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

Q1   Joshua D. ANGRIST and Miikka ROKKANEN

In regression discontinuity (RD) studies of 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 article 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 used to study effects of Boston exam schools for inframarginal applicants. 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. Supplementary materials for this article are available online.

KEY WORDS:

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 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 running variables are 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.

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 (2014) used parametric and nonparametric RD estimators to capture the causal effects of exam school attendance for applicants with index val-

ues 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 (Dobbie and Fryer 2014 reported similar findings for New York exam schools). But applicants who only barely manage to gain admission to, say, the highly selective Boston Latin School (BLS), might be unlikely to benefit from an advanced exam school curriculum. Stronger applicants 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.

Motivated by the question of how exam school attendance affects achievement for inframarginal applicants, this article tackles the theoretical problem of how to capture causal effects for applicants other than those in the immediate neighborhood of admissions cutoffs. 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 is 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. We also experimented with parametric extrapolation. The resulting

estimates are mostly imprecise and sensitive to the polynomial used for extrapolation (see the online appendix for details).

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 2400 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 1700 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 1200 students.

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) test scores. MCAS tests are taken each spring, typically in grades 3–8 and 10. Baseline (i.e., preapplication) scores for 7th grade applicants are from 4th grade. Baseline scores for 9th grade applicants are from 4th grade and from 8th grade math and 7th grade English Language Arts (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 to 2008 or in 9th grade from 2001 to 2007. We focus on applicants enrolled in BPS at the time of application (omitting private school students) because we are interested in how an exam school education compares to regular district schools. Moreover, many private school applicants remain outside the BPS district and hence out of our sample if they fail to get an exam school offer. Applicants who apply to transfer from one exam school to another are also omitted.

2.2 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 school-specific applicant rankings. In practice, however, Boston exam school offers also take account of student preferences over schools. Each student receives at most one offer that is determined by a student-proposing deferred acceptance algorithm. This process complicates our RD analysis because it loosens the direct link between running variables and offers. As in Abdulkadiroğlu, Angrist, and Pathak (2014), our econometric strategy begins by constructing analysis samples, referred to as sharp samples, which restore a deterministic link between exam school offers and running variables, so that offers are sharp around admissions cutoffs. A detailed description of the admissions process and construction of sharp samples is given in Abdulkadiroğlu, Angrist, and Pathak (2014).

The sharp RD treatment variable is an offer dummy, denoted D_{ik}, indicating applicant i is 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 rankings are centered and scaled to produce the following running variable:

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

where N_k is the total number of students who ranked school k (not the number in the sharp sample), τ_k is the admissions cutoff for school k, and c_{ik} is the ranking of student i by school k (with lower numbers constituting a better ranking). The running variables, 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, the running variables give applicants' position in the distribution of applicants to school k. Within sharp samples, we focus on a window limited to applicants with running variables no more than 20 units (percentiles) away from the cutoff. For qualified 9th grade applicants at BLS, this is nonbinding since the BLS cutoff is closer to the top of the 9th grade applicant distribution than the 0.8 quantile.
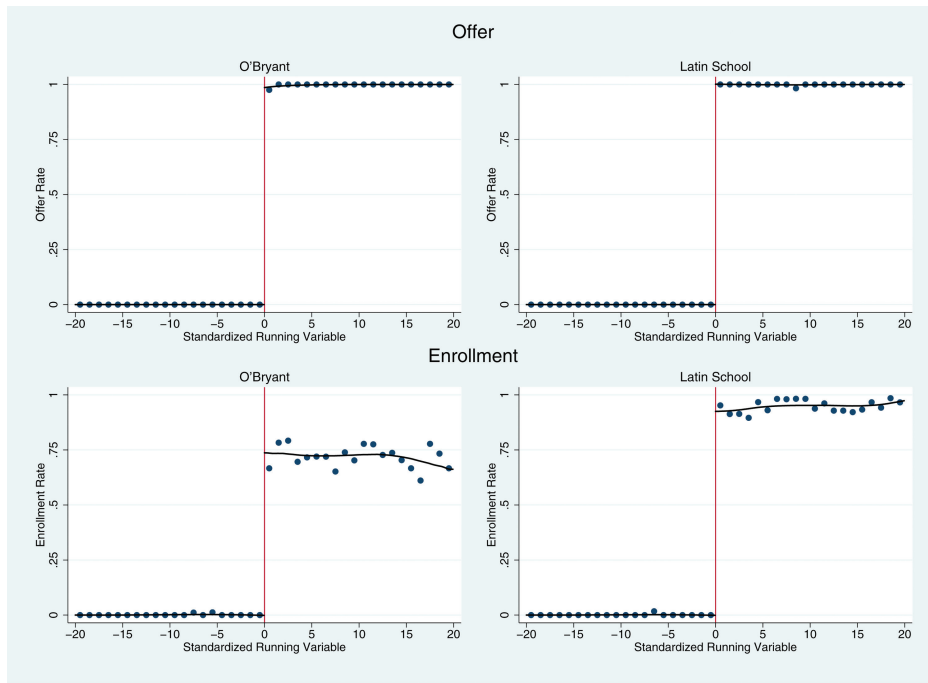
(a) 7th Grade Applicants
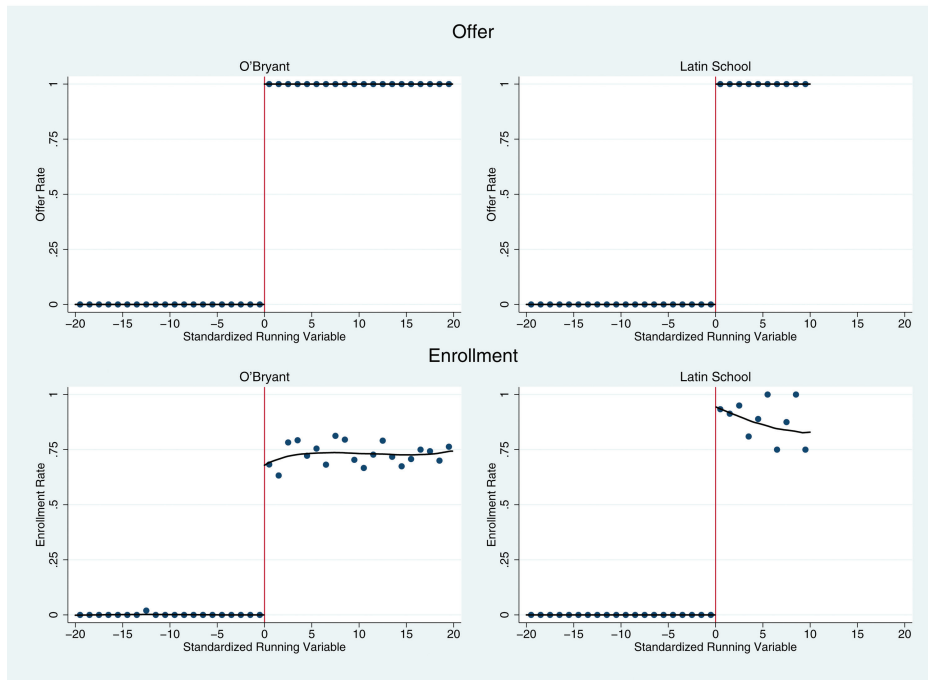
(b) 9th Grade applicants

Figure 1. Offer and enrollment at O'Bryant and Boston Latin School.

In sharp samples, offers are determined by the running variable, but not all offers are accepted. This can be seen in Figure 1(a) and 1(b), 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). Our LLR implementation uses the edge kernel and a version of the DesJardins and McCall (2008) bandwidth (hereafter DM) studied by Imbens and Kalyanaraman (2012).

2.3 Results at the Cutoff

As a benchmark, we begin with estimates for marginal applicants. Figure 2(a) and 2(b) shows 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 figures also hint 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
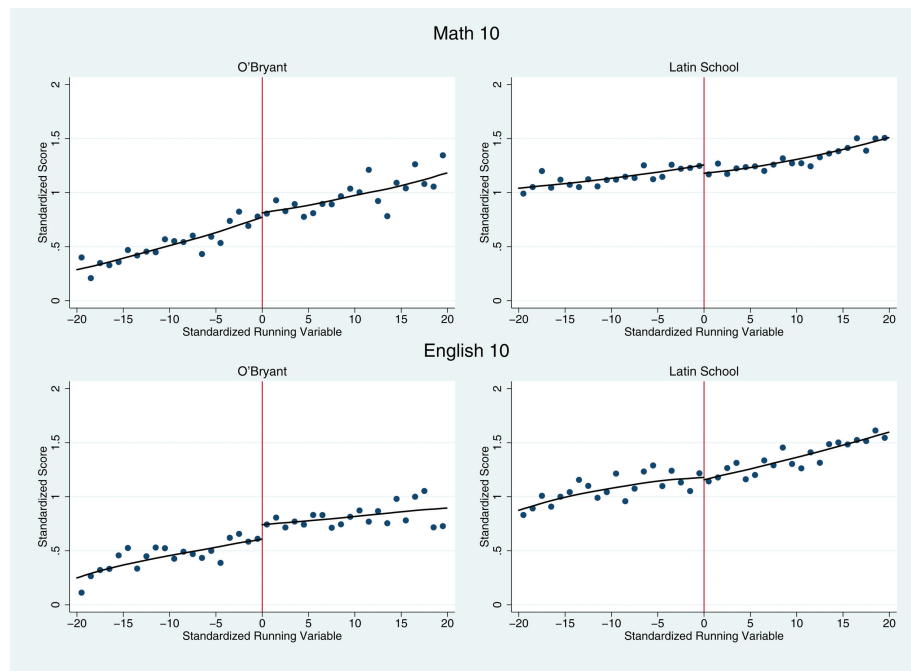
(a) 7th Grade Applicants
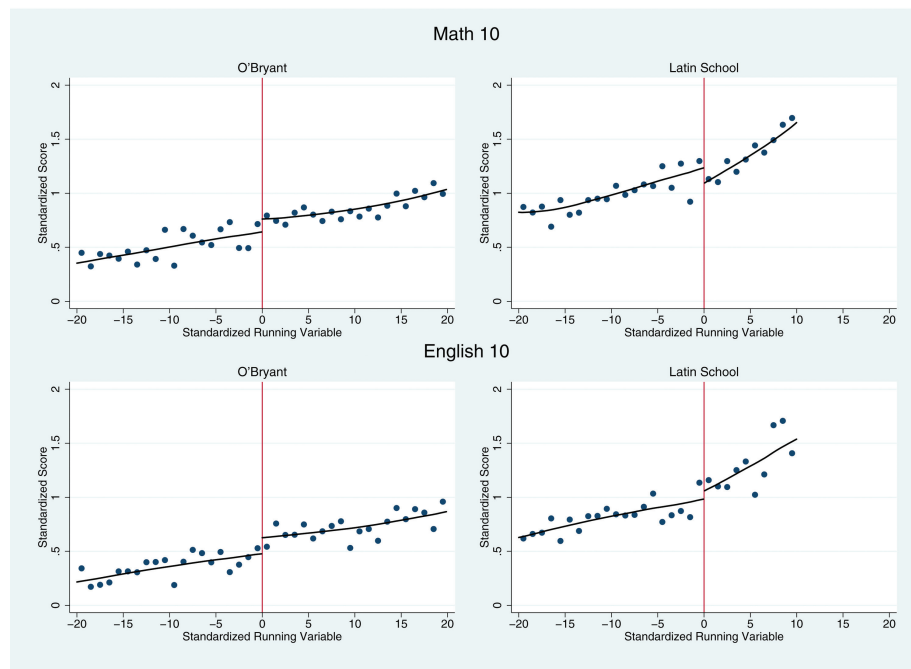
(b) 9th Grade applicants

Figure 2. 10th grade math and ELA scores at O'Bryant and Boston Latin Schools.

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 variable are modeled as

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

210 using polynomials, $f_j(r_i)$; $j = 0, 1$, that have the same intercept.

Substituting polynomials in $E[y_i \mid r_i] = (1 - D_i)$ $E[Y_{0i} \mid r_i] + D_i E[Y_{1i} \mid r_i]$, and allowing for the fact that the estimation sample pools data from different test years and application years, the equation used to construct parametric **215** estimates, fit by ordinary least squares, is

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

This model controls for test year, indexed by t and indicated by dummies w_{it}, 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} (this controls for applicant-preference-group composition effects in the sharp sample; see Abdulkadiroğlu, Angrist, and Pathak 2014 for details). The effects of the running variable are controlled by a pair of third-order polynomials that potentially differ on either side of the cutoff.

Our nonparametric RD estimators differ from parametric in three ways. First, they narrow the estimation window when the optimal data-driven bandwidth falls below 20. The nonparametric estimators also use a tent-shaped edge kernel centered at admissions cutoffs, instead of the uniform kernel implicit in parametric estimation. Finally, our nonparametric models control for linear functions of the running variable only, omitting higher-order terms. The equation used to construct nonparametric estimates, fit by weighted least squares with weights determined by the nonparametric bandwidth, is

$$
\begin{aligned}
y_i = \sum_t \alpha_t w_{it} + \sum_j \beta_j p_{ij} + \sum_\ell \delta_\ell d_{i\ell} + \gamma_0 \left(1 - D_i\right) r_i \\
+ \gamma_1 D_i r_i + \rho D_i + \eta_i.
\end{aligned}
\tag{3}
$$

Consistent with the figures, estimates of ρ in Equations (2) and (3), reported in Table 1, show little in the way of score gains at BLS. But the nonparametric 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. Additional estimates, not reported here, present a broad picture of small effects on 7th grade exam school applicants tested in 7th and 8th grade (see

Table 1. Reduced form estimates for 10th grade MCAS scores

| | Parametric | | Nonparametric | |
	O'Bryant (1)	Latin School (3)	O'Bryant (4)	Latin School (6)
	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. Nonparametric 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%.

Abdulkadiroğlu, Angrist, and Pathak 2014 for 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.

3. CALL IN 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

$$
D_i = 1 \left[r_i > 0 \right].
$$

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

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 predict achievement, we observe preapplication 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.

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)

$$
E \left[Y_{ji} \mid r_i, x_i \right] = E \left[Y_{ji} \mid x_i \right]; \; j = 0, 1.
$$

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:

Common support

$$
0 < P \left[D_i = 1 \mid x_i \right] < 1 \, \text{a.s.}
$$

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

$$
\begin{aligned}
E \left[Y_{0i} \mid D_i = 1 \right] &= E \left\{ E \left[Y_{0i} \mid x_i, D_i = 1 \right] \mid D_i = 1 \right\} \\
&= E \left\{ E \left[y_i \mid x_i, D_i = 0 \right] \mid D_i = 1 \right\},
\end{aligned}
\tag{4}
$$

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

$$
\begin{aligned}
E \left[Y_{1i} - Y_{0i} \mid D_i = 1 \right] \\
= E \left\{ E \left[y_i \mid x_i, D_i = 1 \right] - E \left[y_i \mid x_i, D_i = 0 \right] \mid D_i = 1 \right\}.
\end{aligned}
$$

We can interpret the CIA using notation suggested by Lee (2008). The running variable r_i can be modeled as a function of two parts $g \left(x_i, \epsilon_i \right)$, where x_i is observed and ϵ_i is not. Conditional on x_i the only source of variation in r_i, and consequently in D_i, is ϵ_i. Thus, the CIA requires that, conditional on the

observed x_i, potential outcomes are mean-independent of unobserved determinants of the running variable.

3.1 Testing

Just as with conventional matching strategies (as in, e.g., Heckman, Ichimura, and Todd 1998; 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, under the CIA, we have

$$E\left[Y_{1i} \mid r_i, x_i, r_i > 0\right] = E\left[Y_{1i} \mid x_i\right] = E\left[Y_{1i} \mid x_i, r_i > 0\right],$$

so we should expect that

$$E\left[y_i \mid r_i, x_i, D_i = 1\right] = E\left[y_i \mid x_i, D_i = 1\right], \qquad (5)$$

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

$$E\left[Y_{0i} \mid r_i, x_i, r_i < 0\right] = E\left[Y_{0i} \mid x_i\right] = E\left[Y_{0i} \mid x_i, r_i < 0\right],$$

so we should expect that

$$E\left[y_i \mid r_i, x_i, D_i = 0\right] = E\left[y_i \mid x_i, D_i = 0\right], \qquad (6)$$

to the left of the cutoff.

Regressions of outcomes on x_i and r_i on either side of the cutoff provide a simple test for restrictions (5) and (6). Mean independence is stronger than regression independence, of course, but regression testing procedures can embed flexible, nonlinear conditional mean functions. In practice, simple regression-based tests seem preferable to more elaborate tests that may lack the power to detect violations.

The CIA has a further testable implication: RD estimates and matching-style estimates differ only in how they weight covariate-specific treatment effects. The CIA can therefore be evaluated by comparing RD estimates to properly reweighted (using the distribution of covariates at the cutoff) matching-style estimates. See the online appendix for details and an application of this test in the Boston exam school setting.

3.2 Alternative Approaches

Battistin and Rettore (2008) considered matching estimates in an RD setting, though they focused on fuzzy RD with one-sided noncompliance and did not exploit an RD-specific conditional independence condition. In the spirit of Lalonde (1986), Battistin and Rettore validated a generic matching estimator for average treatment effect on the treated by comparing nonparametric RD estimates with conventional matching estimates constructed at the cutoff. They argued that when matching and RD produce similar results at the cutoff, matching seems worth exploring away from the cutoff as well. Mealli and Rampichini (2012) considered a similar approach that also uses information away from the cutoff.

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 nondegenerate stochastic process that assigns 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, this random running variable interpretation does not seem to offer concrete guidance as to how causal effects might change away from the cutoff. In the spirit of this notion of latent conditioning, however, we might model running variables and the conditioning variables in our CIA assumption as noisy measures of a single underlying ability measure. Rokkanen (2014) explored RD models where identification is based on this sort of latent factor ignorability in a structural econometric framework. Finally, Dong and Lewbel (2012) explored identification of the effect of changing the threshold in an RD design.

3.3 CIA-Based Estimators

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

$$E[Y_{1i} - Y_{0i} \mid r_i = c] = E[E[y_i \mid x_i, D_i = 1] \\ - E[y_i \mid x_i, D_i = 0] \mid r_i = c]. \qquad (7)$$

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:

$$E[Y_{1i} - Y_{0i} \mid 0 < r_i \leq c] = E\{E[y_i \mid x_i, D_i = 1] \\ - E[y_i \mid x_i, D_i = 0] \mid 0 < r_i \leq c\}. \qquad (8)$$

Paralleling this on the left, the bounded effect of treatment on the nontreated is

$$E[Y_{1i} - Y_{0i} \mid -c \leq r_i < 0] = E\{E[y_i \mid x_i, D_i = 1] \\ - E[y_i \mid x_i, D_i = 0] \mid -c \leq r_i < 0\}. \qquad (9)$$

We consider two estimators of Equations (7)–(9). 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 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 as

$$E\left[y_i \mid x_i, D_i = 0\right] = x_i' \beta_0 \qquad (10) \\ E\left[y_i \mid x_i, D_i = 1\right] = x_i' \beta_1.$$

Linearity is not very restrictive since the parameterization for $x_i' \beta_j$ can be rich and flexible. Substituting in Equation (7), we have

$$E\left[Y_{1i} - Y_{0i} \mid r_i = c\right] \\ = (\beta_1 - \beta_0)' E\left[x_i \mid r_i = c\right], \qquad (11)$$

with similar expressions based on Equations (8) and (9).

Let $\lambda(x_i) \equiv E[D_i \mid x_i]$ denote the propensity score. Our propensity score weighting estimator begins with the

observation that the CIA implies

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

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

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

$$E\left[Y_{1i}-Y_{0i}\mid 0<r_i\leq c\right]$$
$$=E\left\{\frac{y_i\left[D_i-\lambda\,(x_i)\right]}{\lambda\,(x_i)\left[1-\lambda\,(x_i)\right]}\times\frac{P\left[0<r_i\leq c\mid x_i\right]}{P\left[0<r_i\leq c\right]}\right\}.\ (12)$$

Similar formulas give the average effect for nontreated applicants and average effects at specific, possibly narrow, ranges of running variable values. The empirical counterpart of Equation (12) requires a model for the probability $P\left[0<r_i\leq c\mid x_i\right]$ as well as for $\lambda\,(x_i)$. It seems natural to use the same parameterization for both. Note also that if c equals the upper bound of the support of r_i, the estimand in Equation (12) simplifies to

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

as in Hirano, Imbens, and Ridder (2003).

When the estimand targets average effects at specific $r_i=c$, as opposed to over an interval, the probabilities $P\left[r_i=c\mid x_i\right]$ and $P\left[r_i=c\right]$ that appear in Equation (12) become densities. Note also that the estimand in Equation (12) can be written as $E\left[\omega_{1i}\,y_i-\omega_{0i}\,y_i\right]$, where $E\left[\omega_{0i}\right]=E\left[\omega_{1i}\right]=1$. As noted by Imbens (2004), however, this need not hold in finite samples. We therefore normalize the sum of these weights to be 1.

4. THE CIA IN ACTION AT BOSTON EXAM SCHOOLS

We start by testing CIA in estimation windows of ± 20 around the O'Bryant and BLS cutoffs. We limit our attention to these narrower windows to avoid bias from changing counterfactuals 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 school 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 see how this correlation arises, note that when estimating a BLS treatment effect, outcomes at BLS provide the relevant Y_{1i}, while those at other schools provide Y_{0i}. Just to the left of the BLS cutoff, most applicants enroll at BLA, generating Y_{0i}^{BLA}. Further to the left, however, below a cutoff, b, BLS applicants no longer qualify for BLA, and therefore end up at O'Bryant. The outcome observed for this group is Y_{0i}^{OBR}. Therefore, for those in the BLS applicant pool, we can write

$$Y_{0i}=Y_{0i}^{\text{BLA}}+\left(Y_{0i}^{\text{OBR}}-Y_{0i}^{\text{BLA}}\right)1\left[r_i<b\right].$$

Under CIA, conditioning on x_i eliminates the dependence of component potential outcomes Y_{0i}^{BLA} and Y_{0i}^{OBR} on r_i. But the switch from Y_{0i}^{BLA} to Y_{0i}^{OBR} at b remains, inducing dependence between Y_{0i} and r_i unless the distinction between BLA and O'Bryant is of no consequence. We therefore insure against

Table 2. Conditional independence tests

| | O'Bryant | | Latin School | |
	$D=0$ (1)	$D=1$ (2)	$D=0$ (3)	$D=1$ (4)
	Panel A. 7th grade applicants			
Math	0.022***	0.015***	0.008***	0.014***
	(0.004)	(0.004)	(0.002)	(0.002)
	838	618	706	748
ELA	0.015***	0.006	0.013***	0.018***
	(0.004)	(0.005)	(0.003)	(0.003)
	840	621	709	750
	Panel B. 9th grade applicants			
Math	0.002	0.005	0.008**	0.018
	(0.004)	(0.003)	(0.003)	(0.028)
	513	486	320	49
ELA	0.003	0.002	0.006	0.055
	(0.004)	(0.004)	(0.005)	(0.053)
	516	489	320	50

NOTES: This table reports regression-based tests of the conditional independence assumption described in the text. Cell entries show the coefficient on the same-subject running variable in models for 10th grade math and ELA scores that control for baseline scores, along with indicators for special education status, limited English proficiency, eligibility for free or reduced price lunch, race (black/Asian/Hispanic), and sex. Estimates use only observations to the left or right of the cutoff as indicated in column headings. Robust standard errors are reported in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

bias from changing counterfactuals by limiting extrapolation to the left of the BLS cutoff when looking at BLS applicants.

The regressions we use to test the CIA include controls 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 score controls include 7th grade ELA scores and 8th grade math scores.

CIA test results, reported in Table 2, 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 are promising. Most test statistics (i.e., 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). 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 in part because the sample for this group is small.

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, the randomizer becomes the running variable for an RD design. The relationship between outcomes and this running variable should be flat, however, except possibly for a jump at the quantile cutoff that determines proportion treated. Our CIA assumption implies this same
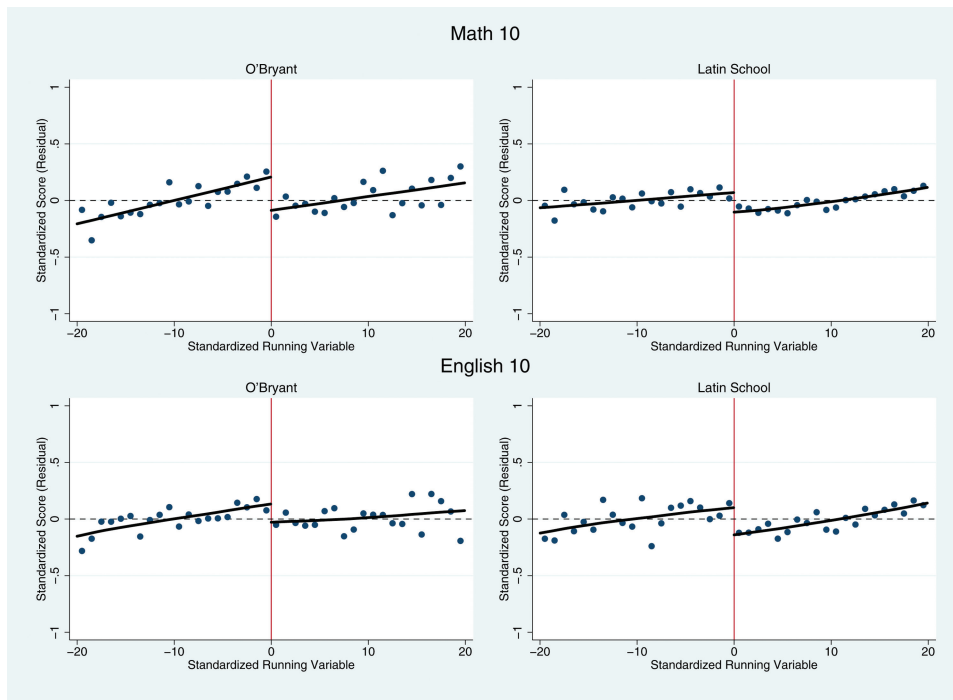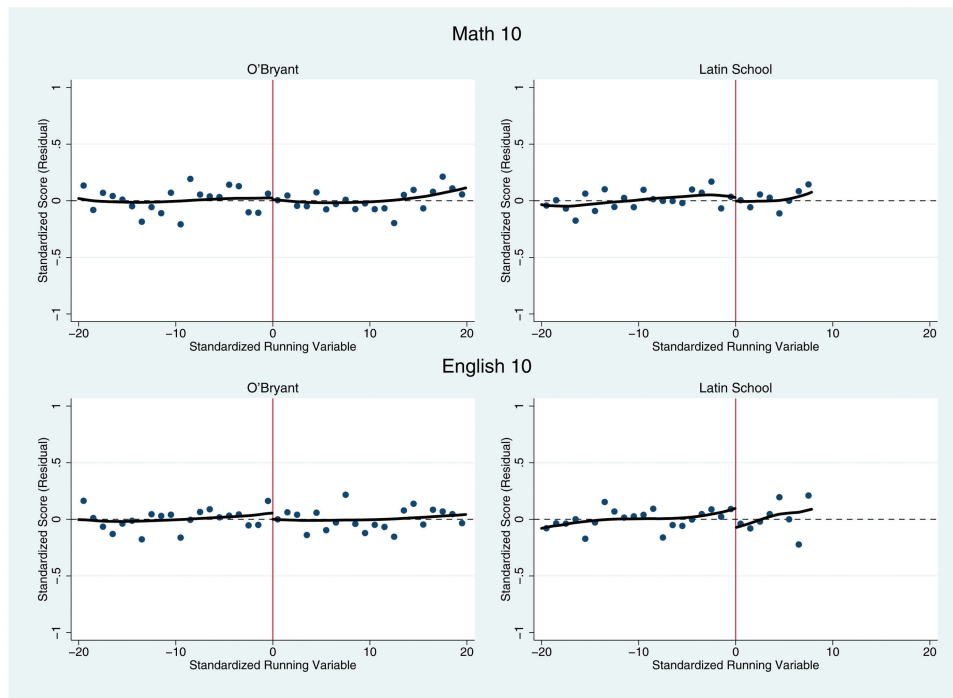
(a) 7th Grade Applicants

(b) 9th Grade Applicants

Figure 3. Visual evaluation of CIA in the window $[-20, 20]$.

pattern. Figure 3 therefore plots 10th grade math and ELA residuals (after regressing the outcomes on x_i) against running variables. 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 2, Figure 3 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.

In combination with demographic control variables and 4th grade scores, 7th and 8th grade MCAS scores appear to do a good job of eliminating the running variable from 9th graders' conditional mean functions for 10th grade scores. The difference in CIA test results for 7th and 9th grade applicants may be

Table 3. CIA estimates of the effect of exam school offers for 9th grade applicants

| | Math | | ELA | |
	O'Bryant (1)	Latin School (2)	O'Bryant (3)	Latin School (4)
Linear reweighting	0.156***	−0.031	0.198***	0.088
	(0.039)	(0.094)	(0.041)	(0.084)
N untreated	513	320	516	320
N treated	486	49	489	50
Propensity score weighting	0.148***	−0.028	0.251***	0.054
	(0.052)	(0.192)	(0.090)	(0.207)
N untreated	509	320	512	320
N treated	482	49	485	50

NOTES: This table reports CIA estimates of the effect of exam school offers on MCAS scores for 9th grade applicants to O'Bryant and BLS. The first row reports results from a linear reweighting estimator, and the second row reports 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 to the left of the admissions cutoff; the BLS estimates are effects on treated applicants 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 the relevant outcome sample. *significant at 10%; **significant at 5%; ***significant at 1%.

because baseline scores for 9th grade applicants come from tests taken in a grade closer to the outcome test grade than baseline scores for 7th grade applicants. By contrast, the most recent baseline test scores available for 7th grade applicants are from 4th grade tests. In view of the results in Table 2 and Figure 3, the CIA-based estimates that follow are for 9th grade applicants only.

The first row of Table 3 reports linear reweighting estimates of average treatment effects. These are esti-

mates of $E[Y_{1i} - Y_{0i} \mid 0 < r_i < 20]$ for BLS applicants and $E[Y_{1i} - Y_{0i} \mid -20 < r_i < 0]$ for O'Bryant applicants. Specifically, the estimand for BLS is

$$E[Y_{1i} - Y_{0i} \mid 0 < r_i \leq 20] = (\beta_1 - \beta_0)' E[x_i \mid 0 < r_i \leq 20], \quad (13)$$

while that for O'Bryant is

$$E[Y_{1i} - Y_{0i} \mid -20 \leq r_i < 0] = (\beta_1 - \beta_0)' E[x_i \mid -20 \leq r_i < 0], \quad (14)$$

where β_0 and β_1 are defined in Equation (10). The BLS estimand is an average effect of treatment on the treated, while the O'Bryant estimand is an average effect of treatment on the nontreated.

As with RD estimates at the cutoff, the CIA results in Table 3 show no evidence of a BLS achievement boost. At the same time, results for inframarginal unqualified O'Bryant applicants offer some evidence of gains. The estimates for math and ELA are 0.16σ and 0.2σ, both significantly different from zero. These CIA estimates are remarkably consistent with the corresponding RD estimates at the cutoff.

Figure 4 completes the picture of effects away from the cutoff by plotting linear reweighting estimates of $E[Y_{1i} \mid r_i = c]$ and $E[Y_{0i} \mid r_i = c]$ for all values of c in the $[-20, 20]$ interval. To the left of the O'Bryant cutoff, the estimates of $E[Y_{0i} \mid r_i = c]$ are fitted values from regression models for observed outcomes, while the estimates of $E[Y_{1i} \mid r_i = c]$ are an extrapolation based on Equation (10). To the right of the BLS cutoff, the estimates of $E[Y_{1i} \mid r_i = c]$ are fitted values while the estimates of $E[Y_{0i} \mid r_i = c]$ are an extrapolation based on Equation (10). The conditional means in this figure were constructed by plugging individual values of x_i into Equation (10) and smoothing the results using local linear regression (using the edge kernel
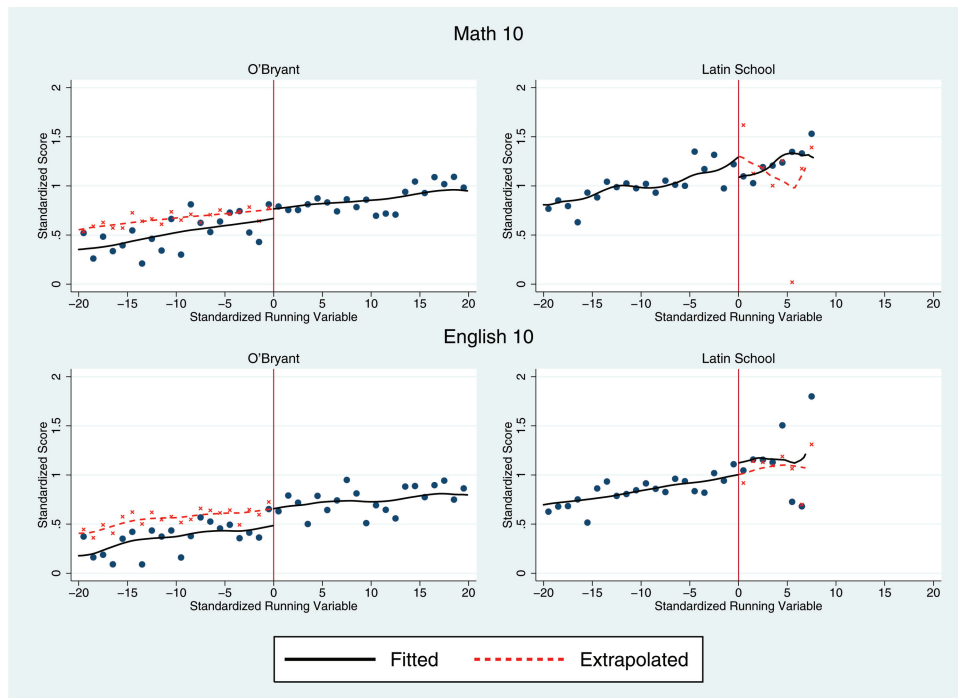

Figure 4. CIA-based estimates of $E[Y_{1i} \mid r_i = c]$ and $E[Y_{0i} \mid r_i = c]$ for c in the window $[-20, 20]$ for 9th grade applicants.
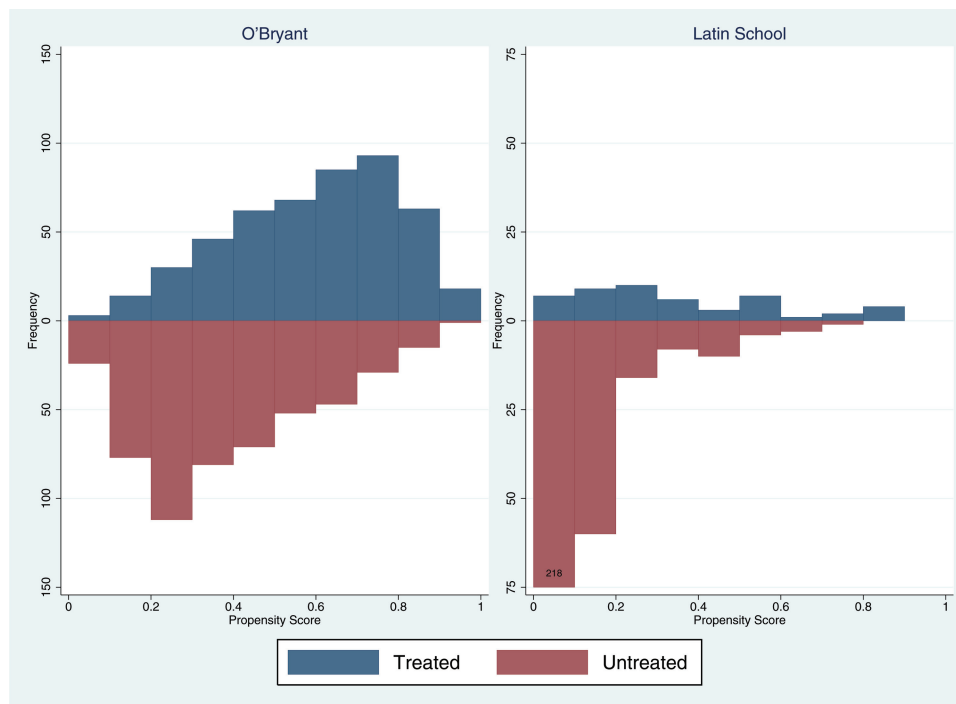
Figure 5. Histograms of estimated propensity scores in the window[−20, 20] for 9th grade applicants to O, Bryant and BLS.

505 with Stata's default bandwidth). The figure presents a picture consistent with that arising from the estimates in Table 3. In particular, the extrapolated BLS effects are small (for ELA) or noisy (for math), while the O'Bryant extrapolation reveals a re-markably stable gain in ELA scores away from the cutoff. The **510** 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

515 CIA-based estimation of the effect of exam school offers seems like a good setting for propensity score methods, since the conditioning set in this case includes multiple continuously distributed control variables. These features of the data compli-cate full covariate matching. Our logit model for the propensity score uses the control variables and parameterization used to **520** construct the tests in Table 2 and the linear reweighting esti-mates in the first row of Table 3 (the logit model for the smaller sample of BLS applicants omits test year and application pref-erence dummies).

The estimated propensity score distributions for admitted and **525** rejected applicants exhibit a substantial degree of overlap. This is documented in Figure 5, which plots the histogram of propen-sity score fitted values 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 **530** sample for highly selective BLS. Most score values for un-treated O'Bryant applicants fall below about 0.6. 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 **535** 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.

The propensity-score-weighted estimates reported in the sec-ond row of Table 3 are consistent with the linear reweighting estimates shown in the first row of the table. In particular, the **540** 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 where they of-fered a seat at O'Bryant. At the same time, the propensity score estimates for BLS applicants are highly imprecise. These BLS **545** 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 (standard errors come from a bootstrap with 500 repli-cations). Linear reweighting looks like an attractive procedure **550** in this context.

5. FUZZY CIA MODELS

Exam school offers affect achievement by facilitating exam school enrollment. The magnitude of reduced form offer com-parisons is therefore easier to interpret when the relevant enroll- **555** ment first-stage estimates scale these effects. If the first stage changes as a function of the running variable, comparisons of reduced form estimates across running variable values are mean-ingful only after rescaling. In principle, IV methods make the appropriate adjustment. A question that arises here, however, is **560** how to interpret IV estimates constructed under the CIA in a world of heterogenous 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 **565** enrollment by adapting the dummy treatment/dummy instru-ment 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, together with the common support assumption given in Section 3, is a generalized version of CIA:

Generalized conditional independence assumption (GCIA)

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

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, it is hard to see why we would have the latter without the former.

5.1 Fuzzy Identification

5.1.1 Local Average Treatment Effects. In a local average treatment effects (LATE) framework with Bernoulli treatment and Bernoulli instrument, the subset of compliers consists of people 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 shifts treatment only one way. Assuming that the instrument D_i satisfies monotonicity with $W_{1i} \geq W_{0i}$, and that for some i the inequality is strict, so there is a first stage, the LATE theorem (Imbens and Angrist 1994) tells us that

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

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) generalized 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

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

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

Theorem 1 (Fuzzy CIA Effects)

$$E[Y_{1i} - Y_{0i} \mid W_{1i} > W_{0i}, 0 < r_i \leq c]$$

$$= \frac{1}{P[W_{1i} > W_{0i} \mid 0 < r_i \leq c]}$$

$$\times E\left\{ \psi(D_i, x_i) \frac{P[0 < r_i \leq c \mid x_i]}{P[0 < r_i \leq c]} y_i \right\}, \quad (15)$$

where $\quad \psi(D_i, x_i) \equiv \dfrac{D_i - \lambda(x_i)}{\lambda(x_i)[1 - \lambda(x_i)]}.$

Estimators based on Equation (15) capture causal effects for compliers with running variable values falling into any range over which there is common support.

At first blush, it is not immediately clear how to estimate the conditional compliance probability, $P[W_{1i} > W_{0i} \mid 0 < r_i \leq c]$, appearing in the denominator of Equation (15). Because everyone to the right of the cutoff is offered treatment, 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 compliance). Paralleling an argument in Abadie (2003), however, the online appendix shows that

$$P[W_{1i} > W_{0i} \mid 0 < r_i \leq c]$$
$$= E\left\{ \kappa(W_i, D_i, x_i) \frac{P[0 < r_i \leq c \mid x_i]}{P[0 < r_i \leq c]} \right\},$$
$$(16)$$

where

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

5.1.2 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 such 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 when $w_i = j$, for $j = 0, 1, 2, \ldots, J$, are denoted by Y_{ji}. We assume also that potential treatments, w_{1i} and w_{0i}, satisfy monotonicity with $w_{1i} \geq w_{0i}$, and that these potential treatments generate a conditional first stage. In other words, we assume that

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

for the same conditioning variables that give us a valid CIA.

The Angrist and Imbens (1995) average causal response (ACR) theorem is a key building block in our analysis of ordered treatment effects identified using the GCIA. This theorem

describes the Wald IV estimand as follows:

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

where
$$v_j = \frac{P\left[w_{1i} \geq j > w_{0i}\right]}{\sum_\ell P\left[w_{1i} \geq \ell > w_{0i}\right]}$$

$$= \frac{P\left[w_i \leq j \mid D_i = 0\right] - P\left[w_i \leq j \mid D_i = 1\right]}{E\left[w_i \mid D_i = 1\right] - E\left[w_i \mid D_i = 0\right]}.$$

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 our online appendix:

Theorem 2 (Fuzzy Average Causal Response)

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

$$\tag{17}$$

where

$$v_{jc} = \frac{P\left[w_{1i} \geq j > w_{0i} \mid 0 < r_i \leq c\right]}{\sum_\ell P\left[w_{1i} \geq \ell > w_{0i} \mid 0 < r_i \leq c\right]}. \tag{18}$$

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, $E[Y_{ji} - Y_{j-1,i} \mid w_{1i} \geq j > w_{0i}, 0 < r_i \leq c]$, is weighted by the conditional (on $0 < r_i \leq c$) 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 Equation (17) by fitting linear models that include 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 Equations (13) and (14); divide by a similarly constructed average first stage. The same procedure can be used to estimate Equation (17) 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.

5.2 Fuzzy Estimates

The first row of Table 4 shows estimates of the effect of exam school offers on exam school enrollment, constructed separately for O'Bryant and BLS applicants using Equation (16). The propensity score model used to construct these estimates is

Table 4. Fuzzy CIA estimates of LATE (exam school enrollment) for 9th grade applicants

	Math		ELA	
	O'Bryant (1)	Latin School (2)	O'Bryant (3)	Latin School (4)
First stage	0.659***	0.898***	0.660***	0.900***
	(0.062)	(0.054)	(0.062)	(0.052)
N untreated	509	320	512	320
N treated	482	49	485	50
LATE	0.225**	−0.031	0.380**	0.060
	(0.088)	(0.217)	(0.183)	(0.231)
N untreated	509	320	512	320
N treated	482	49	485	50

NOTES: This table reports fuzzy CIA 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 to the left of the admissions cutoff; the BLS estimates are for treated applicants to the right of the cutoff. The first-stage estimates in the first row and the estimated causal effects in the second row 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 the relevant outcome sample. *significant at 10%; **significant at 5%; ***significant at 1%.

the same as that used to construct the estimates in Table 3 (Table 4 shows separate first-stage estimates for the math and ELA samples, as these differ slightly). The second row of the table reports estimates of $E[Y_{1i} - Y_{0i} \mid W_{1i} > W_{0i}, 0 < r_i \leq 20]$. The pattern here is consistent with that in Table 3, with small and statistically insignificant effects at BLS and evidence of large effects at O'Bryant. The estimates of O'Bryant effects on ELA and math scores show impressive gains of 0.38σ and 0.23σ.

The estimates 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 school offer through an enrollment dummy. Many who start in an exam school dropout, so we would like to adjust reduced form 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 (17), 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 3).

First-stage estimates for years of exam school enrollment, reported in the first row of Table 5, indicate that successful BLS applicants spent about 1.8 years in BLS between application and test date, while successful O'Bryant applicants spent about 1.4 years at O'Bryant between application and test date. The associated ACR estimates, reported in the second row of Table 5, are in line with the LATE estimates in Table 4, 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 0.05σ, with a standard error of roughly the same magnitude. This compares with an estimate of about the same size in column 4 of Table 4, but the standard error for the latter is five times larger. The precision gain here would seem to come from linearity of the estimator and not from the change in endogenous variable, par-

Table 5. Fuzzy CIA estimates of average causal response (years of exam school enrollment) for 9th grade applicants

	Math		ELA	
	O'Bryant (1)	Latin School (2)	O'Bryant (3)	Latin School (4)
First stage	1.394***	1.816***	1.398***	1.820***
	(0.064)	(0.096)	(0.065)	(0.093)
N untreated	513	320	516	320
N treated	486	49	489	50
ACR	0.112***	−0.017	0.142***	0.048
	(0.029)	(0.050)	(0.030)	(0.045)
N untreated	513	320	516	320
N treated	486	49	489	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 to the left of the admissions cutoff; the BLS estimates are for treated applicants to the right of the cutoff. The first-stage estimates in the first row and the estimated causal effects in the second row 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 the relevant outcome sample. *significant at 10%; **significant at 5%; ***significant at 1%.

alleling precision gains seen in the switch from propensity score to linear reweighting when constructing the sharp estimates in Table 3. The ELA and math estimates for O'Bryant shows gains of about 0.14σ and 0.11σ per year of exam school exposure.

6. SUMMARY AND DIRECTIONS FOR FURTHER WORK

The power of the RD framework comes in part from the fact 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. Although our CIA framework is demanding and would not be appropriate for all RD applications, the conditional independence assumption has testable implications that are easily checked. Specifically, the CIA implies that in samples limited to either treated or control observations, regressions of outcomes on the running variable and the covariates supporting CIA should show no running variable effects.

Among 9th grade applicants to O'Bryant school and BLS, the CIA 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 seem 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 (2014). 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 ELA gains also emerge from the RD estimates of exam school effects reported by Abdulkadiroğlu, Angrist, and Pathak (2014), 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 reintroducing 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.

Our CIA-based extrapolation strategy is likely to prove useful in other RD setting. We have seen our framework applied in recent and ongoing studies of merit scholarships and college enrollment (Bruce and Carruthers 2014), incumbency advantage in elections (Hainmueller, Hall, and Snyder 2014), employment protection and job security (Hijzen, Mondauto, and Scarpetta 2013), publicity requirements in public procurements (Coviello and Mariniello in press), public guarantee schemes to small and medium enterprise borrowing (de Blasio, De Mitri, Alessio D'Ignazio, and Stoppani 2014), and public school funding (Kreisman 2013). Our fuzzy extension also 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. Finally, Rokkanen (2014) developed 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 new context.

SUPPLEMENTARY MATERIALS

Q4

[Received December 2013. Revised January 2015.]

REFERENCES

Abadie, A. (2003), "Semiparametric Instrumental Variables Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263. [10,11,13]

Abadie, A., Angrist, J. D., and Imbens, G. (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117. [13]

Abdulkadiroğlu, A., Angrist, J. D., and Pathak, P. (2014), "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica*, 82, 137–196. [1,2,5,13]

810 Angrist, J. D., and Imbens, G. W. (1995), "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442. [11]

Q5 Barnow, B. S. (1972), "Conditions for the Presence or Absence of a Bias in Treatment Effect: Some Statistical Models for Head Start Evaluation,"
815 Discussion Paper, University of Wisconsin-Madison. [5]

Battistin, E., and Rettore, E. (2008), "Ineligibles and Eligible Non-Participants as a Double Comparison Group in Regression-Discontinuity Designs," *Journal of Econometrics*, 142, 715–730. [6]

Bruce, D. J., and Carruthers, C. K. (2014), "Jackpot? The Impact of Lottery
820 Scholarships on Enrollment in Tennessee," *Journal of Urban Economics*, 81, 30–44. [13]

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, 636–654. [5]

825 Coviello, D., and Mariniello, M. (in press), "Publicity Requirements in Public
Q6 Procurement: Evidence from a Regression Discontinuity Design," *Journal of Public Economics*. [13]

Q7 de Blasio, G., De Mitri, S., Alessio D'Ignazio, P. F., and Stoppani, L. (2014), "Public Guarantees to SME Borrowing. An RDD Evaluation," unpublished
830 manuscript. [13]

Dehejia, R. H., and Wahba, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. [6]

DesJardins, S., and McCall, B. (2008), "The Impact of Gates Millennium Schol-
835 ars Program on the Retention, College Finance- and Work-Related Choices,
Q8 and Future Educational Aspirations of Low-Income Minority Students," unpublished manuscript. [3,5]

DiNardo, J., and Lee, D. S. (2011), "Program Evaluation and Research De-
Q9 signs," in *Handbook of Labor Economics* (Vol. 4), Amsterdam: Elsevier, pp.
840 463–536. [6]

Dobbie, W., and Fryer, R. G. (2014), "The Impact of Attending a School with High-Achieving Peers: Evidence From the New York City Exam Schools," *American Economic Journal: Applied Economics*, 6, 58–75. [1]

Dong, Y., and Lewbel, A. (2012), "Identifying the Effects of Changing the Policy
845 Threshold in Regression Discontinuity Models," unpublished manuscript. [6]

Frandsen, B. R., Frölich, M., and Melly, B. (2012), "Quantile Treatment Effects
Q10 in the Regression Discontinuity Design," *Journal of Econometrics*, 168, 382–395. [13]

850 Goldberger, A. S. (1972a), "Selection Bias in Evaluating Treatment Effects:
Q11 Some Formal Illustrations," unpublished manuscript. [5]
Q12 ——— (1972b), "Selection Bias in Evaluating Treatment Effects: The Case of Interaction," unpublished manuscript. [5]

Hainmueller, J., Hall, A. B., and Snyder, J. M. (2014), "Assessing the External Validity of Election RD Estimates: An Investigation of the Incumbency **855** Advantage," unpublished manuscript. [13] **Q13**

Heckman, J. J., Ichimura, H., and Todd, P. (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294. [6]

Hijzen, A., Mondauto, L., and Scarpetta, S. (2013), "The Perverse Effects of **860** Job-Security Provisions on Job Security in Italy: Results from a Regression Discontinuity Design," IZA Discussion Papers 7594. [13] **Q14**

Hirano, K., Imbens, G. W., and Ridder, G. (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189. [6,7] **865**

Horvitz, D. G., and Thompson, D. J. (1952), "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685. [6]

Imbens, G. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economic Studies*, 86, 4–29. [7] **870**

Imbens, G., and Angrist, J. D. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [11]

Imbens, G., and Kalyanaraman, K. (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, **875** 933–959. [6]

Kline, P. M. (2011), "Oaxaca-Blinder as a Reweighting Estimator," *American Economic Review: Papers and Proceedings*, 101, 532–537. [6,12]

Kreisman, D. (2013), "The Effect of Increased Funding on Budget Allocations and Student Outcomes: RD and IV Estimates From Texas's Small District **880** Adjustment," unpublished manuscript. [13] **Q15**

Lalonde, R. J. (1986), "Evaluating the Econometric Evaluations of Training Programs With Experimental Data," *American Economic Review*, 76, 604–620. [6]

Lee, D. S. (2008), "Randomized Experiments from Non-Random Se- **885** lection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697. [5]

Lee, D. S., and Lemieux, T. (2010), "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355. [6,7]

Lord, F. M., and Novick, M. R. (1972), *Statistical Theories of Mental Test* **890** *Scores*, Addison-Wesley. [5] **Q16**

Mealli, F., and Rampichini, C. (2012), "Evaluating the Effects of University Grants by Using Regression Discontinuity Designs," *Journal of the Royal Statistical Association*, Series A, 175, 775–798. [6]

Rokkanen, M. (2014), "Exam Schools, Ability, and the Effects of Affirma- **895** tive Action: Latent Factor Extrapolation in the Regression Discontinuity Design," unpublished manuscript. [6,13] **Q17**