Statistics*, 27, 1745–1784.
- ÖCKERT, B. (2008): “What’s the value of an acceptance letter? Using admissions data to estimate the return to college,” mimeo, IFAU, Uppsala.