Quantile treatment effects in the regression discontinuity design[☆]

Brigham R. Frandsen [a], Markus Frölich [b], Blaise Melly [c,*]

[a] *Harvard University and Brigham Young University, USA*
[b] *Universität Mannheim and IZA and ZEW, Germany*
[c] *Brown University, USA*

ARTICLE INFO

ABSTRACT

We introduce a nonparametric estimator for local quantile treatment effects in the regression discontinuity (RD) design. The procedure uses local distribution regression to estimate the marginal distributions of the potential outcomes. We illustrate the procedure through Monte Carlo simulations and an application on the distributional effects of a universal pre-K program in Oklahoma. We find that participation in a pre-K program significantly raises the lower end and the middle of the distribution of test scores.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The regression discontinuity design (RDD) was originally developed by Thistlethwaite and Campbell (1960) as a quasi-experimental design for evaluating the impact of an education program in a setting where exposure to a treatment is determined by exceeding some score threshold. 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 and increasing attention in many fields, including labor markets, political economy, health, criminology, environment, and development.[1]

The recent popularity of the RD design appears to be justified in many cases. Black et al. (2007) and Buddelmeyer and Skoufias (2003) compare RD to randomized experiments and find that the RD estimates replicate the experimental results well (see Cook and Wong, 2008, for a summary of studies comparing RD to experiments).

The studies mentioned above and others using the RD design focus on estimating average treatment effects. In many contexts, however, the effect of a treatment on the entire distribution of outcomes is of interest. For example, economists often evaluate the social welfare implications of a policy based on the differences in the distribution of outcomes under various alternatives (Atkinson, 1970). In the field of education (e.g., Leuven et al., 2007; Frölich and Melly, 2010), achievement *disparities* are of large public concern. When analyzing the effects of unemployment insurance on unemployment durations (e.g., Lalive, 2008), the risk of becoming long-term unemployed may be the principal concern. Finally, a zero average effect may mask significant offsetting effects at different points in the distribution, as in the effect of closely-contested unionization elections on employees' earnings (DiNardo and Lee, 2004; Frandsen, 2010).[2]

---

[1] For a summary of recent applications of the RDD, see Cook (2008).

---

[2] 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 RD estimator, a *median* RD estimator can provide more stable estimates when the outcome variable is noisy, e.g. wages or

In this paper, we introduce a procedure to nonparametrically estimate the effects of a potentially endogenous treatment on the distribution of an outcome variable in the RD design. We focus on quantile treatment effects (QTE) as a convenient way to summarize these effects when the outcome is continuous, but more generally our results imply uniformly consistent estimators for any continuous functional of the distribution functions, e.g., the Gini coefficient, the Lorenz curve or distribution treatment effects (see Chernozhukov et al., 2009).[3] The results in this paper apply both to the fuzzy RD design and to the sharp RD design, which we treat as a special case of the fuzzy design. Our procedure is based on local linear estimates of the distribution functions. For an appropriate choice of bandwidth, our estimator is consistent at the $n^{-2/5}$ rate but our results also allow for a sequence of bandwidths that eliminates the asymptotic bias.

Our results are related to the previous theoretical work on the RD design and quantile treatment effects under endogeneity. Hahn et al. (2001) describe this design using the treatment effects framework and formalize the assumptions required to identify average causal effects and provide local linear estimators. Porter (2003) complements their work by considering alternative estimators. Our approach is in the spirit of these papers in using local linear techniques to estimate effects at the threshold. Another approach to estimating distributional effects in the RD context that makes use of local linear techniques is being developed by Baker et al. (2005), and relies on a selection-on-observables identifying assumption at the threshold. This rules out cases where selection into treatment is endogenous even at the threshold of the running variable. Our estimator allows for endogenous treatment selection even in a neighborhood of the threshold, and thus it has an IV interpretation. Frölich (2007) incorporates covariates in a fully nonparametric way and shows that efficiency gains are obtained for mean estimation 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.

As Angrist and Lavy (1999) and Hahn et al. (2001) suggested, the fuzzy RD design leads naturally to instrumental variables (IV) type estimators, where the instrument is an indicator for exceeding a threshold in the running variable. The estimator Hahn et al., develop has an interpretation as a local Wald estimator of a local average treatment effect (LATE). Their insight suggests applying IV quantile treatment effects estimators in order to estimate distributional effects in the RD design.

Two recently developed approaches to IV quantile treatment effects are Chernozhukov and Hansen (2005) and Abadie et al. (2002). These two approaches rely on distinct sets of identifying assumptions, and the interpretations of the estimands differ. An RD quantile treatment effects estimator in the spirit of Chernozhukov and Hansen (2005) is developed by Guiteras (2008). In some contexts, however, the requirement of rank invariance or rank similarity across treatment states in that model may be less desirable than the LATE assumptions of Abadie et al. (2002). In addition, the LATE framework does not require the outcome variable Y to be continuous, such that we can allow masspoints (e.g. of earnings at zero or top coding) or discrete outcome variables. We therefore focus on the LATE framework, although the necessity of controlling for the running (or forcing) variable in the RD design prevents the trivial application of Abadie, Angrist and Imbens' (AAI) estimator.[4]

The remainder of this paper is organized as follows. Section 2 develops the statistical framework. Identification results are established in Section 3, and Section 4 describes the estimation procedure. Section 5 derives the asymptotic distribution for the proposed estimator and discusses inference. We present Monte Carlo simulation results in Section 6. Section 7 applies the procedure to estimate the effect of an Oklahoma universal pre-K program on the distribution of test scores, and Section 8 concludes.

2. Econometric framework

We define causal effects using the potential outcome notation in the framework known as the Neyman–Fisher–Rubin causal model.[5] We are interested in the effect of a binary treatment D on an outcome variable Y. We observe n units, indexed by $i = 1, \ldots, n$, 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 under treatment and no treatment, so the observed outcome is $Y_i = Y_i^0 (1 - D_i) + Y_i^1 D_i$. We do permit arbitrary treatment effect heterogeneity, i.e. there are no restrictions placed on the treatment effects $\delta_i = Y_i^1 - Y_i^0$, and selection into treatment may be endogenous.

The essence of the RD design is the presence of a running variable R which influences the probability of treatment in a discontinuous way when it exceeds some threshold r_0. We define an indicator for exceeding the threshold to be $Z_i = 1 (R_i \geq r_0)$. In the empirical example in Section 7 R_i will be individual i's birthdate and Z_i will be an indicator for meeting an age cutoff for pre-K eligibility. Let unit i's potential treatment status as a function of the running variable be $D_i (r)$, so that observed treatment status is $D_i = D_i (R_i)$. Let the limit (if it exists) of $D_i (r)$ as r approaches r_0 from below be denoted $D_i^0 \equiv \lim_{r \to r_0^-} D_i (r)$ and let the limit (if it exists) from above be $D_i^1 \equiv \lim_{r \to r_0^+} D_i (r)$. Based on the local potential treatment states D_i^0 and D_i^1, we can conceptually classify individuals into one of several mutually exclusive groups, extending the standard concept introduced in Imbens and Angrist (1994). "Local always-takers" (AT) are exposed to treatment whenever the running variable is near the threshold, while "local never-takers" (NT) are never exposed to treatment when the running variable is near the threshold. "Local compliers" (C) are exposed to treatment when the running variable approaches the threshold from above, but not when it approaches from below, and vice versa for "local defiers" (DE). Finally, for "local indefinites," (I) the treatment status as the running variable approaches the threshold from either side does not have a well-defined limit. Formally, these classifications can be expressed as events in a common probability space $(\Omega, \mathcal{F}, P)$:

- Always takers: $AT = \left\{ \omega : D^0 (\omega) = D^1 (\omega) = 1 \right\}$
- Never takers: $NT = \left\{ \omega : D^0 (\omega) = D^1 (\omega) = 0 \right\}$
- Compliers: $C = \left\{ \omega : D^1 (\omega) > D^0 (\omega) \right\}$
- Defiers: $DE = \left\{ \omega : D^1 (\omega) < D^0 (\omega) \right\}$
- Indefinite: $I = \{AT \cup NT \cup C \cup DE\}^c$.

The estimand we primarily consider in this paper is the local quantile treatment effect, or the difference between the marginal distributions of potential outcomes for compliers eval-

---

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 in linear regression models, where the robustness of median regression to outliers was emphasized. This may be particularly relevant for the 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.

[3] For a further discussion on inequality measures, see also Firpo (2008).

[4] See Appendix A for more details on why AAI's approach fails in the RD context.

[5] See Neyman (1935), Fisher (1935) and Rubin (1978).

$$F_{Y^1|C}(y) = \frac{\lim_{r \to r_0^+} E\left[1\left(Y \leq y\right) D | R = r\right] - \lim_{r \to r_0^-} E\left[1\left(Y \leq y\right) D | R = r\right]}{\lim_{r \to r_0^+} E\left[D | R = r\right] - \lim_{r \to r_0^-} E\left[D | R = r\right]} \qquad (2)$$

$$F_{Y^0|C}(y) = \frac{\lim_{r \to r_0^+} E\left[1\left(Y \leq y\right)\left(1 - D\right) | R = r\right] - \lim_{r \to r_0^-} E\left[1\left(Y \leq y\right)\left(1 - D\right) | R = r\right]}{\lim_{r \to r_0^+} E\left[1 - D | R = r\right] - \lim_{r \to r_0^-} E\left[1 - D | R = r\right]} \qquad (3)$$

Box I.

uated at a particular quantile at the threshold level of the running variable

$$\delta_{\text{LQTE}}(\tau) \equiv Q_{Y^1|C,R=r_0}(\tau) - Q_{Y^0|C,R=r_0}(\tau), \qquad (1)$$

where $Q_{Y|C,R=r_0}(\tau)$ is the τ quantile of a random variable Y for the local compliers, i.e. $Q_{Y|C,R=r_0}(\tau) = \inf\left\{u : F_{Y|C,R=r_0}(u) \geq \tau\right\}$. In the following, we suppress explicit conditioning on $R = r_0$ to simplify the notation, i.e. we will write $Q_{Y|C}$ and $F_{Y|C}$. Note that this object reflects the effect of treatment on the distribution, rather than the effect of treatment on any particular individual. Without an additional rank invariance assumption, there is no sense in which (1) represents the treatment effect for a particular individual, since one individual could be at different ranks in the Y^0 and Y^1 distribution. See Koenker (2005) for a discussion of quantile treatment effects and rank invariance with additional references.

3. Identification of treatment effects

Besides those embodied in the notation given in Section 2, we make the following identification assumption.

Assumption I. I1: *RD.* $\lim_{r \to r_0^+} \Pr\left(D = 1 | R = r\right) > \lim_{r \to r_0^-} \Pr(D = 1 | R = r)$.
I2: *Local smoothness.* $F_{Y^d|D^0,D^1,R}\left(y | d^0, d^1, r\right)$ is continuous in r at r_0, for $d^0, d^1 \in \{0, 1\}$. $E\left[D^z | R = r\right]$ is continuous at r_0, for $z \in \{0, 1\}$.
I3: *Monotonicity.* $\lim_{r \to r_0} \Pr\left(D^1 \geq D^0 | R = r\right) = 1$ and $\Pr(\text{Indefinite}) = 0$.
I4: *Density at threshold.* $F_R(r)$ is differentiable at r_0 and $\lim_{r \to r_0} f_R(r) > 0$.

Assumption I1 is the defining feature of the regression discontinuity design: the probability of treatment changes discontinuously at the threshold value of the running variable. In the so-called *sharp* RD design, the difference in the probability of treatment across the threshold is one: treatment status is completely determined by location relative to the threshold. In the sharp design all units are compliers, so the estimand (1) corresponds to the quantile treatment effect for all units at the threshold. In the *fuzzy* RD design the difference in treatment probability is less than one – but still strictly positive – so other factors influence selection into treatment besides the running variable. We focus on the more general fuzzy design, treating the sharp design as a special case.[6]

Assumption I2 is a smoothness condition which, intuitively speaking, ensures that after controlling smoothly for the running variable, differences in the distribution of outcomes on either side of the threshold are due to the change in probability of treatment assumed in Assumption I1.

Assumption I3 is the crucial monotonicity assumption that the response of treatment selection to exceeding the threshold

is monotone.[7] An immediate consequence of this assumption is that the monotonicity condition rules out the existence of defiers and indefinites in a neighborhood around the threshold. Finally, Assumption I4 requires that observations close to r_0 exist.

These assumptions are analogous to Hahn et al.'s (2001) conditions for identifying the local average treatment effect in an RD setting. Assumption I1 here is precisely their RD condition, and Assumption I3 is equivalent to the monotonicity condition in their assumption A3. The smoothness of $F_{Y^d|R}(y|r)$ and $E[D^z|R = r]$ in Assumption I2 are analogous to their assumption A1 and the joint independence condition in their A3. One difference is that Hahn et al. assume only the smoothness of conditional expectations while we require smoothness of the conditional distribution functions because we are identifying quantile treatment effects.

Given our assumptions, at the threshold we can adapt Imbens and Rubin's (1997) and Abadie's (2002) method of identifying counterfactual distributions for compliers. The local quantile treatment effect is then simply the difference between the inferred marginal distributions of the potential outcomes for compliers at a particular quantile. The following lemma shows that the local quantile treatment effect can be written as the horizontal difference between "local Wald ratios", emphasizing the connection with instrumental variables estimation of treatment effects. (The lemma is a special case of Frölich (2007). We provide a simplified proof in Appendix B for notational consistency.)

Lemma 1 (*Identification*). *Under Assumption I the distribution functions of the potential outcomes for the compliers at the threshold $F_{Y^d|C}(y)$ are identified for any y in $\mathbb{R}$ from the joint distribution of (Y, D, R). The expressions for both distribution functions are given in Eqs. (2) and (3) in Box I. The local quantile treatment effect is identified as*

$$\delta_{\text{LQTE}}(\tau) = \inf\left\{u : F_{Y^1|C}(u) \geq \tau\right\} - \inf\left\{u : F_{Y^0|C}(u) \geq \tau\right\}.$$

4. Estimation

The local quantile treatment effect (1) may be consistently estimated in a number of ways. We briefly mention one approach that may appear to be the most obvious, and thereafter spend greater time developing our preferred approach. At first blush, one could think of adapting Abadie et al.'s (2002) IV quantile treatment effects estimator to estimate local quantile treatment effects, treating $Z_i = 1 (R_i \geq r_0)$ as a binary instrument for D_i. This approach requires ignoring any direct effects of R_i other than through Z_i, introducing a sort of omitted variables bias in finite samples and leading to a slower rate of convergence, as discussed in Appendix A. The Monte Carlo results below show that the

---

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

[7] There are several settings in which monotonicity holds automatically, including when non-compliance is one-sided, with either no treatment below the threshold, or 100% treatment above the threshold. Other settings which imply monotonicity are latent index models of selection.

combined bias from ignoring the running variable and boundary effects can be large.

We overcome these bias problems by using local linear techniques to estimate the distribution of the potential outcomes for compliers at the discontinuity threshold. This estimator is not subject to bias from ignoring the running variable and automatically corrects for boundary effects (Fan, 1992). The key task is nonparametric quantile estimation at a boundary, an estimation problem considered by Yu and Jones (1998) in a setting without an endogenous treatment. Their preferred technique, which we adapt here, estimates conditional quantiles by inverting local linear estimates of the conditional distribution function.

Given differentiability of the conditional distributions in a neighborhood of the threshold, a consistent estimator for the local quantile treatment effect, then, is the (horizontal) difference between local linear estimates of the conditional distribution functions (2) and (3) at a particular quantile:

$$\hat{\delta}_{LQTE}(\tau) = \hat{Q}_{Y^1|C}(\tau) - \hat{Q}_{Y^0|C}(\tau),\tag{4}$$

where

$$\hat{Q}_{Y^1|C}(\tau) = \inf\left\{a : \hat{F}_{Y^1|C}(a) \geq \tau\right\},$$

$$\hat{Q}_{Y^0|C}(\tau) = \inf\left\{b : \hat{F}_{Y^0|C}(b) \geq \tau\right\},$$

and $\hat{F}_{Y^1|C}(y)$, $\hat{F}_{Y^0|C}(y)$ are consistent estimators of (2) and (3).

We estimate the sample analogs of the "local Wald ratios" in (2) and (3)

$$\hat{F}_{Y^1|C}(y) = \frac{\hat{m}^+_{1(Y \leq y)D}(r_0) - \hat{m}^-_{1(Y \leq y)D}(r_0)}{\hat{m}^+_D(r_0) - \hat{m}^-_D(r_0)}\tag{5}$$

$$\hat{F}_{Y^0|C}(y) = \frac{\hat{m}^+_{1(Y \leq y)(1-D)}(r_0) - \hat{m}^-_{1(Y \leq y)(1-D)}(r_0)}{\hat{m}^-_D(r_0) - \hat{m}^+_D(r_0)},\tag{6}$$

where $m^+_W(r_0)$ denotes a local linear estimate of the conditional expectation of a generic random variable W at $R = r_0$ from above; i.e. $m^+_W(r_0)$ is estimated as the value of a that solves

$$\arg\min_{a,b} \sum_{j:R_j \geq r_0} (W_i - a - b(R_i - r_0))^2 \cdot K\left(\frac{R_i - r_0}{h}\right).\tag{7}$$

The expression for $m^-_W(r_0)$ is analogous, using only the observations from below r_0.

Rather than separately estimating the four conditional means in (5), $\hat{F}_{Y^1|C}(y)$ can be estimated in one step via local linear weighted two-stage least squares (2SLS), as suggested by Imbens and Lemieux (2008) for mean estimation. Define

$$X_i = \begin{pmatrix} 1 \\ 1(R_i < r_0)(R_i - r_0) \\ 1(R_i \geq r_0)(R_i - r_0) \end{pmatrix}.$$

Then estimating the equation

$$1(Y_i \leq y)D_i = \alpha D_i + X'_i\beta + \varepsilon_i\tag{8}$$

by weighted 2SLS with $Z_i = 1(R_i \geq r_0)$ as the excluded instrument for D_i and weights $K\left(\frac{R_i - r_0}{h}\right)$ results in a coefficient estimate $\hat{\alpha}$ algebraically identical to the nonparametric estimator $\hat{F}_{Y^1|C}(y)$ given by (5). The estimator $\hat{F}_{Y^0|C}(y)$ can be similarly obtained by weighted 2SLS, replacing D_i by $(1 - D_i)$ on both sides of (8). We describe a data-driven choice of bandwidths in Appendix C.

4.1. Finite-sample refinements

In finite samples the conditional cdfs estimated by (5) and (6) are non-monotonic step functions. This poses problems for the inversion of the cdfs to obtain the quantile functions. We follow here the suggestion of Chernozhukov et al. (2010) and monotonize the estimated distribution functions by re-arrangements. This does not affect the asymptotic properties of the estimator but allows it to be inverted. This procedure consists of a sequence of closed-form steps and is fast.[8]

For some applications where the discontinuities are especially large it helps to smooth the problematic indicator function on the left hand side of (8). Yu and Jones (1998) suggest smoothing "in the y-direction", which in our setting amounts to estimating $\hat{F}_{Y^1|C}(y)$ by weighted 2SLS on the following regression equation

$$\Omega\left(\frac{y - Y_i}{h_Y}\right)D_i = \alpha D_i + X'_i\beta + \varepsilon_i,\tag{9}$$

where $\Omega(\cdot)$ is a differentiable distribution function, e.g. the normal cdf. By construction, the estimated distribution function is continuous in finite samples. We follow Yu and Jones (1998) and choose $h_Y \ll h$ so that the asymptotic distribution of the estimator does not depend on the exact value of the bandwidth h_Y used to smooth in the y-direction. As discussed in Appendix C, we also use their operational rules of thumb for choosing the bandwidths.

4.2. Special case: sharp and semi-fuzzy designs

The identification results and estimation procedure described above assume a fuzzy RD design, but they apply equally well to the special cases of sharp and semi-fuzzy designs, where treatment status may be constant on one or both sides of the threshold. Identification is more straightforward, since the monotonicity condition, Assumption I3, is automatically satisfied in the sharp and semi-fuzzy designs. In the sharp design case, the procedure still consists of inverting estimated distribution functions on each side of the threshold, but since now $D_i = Z_i$, the distribution functions can be obtained by simple weighted least squares estimation of (8).

4.3. Including additional covariates

Perhaps the main advantage of the canonical RD design as described above is that identification does not require controlling for additional covariates, a point made by Hahn et al. (2001). However, empirical analysis often includes additional covariates for any of several reasons: to establish identification, to explore how the parameter of interest varies in subgroups, to test the sensitivity of results to incorporating different sets of covariates, to separate direct from indirect effects or to increase precision.

If Assumption I holds conditionally on covariates X, then the identification results stated above now apply immediately to the treatment effect conditionally on X. In the case of discrete covariates, comparing estimation results for different values of the covariates may be of interest, for example, comparing achievement gains for students of different socio-economic backgrounds. Specific values of a continuous covariate may also be conditioned

---

[8] We could also monotonize the intermediate estimators $\hat{m}^+_{1(Y \leq y)D}(r_0)$, $\hat{m}^-_{1(Y \leq y)D}(r_0)$, $\hat{m}^+_{1(Y \leq y)(1-D)}(r_0)$ and $\hat{m}^-_{1(Y \leq y)(1-D)}(r_0)$, for instance using the procedure discussed in Chernozhukov et al. (2010) or the estimators suggested in Hall et al. (1999). This could improve the small sample behavior of the estimator but is neither sufficient nor necessary to guarantee the monotonicity of the distribution functions $\hat{F}_{Y^1|C}(y)$ and $\hat{F}_{Y^0|C}(y)$.

upon, but this would require another smoothing parameter, and a curse of dimensionality would apply, slowing the rate of convergence and reducing power.

In many situations we are however more interested in the unconditional effect, that is, the effect for all compliers irrespective of their value of X. In this case, covariates – discrete or continuous – could be incorporated for sensitivity analysis and precision gains, as Frölich (2007) shows for the case of mean estimation. Because all covariates but the running variable are integrated out, the curse of dimensionality is avoided and the rate of convergence is not affected. Frölich's (2007) results can be extended to the quantile treatment effects but this is outside of the scope of this paper.

5. Asymptotic distribution theory and inference

In this section, we derive the limiting distribution for the local quantile treatment effects estimator, (4), obtained via inverting estimated distribution functions. The regularity assumptions we employ are analogous to those used by Hahn et al. (1999). Define

$$\Delta p = \lim_{r \to r_0^+} E[D|R = r] - \lim_{r \to r_0^-} E[D|R = r],$$

which represents the change in the probability of treatment at the threshold and corresponds to the limiting fraction of compliers. The following assumption comprises the regularity conditions under which the limiting distribution of our estimator is derived.

Assumption E. E1 The left and right limits of the functions $E[1(Y \leq y)(1 - D)|R = r]$, $E[1(Y \leq y)D|R = r]$, and $E[D|R = r]$ are twice left- and right-continuously differentiable with respect to r at r_0 with second derivative Hölder continuous in a left and right ε-neighborhood of r_0, respectively, and are uniformly bounded in $y \in \mathcal{Y}$, where $\mathcal{Y}$ is a compact subset of $\mathbb{R}$, for some $\varepsilon > 0$.
E2 Δp is strictly positive.
E3 The density $f_R(r)$ is bounded away from zero and infinity in a neighborhood of r_0.
E4 $K(\cdot)$ is Borel measurable, bounded, continuous, symmetric, nonnegative-valued with compact support, and integrates to one.
E5 Bandwidth conditions: $nh \to \infty$ and $\sqrt{nh}h^2 \to \gamma < \infty$.

Assumption E1 ensures that the underlying conditional distributions of potential outcomes are sufficiently smooth at the discontinuity. E2 requires that the probability of treatment changes discretely at the threshold. It corresponds to some kind of strong instrument assumption in an instrumental variables framework. E3 ensures that the distribution of the running variable is well behaved near the threshold. E4 imposes standard conditions on the kernel function. Assuming that the kernel is symmetric, bounded, non-negative and integrates to one is not strictly necessary and only imposed in order to provide simpler expressions for the bias and variance. (If a kernel is used that does not satisfy these properties, the expressions for the bias and variance need to be recalculated.) E5 characterizes the sequence of bandwidths, and thus determines the rate of convergence of the estimator. For $\gamma > 0$, squared bias and variance are of the same order, and the estimator minimizes the asymptotic mean squared error, converging at the $n^{-\frac{2}{5}}$ rate. If we choose a bandwidth such that $\gamma = 0$, the bias vanishes asymptotically, but the convergence rate is slower.

Asymptotic normality of the quantile treatment effect estimator (4) follows from the weak convergence as a process of the complier cdf estimators, (5) and (6), and the functional delta method (van der Vaart, 1998). All proofs are given in Appendix B. We summarize the asymptotic results first for the distribution functions of the potential outcomes and thereafter for the quantile functions.

Theorem 2 (*Limit Distribution for Distribution Functions*). *Under Assumptions* E *and* I *the estimators of the compliers' distribution functions jointly converge in law to the following tight Gaussian processes*

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

in $\ell^\infty(\mathcal{Y})$, *where* $y \longmapsto Z_j(y)$ *have mean functions*

$$b_j(y) = \frac{\gamma \lambda'_K}{\Delta p}\left(\frac{\partial^2 m^+_{1(Y \leq y)(D+j-1)}}{\partial r^2} - F_{Y^j|C}(y)\frac{\partial^2 m^+_D}{\partial r^2}\right.$$
$$\left. - \frac{\partial^2 m^-_{1(Y \leq y)(D+j-1)}}{\partial r^2} + F_{Y^j|C}(y)\frac{\partial^2 m^-_D}{\partial r^2}\right),$$

where λ'_K *is a constant that depends on the kernel function,*[9] $\frac{\partial^2 m^+_W}{\partial r^2} = \lim_{r \to r_0^+}\frac{\partial^2 E[W|R=r]}{\partial r^2}$ *for a generic random variable* W *and* $\frac{\partial^2 m^-_W}{\partial r^2}$ *is the analogous left limit function. The covariance functions are, for* $j, k \in \{0, 1\}$,

$$v_{j,k}(y, \tilde{y}) = \frac{\lambda_K}{f_R(r_0)(\Delta p)^2}\left(\omega^+_{j,k}(y, \tilde{y}) + \omega^-_{j,k}(y, \tilde{y})\right)$$

where λ_K *is a constant that depends on the kernel function, and*

$$\omega^+_{j,k}(y, \tilde{y}) = \lim_{r \to r_0^+}\text{Cov}\{(D + j - 1)\left(1(Y \leq y) - F_{Y^j|C}(y)\right),$$
$$(D + k - 1)\left(1(Y \leq \tilde{y}) - F_{Y^k|C}(\tilde{y})\right)|R = r\}$$

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

This result is of interest in its own right, as it is the basis of deriving the limiting distribution of any parameter that is a function of the counterfactual distributions, including distribution treatment effects, counterfactual densities, and comparisons of stochastic dominance. It also provides the basis for deriving the limiting distribution of the quantile treatment effects process below.

For deriving the above limit processes of the distribution function estimators and any functional thereof, e.g. the Gini coefficient or other inequality measures (see Frölich and Melly, 2010), we can permit Y to be discrete or continuous or mixed discrete–continuous, e.g. continuous with some mass points. If Y is not continuous, however, the inverse of the cdf is not uniquely defined everywhere. This does not pose a problem for *identifying* quantile treatment effects, with quantiles defined as $Q_Y(\tau) = \inf\{u : F_Y(u) \geq \tau\}$, but the asymptotic properties of the *estimated* QTE are different when Y is not continuous. In the following, we derive the asymptotic properties for the quantile treatment effects estimator for the case where Y is continuous, making an additional assumption on the uniqueness of the quantiles of the potential outcomes.

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

With this additional assumption, we establish the limiting process for the QTE estimator.

---

[9] The exact formula can be found in Appendix B. $\lambda'_K = -\frac{11}{190}$ for the Epanechnikov kernel function and $-\frac{1}{12}$ for the uniform kernel with support $[-1, 1]$. $\lambda_K = \frac{56832}{12635}$ for the Epanechnikov kernel function and 4 for the uniform kernel with support $[-1, 1]$.

Theorem 3 (*Limit Distribution for Quantile Functions*)**.** *Under Assumptions* E, Q *and* I *the estimators* $\hat{Q}_{Y^0|C}(\tau)$ *and* $\hat{Q}_{Y^1|C}(\tau)$ *jointly converge to the following Gaussian processes:*

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

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

Corollary 4 (*Limit Distribution for Quantile Treatment Effects*)**.** *Under Assumptions* E, Q *and* I, *the local quantile treatment effects estimator converges to the following Gaussian process*

$$\sqrt{nh}\left(\hat{Q}_{Y^1|C}(\tau) - \hat{Q}_{Y^0|C}(\tau) - \delta_{\mathrm{LQTE}}(\tau)\right) \implies W^1(\tau) - W^0(\tau)$$

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

A corollary of this result is that the quantile treatment effects estimator for any finite set of indices $\{\tau_k\}$ has a jointly normal limiting distribution. Inference may be based on consistent estimates of the asymptotic variance and the asymptotic normality.

Consistent estimation of the asymptotic variance requires consistent estimates of the quantities in the variance formulas in Theorem 3 and Corollary 4. The conditional expectations in the variance formulas can be estimated via local linear regression. The conditional density $f_{Y^1|C}$ can be estimated via weighted 2SLS on the following equation:

$$\frac{1}{h_Y}\phi\left(\frac{y - Y_i}{h_Y}\right) D_i = \alpha D_i + X_i'\beta + \varepsilon_i,$$

where $\phi(\cdot)$ is a density function, e.g. the normal density. The density $f_{Y^0|C}$ can be estimated analogously. Finally, $f_R(r_0)$ can be estimated using, for example, Fan et al.'s (1996) local linear density estimator. After constructing consistent estimates of the asymptotic variance from these intermediate consistent estimates, hypothesis tests or confidence intervals based on the normal distribution will be consistent.

6. Monte Carlo results

To illustrate the practical performance of our estimation procedure, in this section we present the results of Monte Carlo simulations. The primitive of the model underlying the simulations is the joint distribution of (Y^0, Y^1, D, R), which we specify as follows: $R \sim N\left(0, \sigma_R^2\right)$ and $Y^0 = R + \varepsilon_0, Y^1 = Y^0 - \varepsilon_1$ and $D = 1\left(Y^1 - Y^0 + \alpha \cdot 1(R \geq 0) \geq \varepsilon_D\right)$, and the disturbance terms $(\varepsilon_0, \varepsilon_1, \varepsilon_D)$ are jointly normal and independent with mean zero and variances σ_0^2, σ_1^2, and σ_D^2, respectively. This model could be interpreted as a simple Roy model of selection on gains where exceeding the threshold $r_0 = 0$ reduces the gross cost of treatment ε_D by α. It exhibits the key features of the RD design with heterogeneous treatment effects. Note that the average treatment effect (ATE) is zero. In this model, the complier group is $C = \{0 < \varepsilon_D + \varepsilon_1 \leq \alpha\}$, and for positive α, the local average treatment effect (LATE) is negative.

Two key parameters affecting the performance of the estimators are the sample size, N, and the change in the probability of treatment at the threshold, Δp. The greater the magnitude of Δp,

the higher the precision of the estimator. In the simulation model, the parameter α controls this change in probability:

$$\Delta p = \Phi\left(\frac{\alpha}{\sqrt{\sigma_1^2 + \sigma_D^2}}\right) - \Phi(0).$$

We illustrate the performance of the estimators for several scenarios which are broadly representative of actual empirical examples. For each scenario, we perform 500 repetitions with parameter values $\sigma_R = \sigma_0 = \sigma_1 = \sigma_D = 1$, with $\alpha = 0.5$ ("small Δp") or $\alpha = 3$ ("large Δp") and a sample size of either $N = 10,000$ ("small N") or $N = 100,000$ ("large N"). To call a sample size of 10,000 "small" of course reflects the demands on the data nonparametric estimators require in general. We use a uniform kernel with bandwidths chosen as described in Appendix C.

We compare our local linear approach to estimating quantiles of the compliers' potential outcomes to a "local constant" approach described in Appendix A that applies instrumental variables quantile treatment effects estimation using kernel weights to narrow in on the threshold, ignoring the running variable (Abadie LQTE).[10]

The first simulation scenario, "large N, large Δp", represents the most favorable conditions for the estimation procedures. We set $\alpha = 3$, which implies a jump in the probability of treatment at the threshold of about 48%, and we use a sample size of 100,000. This change in the probability and the sample size are on the order of those found in several recent RD studies, including Matsudaira (2008) and Jacob and Lefgren (2004). In this and all scenarios, bandwidths for the preferred approach were chosen in each repetition according to the procedure outlined in Appendix C. Typical values for the bandwidth were in the range 0.15–0.3. The bandwidths for the local constant approach were chosen to minimize the simulated mean squared error. While this choice is infeasible in practice, it gives an upper bound on how well a local constant approach might perform relative to local linear approaches.

Fig. 1 shows the results of the simulation under this scenario for RD LQTE. The figure shows the average point estimate for each quantile index, as well as the pointwise (in the quantile index) 90% confidence interval spanned by the fifth and 95th percentile estimate. The figure shows that the bias is very small, despite the fact that the estimator consists of nonlinear functions of estimated quantities. The estimates also appear to be quite precise, although the simulations are not calibrated to any particular economic context to give the scale meaning. By way of comparison, Fig. 2 illustrates the performance of applying the locally constant Abadie LQTE approach to the RD setting. The confidence intervals are somewhat wider, and the estimates are substantially biased. Thus in terms of bias and variance, the local linear RD LQTE approach appears to do strictly better, although for samples this large and discontinuities of this size the difference is not extremely large.

The second simulation scenario, "small N, large Δp" illustrates the implications for the performance of the estimators when the sample is smaller. The change in the probability of treatment remains at 48%, but we use a sample size of 10,000. This corresponds roughly to the empirical application in Section 7 based on Gormley et al. (2005). Fig. 3 shows that the confidence intervals for RD LQTE are substantially wider than for the large N case, but the bias remains negligible. Fig. 4 shows that Abadie LQTE, on the

---

[10] Although we refer to the application of Abadie et al.'s (2002) IV quantile treatment effects estimator to the RD setting using the shorthand "Abadie LQTE", we emphasize those authors did not propose applying their estimator to the RD design.
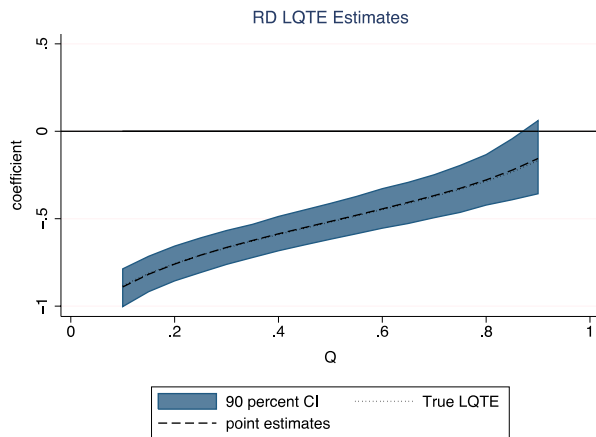
Fig. 1. RD LQTE Monte Carlo results: large *N*, large △*p*. The figure shows point estimates and confidence intervals from a Monte Carlo simulation of RD LQTE with 500 repetitions, a sample size of 100,000, and a discontinuity in the probability of treatment at the threshold of 48%.
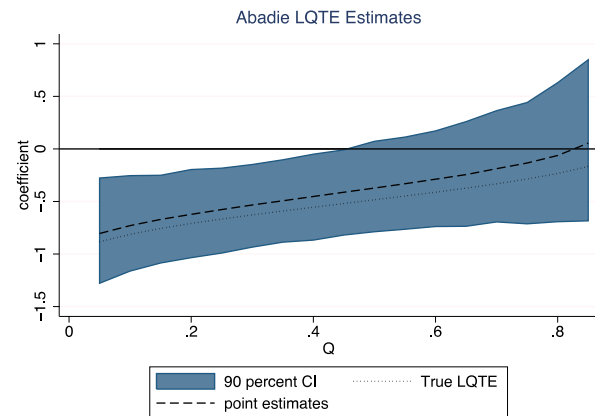

Fig. 2. Abadie LQTE Monte Carlo results: large *N*, large △*p*. The figure shows point estimates and confidence intervals from a Monte Carlo simulation of kernel-weighted Abadie quantile regression with 500 repetitions, a sample size of 100,000, and a discontinuity in the probability of treatment at the threshold of 48%.
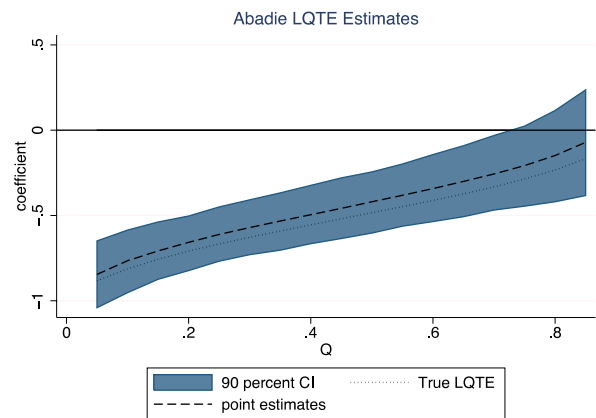

Fig. 3. RD LQTE Monte Carlo results: small *N*, large △*p*. The figure shows point estimates and confidence intervals from a Monte Carlo simulation of RD LQTE with 500 repetitions, a sample size of 10,000, and a discontinuity in the probability of treatment at the threshold of 48%.

Fig. 4. Abadie LQTE Monte Carlo results: small *N*, large △*p*. The figure shows point estimates and confidence intervals from a Monte Carlo simulation of kernel-weighted Abadie quantile regression with 500 repetitions, a sample size of 10,000, and a discontinuity in the probability of treatment at the threshold of 48%.
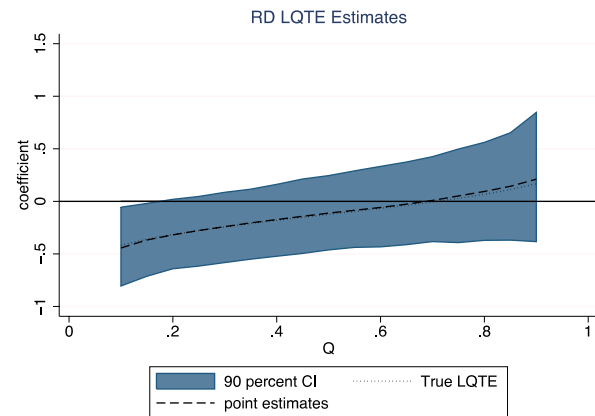

Fig. 5. RD LQTE Monte Carlo results: large *N*, small △*p*. The figure shows point estimates and confidence intervals from a Monte Carlo simulation of RD LQTE with 500 repetitions, a sample size of 100,000, and a discontinuity in the probability of treatment at the threshold of 14%.
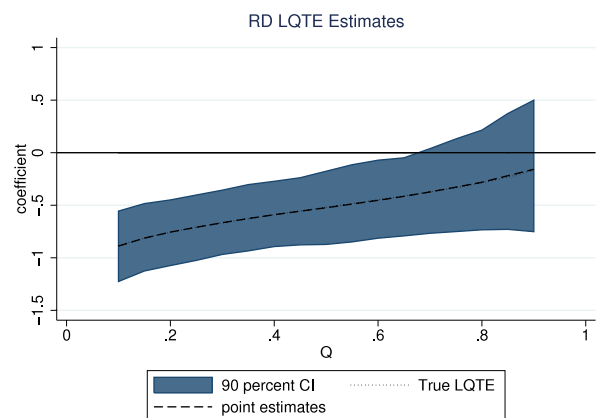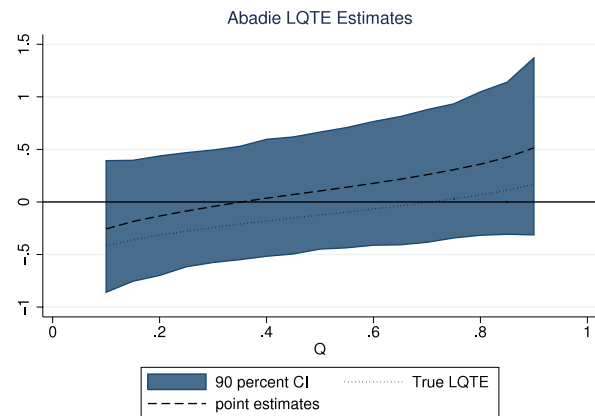

Fig. 6. Abadie LQTE Monte Carlo results: large *N*, small △*p*. The figure shows point estimates and confidence intervals from a Monte Carlo simulation of kernel-weighted Abadie quantile regression with 500 repetitions, a sample size of 100,000, and a discontinuity in the probability of treatment at the threshold of 14%.

other hand, has confidence intervals around 30% wider still than RD LQTE, and the bias is also larger than for the large *N* case.

Finally, the third scenario, "large *N*, small △*p*" preserves the large sample size of 100,000, but sets $\alpha = 0.5$, implying a change in the probability of treatment at the threshold of about 14%. Fig. 5 shows the results from this scenario for RD LQTE. Despite the

large sample size in this scenario, the smaller △*p* results in the widest confidence intervals for any of the simulation scenarios we consider. The bias, however, remains negligible, even for a much smaller △*p*. As Fig. 6 shows, the confidence intervals for applying Abadie LQTE to RD are also widest for this scenario, and the bias is the greatest as well. Thus even for relatively large sample sizes,
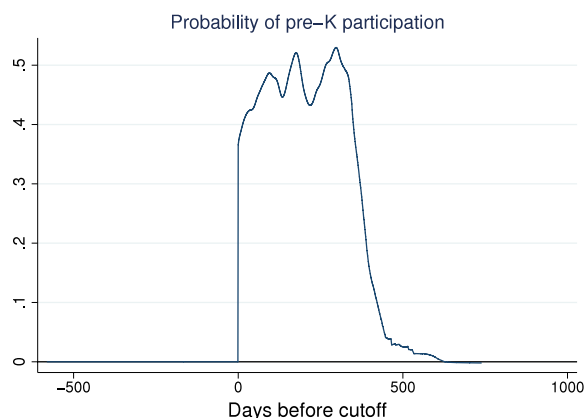
Fig. 7. The figure plots the probability of attending TPS pre-K in 2002–2003 as a function of birthdate relative to cutoff.
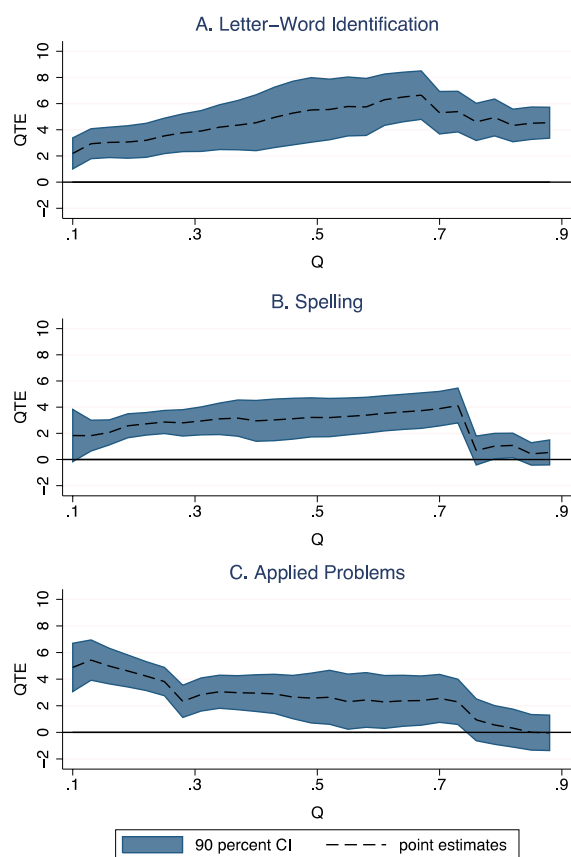

Fig. 8. The figure plots point estimates and 90% confidence intervals (based on the asymptotic variance) for the effect of TPS pre-K participation on the distribution of scores on three Woodcock–Johnson subtests. A uniform kernel was used and bandwidths for each quantile index were chosen as described in Appendix C. The estimated bandwidths were on the order of 30–40 days.

a small jump at the threshold in the probability of treatment can result in significant bias in the local constant approach.

Not surprisingly, the precision of the estimator was best when the sample size and/or the discontinuity was large. In all cases, however, the bias was minimal. The simulations also highlight that the local linear approach perform unambiguously better than an approach ignoring the effects of R within a window around the discontinuity.

The simulations also confirmed that the approximation given by the asymptotic theory seems to perform well. When we increased the sample size from 10,000 to 100,000, the length of the confidence intervals reduced to 40% of the previous length,

which is in line with $10^{-\frac{2}{5}}$ as predicted by the asymptotic theory. Furthermore, according to asymptotic theory the length of the confidence interval should also be inversely proportional to Δp, which is also what we found in the simulations when comparing the "large N, large Δp" design ($\Delta p = 48\%$) to the length of the confidence intervals in the "large N, small Δp" design ($\Delta p = 14\%$).

7. Application: effects of universal pre-K

In this section, we apply the RD quantile treatment effects procedure to an example from the literature which will both illustrate how the procedure might be applied to real-life questions, as well as point out some challenges faced by nonparametric estimation of distributional effects.

Interventions designed to improve educational performance are one setting in which distributional effects may be very important to policy makers. One such policy that specifically targets the lower end of the distribution is the introduction of universal pre-K programs. Gormley et al. (2005) use a regression discontinuity design to analyze an Oklahoma universal pre-K program, and find significant positive effects on average test scores measuring cognitive development along a variety of dimensions. By conditioning on various socio-economic status indicators, they find indirect suggestive evidence that the program also has positive effects on the lower end of the distribution. The quantile treatment effects estimator developed in this paper allows direct investigation of the effect of the policy on the lower end of the distribution.

Oklahoma introduced a universal pre-K program for four-year-olds in 1998, and by 2002–2003 (the period we analyze) 91% of the state's school districts were participating, including Tulsa Public Schools (TPS), the largest district in the state, and the district from which our sample is drawn.

A child's participation in the pre-K program is voluntary (on the part of the parents), but is subject to a birthday cutoff eligibility rule. Children who had turned four years old by September 1, 2002 were eligible for the program, while younger children were not. Fig. 7 shows the discontinuity in probability of treatment that the eligibility rule induced. Because the participation among children who missed the cutoff is essentially nil, local treatment effects in this setting correspond to the effect of treatment on the treated.

At the start of the 2003–2004 school year, all incoming kindergartners and TPS pre-K participants were given the Woodcock–Johnson Achievement Test, a nationally normed test that has been widely used in studies of early education. Treated students are those who participated in a TPS pre-K program the previous year.

We use a sample of 4710 incoming TPS kindergartners and pre-K participants. The dataset includes exact birthdate, an indicator for participation in TPS pre-K the previous year (the treatment variable), and scores on the three Woodcock–Johnson subtests: Letter-word, Spelling, and Applied Problems.

Using a uniform kernel and bandwidths chosen by the plug-in method described in Appendix C, we estimated the local treatment effects of attending the Pre-K program on the three subtests of the Woodcock-Johnson test. The estimated bandwidths were on the order of 30–40 days. The estimated local quantile treatment effects of TPS pre-K programs on scores on the three subtests of the Woodcock-Johnson tests are plotted in Fig. 8. The figure shows the point estimates for each quantile index, as well as pointwise (in the quantile index) 90% confidence intervals based on the asymptotic distribution.

Panel A shows a relatively precisely estimated two to four point (around 80% of a standard deviation) effect on the lower end of the distribution of the Letter-Word Identification score. The estimated effect on the middle of the distribution is a somewhat larger five

to eight points, but less precisely estimated, and the effect on the upper end of the distribution is estimated to be close to four points. Effects on the distribution of the Spelling score are plotted in panel B. Similar to the Letter-Word test, the effect on the lower end of the distribution is a relatively precisely estimated two to four points. The effect above the 75th percentile is smaller and not significantly different from zero. Panel C shows the effects of pre-K participation on the Applied Problems score. The point estimates are largest and most precisely estimated for the bottom end of the distribution, and the point estimates of the effects for the top of the distribution are smaller and not significantly different from zero. For reference, the local average treatment effects (LATE) are estimated to be 3.66 for Letter-Word Identification, 1.93 for Spelling, and 3.44 for applied problems.[11] An application of the local constant, Abadie-weighted approach (not reported) yields a similar pattern across quantiles, but with point estimates higher by about two tenths of a point, which is consistent with the bias observed in the simulations in Section 6.

The estimation results imply that universal pre-K in Oklahoma succeeded in significantly raising the lower end of the distribution of test scores, especially for the Applied Problems subtest. These results are consistent with Gormley et al.'s (2005) findings that point estimates of average effects were larger for children from potentially disadvantaged socio-economic groups. These results are subject to the caveat that they measure the net effect of participating in a TPS pre-K program versus alternatives parents might have chosen in the absence of the program. The alternatives might have been different for children at different points in the distribution, and thus we cannot draw conclusions about the gross impact of universal pre-K programs on the distribution of outcomes. An additional caveat is that these results reflect the short-term effect. It is possible that children who did not participate may catch up over time, although evidence from the Perry Preschool Study (Schweinhart et al., 1993; Anderson, 2008) suggests that there may be significant long-term impacts of pre-K programs.

8. Conclusion

In this paper, we describe how the regression discontinuity design can be used to evaluate the impact of endogenous treatments on the entire distribution of outcome variables. We show that both potential outcome distributions are identified for the population affected by the discontinuity. We introduce estimators for these two distribution functions and show their joint convergence to continuous Gaussian processes. We also consider 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. Monte Carlo simulations confirm that the bias of the approach we suggest is small.

An application of the procedure to estimating the effects of an Oklahoma universal pre-K program across the distribution of test scores shows that the lower end of the distribution is significantly raised, while estimates at the top of the distribution are smaller and less precise. Other possibilities for applying the methodology are numerous, and include the study of remedial education programs by Jacob and Lefgren (2004) and Matsudaira (2008), the study of the UI Worker Profiling and Reemployment Services program by Black et al. (2003), and the effect of unions on wages by DiNardo and Lee (2004).

Appendix A. Relationship to Abadie et al. (2002)

In this appendix, we discuss the relationship of our estimator to an approach based on adapting Abadie et al. (2002) to an RD setting. (However, note that they never proposed to use their method for an RD setting.) Our Assumption I is analogous to those required for Abadie et al.'s (2002) local quantile treatment effects estimator. Instead of independence between an instrument and potential outcomes and potential treatment status, we make continuity assumptions on the distribution of potential outcomes and potential treatment status. The LATE first stage assumption is replaced by the analogous RD assumption that the probability of treatment jumps discretely as the running variable hits the threshold value. Assumption I3, local monotonicity, is directly analogous to the monotonicity assumption in the LATE framework.

The most striking difference between our assumptions and Abadie et al.'s (2002) assumptions is the absence here of the "Non-trivial assignment" condition which they require. Indeed, the principal challenge of applying Abadie et al.'s (2002) quantile treatment effects estimator in an RD setting is that the non-trivial assignment condition fails here, since conditional on the running variable, the "instrument", $Z = 1\,(R \geq r_0)$, is deterministically either zero or one.

An adaptation of Abadie et al.'s (2002) IV quantile treatment effects estimator to estimating the local quantile treatment effect would give the following. This estimator treats $Z = 1\,(R \geq r_0)$ as a binary instrument for D and combines kernel weights which narrow in on the threshold with "complier finding" weights:

$$\kappa_v = 1 - \frac{D\,(1 - E\,[Z|Y, D])}{1 - E\,[Z]} - \frac{(1 - D)\,E\,[Z|Y, D]}{E\,[Z]}.$$

These weights can be estimated via kernel regression centered on $R = r_0$ to form $\hat{\kappa}_{vi}$, and the estimator would take the form:

$$\left(\hat{a}, \hat{\delta}_{\text{ALQTE}}\,(\tau)\right)$$
$$= \arg\min_{a,d} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau\,(Y_i - a - dD_i) \cdot \hat{\kappa}_{vi} \cdot K\left(\frac{R_i - r_0}{h_n}\right), \qquad (10)$$

where $K\,(\cdot)$ is a kernel function, $\rho_\tau\,(u) = (\tau - 1\,(u < 0))\,u$ and h_n is a smoothing parameter. Strictly speaking, all expectations in the definition of κ_v should be taken conditional on the running variable, R. In that case a zero would appear in one of the denominators, since Z is a deterministic function of R, violating the so-called non-trivial assignment condition which requires that there be variation in the instrument conditional on included variables. However, as the sample size grows, in the limit we are conditioning on $R = r_0$, and so there is no longer any need to include R as a regressor, and thus the technique can be applied. However, in finite samples, there will be variation in R over the window defined by the kernel weights. This leads to a

---

sort of omitted variables bias, since the instrument Z may not be independent of the potential outcomes when we do not condition on R. This finite sample bias is analogous to the inflated bias of locally constant kernel regression estimators, noted by Fan (1992). For the case of quantile regression, Yu and Jones (1997) show that the bias in this approach is proportional to the slope of the conditional quantile function estimated. Another source of bias in this approach stems from implicitly estimating quantile functions at a boundary,[12] leading to a convergence rate of $n^{-\frac{1}{5}}$ instead of the rate $n^{-\frac{2}{5}}$. The Monte Carlo results of Section 6 showed that the combined bias from ignoring the running variable and boundary effects can be quite large.

Appendix B. Proofs of lemmas and theorems

B.1. Proofs for identification

The main identification result, Lemma 1, follows from a special case of Frölich (2007). We provide here a simplified proof to maintain consistency with the notation and assumptions in this paper. We show that

$$\lim_{r \to r_0^+} E\left[1\left(Y \leq y\right)D|R=r\right] - \lim_{r \to r_0^-} E\left[1\left(Y \leq y\right)D|R=r\right]$$
$$= E\left[1\left(Y^1 \leq y\right)|C, R=r_0\right]\Pr\left(C|R=r_0\right)$$

and that

$$\lim_{r \to r_0^+} E\left[D|R=r\right] - \lim_{r \to r_0^-} E\left[D|R=r\right] = \Pr\left(C|R=r_0\right).$$

Combining these pieces and with analogous derivations when replacing D with $1-D$ the formulas of Lemma 1 are obtained.

First we show that the probability of a local complier is identified as

$$\lim_{r \to r_0^+} E\left[D|R=r\right] - \lim_{r \to r_0^-} E\left[D|R=r\right]$$
$$= E\left[D^1|R=r_0\right] - E\left[D^0|R=r_0\right]$$
$$= E\left[D^1 - D^0|R=r_0\right]$$
$$= \Pr\left(D^1 > D^0|R=r_0\right) = \Pr\left(C|R=r_0\right).$$

The first equality follows from the definition of D^1 and D^0, the continuity Assumption I2, and the Assumption I3 that indefinites are of zero measure. The third equality follows from the fact that $D^1 - D^0$ equals one when $D^1 > D^0$ and zero when $D^1 = D^0$, and that by the Assumption I3: $\Pr\left(D^1 < D^0|R=r_0\right) = 0$. The final equality reflects the definition of a complier.

With similar derivations we obtain

$$\lim_{r \to r_0^+} E\left[1\left(Y \leq y\right)D|R=r\right] - \lim_{r \to r_0^-} E\left[1\left(Y \leq y\right)D|R=r\right]$$
$$= \lim_{r \to r_0^+} E\left[1\left(Y^1 \leq y\right)D|R=r\right] - \lim_{r \to r_0^-} E\left[1\left(Y^1 \leq y\right)D|R=r\right]$$
$$= E\left[1\left(Y^1 \leq y\right)D^1|R=r_0\right] - E\left[1\left(Y^1 \leq y\right)D^0|R=r_0\right]$$
$$= E\left[1\left(Y^1 \leq y\right)\left(D^1 - D^0\right)|R=r_0\right]$$
$$= E\left[1\left(Y^1 \leq y\right)|D^1 > D^0, R=r_0\right]\Pr\left(D^1 > D^0|R=r_0\right)$$
$$= E\left[1\left(Y^1 \leq y\right)|C, R=r_0\right]\Pr\left(C|R=r_0\right).$$

The second equality follows from the definition of the potential treatment status, the continuity Assumption I2 and the inexistence of indefinites (Assumption I3).

B.2. Proofs for asymptotic distribution

The following lemma will be helpful in deriving the asymptotic properties of the complier cdf and quantile function estimators. Asymptotic normality of the quantile treatment effect estimator (4) follows from the weak convergence as a process of conditional expectation estimators of the form (7) and the functional delta method (van der Vaart, 1998). The following lemma establishes the weak convergence as a process of the local linear conditional expectation estimator $\hat{m}_{1(Y \leq y)D}^+(r_0)$. An analogous result holds for the corresponding left limit, as well as for replacing D by $(1-D)$. Having established that the local linear conditional expectation estimators converge as processes, we can afterwards apply a functional delta method to derive the limiting distribution of the complier distribution function estimators (5) and (6). (A comment on notation: in the proofs we occasionally denote the bandwidth h as h_n to emphasize the dependence on the sample.)

Lemma 5. *Under Assumptions* E *and* I, *the sequence*

$$\left\{\sqrt{h_n n}\left(\hat{m}_{1(Y \leq y)D}^+(r_0) - m_{1(Y \leq y)D}^+(r_0)\right) : y \in \mathbb{R}\right\}$$

weakly converges to a tight Gaussian process in $\ell^\infty(\mathbb{R})$.

Proof of Lemma 5. General comment: we will apply Lyapunov to obtain a central limit theorem. Therefore, we will need the assumption that $\lim_{r \to r_0} E\left[|1\left(Y \leq y\right)D - m_{1(Y \leq y)D}^+(r)|^3 \Big| R=r\right]$ exists, uniformly in y. Note that this property already follows from Assumption E. $\square$

Proof. The local linear conditional expectation estimator for $1\left(Y \leq y\right)D$, of the form (7), can be written as a function of sample averages.

$$\hat{m}_{1(Y \leq y)D}^+(r_0) = \frac{A_{n,2}B_{n,0}(y) - A_{n,1}B_{n,1}(y)}{A_{n,2}A_{n,0} - A_{n,1}^2}, \tag{11}$$

$$A_{n,l} \equiv \frac{1}{n}\sum_{i=1}^n \frac{1}{h_n}K\left(\frac{R_i - r_0}{h_n}\right)\left(\frac{R_i - r_0}{h_n}\right)^l$$

$$B_{n,l}(y) \equiv \frac{1}{n}\sum_{j=1}^n \frac{1}{h_n}1\left(Y_j \leq y\right)D_j K\left(\frac{R_j - r_0}{h_n}\right)\left(\frac{R_j - r_0}{h_n}\right)^l.$$

For ease of notation, in this proof we consider a sample drawn from the joint distribution of (Y, D, R) conditional on $R \geq r_0$, with a sample size of n, and we therefore omit explicit conditioning on $R \geq r_0$ in sums (i.e. n refers here to the number of observations with $R_i \geq r_0$). We write the bandwidth as h_n to emphasize dependence on the sample size. To show the weak convergence of $m_{1(Y \leq y)D}^+(r_0)$ we establish that each of the terms $A_{n,0}, A_{n,1}, A_{n,2}, B_{n,0}(y), B_{n,1}(y)$ converge weakly as processes, and apply a functional delta method. We start by establishing the convergence as a process of

$$\sqrt{nh_n}\left(B_{n,l}(y) - E\left[B_{n,l}(y)\right]\right), \quad l = 0, 1, \tag{12}$$

since the $A_{n,l}$ terms are trivial functions of y. Define a vector of random variables, X_i, and indexing set T.

$$X_i = \begin{pmatrix} Y_i \\ R_i \end{pmatrix},$$

$$T = \mathbb{R}.$$

Define the set of functions $\mathcal{F}_n = \left\{f_{n,t} : t \in T\right\}$, with

$$f_{n,t}(X_i) = 1\left(Y_i \leq t\right)D_i \frac{1}{\sqrt{h_n}}K\left(\frac{R_i - r_0}{h_n}\right)\left(\frac{R_i - r_0}{h_n}\right)^l, \quad l = 0, 1.$$

Then the process (12) can be written as

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(f_{n,t}(X_i) - Pf_{n,t} \right) : t \in T,$$

which corresponds to van der Vaart and Wellner's (1996) setup for Theorem 2.11.22 for convergence of processes indexed by classes of functions changing with n. Letting P and P^* denote measure and outer measure, respectively, and $\rho(s, t)$ a pseudonorm on $\mathbb{R}$, the conditions needed for convergence are the following.

1. There exist envelope functions $F_n : \left| f_{n,t}(x) \right| \le F_n(x) \quad \forall x, f, n$ which satisfy
   (a) $P^* F_n^2 = O(1)$, and
   (b) $P^* F_n^2 \{F_n > \eta \sqrt{n}\} \to 0$, for every $\eta > 0$.
2. $\mathcal{F}_{n,\delta} = \{f_{n,s} - f_{n,t} : \rho(s, t) < \delta\}$ and $\mathcal{F}_{n,\delta}^2$ are P-measurable for every $\delta > 0$.
3. $f_{n,t}$ satisfy

$$\sup_{\rho(s,t)<\delta_n} P \left(f_{n,s} - f_{n,t} \right)^2 \to 0, \quad \text{for every } \delta_n \downarrow 0,$$

4. The uniform entropy condition on page 220 of van der Vaart and Wellner holds.

Start with the first condition (envelope functions). Define a set of envelope functions to be

$$F_n = \left| \frac{1}{\sqrt{h_n}} K \left(\frac{R_j - r_0}{h_n} \right) \left(\frac{R_j - r_0}{h_n} \right)^l \right|, \quad l = 0, 1.$$

Clearly, these are envelope functions for class $\mathcal{F}_n$. Under the measurability assumption, condition 1a can be written as

$$PF_n^2 = \int \left(\frac{1}{\sqrt{h_n}} K \left(\frac{r - r_0}{h_n} \right) \left(\frac{r - r_0}{h_n} \right)^l \right)^2$$
$$\times dF_{R|R \ge r_0, D=0}(r), \quad l = 0, 1$$
$$= \int \left(K(u) u^l \right)^2 f_R(h_n u) du, \quad l = 0, 1,$$

making the change of variables $u = \frac{r - r_0}{h_n}$. Condition 1a then holds under our boundedness assumptions on R and $K(\cdot)$. Condition 1b holds trivially for $l = 0$ for bounded $K(\cdot)$. For $l = 1$, 1b is essentially the Lindeberg–Feller condition, and holds if, for example, R is bounded. Condition 2 is implied by our assumption that $K(\cdot)$ is measurable.

The quantity in condition 3 can be written as

$$\sup_{\rho(s,t)<\delta_n} P \left(f_{n,s} - f_{n,t} \right)^2$$

$$= \sup_{\rho(s,t)<\delta_n} P \left((1(Y_i \le s) - 1(Y_i \le t)) D_i \right.$$

$$\left. \times \frac{1}{\sqrt{h_n}} K \left(\frac{R_i - r_0}{h_n} \right) \left(\frac{R_i - r_0}{h_n} \right)^l \right)^2$$

$$= \int_{r \ge r_0} \left\{ \sup_{\rho(s,t)<\delta_n} \int_y (1(y \le s) - 1(y \le t))^2 dF_{Y|R=r \ge r_0, D=1}(y) \right\}$$

$$\times E[D|R = r \ge r_0] \left(\frac{1}{\sqrt{h_n}} K \left(\frac{r}{h_n} \right) \left(\frac{r}{h_n} \right)^l \right)^2 dF_{R|R \ge r_0, D=1}(r).$$

In view of condition 1a holding, condition 3 holds if we have

$$\sup_{\rho(s,t)<\delta_n} \int_y (1(y \le s) - 1(y \le t))^2 dF_{Y|R=r \ge r_0, D=1}(y)$$

$$= \sup_{\rho(s,t)<\delta_n} F_{Y|R=r \ge r_0, D=1}(s) - 2F_{Y|R=r \ge r_0, D=1}(s \wedge t)$$
$$+ F_{Y|R=r \ge r_0, D=1}(t)$$
$$= \sup_{\rho(s,t)<\delta_n} F_{Y|R=r \ge r_0, D=1}(s \vee t) - F_{Y|R=r \ge r_0, D=1}(s \wedge t) \to 0,$$

for every $\delta_n \downarrow 0$. This holds under uniform continuity of $F_{Y|R \ge r_0, D=1}$, which follows from E1.

Finally, by example 2.11.24 on page 221 of van der Vaart and Wellner, condition 4 is satisfied since $\mathcal{F}_n$ is a VC class with a VC index of 2. To see this, note that every one-point set is shattered, but a two-point set

$$\{x_1, x_2\} = \left\{ \begin{pmatrix} y_1 \\ r_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ r_2 \end{pmatrix} \right\},$$

with, say, $y_1 < y_2$ is not shattered because the function cannot pick out $\{x_2\}$. This establishes that the $B_{n,l}(y)$ terms converge. A similar argument applies to the $A_{n,l}$ terms. By the Cramér–Wold device the terms converge jointly. Finally, we need to establish the Hadamard differentiability of (11). Define (11) as a map ϕ : $\left(\mathbb{R}^3 \times \ell^\infty(\mathbb{R})^2 \right) \to (0, 1) \subset \mathbb{R}$, and define $\ell^\infty(\mathbb{R})$ as the set of all uniformly bounded real functions on the real line, and $C(\mathbb{R})$ as the space of continuous functions on the real line. As a map from $\mathbb{R}^5$ to $(0, 1) \subset \mathbb{R}$ (for a fixed value of y) the usual differentiability of ϕ implies Hadamard differentiability tangentially to the subspace $\mathbb{D}_0 = \left(\mathbb{R}^3 \times C(\mathbb{R})^2 \right)$, and the conclusion follows by the functional delta method. $\quad \square$

Proof of Theorem 2. For the following proofs, we define

$$\lambda_K' = \frac{1}{2} \frac{s_2^2 - s_3 s_1}{s_2 s_0 - s_1^2}, \tag{13}$$

$$\lambda_K = \frac{\int_0^\infty (s_2 - s_1 u)^2 K(u)^2 du}{\left(s_2 s_0 - s_1^2 \right)^2}, \tag{14}$$

where $s_l = \int_0^\infty K(u) u^l du$. For the uniform kernel with support $[-1, 1]$, i.e. $K(u) = \frac{1(|u| \le 1)}{2}$, these constants are $\lambda_K' = -\frac{1}{12}$ and $\lambda_K = 4$. $\quad \square$

Proof. The estimators (5) and (6) are (Hadamard) differentiable functions of several intermediate local linear estimators of the form (7). Let the vector variables whose expectations compose (5) and (6) be denoted as

$$W(y) = \begin{pmatrix} 1(Y \le y) D \\ 1(Y \le y)(1 - D) \\ D \\ 1 - D \end{pmatrix}.$$

Then the vector of component conditional expectations in (5) and (6) can be written as

$$M(r_0, y) = \begin{pmatrix} M^+(r_0, y) \\ M^-(r_0, y) \end{pmatrix},$$

where $M^+(r_0, y) = \lim_{r \to r_0^+} E[W(y)|R = r]$ and $M^-(r_0, y)$ is the corresponding left limit. Let $\hat{M}(r_0, y)$ be the vector of corresponding local linear estimators. Hahn et al. (1999) established the joint convergence in distribution of local linear estimators of this type. Given the regularity conditions in Assumption E, which satisfy their conditions, we can combine their result with the previous lemma to establish the joint weak convergence of $\sqrt{nh} \left(\hat{M}(r_0, y) - M(r_0, y) \right)$ to a tight Gaussian element with mean function $B_P(y) = \gamma \lambda_K' \partial^2 M(r_0, y) / \partial r^2$ and block-diagonal covariance function $\Sigma_P(y, \tilde{y})$ with the upper block given by

$$
J_P^j(y) = \frac{1}{\Delta p} \begin{pmatrix} j & j-1 & -jF_{Y^j|C}(y) & (1-j)F_{Y^j|C}(y) & -j & 1-j & jF_{Y^j|C}(y) & (j-1)F_{Y^j|C}(y) \end{pmatrix},
$$

Box II.

$$
E\left[Y^1|C, R=r_0\right] = \frac{\lim\limits_{r \to r_0^+} E\left[Y|R=r, D=1\right] \lim\limits_{r \to r_0^+} E\left[D|R=r\right] - \lim\limits_{r \to r_0^-} E\left[Y|R=r, D=1\right] \lim\limits_{r \to r_0^-} E\left[D|R=r\right]}{\lim\limits_{r \to r_0^+} E\left[D|R=r\right] - \lim\limits_{r \to r_0^-} E\left[D|R=r\right]}.
$$

Box III.

$$
\sigma^+(y, \tilde{y}) = \frac{\lambda_K}{f_R(r_0)} \lim_{r \to r_0^+} \text{Cov}\left(W(y), W(\tilde{y}) | R = r_0\right),
$$

and the lower block is the analogous left limit.

Next we turn to the joint limiting distribution of the local linear conditional distribution function estimators $\hat{F}_{Y^j|C, R=r_0}(y)$ which are differentiable functions of $\hat{M}(r_0, y)$ with Jacobians $J_P^j(y)$ given in Box II, so by the multivariate functional delta method and Lemma 5 we have that the sequences

$$
\sqrt{nh}\left[\hat{F}_{Y^j|C}(y) - F_{Y^j|C}(y)\right]
$$

converge jointly as processes to a tight Gaussian element with mean functions $J_P^j(y) B_P(y)$ and covariance function $v_{j,k}(y, \tilde{y}) = J_P^j(y) \Sigma_P(y, \tilde{y}) J_P^k(\tilde{y})'$. Multiplying these matrix expressions and simplifying yields the result as stated in the theorem. □

Proof of Theorem 3 and Corollary 4.

Proof. These results follow from Theorem 2 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). □

Appendix C. Bandwidth selection

The procedure we propose for estimating quantile treatment effects involves choosing a set of bandwidths for each quantile index τ being estimated. In principle, the asymptotically optimal bandwidths for each τ can be derived using the limiting distributions shown in Corollary 4. While straightforward, we suggest an alternative approach that avoids estimating second-order derivatives for each τ. We pursue an approach similar to Yu and Jones (1998), making simplifying assumptions that justify operational rules of thumb for choosing these bandwidths based on a single reference set of bandwidths in the context of local linear quantile regression. To simplify notation in the main text, we assumed a single bandwidth, h. In practice, however, it may be desirable to allow bandwidths to differ for estimation of $Q_{Y^0|C}(\tau)$ and $Q_{Y^1|C}(\tau)$ and on either side of the threshold. We therefore propose choosing two bandwidths $\left(h_\tau^{1+}, h_\tau^{1-}\right)$ for estimating $Q_{Y^1|C}(\tau)$ and bandwidths $\left(h_\tau^{0+}, h_\tau^{0-}\right)$ for $Q_{Y^0|C}(\tau)$. It is straightforward to adapt the expressions for the limiting distributions in Section 5 to allow for differing bandwidths, so long as they shrink at the same rate with respect to the sample size. Estimation also remains simple: the kernel weights $K\left(\frac{R_i - r_0}{h}\right)$ for the local linear two-stage least squares estimating scheme in Section 4 are simply replaced by $K\left(\frac{R_i - r_0}{h_\tau^{1+}}\right)^{Z_i} K\left(\frac{R_i - r_0}{h_\tau^{1-}}\right)^{1-Z_i}$ for $Q_{Y^1|C}(\tau)$ and similarly for $Q_{Y^0|C}(\tau)$. In the following, we describe the steps for estimating the bandwidths for $Q_{Y^1|C}(\tau)$, and note that the approach is analogous for $Q_{Y^0|C}(\tau)$.

The bandwidths for $Q_{Y^1|C}(\tau)$ are specified in terms of a reference bandwidth for estimating a mean

$$
h_\tau^{1+} = h_{\text{mean}}^{1+} \left(\frac{\tau(1-\tau)}{\phi\left(\Phi^{-1}(\tau)\right)^2}\right)^{1/5},
$$

$$
h_\tau^{1-} = h_{\text{mean}}^{1-} \left(\frac{\tau(1-\tau)}{\phi\left(\Phi^{-1}(\tau)\right)^2}\right)^{1/5},
$$

where h_{mean}^{1+} and h_{mean}^{1-} are suitable bandwidths for estimating $E\left[Y^1|C, R=r_0\right]$. This rule follows from assuming that the second order derivatives of all quantile functions are the same and that the error terms are normally distributed. We make these simplifying assumptions only to relate the optimal bandwidth for one quantile to the optimal bandwidth for the mean function. To choose a suitable pair $\left(h_{\text{mean}}^{1+}, h_{\text{mean}}^{1-}\right)$, we write the complier mean of Y^1 as $E\left[Y^1|C, R=r_0\right]$ which is given in Box III. In principle, a different bandwidth could be chosen to estimate each of the four conditional means in this expression, but following Imbens and Kalyanaraman (2009), we choose h_{mean}^{1+} to be optimal for estimating $\lim_{r \to r_0^+} E[Y|R=r, D=1]$ and h_{mean}^{1-} to be optimal for estimating $\lim_{r \to r_0^-} E[Y|R=r, D=1]$, as $\lim_{r \to r_0^+} E[D|R=r]$ and $\lim_{r \to r_0^-} E[D|R=r]$ are typically more precisely estimated.[13] We describe in the following how to choose h_{mean}^{1+}, noting that the procedure for choosing h_{mean}^{1-} is analogous.

Let $\mu_1(r_0+) \equiv \lim_{r \to r_0^+} E[Y|R=r, D=1]$. As Imbens and Lemieux (2008) suggest for the RD context, we choose h_{mean}^{1+} to minimize the approximate mean squared error of a local linear estimator of this quantity. However, instead of the cross-validation method they propose, we suggest the plug-in method. This has been shown to have better convergence properties theoretically (see Ruppert et al., 1995, for discussion), and in practice we have found plug-in to be more reliable than cross-validation. Fan (1992) shows that the conditional mean squared error of a local linear regression estimator with a bandwidth h at the boundary can be written as

$$
E\left[\left(\hat{\mu}_1(r_0+) - \mu_1(r_0+)\right)^2 \Big| \{R_i\}_{i=1}^{n_1^+}, D_i = 1, R_i \geq r_0\right]
$$

$$
= \left(\lambda_K' \mu_1''(r_0+)\right)^2 h^4 + \frac{\lambda_K}{n_1^+ h} \frac{\sigma_1^2(r_0+)}{f_{R|D=1, R \geq r_0}(r_0)} + o_p\left(h^4 + \frac{1}{n_1^+ h}\right),
$$

where n_1^+ is the number of observations above the threshold with $D = 1$, and where λ_K and λ_K' have been defined in (13) and (14).

---

[13] Imbens and Kalyanaraman (2009) go one step further and also use the same bandwidth on either side of the threshold.

Simple algebra shows that the optimal bandwidth is

$$h_{\text{mean}}^{1+} = \left(n_1^+\right)^{-1/5} \left(\frac{1}{4} \frac{\lambda_K}{\lambda_K'^2} \frac{\sigma_1^2(r_0+)}{\overline{f}_{R|D=1,R \ge r_0}(r_0)} \right)^{1/5}.$$

The plugin method (Hall et al., 1991) consists of estimating the quantities in this expression to compute the optimal bandwidth. The quantities to be estimated are $\mu_1''(r_0+) \equiv \lim_{r \to r_0^+} \frac{\partial^2}{\partial r^2} E[Y|D=1, R=r]$ and $\sigma_1^2(r_0+) \equiv \lim_{r \to r_0^+} \text{Var}(Y|D=1, R=r)$ and $f_{R|D=1,R \ge r_0}(r_0)$. For $\sigma_1^2(r_0+)$ and $\mu_1''(r_0+)$ we adapt the methods of Ruppert et al. (1995). The estimator for $\mu_1''(r_0+)$ is twice the coefficient on $(R_i - r_0)^2$ from the following least-squares quartic fit in the $(R \ge r_0, D = 1)$ cell

$$\min_{b_0,b_1,b_2,b_3,b_4} \sum_{i:R_i \ge r_0, D_i=1} \left(Y_i - \sum_{p=0}^4 b_p (R_i - r_0)^p \right)^2.$$

The estimator for $\sigma_1^2(r_0+)$ is obtained using the residuals $\hat{\varepsilon}_i$ from this same quartic fit.

$$\hat{\sigma}_1^2(r_0+) = \frac{1}{n_1^+ - 5} \sum_{i=1}^{n_1^+} \hat{\varepsilon}_i^2.$$

Finally, for $f_{R|D=1,R \ge r_0}(r_0)$ we use the following boundary kernel estimator (Jones, 1993)

$$\hat{\overline{f}}_{R|D=1,R \ge r_0}(r_0) = \frac{1}{n_1^+ h_f} \sum_{i=1}^{n_1^+} \overline{K}\left(\frac{R_i - r_0}{h_f} \right),$$

where the boundary kernel $\overline{K}$ is given by

$$\overline{K}(r) = \frac{s_2 - s_1 r}{s_0 s_2 - s_1^2} K(r),$$

where $s_l = \int_0^\infty K(u) u^l du$ and where the pilot bandwidth h_f is chosen by a simple rule of thumb (Silverman's rule).

Plugging these estimates in, we compute the bandwidth for the mean as

$$\hat{h}_{\text{mean}}^{1+} = \left(n_1^+\right)^{-1/5} \left(\frac{1}{4} \frac{\lambda_K}{\lambda_K'^2} \frac{\hat{\sigma}_1^2(r_0+)}{\hat{\overline{f}}_{R|D=1,R \ge r_0}(r_0)} \right)^{1/5}.$$

In the case of the uniform kernel, $K(u) = 0.5 \times 1(|u| \le 1)$, the constants in this expression are $\lambda_K' = -1/12$ and $\lambda_K = 4$.

Finally, in order to ensure that the estimated quantile functions are continuous, we can adopt the suggestion of Yu and Jones (1998) to also smooth "in the y-direction," as described in Section 4.1. The bandwidth for smoothing in the y-direction is chosen by the following rule of thumb $h_{Y,\tau}^1 = \min\{h_{Y,\tau}^{1+}, h_{Y,\tau}^{1-}\}$, where $h_{Y,\tau}^{1+}$ is given by the following

$$h_{Y,\tau}^{1+} = \max\left(\frac{\left(h_{0.5}^{1+}\right)^5}{\left(h_\tau^{1+}\right)^3}, \frac{h_\tau^{1+}}{10} \right) \quad \text{if } h_{0.5}^{1+} < 1$$

$$\text{and } \frac{\left(h_{0.5}^{1+}\right)^4}{\left(h_\tau^{1+}\right)^3} \quad \text{otherwise,}$$

and analogously for $h_{Y,\tau}^{1-}$.

References

Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association 97 (457), 284–292.

Abadie, A., Angrist, J.D., Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. Econometrica 70, 91–117.

Anderson, M., 2008. Multiple inference and gender differences in the effect of early intervention: a reevaluation of the abecedarian, perry preschool, and early training projects. Journal of the American Statistical Association 103 (484), 1481–1495.

Angrist, J.D., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly Journal of Economics 114 (2), 533–575.

Atkinson, A.B., 1970. On the measurement of inequality. Journal of Economic Theory 2, 244–263.

Baker, M., Firpo, S., Milligan, K., 2005. Identification and estimation of distributional effects under a regression discontinuity design with application to changes in maternity leave mandates. Unpublished Manuscript.

Battistin, E., Rettore, E., 2008. Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs. Journal of Econometrics 142, 715–730.

Black, D., Galdo, J., Smith, J., 2007. Evaluating the bias of the regression discontinuity design using experimental data. Mimeo. University of Chicago.

Black, D.A., Smith, J.A., Berger, M.C., Noel, B.J., 2003. Is the threat of reemployment services more effective than the services themselves? evidence from random assignment in the UI system. The American Economic Review 93 (4), 1313–1327.

Buddelmeyer, H., Skoufias, E., 2003. An evaluation of the performance of regression discontinuity design on PROGRESA. IZA, Bonn, Germany.

Campbell, D.T., Stanley, J.C., 1963. Experimental and Quasi-Experimental Designs for Research. Rand McNally, Chicago.

Chernozhukov, V., Fernández-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78 (3), 1093–1125.

Chernozhukov, V., Fernández-Val, I., Melly, B., 2009. Inference on counterfactual distributions. Unpublished Manuscript. MIT Department of Economics.

Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. Econometrica 73 (1), 245–261.

Cook, T.D., 2008. Waiting for life to arrive: a history of the regression-discontinuity design in psychology, statistics, and economics. Journal of Econometrics 142 (2), 636–654.

Cook, T.D., Wong, V.C., 2008. Empirical tests of the validity of the regression-discontinuity design. Annales d'Economie et de Statistique (91–92), 127–150.

DiNardo, J., Lee, D.S., 2004. Economic impacts of new unionization on private sector employers: 1984–2001. Quarterly Journal of Economics 119 (4), 1383–1441.

Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Statistical Association 87 (420), 998–1004.

Fan, J., Yao, Q., Tong, H., 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika 83 (1), 189–206.

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.

Frandsen, B.R., 2010. Union wage setting and the distribution of employees' earnings: evidence from certification elections. MIT Doctoral Dissertation.

Frölich, M., 2007. Regression discontinuity design with covariates. IZA Discussion Paper No. 3024.

Frölich, M., Melly, B., 2010. Quantile treatment effects in the regression discontinuity design: process results and gini coefficient. IZA Discussion Paper 4993. Institute for the Study of Labor, IZA.

Gormley Jr., W.T., Gayer, T., Phillips, D., Dawson, B., 2005. The effects of universal pre-K on cognitive development. Developmental Psychology 41 (6), 872–884.

Guiteras, R., 2008. Estimating quantile treatment effects in a regression discontinuity design. Mimeo. MIT.

Hahn, J., Todd, P., van der Klaauw, W., 1999. Evaluating the effect of an antidiscrimination law using a regression-discontinuity design. Working Paper 7131. National Bureau of Economic Research.

Hahn, J., Todd, P., van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–209.

Hall, P., Sheather, S.J., Jones, M.C., Marron, J.S., 1991. On optimal data-based bandwidth selection in kernel density estimation. Biometrika 78, 521–530.

Hall, P., Wolff, R.C.L., Yao, Q., 1999. Methods for estimating a conditional distribution function. Journal of the American Statistical Association 94 (445), 154–163.

Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), 467–475.

Imbens, G., Kalyanaraman, K., 2009. Optimal bandwidth choice for the regression discontinuity estimator. Working Paper 14726. National Bureau of Economic Research.

Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. Journal of Econometrics 142 (2), 615–635.

Imbens, G.W., Rubin, D.B., 1997. Estimating outcome distributions for compliers in instrumental variables models. The Review of Economic Studies 64 (4), 555–574.

Jacob, B.A., Lefgren, L., 2004. Remedial education and student achievement: a regression discontinuity analysis. The Review of Economics and Statistics 86 (1), 226–244.

Jones, M.C., 1993. Simple boundary correction for kernel density estimation. Statistics and Computing 3, 135–146.

Koenker, R., 2005. Quantile Regression. Cambridge University Press.

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.

Lee, D.S., Lemieux, T., 2009. Regression discontinuity designs in economics. NBER working paper No. 14723.

Leuven, E., Lindahl, M., Oosterbeek, H., Webbink, D., 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 (2), 829–850.

Neyman, J., 1935. Statistical problems in agricultural experiments. Supplement to the Journal of the Royal Statistical Society II, 107–180.

Porter, J., 2003. Estimation in the regression discontinuity model. Unpublished Paper. Department of Economics, Harvard University.

Rubin, D., 1978. Bayesian inference for causal effects: the role of randomization. Annals of Statistics VI, 34–58.

Ruppert, D., Sheather, S.J., Wand, M.P., 1995. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association 90 (432), 1257–1270.

Schweinhart, L., Barnes, H., Weikart, D., Barnett, W.S., Epstein, A., 1993. Significant benefits: Vol. 10. The High/Scope Perry Preschool Study Through Age 27. High/Scope Press, Ypsilanti, MI.

Thistlethwaite, D.L., Campbell, D.T., 1960. Regression-discontinuity analysis: an alternative to the ex post facto experiment. Journal of Educational Psychology 51, 309–317.

van der Klaauw, W., 2008. Regression-discontinuity analysis: a survey of recent developments in economics. Labour 22, 219–245.

van der Vaart, A., 1998. Asymptotic Statistics. Cambridge University Press, New York.

van der Vaart, A., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag, New York.

Yu, K., Jones, M.C., 1997. A comparison of local constant and local linear regression quantile estimators. Computational Statistics & Data Analysis 25, 159–166.

Yu, K., Jones, M.C., 1998. Local linear quantile regression. Journal of the American Statistical Association 93 (441), 228–237.