

# Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago’s Exam Schools

By ATILA ABDULKADIROĞLU, JOSHUA D. ANGRIST, YUSUKE NARITA, PARAG A. PATHAK,  
AND ROMAN A. ZARATE\*

Many school and college admission systems use centralized mechanisms to allocate seats in a manner that reflects applicant preferences and school priorities. Abdulkadiroğlu et al. (2015) show how lottery-based tie-breaking creates a stratified randomized trial, where the strata are preferences and priorities and the tie-breaker is the randomizer. In many settings, however, tie-breaking uses non-randomly assigned criteria like distance or a test score. Under non-lottery tie-breaking, applicants with the same preferences and priorities are no longer comparable.

Although non-lottery tie-breaking produces assignments that are correlated with applicants’ potential outcomes, the non-lottery scenario induces a kind of local random assignment. This opens the door to quasi-experimental regression discontinuity (RD) designs to measure school effects. This paper introduces a hybrid RD/propensity score empirical strategy that exploits the experiments embedded in serial dictatorship (SD), a mechanism widely used for college and selective K-12 school admissions. The key to our analysis is an RD-SD propensity score that controls for the local probability of school assignment. We use the RD-SD propensity score to estimate achievement effects of Chicago’s exam schools.

\* Abdulkadiroğlu: Duke and NBER. email: aa88@duke.edu Angrist: MIT and NBER. email: angrist@mit.edu Narita: Yale. email: yusuke.narita@yale.edu. Pathak: MIT and NBER, email: ppathak@mit.edu. Zarate: MIT, email: razarate@mit.edu We gratefully acknowledge financial support from the National Science Foundation, the William T. Grant Foundation, and the Spencer Foundation. Thanks go to Ignacio Rodriguez for outstanding research assistance and to SEII research manager Eryn Heying, who ably centralizes our many assignments. Thanks also go to our discussant Eduardo Azevedo. The findings here do not represent the views of Chicago Public Schools.

## I. Characterizing Serial Dictatorship

Serial dictatorship with exam-score tie-breaking assigns applicants one at a time in the order of their exam scores to their most preferred schools with available seats. We assume (without loss of generality) that SD processes applicants in ascending order of exam scores, referred to here as the running variable and denoted by  $R_i$  for applicant  $i$ . SD assignments are characterized by a set of admissions cutoffs. Let  $\mathbf{c} = (c_1, \dots, c_S)$  denote admissions cutoffs, where  $c_s$  is the cutoff at school  $s \in \{1, \dots, S\}$ . SD assigns applicant  $i$  her most preferred school for which  $R_i < c_s$ . With a continuum of applicants and school seats, these cutoffs are known to be constant, that is, fixed in repeated draws of the tie-breaker.

As in Abdulkadiroğlu et al. (2015), our goal is to learn about school effects using offers of school seats as instrumental variables for school attendance. Applicant type or preference order (denoted by  $\theta_i$ ) is a source of omitted variables bias (OVB) in such comparisons because applicants who rank schools differently tend to have different socioeconomic characteristics and therefore different outcomes. Type conditioning eliminates this source of OVB, but is unattractive when there are many types (5,776 applicants with non-trivial risk of an offer from Chicago’s nine exam schools includes 4,580 types). Our framework exploits the fact that the OVB induced by the correlation between type and offers is controlled by conditioning on a scalar function of type, the propensity score (Rosenbaum and Rubin, 1983). This function is the conditional probability of assignment,

$$p_s(\theta) = E[D_{is} | \theta_i = \theta],$$

where  $D_{is}$  indicates the SD-generated offer

of a seat at school  $s$  to applicant  $i$ .

In general,  $p_s(\theta)$  is an unrestricted function of type, so score conditioning would appear to have little advantage over full type conditioning. But the asymptotic approximation developed in Abdulkadiroğlu et al. (2015) yields a large market score for markets with lottery tie-breaking that is determined by only two statistics. Here, we derive a large market propensity score for SD mechanisms with non-random tie-breakers. In contrast with the lottery-based score, the large market approximation for SD with running variable tie-breaking is not necessarily coarser than type. Therefore, after deriving the global RD-SD propensity score, we derive a local propensity score that equalizes assignment risk across types while also controlling bias from the running variable.

## II. The RD-SD Propensity Score

We model assignment risk as generated by draws from the running variable distribution, fixing the number of applicants and their preferences. Assume that  $R_i$  is distributed over  $[0, 1]$ , with cumulative distribution function  $F_R^i$ . Running variables  $R_i$  and  $R_j$  for applicants  $i$  and  $j$  are independent, but, in contrast with the lottery case, not necessarily identically distributed.

The large market RD-SD propensity score depends on at most two cutoffs. The first is the cutoff at  $s$ . The second, called the most informative disqualification and denoted by  $MID_{\theta s}$ , varies with type.  $MID_{\theta s}$  equals zero when  $s$  is type  $\theta$ 's first choice, but is otherwise the most forgiving (i.e. the maximum) cutoff among the schools type  $\theta$  ranks ahead of  $s$ .  $MID_{\theta s}$  captures the effect of truncation induced by qualification at schools preferred to  $s$  on assignment risk at  $s$ : students who qualify at a school they prefer to  $s$  are never offered seats at  $s$ .

By the law of iterated expectations, the probability a type  $\theta$  applicant has a running variable below any value  $r_0$  is

$$F_R(r_0|\theta) = E[F_R^i(r_0)|\theta_i = \theta],$$

where  $F_R^i(r_0)$  is  $F_R^i$  evaluated at  $r_0$ . Our

first result, implied by a more general result in Abdulkadiroğlu et al. (2017), uses  $F_R(r_0|\theta)$  to derive the RD-SD propensity score by plugging cutoffs and MID values in for  $r_0$ :

**PROPOSITION 1:** *For all  $s$  and  $\theta$  in any continuum economy, we have:*

$$p_s(\theta) = (1 - F_R(MID_{\theta s}|\theta)) \times \max \left\{ 0, \frac{F_R(c_s|\theta) - F_R(MID_{\theta s}|\theta)}{1 - F_R(MID_{\theta s}|\theta)} \right\},$$

where we set  $p_s(\theta) = 0$  when  $MID_{\theta s} = 1$ .

This proposition reflects the forces of qualification and disqualification that determine SD-generated assignment risk. Applicant  $i$  of type  $\theta$  is assigned a school she prefers to  $s$  when  $r_i < MID_{\theta s}$ . Therefore, fraction  $1 - F_R(MID_{\theta s}|\theta)$  of type  $\theta$  applicants are considered for  $s$ . The second line is the probability of being assigned  $s$ , an event that occurs if and only if  $MID_{\theta s} < r_i \leq c_s$ , conditional on not being assigned a more preferred choice. Finally, applicants for whom  $MID_{\theta s} > c_s$  are never seated at  $s$  because those who fail to clear  $MID_{\theta s}$  are surely disqualified at  $s$  as well. This result generalizes Corollary 1 of Abdulkadiroğlu et al. (2015), which gives the large market score for SD with lottery tie-breaking, to cover arbitrary distributions of  $R_i$ . When  $F_R(\cdot|\theta)$  is standard uniform, Proposition 1 implies the corollary.

Control for the RD-SD propensity score in Proposition 1 eliminates OVB due to the association between type and potential outcomes. But Proposition 1 raises three empirical challenges not encountered under lottery tie-breaking. First, because  $F_R(\cdot|\theta)$  depends on  $\theta$  in an unrestricted manner, the score in Proposition 1 need not have coarser support than  $\theta$ . This is in spite of the fact applicants with different values of  $\theta$  have the same  $MID_{\theta s}$ . Second, the conditional running variable distribution,  $F_R(\cdot|\theta)$ , is unknown and must be estimated for each  $\theta$ . Third, while control for the propensity score eliminates confounding from type, conditional on  $p_s(\theta)$ , assignment is still correlated with potential outcomes because  $D_{is}$  is a function of  $R_i$ .

We tackle these three problems by focusing on applicants with running variable values in a  $\delta$ -neighborhood of admissions cutoffs. Specifically, define the probability of an offer from school  $s$  for applicants in a neighborhood of  $r_0$  as

$$p_s(\theta; r_0, \delta) = E[D_{is} | \theta_i = \theta, R_i \in (r_0 - \delta, r_0 + \delta)]$$

for  $\delta > 0$ . For small enough  $\delta$ , the restriction to applicants with admissions scores in  $(r_0 - \delta, r_0 + \delta)$  eliminates OVB from the running variable, while conditioning on values of  $p_s(\theta; r_0, \delta)$  eliminates confounding from applicant preferences.

Our second theoretical result characterizes the local RD-SD propensity score as the limit of  $p_s(\theta; r_0, \delta)$  as  $\delta$  goes to 0.

**PROPOSITION 2:** *Suppose  $F_R(r_0 | \theta)$  is differentiable everywhere and that  $c_s \neq c_{s'}$  for any  $s \neq s'$ . Then for all  $s, \theta$  in a continuum economy,*

$$\lim_{\delta \rightarrow 0} p_s(\theta; r_0, \delta) = \begin{cases} 0 & \text{if } c_s < MID_{\theta s}, \\ 0.5 & \text{if } MID_{\theta s} < c_s, \end{cases}$$

for  $r_0 = MID_{\theta s}, c_s$ , and

$$\lim_{\delta \rightarrow 0} p_s(\theta; r_0, \delta) = \begin{cases} 1 & \text{if } r_0 \in (MID_{\theta s}, c_s), \\ 0 & \text{otherwise} \end{cases}$$

for  $r_0 \neq MID_{\theta s}, c_s$ .

The local propensity score in Proposition 2 is constant for applicants with non-trivial assignment risk, obviating the need to estimate  $F_R(r_0 | \theta)$ . Proposition 2 also reveals the school-level RD-style experiments embedded in SD. In particular, consider type  $\theta$  applicants to  $s$  with exam scores in non-overlapping intervals around the cutoff  $c_s$  and  $MID_{\theta s}$ . Proposition 2 says that offers to applicants in this group are approximately determined by a coin toss.

Figure 1 depicts the cutoffs for 373 applicants to King College Prep High School for whom  $MID_{\theta s}$  is the cutoff at (more selective) Brooks. The Brooks cutoff is indicated with a left vertical line; applicants with MID at King equal to the Brooks cutoff are never seated at King when they qualify at more highly ranked Brooks. The King

cutoff is indicated with the right vertical line; applicants with running variable values above this are likewise never seated at King. Applicant with values between the King and Brooks cutoffs are offered seats at King. Dots in the figure identify average offer rates as a function of the running variable. An important consequence of Proposition 2 is that marginal applicants at King include two groups: applicants with running variable values near the King cutoff, and a group well away from the King cutoff, near the Brook's cutoff instead. Exam school effects might differ for these two groups, a possibility we're exploring in ongoing work.

The fact that offers are randomized while enrollment remains a choice motivates our two-stage least squares (2SLS) estimation strategy using offer dummies to instrument enrollment. Many King offers are declined; this can be seen in the enrollment rates plotted with triangles in Figure 1. No one not offered a seat at King enrolls there, while the King first-stage, or offer take-up rate, averages around 0.35.

We're often interested in an overall school sector effect, rather than the effect of enrollment at specific schools. For example, Abdulkadiroğlu, Angrist and Pathak (2014) looks at the effects of any exam school enrollment in Boston and New York. It's therefore natural to look at any-exam-school effects in Chicago as well.

Under SD, the risk of receiving any exam school offer *somewhere* is determining by the cutoff at the least selective school an applicant ranks. Formally, let  $S_\theta$  be the set of exam schools that type  $\theta$  ranks and define the qualifying cutoff to be the most forgiving cutoff among schools in  $S_\theta$ :

$$QC_\theta = \max_{s \in S_\theta} c_s.$$

An indicator for any exam school offer can then be coded as

$$D_i = 1[r_i < QC_{\theta_i}] = \sum_s D_{is},$$

where the second equality reminds us that, because SD is a single-offer system, the any

offer dummy equals the sum of all single offer dummies.

As for Proposition 2, the any-offer propensity score is derived after first defining a local assignment probability around value  $r_0$ :

$$q_s(\theta; r_0, \delta) = E[D_i | \theta_i = \theta, R_i \in (r_0 - \delta, r_0 + \delta)].$$

Using this notation, we have:

**PROPOSITION 3:** *If  $F_R$  is differentiable everywhere and  $c_s \neq c_{s'}$  for any  $s \neq s'$ , then for all  $\theta$  in a continuum economy and for any  $r_0 \in [0, 1]$ ,*

$$\lim_{\delta \rightarrow 0} q(\theta; r_0, \delta) = \begin{cases} 0 & \text{if } r_0 > QC_\theta, \\ 0.5 & \text{if } r_0 = QC_\theta, \\ 1 & \text{if } r_0 < QC_\theta. \end{cases}$$

Proposition 3 reflects the simplified nature of the risk behind  $D_i$ : applicants with a running variable value above their qualifying cutoff are sure to get an offer somewhere, though they may do better than the school that determines qualification. The limiting score treats qualification as random for those with values near the cutoff; this local risk is again a coin toss.

### III. Empirical Strategies and Estimates

Proposition 2 and 3 provide a foundation for identification strategies that capture the causal effect of enrollment at Chicago's exam schools on achievement, as measured by 10th grade PLAN and 11th grade ACT tests. Chicago students apply for seats in 8th grade, hoping to enroll in 9th grade.

In our sample period (2011-12), Chicago Public Schools (CPS) operates nine exam schools. Applicants rank up to six schools. Exam schools prioritize applicants using a common composite index formed from an admissions test, GPA, and grade 7 standardized test scores. This composite is the running variable.

The CPS exam school assignment mechanism incorporates place-based affirmative action, in which applicant addresses are classified into one of four tiers by the socioeconomic status of the census tract in which they live. Schools divide 70% of their

seats equally between applicants from each of the four tiers, with each quarter treated as a sub-school that assigns priority to one tier. The remaining 30%, said to be merit seats, are assigned without priorities.

In practice, applicants from a given tier are almost always offered either a merit seat or one of the seats prioritizing for their tier. We can therefore analyze Chicago's assignment system as a serial dictatorship in which each school is split into five sub-schools. Applicants to school  $s$  are then treated as if they apply to both the sub-school containing merit seats and the sub-school containing seats reserved for their tier.<sup>1</sup> Our notation for empirical models below ignores tiers; empirically, each school indexed by  $s$  is a school-tier combination.

We use Propositions 2 and 3 to classify applicants by risk for school-specific and any-school offers, and to find students in the neighborhood of each school's cutoff. The realized CPS allocation for school year 2011-12 is used to compute these cutoffs. These in turn determine  $MID_{\theta_{is}}$ .

Individual school offer dummies,  $D_{is}$ , indicate  $r_i \in [MID_{\theta_{is}}, c_s]$ . Given a cutoff-specific bandwidth  $\delta_s$ , the estimated local RD-SD propensity score for each school-specific offer,  $\hat{p}_{is}$ , is computed as follows:

$$\hat{p}_{is} = \begin{cases} 0.5 & \text{if } MID_{\theta_{is}} < c_s \text{ and} \\ & r_i \in (c_s - \delta_s, c_s + \delta_s) \text{ or} \\ & r_i \in (MID_{\theta_{is}} - \delta_s, MID_{\theta_{is}} + \delta_s), \\ 1 & \text{if } MID_{\theta_{is}} < c_s \text{ and} \\ & r_i \in (MID_{\theta_{is}} + \delta_s, c_s - \delta_s), \\ 0 & \text{if } MID_{\theta_{is}} > c_s \text{ or} \\ & r_i \notin (MID_{\theta_{is}} - \delta_s, c_s + \delta_s). \end{cases}$$

Similarly, the any offer score for any offer,

<sup>1</sup>Specifically, each school is split into five sub-schools as follows: 30% merit, and four equal-sized tier schools, each with size 17.5% of seats. An applicant from a given tier ranks the merit school, followed by the school corresponding to their tier. This procedure matches 99.7% of CPS assignments. Dur, Pathak and Sönmez (2016) give a detailed account of the CPS assignment scheme.

$D_i$ , denoted  $\hat{q}_i$ , is computed as:

$$\hat{q}_i = \begin{cases} 0.5 & \text{if } r_i \in [QC_i - \delta_s, QC_i + \delta_s], \\ 1 & \text{if } r_i < QC_i - \delta_s, \\ 0 & \text{if } r_i > QC_i + \delta_s. \end{cases}$$

Because each school offer potentially has a different effect on any-exam enrollment, we use the offers individually to construct over-identified 2SLS estimates of the effects of any exam school enrollment, indicated by  $C_i$ . For outcome variable  $Y_i$ , the 2SLS first and second stages can be written:

$$(1) \quad \begin{aligned} C_i &= \sum_s \gamma_{1s} D_{is} \\ &+ \sum_s \sum_{p=0,0.5,1} \eta_{1ps} 1\{\hat{p}_{is} = p\} + h(r_i) + \nu_i \\ Y_i &= \gamma_2 C_i \\ &+ \sum_s \sum_{p=0,0.5,1} \eta_{2ps} 1\{\hat{p}_{is} = p\} + h(r_i) + \epsilon_i, \end{aligned}$$

where  $h(r_i)$  is a running variable control described below. The  $\eta_{1ps}$  and  $\eta_{2ps}$  terms control for the propensity score associated with each offer dummy in the first and second stages. The sample consists of applicants for whom  $\hat{p}_{is} = 0.5$  for at least one  $s$ .

Using a single any-offer instrument and the propensity controls suggested by Proposition 3 generates the following just-identified 2SLS setup:

$$(2) \quad \begin{aligned} C_i &= \gamma_1 D_i \\ &+ \sum_s \alpha_{1s} 1\{s = s_\theta^q\} + h(r_i) + \nu_i, \\ Y_i &= \gamma_2 C_i \\ &+ \sum_s \alpha_{2s} 1\{s = s_\theta^q\} + h(r_i) + \epsilon_i, \end{aligned}$$

where  $s_\theta^q$  identifies the school that determines the qualifying cutoff for type  $\theta$ . This model omits score controls because the estimation sample is limited to applicants with  $\hat{q}_i = 0.5$ . These applicants are in the bandwidth around their qualifying cutoff.

We report 2SLS estimates using two specifications of the running variable control

function,  $h(r)$ :

$$\begin{aligned} h_1(r) &= \phi_0 r + \sum_s \phi_s \max\{0, r - c_s\}, \\ h_2(r) &= \sum_{k=0}^4 \phi_k r^k. \end{aligned}$$

The first,  $h_1(r)$ , specifies a piecewise linear function of the running variable, with slope changes at each cutoff (in practice, these are school and tier specific). This control function is motivated by commonly employed RD implementations using local linear control for the running with slope changes at the cutoff. The second function,  $h_2(r)$ , specifies a quartic polynomial with common polynomial coefficients throughout the support of the running variables.

All models are estimated in a sample of applicants with running variable values in a set of cutoff-specific bandwidths. This is motivated by the limiting argument behind Propositions 2 and 3. Within these bandwidths, propensity scores are fixed at 0.5 for applicants with non-trivial risk of assignment; no further controls should therefore be necessary. Not surprisingly, however, and as in other RD applications, bandwidths are large enough to require control for running variable effects; this is accomplished here by including the control function,  $h(r)$ . We also add a set of four tier dummies to the running variable controls. These improve both precision and covariate balance for parsimonious specifications of  $h(r)$ . Finally, because bandwidths are cutoff-specific, the risk of any exam school offer varies (in our finite sample) with the identity of the qualifying cutoff. Just-identified models therefore include qualifying-cutoff fixed effects ( $\alpha_{1s}$  and  $\alpha_{2s}$ ).

Almost all applicants with  $D_i = 1$  obtain an offer; the any-offer first stage is 0.90. This value is less than one because while our offer dummy codes the offers produced by SD, the CPS assignment mechanism isn't quite SD. Also, a few applicants receive offers outside the mechanism. Consistent with the take-up rates plotted in Figure 1, the overall first stage for enrollment is about 0.38. First stages for indi-

vidual school offers range from 0.26-0.69.

2SLS estimates from both over-identified and just-identified models, and for different choices of the running variable control, each suggest exam schools have no effect on student achievement. These results can be seen in Panel A of Table 1, which reports over-identified estimates based on Proposition 2 in the first two rows for two choices of  $h(r)$ . Exam school effects on math range from  $-0.11$  to  $-0.16$ , while effects on reading are very close to zero. Just-identified estimates using a single any-offer instrument, reported in the third row of the table, are similar though less precise. For example, the ACT math standard error increases from 0.087 to 0.133 between row 1 and row 3.<sup>2</sup> These findings echo those for Boston and New York exam schools (Abdulkadiroğlu, Angrist and Pathak (2014)), and in other analyses of CPS exam schools (Barrow, Sartain and de la Torre (2016)).

The advantage of propensity score control for applicant risk can be seen in Panel B of Table 1. This panel reports estimates of models (1) and (2) that include a full set of controls for applicant type. The Panel A sample is already limited to applicants with non-trivial assignment risk, that is, with local propensity score of 0.5. Yet, full type control eliminates many of these applicants because within-type there is no treatment variation, resulting in a two-third reduction in the estimation sample size. Consequently, the results are far less conclusive. We see, for example, large positive and negative effects. None of these are significantly different from zero, since the standard errors are more than twice those in Panel A.

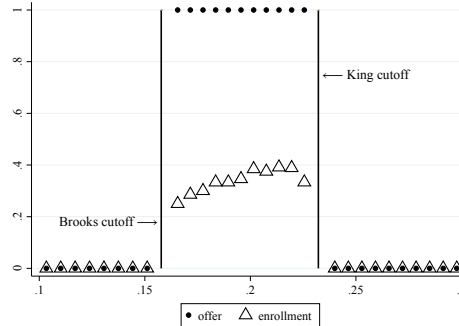
#### IV. Summary

Large urban districts increasingly use centralized assignment schemes, which generate a wealth of data that can be used to answer questions about school effectiveness. We show here that this opportunity for impact evaluation is not limited to systems

<sup>2</sup>Estimates in this table were computed using the Imbens and Kalyanaraman (2011) procedure to compute bandwidths. Specification (1) uses a uniform kernel while specification (2) uses the edge kernel.

using lotteries. The methods and models needed for efficient and informative evaluation using RD tie-breaking differ from the lottery setting. Application of the former to an analysis of Chicago exam schools yields precisely estimated effects showing little evidence of an exam school advantage.

FIGURE 1: OFFERS AND ENROLLMENT AT KING



Notes: Offers and enrollment for 373 applicants to King with MID given by the Brooks cutoff.

TABLE 1: 2SLS ESTIMATES EXAM SCHOOL EFFECTS

Instrument	PLAN Math (1)	PLAN Reading (2)	ACT Math (3)	ACT Reading (4)
A. Propensity Score Conditioning				
School-Specific Offers	-0.137 (0.096)	-0.003 (0.096)	-0.114 (0.087)	0.058 (0.093)
School-Specific Offers (Polynomial controls)	-0.163 (0.095)	-0.022 (0.094)	-0.151 (0.085)	0.036 (0.088)
N	5387	4969	5275	4624
Any Offer	-0.035 (0.131)	-0.094 (0.133)	-0.220 (0.133)	0.050 (0.137)
N	4815	4301	4639	3816
B. Full Type Conditioning				
School-Specific Offers	-0.053 (0.216)	0.108 (0.228)	-0.018 (0.202)	0.301 (0.218)
N	1747	1532	1657	1489
Any Offer	-0.109 (0.260)	0.275 (0.271)	-0.194 (0.277)	0.511 (0.295)
N	1406	1162	1269	1056

Note: This table reports 2SLS estimates of exam school enrollment effects for four outcomes. Panel A shows estimates using propensity score controls; Panel B reports 2SLS estimates controlling for preferences and tier. Models using school-specific offers are over-identified.

#### REFERENCES

- Abdulkadiroğlu, Atila, Josh Angrist, and Parag Pathak. 2014. “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools.” *Econometrica*, 82(1): 137–196.
- Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2015. “Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation.” NBER Working Paper 21705.

**Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, Parag A. Pathak, and Román A. Zárate.** 2017. “Market Design Meets Regression Discontinuity Design.” Working Paper.

**Barrow, Lisa, Lauren Sartain, and Marisa de la Torre.** 2016. “The Role of Selective High Schools in Equalizing Educational Outcomes: Heterogeneous Effects by Neighborhood Socioeconomic Status.” Federal Reserve Bank of Chicago, Working Paper 2016-17.

**Dur, Umut, Parag A. Pathak, and Tayfun Sönmez.** 2016. “Explicit vs. Statistical Targeting in Affirmative Action: Theory and Evidence from Chicago’s Exam Schools.” NBER Working Paper 22109.

**Imbens, Guido, and Karthik Kalyanaraman.** 2011. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator.” *Review of Economic Studies*, 79(3): 933–959.

**Rosenbaum, Paul R., and Donald B. Rubin.** 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, 41–55.