Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff

NBER WORKING PAPER SERIES

WANNA GET AWAY? RD IDENTIFICATION AWAY FROM THE CUTOFF

Joshua Angrist
Miikka Rokkanen

Working Paper 18662
http://www.nber.org/papers/w18662

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2012

Our thanks to Parag Pathak for many helpful discussions and comments, and to seminar participants at Berkeley, CREST, and Stanford for helpful comments. Thanks also go to Peter Hull for expert research assistance. Angrist gratefully acknowledges funding from the Institute for Education Sciences. The views expressed here are those of the authors alone and do not necessarily reflect the views of the National Bureau of Economic Research or The Institute for Education Sciences.

Wanna Get Away? RD Identification Away from the Cutoff
Joshua Angrist and Miikka Rokkanen
NBER Working Paper No. 18662
December 2012
JEL No. C26,C31,C36,I21,I24,I28,J24

ABSTRACT

In the canonical regression discontinuity (RD) design for applicants who face an award or admissions cutoff, causal effects are nonparametrically identified for those near the cutoff. The impact of treatment on inframarginal applicants is also of interest, but identification of such effects requires stronger assumptions than are required for identification at the cutoff. This paper discusses RD identification away from the cutoff. Our identification strategy exploits the availability of dependent variable predictors other than the running variable. Conditional on these predictors, the running variable is assumed to be ignorable. This identification strategy is illustrated with data on applicants to Boston exam schools. Functional-form-based extrapolation generates unsatisfying results in this context, either noisy or not very robust. By contrast, identification based on RD-specific conditional independence assumptions produces reasonably precise and surprisingly robust estimates of the effects of exam school attendance on inframarginal applicants. These estimates suggest that the causal effects of exam school attendance for 9th grade applicants with running variable values well away from admissions cutoffs differ little from those for applicants with values that put them on the margin of acceptance. An extension to fuzzy designs is shown to identify causal effects for compliers away from the cutoff.

Joshua Angrist
Department of Economics
MIT, E52-353
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER
angrist@mit.edu

Miikka Rokkanen
Department of Economics
MIT, E52-391
50 Memorial Drive
Cambridge, MA 02142-1347
rokkanen@mit.edu

> Both the tie-breaking experiment and the regression-discontinuity analysis are particularly subject to the external validity limitation of selection-X interaction in that the effect has been demonstrated only for a very narrow band of talent, i.e., only for those at the cutting score... Broader generalizations involve the extrapolation of the below-X fit across the entire range of X values, and at each greater degree of extrapolation, the number of plausible rival hypotheses becomes greater.
>
> – Donald T. Campbell and Julian Stanley (*1963; Experimental and Quasi-Experimental Designs for Research*)

1 Introduction

In a regression discontinuity (RD) framework, treatment status changes discontinuously as a function of an underlying covariate, commonly known as the running variable. Provided conditional mean functions for potential outcomes given the running variable are reasonably smooth, changes in outcome distributions at the assignment cutoff must be driven by discontinuities in the likelihood of treatment. RD identification comes from a kind of virtual random assignment, where small and presumably serendipitous variation in the running variable across the point of discontinuity manipulates treatment. On the other hand, because the running variable is typically related to outcomes, claims for unconditional "as-if random assignment" are most credible for observations near the point of discontinuity. RD methods need not identify causal effects for larger and perhaps more representative groups of subjects. Our epigraph suggests this point was no less apparent to RD's inventors than to today's RD nonparametricians.

A recent study of causal effects at Boston's selective public schools – known as "exam schools" – highlights the possibly local and potentially limiting nature of RD findings. Boston exam schools choose their students based on an index that combines admissions test scores with a student's grade point average (GPA). Abdulkadiroğlu, Angrist, and Pathak (2012) use parametric and nonparametric RD estimators to capture the causal effects of exam school attendance for exam school applications with index values in the neighborhood of admissions cutoffs. In this case, nonparametric RD compares students just to the left and just to the right of each cutoff. For most of these marginal students, the resulting estimates suggest that exam school attendance does little to boost achievement.[1] But applicants who only barely manage to gain admission to, say, the highly selective Boston Latin School, might be unlikely to benefit from an advanced exam school curriculum.

[1] Dobbie and Fryer (2012) report similar results for applicants to New York exam schools.

Stronger applicants who qualify more easily may benefit more from an elite public school education. Debates over affirmative action also focus attention on inframarginal applicants, including some who stand to gain seats and some who stand to lose should affirmative action considerations be introduced.

Motivated by the question of how exam school attendance affects achievement for inframarginal applicants, this paper tackles the theoretical problem of how to identify causal effects in an RD setting for applicants other than those in the immediate neighborhood of admissions cutoffs. Our first tack extrapolates parametric models for conditional mean functions estimated to the left and right of cutoffs. As noted by Angrist and Pischke (2009), in a parametric RD world, extrapolation is easy. In a closely-related approach, inspired by an argument in Dong and Lewbel (2011), we also use derivatives estimated at the cutoff to extrapolate treatment effects. Dong and Lewbel exploit the fact that derivatives at the cutoff are non-parametrically identified, though, in practice, our implementation of derivative-based extrapolation looks a lot like a parametric polynomial. We therefore see parametric and derivative-based extrapolation as two versions of a common functional-form-based approach.

As it turns out, functional-form-based estimation procedures fail to produce compelling results for the empirical questions that motivate our theoretical inquiry. The resulting estimates of exam school effects away from the cutoff are mostly imprecise and sensitive to the polynomial used for extrapolation, with or without the implicit weighting induced by a nonparametric bandwidth. We therefore turn to a conditional independence argument that exploits a key feature of most RD assignment mechanisms: RD treatments are assigned as a deterministic function of a single observed covariate, the running variable. The association between running variable and outcome variables is therefore the *only* source of omitted variables bias in RD estimates. If, for example, the running variable were randomly assigned, or otherwise made independent of potential outcomes, we could ignore it and analyze data from RD designs as if from a randomized trial.

This unique feature of RD assignment leads us to a conditional independence assumption (CIA) that identifies causal effects by conditioning on covariates besides the running variable, with an eye to eliminating the relationship between running variable and outcomes. It's not always possible to find such good controls, of course, but, as we show below, a straightforward statistical test isolates promising candidates. As an empirical matter, we show that conditioning on baseline scores and demographic variables largely eliminates the relationship between running variables and test score outcomes for 9th grade applicants to Boston exam schools. This finding lays the foundation for a matching strategy that identifies causal effects for inframarginal applicants.

Our CIA-driven matching strategy is used here to answer two questions about exam school treatment effects away from admissions cutoffs. The first concerns the impact of attendance at Boston's most selective exam school, the Boston Latin School (BLS), on applicants with running variable values that put them well above the BLS admissions cutoff. The second question concerns the impact of attendance at Boston's least selective exam school, the O'Bryant School of Math and Science, on applicants who fall a decile or more below the relevant admissions cutoff. The resulting estimates are mostly in line with RD estimates of causal effects at the cutoff. In particular, both strategies suggest BLS attendance generates little effect on either math or English achievement, while O'Bryant may generate some gains, especially in English Language Arts (ELA). The ELA gains for successful O'Bryant applicants approach one-fifth of a standard deviation. Interestingly, therefore, those who seem most likely to gain from any expansion in exam school seats are weak applicants who currently fail to gain admission to Boston's least selective exam school. Ultra-high ability applicants, that is, BLS applicants who easily clear the threshold for Boston's most selective public school, are likely to do well with our without a BLS experience, at least as far as standardized test scores go.

2 Motivation: Causal Effects at Boston Exam Schools

Boston's three exam schools serve grades 7-12. The high-profile Boston Latin School (BLS), which enrolls about 2,400 students, is the oldest American high school, founded in 1635. BLS is a model for other exam schools, including New York's well-known selective high schools. The second oldest Boston exam school is Boston Latin Academy (BLA), formerly Girls' Latin School. Opened in 1877, BLA first admitted boys in 1972 and currently enrolls about 1,700 students. The John D. O'Bryant High School of Mathematics and Science (formerly Boston Technical High) is Boston's third exam school; O'Bryant opened in 1893 and now enrolls about 1,200 students.

The Boston Public School (BPS) system spans a wide range of peer achievement. Like urban students elsewhere in the U.S., Boston exam school applicants who fail to enroll in an exam school end up at schools with average SAT scores well below the state average, in this case, at schools close to the 5th percentile of the distribution of school averages in the state. By contrast, O'Bryant's average SAT score falls at about the 40th percentile of the state distribution of averages, a big step up from the overall BPS average, but not elite in an absolute sense. Successful Boston BLA applicants find themselves at a school with average score around the 80th percentile of the distribution of school means, while the average SAT score at BLS is the fourth highest among public schools in

Massachusetts.

Abdulkadiroğlu, Angrist, and Pathak (2012) investigate the causal effects of exam school attendance in a fuzzy RD setup, where exam school offers are used as instrumental variables for mediating channels that might explain exam school impacts. Because Boston's exams are ordered by selectivity, unsuccessful applicants to BLS and BLA mostly find themselves in another exam school. Still, applicants admitted to each of the three schools are exposed to marked changes in peer composition. An O'Bryant offer increases average baseline (4th or 8th grade) peer scores by roughly three-fourths of a standard deviation, while the peer achievement gain is almost $.4\sigma$ at the BLA cutoff, and about $.7\sigma$ at the BLS cutoff. First-stages for racial composition also show that exam school offers induce a 12-24 percentage point reduction in the proportion of non-white classmates at each exam school cutoff. Peer achievement and racial composition are not the only channels by which an exam school education might matter, but they are clearly important features of the exam school experiment.

Here, we're initially interested in the effects of an exam school *offer* for students away from admissions cutoffs, without the complication of adjustment for possible mediators. We therefore begin by focusing on what amounts to the reduced form relation in the Abdulkadiroğlu, Angrist, and Pathak (2012) analysis. In the background, however, is the idea that mediators like exam school enrollment and peer composition explain why we might expect an exam school effect in the first place. Following the discussion of reduced form offer effects, we also consider an extension to fuzzy RD with causal mediators, in this case, a dummy for exam school enrollment and an ordered treatment variable that counts years of exam school attendance. Extensions to other mediators, such as peer composition, are straightforward in our fuzzy CIA framework.

2.1 Data

The analysis file used here comes from a data set constructed by Abdulkadiroğlu, Angrist, and Pathak (2012). This data set merges BPS enrollment and demographic information with Massachusetts Comprehensive Assessment System (MCAS) scores. MCAS tests are taken each spring, typically in grades 3-8 and 10. Baseline (i.e., pre-application) scores for grade 7 applicants are from 4th grade. Baseline English scores for 9th grade applicants come from 8th grade math and 7th grade ELA tests (the 8th grade English exam was introduced in 2006). We lose some applicants with missing baseline scores. For the purposes of our analysis, scores were standardized by subject, grade, and year to have mean zero and unit variance in the BPS population.

Data on student enrollment, demographics and test scores were combined with the BPS exam school applicant file. This file records applicants' current grade and school enrolled, applicants' preference ordering over exam schools, and applicants' Independent Schools Entrance Exam (ISEE) test scores, along with each exam schools' ranking of its applicants as determined by ISEE scores and GPA. These school-specific rankings become the exam school running variables in our setup.

Our initial analysis sample includes BPS-enrolled students who applied for exam school seats in 7th grade from 1997-2008 or in 9th grade from 2001-2007. For reasons explained below, most of the estimates reported here use 9th grade applicants only. We focus on applicants enrolled in BPS at the time of application (as opposed to private school students) because we're interested in how an exam school education compares to a traditional BPS education. Moreover, private school applicants are much more likely to remain outside the BPS district and hence out of our sample if they fail to get an exam school offer. Applicants who apply to transfer from one exam school to another are also omitted.[2]

2.2 Exam School Admissions: Defining Sharp Samples

Our sharp CIA-based estimation strategy is predicated on the notion that exam school effects are a deterministic function of exam school running variables. Exam school running variables are constructed by ranking a weighted average of ISEE scores and applicants' GPAs at the time of application. In practice, however, Boston exam school offers take account of student preferences over schools as well as ISEE scores and GPAs. Students list up to three exam schools for which they wish to be considered, in order of preference. Admissions offers are determined by a student-proposing deferred acceptance (DA) algorithm, using student preferences and school-specific running variables as inputs (for more on DA, see Abdulkadiroğlu, Pathak, and Roth (2009), Pathak and Sönmez (2008), and Pathak and Sönmez (2011)). The DA matching process complicates our RD analysis because it loosens the direct link between running variables and admissions offers. As in Abdulkadiroğlu, Angrist, and Pathak (2012), our econometric strategy begins by constructing analysis samples that restore a deterministic link between exam school offers and running variables, so that offers are sharp around admissions cutoffs.

Applicants are ranked only for schools to which they've applied, so applicants with the same GPA and ISEE scores might be ranked differently at different schools depending on where they fall in each school's applicant pool (each schools also weights ISEE and GPA a little differently). Applicants are ranked at every school to which they apply, regardless of how they've ordered schools. Student-proposing DA generates offers from student preference and school-specific rankings as follows:

- In round 1: Each student applies to his first choice school. Each school rejects its lowest-ranked

[2] For more on data, see the appendix to Abdulkadiroğlu, Angrist, and Pathak (2012).

applicants in excess of its capacity, with the rest provisionally admitted (students not rejected at this step may be rejected in later steps.)

- In general, at round $\ell > 1$: Students rejected in Round ℓ-1 apply to their next most preferred school (if any). Each school considers these students *and* provisionally admitted students from the previous round together, rejecting the lowest-ranked applicants in excess of capacity from this combined pool, producing a new provisional admit list (again, students not rejected at this step may be rejected in later steps.)

The DA algorithm terminates when either every student is matched to a school or every unmatched student has been rejected by every school he has ranked.

Let τ_k denote the rank of the last applicant offered a seat at school k; let c_{ik} denote student i's composite score at school k; and write the vector of composite scores as $\mathbf{c}_i = (c_{i1}, c_{i2}, c_{i3})$, where c_{ik} is missing if student i did not rank school k. A dummy variable $q_i(k) = 1[c_{ik} \leq \tau_k]$ indicates that student i qualified for school k by clearing τ_k (rank and qualification at k are missing for applicants who did not rank k). Finally, let p_{ik} denote student i's kth choice and represent i's preference list by $\mathbf{p}_i = (p_{i1}, p_{i2}, p_{i3})$, where $p_{ik} = 0$ if the list is incomplete. Students who ranked and qualified for a school will not be offered a seat at that school if they get an offer from a more preferred school. With three schools ranked, applicant i is offered a seat at school k in one of three ways:

- The applicant ranks school k first and qualifies: $(\{p_{i1} = k\} \cap \{q_i(k) = 1\})$.
- The applicant doesn't qualify for his first choice, ranks school k second and qualifies there: $(\{q_i(p_{i1}) = 0\} \cap \{p_{i2} = k\} \cap \{q_i(k) = 1\})$.
- The applicant doesn't qualify at his top two choices, ranks school k third, and qualifies there: $(\{q_i(p_{i1}) = q_i(p_{i2}) = 0\} \cap \{p_{i3} = k\} \cap \{q_i(k) = 1\})$.

We summarize the relationship between composite scores, cutoffs, and offers by letting O_i denote the identity of student i's offer, with the convention that $O_i = 0$ means no offer. DA determines O_i as follows:

$$O_i = \sum_{j=1}^{J} p_{ij} q_i(p_{ij}) \left[\prod_{\ell=1}^{j-1} (1 - q_i(p_{i\ell})) \right].$$

The formulation shows that the sample for which offers at school k are deterministically linked with the school-k composite score, a group we refer to as the *sharp sample* for school k, is the union of three sets of applicants:

- Applicants who rank k first, so $(p_{i1} = k)$

- Applicants unqualified for their first choice, ranking k second, so $(q_i(p_{i1}) = 0 \cap p_{i2} = k)$
- Applicants unqualified for their top two choices, ranking k third, so $((q_i(p_{i1}) = q_i(p_{i2}) = 0) \cap p_{i3} = k)$.

All applicants are in at least one sharp sample (at the exam school they rank first), but can be in more than one. For example, a student who ranked BLS first, but did not qualify there, is also in the sharp sample for BLA if he ranked BLA second.

The sharp RD treatment is an offer dummy, denoted D_{ik}, indicating applicants offered a seat at school k, defined separately for applicants in each school-specific sharp sample. For the purposes of empirical work, school-specific ranks are centered and scaled to produce the following running variable:

$$r_{ik} = \frac{100}{N_k} \times (\tau_k - c_{ik}), \tag{1}$$

where N_k is the total number of students who ranked school k (not the number in the sharp sample). Scaled school-specific ranks, r_{ik}, equal zero at the cutoff rank for school k, with positive values indicating students who ranked and qualified for admission at that school. Absent centering, scaled ranks give applicants' percentile position in the distribution of applicants to school k. Within sharp samples, we focus on a window limited to applicants with running variables no more than 20 units (percentiles) away from the cutoff. For qualified 9th grade applicants at BLS, this is non-binding since the BLS cutoff is closer to the top of the 9th grade applicant distribution than the .8 quantile.

Figure 1a plots offers as a function of r_{ik}, confirming the sharpness of offers in sharp samples. Plotted points in this figure are conditional means for all applicants in a one-unit binwidth along with a conditional mean function smoothed using local linear regression (LLR). Specifically, for school k, data in the estimation window were used to construct estimates of $\hat{\mathrm{E}}[y_i|r_{ik}]$, where y_i is the dependent variable and r_{ik} is the running variable. The LLR smoother uses the edge kernel,

$$K_h(r_{ik}) = \mathbf{1}\left\{\left|\frac{r_{ik}}{h}\right| \leq 1\right\} \cdot \left(1 - \left|\frac{r_{ik}}{h}\right|\right),$$

where h is the bandwidth. In a RD context, LLR has been shown to produce estimates with good properties at boundary points.[3]

In sharp samples, offers are determined by the running variable, but not all offers are accepted. Figure 1b plots school-specific enrollment rates around cutoffs, while Table 1 reports some of the associated summary statistics. Specifically, Table 1 shows LLR estimates of school-specific enrollment

[3]The bandwidth used here is a version of the DesJardins and McCall (2008) bandwidth (hereafter, DM) studied by Imbens and Kalyanaraman (2012), who derive optimal bandwidths for sharp RD using a mean square-error loss function with a regularization adjustment. The DM smoother (which generates somewhat more stable estimates in our application than the bandwidth Imbens and Kalyanaraman (2012) prefer) is also used to construct nonparametric RD estimates, below.

rates in the neighborhood of each school's cutoff. Among qualifying 7th grade applicants in the O'Bryant sharp sample, 72% enroll in O'Bryant, while the remaining 28% enroll in a regular BPS school. 91% of those qualifying at BLA enroll there the following fall, while 93% qualifying at BLS enroll there. Many applicants not offered seats at one exam school end up in another, mostly the next school down in the hierarchy of school selectivity. At the same time, Abdulkadiroğlu, Angrist, and Pathak (2012) document the fact that movement up the ladder of exam school selectivity is associated with dramatic changes in peer composition.[4]

2.3 Results at the Cutoff

As a starting point, we recap the RD estimation strategy that Abdulkadiroğlu, Angrist, and Pathak (2012) use to estimate the effects of exam school offers at the cutoff. Because there are no estimation issues unique to a particular school, the k subscript indexing schools is omitted going forward, though running variables and the conditional expectation functions they generate should still be understood to be school-specific. Thus, D_i indicates a generic exam school offer, from a school with running variable r_i.

Our graphical analysis begins with plots of binwise and smoothed (LLR) estimates of $\hat{\mathrm{E}}[y_i|r_i]$, where y_i is a dependent variable and r_i is the relevant (school-specific) running variable. These graphs and the associated estimates of jumps at the cutoff capture causal effects for marginal applicants. Except in the plots for 10th grade English, which show jumps at two out of three cutoffs, we see little evidence of marked discontinuities in MCAS scores at school-specific cutoffs. This is apparent in Figures 2a and 2b for middle school and Figures 3a and 3b for high school.

Parametric and non-parametric RD estimates of the effect of an exam school offer using the centered and scaled composite ranks are derived from models for potential outcomes as follows. Let Y_{1i} and Y_{0i} denote potential outcomes in treated and untreated states, with the observed outcome determined by

$$y_i = Y_{0i} + [Y_{1i} - Y_{0i}]D_i.$$

[4]A possible concern with the sharp sample strategy originates in the fact that the sharp sample itself may change discontinuously at the cutoff. Suppose , for example, there are two schools with the same cutoff and a common running variable. Some students rank two ahead of one and some rank one ahead of two. The sharp sample for school one includes those who rank 1 first and those who rank two first but are disqualfied there. This second group appears only to the left of the common cutoff, changing the composition of the sample. In practice, however, Boston school-specific running variables are distinct and school-specific cutoffs are separated. Not surprsingly, therefore, we find no evidence of discontinuities in sharp sample partiticipation at each cutoff.

The conditional mean functions for potential outcomes given the running variable are modeled as:

$$\begin{aligned} E[Y_{0i}|r_i] &= f_0(r_i) \\ E[Y_{1i}|r_i] &= \rho + f_1(r_i), \end{aligned}$$

using polynomials for $f_j(r_i); j = 0, 1$.

Substituting these expressions in $E[y_i|r_i] = E[Y_{0i}|r_i] + E[Y_{1i} - Y_{0i}|r_i]D_i$, and allowing for the fact that the estimates pool data from different test years and from different application years and entry grades, the parametric estimating equation for applicants in the sharp sample of applicants to school k is

$$y_{it} = \alpha_t + \sum_{\ell} \delta_\ell d_{i\ell} + (1 - D_i)f_0(r_i) + D_i f_1(r_i) + \rho D_i + \eta_{it}, \tag{2}$$

where the coefficient of interest is ρ. Equation (2) controls for test year effects, denoted α_t, and for the year and grade of application, indexed by ℓ and indicated by dummies, $d_{i\ell}$. The effects of the running variable are controlled by a pair of pth-order polynomials that differ on either side of the cutoff, specifically

$$f_j(r_i) = \pi_{1j}r_i + \pi_{2j}r_i^2 + ... + \pi_{pj}r_i^p; \quad j = 0, 1. \tag{3}$$

The benchmark estimates set $p = 3$.

Non-parametric estimators differ from parametric in three ways. First, they narrow the estimation window when the optimal data-driven bandwidth falls below 20, as it usually does. Non-parametric estimators also use a tent-shaped edge kernel centered at admissions cutoffs instead of the uniform kernel implicit in parametric estimation. Finally, non-parametric models control for linear functions of the running variable only, omitting higher-order terms. The nonparametric estimating equation is

$$\begin{aligned} y_{it} &= \alpha_t + \sum_{\ell} \delta_\ell d_{i\ell} + \gamma_0(1 - D)r_i + \gamma_1 D_i r_i + \rho D_i + \eta_{it} \\ &= \alpha_t + \sum_{\ell} \delta_\ell d_{i\ell} + \gamma_0 r_i + \gamma^* D_i r_i + \rho D_i + \eta_{it} \end{aligned} \tag{4}$$

Non-parametric RD estimates come from a kernel-weighted LLR fit of equation (4), estimated separately in the sharp sample of applicants to each of Boston's three exam schools.

Estimates of (2) and (4), reported in Table 2, are mostly small and some are precise enough to support a conclusion of no effect (estimates that pool applicants to the three schools are more precise; see Abdulkadiroğlu, Angrist, and Pathak (2012) for details). The statistically significant effects at Latin School are negative (for example, Latin School effects on 10th grade math and middle school

ELA). On the other hand, as suggested by the figures, Table 2 shows a statistically significant gain in 10th grade ELA scores at BLA and O'Bryant. This positive finding emerges even more clearly in an analysis that distinguishes 7th and 9th grade applicants, as we show below.

2.4 To Infinity and Beyond: Extrapolating Across the Cutoff

Researchers relying on RD to identify causal effects must be willing to extrapolate. This is highlighted by Figure 4, which illustrates the identification problem for the effects of BLS attendance on applicants with running variable values both at and away from the cutoff. The running variable is the star covariate in any RD scene, but plays a role distinct from that played by covariates in matching and regression-control strategies. In the latter, we look to comparisons of treated and non-treated observations *conditional* on covariates to eliminate omitted variables bias. As Figure 4 highlights, however, in an RD design, there is *no* value of the running variable at which both treatment and control subjects are observed. Nonparametric identification comes from infinitesimal movement across the RD cutoff. In empirical work, nonparametric inference compares applicants in a small neighborhood to the left of the cutoff to those in a small neighborhood to the right in an effort to ensure that identification involves only modest extrapolation. Identification of causal effects away from the cutoff necessarily requires a more substantial extrapolatory leap.

In a parametric setup such as described by the model in (2) and (3), extrapolation is easy though not necessarily credible. For any relevant value of c, we have

$$\rho(c) \equiv E[Y_{1i} - Y_{0i} | r_i = c] = \rho + \pi_1^* c + \pi_2^* c^2 + ... + \pi_p^* c^p, \tag{5}$$

where $\pi_1^* = \pi_{11} - \pi_{10}$, and so on. But the notation in (5) buries the extrapolation problem inherent in identification away from the cutoff: potential outcomes in the treated state are observed for $r_i = c > 0$, but the value of $E[Y_{0i}|r_i = c]$ for positive c is fundamentally counterfactual. The dotted lines in Figure 4 illustrate two possibilities, implying different causal effects at $r_i = c$. It seems natural to use observations to the left of the cutoff to pin down functional form, and then extrapolate this to impute $E[Y_{0i}|r_i = c]$. With enough data, and sufficiently regular conditional mean functions, $f_0(c)$ is identified for all values of c, including those never seen in the data. It's also easy to see, however, why this approach may not generate robust or convincing findings.

We experimented with parametric extrapolation for the effects of exam school attendance on the 10th grade math scores of 7th and 9th grade applicants to O'Bryant and BLS. This inquiry revolves around two focused questions, which also anchor our CIA-based analysis. Specifically, we ask:

1. How would inframarginal low-scoring O'Bryant applicants do if they were lucky enough to

find seats at O'Bryant; in other words, what if poorly qualified O'Bryant applicants now at a regular BPS school were given the opportunity to attend O'Bryant?

2. How would inframarginal high-scoring BLS applicants do if their BLS offers were withdrawn; in other words, what if highly qualified applicants now at BLS had to settle for BLA?

Both questions can be motivated by debates over affirmative action in exam schools. Between 1974 and 1998, Boston exam schools reserved seats for minority applicants. Though quotas are no longer in place, the role of race in exam school admissions continues to be debated in Boston and is currently the subject of litigation in New York. The first question above addresses the impact of exam school attendance on applicants who currently fail to make the cut but might do so with minority preferences restored. The second question applies to applicants like Julia McLaughlin, whose 1996 lawsuit ended racial quotas at Boston exam schools. McLaughlin was offered a seat at BLA, but sued for a seat at BLS, arguing, ultimately successfully, that she was kept out of BLS by unconstitutional racial quotas. The counterfactual here sends BLS students like McLaughlin back to BLA.

The unsatisfying nature of parametric extrapolation emerges in Figure 5. This figure shows imputed counterfactuals for 10th grade math scores using linear, quadratic, and cubic specifications for $f_j(r_i)$ in a sample that includes both 7th and 9th grade applicants. These models generate a wide range of estimates, especially as distance from the cutoff grows. The estimated effect of BLS attendance 15 units out change sign when the polynomial goes from second to third degree. This variability is not surprising, perhaps, and consistent with the Campbell and Stanley (1963) observation that, "at each greater degree of extrapolation, the number of plausible rival hypotheses becomes greater." On the other hand, given that $f_0(r_i)$ looks reasonably linear for $r_i < 0$ and $f_1(r_i)$ looks reasonably linear for $r_i > 0$, we might have hoped for results consistent with those from linear models even when the specification allows something more elaborate.

Panel A of Table 3, which reports estimates and standard errors from the models used to construct the fitted values plotted in Figure 5, shows that part of the problem uncovered in the figure is imprecision. Estimates constructed with $p = 3$ are clearly too noisy to be useful at $c = 5$ or higher. Models setting $p = 2$ generate much more precise estimates of $\rho(c)$ than when $p = 3$, though still fairly imprecise for $c \geq 10$. On the other hand, for very modest extrapolation ($c = 1$), a reasonably consistent picture emerges. Like RD estimates at the cutoff, this slight extrapolation generates small positive estimates for O'Bryant and small negative effects at BLS, some on the margin of statistical significance.[5]

[5] The estimates in Table 3 parallel those in Figure 5 and are from models omitting controls for test year and

Using Derivatives Instead

Dong and Lewbel (2011) propose an alternative to parametric extrapolation based on the insight that the derivatives of conditional mean functions at the cutoff are nonparametrically identified (a similar idea appears in Section 3.3.2 of DiNardo and Lee, 2011). Using the fact that

$$f_j(c) \approx f_j(0) + f_j'(0) \cdot c, \tag{6}$$

the necessary conditional mean functions can be approximated for any value of c, though the quality of the approximation is likely to shrink as c grows.

The components of (6) are consistently estimated by fitting a linear model to $f_j(r_i)$ in a neighborhood of the cutoff using a data-driven bandwidth. Specifically, the effect of an offer at cutoff value c can be approximated as

$$\rho(c) \approx \rho + \gamma^* \cdot c, \tag{7}$$

with parameters taken from equation (4). The innovation here relative to LLR estimation of (4) is in the interpretation of the interaction term, γ^*. Instead of a bias-reducing nuisance parameter, γ^* is seen here as as identifying a derivative of substantive interest. Note also that $f_j(c)$ can be approximated to higher order by setting $p > 1$ in 2, while still weighting the LLR minimand with a data-driven bandwidth to keep the estimation local.

The results of linear extrapolation with local derivative estimates, plotted in Figure 6, seem mostly like the corresponding linear parametric results plotted in Figure 5. Interestingly, however, Figure 6 also documents more robustness to the degree of approximation used to estimate effects to right of the BLS cutoff, with quadratic and cubic results consistently showing no effect for any positive c. On the other hand, derivative-based extrapolation of O'Bryant effects to the left of the cutoff appear to be no less sensitive to the choice of p than the corresponding parametric estimates. This sensitivity emerges is in spite of the fact that Figure 6 offers little evidence of curvature in conditional mean functions for applicants to either school. The estimates that go with Figure 6, reported in Panel B of Table **3**, suggest once again a big part of the problem here is lack of precision. Perhaps not surprisingly, treatment effects imputed using local derivative estimates are uniformly less precise than the corresponding parametric estimates.

application year and grade. Estimates from models with these controls differ little from those reported in the table. Use of orthogonal polynomials as a model selection device generates no highly significant polynomial coefficients beyond linear. Still, many of the extrapolated linear estimates seem too imprecise to be useful.

Related Work

Related discussions of RD identification away from the cutoff include DiNardo and Lee (2011) and Lee and Lemieux (2010), both of which note that the local interpretation of nonparametric RD estimates can be relaxed by treating the running variable as random rather than conditioning on it. In this view, observed running variable values are the realization of a non-degenerate stochastic process assigning values to individuals of an underlying type. Each type contributes to local-to-cutoff average treatment effects in proportion to the type's likelihood of being represented at the cutoff.

Since "type" is an inherently latent construct, the DiNardo-Lee-Lemieux interpretation doesn't seem to offer concrete guidance as to how general any particular RD estimate might be. The distinction between fixed and random running variables parallels that between inference with fixed and stochastic regressors in classical regression theory. In practice, this distinction offers researchers the opportunity to fix or ignore the marginal distribution of regressors observed in any particular sample. The empirical consequences of such regressor conditioning boil down to a modest adjustment of standard errors.[6] At the same time, Lee and Lemieux (2010, p. 298-299) note that *observed* covariates may provide a useful lever for the RD extrapolation problem: "It remains to be seen whether or not and how information on the reliability [of a test-based running variable], or a second test measurement, or other covariates that can predict assignment could be used in conjunction with the RD gap to learn about average treatment effects for the overall population." The next section takes up this challenge.[7]

3 Using Conditional Independence

RD designs take the mystery out of treatment assignment. In sharp samples of applicants to Boston exam schools, we know that offers at a given school and applicant cohort are determined by

$$D_i = 1[r_i > 0].$$

[6]See Abadie, Imbens, and Zheng (2011) for a recent discussion of this point.

[7]Also noteworthy is the study by Jackson (2010), which tackles inframarginal RD identification by exploiting cross-sectional variation in cutoffs. Jackson identifies inframarginal exam school effects at schools in Trinidad and Tobago by exploiting the fact that students with the same running variable (a test score) can end up at different schools depending on their preferences. Effects away from the cutoff are identified by differences-in-differences style contrasts between infra-marginal high- and low-scoring applicants with different rankings. A differences-in-differences approach might also work for Boston exam schools, as here too offers are partly determined by student preferences. We leave application of this approach to Boston exam schools for future work as the econometric issues seem distinct from those raised by the conditional independence framework detailed below.

This signal feature of the RD design implies that failure to control for r_i is the only possible source of omitted variables bias.[8]

Armed with precise knowledge of the source of omitted variables bias, we propose to identify causal effects by means of a conditional independence argument. In sharp samples, Boston exam school offers are determined by measures of past achievement, specifically ISEE scores and students' GPAs. But these are not the only measures of lagged achievement available. In addition to demographic variables that are highly predictive of achievement, we observe (pre-application) scores on MCAS tests taken in 4th grade and, for high school applicants, in 7th or 8th grade. Conditioning on this rich and relevant set of controls may serve to break the link between running variables and outcomes.

Gather the set of available controls in a covariate vector, x_i. Our conditional independence assumption (CIA) asserts that:

CONDITIONAL INDEPENDENCE ASSUMPTION (CIA)

$$E[Y_{ji}|r_i, x_i] = E[Y_{ji}|x_i];\ j = 0, 1$$

In other words, the CIA says that potential outcomes are mean-independent of the running variable conditional on x_i. This strong assumption identifies any counterfactual average of interest. For example, the average of Y_{0i} to the right of the cutoff is:

$$E[Y_{0i}|D_i = 1] = E\{E[Y_{0i}|x_i, D_i = 1]|D_i = 1\} = E\{E[y_i|x_i, D_i = 0]|D_i = 1\}, \tag{8}$$

while the average treatment effect on the treated is identified by a matching-style estimand:

$$E[Y_{1i} - Y_{0i}|D_i = 1] = E\{E[y_i|x_i, D_i = 1] - E[y_i|x_i, D_i = 0]|D_i = 1\}.$$

3.1 Testing the CIA

CIA-type assumptions break the link between treatment status and potential outcomes, opening the door to identification of a wide range of average causal effects (as in Heckman, Ichimura, and Todd (1998) and Dehejia and Wahba (1999)). Have we thrown RD out the window in favor of a conventional matching style conditional independence argument? A key distinction between the CIA assumption invoked here and conventional matching is exploitation of the deterministic nature of the RD assignment mechanism.

[8]Cook (2008) credits Goldberger (1972a) and Goldberger (1972b) for showing that with selection on a pretest, regression control for the pretest eliminates omitted variables bias. Goldberger credits Barnow (1972) and Lord and Novick (1972) for similar insights.

Knowledge of the assignment mechanism is key to our approach because it guides specification of the covariate vector, x_i. By virtue of the conditional independence relation implied by the CIA, we have:

$$E[Y_{1i}|r_i, x_i, r_i > 0] = E[Y_{1i}|x_i] = E[Y_{1i}|x_i, r_i > 0],$$

so we should expect that

$$E[y_i|r_i, x_i, D_i = 1] = E[y_i|x_i, D_i = 1], \tag{9}$$

to the right of the cutoff. Likewise, the CIA also implies:

$$E[Y_{0i}|r_i, x_i, r_i < 0] = E[Y_{0i}|x_i] = E[Y_{0i}|x_i, r_i < 0],$$

suggesting we look for

$$E[y_i|r_i, x_i, D_i = 0] = E[y_i|x_i, D_i = 0], \tag{10}$$

to the left of the cutoff.

Regressions of outcomes on x_i and the running variable on either side of the cutoff provide a natural test for (9) and (10). These testable implications give the CIA a big plausibility boost when substantiated, while the CIA strategy seems implausible when they fail. Mean independence is stronger than regression independence but regression testing procedures can easily embed flexible models that approximate conditional mean functions. As always, researchers face a trade-off between power and specification error as the conditional mean parameterization becomes more elaborate, a fact that favors parsimony. In practice, simple models seem likely to provide the most useful evidence since any simple model is likely to fail conditional independence tests in the face of strong dependence between outcomes and running variable, while more elaborate specifications with many free parameters may come in under the radar when parameter estimates are poorly determined.

Concerns about power notwithstanding, the CIA is demanding and may be hard to satisfy. A weaker and perhaps more realistic version limits the range of running variable values for which the CIA is maintained. A bounded conditional independence assumption asserts that the CIA holds only over a limited range, as follows:

BOUNDED CONDITIONAL INDEPENDENCE ASSUMPTION (BCIA)

$$E[Y_{ji}|r_i, x_i, \mid r_i \mid < d] = E[Y_{ji}|x_i, \mid r_i \mid < d];\ j = 0, 1$$

Bounded CIA says that potential outcomes are mean-independent of the running variable conditional on x_i, but only in a d-neighborhood of the cutoff. Testing BCIA, we look for

$$E[y_i|r_i, x_i, 0 < r_i < d] = E[y_i|x_i, 0 < r_i < d] \tag{11}$$

to the right of the cutoff, and

$$E[y_i|r_i, x_i, -d < r_i < 0] = E[y_i|x_i, -d < r_i < 0] \tag{12}$$

to the left.

The BCIA is reminiscent of nonparametric RD identification in that it leads to estimation of casual effects inside an implicit bandwidth around the cutoff. An important distinction, however, is the absence of any promise to make the d-neighborhood smaller as the sample size grows. Likewise, there's no parallel effort to choose bandwidth or local polynomial smoothers with an eye on bias-variance trade-offs. Researchers need only choose values of d that are large enough to be useful or interesting in the application at hand. The largest value of d that appears to satisfy BCIA defines the playing field for CIA-based matching.

CEI vs CIA

A natural alternative to the CIA asserts mean independence of individual causal effects, instead of potential outcomes. In an RD context, this weaker Conditional Effect Ignorability (CEI) assumption, similar to the one introduced by Angrist and Fernandez-Val (2010), says:

CONDITIONAL EFFECT IGNORABILITY (CEI)

$$E[Y_{1i} - Y_{0i}|r_i, x_i] = E[Y_{1i} - Y_{0i}|x_i]$$

CEI means that - conditional on x_i - we can ignore the running variable when computing average causal effects, even if potential outcomes are not marginally mean-independent of the running variable.[9]

CEI has much of the identifying power of the CIA. In particular, given CEI, the effect of treatment on the treated can be written:

$$E[Y_{1i} - Y_{0i}|D_i = 1] = E\{E[y_i|x_i, r_i = 0^+] - E[y_i|x_i, r_i = 0^-]|D_i = 1\}, \tag{13}$$

where $E[y_i|x_i, r_i = 0^+]$ and $E[y_i|x_i, r_i = 0^-]$ denote right- and left-hand limits of conditional-on-x_i conditional expectation functions for outcomes at the cutoff. In other words, the CEI identifies causal effects away from the cutoff by reweighting nonparametrically identified conditional-on-covariates effects at the cutoff.

[9] Lewbel (2007) invokes a similar assumption in a setup using exclusion restrictions to correct for classifcation error in treatment status.

The CIA is stronger than CEI and therefore more powerful. In practice, CIA-based estimates also seem likely to be more useful than those derived from equation (13). For one thing, the CIA can be expected to generate more precise estimates since, not being limited to identification near the cutoff, it uses more data. Second, CEI relies on the ability to find a fair number of observations near the cutoff for all relevant covariate values, a tall order in many empirical applications. Finally, the CEI framework lacks the strong testable assumptions that fall out of the CIA. Conditional independence tests based on (11) and (12) make CIA-based estimation credible. Relying as it does on a kernel of covariate-specific nonparametric estimates at the cutoff, the CEI is hard to assess. Finally, as a practical matter, our experiments with CEI estimators for Boston exam school applicants failed to produce estimates that seem precise enough to be useful.

3.2 CIA-based Estimators

We economize on notation by omitting explicit conditioning on running variable values in the $[-d, d]$ interval. All expectations in this section should be understood to be conditional on the largest value of d that satisfies BCIA. Where relevant, the constant c is assumed to be no bigger than d in absolute value.

At specific running variable values, CIA leads to the following matching-style estimand:

$$\begin{aligned} &E[Y_{1i} - Y_{0i} | r_i = c] = \\ &\qquad E\{E[y_i|x_i, D_i = 1] - E[y_i|x_i, D_i = 0]|r_i = c\} \end{aligned} \tag{14}$$

Alternately, on the right-hand side of the cutoff, we might consider causal effects averaged over all positive values up to c, a bounded effect of treatment on the treated:

$$\begin{aligned} &E[Y_{1i} - Y_{0i} | 0 < r_i \leq c] = \\ &\qquad E\{E[y_i|x_i, D_i = 1] - E[y_i|x_i, D_i = 0]|0 < r_i \leq c\} \end{aligned} \tag{15}$$

Paralleling this on the left, the bounded effect of treatment on the non-treated is:

$$\begin{aligned} &E[Y_{1i} - Y_{0i} | -c \leq r_i < 0] = \\ &\qquad E\{E[y_i|x_i, D_i = 1] - E[y_i|x_i, D_i = 0]| -c \leq r_i < 0\} \end{aligned} \tag{16}$$

We consider two sorts of estimators of (14), (15) and (16). The first is the linear reweighting estimator discussed by Kline (2011). The second is a version of the Hirano, Imbens, and Ridder (2003) propensity score estimator based on Horvitz and Thompson (1952). We also use the estimated propensity score to document common support, as in Dehejia and Wahba's (1999) pioneering propensity score study of the effect of a training program on earnings.

Kline's reweighting estimator begins with linear models for conditional means, which can be written:

$$\begin{aligned} E[y_i|x_i] &= x_i'\beta_0 \\ E[y_i|x_i] &= x_i'\beta_1 \end{aligned} \tag{17}$$

Linearity is not really restrictive since the parametrization for $x_i'\beta_j$ can be rich and flexible. Substituting in (14), we have

$$\begin{aligned} &E[Y_{1i} - Y_{0i}|r_i = c] \\ &\quad = (\beta_1 - \beta_0)'E[x_i|r_i = c], \end{aligned} \tag{18}$$

with similar expressions based on (15) and (16).

Let $\lambda(x_i) \equiv E[D_i|x_i]$ denote the propensity score. Our propensity score weighting estimator begins with the observation that the CIA implies

$$\begin{aligned} E\left[\frac{y_i(1-D_i)}{1-\lambda(x_i)}|x_i\right] &= E[Y_{0i}|x_i] \\ E\left[\frac{y_iD_i}{\lambda(x_i)}|x_i\right] &= E[Y_{1i}|x_i] \end{aligned}$$

Bringing these expressions inside a single expectation and over a common denominator, the treatment effect on the treated for those with $0 < r_i < c$ is given by

$$E[Y_{1i} - Y_{0i}|0 < r_i \leq c] = E\left\{\frac{y_i[D_i - \lambda(x_i)]}{\lambda(x_i)\left[1-\lambda(x_i)\right]} \cdot \frac{P[0 < r_i \leq c|x_i]}{P[0 < r_i \leq c]}\right\}. \tag{19}$$

Similar formulas give the average effect for non-treated applicants and average effects at specific, possibly narrow, ranges of running variable values. Note that implementation of (19) requires a model for the probability $P[0 < r_i \leq c|x_i]$ as well as for $\lambda(x_i)$. It seems natural to use the same parameterization for both. If $c = d$, the estimand in (19) simplifies to

$$E[Y_{1i} - Y_{0i}|D_i = 1] = E\left\{\frac{y_i[D_i - \lambda(x_i)]}{[1-\lambda(x_i)]E[D_i]}\right\},$$

as in Hirano, Imbens, and Ridder (2003).[10]

[10]The expectations and conditioning here refer to distributions in the sharp sample of applicants for each school. Thus, treatment effects on the treated are for the treated applicants in a school-k sharp sample. Assuming the running variable is continuous, when the estimand targets average effects at $r_i = c$, the probabilities $P[r_i = c|x_i]$ and $P[r_i = c]$ needed for (19) become densities.

4 The CIA in Action at Boston Exam Schools

We continue to focus on counterfactual scenarios involving inframarginal unqualified O'Bryant applicants and highly qualified applicants admitted to BLS. The sample here is limited to 9th grade applicants because only for this group do we have close-to-outcome-date baseline scores. In combination with demographic control variables and 4th grade scores, 7th or 8th grade MCAS scores do a good job of eliminating the running variable from outcome conditional mean functions. By contrast, the most recent lagged test score available for 7th grade applicants is a 4th grade MCAS score. For 7th grade applicants, the available conditioning variables fail to satisfy bounded CIA for empirically interesting values of d.

As a benchmark, Figures 7a and 7b plot 10th grade scores against school-specific running variables in the sample of 9th grade applicants; the corresponding RD estimates of effects at the cutoff are reported in Table 4. Consistent with the results for all applicants reported in Table 2, the estimates and figures for 9th grade applicants offer little evidence of a gain in scores at the BLS cutoff. On the other hand, both figures and estimates point to a possible score gains for marginal O'Bryant applicants, especially in English. The nonparametric ELA estimate at the O'Bryant cutoff, reported in column 4 of Table 4, is a reasonably precise $.17\sigma$.

Bounded CIA tests for 9th grade applicants come from models that control for 4th and 7th grade ELA scores and 4th and 8th grade math scores, along with indicators of special education status, limited English proficiency, eligibility for free or reduced price lunch, race (black/Asian/Hispanic) and sex.[11] The estimation windows set d equal to 10, 15, and 20. Test results, reported in Table 5, offer only scattered evidence of CIA violations, even where $d = 20$. Results for $d = 10$ show mostly small coefficients, none significantly different from zero, while results with $d = 15$ produce only one rejection. Results for $d = 20$ are more worrying, but only one of three rejections here is stronger than marginal. This scattered pattern of violations allows us to check for changes in impact estimates across subgroups where the CIA is in doubt. For example, the marginally significant .005 in column 2 when estimated in a window of width 20 (for the math scores of treated O'Bryant applicants) disappears when the window is narrowed. When findings for this group are similar in estimation windows of 10, 15 and 20, bias from failure of the CIA seems unlikely. The coefficients in the last column of Table 5, for the ELA scores of admitted BLS applicants, are large enough to be worrying even if not significantly different from zero. Fortunately, however, the corresponding test results for math (reported in column 4) are much smaller.[12]

[11] Results with quadratic terms in lagged scores and cross-subject interaction terms are similar.

[12] The unchanging sample size to the right of the BLS cutoff as d shrinks reflects the high BLS admissions threshold

Columns 1-4 Table 6 report linear reweighting estimates of $E[Y_{1i} - Y_{0i}|0 < r_i < d]$ for BLS applicants and $E[Y_{1i} - Y_{0i}| - d < r_i < 0]$ for O'Bryant applicants, in samples that set d equal to 10, 15, and 20. The estimand for BLS is

$$\begin{aligned} &E[Y_{1i} - Y_{0i}|0 < r_i \leq d] \\ &\quad = (\beta_1 - \beta_0)' E[x_i|0 < r_i \leq d], \end{aligned} \tag{20}$$

while that for O'Bryant is

$$\begin{aligned} &E[Y_{1i} - Y_{0i}| - d \leq r_i < 0] \\ &\quad = (\beta_1 - \beta_0)' E[x_i| - d \leq r_i < 0], \end{aligned} \tag{21}$$

where β_0 and β_1 are as defined in (17). The BLS estimand is an average effect of treatment on the treated, since treated observations in the estimation window must have positive running variables. Similarly, the O'Bryant estimand is an average effect of treatment on the non-treated.

As with RD estimates at the cutoff, the CIA results in Table 6 show no evidence of a BLS achievement boost. At the same time, results for inframarginal unqualified O'Bryant applicants offer some evidence of gains, especially in ELA. The math estimates range from $.08\sigma$ when $d = 10$ to $.16\sigma$ when $d = 20$, though only those for $d = 15$ and $d = 20$ are significantly different from zero. Results for ELA are clear cut, ranging from $.18\sigma$ to $.2\sigma$, and significantly different from zero for each choice of d. The CIA estimates are remarkably consistent with the corresponding RD estimates at the cutoff: compare, for example, the estimates in columns 1 and 3 of Table 6 to nonparametric O'Bryant RD estimates at the cutoff of $.12\sigma$ (SE=.07) in math and $.17\sigma$ (SE=.07) for ELA.

Figure 8 fills in CIA-based estimates over a range of cutoff values by plotting linear reweighting estimates of $E[Y_{1i}|r_i = c]$ and $E[Y_{0i}|r_i = c]$ for all values of c in the $[-20, 20]$ interval. To the left of the O'Bryant cutoff, the estimates of $E[Y_{0i}|r_i = c]$ are essentially fitted values for observed outcomes (the line labelled "Estimated" in the figure) while the estimates of $E[Y_{1i}|r_i = c]$ are effectively an extrapolation and labelled accordingly. To the right of the BLS cutoff, the estimates of $E[Y_{1i}|r_i = c]$ are fitted values while the estimates of $E[Y_{0i}|r_i = c]$ are an extrapolation. The conditional means in this figure were constructed by plugging individual values of x_i into (17) and smoothing the resulting individual-level treatment effects using local linear regression.[13] The figure presents a picture consistent with that arising from the estimates in Table 6. In particular, the extrapolated BLS effects are small (for ELA) or noisy (for math), while the O'Bryant extrapolation reveals a

for 9th grade applicants; setting $d = 10$ is not binding for BLS on the right.

[13]This uses the edge kernel with Stata's default bandwidth.

remarkably stable increase in ELA scores away from the cutoff. The extrapolated effect of O'Bryant offers on math scores appears to increase modestly as a function of distance from the cutoff, a finding probed further below.

4.1 Propensity Score Estimates

CIA-based identification strategies for the effect of exam school offers seem like a good application for propensity score estimators since the covariates generating conditional independence include multiple continuously distributed control variables. These features of the data complicate full covariate matching. Our logit model for the propensity score uses the same control variables and parametrization as were used to construct the tests in Table 5 and the reweighting estimates in columns 1-4 of Table 6.[14]

Estimated score distributions for treated and control observations exhibit a substantial degree of overlap. This is documented in Figure 9, which plots the histogram of estimated scores for treated and control observations above and below a common horizontal axis. Not surprisingly, the larger sample for O'Bryant generates more overlap than the sample for highly selective BLS. Most score values for untreated O'Bryant applicants fall below about .6. Each decile in the O'Bryant score distribution contains at least a few treated observations, however; above the first decile, there appear to be more than enough for accurate inference. By contrast, few non-treated BLS applicants have covariate values for which a BLS offer is highly likely. We should therefore expect the BLS counterfactual to be estimated less precisely than that for O'Bryant.

It's also worth noting that because the sample contains no BLS controls with propensity score values above .8 (or .85 in one window), the BLS estimates fail to reflect outcomes for applicants with admissions probabilities above this value. Figure 9 documents a related and probably unsurprising feature of propensity score strategies: the data are likely to be thinnest where we need them most. The O'Bryant treatment effect on the non-treated implicitly compares the many non-treated applicants with low scores to the fewer (though still plentiful) treated O'Bryant applicants with scores in this range. The BLS treatment effect on the treated compares a modest number of treated applicants, more or less uniformly distributed across score values, with the corresponding untreated observations, of which many more are low-scoring than high.

Concerns about limited covariate support notwithstanding, the propensity-score-weighted estimates reported in columns 5-8 of Table 6 are remarkably consistent with the linear reweighting estimates in columns 1-4 of the table. In particular, the estimates here suggest most BLS students

[14]Propensity score models omit test date dummies because application and test dates are nearly collinear.

would lose little if they had had to go to BLA instead, while low scoring O'Bryant applicants might enjoy substantial gains in ELA were they given the chance to attend O'Bryant. At the same time, the BLS propensity score estimates in columns 6 and 8 are highly imprecise. BLS propensity score estimates are not only much less precise than the corresponding O'Bryant estimates, the standard errors here are two-four times larger than those generated by linear reweighting for the same samples of applicants. Linear reweighting looks like an attractive procedure in this context.[15]

5 Fuzzy CIA Models

Effects of O'Bryant offers on the ELA scores of 9th grade applicants are reasonably stable as distance from the cutoff grows. By contrast, the effect of O'Bryant offers on math scores appears to increase as window width or distance from the cutoff increases. In a window of width 10, estimated O'Bryant math effects are not significantly different from zero, while the estimate in a window of width 20 is fifty percent larger and significant ($.13\sigma$ with a standard error of $.05\sigma$). Taken at face value, this finding suggests that the weakest 9th grade applicants stand to gain the most from O'Bryant admission, an interesting substantive finding. Omitted variables bias (failure of CIA) seems unlikely to explain this pattern since the relevant conditional independence tests, reported in columns 1 and 5 of Table 5, show only one (marginally significant) violation.

An alternative explanation for the pattern of O'Bryant math estimates plotted in Figure 8 begins with the observation that exam school offers affect achievement by facilitating exam school enrollment. Assuming, as seems plausible, that exam school offers affect outcomes solely through enrollment (that is, other causal channels, such as peer effects, are downstream to enrollment), the estimates in Table 6 can be interpreted as the reduced form for an instrumental variables (IV) procedure in which exam school enrollment is the endogenous variable. The magnitude of reduced form comparisons is easier to interpret when the relevant first stage estimates scale these effects. If the first stage changes as a function of the running variable, comparisons of estimates across running variable values are informative only after appropriate rescaling. In principle, IV methods make the appropriate adjustment. A subtlety here, however, is how to interpret IV estimates constructed under the CIA in a world of heterogeneous potential outcomes, where the average causal effects identified by IV potentially vary with the running variable.

[15] The standard errors reported in this table use a bootstrap with 500 replications. Bootstrap standard errors provide asymptotically valid confidence intervals for estimators like (19) since, as note by Hirano, Imbens, and Ridder (2003), the propensity-score-weighting estimator is asymptotically linear. As noted at the end of Section 3.1, estimates based on CEI instead of CIA are imprecise. Still, the general pattern is similar, suggesting positive effects at O'Bryant, and nothing at BLS.

We estimate and interpret the causal effects of exam school enrollment by adapting the causal IV framework outlined in Abadie (2003). This framework allows for unrestricted treatment effect heterogeneity while accommodating covariates without functional form restrictions. The starting point is notation for potential treatment assignments, W_{0i} and W_{1i}, indexed against the instrument, in this case, exam school offers indicated by D_i. Thus, W_{0i} indicates (eventual) exam school enrollment among those not offered a seat, while W_{1i} indicates (eventual) exam school enrollment among those offered a seat. Observed enrollment status is

$$W_i = W_{0i}(1 - D_i) + W_{1i}D_i.$$

The core identifying assumption in our IV framework is a generalized version of CIA:

GENERALIZED CONDITIONAL INDEPENDENCE ASSUMPTION (GCIA)

$$(Y_{0i}, Y_{1i}, W_{0i}, W_{1i}) \amalg r_{ik} \mid x_i$$

GCIA can be assumed to hold in a $d-$neighborhood of the cutoff as with BCIA.

The GCIA generalizes simple CIA in three ways. First, GCIA imposes full independence instead of mean independence; this seems innocuous since any behavioral or assignment mechanism satisfying the latter is likely to satisfy the former. Second, along with potential outcomes, the pair of potential treatment assignments (W_{0i} and W_{1i}) is taken to be conditionally independent of the running variable. Finally, GCIA requires joint independence of all outcome and assignment variables, while the CIA in Section 3 requires only marginal (mean) independence. Again, its hard to see why we'd have the latter without the former.

5.1 Fuzzy Identification

As in Section 3.2, all expectations below and in the proofs of the theorems in this section should be understood to be conditional on the largest value of d that satisfies GCIA.

Local Average Treatment Effects

In a local average treatment effects (LATE) framework with Bernoulli treatment and Bernoulli instruments, the subset of *compliers* consists of individuals whose treatment status can be changed by changing the instrument. This group is defined here by $W_{1i} > W_{0i}$. A key identifying assumption in the LATE framework is *monotonicity*: the instrument can only make treatment more or less likely. Assuming that the instrument D_i satisfies monotonicity with $W_{1i} \geq W_{0i}$, and that for some the inequality is strong so there is a first-stage, the LATE theorem (Imbens and Angrist, 1994) tells us that

$$\frac{E[y_i|D_i=1]-E[y_i|D_i=0]}{E[W_i|D_i=1]-E[W_i|D_i=0]}=E[Y_{1i}-Y_{0i}|W_{1i}>W_{0i}]$$

In other words, a simple Wald-type IV estimator captures average causal effects on exam school applicants who enroll when they receive an offer but not otherwise.

Abadie (2003) generalizes the LATE theorem by showing that the expectation of any measurable function of treatment, covariates, and outcomes is identified for compliers. This result opens the door to IV estimation using a wide range of causal models, including nonlinear models such as those based on the propensity score. Here, we adapt the Abadie (2003) result for a fuzzy RD setup that identifies causal effects away from the cutoff. To make this adaptation, we need a conditional first stage in addition to the GCIA and monotonicity:

CONDITIONAL FIRST STAGE

$$0<P[D_i=1|x_i]<1 \quad and \quad P[W_{1i}=1|x_i]>P[W_{0i}=1|x_i]\, a.s.$$

Given GCIA, monotonicity, and a conditional first stage, the appendix proves the following theorem:

THEOREM 1 (FUZZY CIA EFFECTS)

$$\begin{aligned} E[Y_{1i}-Y_{0i}|W_{1i}>W_{0i}, 0<r_i\leq c] & \\ &= \frac{1}{P[W_{1i}>W_{0i}|0<r_i\leq c]}E\left\{\psi(D_i,x_i)\frac{P[0<r_i\leq c|x_i]}{P[0<r_i\leq c]}y_i\right\} \quad (22)\\ for \quad \psi(D_i,x_i) &\equiv \frac{D_i-\lambda(x_i)}{\lambda(x_i)[1-\lambda(x_i)]} \quad (23)\end{aligned}$$

This theorem leads to estimators that capture causal effects for compliers with running variable values falling into any range over which there's common support.[16]

At first blush, it's not immediately clear how to identify the conditional compliance probability, $P[W_{1i}>W_{0i}|0<r_i\leq c]$, appearing in the denominator of (22). Because everyone to the right of the cutoff is treated, there would seem to be no lever available to construct compliance rates conditional on $0<r_i\leq c$ (in the original LATE framework, the IV first stage measures the probability of compliance). However, paralleling an argument in Abadie (2003), the appendix shows that

$$P[W_{1i}>W_{0i}|0<r_i\leq c]=E\left\{\kappa\left(W_i,D_ix_i\right)\frac{P\left[0\leq r_i\leq c\mid x_i\right]}{P\left[0\leq r_i\leq c\right]}\right\} \quad (24)$$

[16]The weighting function in the numerator is much like that used to construct average treatment effects in Hirano, Imbens, and Ridder (2003) and Abadie (2005). Extensions of the theorem along the lines suggested by Theorem 3.1 in Abadie (2003) identify the marginal distribution of Y_{0i} and Y_{1i}.

where

$$\kappa\left(W_i, D_i x_i\right) = 1-\frac{W_i\left(1-D_i\right)}{1-\lambda(x_i)}-\frac{\left(1-W_i\right)D_i}{\lambda(x_i)}.$$

Average Causal Response

The causal framework leading to Theorem 1 is limited to Bernoulli endogenous variables. For many applicants, however, the exam school treatment is mediated by years of potential attendance rather than a simple go/no-go decision. We develop a fuzzy CIA framework for ordered treatments by adapting a result from Angrist and Imbens (1995). This framework relies on potential outcomes indexed against an ordered treatment, w_i. In this context, potential outcomes are denoted by Y_{ji} when $w_i = j$, for $j = 0, 1, 2, ..., J$. Assume also that potential treatments, w_{1i} and w_{0i}, satisfy monotonicity with $w_{1i} \geq w_{0i}$ and that there's a conditional first stage:

$$E[w_{1i}|x_i] \neq E[w_{0i}|x_i]$$

The Angrist and Imbens (1995) Average Causal Response (ACR) theorem describes the Wald IV estimand as follows:

$$\frac{E\left[y_i \mid D_i=1\right]-E\left[y_i \mid D_i=0\right]}{E\left[w_i \mid D_i=1\right]-E\left[w_i \mid D_i=0\right]} = \sum_j \nu_j E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}\right]$$

where

$$\begin{aligned}\nu_j &= \frac{P\left[w_{1i} \geq j > w_{0i}\right]}{\sum_{\ell} P\left[w_{1i} \geq \ell > w_{0i}\right]}\\ &= \frac{P\left[w_i \leq j \mid D_i=0\right]-P\left[w_i \leq j \mid D_i=1\right]}{E\left[w_i \mid D_i=1\right]-E\left[w_i \mid D_i=0\right]}\end{aligned}$$

In other words, a simple Wald-type IV estimator captures a weighted average of the unit causal treatment effects for compliers whose treatment status is moved by the instrument from below j to above j. The weights, given by the impact of the instrument on the CDF of the endogenous variable at each point, are positive and sum to one. The numerator here captures the extent to which the instrument shifts the distribution of potential treatments at each point of support.

The GCIA assumption allows us to adapt the Angrist and Imbens (1995) result to identify ACR away from the cutoff in a fuzzy RD setup with an ordered treatment. The appendix proves the following:

THEOREM 2 (FUZZY AVERAGE CAUSAL RESPONSE)

$$
\begin{aligned}
&\frac{E\left\{E\left[y_i \mid D_i=1, x_i\right]-E\left[y_i \mid D_i=0, x_i\right] \mid 0 \leq r_i \leq c\right\}}{\left\{E\left[w_i \mid D_i=1, x_i\right]-E\left[w_i \mid D_i=0, x_i\right] \mid 0 \leq r_i \leq c\right\}} \\
&=\sum_j \nu_{jc} E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j>w_{0i}, 0 \leq r_i \leq c\right] \qquad (25)
\end{aligned}
$$

where

$$
\nu_{jc} \quad = \quad \frac{P\left[w_{1i} \geq j>w_{0i} \mid 0 \leq r_i \leq c\right]}{\sum_{\ell} P\left[w_{1i} \geq \ell>w_{0i} \mid 0 \leq r_i \leq c\right]} \qquad (26)
$$

Theorem 2 says that a Wald-type estimator constructed by averaging covariate-specific first-stages and reduced forms can be interpreted as a weighted average causal response for compliers with running variable values in the desired range. The weights applied to each incremental average causal response, $E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j>w_{0i}, 0 \leq r_i \leq c\right]$, are given by the conditional probability the instruments moves the ordered treatment through the point at which the increment is evaluated.

In practice, we estimate the left hand side of (25) by fitting linear models to the reduced form and first stage using a version of Kline's (2011) reweighting procedure. The fuzzy reweighting procedure works as follows: Estimate linear reduced forms based on (17); use these estimates to construct the desired average reduced form effect as in (20) and (21); divide by a similarly constructed average first stage.[17] The same procedure can be used to estimate (25) for a Bernoulli treatment like W_i, in which case the average causal response defined by Theorem 2 becomes the same average causal effect identified by Theorem 1.

5.2 Fuzzy Estimates

We estimated fuzzy enrollment effects to the left of the O'Bryant cutoff and to the right of the BLS cutoff in windows setting d equal to 10, 15 and 20. The enrollment first stage changes remarkably little as distance from the cutoff grows. This can be seen columns 1-4 of Table 7, which report estimates of the effect of exam school offers on exam school enrollment, constructed separately for O'Bryant and BLS applicants using equation (24). The propensity score model is the same as that used to construct the estimates in Table 6 (Table 7 shows separate first stage estimates for the math and ELA samples, which differ slightly). Given this stable first stage, its unsurprising that the estimates of $E[Y_{1i}-Y_{0i}|W_{1i}>W_{0i}, 0<r_i \leq d]$, reported in columns 5-8 of the table, change qualitatively little as a function of d. The pattern here is consistent with that in Table 6, showing small and statistically insignificant effects at BLS, with evidence of large effects at O'Bryant. Estimates of O'Bryant effects on ELA are large, ranging from an impressive gain of $.37\sigma$ when $d=20$,

[17]Specifically, let ϕ_0 be the main effect of D_i and let ϕ_1 be the vector of interactions with x_i in a first stage regression of w_i on D_i, x_i, and $D_i x_i$. The denominator of (25) is $\phi_0+\phi_1' \mu_{xc}$, where $\mu_{xc}=E[x_i|0 \leq r_i \leq c]$.

to a still-substantial gain of $.27\sigma$ when the window is half as wide.

The gains for inframarginal applicants who enroll at O'Bryant are perhaps too large to be credible and may signal failure of the underlying exclusion restriction. Not everyone who starts in an exam school completes their schooling there, so we'd like to adjust for years of exam school exposure. We therefore treat years of exam school enrollment as the endogenous variable and estimate the ACR parameter on the left-hand side of equation (25), using the modified Kline reweighting procedure described above. The covariate parameterization used to construct both reduced form and first stage estimates is the same as that used to construct the sharp estimates in Table 6.

The first stage estimates for average causal response, reported in columns 1-4 of Table 8, indicate that successful BLS applicants spend about 1.8 years in BLS between application and test date, while successful O'Bryant applicants spend about 1.4 years at O'Bryant between application and test date. The associated 2SLS estimates, reported in columns 5-8 of the table, are in line with those in Table 7, but considerably more precise. For example the effect of a year of BLS exposure on ELA scores is estimated to be no more than about .04, with a standard of roughly the same magnitude. This compares with estimates of about the same size in column 8 of Table 7, but standard errors for the latter are five or more times larger. The precision gain here comes from linearity of the estimator and not the change in endogenous variable, paralleling precision gains seen in the switch from propensity score to linear reweighting to construct the sharp estimates in Table 6. The estimate ELA effects at O'Bryant consistently suggest gains of about $.14\sigma$ per year of exam school exposure, a finding that's more stable across window width than the Bernoulli estimates in column 7 of Table 7. This suggests that some of the variability in the Table 7 estimates comes from changes in the underlying first stage for years of enrollment in column 3 of Table 8. At the same time, the estimated O'Bryant math gains in column 5 of Table 8 still seem to fade in a narrower window, the same pattern seen in Tables 6 and 7.

6 Summary and Directions for Further Work

RD estimates of the effect of an offer at Boston exam schools generate little evidence of an achievement gain for most applicants on the margin of admission, but these results need not be relevant for applicants with running variable values well above or well below admissions cutoffs. We therefore develop RD-inspired identification strategies for causal effects away from the cutoff. Parametric extrapolation seems like a natural first step, but this generates unsatisfying estimates of the effects of exam school offers, mostly because many of the resulting estimates are too imprecise to be useful and sensitive to functional form. We therefore turn to identification strategies based on a conditional independence assumption that focuses on the running variable.

A key insight emerging from the RD framework is that the only source of omitted variables bias is the running variable. Our conditional independence assumption makes the running variable ignorable, that is, independent of potential outcomes, by conditioning on other predictors of outcomes. When the running variable is ignorable, treatment is ignorable. The conditional independence assumption invoked here has strong testable implications that are easily checked. Specifically, the CIA implies that in samples limited to either treated or control observations, regressions of outcomes on the running variable and the covariate vector supporting CIA should show no running variable effects. A modified or bounded version of the CIA asserts that this conditional independence relation holds only in a neighborhood of the cutoff.

Among 9th grade applicants to the O'Bryant school and Boston Latin School, bounded conditional independence appears to hold over a reasonably wide interval. Importantly, the conditioning variables supporting this result include 7th or 8th grade and 4th grade MCAS scores, all lagged versions of the 10th grade outcome variable. Lagged middle school scores in particular seems like a key control, probably because relatively recent lags are a powerful predictor of future scores. Lagged outcomes are better predictors, in fact, than the running variable, which is a composite constructed from applicants' GPAs and a distinct exam school admissions test.

Results based on the CIA suggest that inframarginal high-scoring BLS applicants gain little (in terms of achievement) from BLS attendance, a result consistent with the RD estimates of BLS effects at the cutoff reported in Abdulkadiroğlu, Angrist, and Pathak (2012). At the same time, CIA-based estimation using both linear and propensity score models generate robust evidence of strong gains in English for unqualified inframarginal O'Bryant applicants. Findings showing 10th grade ELA gains also emerge from the RD estimates reported by Abdulkadiroğlu, Angrist, and Pathak (2012), especially for nonwhites. The CIA-based estimates reported here suggest similar gains would likely be observed should the O'Bryant cutoff be reduced to accommodate currently inframarginal high school applicants, perhaps as a result of changing affirmative action policies. On the other hand, most exam school seats are offered to rising 7th graders. Unfortunately, CIA-based estimation strategies have little to say about causal effects in this larger group.

An additional contribution here is a modification of CIA-based identification strategies for fuzzy RD. We use this modification to estimate the effects of exam school offers and years of exam school attendance, in addition to the (reduced form) effects of offers themselves. The fuzzy extension allows us to explore the possibility that variations in reduced form causal effects as a function of the running variable might be driven by changes in an underlying first stage. An important by-product of the fuzzy extension, though one we leave for future work, is that this opens the door to Abadie (2003)-style identification of causal effects for compliers in RD models for quantile treatment effects. As noted recently by Frandsen, Frölich, and Melly (2012), the weighting approach used by Abadie, Angrist, and Imbens (2002) breaks down in the standard RD framework because the instrument distribution is degenerate conditional on the running variable. By taking the running variable out of the equation, our framework would seem to circumvent this problem.

In other work in progress, we're adapting the identification strategy introduced here for use in factor models in which the necessary conditioning variable is latent. The availability of multiple noisy indicators of this latent factor gives leverage in this approach. Finally, an important unsolved problem implicit in our empirical strategy is causal inference conditional on a pretest. Estimators that condition on the results of a specification or model-selection test may have sampling distributions for which conventional asymptotic formulas provide a poor approximation. Pretesting is a challenging and virtually ubiquitous problem in applied econometrics. It remains to be seen whether recent theoretical progress on the pretesting problem (e.g., Andrews and Guggenberger (2009); Belloni, Chernozhukov, and Hansen (2012)) can be fruitfully applied in our context.

Table 1: Boston Destinations

	All Applicants					
	O'Bryant		Latin Academy		Latin School	
	Z=0	Z=1	Z=0	Z=1	Z=0	Z=1
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A. 7th Grade Applicants			
Traditional Boston public schools	1.00	0.28	0.24	0.09	0.08	0.05
O'Bryant	0.00	0.72	0.75	0.00	0.06	0.00
Latin Academy	0.00	0.00	0.00	0.91	0.86	0.01
Latin School	...	...	...	...	0.00	0.93
			Panel B. 9th Grade Applicants			
Traditional Boston public schools	1.00	0.34	0.28	0.14	0.15	0.04
O'Bryant	0.00	0.66	0.72	0.00	0.00	0.00
Latin Academy	...	...	0.00	0.87	0.86	0.02
Latin School	...	...	0.00	-0.01	0.00	0.94

Notes: This table (excerpted from Abdulkadiroglu, Angrist, and Pathak, 2012) describes the destination schools of Boston exam school applicants. Enrollment rates are measured in the fall admissions cycle following exam school application and estimated using local linear smoothing. The sample of Boston 7th grade applicants includes students who applieds for an exam school seat between 1999-2008. The sample of Boston 9th grade applicants includes students who applied for an exam school seat between 2001-2007.

Table 2: Boston Reduced Form Estimates - MCAS Math and English

		Parametric Estimates			Non-parametric (DM) Estimates		
			Latin	Latin		Latin	Latin
Application	Test	O'Bryant	Academy	School	O'Bryant	Academy	School
Grade	Grade	(1)	(2)	(3)	(5)	(6)	(7)
				Panel A. Math			
7th	7th and 8th	-0.128	-0.081	-0.015	-0.087	-0.143*	-0.001
		(0.101)	(0.092)	(0.098)	(0.069)	(0.076)	(0.062)
		4035	4194	3776	3621	3986	3066
7th and 9th	10th	0.070	-0.090	-0.053	0.060	-0.050	-0.076**
		(0.070)	(0.080)	(0.055)	(0.046)	(0.044)	(0.033)
		3370	2702	2457	3067	2022	1825
				Panel B. ELA			
7th	7th and 8th	-0.060	-0.092	-0.183***	-0.072*	0.011	-0.136***
		(0.078)	(0.067)	(0.067)	(0.040)	(0.041)	(0.038)
		4139	4302	3790	3930	3743	3524
7th and 9th	10th	0.125	0.159	0.040	0.152***	0.181***	0.012
		(0.083)	(0.105)	(0.089)	(0.047)	(0.064)	(0.066)
		3379	2707	2459	3289	1781	1913

Notes: This table (excerpted from Abdulkadiroglu, Angrist, and Pathak, 2012) reports estimates of the effects of exam school offers on MCAS scores. The sample covers students within 20 standardized units of offer cutoffs. Parametric models include a cubic function of the running variable, allowed to differ on either side of offer cutoffs. Non-parametric estimates use the edge kernel, with bandwidth computed following DesJardins and McCall (2008) and Imbens and Kalyanaraman (2012). Optimal bandwidths were computed separately for each school. Robust standard errors, clustered on year and school are shown in parentheses. The number of observations is reported below standard errors.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Extrapolation Estimates for 10th Grade Math (7th & 9th grade applicants)

	O'Bryant				Latin School			
	c = -1	*c* = -5	*c* = -10	*c* = -15	*c* = 1	*c* = 5	*c* = 10	*c* = 15
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Parametric				
Linear	0.081**	0.108***	0.142***	0.176***	-0.073**	-0.069**	-0.064	-0.059
	(0.039)	(0.041)	(0.050)	(0.063)	(0.030)	(0.034)	(0.042)	(0.052)
	3391	3391	3391	3391	2460	2460	2460	2460
Quadratic	0.109*	0.199**	0.325*	0.466*	-0.093**	-0.162**	-0.247*	-0.331
	(0.057)	(0.091)	(0.166)	(0.264)	(0.045)	(0.077)	(0.143)	(0.229)
	3391	3391	3391	3391	2460	2460	2460	2460
Cubic	0.112	0.211	0.360	0.539	-0.075	-0.055	0.017	0.207
	(0.081)	(0.234)	(0.626)	(1.293)	(0.065)	(0.197)	(0.532)	(1.103)
	3391	3391	3391	3391	2460	2460	2460	2460
				Panel B: Derivative-based (Dong & Lewbel 2011)				
Linear	0.091**	0.132***	0.178***	0.222***	-0.079**	-0.100**	-0.117**	-0.122*
	(0.042)	(0.048)	(0.062)	(0.079)	(0.035)	(0.041)	(0.054)	(0.070)
	3360	3360	3360	3360	1987	2152	2152	2152
Quadratic	0.111*	0.202*	0.328	0.475	-0.078	-0.095	-0.110	-0.091
	(0.062)	(0.114)	(0.222)	(0.365)	(0.052)	(0.109)	(0.225)	(0.382)
	3360	3360	3360	3360	1987	2152	2152	2152
Cubic	0.120	0.270	0.528	0.920	-0.081	-0.071	-0.070	-0.000
	(0.090)	(0.294)	(0.834)	(1.780)	(0.079)	(0.308)	(0.938)	(2.076)
	3360	3360	3360	3360	1987	2152	2152	2152

Notes: This table reports estimated exam school offer effects on 10th grade Math scores extrapolated to points away from the cutoff indicated in the column headings. Columns 1-4 report estimates of the effect of O'Bryant attendance on unqualified O'Bryant applicants. Columns 5-8 report the effects of BLS attendance on qualified BLS applicants. Panel A shows parametric estimates computed using first, second, and third order polynomials, as indicated in rows of the table. Panel B shows estimates based on nonparametric estimates of derivatives for polynomial approximations of the same degree. The nonparametric bandwidth was constructed as for Table 2. Robust standard errors are shown in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: RD Estimates for 9th Grade Applicants

	Parametric		Nonparametric	
	O'Bryant	Latin School	O'Bryant	Latin School
	(1)	(2)	(3)	(4)
Math	0.160	-0.131	0.115*	-0.149*
	(0.109)	(0.116)	(0.066)	(0.077)
	1559	606	1386	361
ELA	0.190*	0.094	0.171***	0.069
	(0.110)	(0.179)	(0.066)	(0.104)
	1564	607	1532	458

Notes: This table reports parametric and nonparametric RD estimates of the effects of exam school attendance on 10th grade MCAS scores in the sample of 9th grade applicants. Models and estimators are the same as used to construct the estimates in Table 2. Robust standard errors are reported in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Conditional Independence Tests

	Math				ELA			
	O'Bryant		Latin School		O'Bryant		Latin School	
	D = 0	D = 1	D = 0	D = 1	D = 0	D = 1	D = 0	D = 1
Window	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
20	0.003	0.005*	0.009***	0.009	0.004	0.002	0.008*	0.054
	(0.004)	(0.003)	(0.004)	(0.024)	(0.004)	(0.004)	(0.005)	(0.044)
	513	486	320	49	516	489	320	50
15	0.012**	-0.000	0.008	0.009	0.009	0.001	0.004	0.054
	(0.006)	(0.005)	(0.005)	(0.024)	(0.006)	(0.006)	(0.006)	(0.044)
	375	373	228	49	376	374	229	50
10	0.002	0.000	0.007	0.009	0.013	-0.004	0.019	0.054
	(0.011)	(0.009)	(0.009)	(0.024)	(0.011)	(0.010)	(0.014)	(0.044)
	253	260	142	49	253	261	142	50

Notes: This table reports regression-based tests of the conditional independence assumption described in the text. Tests statistics consist of the coefficient on the same-subject running variable in models for 10th grade math and ELA scores that control for baseline scores, along with indicators for special education status, limited English proficiency, eligibility for free or reduced price lunch, race (black/Asian/Hispanic) and sex. Estimates use only observations to the left or right of the cutoff as indicated in column headings, and were computed in the window width indicated at left. Robust standard errors are reported in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: CIA Estimates of the Effect of Exam School offers

	Linear Reweighting				Propensity Score Weighting			
	Math		ELA		Math		ELA	
	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School
Window	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
20	0.157***	-0.023	0.205***	0.077	0.134**	-0.028	0.247***	0.054
	(0.040)	(0.086)	(0.041)	(0.081)	(0.052)	(0.167)	(0.072)	(0.173)
N untreated	513	320	516	320	513	320	516	320
N treated	486	49	489	50	486	49	489	50
15	0.127***	-0.070	0.185***	0.050	0.103*	-0.076	0.200***	0.018
	(0.043)	(0.049)	(0.044)	(0.075)	(0.053)	(0.213)	(0.062)	(0.233)
N untreated	375	228	376	229	375	228	376	229
N treated	373	49	374	50	373	49	374	50
10	0.083	-0.088	0.186***	0.005	0.094	-0.093	0.186**	-0.052
	(0.054)	(0.054)	(0.053)	(0.090)	(0.060)	(0.251)	(0.075)	(0.347)
N untreated	253	142	253	142	253	142	253	142
N treated	260	49	261	50	260	49	261	50

Notes: This table reports estimates of the effect if exam school offers on MCAS scores for 9th grade applicants to O'Bryant and BLS. Columns 1-4 report results from a linear reweighting estimator, whiles columns 5-8 report results from inverse propesnity score weighting, as described in the text. Controls are the same as used to construct the test statistics except that the propensity score models omit test year dummies. The O'Bryant estimates are effects on nontreated applicants in windows to the left of the admissions cutoff; the BLS estimates are are effects on treated applicants in windows to the right of the cutoff. Standard errors (shown in parentheses) were computed using a nonparametric bootstrap with 500 replications. The table also reports the number of treated and untreated (offered and not offered) observations in each window, in the relevant outcome sample.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Fuzzy CIA Estimates of LATE (Exam School Enrollment)

	First Stage				LATE			
	Math		ELA		Math		ELA	
	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School
Window	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
20	0.661***	0.898***	0.662***	0.900***	0.203**	-0.031	0.372***	0.060
	(0.056)	(0.052)	(0.058)	(0.050)	(0.085)	(0.180)	(0.145)	(0.167)
N untreated	513	320	516	320	513	320	516	320
N treated	486	49	489	50	486	49	489	50
15	0.668***	0.898***	0.669***	0.900***	0.154**	-0.085	0.299***	0.020
	(0.048)	(0.052)	(0.049)	(0.049)	(0.078)	(0.222)	(0.113)	(0.238)
N untreated	375	228	376	229	375	228	376	229
N treated	373	49	374	50	373	49	374	50
10	0.676***	0.898***	0.680***	0.900***	0.139	-0.104	0.273**	-0.058
	(0.048)	(0.050)	(0.048)	(0.047)	(0.086)	(0.303)	(0.134)	(0.411)
N untreated	253	142	253	142	253	142	253	142
N treated	260	49	261	50	260	49	261	50

Notes: This table reports fuzzy RD estimates of the effect of exam school enrollment on MCAS scores for 9th grade applicants to O'Bryant and BLS. The O'Bryant estimates are effects on nontreated applicants in windows to the left of the admissions cutoff; the BLS estimates are for treated applicants in windows to the right of the cutoff. The first stage estimates in columns 1-4 and the estimated causal effects in columns 5-8 are from a modified propensity-score style weighting estimator described in the text. Standard errors (shown in parentheses) were computed using a nonparametric bootstrap with 500 replications. The table also reports the number of treated and untreated (offered and not offered) observations in each window, in the relevant outcome sample.

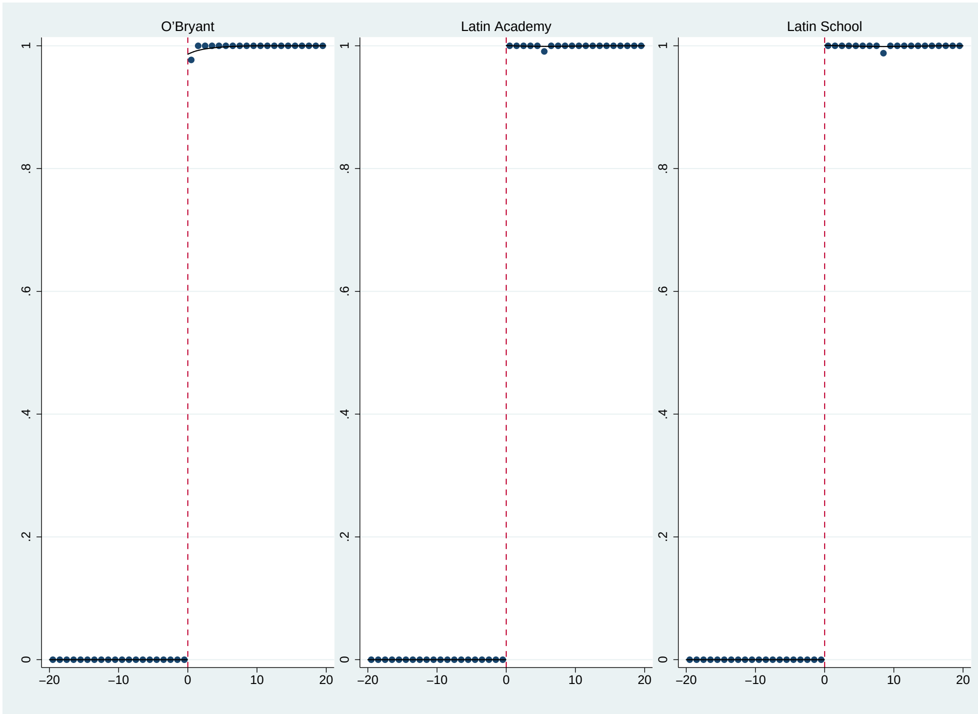
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: Fuzzy CIA Estimates of Average Causal Response (Years of Exam School Enrollment)

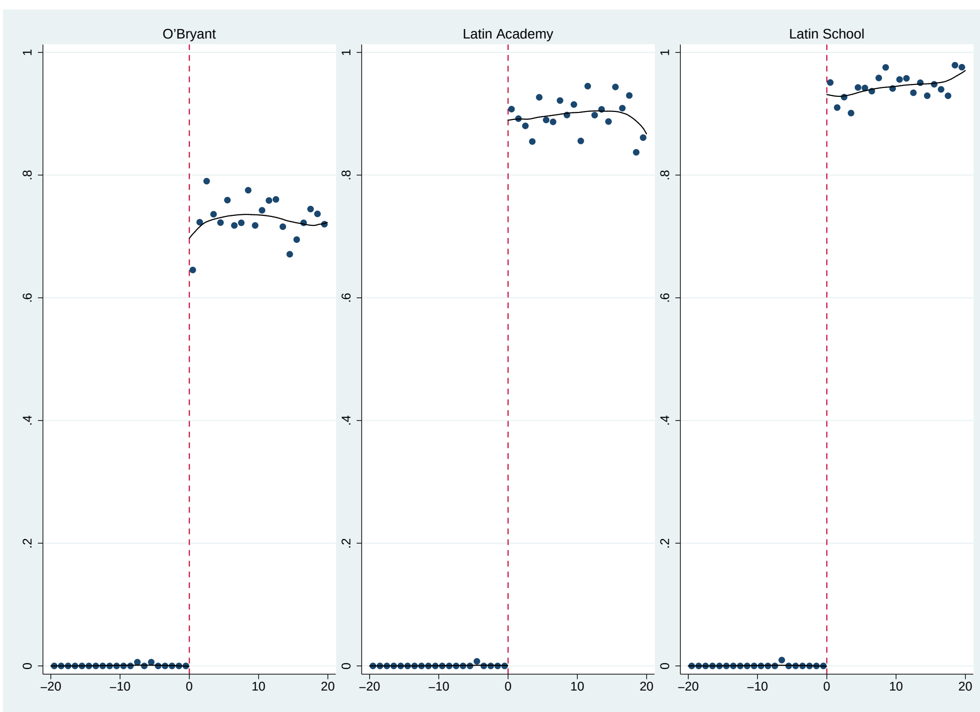
	First Stage				ACR			
	Math		ELA		Math		ELA	
	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School
Window	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
20	1.388***	1.816***	1.394***	1.820***	0.113***	-0.013	0.147***	0.042
	(0.064)	(0.095)	(0.065)	(0.090)	(0.029)	(0.046)	(0.030)	(0.043)
N untreated	513	320	516	320	513	320	516	320
N treated	486	49	489	50	486	49	489	50
15	1.347***	1.816***	1.353***	1.820***	0.094***	-0.039	0.137***	0.027
	(0.064)	(0.096)	(0.061)	(0.093)	(0.032)	(0.028)	(0.034)	(0.042)
N untreated	375	228	376	229	375	228	376	229
N treated	373	49	374	50	373	49	374	50
10	1.317***	1.816***	1.308***	1.820***	0.063	-0.048	0.142***	0.003
	(0.076)	(0.096)	(0.077)	(0.096)	(0.039)	(0.031)	(0.043)	(0.052)
N untreated	253	142	253	142	253	142	253	142
N treated	260	49	261	50	260	49	261	50

Notes: This table reports fuzzy RD estimates of the effect of years of exam school enrollment on MCAS scores for 9th grade applicants to O'Bryant and BLS. The O'Bryant estimates are effects on nontreated applicants in windows to the left of the admissions cutoff; the BLS estimates are for treated applicants in windows to the right of the cutoff. The first stage estimates in columns 1-4 and the estimated causal effects in columns 5-8 are from a modified linear 2SLS estimator described in the text. Standard errors (shown in parentheses) were computed using a nonparametric bootstrap with 500 replications. The table also reports the number of treated and untreated (offered and not offered) observations in each window, in the relevant outcome sample.

* significant at 10%; ** significant at 5%; *** significant at 1%

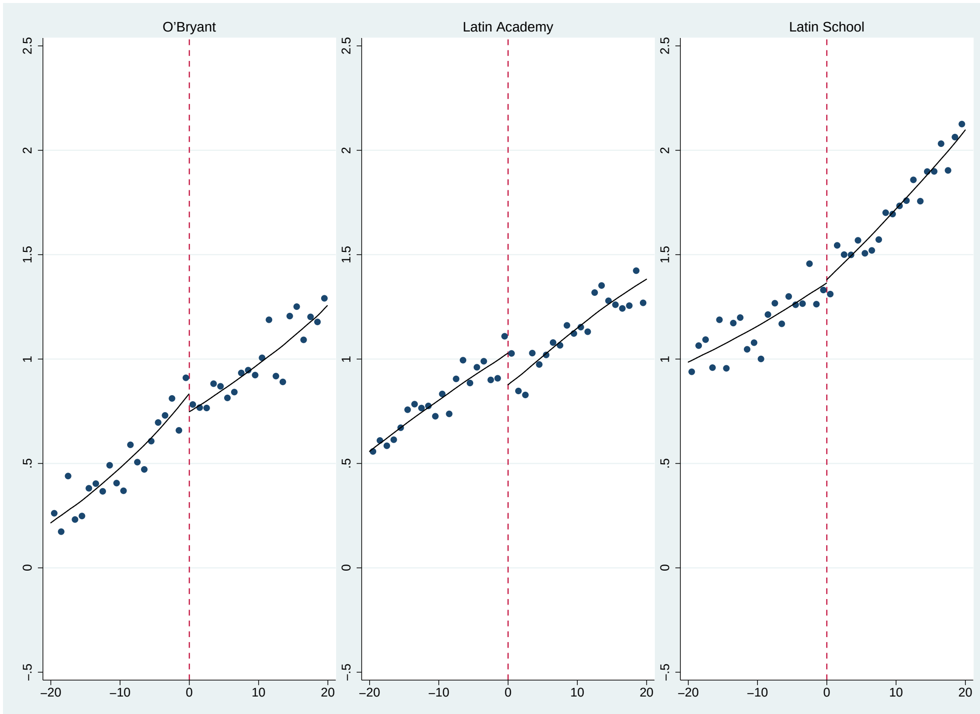


(a) Offers

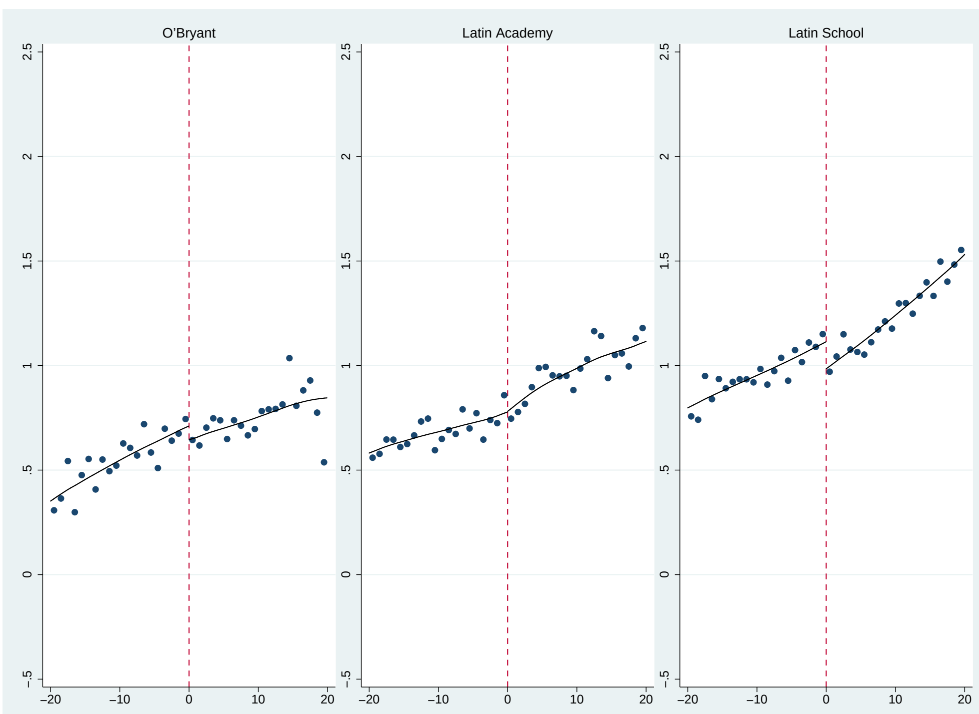


(b) Enrollment

Figure 1: Offers and Enrollment at Each Boston Exam School for 7th (1999-2008) and 9th (2001-2007) Grade Applicants

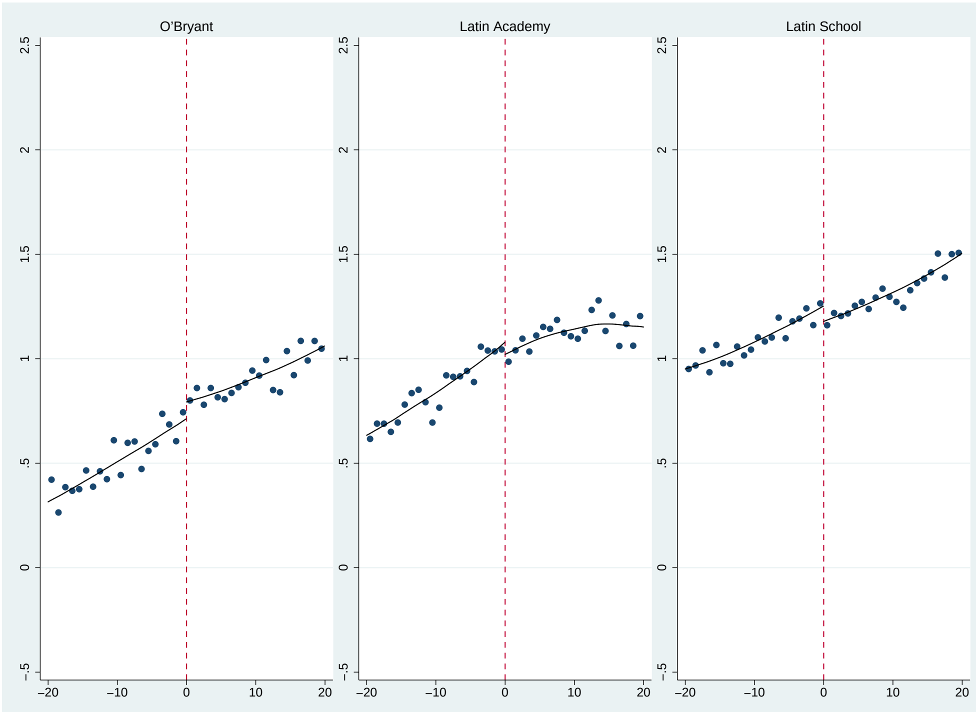


(a) Math

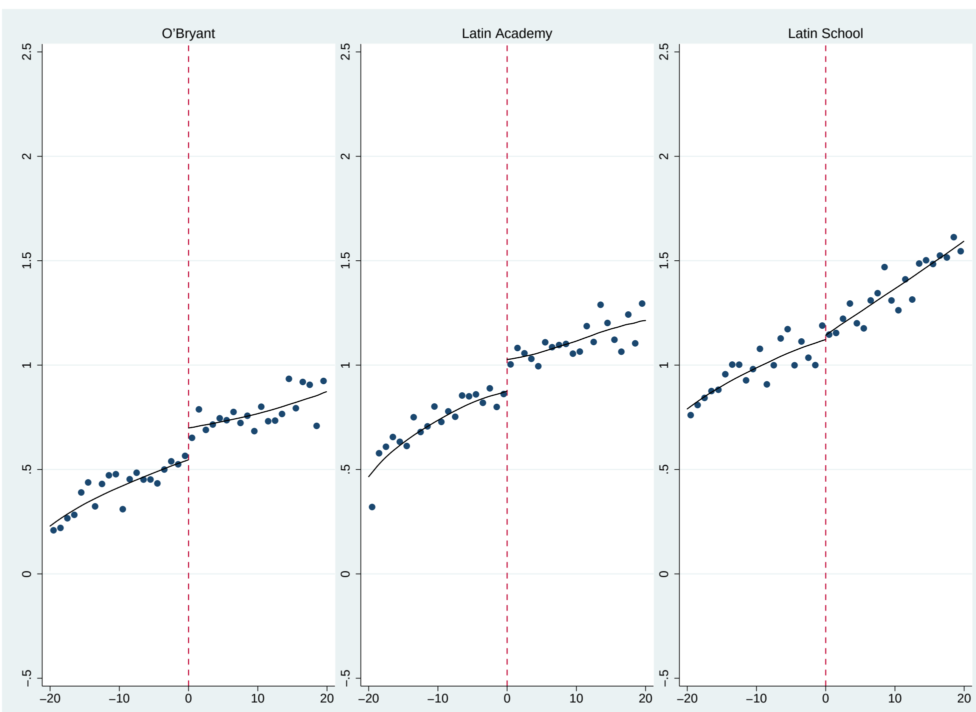


(b) English

Figure 2: 7th and 8th Grade Math and ELA Scores at Boston Exam Schools for 7th Grade Applicants



(a) Math



(b) English

Figure 3: 10th Grade Math and ELA Scores at Boston Exam Schools for 7th and 9th Grade Applicants

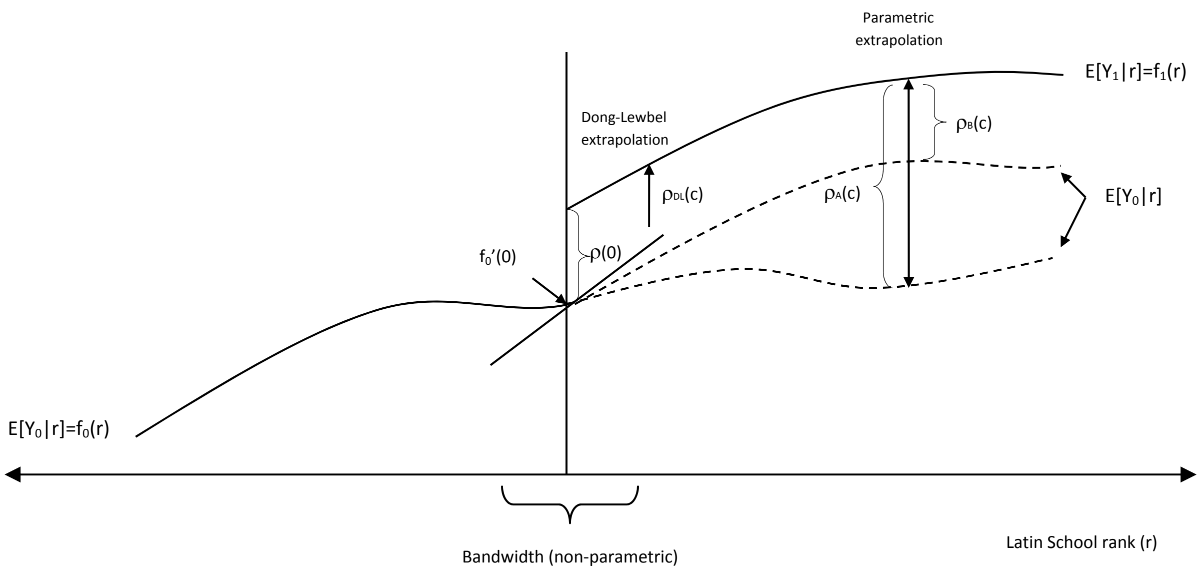


Figure 4: Identification of Boston Latin School Effects At and Away from the Cutoff. $\rho(0)$ is an effect at the cutoff; $\rho_{DL}(c)$ is an effect near the cutoff approximated using a first derivative; $\rho_A(c)$ and $\rho_B(c)$ are possible effects well away from the cutoff.

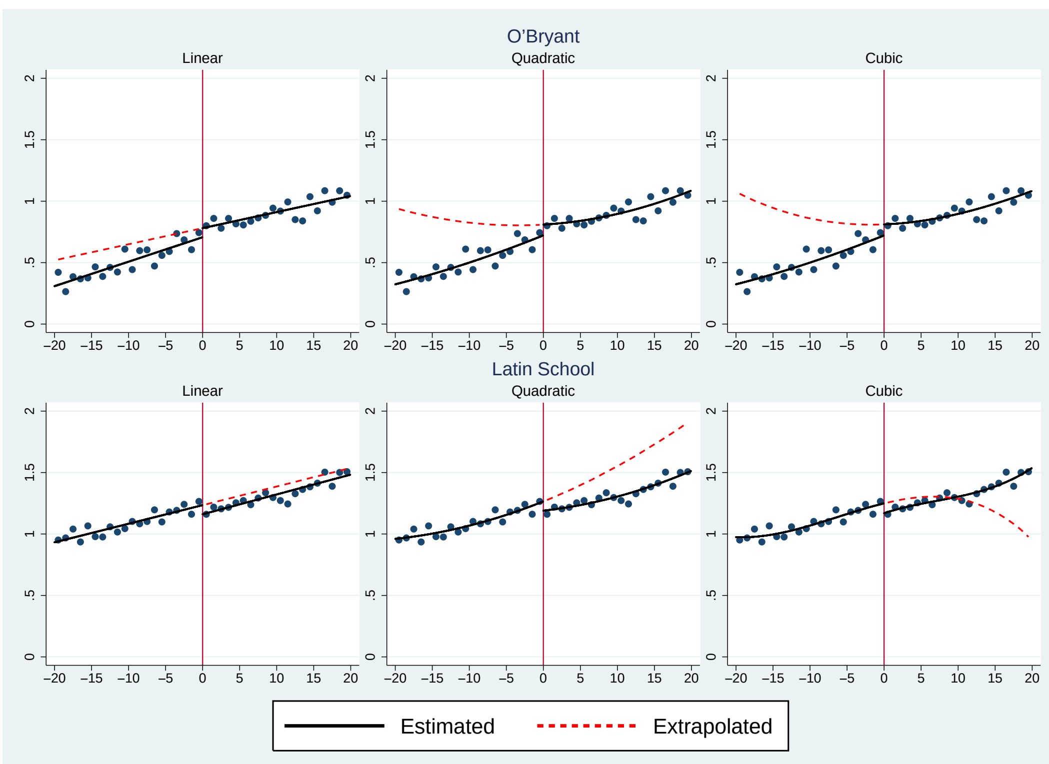


Figure 5: Parametric Extrapolation at O'Bryant and Boston Latin School. O'Bryant extrapolation is for $E[Y_{1i}|r_i = c]$; BLS extrapolation is for $E[Y_{0i}|r_i = c]$.

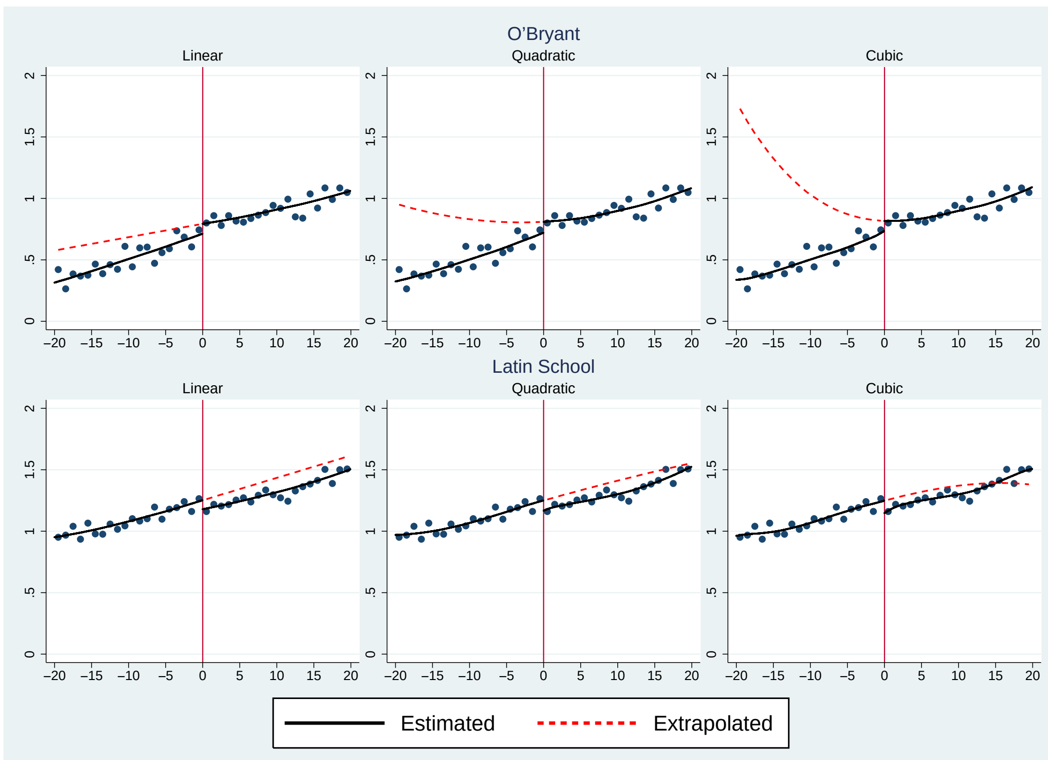
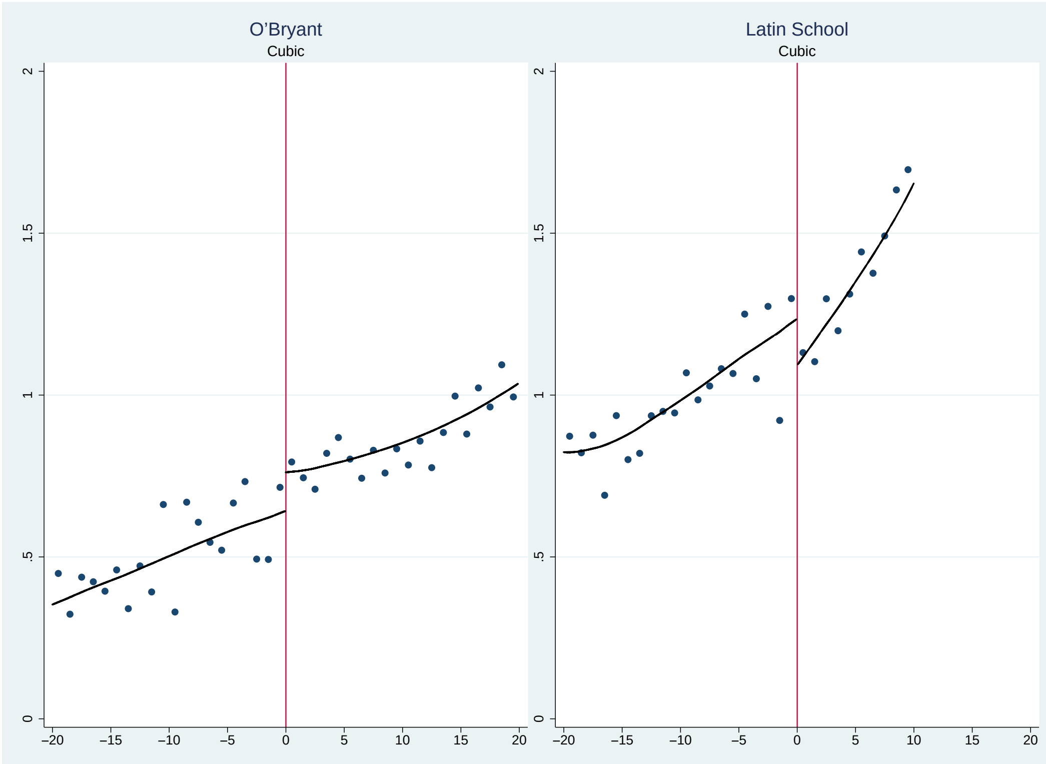
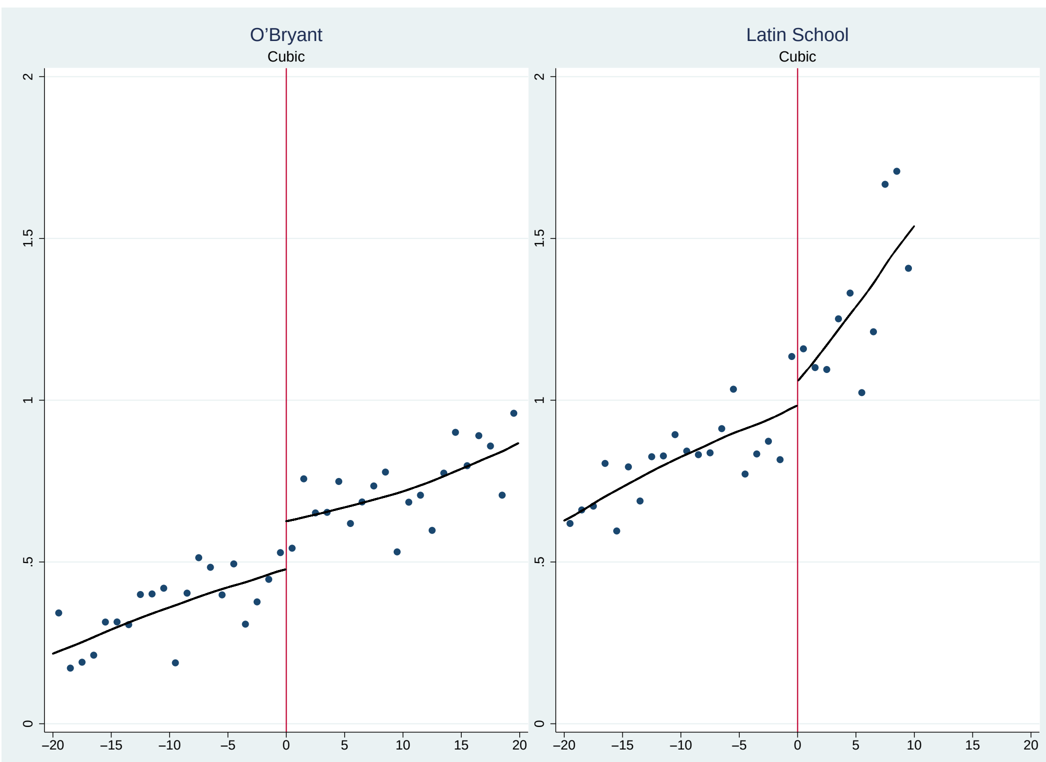


Figure 6: Derivative-based Extrapolation at O'Bryant and Boston Latin School. O'Bryant extrapolation is for $E[Y_{1i}|r_i = c]$; BLS extrapolation is for $E[Y_{0i}|r_i = c]$.



(a) 10th Grade Math



(b) 10th Grade English

Figure 7: 10th Grade Math and ELA Scores for 9th Grade Applicants to O'Bryant and Boston Latin School

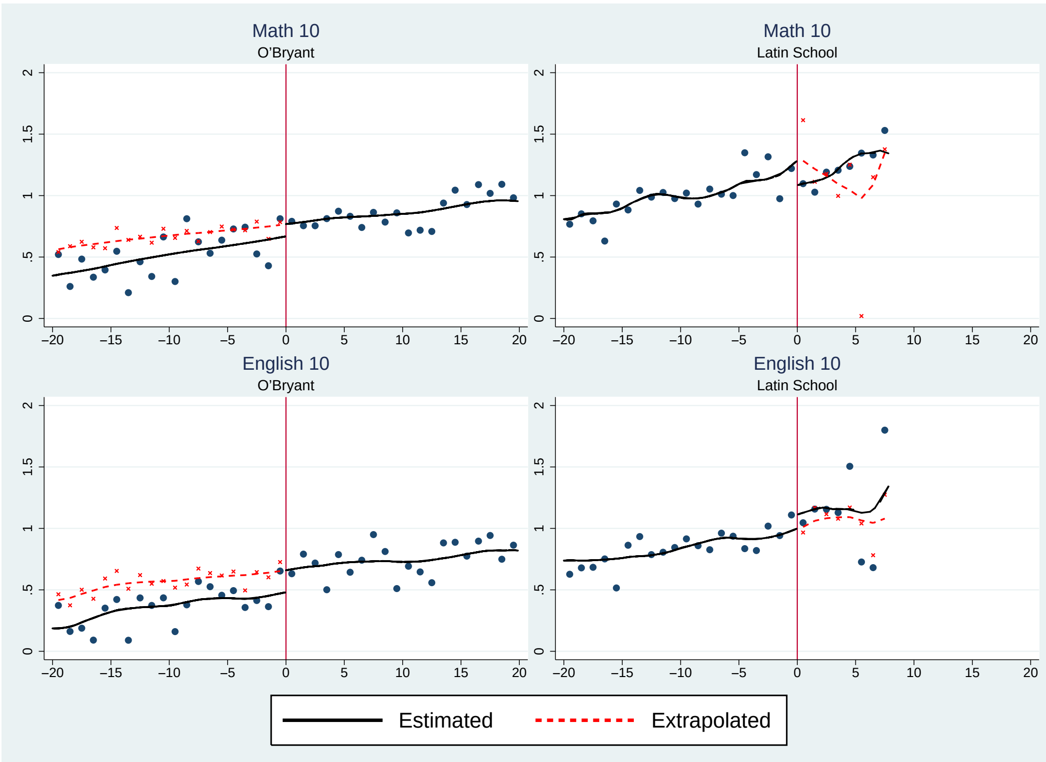


Figure 8: CIA-based Estimates of $E[Y_{1i}|r_i = c]$ and $E[Y_{0i}|r_i = c]$ for c in $[-20, 20]$. To the left of the O'Bryant cutoff, the estimates of $E[Y_{0i}|r_i = c]$ are fitted values for observed outcomes (the line labelled "Estimated") while the estimates of $E[Y_{1i}|r_i = c]$ are extrapolations. To the right of the BLS cutoff, the estimates of $E[Y_{1i}|r_i = c]$ are fitted values while the estimates of $E[Y_{0i}|r_i = c]$ are extrapolations. These estimates were constructed using the linear reweighting estimator discussed in the text.

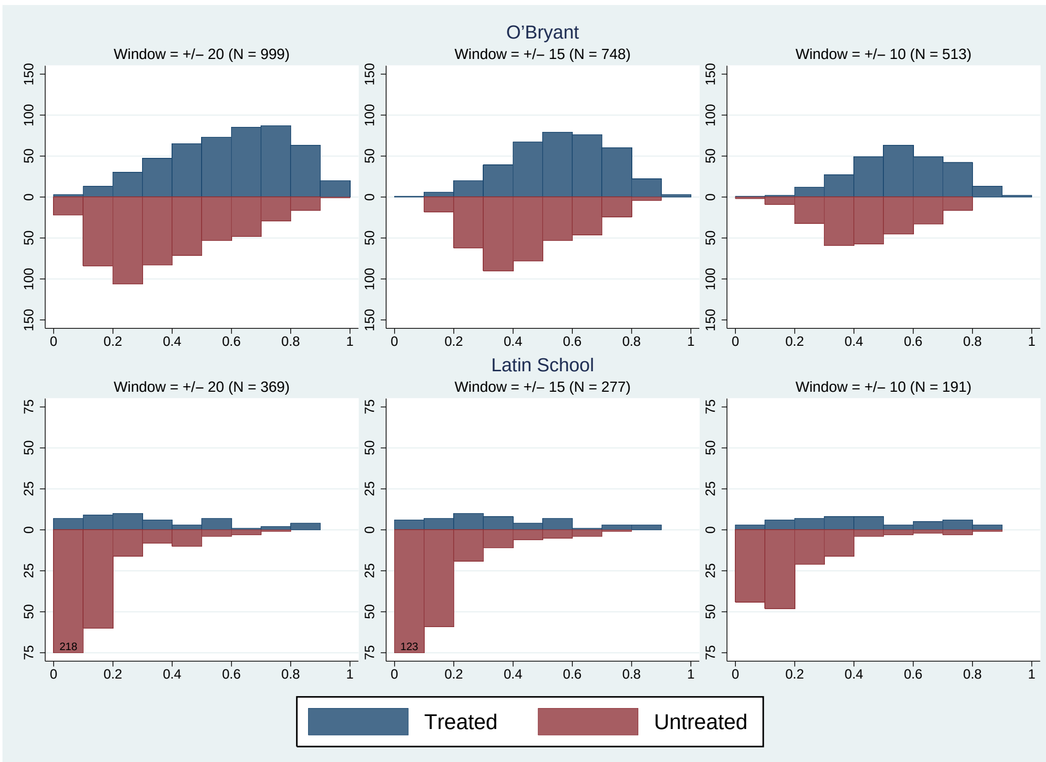


Figure 9: Histograms of Estimated Propensity Scores for 9th Grade Applicants to O'Bryant and BLS. The y-axis shows the observation count in each score-distribution decile. These estimates were constructed using a logit model fitted in the sample window indicated above each panel.

References

ABADIE, A. (2003): "Semiparametric Instrumental Variables Estimation of Treatment Response Models," *Journal of Econometrics*, 113(2), 231–263.

ABADIE, A., J. D. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

ABADIE, A., G. IMBENS, AND F. ZHENG (2011): "Robust Inference for Misspecified Models Conditional on Covariates," NBER Working Paper, 17442.

ABDULKADIROĞLU, A., J. D. ANGRIST, AND P. PATHAK (2012): "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," IZA Discussion Papers 6790, Institute for the Study of Labor.

ABDULKADIROĞLU, A., P. A. PATHAK, AND A. E. ROTH (2009): "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match," *American Economic Review*, 99(5), 1954–1978.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): "Hybrid and Size-Corrected Subsampling Methods," *Econometrica*, 77(3).

ANGRIST, J. D., AND I. FERNANDEZ-VAL (2010): "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," NBER Working Paper, 16566.

ANGRIST, J. D., AND G. W. IMBENS (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

ANGRIST, J. D., AND J.-S. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

BARNOW, B. S. (1972): *Conditions for the Presence or Absence of a Bias in Treatment Effect: Some Statistical Models for Head Start Evaluation*. University of Wisconsin-Madison.

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2012): "Inference on Treatment Effects After Selection Amongst High-Dimensional Controls," ArXiv:1201.02243.

CAMPBELL, D. T., AND J. STANLEY (1963): *Experimental and Quasi-experimental Designs for Research*. Rand McNally.

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.

Dehejia, R. H., and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94(448), 1053–1062.

DesJardins, S., and B. McCall (2008): "The Impact of Gates Millennium Scholars Program on the Retenion, College Finance- and Work-Related Choices, and Future Educational Aspirations of Low-Income Minority Students," Unpublished manuscript, University of Michigan.

DiNardo, J., and D. S. Lee (2011): *Program Evaluation and Research Designs*vol. 4 of *Handbook of Labor Economics*, chap. 5, pp. 463–536. Elsevier.

Dobbie, W., and R. G. Fryer (2012): "Exam High Schools and Academic Achievement: Evidence from New York City," NBER Working Paper, 17286.

Dong, Y., and A. Lewbel (2011): "Regression Discontinuity Marginal Threshold Treatment Effects," Working paper, Boston College.

Frandsen, B. R., M. Frölich, and B. Melly (2012): "Quantile Treatment Effects in the Regression Discontinuity Design," *Journal of Econometrics*, 168(2).

Goldberger, A. S. (1972a): "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations," Unpublished manuscript.

——— (1972b): "Selection Bias in Evaluating Treatment Effects: the Case of Interaction," Unpublished manuscript.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65(2), 261–294.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71(4), 1161–1189.

Horvitz, D. G., and D. J. Thompson (1952): "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47(260), 663–685.

Imbens, G., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Imbens, G., and K. Kalyanaraman (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79(3), 933–959.

Jackson, K. (2010): "Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago," *Economic Journal*, 120, 1399–1429.

Kline, P. M. (2011): "Oaxaca-Blinder as a Reweighting Estimator," *American Economic Review: Papers and Proceedings*, 101(3).

Lee, D. S., and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

Lewbel, A. (2007): "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75(2), 537–551.

Lord, F. M., and M. R. Novick (1972): *Statistical Theories of Mental Test Scores*. Addison-Welsley.

Pathak, P. A., and T. Sönmez (2008): "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *American Economic Review*, 98(4), 1636–1652.

——— (2011): "School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation," forthcoming, *American Economic Review*.

Appendix

Proof of Theorem 1

We continue to assume that GCIA and other LATE assumptions hold. Given these assumptions, Theorem 3.1 in Abadie (2003) implies that for any measurable function, $g(y_i, W_i, x_i)$, we have

$$E[g(y_i, W_i, x_i) \mid x_i, W_{1i} > W_{0i}] = \frac{1}{P[W_{1i} > W_{0i} \mid x_i]} E[\kappa(W_i, D_i, x_i) g(y_i, W_i, x_i) \mid x_i] \quad (27)$$

where

$$\kappa(W_i, D_i, x_i) = 1 - \frac{W_i(1-D_i)}{1 - P[D_i = 1 \mid x_i]} - \frac{(1-W_i) D_i}{P[D_i = 1 \mid x_i]}$$

and

$$E[g(Y_{Wi}, x_i) \mid x_i, W_{1i} > W_{0i}] = \frac{1}{P[W_{1i} > W_{0i} \mid x_i]} E[\kappa_W(W_i, D_i, x_i) g(y_i, x_i) \mid x_i],$$

where $W \in \{0, 1\}$ and

$$\begin{aligned} \kappa_0(W_i, D_i, x_i) &= (1 - W_i) \frac{P[D_i = 1 \mid x_i] - D_i}{(1 - P[D_i = 1 \mid x_i]) P[D_i = 1 \mid x_i]} \\ \kappa_1(W_i, D_i, x_i) &= W_i \frac{D_i - P[D_i = 1 \mid x_i]}{(1 - P[D_i = 1 \mid x_i]) P[D_i = 1 \mid x_i]}. \end{aligned}$$

Using the GCIA, we can simplify as follows:

$$\begin{aligned} & E[g(Y_{Wi}, x_i) \mid W_{1i} > W_{0i}, 0 \leq r_i \leq c] \\ &= E\{E[g(Y_{Wi}, x_i) \mid x_i, W_{1i} > W_{0i}] \mid W_{1i} > W_{0i}, 0 \leq r_i \leq c\} \\ &= \int \frac{1}{P[W_{1i} > W_{0i} \mid x_i]} E[\kappa_W(W_i, D_i, x_i) g(y_i, x_i) \mid X] \, dP[x_i \mid W_{1i} > W_{0i}, 0 \leq r_i \leq c] \\ &= \frac{1}{P[W_{1i} > W_{0i} \mid 0 \leq r_i \leq c]} \int E[\kappa_W(W_i, D_i, x_i) g(y, x_i) \mid x_i] \frac{P[0 \leq r_i \leq c \mid x_i]}{P[0 \leq r_i \leq c]} dP[x_i] \quad (28) \\ &= \frac{1}{P[W_{1i} > W_{0i} \mid 0 \leq r_i \leq c]} E\left[\kappa_W(W_i, D_i, x_i) \frac{P[0 \leq r_i \leq c \mid x_i]}{P[0 \leq r_i \leq c]} g(y_i, x_i)\right]. \end{aligned}$$

This implies that LATE can be written:

$$\begin{aligned} & E[Y_{1i} - Y_{0i} \mid W_{1i} > W_{0i}, 0 \leq r_i \leq c] \\ &= E[Y_{1i} \mid W_{1i} > W_{0i}, 0 \leq r_i \leq c] - E[Y_{0i} \mid W_{1i} > W_{0i}, 0 \leq r_i \leq c] \\ &= \frac{1}{P[W_{1i} > W_{0i} \mid 0 \leq r_i \leq c]} E\left[\psi(D_i, x_i) \frac{P[0 \leq r_i \leq c \mid x_i]}{P[0 \leq r_i \leq c]} y_i\right] \end{aligned}$$

where

$$\begin{aligned}\psi\left(D_i, x_i\right) &= \kappa_1\left(W_i, D_i, x_i\right)-\kappa_0\left(W_i, D_i, x_i\right) \\ &= \frac{D_i-P\left[D_i=1 \mid x_i\right]}{\left(1-P\left[D_i=1 \mid x_i\right]\right) P\left[D_i=1 \mid x_i\right]}.\end{aligned}$$

Finally, by setting $g\left(y_i, W_i, x_i\right)=1$ in equation (27) we get:

$$P\left[W_{1i}>W_{0i} \mid x_i\right] = E\left[\kappa\left(W_i, D_i, x_i\right) \mid x_i\right].$$

Using the same steps as in equation (28), the GCIA implies:

$$\begin{aligned}P\left[W_{1i}>W_{0i} \mid 0 \leq r_i \leq c\right] &= E\left\{P\left[W_{1i}>W_{0i} \mid x_i\right] \mid 0 \leq r_i \leq c\right\} \\ &= E\left[\kappa\left(W_i, D_i, x_i\right) \frac{P\left[0 \leq r_i \leq c \mid x_i\right]}{P\left[0 \leq x_i \leq c\right]}\right].\end{aligned}$$

Proof of Theorem 2

Theorem 1 in Angrist and Imbens (1995) implies:

$$\begin{aligned}E\left[y_i \mid D_i=1, x_i\right]-E\left[y_i \mid D_i=0, x_i\right] &= \sum_j P\left[w_{1i} \geq j>w_{0i} \mid x_i\right] E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j>w_{0i}, x_i\right] \\ E\left[w_i \mid D_i=1, x_i\right]-E\left[w_i \mid D_i=0, x_i\right] &= \sum_j P\left[w_{1i} \geq j>w_{0i} \mid x_i\right].\end{aligned}$$

Given the GCIA, we have:

$$\begin{aligned}& E\left\{E\left[y_i \mid D_i=1, x_i\right]-E\left[y_i \mid D_i=0, x_i\right] \mid 0 \leq r_i \leq c\right\} \\ & =\sum_j \int P\left[w_{1i} \geq j>w_{0i} \mid x_i\right] E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j>w_{0i}, x_i\right] dP\left[x_i \mid 0 \leq r_i \leq c\right] \\ & =\sum_j \int P\left[w_{1i} \geq j>w_{0i} \mid x_i, 0 \leq r_i \leq c\right] E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j>w_{0i}, x_i\right] dP\left[x_i \mid 0 \leq r_i \leq c\right] \\ & =\sum_j P\left[w_{1i} \geq j>w_{0i} \mid 0 \leq r_i \leq c\right] \\ & \quad \times \int E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j>w_{0i}, x_i\right] dP\left[x_i \mid w_{1i} \geq j>w_{0i}, 0 \leq r_i \leq c\right] \\ & =\sum_j P\left[w_{1i} \geq j>w_{0i} \mid 0 \leq r_i \leq c\right] E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i} \geq j>w_{0i}, 0 \leq r_i \leq c\right].\end{aligned}$$

The GCIA can similarly be shown to imply:

$$\begin{aligned}& E\left[E\left[w_i \mid D_i=1, x_i\right]-E\left[w_i \mid D_i=0, x_i\right] \mid 0 \leq r_i \leq c\right] \\ & =\sum_j P\left[w_{1i} \geq j>w_{0i} \mid 0 \leq r_i \leq c\right].\end{aligned}$$

Combining these results, the ACR can be written:

$$\frac{E\left\{E\left[y_i \mid D_i=1,x_i\right]-E\left[y_i \mid D_i=0,x_i\right] \mid 0\leq r_i\leq c\right\}}{E\left\{E\left[w_i \mid D_i=1,x_i\right]-E\left[w_i \mid D_i=0,x_i\right] \mid 0\leq r_i\leq c\right\}}$$
$$=\sum_j \nu_{jc} E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i}\geq j>w_{0i}, 0\leq r_i\leq c\right]$$

where

$$\nu_{ijc} \quad = \quad \frac{P\left[w_{1i}\geq j>w_{0i} \mid 0\leq r_i\leq c\right]}{\sum_\ell P\left[w_{1i}\geq \ell>w_{0i} \mid 0\leq r_i\leq c\right]}.$$