DISCUSSION PAPER SERIES

IZA DP No. 7429

Wanna Get Away?
RD Identification Away from the Cutoff

Joshua D. Angrist
Miikka Rokkanen

May 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Wanna Get Away?
RD Identification Away from the Cutoff

Joshua D. Angrist
MIT, NBER and IZA

Miikka Rokkanen
MIT

Discussion Paper No. 7429
May 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 7429
May 2013

ABSTRACT

Wanna Get Away? RD Identification Away from the Cutoff*

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 effect of treatment on inframarginal applicants is also of interest, but identification of such effects requires stronger assumptions than those 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.

JEL Classification: I21, I28, C21, C31

Keywords: regression discontinuity, human capital, school quality

Corresponding author:

Joshua D. Angrist
MIT Department of Economics
50 Memorial Drive
Building E52, Room 353
Cambridge, MA 02142-1347
USA
E-mail: angrist@mit.edu

* Our thanks to Parag Pathak for many helpful discussions and comments, and to Alberto Abadie, Victor Chernozhukov, Yingying Dong, Ivan Fernandez-Val, Patrick Kline, Arthur Lewbel and seminar participants at Berkeley, CREST, HECER, Stanford, the January 2013 Lech am Arlberg Labor Conference, and the 2013 SOLE meetings 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. This is a revision of NBER Working Paper 18662, December 2012.

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, often called 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 manipulates treatment. On the other hand, because the running variable is usually related to outcomes, claims for unconditional "as-if random assignment" are most credible for samples 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 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 (forthcoming) use parametric and nonparametric RD estimators to capture the causal effects of exam school attendance for applicants 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. Stronger applicants

[1] In an RD study of New York exam schools, Dobbie and Fryer (2012) similarly find little evidence of gains for admitted applicants at the cutoff.

who qualify more easily may get 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 their seats should affirmative action considerations be brought in to the admissions process.[2]

Motivated by the question of how exam school attendance affects achievement for inframarginal applicants, this paper tackles the theoretical problem of RD identification 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 framework, extrapolation is easy.

As it turns out, functional-form-based estimation procedures fail to produce compelling results for the empirical question that motivates 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: treatments is 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.

The special nature 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, though not for 7th grade applicants (for whom the available controls are not as good). These results lay the foundation for a matching strategy that identifies causal effects for inframarginal 9th applicants.

Our estimates of effects away from the cutoff are mostly in line with RD estimates of causal effects at the cutoff. In particular, away-from-the-cutoff estimates suggest BLS attendance has little effect on either math or English achievement, while the O'Bryant school may generate some

[2] In a study of tracking at Kenyan elementary schools, Duflo, Dupas, and Kremer (2011) use a combination of RD and a randomized trial to document treatment effect heterogeneity as a function of the running variable used to track students.

gains, especially in English Language Arts (ELA). The ELA gains for successful O'Bryant applicants approach one-fifth of a standard deviation. Perhaps surprisingly, therefore, those who seem most likely to gain from any expansion in exam school seats are relatively 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 the benefit of a BLS experience, at least as far as standardized test scores go.

2 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 many 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 scores fall near 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 scores 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.

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 the subject of ongoing litigation in New York. Our CIA-driven matching strategy is used here to answer two questions about the most- and least-selective of Boston's three exam schools; both questions are motivated by the contemporary debate over affirmative action in exam school admissions. Specifically, we ask:

1. How would inframarginal low-scoring applicants to O'Bryant, Boston's least selective exam

school, do if they were lucky enough to find seats at O'Bryant in spite of falling a decile or more below today's O'Bryant cutoff? 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 applicants to BLS, Boston's most selective exam school and one of the most selective in the country, fare if their BLS offers were withdrawn in spite of the fact that they qualify easily by today's standards? In other words, what if highly qualified applicants now at BLS had to settle for BLA?

The first of these questions addresses the impact of exam school attendance on applicants who currently fail to make the cut for any school but might do so with minority preferences restored or exam school seats added in an effort to boost minority enrollment. 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 solely by unconstitutional racial quotas. The thought experiment implicit in our second question sends high-scoring BLS students like McLaughlin back to BLA.

2.1 Data

The data used here merge 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. 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 1999-2008 or in 9th grade from 2001-2007. We focus on applicants enrolled in BPS at the time of application (omitting private school students) because we're interested in how an exam school education compares to regular district schools. 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.[3]

2.2 Exam School Admissions

The sharp CIA-based estimation strategy developed here is predicated on the notion that exam school offers 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 their 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. 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 (forthcoming), 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. A description of the manner in which these *sharp samples* are constructed appears in the appendix.[4]

The sharp RD treatment variable is an offer dummy, denoted D_{ik}, indicating applicants offered a seat at school k, determined separately as a function of rank 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

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

[4] Instead of defining sharp samples, a dummy for threshold crossing (qualification) can be used to instrument fuzzy offers. The extension of our CIA approach to fuzzy designs is discussed in Section 5, below. The construction of sharp sample produces an asymptotic efficiency gain, however, since those in the sharp sample are compliers in a setup that uses qualification as an instrument for offers (this is implied by results in Frolich (2007) and Hong and Nekipelov (2010), which show that the ability to predict compliance reduces the semiparametric efficiency bound for local average treatment effects.)

(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.

In sharp samples, offers are determined by the running variable, but not all offers are accepted. This can be seen in Figures 1a and 1b, which plot school-specific offer and enrollment rates around O'Bryant and BLS admissions cutoffs. Specifically, the figures show conditional means for sharp sample applicants in a one-unit binwidth, along with a conditional mean function smoothed using local linear regression (LLR).[5] As can be seen in Table 1, which reports estimates that go with these figures, 72% of 7th graders offered a seat at O'Bryant enroll there, while among 9th grade applicants offered an O'Bryant seat, 66% enroll. Enrollment rates are much higher for those offered a seat at BLS, while many applicants not offered a seat at BLS end up at Boston's second most selective exam school, BLA. At the same time, movement up the ladder of exam school selectivity is associated with dramatic changes in peer composition. This can be seen in Figure 2a and 2b, which plot peer achievement of applicants' classmates (as measured by baseline MCAS scores), for applicants within 20 percentile points of the O'Bryant and BLS cutoffs.

2.3 Results at the Cutoff

As a benchmark, we begin with estimates for marginal applicants. Figures 3a and 3b show little evidence of gains in 10th grade math scores for 7th grade applicants offered exam school seats. On the other hand, among both 7th and 9th grade applicants, 10th grade ELA scores seem to jump at the O'Bryant cutoff. The figure also hints at an O'Bryant-induced gain in math scores, though only for 9th grade applicants.

Our estimators of the effect of an exam school offer are derived from models for potential outcomes. Let Y_{1i} and Y_{0i} denote potential outcomes in treated and untreated states, with the observed outcome determined by

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

In a parametric setup, the conditional mean functions for potential outcomes given the running

[5] For school k, data in the estimation window were used to construct estimates of $\hat{E}[y_i|r_{ik}]$, 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. 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.

variable are modeled as:

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

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

Substituting polynomials 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 estimation sample pools data from different test years and application years, the parametric estimating equation for applicant i observed in year t is:

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

This model controls for test year effects, denoted α_t, and for application year, indexed by ℓ and indicated by dummies, $d_{i\ell}$. The model also includes a full set of application preference dummies, denoted p_{ij}.[6] 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 RD estimators differ from parametric in three ways. First, they narrow the estimation window when the optimal data-driven bandwidth falls below 20. 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:

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

Non-parametric RD estimates come from a kernel-weighted LLR fit of equation (4), estimated separately in the sharp sample of applicants to O'Bryant and BLS.

Consistent with the figures, estimates of (2) and (4), reported in Table 2, show little in the way of score gains at BLS. But the non-parametric estimates suggest an O'Bryant offer may boost 10th grade ELA scores for both 7th and 9th grade applicants. Other estimates are either smaller or less precise, though among 9th grade O'Bryant applicants, we see a marginally significant effect on math.

[6] As explained in the appendix, this controls for applicant-preference-group composition effects in the sharp sample.

Other estimates, not reported here, present a broad picture of small effects on 7th grade exam school applicants tested in 7th and 8th grade (see Abdulkadiroğlu, Angrist, and Pathak (forthcoming) for nonparametric estimates of effects on middle school scores.) Results for the 10th grade ELA scores of O'Bryant applicants offer the strongest evidence of an exam school gain.

2.4 To Infinity and Beyond: Parametric Extrapolation

The running variable is the star covariate in any RD scene, but the role played by the running variable is 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 changes in covariate values across the RD cutoff. As a practical matter, however, nonparametric inference procedures compare applicants with covariate values in a small - though not infinitesimal - neighborhood to the left of the cutoff with applicants whose covariate values put them in a small neighborhood to the right. This empirical comparison requires some extrapolation, however modest. Identification of causal effects away from the cutoff requires a more substantial extrapolative leap.

In a parametric setup such as described by (2) and (3), extrapolation is easy though not necessarily credible. For any distance, 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. The notation in (5) masks the extrapolation challenge 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 never seen. The dotted lines in Figure 4 show two equally plausible possibilities, implying different causal effects at $r_i = c$. It seems natural to use observations to the left of the cutoff in an effort to pin down functional form, and then extrapolate this to impute $E[Y_{0i}|r_i = c]$. With enough data, and sufficiently well-behaved conditional mean functions, $f_0(c)$ is identified for all values of c, including those never seen in the data. It's easy to see, however, why this approach may not generate robust or convincing findings.

The unsatisfying nature of parametric extrapolation emerges in Figures 5a and 5b. These figures show observed and imputed counterfactual 10th grade math scores for 7th and 9th grade applicants. Specifically, the figures plot nonparametric estimates of the observed conditional mean function $E[Y_{0i}|r_i = c]$ for O'Bryant applicants to the left of the cutoff, along with imputed $E[Y_{1i}|r_i = c]$

to the left. Similarly, for BLS applicants, the figures plot nonparametric estimates of observed $E[Y_{1i}|r_i = c]$ for applicants to the right of the cutoff, along with imputed $E[Y_{0i}|r_i = c]$ to the right. The imputations use linear, quadratic, and cubic specifications for $f_j(r_i)$. These models generate a wide range of estimates, especially as distance from the cutoff grows. For instance, the estimated effect of BLS attendance to the right of the cutoff for 9th grade applicants changes sign when the polynomial goes from second to third degree. This variability seems unsurprising and consistent with Campbell and Stanley (1963)'s 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.

Table 3, which reports the 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 too noisy to be useful at $c = +/-5$ or higher. Models setting $p = 2$ generate more precise estimates 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 at O'Bryant and small negative effects at BLS for both 7th and 9th grade applicants, though few of these estimates are significantly different from zero.[7]

Using Derivatives Instead

Dong and Lewbel (2012) propose an alternative to parametric extrapolation based on the insight that the derivatives of conditional mean functions are nonparametrically identified at the cutoff (a similar idea appears in Section 3.3.2 of DiNardo and Lee, 2011). First-order derivative-based extrapolation exploits the fact that

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

This approximation can be implemented using a nonparametric estimate of $f_j'(0)$.

The components of (6) are estimated consistently by fitting linear models to $f_j(r_i)$ in a neighborhood of the cutoff, using a data-driven bandwidth and slope terms that vary across the cutoff.

[7]To parallel Figure 5, the estimates in Table 3 are from models omitting controls for test year, application year and application preferences. Estimates from models with these controls differ little from those reported in the table.

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 estimated using equation (4). The innovation in this procedure 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 in this context as identifying a derivative that facilitates extrapolation. As a practical matter, the picture that emerges from derivative-based extrapolation of exam school effects is similar to that shown in Figure 5.

3 Calling on the CIA

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

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

This signal feature of the RD design implies that failure to control for r_i is the only source of omitted variables bias in estimates of the causal effect of D_i.

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 lagged achievement measures 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.[8]

To formalize this identification strategy, we 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, potential outcomes are assumed to be mean-independent of the running variable conditional on x_i. We also require treatment status to vary conditional on x_i:

[8] Cook (2008) credits Goldberger (1972a) and Goldberger (1972b) for the observation that when treatment status is determined solely by a pre-treatment test score, regression control for pre-treatment scores eliminates omitted variables bias. Goldberger credits Barnow (1972) and Lord and Novick (1972) for similar insights.

COMMON SUPPORT

$$0 < P[D_i = 1|x_i] < 1\, a.s.$$

The CIA and common support assumptions identify 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 and Bounding

Just as with conventional matching strategies (as in, for example, Heckman, Ichimura, and Todd (1998) and Dehejia and Wahba (1999)), the CIA assumption invoked here breaks the link between treatment status and potential outcomes, opening the door to identification of a wide range of average causal effects. In this case, however, the prior information inherent in an RD design is also available to guide our choice of the conditioning vector, x_i. Specifically, 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 simple test for (9) and (10). Mean independence is stronger than regression independence, of course, but regression testing procedures can embed flexible models that approximate nonlinear conditional mean functions. In practice, simple regression-based tests seem likely to provide the most useful specification check since such tests are likely to reject in the face of any sort of dependence between outcomes and running variable, while more elaborate specifications with many free parameters may lack the power to detect violations.[9]

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. This weaker bounded conditional independence assumption asserts that the CIA holds only over a limited range:

BOUNDED CONDITIONAL INDEPENDENCE ASSUMPTION (BCIA)

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

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 of the cutoff.

At first blush, the BCIA evokes 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, BCIA requires no choice of bandwidth or local polynomial smoothers with an eye to bias-variance trade-offs. Rather, the *largest* value of d that appears to satisfy BCIA defines the playing field for CIA-based estimation.

Beyond providing an opportunistic weakening of the CIA, the BCIA assumption allows us to avoid bias from counterfactual composition effects as distance from the cutoff grows. Moving, say, to the left of the BLS cutoff, BLS applicants start to fall below the BLA cutoff as well, thereby changing the relevant counterfactual from BLA to O'Bryant for BLS applicants not offered a seat there. The resulting change in Y_{0i} (where potential outcomes are indexed against BLS offers) is likely to be correlated with the BLS running variable with or without conditioning on x_i. To argue otherwise requires the distinction between BLA and O'Bryant to be of no consequence. BCIA avoids the resulting composition bias by requiring that we not extrapolate too far to the left of the BLS cutoff when looking at BLS applicants.

[9] Fan and Li (1996), Lavergne and Vuong (2000), Ait-Sahalia, Bickel, and Stoker (2001), and Angrist and Kuersteiner (2011) develop nonparametric conditional independence tests.

3.2 Alternative Assumptions and Approaches

CEI vs CIA

A weaker alternative to the CIA asserts conditional independence between average causal effects and the running variable, instead of between potential outcomes and the running variable. A Conditional Effect Ignorability (CEI) assumption, similar to that introduced by Angrist and Fernandez-Val (2010) in an instrumental variables setting, describes this as follows:

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 individually mean-independent of the running variable.[10]

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

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

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 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.

In practice, CIA-based estimates seem likely to be more useful than those derived from equation (13). For one thing, not being limited to identification near the cutoff, CIA-based estimation 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 applications. Finally, the CEI is harder to assess. CEI implies that the derivative of the conditional average treatment effect given covariates should be zero at the cutoff; as noted by Dong and Lewbel (2012), this derivative is non-parametrically identified (and given by the interaction term in the nonparametric estimating equation, (4)). In practice, however, samples large enough for reliable nonparametric estimates of conditional mean functions may still generate inconclusive results for derivatives. Not surprisingly, therefore, our experiments with CEI estimators for Boston exam school applicants failed to produce estimates that seem precise enough to be useful.

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

CIA in a Latent Factor Model

Instead of being intrinsically meaningful, running variables and the conditioning variables in our CIA assumption can be modeled as noisy measures of a single underlying ability measure, θ_i. Suppose that a version of the CIA holds for latent ability. In other words, we're prepared to assume only that

$$E[Y_{ji}|r_i, \theta_i] = E[Y_{ji}|\theta_i];\ j = 0, 1 \tag{14}$$

for unobserved θ_i. Not knowing θ_i, however, we condition on a vector of proxies, x_i, instead. Given (14), when is proxy conditioning good enough?

A sufficient condition for proxy conditioning to support identification in a latent factor model is that

$$r_i \perp\!\!\!\perp \theta_i \mid x_i. \tag{15}$$

In the exam school setting, (15) says that the running variable contains no further information about ability after conditioning on baseline scores. Conditions (14) and (15) are easily seen to imply covariate-based CIA. In ongoing work, Rokkanen (2013) explores identification strategies for RD in more general latent factor models where identification is based on (14), without also assuming (15).

Other Related Work

Battistin and Rettore (2008) also consider matching estimates in an RD setting, though they don't exploit an RD-specific conditional independence condition. Rather, in the spirit of Lalonde (1986), Battistin and Rettore propose to validate a generic matching estimator by comparing non-parametric RD estimates with conventional matching estimates constructed at the cutoff. If matching and RD produce similar results at the cutoff, matching seems worth exploring away from the cutoff as well.

Other 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 that 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 causal effects might change away from the cutoff. 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 option to fix the marginal distribution of regressors observed in any particular sample. The empirical consequences of regressor conditioning boil down to an adjustment of standard errors.[11] 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." Our approach takes up this challenge.[12]

3.3 CIA-based Estimators

We economize on notation by omitting explicit conditioning on running variable values falling in the $[-d, d]$ interval; 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, the CIA leads to the following matching-style estimand:

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

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:

$$\begin{aligned} &E[Y_{1i} - Y_{0i}|0 < r_i \leq c] = \\ &\qquad E\{E[y_i|x_i, D_i = 1] - E[y_i|x_i, D_i = 0]|0 < r_i \leq c\} \end{aligned} \tag{17}$$

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

$$\begin{aligned} &E[Y_{1i} - Y_{0i}| - c \leq r_i < 0] = \\ &\qquad E\{E[y_i|x_i, D_i = 1] - E[y_i|x_i, D_i = 0]| - c \leq r_i < 0\} \end{aligned} \tag{18}$$

We consider two estimators of (16), (17) and (18). The first is a 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

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

[12] Moving in a different direction, Jackson (2010) outlines an extrapolation approach that identifies inframarginal effects at exam 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. Jackson (2010) identifies effects away from the cutoff by differences-in-differences style contrasts between infra-marginal high- and low-scoring applicants with different rankings. Cook and Wing (2012) explore a similar idea, offering supportive Monte Carlo evidence for this approach.

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:

$$\begin{aligned} E[y_i|x_i, D_i = 0] &= x_i'\beta_0 \\ E[y_i|x_i, D_i = 1] &= x_i'\beta_1 \end{aligned} \tag{19}$$

Linearity is not really restrictive since the parametrization for $x_i'\beta_j$ can be rich and flexible. Substituting in (16), we have

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

with similar expressions based on (17) and (18).

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

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

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

$$E[Y_{1i} - Y_{0i}|0 < r_i \leq c] = E\left\{\frac{y_i[D_i - \lambda(x_i)]}{\lambda(x_i)\left[1-\lambda(x_i)\right]} \cdot \frac{P[0 < r_i \leq c|x_i]}{P[0 < r_i \leq c]}\right\}. \tag{21}$$

Similar formulas give the average effect for non-treated applicants and average effects at specific, possibly narrow, ranges of running variable values. The empirical counterpart of (21) 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. Note also that if $c = d$, the estimand in (21) 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).[13]

[13] 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 treated applicants in a school-k sharp sample. When the estimand targets average effects at specific $r_i = c$, as opposed to over an interval, the probabilities $P[r_i = c|x_i]$ and $P[r_i = c]$ needed for (21) become densities.

4 The CIA in Action at Boston Exam Schools

We start by testing BCIA in estimation windows that set d equal to 10, 15, and 20. Regressions used for testing control for baseline test scores along with indicators of special education status, limited English proficiency, eligibility for free or reduced price lunch, race (black/Asian/Hispanic) and sex, as well as indicators for test year, application year and application preferences. Baseline score controls for 7th grade applicants include 4th grade math and ELA scores, while for 9th grade applicants, baseline scores include 7th grade ELA scores and 8th grade math scores.

CIA test results, reported in Table 4, show that conditioning fails to eliminate the relationship between running variables and potential outcomes for 7th grade applicants; most of the estimated coefficients are significantly different from zero for both 10th grade math and ELA scores. At the same time, test results for 9th grade applicants seem promising. Most test statistics (that is, running variable coefficient estimates) for 9th grade applicants are smaller than the corresponding statistics for 7th grade applicants, and only one is significantly different from zero (this is for math scores to the left of the BLS cutoff in the $d = 20$ window). It should be noted, however, that few 9th grade applicants fall to the right of the BLS cutoff. CIA tests for BLS applicants with $D_i = 1$ are forgiving because the sample for this group is small.[14]

We complement formal CIA testing with a graphical tool motivated by an observation in Lee and Lemieux (2010): in a randomized trial using a uniformly distributed random number to determine treatment assignment, this number becomes the running variable for an RD design. The relationship between outcomes and running variable should be flat, however, except possibly for a jump at the quantile cutoff which determines proportion treated. Our CIA assumption implies this same pattern. Figure 6 therefore plots 10th grade math and ELA residuals constructed by partialing out x_i against running variables in a $d = 20$ window. The figure shows conditional means for all applicants in one-unit binwidths, along with conditional mean functions smoothed using local linear regression. Consistent with the test results reported in Table 4, Figure 6 shows a strong positive relationship between outcome residuals and running variables for 7th grade applicants. For 9th grade applicants, however, the relationship between outcome residuals and running variables is essentially flat, except perhaps for ELA scores in the BLS sample.

The difference in CIA test results for 7th and 9th grade applicants may be due to the fact that baseline scores for 9th grade applicants come from a grade closer to the outcome test grade than for 7th grade applicants. In combination with demographic control variables and 4th grade scores,

[14] The unchanging sample size to the right of the BLS cutoff as d shrinks reflects the high BLS admissions threshold for 9th grade applicants: the $d = 10$ limit isn't binding for BLS on the right.

7th or 8th grade MCAS scores do a good job of eliminating the running variable from 9th graders' conditional mean functions for 10th grade scores. By contrast, the most recent baseline test scores available for 7th grade applicants are from 4th grade tests.[15] In view of the results in Table 4 and Figure 6, the CIA-based estimates that follow are for 9th grade applicants only.

Columns 1-4 of Table 5 report linear reweighting estimates of average treatment effects. These are 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

$$\begin{aligned} &E[Y_{1i} - Y_{0i}|0 < r_i \leq d] \\ &\quad = (\beta_1 - \beta_0)' E[x_i|0 < r_i \leq d], \end{aligned} \tag{22}$$

while that for O'Bryant is

$$\begin{aligned} &E[Y_{1i} - Y_{0i}| - d \leq r_i < 0] \\ &\quad = (\beta_1 - \beta_0)' E[x_i| - d \leq r_i < 0], \end{aligned} \tag{23}$$

where β_0 and β_1 are defined in (19). The BLS estimand is an average effect of treatment on the treated, since treated observations in the estimation window must have positive running variables. Likewise, 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 5 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 $.09\sigma$ when $d = 10$ to $.16\sigma$ when $d = 20$, though the estimate effect for $d = 10$ is only marginally significantly different from zero. Linear reweighting results for the ELA scores of O'Bryant applicants are clear cut, however, 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 CIA estimates in columns 1 and 3 of Table 5 to the nonparametric O'Bryant RD estimates at the cutoff of $.13\sigma$ (SE=.07) in math and $.18\sigma$ (SE=.07) for ELA, shown in column 3 of Table 2.

Figure 7 completes the picture on effects away from the cutoff 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 fitted values from regression models for observed outcomes, while the estimates of $E[Y_{1i}|r_i = c]$ are implicitly 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

[15] The addition of quadratic and cross-subject interaction terms in baseline scores fails to improve CIA test results for 7th grade applicants.

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 (19) and smoothing the results using local linear regression.[16] The figure presents a picture consistent with that arising from the estimates in Table 5. In particular, the extrapolated BLS effects are small (for ELA) or noisy (for math), while the O'Bryant extrapolation reveals a remarkably stable gain 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 estimation of the effect of exam school offers seems like a good setting for propensity score methods, since the conditioning set includes 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 4 and the linear reweighting estimates in columns 1-4 of Table 5.[17]

The estimated propensity score distributions for admitted and rejected applicants exhibit a substantial degree of overlap. This is documented in Figure 8, which plots the histogram of estimated scores for treated and control observations above and below a common horizontal axis. Not surprisingly, the larger sample of O'Bryant applicants 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; above the first decile, there appear to be more than enough for accurate inference. By contrast, few untreated 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 .9 in one window), the BLS estimates fail to reflect outcomes for applicants with admissions probabilities above this value. Figure 8 documents other noteworthy features of conditional-on-score comparisons: 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 corresponding untreated observations, of which many more are low-scoring than high.

[16] Smoothing here uses the edge kernel with Stata's default bandwidth.

[17] Propensity score models for the smaller sample of BLS applicants omit test date and application preference dummies.

The propensity-score-weighted estimates reported in columns 5-8 of Table 5 are remarkably consistent with the linear reweighting estimates shown in columns 1-4 of the table. In particular, the estimates here suggest most BLS students would do no worse if they had had to go to BLA instead, while low scoring O'Bryant applicants might enjoy substantial gains in ELA were they offered a seat at O'Bryant. At the same time, the propensity score estimates for BLS applicants reported in columns 6 and 8 are highly imprecise. These BLS 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. Linear reweighting looks like an attractive procedure in this context.[18]

5 Fuzzy CIA Models

Estimates of the effect of O'Bryant offers on the ELA scores of 9th grade applicants are reasonably stable as distance from the cutoff grows. By contrast, the estimated 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, for example, estimated O'Bryant math effects are only marginally significantly different from zero, while the estimate in a window of width 20 is almost twice as large and significant (at $.16\sigma$ 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 4, show no violations of CIA.

An alternative explanation for the pattern of O'Bryant math estimates plotted in Figure 7 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 5 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 reduced form estimates

[18]The standard errors reported in this table use a bootstrap with 500 replications. Bootstrap standard errors provide asymptotically valid confidence intervals for estimators like (21) 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 the CIA are imprecise. Still, the general pattern is similar, suggesting positive effects at O'Bryant only.

across running variable values are meaningful only after rescaling. In principle, IV methods make the appropriate adjustment. A question that arises 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.

We estimate and interpret the causal effects of exam school enrollment by adapting the dummy treatment/dummy instrument framework outlined in Abadie (2003). This framework allows for unrestricted treatment effect heterogeneity in potentially nonlinear IV models with covariates. 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 setup is a generalized version of CIA:

GENERALIZED CONDITIONAL INDEPENDENCE ASSUMPTION (GCIA)

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

GCIA can be assumed to hold in a $d-$neighborhood of the cutoff as with BCIA. We also maintain the common support assumption given in Section 3.

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 3.3, the expectations 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 shift treatment one way. Assuming that the instrument D_i satisfies monotonicity with $W_{1i} \geq W_{0i}$, and that for some i 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 facilitates 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 to a fuzzy RD setup that identifies causal effects away from the cutoff. This requires a conditional first stage, described below:

CONDITIONAL FIRST STAGE

$$P[W_{1i} = 1|x_i] > P[W_{0i} = 1|x_i] \; a.s.$$

Given GCIA, common support, monotonicity, and a conditional first stage, the following identification result can be established (see the appendix for details):

THEOREM 1 (FUZZY CIA EFFECTS)

$$\begin{aligned} &E[Y_{1i} - Y_{0i}|W_{1i} > W_{0i}, 0 < r_i \leq c] \\ &\qquad = \frac{1}{P[W_{1i} > W_{0i}|0 < r_i \leq c]} E\left\{\psi(D_i, x_i)\frac{P[0 < r_i \leq c|x_i]}{P[0 < r_i \leq c]} y_i\right\} \qquad (24) \\ &for \quad \psi(D_i, x_i) \equiv \frac{D_i - \lambda(x_i)}{\lambda(x_i)[1 - \lambda(x_i)]} \qquad (25) \end{aligned}$$

Estimators based on (24) capture causal effects for compliers with running variable values falling into any range over which there's common support.[19]

At first blush, it's not immediately clear how to estimate the conditional compliance probability, $P[W_{1i} > W_{0i}|0 < r_i \leq c]$, appearing in the denominator of (24). Because everyone to the right of the cutoff is treated, there would seem to be no data available to estimate compliance rates conditional on $0 < r_i \leq c$ (in the original LATE framework, the IV first stage measures the probability of

[19]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 this theorem along the lines suggested by Theorem 3.1 in Abadie (2003) identify the marginal distributions of Y_{0i} and Y_{1i}.

compliance). Paralleling an argument in Abadie (2003), however, the appendix shows that

$$P[W_{1i} > W_{0i} | 0 < r_i \leq c] = E\left\{\kappa\left(W_i, D_i x_i\right) \frac{P\left[0 < r_i \leq c \mid x_i\right]}{P\left[0 < r_i \leq c\right]}\right\} \tag{26}$$

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 some applicants, however, the exam school treatment is mediated by years of attendance rather than a simple go/no-go decision. We develop a fuzzy CIA estimator for ordered treatments by adapting a result from Angrist and Imbens (1995). The ordered treatment 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 $j = 0, 1, 2, ..., J$. We assume also that potential treatments, w_{1i} and w_{0i}, satisfy monotonicity with $w_{1i} \geq w_{0i}$ and generate 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

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

Wald-type IV estimators therefore capture a weighted average of the average causal effect of increasing w_i from $j-1$ to j, for compliers whose treatment intensity is moved by the instrument from below j to above j. The weights are given by the impact of the instrument on the cumulative distribution function (CDF) of the endogenous variable at each intensity.

The GCIA assumption allows us to establish a similar result in a fuzzy RD setup with an ordered treatment. The following is shown in the appendix:

THEOREM 2 (FUZZY AVERAGE CAUSAL RESPONSE)

$$\frac{E\left\{E\left[y_i \mid D_i=1,x_i\right]-E\left[y_i \mid D_i=0,x_i\right] \mid 0<r_i\leq c\right\}}{E\left\{E\left[w_i \mid D_i=1,x_i\right]-E\left[w_i \mid D_i=0,x_i\right] \mid 0<r_i\leq c\right\}} = \sum_j \nu_{jc} E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i}\geq j>w_{0i}, 0<r_i\leq c\right] \quad (27)$$

where

$$\nu_{jc} = \frac{P\left[w_{1i}\geq j>w_{0i} \mid 0<r_i\leq c\right]}{\sum_\ell P\left[w_{1i}\geq \ell>w_{0i} \mid 0<r_i\leq c\right]} \quad (28)$$

This theorem 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 incremental average causal response, $E\left[Y_{ji}-Y_{j-1,i} \mid w_{1i}\geq j>w_{0i}, 0<r_i\leq c\right]$, is weighted by the conditional probability the instrument moves the ordered treatment through the point at which the incremental effect is evaluated.

In practice, we estimate the left hand side of (27) by fitting linear models with covariate interactions to the reduced form and first stage. The resulting estimation procedure adapts Kline (2011) to an ordered treatment and works as follows: estimate conditional linear reduced forms interacting D_i and x_i; use these estimates to construct the desired average reduced form effect as in (22) and (23); divide by a similarly constructed average first stage.[20] The same procedure can be used to estimate (27) for a Bernoulli treatment like W_i, in which case the average causal response identified by Theorem 2 becomes the average causal effect identified by Theorem 1 (though the corresponding estimates won't be algebraically the same unless the propensity score model used under Theorem 1 is linear).

5.2 Fuzzy Estimates

As with the sharp estimates discussed in Section 4, fuzzy enrollment effects are estimated for applicants 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 in columns 1-4 of Table 6, which report estimates of the effect of exam school offers on exam school enrollment, constructed separately for O'Bryant and BLS applicants using equation (26). The propensity score model is the same as that used to construct the estimates in Table 5 (Table 6 shows separate first stage estimates for the math and ELA samples, as these differ slightly). Given this stable first stage, its unsurprising that 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 little as a function

[20]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 (27) is $\phi_0+\phi_1'\mu_{xc}$, where $\mu_{xc}=E[x_i|0\leq r_i\leq c]$.

of d. The pattern here is consistent with that in Table 5, with small and statistically insignificant effects at BLS, and evidence of large effects at O'Bryant. Estimates of O'Bryant effects on ELA scores range from an impressive gain of $.38\sigma$ when $d = 20$, to a still-substantial $.27\sigma$ when the window is half as wide. The estimated O'Bryant effects on math scores are also considerable, varying from $.17\sigma$ to $.23\sigma$.

The gains for inframarginal applicants who enroll at O'Bryant are perhaps too large to be credible and may therefore signal failure of the underlying exclusion restriction, which channels all causal effects of an exam through an enrollment dummy. Many who start in an exam school drop out, so we'd like to adjust these estimates 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 right-hand side of equation (27), using the modified linear 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 5.

First stage estimates for years of exam school enrollment, reported in columns 1-4 of Table 7, 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 ACR estimates, reported in columns 5-8 of the table, are in line with those in Table 6, 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 $.05\sigma$, with a standard error of roughly the same magnitude. This compares with estimates of about the same size in column 8 of Table 6, but standard errors for the latter are five or more times larger. The precision gain here would seem to come 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 when constructing the sharp estimates in Table 5.

ELA estimates for O'Bryant show gains of about $.14\sigma$ per year of exam school exposure, a finding that appears to be more stable across window width than the corresponding dummy enrollment estimates in column 7 of Table 6. This comparison suggests that some of the variability seen in the estimates in Table 6 comes from a failure to adjust for small changes in the underlying first stage for years of enrollment across windows (as can be seen in column 3 of Table 7). At the same time, the estimated O'Bryant math gains in column 5 of Table 7 still fade in a narrower window, a pattern seen for the O'Bryant math estimates in Tables 5 and 6.

6 Summary and Directions for Further Work

RD estimates of the effect of Boston exam school offers 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. This observation motivates RD-inspired identification strategies for causal effects away from the cutoff. Parametric extrapolation seems like a natural first step, but a parametric approach generates unsatisfying estimates of the effects of exam school offers, sensitive to functional form and too imprecise to be useful. 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 therefore 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 underlying ignorability has strong testable implications that are easily checked in this context. 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 the 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 these relatively recent baseline tests are a powerful predictor of future scores. Lagged outcomes are better predictors, in fact, than the running variable itself, 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 (forthcoming). At the same time, CIA-based estimates using both linear and propensity score models generate robust evidence of strong gains in English for unqualified inframarginal O'Bryant applicants. Evidence of 10th grade grade ELA gains also emerge from the RD estimates of exam school effects reported by Abdulkadiroğlu, Angrist, and Pathak (forthcoming), 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 re-introducing affirmative action considerations in exam school admissions.

We also modify CIA-based identification strategies for fuzzy RD and use this modification to estimate the effects of exam school enrollment and years of exam school attendance, in addition to the reduced form effects of exam school admissions offers. A fuzzy analysis allows us to explore the possibility that changes in reduced form offer effects as a function of the running variable are driven by changes in an underlying first stage for exam school exposure. Interestingly, the fuzzy extension opens the door to 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) and Abadie (2003) breaks down in a conventional RD framework because the distribution of treatment status is degenerate conditional on the running variable. By taking the running variable out of the equation, our framework circumvents this problem, a feature we plan to exploit in future work on distributional outcomes.

In a parallel and ongoing investigation, Rokkanen (2013) develops identification strategies for RD designs in which the CIA conditioning variable is an unobserved latent factor. Multiple noisy indicators of the underlying latent factor provide the key to away-from-the-cutoff identification in this context. An important unsolved econometric problem implicit in our empirical strategy is causal inference conditional on a pretest. Estimators that condition on the results of a specification test may have sampling distributions for which conventional asymptotic approximations are poor. 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 applied fruitfully in this context.

Finally, the mixed results reported here raise the question of what might explain the variation in our estimates across schools. In a pair of recent papers, Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak (2011) and Angrist, Cohodes, Dynarski, Pathak, and Walters (2013) document large gains at Boston charter high schools when using admissions lotteries to estimate the effects of charter attendance relative to regular district schools. These gains appear to vary inversely with students' baseline achievement, suggesting that the quality of the implicit counterfactual may be an important driver of the treatment effects arising from school choice. The fallback school for most O'Bryant applicants (a regular district school) may have lower value-added than the fallback school for BLS applicants (mostly the BLA exam school), even though the gain in peer quality is larger at the admissions cutoff for the latter. In ongoing work, we're continuing to explore the nexus linking school choice, school quality, and measures of students' baseline ability.

Table 1: Destinations of Applicants to O'Bryant and Boston Latin School

	O'Bryant		Latin School	
	D=0	D=1	D=0	D=1
	(1)	(2)	(3)	(4)
	Panel A. 7th Grade Applicants			
Traditional Boston public schools	1.00	0.28	0.08	0.05
O'Bryant	0.00	0.72	0.06	0.00
Latin Academy	0.00	0.00	0.86	0.01
Latin School	...	...	0.00	0.93
	Panel B. 9th Grade Applicants			
Traditional Boston public schools	1.00	0.34	0.15	0.04
O'Bryant	0.00	0.66	0.00	0.00
Latin Academy	...	...	0.86	0.02
Latin School	...	...	0.00	0.94

Notes: This table 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 applied 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: Reduced Form Estimates for 10th Grade MCAS Scores

	Parametric		Nonparametric	
	O'Bryant	Latin School	O'Bryant	Latin School
	(1)	(2)	(3)	(4)
	Panel A. 7th Grade Applicants			
Math	-0.011	-0.034	0.034	-0.055
	(0.100)	(0.060)	(0.056)	(0.039)
	1832	1854	1699	1467
ELA	0.059	0.021	0.125**	0.000
	(0.103)	(0.095)	(0.059)	(0.061)
	1836	1857	1778	1459
	Panel B. 9th Grade Applicants			
Math	0.166	-0.128	0.128*	-0.144*
	(0.109)	(0.117)	(0.066)	(0.076)
	1559	606	1386	361
ELA	0.191*	0.097	0.180***	0.048
	(0.112)	(0.187)	(0.066)	(0.106)
	1564	607	1532	458

Notes: This table reports estimates of the effects of exam school offers on 10th grade 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 are shown in parentheses. The number of observations is reported below standard errors.

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

Table 3: Parametric Extrapolation Estimates for 10th Grade Math

	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: 7th Grade Applicants				
Linear	0.041	0.061	0.085	0.110	-0.076**	-0.051	-0.021	0.010
	(0.052)	(0.057)	(0.072)	(0.093)	(0.035)	(0.040)	(0.049)	(0.061)
	1832	1832	1832	1832	1854	1854	1854	1854
Quadratic	0.063	0.204	0.391*	0.588	-0.056	-0.111	-0.152	-0.161
	(0.075)	(0.125)	(0.237)	(0.384)	(0.051)	(0.088)	(0.162)	(0.261)
	1832	1832	1832	1832	1854	1854	1854	1854
Cubic	0.034	0.167	0.247	0.266	-0.050	-0.096	-0.106	-0.065
	(0.110)	(0.336)	(0.921)	(1.927)	(0.073)	(0.220)	(0.589)	(1.215)
	1832	1832	1832	1832	1854	1854	1854	1854
				Panel B: 9th Grade Applicants				
Linear	0.088	0.083	0.077	0.071	-0.090	0.079	0.291***	0.502***
	(0.057)	(0.059)	(0.070)	(0.088)	(0.065)	(0.063)	(0.108)	(0.168)
	1559	1559	1559	1559	606	606	606	606
Quadratic	0.170**	0.264**	0.427*	0.639*	-0.147*	-0.106	0.078	0.409
	(0.085)	(0.133)	(0.237)	(0.372)	(0.088)	(0.142)	(0.303)	(0.713)
	1559	1559	1559	1559	606	606	606	606
Cubic	0.143	0.069	-0.059	-0.355	-0.061	0.196	0.996	3.094
	(0.119)	(0.327)	(0.851)	(1.735)	(0.118)	(0.338)	(0.910)	(2.543)
	1559	1559	1559	1559	606	606	606	606

Notes: This table reports estimates of effects on 10th grade Math scores away from the RD cutoff at points indicated in the column heading. 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. The estimates are based on first, second, and third order polynomials, as indicated in rows of the table. Robust standard errors are shown in parentheses.

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

Table 4: 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)
				Panel A. 7th Grade Applicants				
20	0.022***	0.015***	0.008***	0.014***	0.015***	0.006	0.013***	0.018***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.004)	(0.005)	(0.003)	(0.003)
	838	618	706	748	840	621	709	750
15	0.023***	0.015***	0.010***	0.012***	0.014**	0.006	0.007	0.015***
	(0.006)	(0.005)	(0.003)	(0.003)	(0.005)	(0.006)	(0.005)	(0.005)
	638	587	511	517	638	590	514	519
10	0.030***	0.016**	0.010*	0.007	0.024**	0.001	0.012	0.012
	(0.009)	(0.008)	(0.006)	(0.005)	(0.010)	(0.009)	(0.010)	(0.008)
	419	445	335	347	421	447	338	348
				Panel B. 9th Grade Applicants				
20	0.002	0.005	0.008**	0.018	0.003	0.002	0.006	0.055
	(0.004)	(0.003)	(0.003)	(0.028)	(0.004)	(0.004)	(0.005)	(0.053)
	513	486	320	49	516	489	320	50
15	0.010	0.000	0.006	0.018	0.009	-0.000	0.000	0.055
	(0.006)	(0.005)	(0.006)	(0.028)	(0.006)	(0.006)	(0.007)	(0.053)
	375	373	228	49	376	374	229	50
10	0.003	-0.001	0.007	0.018	0.014	-0.004	0.014	0.055
	(0.011)	(0.009)	(0.009)	(0.028)	(0.011)	(0.010)	(0.015)	(0.053)
	253	260	142	49	253	261	142	50

Notes: This table reports regression-based tests of the conditional independence assumption described in the text. Cell entries show the coefficient on the 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, as well as indicators for test year, application year and application preferences. 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 5: CIA Estimates of the Effect of Exam School Offers for 9th Grade Applicants

	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	0.156***	-0.031	0.198***	0.088	0.148***	-0.028	0.251***	0.054
	(0.039)	(0.094)	(0.041)	(0.084)	(0.052)	(0.192)	(0.090)	(0.207)
N untreated	513	320	516	320	509	320	512	320
N treated	486	49	489	50	482	49	485	50
15	0.129***	-0.080	0.181***	0.051	0.116**	-0.076	0.202***	0.018
	(0.043)	(0.055)	(0.047)	(0.088)	(0.052)	(0.161)	(0.069)	(0.204)
N untreated	375	228	376	229	373	228	374	229
N treated	373	49	374	50	370	49	371	50
10	0.091*	-0.065	0.191***	-0.000	0.123*	-0.093	0.186**	-0.052
	(0.054)	(0.054)	(0.055)	(0.097)	(0.070)	(0.249)	(0.073)	(0.356)
N untreated	253	142	253	142	253	142	253	142
N treated	260	49	261	50	258	49	259	50

Notes: This table reports estimates of the effect of 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, while columns 5-8 report results from inverse propensity score weighting, as described in the text. Controls are the same as used to construct the test statistics except that the propensity score models for Latin School omit test year and application preference 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 6: Fuzzy CIA Estimates of LATE (Exam School Enrollment) for 9th Grade Applicants

	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	0.659***	0.898***	0.660***	0.900***	0.225**	-0.031	0.380**	0.060
	(0.062)	(0.054)	(0.062)	(0.052)	(0.088)	(0.217)	(0.183)	(0.231)
N untreated	509	320	512	320	509	320	512	320
N treated	482	49	485	50	482	49	485	50
15	0.666***	0.898***	0.667***	0.900***	0.174**	-0.085	0.302**	0.020
	(0.047)	(0.048)	(0.050)	(0.047)	(0.080)	(0.177)	(0.125)	(0.225)
N untreated	373	228	374	229	373	228	374	229
N treated	370	49	371	50	370	49	371	50
10	0.670***	0.898***	0.678***	0.900***	0.184*	-0.104	0.274**	-0.058
	(0.055)	(0.048)	(0.050)	(0.047)	(0.108)	(0.274)	(0.121)	(0.402)
N untreated	253	142	253	142	253	142	253	142
N treated	258	49	259	50	258	49	259	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.

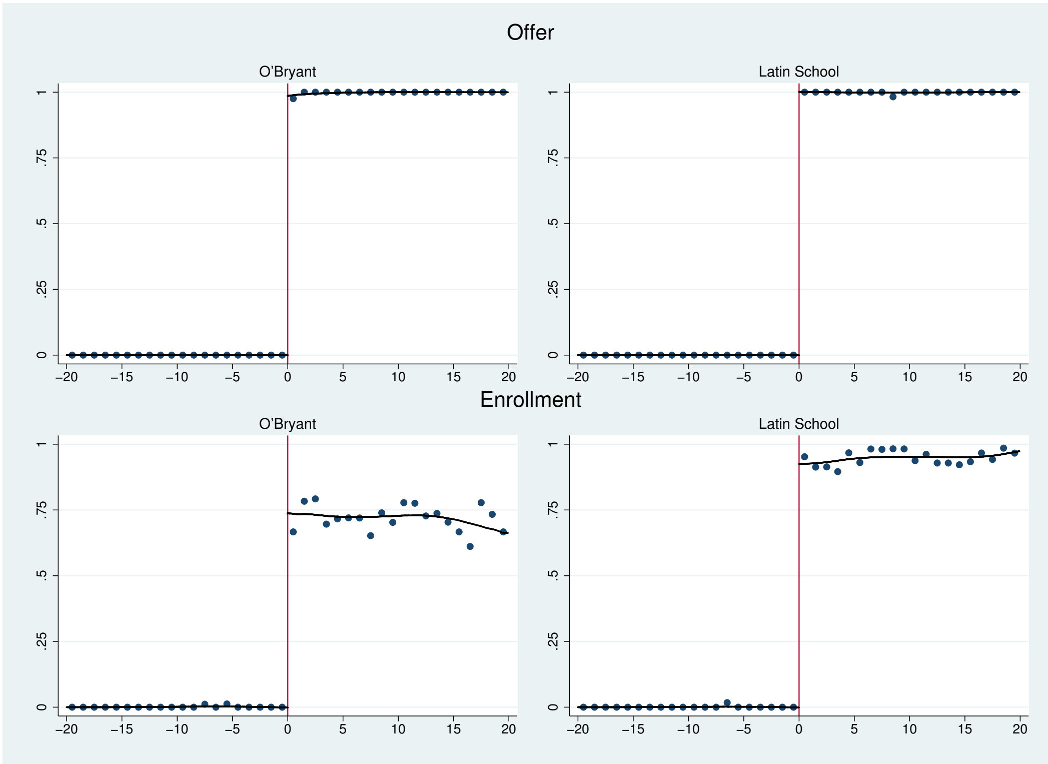
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Fuzzy CIA Estimates of Average Causal Response (Years of Exam School Enrollment) for 9th Grade Applicants

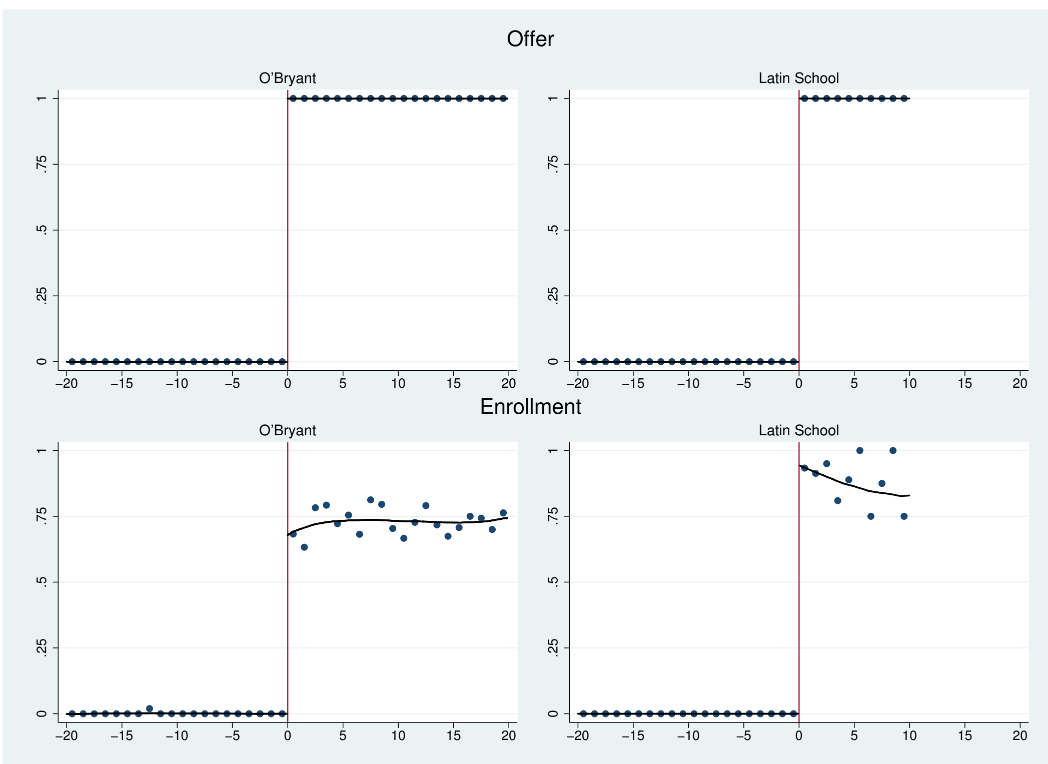
	First Stage				ACR			
	Math		ELA		Math		ELA	
	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School	O'Bryant	Latin School
Window	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
20	1.394***	1.816***	1.398***	1.820***	0.112***	-0.017	0.142***	0.048
	(0.064)	(0.096)	(0.065)	(0.093)	(0.029)	(0.050)	(0.030)	(0.045)
N untreated	513	320	516	320	513	320	516	320
N treated	486	49	489	50	486	49	489	50
15	1.359***	1.816***	1.363***	1.820***	0.095***	-0.044	0.133***	0.028
	(0.064)	(0.099)	(0.064)	(0.089)	(0.032)	(0.031)	(0.034)	(0.047)
N untreated	375	228	376	229	375	228	376	229
N treated	373	49	374	50	373	49	374	50
10	1.320***	1.816***	1.312***	1.820***	0.069	-0.036	0.145***	-0.000
	(0.080)	(0.095)	(0.080)	(0.089)	(0.043)	(0.031)	(0.041)	(0.054)
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%

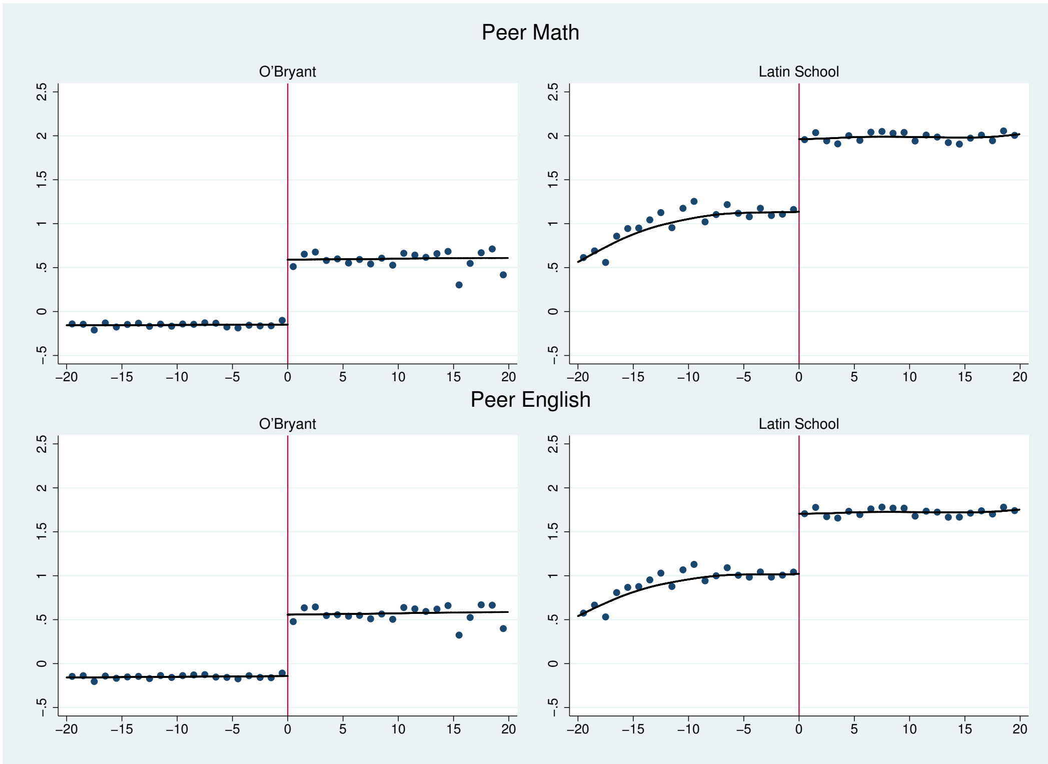


(a) 7th Grade Applicants

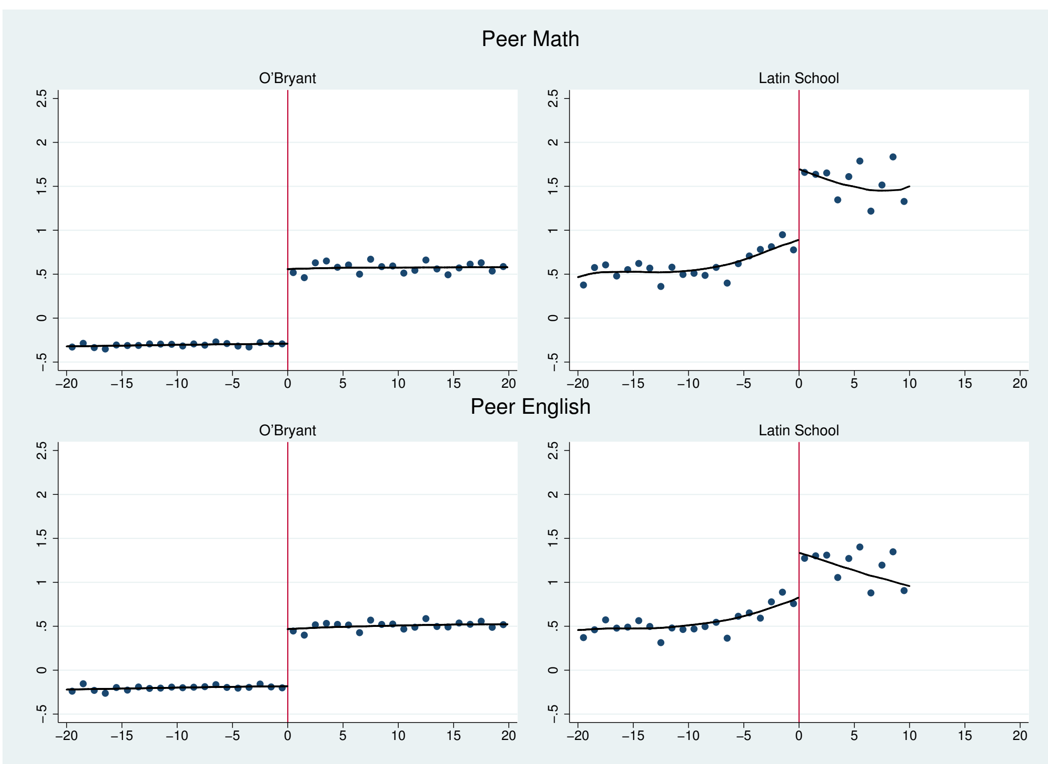


(b) 9th Grade applicants

Figure 1: Offer and Enrollment at O'Bryant and Boston Latin School

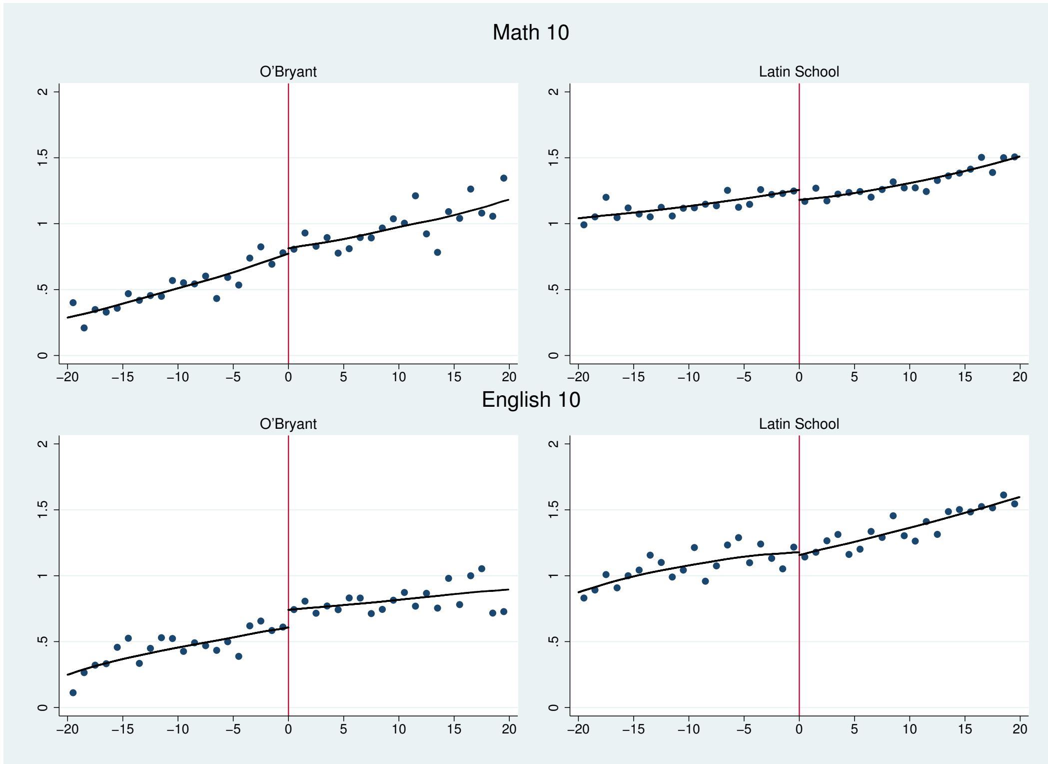


(a) 7th Grade Applicants

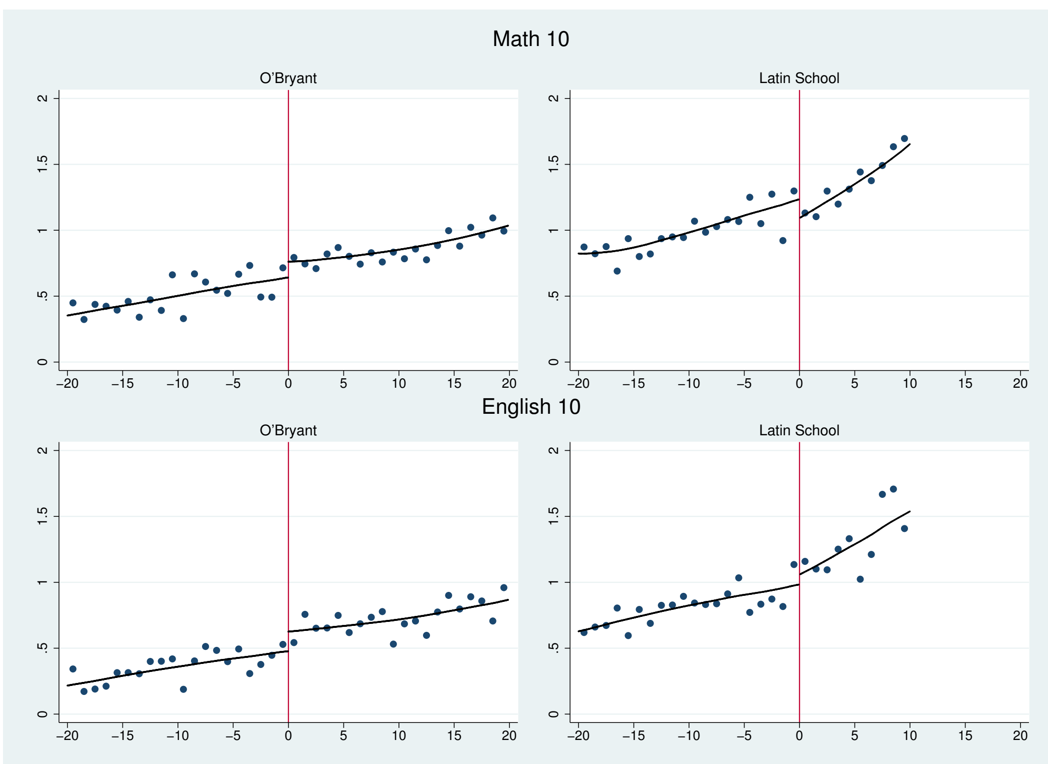


(b) 9th Grade applicants

Figure 2: Peer Achievement at O'Bryant and Boston Latin School



(a) 7th Grade Applicants



(b) 9th Grade applicants

Figure 3: 10th Grade Math and ELA Scores at O'Bryant and Boston Latin Schools

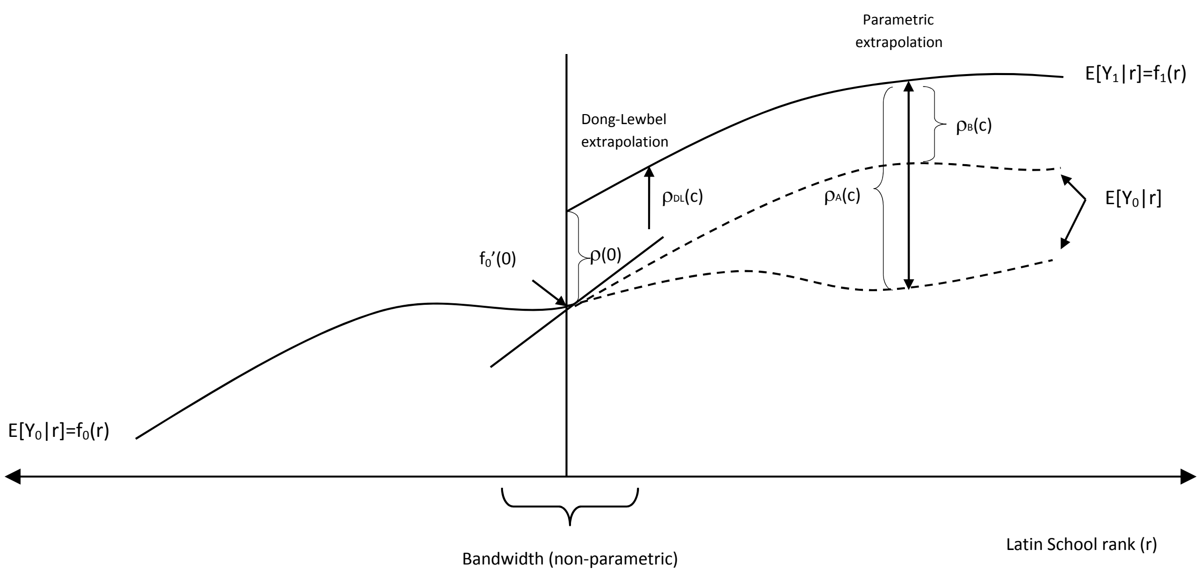
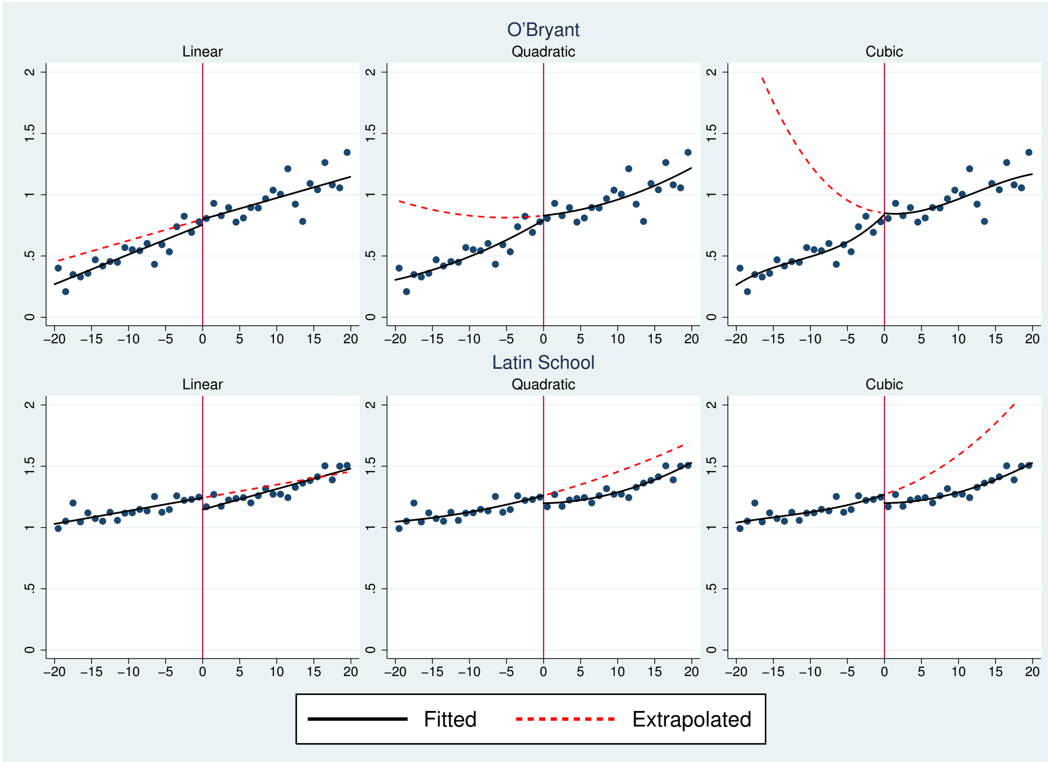
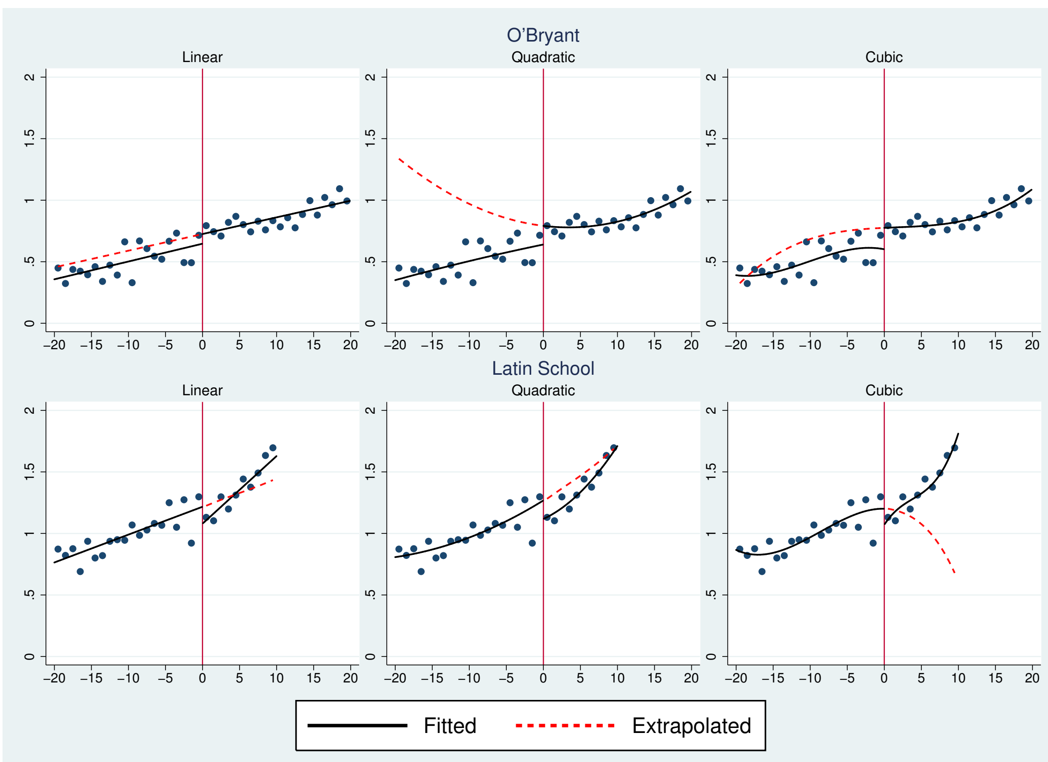


Figure 4: Identification of Boston Latin School Effects At and Away from the Cutoff. $\rho(0)$ is an effect at the cutoff; $\rho_{DL}(c)$ is an effect near the cutoff approximated using a first derivative; $\rho_A(c)$ and $\rho_B(c)$ are possible effects well away from the cutoff.

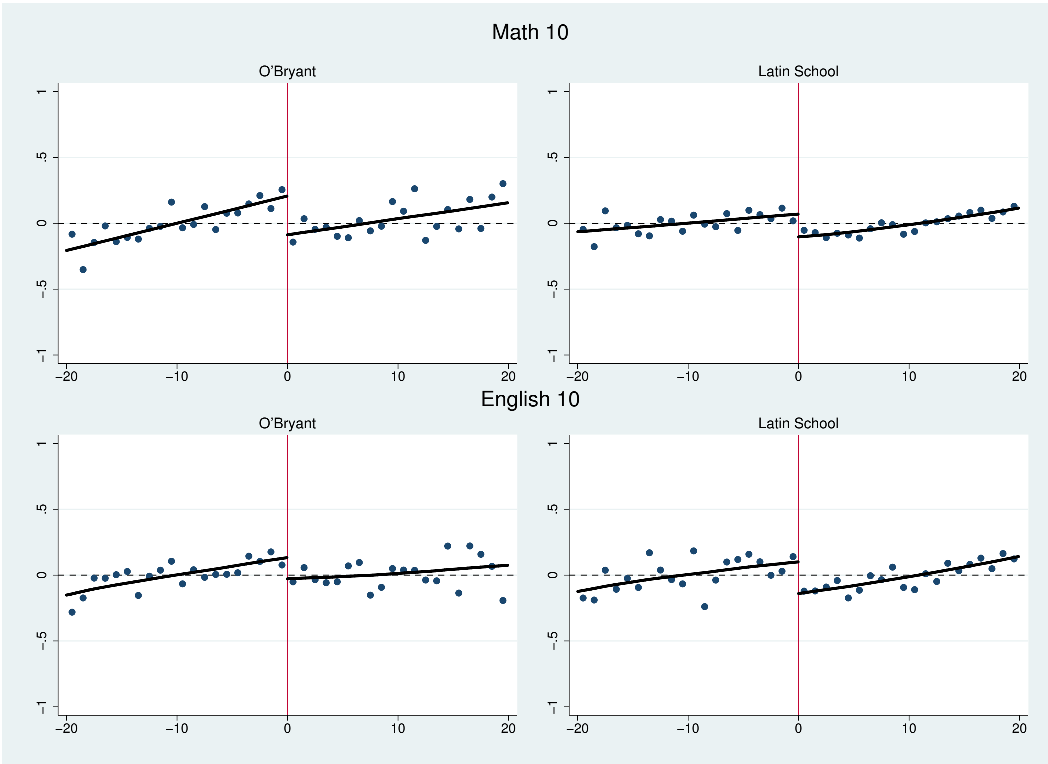


(a) 7th Grade Applicants

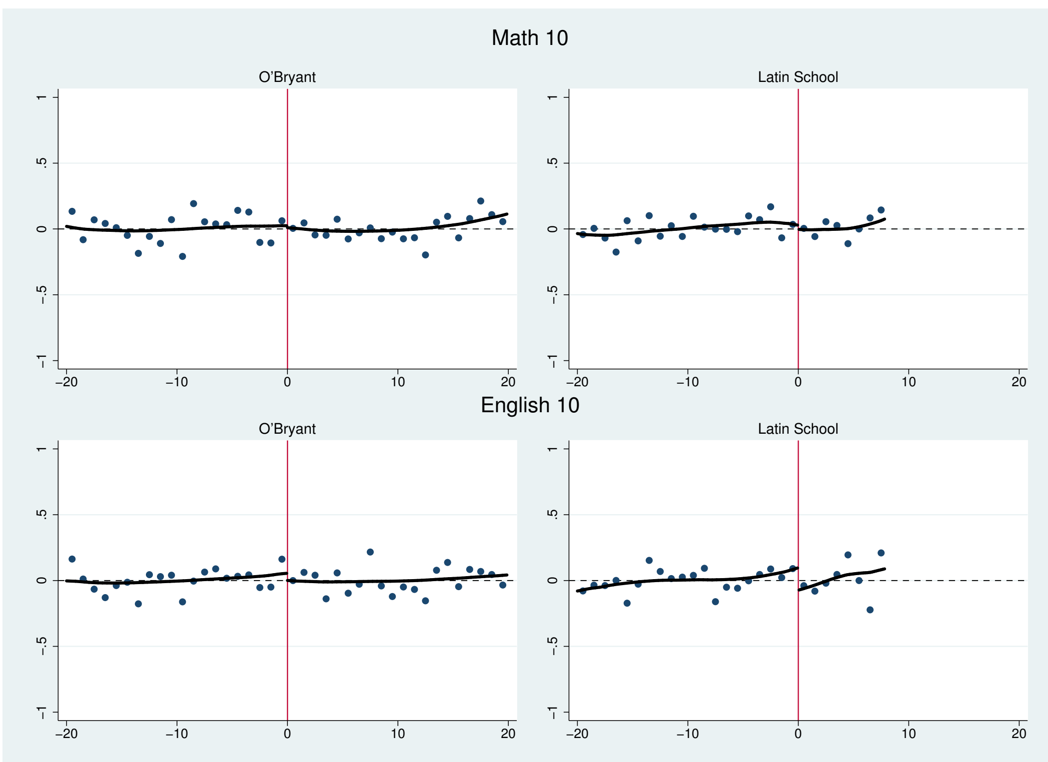


(b) 9th Grade Applicants

Figure 5: Parametric Extrapolation at O'Bryant and Boston Latin School for 10th Grade Math. O'Bryant extrapolation is for $E[Y_{1i}|r_i = c]$; BLS extrapolation is for $E[Y_{0i}|r_i = c]$.



(a) 7th Grade Applicants



(b) 9th Grade Applicants

Figure 6: Visual Evaluation of CIA in the Window $[-20, 20]$

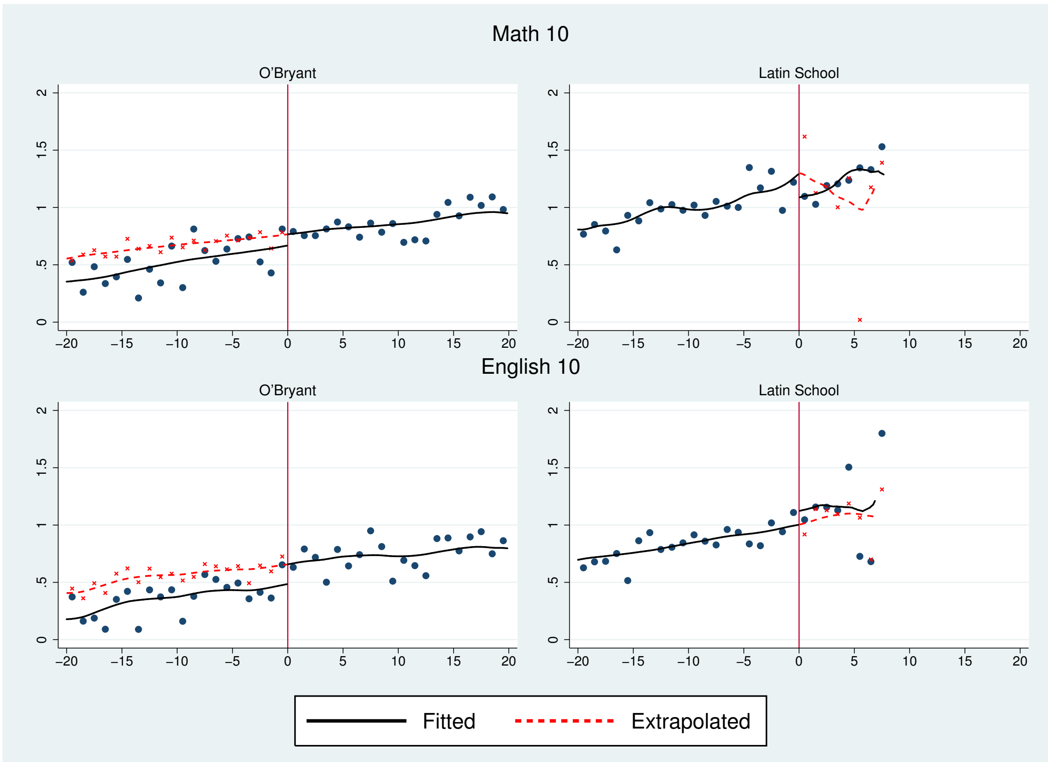


Figure 7: CIA-based Estimates of $E[Y_{1i}|r_i = c]$ and $E[Y_{0i}|r_i = c]$ for c in $[-20, 20]$ for 9th Grade Applicants. To the left of the O'Bryant cutoff, the estimates of $E[Y_{0i}|r_i = c]$ are fitted values for observed outcomes while the estimates of $E[Y_{1i}|r_i = c]$ are extrapolations. To 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.

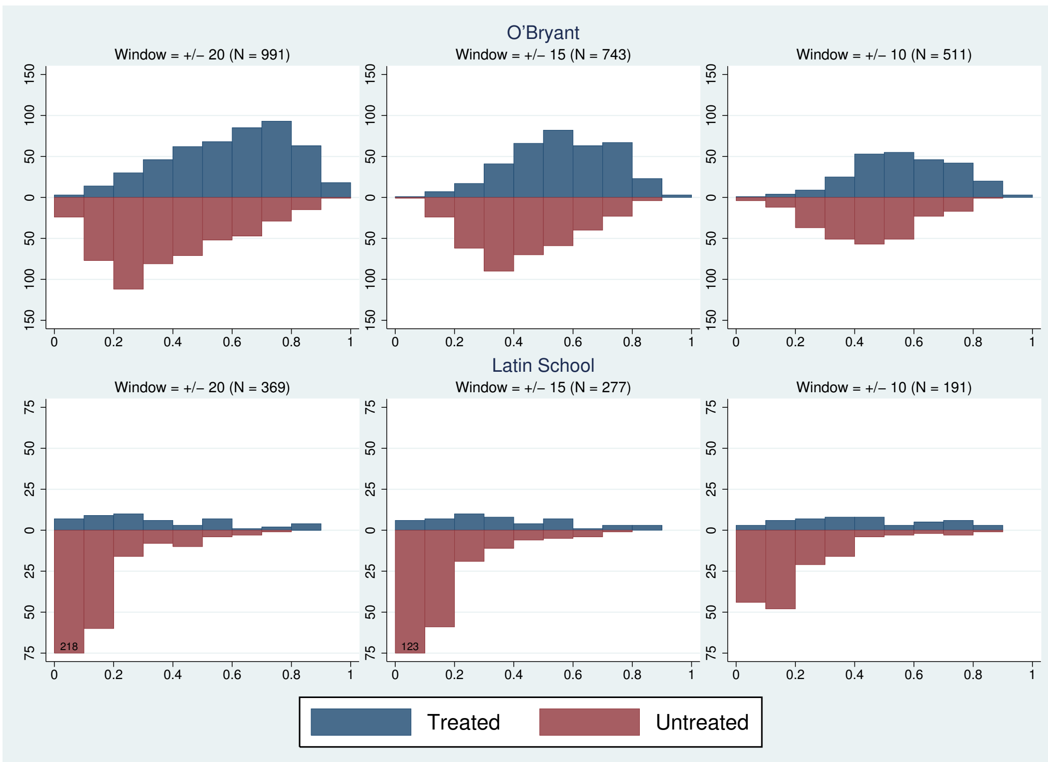


Figure 8: Histograms of Estimated Propensity Scores for 9th Grade Applicants to O'Bryant and BLS. The y-axis shows the observation count in each score-distribution decile. These estimates were constructed using a logit model fitted in the sample window indicated above each panel.

References

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

——— (2005): "Semiparametric Difference-in-Difference Estimators," *Review of Economic Studies*, 72(1), 1–19.

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, S. M. Dynarski, T. J. Kane, and P. Pathak (2011): "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *Quarterly Journal of Economics*, 126(2), 699–748.

Abdulkadiroğlu, A., J. D. Angrist, and P. Pathak (forthcoming): "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica*.

Ait-Sahalia, Y., P. J. Bickel, and T. M. Stoker (2001): "Goodness-of-Fit Tests for Kernel Regression with an Application to Option Implied Volatilities," *Journal of Econometrics*, 105(2), 363–412.

Andrews, D. W. K., and P. Guggenberger (2009): "Hybrid and Size-Corrected Subsampling Methods," *Econometrica*, 77(3), 721–762.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. D. Walters (2013): "Charter Schools and the Road to College Readiness: The Effects on College Preparation, Attendance and Choice," Research Report Prepared for the Boston Foundation and NewSchools Venture Fund.

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 G. M. Kuersteiner (2011): "Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score," *Review of Economics and Statistics*, 93(3), 725–747.

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.

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

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.

Cook, T. D., and C. Wing (2012): "Strengthening the Regression Discontinuity Design Using Additional Design Elements - A Within-Study Comparison," Unpublished manuscript.

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*chap. 5, pp. 463–536, Handbook of Labor Economics 4. 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 (2012): "Identifying the Effects of Changing the Policy Threshold in Regression Discontinuity Models," Unpublished manuscript.

Duflo, E., P. Dupas, and M. Kremer (2011): "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101(5), 1739–1774.

Fan, Y., and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64(4), 865–890.

Frandsen, B. R., M. Frölich, and B. Melly (2012): "Quantile Treatment Effects in the Regression Discontinuity Design," *Journal of Econometrics*, 168(2), 382–395.

Frolich, M. (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139(1), 35–75.

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.

Hong, H., and D. Nekipelov (2010): "Semiparametric Efficiency in Nonlinear LATE Models," *Quantitative Economics*, 1(2), 279–304.

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), 532–537.

Lalonde, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76(4), 604–620.

Lavergne, P., and Q. Vuong (2000): "Nonparametric Significance Testing," *Econometric Theory*, 16(4), 576–601.

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*, 2(3), 537–551.

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

Rokkanen, M. (2013): "Latent Factor-Based Extrapolation in the Regression Discontinuity Design," Unpublished manuscript.

Appendix

Defining Sharp Samples

Boston exam school applicants rank up to three schools in order of preference, while schools rank their applicants according to an average of GPA and ISEE scores. 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 also school 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 deferred acceptance (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 the lowest-ranked applicants in excess of capacity, with the rest provisionally admitted (students not rejected at this step may be rejected in later steps.)

- 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, rejecting the lowest-ranked applicants in excess of capacity from this combined pool and 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 be 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 - 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.

A possible concern with nonparametric identification strategies using sharp samples arises from the fact that the sharp sample itself may change discontinuously at the cutoff. Suppose, for example, that two schools have the same cutoff and a common running variable. Some students rank school 2 ahead of school 1 and some rank school 1 ahead of school 2. The sharp sample for school 1 includes both those who rank 1 first and those who rank 2 first but are disqualified there. This second group appears only to the left of the common cutoff, changing the composition of the sharp sample for school 1 (with a similar argument applying to the sharp sample for school 2). In view of this possibility, all estimating equations include dummies for applicants' preference orderings over schools.

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

$$E\left[g(y_i, W_i, x_i) \mid x_i, W_{1i} > W_{0i}\right] = \frac{1}{P\left[W_{1i} > W_{0i} \mid x_i\right]} E\left[\kappa(W_i, D_i, x_i)\, g(y_i, W_i, x_i) \mid x_i\right] \quad (29)$$

where

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

and

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

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

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

Using the GCIA, we can simplify as follows:

$$\begin{aligned}
& E\left[g\left(Y_{Wi}, x_i\right) \mid W_{1i}>W_{0i}, 0<r_i \leq c\right] \\
& =E\left\{E\left[g\left(Y_{Wi}, x_i\right) \mid x_i, W_{1i}>W_{0i}\right] \mid W_{1i}>W_{0i}, 0<r_i \leq c\right\} \\
& =\int \frac{1}{P\left[W_{1i}>W_{0i} \mid x_i\right]} E\left[\kappa_W\left(W_i, D_i, x_i\right) g\left(y_i, x_i\right) \mid X\right] d P\left[x_i \mid W_{1i}>W_{0i}, 0<r_i \leq c\right] \\
& =\frac{1}{P\left[W_{1i}>W_{0i} \mid 0<r_i \leq c\right]} \int E\left[\kappa_W\left(W_i, D_i, x_i\right) g\left(y, x_i\right) \mid x_i\right] \frac{P\left[0<r_i \leq c \mid x_i\right]}{P\left[0<r_i \leq c\right]} d P\left[x_i\right] \qquad (30) \\
& =\frac{1}{P\left[W_{1i}>W_{0i} \mid 0<r_i \leq c\right]} E\left[\kappa_W\left(W_i, D_i, x_i\right) \frac{P\left[0<r_i \leq c \mid x_i\right]}{P\left[0<r_i \leq c\right]} g\left(y_i, x_i\right)\right].
\end{aligned}$$

This implies that LATE can be written:

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

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 (29) 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 (30), the GCIA implies:

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

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:

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

The GCIA can similarly be shown to imply:

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

Combining these results, the ACR can be written:

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

where

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