Breaking Ties:
Regression Discontinuity Design Meets Market Design*

Atila Abdulkadiroğlu Joshua D. Angrist Yusuke Narita Parag A. Pathak†

December 27, 2018

Abstract

Centralized school assignment algorithms employ non-lottery tie-breakers like test scores, randomly assigned lottery numbers, or both. The New York City public high school match illustrates the latter, using test scores, grades, and interviews to rank applicants to screened schools, combined with lottery tie-breaking at unscreened schools. We show how to identify causal effects of school attendance in such settings. Our approach generalizes regression discontinuity designs to allow for multiple treatments and multiple running variables, some of which are randomly assigned. Lotteries generate assignment risk at screened as well as unscreened schools. These results are used to assess the predictive value of New York City's school report cards. Grade A schools improve SAT math scores and increase the likelihood of graduating, though by less than OLS estimates suggest. Grade A attendance also boosts measures of college and career readiness. Estimation strategies that exploit the combination of lottery and non-lottery risk increase precision markedly. Grade A effects are similar when identified by screened and unscreened (lottery) tie-breakers and for screened and lottery schools. Selection bias in OLS estimates is egregious for Grade A screened schools.

*This paper is a revision of our NBER Working Paper 24172, "Impact Evaluation in Matching Markets with General Tie-Breaking." We thank Nadiya Chadha, Sonali Murarka, Lianna Wright, and the staff of the New York City Department of Education for answering our questions and facilitating access to data. Don Andrews, Tim Armstrong, Eduardo Azevedo, Yeon-Koo Che, Glenn Ellison, Brigham Frandsen, Guido Imbens, Jacob Leshno, Whitney Newey, Pedro Sant'Anna, Hal Varian and seminar participants at Columbia, Montreal, Harvard, Hebrew University, Google, the NBER Summer Institute, Columbia, the NBER Market Design Working Group, the FRB of Minneapolis, CUNY, Yale, Hitotsubashi, and Tokyo provided helpful feedback. We're especially indebted to Adrian Blattner, Ignacio Rodriguez, and Suhas Vijaykumar for expert research assistance and to MIT SEII program manager Eryn Heying for invaluable administrative support. We gratefully acknowledge funding from the Laura and John Arnold Foundation, the National Science Foundation (under awards SES-1056325 and SES-1426541), and the WT Grant Foundation. Abdulkadiroğlu and Pathak are Scientific Advisory Board members of the Institute for Innovation in Public School Choice. Angrist's daughter teaches at a Boston charter school.

†Abdulkadiroğlu: Department of Economics, Duke University and NBER, email: atila.abdulkadiroglu@duke.edu. Angrist: Department of Economics, MIT and NBER, email: angrist@mit.edu. Narita: Department of Economics and Cowles Foundation, Yale University, email: yusuke.narita@yale.edu. Pathak: Department of Economics, MIT and NBER, email: ppathak@mit.edu.

1 Introduction

Large urban school districts increasingly use sophisticated matching mechanisms to assign their seats. In addition to producing fair and transparent admissions decisions, centralized assignment schemes offer a unique resource for research and accountability: the data they generate can be used to construct causal estimates of school value-added. This research dividend arises from the *tie-breaking* embedded in centralized matching. A commonly used school matching scheme, deferred acceptance (DA), takes as input information on applicant preferences and school priorities. In settings where slots are scarce, tie-breaking variables distinguish between applicants who have the same preferences and are subject to the same priorities. Holding preferences and priorities fixed, stochastic tie-breakers become a source of quasi-experimental variation in school assignment.

Many districts break ties with a single random variable, often described as a "lottery number". Abdulkadiroğlu, Angrist, Narita and Pathak (2017b) show that lottery tie-breaking assigns students to schools as in a stratified randomized trial. That is, conditional on preferences and priorities, admission offers generated by such systems are randomly assigned and therefore independent of potential outcomes. In practice, however, preferences and priorities, which we call applicant *type*, are too finely distributed for full non-parametric conditioning to be useful. The key to a feasible DA-based research design is the *DA propensity score*, defined as the probability of school assignment conditional on preferences and priorities. In a match with lottery tie-breaking, conditioning on the scalar DA propensity score is sufficient to make assignment ignorable, that is, independent of potential outcomes. Moreover, because the DA propensity score for a market with lottery tie-breaking depends on only a few school-level *cutoffs*, the score distribution is much coarser than the distribution of types.

We turn here to the problem of crafting research designs from a broad class of assignment mechanisms in which the tie-breaking variable is non-random and potentially correlated with unobserved potential outcomes. Non-random tie-breaking, used for school assignment in Boston, Chicago, and New York City, raises important challenges for causal inference in matching markets.[1] Most importantly, seat assignment under non-random tie-breaking is no longer ignorable conditional on type. Exam schools, for instance, select students with higher test scores, and these high-scoring students can be expected to do well no matter where they go to school. In regression discontinuity (RD) parlance, the *running variable* used to distinguish between applicants of the same type is a source of omitted variables bias (OVB).

Other barriers to causal inference in this setting are raised by the fact that the propensity score in a general tie-breaking scenario depends on the unknown distribution of tie-breakers

[1]Non-lottery tie-breaking embedded in centralized assignment schemes has been used in econometric research on schools in Chile (Hastings, Neilson and Zimmerman, 2013; Zimmerman, forthcoming), Italy (Fort, Ichino and Zanella, 2016), Ghana (Ajayi, 2014), Kenya (Lucas and Mbiti, 2014), Norway (Kirkeboen, Leuven and Mogstad, 2016), Romania (Pop-Eleches and Urquiola, 2013), Trinidad and Tobago (Jackson, 2010, 2012; Beuermann, Jackson and Sierra, 2016), and the U.S. (Abdulkadiroğlu, Angrist and Pathak, 2014; Dobbie and Fryer, 2014; Barrow, Sartain and de la Torre, 2016). These studies treat different schools and tie-breakers in isolation, without exploiting centralized assignment. Other related work considers estimation methods in regression discontinuity designs with multiple assignment variables and multiple cutoffs (Papay, Willett and Murnane, 2011; Zajonc, 2012; Wong, Steiner and Cook, 2013; Cattaneo, Titiunik, Vazquez-Bare and Keele, 2016).

for each applicant type. This means that the propensity score under general tie-breaking may be no coarser than the underlying type distribution. Moreover, with an unknown distribution of tie-breakers, we cannot easily estimate the propensity score by simulation. These problems are solved here by integrating the non-parametric RD framework introduced by Hahn, Todd and Van der Klaauw (2001) with the large-market matching used to study random tie-breaking in Abdulkadiroğlu *et al.* (2017b).[2] Our results provide an easily-implemented framework for a wide variety of assignment schemes with multiple cutoffs and multiple running variables, some of which may be randomly assigned.[3]

The research value of a matching market with general tie-breaking is demonstrated through an investigation of value-added by New York City (NYC) high schools. Specifically, we exploit variation generated by the NYC high school match, which uses a DA mechanism that integrates distinct non-lottery "screened school" tie-breaking with a common lottery tie-breaker to assign 9th graders to high schools. NYC screened schools use an entrance exam or other criteria to rank applicants instead of ranking by lottery. The quasi-experimental assignment variation generated by this system is used here to answer questions about school quality.

Our results show that attendance at one of New York's "Grade A schools" boosts SAT math scores modestly and may have a small effect on high school graduation. These effects are smaller than the corresponding ordinary least squares (OLS) estimates of Grade A value-added. Grade A attendance also boosts measures of college and career readiness. The practical utility of our approach is seen in the markedly increased precision of estimates that exploit all sources of risk. We also address concerns that RD effects identified solely for applicants closed to screened school cutoffs might be idiosyncratic: estimates identified by lottery risk alone differ little from estimates that exploit screening. Finally, motivated by the ongoing debate over screened admissions policies in public schools, we contrast Grade A effects estimated separately for screened and lottery schools. These are also similar. OLS estimates showing a large Grade A screened school advantage are especially misleading.

2 School Choice Experiments

School assignment problems are defined by a set of applicants, schools, and school capacities. Applicants have preferences over schools while schools have priorities over applicants. For example, schools may prioritize applicants who live nearby or with currently enrolled siblings. Let $s = 0, 1, ..., S$ index schools, where $s = 0$ represents an outside option. The letter I denotes a set of applicants, indexed by i. I may be finite or, in our large-market model, a continuum, with applicants indexed by values in the unit interval. Seating is constrained by a capacity vector, $\mathbf{q} = (q_0, q_1, q_2, ..., q_S)$, where q_s is defined as the proportion of I that can be seated at school s. We assume $q_0 = 1$.

[2] We also build upon the "local random assignment" interpretation of nonparametric RD, discussed by Frölich (2007); Cattaneo, Frandsen and Titiunik (2015); Cattaneo, Titiunik and Vazquez-Bare (2017); Frandsen (2017) and Sekhon and Titiunik (2017). See Lee and Lemieux (2010) for a survey of RD methods.

[3] Large-market results for the special case of serial dictatorship with a single non-random tie-breaker are sketched in Abdulkadiroğlu, Angrist, Narita, Pathak, and Zarate (2017a).

Applicant i's preferences over schools constitute a partial ordering, $\succ_i$, where $a \succ_i b$ means that i prefers school a to school b. Each applicant is also granted a priority at every school. Let $\rho_{is} \in \{1, ..., K, \infty\}$ denote applicant i's priority at school s, where $\rho_{is} < \rho_{js}$ means school s prioritizes i over j. For instance, $\rho_{is} = 1$ might encode the fact that applicant i has sibling priority at school s, while $\rho_{is} = 2$ encodes neighborhood priority, and $\rho_{is} = 3$ for everyone else. We use $\rho_{is} = \infty$ to indicate that i is ineligible for school s. Many applicants share priorities at a given school, in which case $\rho_{is} = \rho_{js}$ for some $i \neq j$. The vector $\boldsymbol{\rho_i} = (\rho_{i1}, ..., \rho_{iS})$ records applicant i's priorities at each school.

Applicant type is defined as $\theta_i = (\succ_i, \boldsymbol{\rho_i})$, that is, the combination of an applicant's preferences and priorities at all schools. We say that an applicant of type θ has preferences $\succ_\theta$ and priorities ρ_θ. Θ denotes the set of possible types. A *mechanism* is a set of rules determining assignment as a function of type and a set of tie-breaking variables that schools use to discriminate between applicants of the same type.

In our framework, tie-breakers and priorities are distinct because the latter are fixed, while the former are modeled as random variables. Resampling tie-breakers makes the mechanisms of interest to us stochastic: the assignment distribution generated by any stochastic mechanism is induced by the distribution of tie-breakers. In particular, stochastic mechanisms generate a probability or "risk" of assignment for each applicant to each school. *Assignment risk* is created by repeatedly drawing tie-breakers from each applicant's tie-breaker distribution and re-running the match, fixing other market features.

Tie-breakers may be uniformly distributed lottery numbers, in which case they're distributed independently of type, or variables like entrance exam scores, that depend on type. With lottery tie-breaking, the relevant distribution is a permutation distribution under which all applicant orderings are equally likely. Tie-breakers overlap with the concept of a running variable in simple RD-style research designs. We prefer "tie-breaker" because this terminology highlights the role played by multiple running variables in centralized matching. As is typical of RD, non-lottery tie-breakers in school choice are not uniformly distributed, and may depend on applicant characteristics like race or potential outcomes, as well as on type.

To describe assignment risk more formally, consider first a market with a single continuously distributed tie-breaker common to all schools, denoted R_i for applicant i. Although R_i is not necessarily uniform, we assume that it's scaled (preserving position or rank) to be distributed over $[0, 1]$, with continuously differentiable cumulative distribution function F_R^i (an assumption we maintain throughout). These common support and smoothness assumptions notwithstanding, tie-breakers may be correlated with type, so that R_i and R_j for applicants i and j are not necessarily identically or uniformly distributed, though they're assumed to be independent of one another. We assume that sets of applicants of the form $\{i | a < r_i \leq b\}$ (where r_i is the realized value of R_i and a and b are constants) are measurable.[4]

By the law of iterated expectations, the probability type θ applicants have a running variable below any value r is $F_R(r|\theta) \equiv E[F_R^i(r)|\theta_i = \theta]$, where $F_R^i(r)$ is F_R^i evaluated at r and the

[4]A sufficient condition for this measurability assumption is that the mapping from $i \in [0, 1]$ to $r_i \in [0, 1]$ is left continuous (Aliprantis and Border, 2006). This sufficient condition is satisfied by reordering applicants by their tie-breaker realizations.

expectation is assumed to exist. To be concrete, suppose that the running variable is a test score. Suppose also that type θ_0 applicants do exceptionally well on tests and therefore have running variable values drawn from a distribution with higher mean than the score distribution for type θ_1. This implies $F_R(r|\theta_0) \neq F_R(r|\theta_1)$. By contrast, when R_i is a lottery number drawn independently from the same distribution for all applicants, $F_R(r|\theta) = F_R^i(r) = r$ for any $r \in [0,1]$ and for all i and θ. Although lottery tie-breaking is important, many real-world markets diverge from this.

2.1 OVB from Type and Tie-Breakers

Suppose we'd like to estimate the causal effect of attendance at school s on the likelihood of high school graduation. Under centralized assignment, offers of a seat at s are determined solely by type and tie-breakers. These variables are therefore the only confounding factors that might compromise causal inference. Provided we can eliminate OVB from these two sources, the offers generated by centralized assignment become powerful instrumental variables that identify causal effects of school attendance.

It's useful to begin with strategies that eliminate OVB from type. Even in a market with lottery tie-breaking, students who list schools differently are likely to have different potential outcomes (many applicants prefer a neighborhood school, for example). On the other hand, since lottery tie-breakers are independent of potential outcomes, type is the only source of OVB in this case. Full-type conditioning therefore eliminates OVB in markets with lottery tie-breaking. In practice, however, matching markets typically have many types (almost as many as applicants in some cases), rendering full-type conditioning impractical. We therefore exploit the fact that the OVB induced by correlation between type and school offers is controlled by conditioning on a scalar function of type, the propensity score.[5]

To formalize the argument for propensity score conditioning in analyses of school choice, let $D_i(s)$ indicate whether applicant i is offered a seat at school s. The propensity score for school assignment is the conditional probability of assignment to s, which can be written

$$p_s(\theta) = E[D_i(s)|\theta_i = \theta].$$

The expectation here is computed using the distribution of tie-breakers. The probability $p_s(\theta)$ quantifies the "risk" of assignment to s faced by an applicant of type θ in repeated executions of a match, drawing tie-breakers anew each time; empirical models that control for $p_s(\theta)$ are likewise said to "control for risk."

Now, let W_i be any random variable independent of lottery numbers. This includes potential outcomes as well as applicant demographic characteristics. Lottery tie-breaking implies

$$P[D_i(s) = 1|\theta_i = \theta, W_i] = E[D_i(s) = 1|\theta_i = \theta] = p_s(\theta), \tag{1}$$

where $P[D_i(s) = 1|\cdot]$ is the conditional relative frequency of assignment to s determined by all possible lottery draws for subsets of applicants. Iterating expectations over type, (1) yields

$$P[D_i(s) = 1|p_s(\theta) = p, W_i] = p. \tag{2}$$

[5] Use of propensity score conditioning to control OVB originates with Rosenbaum and Rubin (1983).

In other words, control for risk makes assignment independent of W_i, eliminating OVB. This conditional independence (CI) relation means that in school choice markets with lottery tie-breaking, empirical strategies that control for risk identify causal effects.

Equation (2) provides a valuable foundation for causal inference. With lottery tie-breaking, $p_s(\theta)$ is typically a function of a few key cutoffs. This coarseness makes score-conditioning preferable to full type conditioning. With non-lottery tie-breaking, however, control for the propensity score fails to eliminate all sources of OVB: the tie-breaker itself is an omitted variable. Moreover, it no longer need be true that $p_s(\theta)$ has support coarser than θ. And, with unknown tie-breaker distributions, $p_s(\theta)$ is hard to estimate reliably. These problems are solved here by (a) using a theoretical propensity score to isolate the set of cutoffs that generate assignment risk; (b) focusing on applicants near these cutoffs. In a limit computed by shrinking bandwidths around relevant cutoffs, applicants have constant non-degenerate risk even when tie-breakers are variables like test scores that are correlated with potential outcomes, and for which distributions are unknown and fully dependent on type.

We illustrate this fundamental result in a simple scenario with three screened schools, A, B, and C, each of which uses a common non-lottery tie-breaker, a test score, say, to select applicants. Let R_i denote the tie-breaker. The assignment mechanism in this example is *serial dictatorship* (SD), with applicants ordered by the tie-breaker.

SD, a version of DA without priorities, works like this:

> Order applicants by tie-breaker. Proceeding in order, offer each applicant his or her most preferred school with seats remaining.

Like any mechanism in the DA family (defined below), SD generates a set of *randomization cutoffs*, denoted τ_s for school s. For any school s that ends up full, cutoff τ_s is given by the tie-breaker of the last student offered a seat at s. Otherwise, $\tau_s = 1$. Finite-market cutoffs are typically random, that is, they depend on the distribution of lottery draws. In large "continuum" markets, however, cutoffs are constant, a result that motivates our use of the continuum model.[6]

Suppose applicants differ in their preferences over B and C, but all list A first and there are more applicants than seats at A (imagine this is a prestigious selective school). This market has applicants of two types, those who list B second and those who list C second. With everyone listing A first, SD assigns A to any applicant with R_i below the school-A randomization cutoff, τ_A. The propensity score for assignment to school A is therefore

$$p_A(\theta) = E[1(R_i \leq \tau_A)|\theta] = F_R(\tau_A|\theta).$$

This simple score nevertheless depends on the unknown distribution $F_R(\tau_A|\theta)$, itself a function of θ. Type is therefore a source of omitted variables bias; applicants preferring B to C might live in better neighborhoods and have higher test scores, for example. It's also clear that *any* applicant who does well on tests is more likely to be offered a seat at A. Nevertheless, Proposition 1 below shows that for applicants in a δ-neighborhood of τ_A, assignment risk converges to 0.5 as δ goes to zero, and equals 0 or 1 otherwise.

[6] Abdulkadiroğlu *et al.* (2017b) explores alternative justifications of the continuum model.

The "local risk" of qualification at A is formalized by partitioning the support of tie-breaker R_i into intervals around τ_A. Given bandwidth δ, these intervals are encoded by

$$t_{iA}(\delta) = \begin{cases} n & \text{if } R_i > \tau_A + \delta \\ a & \text{if } R_i < \tau_A - \delta \\ c & \text{if } R_i \in [\tau_A - \delta, \tau_A + \delta]. \end{cases} \tag{3}$$

To establish the conditional independence properties of local risk, let W_i be any applicant characteristic, like demographic characteristics and potential outcomes, that's *unchanged by school assignment*. This includes tie-breakers other than the one in use at school s.[7] Proposition 1 shows that for all applicant types and conditional on W_i, local risk is constant at 0.5 or degenerate:

Proposition 1. *Assume that τ_A is fixed. Let $F_R(\cdot|\theta, w) = E[F_R^i(\cdot)|\theta_i = \theta, W_i = w]$ and note that $F_R(\cdot|\theta, w)$ is differentiable at τ_A for every θ and w by virtue of continuous differentiability of $F_R^i(r)$. We also assume that $F_R'(\tau_A|\theta, w) \neq 0$. Then, for $t \in \{n, a, c\}$, all θ, and all w,*

$$\lim_{\delta \to 0} E[1(R_i \leq \tau_A)|\theta_i = \theta, t_{iA}(\delta) = t, W_i = w] = \psi_A(\theta, t),$$

where

$$\psi_A(\theta, t) = \begin{cases} 0 & \text{if } t = n \\ 1 & \text{if } t = a \\ 0.5 & \text{if } t = c. \end{cases} \tag{4}$$

Proposition 1 is a restatement of results in Frölich (2007), which shows that limiting qualification risk at a single cutoff is constant at one-half, and in an unpublished draft of Frandsen (2017), which shows something similar for an asymmetric bandwidth. These earlier results omit conditioning variables and degenerate cases; for reference, our version is proved in the appendix.

The arguments of function $\psi_A(\theta, t)$ include applicant type because risk in more complicated matches (and for applicants who list A below first in this simple example) depends on type. Our formulation of Proposition 1 highlights the fact that risk is independent of confounding variables, potential outcomes, and other tie-breakers. The latter property helps us describe risk concisely in models with multiple tie-breakers. Proposition 1 can also be rewritten to show local conditional independence given the propensity score, a result stated below as a corollary:

Corollary 1 (Local Conditional Independence)**.** *Let $D_i(A) = 1(R_i \leq \tau_A)$. Then,*

$$\lim_{\delta \to 0} P[D_i(A) = 1|\theta_i = \theta, t_{iA}(\delta) = t, W_i = w, \psi_A(\theta, t) = p] = p$$

for $p \in \{0, 0.5, 1\}$.

This follows by observing that

$$P[D_i(A) = 1|\theta_i = \theta, t_{iA}(\delta) = t, W_i = w, \psi_A(\theta, t)] = P[D_i(A) = 1|\theta_i = \theta; t_{iA}(\delta) = t, W_i = w],$$

[7] Let $W_i = W_{0i}(1 - D_i(s)) + W_{1i}D_i(s)$, where W_{0i} is the potential value of W_i revealed when $D_i(s) = 0$, and W_{1i} is the potential value revealed when $D_i(s) = 1$. Then W_i is unchanged by school assignment when $W_{0i} = W_{1i}$ for all i. Covariates unchanged by school assignment are independent of lottery tie-breakers.

and then taking the limit of the right hand side. In this simple example, to know t is to know p, but the conditional independence described in the corollary carries over to elaborate matches.

Corollary 1 formalizes the idea of "local random assignment" suggested by Cattaneo *et al.* (2015, 2017) and Sekhon and Titiunik (2017). As noted by Sekhon and Titiunik (2017), most theoretical work on nonparametric RD identification relies on continuity of conditional expectation functions for potential outcomes rather than restrictions on the assignment mechanism. Here, random assignment is a consequence of the fact that, given continuous differentiability of the running variable distribution function, the running variable density is approximately uniform in small enough neighborhoods around the cutoff.[8]

Proposition 1 is a key building block for more elaborate statements of risk. The limiting nature of this theoretical result raises the question of whether Proposition 1 and its corollary have an operational, empirical counterpart. We demonstrate the empirical conditional independence property stated in the corollary by evaluating qualification risk for a particular school in windows of various sizes around this school's cutoff (we say an applicant is empirically qualified at school s when he or she clears τ_s, without regard to school assignment).

Figure 1 describes qualification risk (rates) for applicants to one of NYC's most selective screened schools, Townsend Harris (TH). The top panel of the figure compares the probability of clearing the TH cutoff for two applicant types, those who list TH first and those who list it lower.[9] As can be seen in the left pair of bars in the top panel, applicants who list TH first tend to be high achievers and are therefore more likely than others to qualify for a seat at TH.

In a sample of applicants near the TH cutoff, qualification rates for the two types are closer. Specifically, for the sample of TH applicants with running variable values inside an Imbens and Kalyanaraman (2012) (IK) bandwidth around the cutoff, qualification rates differ by only a few points. Moreover, cutting the window width in half and then in half again leads to further convergence in qualification rates, with rates in both of these narrower groups remarkably close to 0.5. This is the convergence in assignment rates predicted by Proposition 1.

The middle and bottom panels of Figure 1 document qualification rate equalization near cutoffs for groups of TH applicants defined by baseline scores rather than by type. The leftmost pair of bars compares all TH applicants in the upper and lower quartiles of the baseline math and ELA (reading) score distributions, without regard to cutoff proximity. Not surprisingly, applicants with high baseline math scores are far more likely to qualify for a seat at TH than are applicants with low baseline math scores. The qualification gap by baseline scores narrows for applicants with tie-breaker values in an IK bandwidth, however, and again approaches 0.5 for both groups as the window width is halved and then halved again.

It's noteworthy that the IK bandwidth in this case is insufficiently narrow to equalize qualification rates across baseline score groups. In practice, most RD applications use a data-driven

[8] The smoothness conditions behind Proposition 1 allow running variable distributions to have holes and spikes as long as these are away from the cutoff. In practice, NYC screened school tie-breakers are positions defined as a function of underlying variables like exam scores. Although scaling to positions masks gaps in underlying tie-breaker distributions, this should not be a problem if it does not lead to spikes near cutoffs. We discusses running variable continuity in the application section, below.

[9] Although TH runs only one program, it has a new cutoff each year. Qualified applicants in the figure clear the cutoff for the year they apply.

bandwidth combined with local linear regression to minimize bias. Our empirical strategy likewise uses an IK bandwidth to compute locally regression-adjusted comparisons that also condition on the score. As in Robins (2000) and Okui, Small, Tan and Robins (2012), this strategy amounts to a doubly-robust propensity score and regression-based estimator. We control for theoretical propensity scores, while also regression-adjusting for running variable effects in case score control is imperfect. The covariate balance tests and robustness checks reported below suggest this approach works well.

2.2 Risk in Serial Dictatorship

Our TH example illustrates local risk. But real school matching problems involves many cutoffs and a rich variety of types. We explain real-world risk determination in two steps. First, as in Abdulkadiroğlu, Che and Yasuda (2015) and Azevedo and Leshno (2016), we employ a large-market model with a unit continuum of applicants to characterize global assignment risk. The continuum can be interpreted as the limit of a sequence that repeatedly doubles the number of applicants of each type while doubling each school's capacity. In the continuum, randomization cutoffs are fixed, that is, cutoffs are the same across repeated executions of the match with tie-breakers re-drawn each time. As in Abdulkadiroğlu *et al.* (2017b), the continuum model reveals which randomization cutoffs matter for each applicant facing risk at school s. Having identified which of these cutoffs are relevant for risk determination, we evaluate risk for applicants with running variables close to them.

This strategy is outlined first for a realistic version of SD with many schools and types. In SD, applicants seated at school s qualify there and are (necessarily) disqualified at schools they like better. The building blocks for risk at school s are therefore (a) the cutoff at s and (b) cutoffs at schools preferred to s. The latter are characterized by a quantity we call *most informative disqualification* (MID), which tells us how the tie-breaker distribution among type θ applicants to s is truncated by offers at schools θ prefers to s. Formally, let Θ_s denote the set of applicant types who list s and let

$$B_{\theta s} = \{s' \in S \mid s' \succ_\theta s\} \text{ for } \theta \in \Theta_s \tag{5}$$

denote the set of schools type θ prefers to s. For each type and school, $MID_{\theta s}$ is a function of randomization cutoffs at schools in $B_{\theta s}$, specifically:

$$MID_{\theta s} \equiv \begin{cases} 0 & \text{if } B_{\theta s} = \emptyset \\ \max\{\tau_b \mid b \in B_{\theta s}\} & \text{otherwise.} \end{cases} \tag{6}$$

$MID_{\theta s}$ is zero when school s is listed first since all who list s first compete for a seat there. The second line reflects the fact that an applicant who lists s second is seated there only when disqualified at the school they've listed first, while applicants who list s third are seated there when disqualified at their first and second choices, and so on. Moreover, anyone who fails to clear cutoff τ_b is surely disqualified at schools with lower (less forgiving) cutoffs. For example, applicants who fail to qualify at a school with a cutoff of 0.5 are disqualified at schools with cutoffs below 0.5. We can therefore quantify the truncation induced by disqualification at schools preferred to s by recording the most forgiving cutoff among them.

Type θ cannot be seated as s when $MID_{\theta s} > \tau_s$ because those qualified at s can do better (they qualify at the school that determines $MID_{\theta s}$). This scenario is sketched in the top panel of Figure 2. Assignment risk when $MID_{\theta s} \leq \tau_s$ is the probability that

$$MID_{\theta s} < R_i \leq \tau_s,$$

an event sketched in the middle panel of Figure 2. We summarize these facts in the following proposition, which is implied by a more general result for DA derived in the next section.

Proposition 2 (Global Score in Serial Dictatorship)**.** *Consider a serial dictatorship in a continuum market. For all s and $\theta \in \Theta_s$, we have:*

$$p_s(\theta) = \max\left\{0, F_R(\tau_s|\theta) - F_R(MID_{\theta s}|\theta)\right\}.$$

SD assignment risk, which is positive only when when the randomization cutoff at s exceeds $MID_{\theta s}$, is given by the size of the group with R_i between $MID_{\theta s}$ and τ_s. This is

$$F_R(\tau_s|\theta) - F_R(MID_{\theta s}).$$

With lottery tie-breaking (and a uniformly distributed lottery number), the SD risk formula simplifies to $\tau_s - MID_{\theta s}$. With non-random tie-breaking, the SD propensity score depends on the conditional distribution function, $F_R(\cdot|\theta)$, evaluated at τ_s and $MID_{\theta s}$.

Proposition 2 leaves us with three empirical challenges not encountered in markets with lottery tie-breaking. First, with non-random tie-breakers like test scores, conditional running variable distributions, $F_R(.|\theta)$, are likely to depend on θ, so the score in Proposition 2 need not have coarser support than does θ. This is in spite of the fact many applicants with different values of θ share the same $MID_{\theta s}$. Second, $F_R(.|\theta)$ is typically unknown. This precludes straightforward computation of the propensity score by repeatedly sampling from $F_R(.|\theta)$. Finally, while control for the propensity score eliminates confounding from type, assignments are a function of running variables as well as type, and non-lottery running variables are likely to be correlated with potential outcomes.

As in the simple example in the previous section, we address these challenges by evaluating risk for applicants close to cutoffs. Proposition 2 identifies the relevant cutoffs in markets with many schools and types. As before, intervals around each cutoff are encoded by relation (3), but now replacing $t_{iA}(\delta)$ with $t_{is(\delta)}$ for each school, s. We collect the set of these for all schools in the vector

$$T_i(\delta) = [t_{i1}(\delta), ..., t_{is}(\delta), ..., t_{iS}(\delta)]'.$$

The following is a consequence of Theorem 1 in the next section, which characterizes local risk for any DA match.

Proposition 3 (Local Score in Serial Dictatorship)**.** *Consider a serial dictatorship in a continuum market. Assume that cutoffs τ_s are distinct. For each $s \in S$ and $\theta \in \Theta_s$ in this match such that $MID_{\theta s} \neq 0$, suppose $MID_{\theta s} = \tau_{s'}$ for $s' \neq s$. For $T = [t_1, ..., t_s, ..., t_S]' \in \{n, a, c\}^S$, all $\delta > 0$, and all w,*

$$P[D_i(s) = 1|\theta_i = \theta, T_i(\delta) = T, W_i = w] = 0 \text{ if } \tau_{s'} > \tau_s.$$

Otherwise,

$$\lim_{\delta \to 0} P[D_i(s) = 1 | \theta_i = \theta, T_i(\delta) = T, W_i = w] = \begin{cases} 0 & \text{if } t_s = n \text{ or } t_{s'} = a \\ 1 & \text{if } t_s = a \text{ and } t_{s'} = n \\ 0.5 & \text{if } t_s = c \text{ or } t_{s'} = c \end{cases}$$

When $MID_{\theta s} = 0$, $t_{s'} = n$ and risk is determined by t_s alone.

Like Proposition 1 and its corollary, Proposition 3 establishes a key conditional independence result: limiting SD assignment risk depends only on tie-breaker proximity to the cutoff at s and to $MID_{\theta s}$; risk is otherwise unrelated to applicant characteristics.[10] Panel C in Figure 2 interprets this result. Type θ applicants with tie-breakers near either $MID_{\theta s}$ or the cutoff at s face risk of one-half. This fact is an extension of Proposition 1, applied here to the *pair* of cutoffs driving SD risk for each type. Applicants with $t_s = a$ and $t_{s'} = n$ have tie-breakers strictly between $MID_{\theta s}$ and τ_s, meaning they're disqualified at s' but qualified at s. Finally, applicants with $t_s = n$ or $t_{s'} = a$ cannot be seated at s, either because they're disqualified there or because they qualify at s'.

In the empirical (as opposed to theoretical) world, almost all applicants necessarily have tie-breaker values that are strictly above or below any particular randomization cutoff. We treat applicants with tie-breakers close to either $MID_{\theta s}$ or the cutoff at s differently because it is these applicants for whom qualification is (almost) randomly assigned.

3 The DA Score with General Tie-Breaking

SD is a version of DA without priorities. *Student-proposing DA*, which nests all school choice mechanisms in wide use, works like this:

> Each applicant proposes to his most preferred school. Each school ranks these proposals, first by priority then by tie-breaker within priority groups, *provisionally* admitting the highest-ranked applicants in this order up to its capacity. Other applicants are rejected.
>
> Each rejected applicant proposes to his next most preferred school. Each school ranks these new proposals *together with applicants admitted provisionally in the previous round*, first by priority and then by tie-breaker. From this pool, the school again provisionally admits those ranked highest up to capacity, rejecting the rest.
>
> The algorithm terminates when there are no new proposals (some applicants may remain unassigned).

With multiple tie-breakers, different schools may rank applicants differently, but the DA algorithm is otherwise unchanged. For example, NYC runs a centralized DA match for most of its high schools, a match that includes a diverse set of screened schools (Abdulkadiroğlu, Pathak and Roth, 2005, 2009). These schools rank applicant proposals using (mostly) school-specific

[10] Abdulkadiroğlu *et al.* (2017a) reference a version of Proposition 3 in a brief analysis of Chicago exam schools.

tie-breakers derived from interviews, auditions, or GPA in earlier grades, as well as test scores. A few screened-school tie-breakers are shared by multiple programs. The NYC match also includes many "unscreened schools" that use a common lottery tie-breaker.[11]

Formal analysis of markets with general tie-breaking requires notation to keep track of the tie-breakers. Let $v \in \{0, 1, ..., V\}$ index tie-breakers and let $\{S_v : v \in \{0, 1, ..., V\}\}$ be a partition of schools such that S_v is the set of schools using tie-breaker v. Schools s and s' use the same tie-breaker if and only if $s, s' \in S_v$ for some v. The random variable R_{iv} denotes applicant i's tie-breaker at schools in S_v. For any v and students $i \neq j$, tie-breakers R_{iv} and R_{jv} are assumed to be independent when both exist, though not necessarily identically distributed. Likewise, for $v \neq v'$, tie-breakers R_{iv} and $R_{iv'}$ are initially assumed to be independent, an assumption relaxed below.[12]

Define the function $v(s)$ to be the index of the tie-breaker used at school s. By definition, $s \in S_{v(s)}$. We adopt the convention that $v = 0$ identifies the lottery tie-breaker, so S_0 denotes the set of lottery schools.

With a continuum of applicants, DA assignment risk depends on priorities as well as on tie-breakers and cutoffs. We therefore combine applicants' priority status and tie-breaking variables into a single number for each school, called *applicant position* at school s:

$$\pi_{is} = \rho_{is} + R_{iv(s)}.$$

Since the difference between any two priorities is at least 1 and tie-breaking variables are between 0 and 1, applicant position at s is a lexicographic ordering, first by priority then by tie-breaker. We also generalize cutoffs to incorporate priorities; these *DA cutoffs* are denoted ξ_s. For any s that ends up full, ξ_s is given by the position of the last student offered a seat at s. Otherwise, $\xi_s = K + 1$.

Our characterization of large-market DA with general tie-breakers follows from the large market model in Abdulkadiroğlu *et al.* (2017b), replacing position as function of a single tie-breaker $(\rho_{is} + R_i)$ with the tie-breaker-specific π_{is} defined above.

In the large-market model, DA sets the cutoff to $K + 1$ at any school that remains unfilled and offers a seat at s to any applicant listing s who has

$$\pi_{is} \leq \xi_s \text{ and } \pi_{ib} > \xi_b \text{ for all } b \succ_i s. \tag{7}$$

This is a consequence of the fact that the student-proposing DA mechanism is stable. In particular, if I'm seated at s but I prefer b, I must be qualified at s and not have been offered a seat at b. Moreover, since DA offers at b are made in order of position, the fact that I wasn't offered a seat at b means I'm disqualified there.

Condition (7) nests our characterization of assignments under SD, since we can set $\rho_{is} = 0$ for all applicants and use a single tie-breaker to determine position. Statement (7) then amounts

[11] The NYC high school match omits charter schools and specialized (exam) schools like Stuyvesant. NYC charter schools make individual offers. The specialized sector runs a separate DA match.

[12] Real-world tie-breakers, including those in New York City, are often coded as ranks that may be correlated across applicants, even when the underlying criteria being ranked are independent. For example, in a sample of two, only one can be first. Such dependence vanishes as the number of applicants grows, as we show in Appendix B. Running variable positions therefore satisfy our independence assumption in a continuum market.

to saying that $R_i \leq \tau_s$ and $R_i > MID_{\theta s}$ for applicants with $\theta_i = \theta$. In finite markets, cutoffs ξ_s are stochastic, varying from tie-breaker draw to tie-breaker draw in repeated executions of the match. In large (continuum) markets, however, ξ_s is fixed. Equation (7) therefore yields a characterization of assignment risk determined by fixed cutoffs and priorities and the distribution of stochastic tie-breakers.

Our characterization of DA assignment risk covers all mechanisms in the *DA class*. Assignments using mechanisms in this class can be *computed* by student-proposing DA, possibly with applicant priorities replaced by $\boldsymbol{\phi}(\theta_i)$, where $\boldsymbol{\phi} : \Theta \rightarrow \mathbb{N}^{|S|}$ is a function of actual priorities. The DA class includes student- and school-proposing DA, serial dictatorship, and the immediate acceptance (Boston) mechanism. This class omits TTC, which need not satisfy equation (7).[13] After any transformation needed to facilitate DA computation, applicant position at school s is

$$\pi_{is} = \boldsymbol{\phi}_s(\theta_i) + R_{iv(s)}.$$

The propensity score can then be computed using this transformed position data. In what follows, we ignore this step, continuing to denote priorities by ρ_{is}.

The propensity score for DA uses the notion of *marginal priority* at school s, denoted ρ_s and defined as $\text{int}(\xi_s)$, that is, the integer part of the DA cutoff. Applicants for whom seats are rationed by tie-breakers have priority ρ_s. Conditional on rejection by all more preferred schools, applicants to s are assigned s with certainty if $\rho_{is} < \rho_s$, that is, if they clear marginal priority. Applicants with $\rho_{is} > \rho_s$ have no chance of finding a seat at s. Applicants for whom $\rho_{is} = \rho_s$ are marginal: these applicants are seated at s when their running variable values fall below randomization cutoff τ_s, which can now be written as the decimal part of the DA cutoff:

$$\tau_s = \xi_s - \rho_s = \text{frac}(\xi_s).$$

When $\rho_{is} = \rho_s$,

$$\pi_{is} \leq \xi_s \Leftrightarrow R_{iv(s)} \leq \tau_s.$$

Again, this covers SD, since ρ_{is} can be fixed at zero for everyone.

These observations motivate a partition of the set of applicant types. Specifically, partition Θ_s, the set of applicant types who list s, according to:

i) $\Theta_s^n = \{\theta \in \Theta_s \mid \rho_{\theta s} > \rho_s\}$, (*never seated*)

ii) $\Theta_s^a = \{\theta \in \Theta_s \mid \rho_{\theta s} < \rho_s\}$, (*always seated*)

iii) $\Theta_s^c = \{\theta \in \Theta_s \mid \rho_{\theta s} = \rho_s\}$. (*conditionally seated*)

[13] Under TTC, equation (7) need not be satisfied for all matching problems. But the DA class includes China's parallel mechanisms (Chen and Kesten, 2017), England's first-preference-first mechanisms (Pathak and Sönmez, 2013), and the Taiwan mechanism (Dur, Pathak, Song and Sönmez, 2018). In large markets satisfying regularity conditions that imply a unique stable matching, the DA class includes school-proposing as well as applicant-proposing DA (these conditions are spelled out in Azevedo and Leshno (2016)). For serial dictatorship, $\boldsymbol{\phi}(\theta) = (0, ..., 0)$ for all $\theta \in \Theta$. For immediate acceptance, $\boldsymbol{\phi}_s(\theta_i) < \boldsymbol{\phi}_s(\theta_j)$ if i ranks s ahead of j, and $\boldsymbol{\phi}_s(\theta_i) < \boldsymbol{\phi}_s(\theta_j)$ if and only if i and j rank s the same and $\rho_{is} < \rho_{js}$ (Ergin and Sönmez, 2006).

Never seated applicants have worse-than-marginal priority at s, so no one in this group is assigned to s. *Always seated* applicants clear marginal priority at s. Some of these applicants may end up seated at a school they prefer to s, but they're assigned s for sure if they fail to find a seat at any school they've listed more highly. Finally, *conditionally seated* applicants are marginal at s. These applicants are assigned s when not assigned a higher choice *and* when they draw a tie-breaker that clears the randomization cutoff at s. Under SD, all applicants are in Θ_s^c.

3.1 Global DA Risk

Let $F_v^i(r)$ denote the CDF of R_{iv} evaluated at r and define

$$F_v(r|\theta) = E[F_v^i(r)|\theta_i = \theta]. \tag{8}$$

This is the fraction of type θ applicants with tie-breaker v below r (set to zero when type θ lists no schools using tie-breaker v). We again assume running variables have support $[0, 1]$. As with a single tie-breaker, distributions of normalized R_{iv} depend on type.

With multiple tie-breakers, qualification at higher-listed choices may truncate the distribution of any or all R_{iv}. We therefore define running-variable-specific MIDs for each S_v. To this end, partition $B_{\theta s}$ into disjoint sets denoted by

$$B_{\theta s}^v = B_{\theta s} \cap S_v,$$

for each v. This partition is used to construct tie-breaker-specific MIDs:

$$MID_{\theta s}^v = \begin{cases} 0 & \text{if } \theta \in \Theta_b^n \text{ for all } b \in B_{\theta s}^v \text{ or if } B_{\theta s}^v = \emptyset \\ 1 & \text{if } \theta \in \Theta_b^a \text{ for some } b \in B_{\theta s}^v \\ \max\{\tau_b \mid b \in B_{\theta s}^v \text{ and } \theta \in \Theta_b^c\} & \text{otherwise} \end{cases}$$

This extends $MID_{\theta s}$ defined in (6) in two ways. In addition to capturing tie-breaker specificity, $MID_{\theta s}^v$ allows for *complete* truncation of R_{iv} when θ clears marginal priority at a school in $B_{\theta s}^v$.

$MID_{\theta s}^v$ and the partition of Θ_s by priority status determine global DA risk with general tie-breakers:

Proposition 4 (Global Score with General Tie-breaking)**.** *Consider continuum DA with multiple tie-breakers indexed by v, distributed independently of one another according to $F_v(r|\theta)$. For all s and θ in this match,*

$$p_s(\theta) =$$
$$\begin{cases} 0 & \textit{if } \theta \in \Theta_s^n \\ \prod_v (1 - F_v(MID_{\theta s}^v|\theta)) & \textit{if } \theta \in \Theta_s^a \\ \prod_{v \neq v(s)} (1 - F_v(MID_{\theta s}^v|\theta)) \times \max\left\{0, F_{v(s)}(\tau_s|\theta) - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)\right\} & \textit{if } \theta \in \Theta_s^c \end{cases}$$

where $F_{v(s)}(\tau_s|\theta) = \tau_s$ and $F_{v(s)}(MID_{\theta s}^{v(s)}|\theta) = MID_{\theta s}^0$ when $v(s) = 0$.

Proposition 4, which generalizes an earlier multiple lottery tie-breaker result in Abdulkadiroğlu *et al.* (2017b), covers three sorts of applicants, corresponding to the partition of Θ_s.

First, applicants with less-than-marginal priority at s have no chance have being seated there. The second line of the theorem reflects the likelihood of qualification at schools preferred to s among applicants surely seated at s when they can't do better. Since running variables are assumed independent, the probability of not doing better than s is described by a product over tie-breakers, $\prod_v (1 - F_v(MID^v_{\theta s}|\theta))$. If type θ is sure to do better than s, then $MID^v_{\theta s} = 1$ and risk at s is zero.

Finally, risk for applicants in Θ^c_s multiplies the term

$$\prod_{v \neq v(s)} (1 - F_v(MID^v_{\theta s}|\theta))$$

by

$$\max\left\{0, F_{v(s)}(\tau_s|\theta) - F_{v(s)}(MID^{v(s)}_{\theta s}|\theta)\right\}.$$

The first of these is the probability of failing to improve on s by virtue of being seated at schools using a tie-breaker *other* than $v(s)$. The second parallels assignment risk in single-tie-breaker SD: to be seated at s, applicants in Θ^c_s must have $R_{iv(s)}$ between $MID^{v(s)}_{\theta s}$ and τ_s.

Proposition 4 allows for single tie-breaking, lottery tie-breaking, or a mix of non-lottery and lottery tie-breakers as in the NYC high school match. With a single tie-breaker, the risk formula simplifies, omitting product terms over v:

Corollary 2 (Abdulkadiroğlu *et al.* (2017b))**.** *Consider a continuum DA match using a single tie-breaker, R_i, distributed according to $F_R(r|\theta)$ for type θ. For all s and θ in this market, we have:*

$$p_s(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta^n_s, \\ 1 - F_R(MID_{\theta s}|\theta) & \text{if } \theta \in \Theta^a_s, \\ (1 - F_R(MID_{\theta s}|\theta)) \times \max\left\{0, \dfrac{F_R(\tau_s|\theta) - F_R(MID_{\theta s}|\theta)}{1 - F_R(MID_{\theta s}|\theta)}\right\} & \text{if } \theta \in \Theta^c_s, \end{cases}$$

where $p_s(\theta) = 0$ when $MID_{\theta s} = 1$ and $\theta \in \Theta^c_s$, and $MID_{\theta s}$ is as defined in Section 2.2, applied to a single tie-breaker.

Common lottery tie-breaking for all schools further simplifies the DA propensity score. When $v(s) = 0$ for all s, $F_R(MID_{\theta s}) = MID_{\theta s}$ and $F_R(\tau_s|\theta) = \tau_s$, as in the Denver match analyzed by Abdulkadiroğlu *et al.* (2017b). In this case, the DA propensity score is a function only of $MID_{\theta s}$ and the partition of Θ_s into applicants that are never, always, and conditionally seated. This contrasts with the scores in Proposition 2 and Proposition 4, which depend on the unknown and unrestricted conditional distributions of tie-breakers given type ($F_R(\tau_s|\theta)$ and $F_R(MID_{\theta s}|\theta)$ with a single tie-breaker; $F_v(\tau_s|\theta)$ and $F_v(MID_{\theta s}|\theta)$ with general tie-breakers). We therefore turn again to local risk to isolate offers that are independent of type and potential outcomes.

3.2 DA Goes Local

Under general DA, local risk is defined only in marginal priority groups. We therefore modify the set of t_{is} variables to be

$$t_{is}(\delta) = \begin{cases} n & \text{if } \theta \in \Theta_s^n \text{ or, if } v(s) \neq 0,\ \theta \in \Theta_s^c \text{ and } R_{iv(s)} > \tau_s + \delta \\ a & \text{if } \theta \in \Theta_s^a \text{ or, if } v(s) \neq 0,\ \theta \in \Theta_s^c \text{ and } R_{iv(s)} < \tau_s - \delta \\ c & \text{if } \theta \in \Theta_s^c \text{ and, if } v(s) \neq 0,\ R_{iv(s)} \in [\tau_s - \delta, \tau_s + \delta] \end{cases}$$

for each applicant and school. These are collected in a vector,

$$T_i(\delta) = [t_{i1}(\delta), ..., t_{is}(\delta), ..., t_{iS}(\delta)]'.$$

This expands the classification of applicants to school s into $t_{is}(\delta) = a$, n, or c by including those who fail to clear marginal priority at s in group n and including those who clear marginal priority at s in group a.

The *local DA propensity score* is defined as a function of type and cutoff proximity, as summarized by $T_i(\delta)$:

$$\psi_s(\theta, T) = \lim_{\delta \to 0} E[D_i(s) | \theta_i = \theta, T_i(\delta) = T],$$

for $T = [t_1, ..., t_s, ..., t_S]' \in \{n, a, c\}^S$. This describes assignment risk for applicants with tie-breaker values above, below, and near cutoffs for any and all schools in the match. We again require that all tie-breaker distributions be continuously differentiable at randomization cutoffs and that these cutoffs be distinct:

Assumption 1. *(a) For every v and for $r = \tau_1, ..., \tau_S$, $F_v^i(r|e)$ is continuously differentiable with $F_v^{i'}(r|e) > 0$ given any event e of the form that $\theta_i = \theta, R_{iu} > r_u$ for $u = 1, ..., v-1$, and R_i is contained by a closed ball containing r. (b) $\tau_s \neq \tau_{s'}$ for any schools $s \neq s'$ with $\tau_s \neq 0$ and $\tau_{s'} \neq 0$.*

This set-up yields a compact and useful characterization of local assignment risk in continuum DA with general tie-breaking:

Theorem 1 (Local Score with General Tie-breaking)**.** *Consider continuum DA with multiple tie-breakers indexed by v, distributed according to $F_v(r|\theta)$, and suppose Assumption 1 holds. For all $s \in S$, $\theta \in \Theta_s$, $T = [t_1, ..., t_s, ..., t_S]' \in \{n, a, c\}^S$, and all w, we have*

$$\lim_{\delta \to 0} E[D_i(s) | \theta_i = \theta, T_i(\delta) = T, W_i = w] = \psi_s(\theta, T),$$

where $\psi_s(\theta, T) = 0$ if (a) $t_s = n$; or (b) $t_b = a$ for some $b \in B_{\theta s}$. Otherwise,

$$\psi_s(\theta, T) = \begin{cases} 0.5^{m_s(\theta,T)}(1 - MID_{\theta s}^0) & \text{if } t_s = a \\ 0.5^{m_s(\theta,T)} \max\left\{0, \tau_s - MID_{\theta s}^0\right\} & \text{if } t_s = c \text{ and } v(s) = 0 \\ 0.5^{1+m_s(\theta,T)}(1 - MID_{\theta s}^0) & \text{if } t_s = c \text{ and } v(s) > 0. \end{cases} \quad (9)$$

where $m_s(\theta, T) = |\{v > 0 :$ for some $b \in B_{\theta s}^v, MID_{\theta s}^v = \tau_b$ and $t_b = c\}|$.

The local DA score for type θ applicants is determined in part by the screened schools θ prefers to s. Relevant screened schools are those at which applicants to s are in the marginal priority group with a tie-breaker close to randomization cutoffs. The variable $m_s(\theta, T)$ counts the number of tie-breakers involved in such close encounters. As expressed in equation (4) for the single-school case, applicants drawing screened school tie-breakers close to τ_b for some $b \in B^v_{\theta s}$ face multiplicative risk of 0.5.

Theorem 1 starts with a scenario where applicants to s are either sure to do better or are never seated at s and therefore face no risk there. In this case, we need not worry about whether s is a screened or lottery school. In other scenarios, where applicants fail to improve on s, risk at any lottery s is determined in part by truncation of the lottery tie-breaker at more preferred lottery schools and by possible qualification at more preferred screened schools, where qualification risk is 0.5. These sources of risk combine to produce the second line of (9). Similarly, risk at any screened s is determined by possible qualification at more preferred schools (lottery and screened) plus an additional 0.5 risk term for those marginal at s. This explains the addition of 1 to the exponent in the third line of equation (9).

This theorem also yields a general conditional independence relation, similar to Corollary 1:

$$\lim_{\delta \to 0} P[D_i(s) = 1 | \theta_i = \theta, T_i(\delta) = T, W_i = w, \psi_s(\theta, T) = p] = p, \tag{10}$$

for $p \in [0, 1]$. In other words, fixing $\psi_s(\theta, T)$, DA-generated offers are independent of type and any W_i that's unaffected by treatment. Local conditional independence allows us to eliminate OVB by conditioning on $\psi_s(\theta, T)$. Moreover, $\psi_s(\theta, T)$ is typically far coarser than the underlying type distribution. Appendix A illustrates risk calculations based on Theorem 1 for a number of simple examples.

Isolating Lottery Risk

An important implication of Theorem 1 is that lottery schools create assignment risk at screened schools for applicants with tie-breaker values away from screened school cutoffs. We isolate this *lottery risk* with the aid of the following applicant classifier:

$$\ell_{is} = \begin{cases} n & \text{if } t_{is}(0) = n \\ a & \text{if } t_{is}(0) = a \text{ or if } t_{is}(0) = c \text{ and } v(s) > 0 \\ c & \text{if } t_{is}(0) = c \text{ and } v(s) = 0. \end{cases} \tag{11}$$

Classification variable ℓ_{is} sets $\delta = 0$, effectively turning screened-school tie-breakers into priorities. As with the t_{is}, collect the group of ℓ_{is} in a vector,

$$L_i = [\ell_{i1}, ..., \ell_{is}, ..., \ell_{iS}]',$$

and define

$$\lambda_s(\theta, L) = E[D_i(s) | \theta_i = \theta, L_i = L],$$

for $L = [\ell_1, ..., \ell_s, ..., \ell_S]' \in \{n, a, c\}^S$. Note that, having fixed $\delta = 0$, we no longer need be concerned with limiting risk. Then,

$$\lambda_s(\theta, L) = \begin{cases} 0 & \text{if } \ell_s = n \text{ or if } \ell_b = a \text{ for some } b \in B_{\theta s} \\ (1 - MID^0_{\theta s}) & \text{if } \ell_s = a \\ \max\left\{0, \tau_s - MID^0_{\theta s}\right\} & \text{if } \ell_s = c. \end{cases} \tag{12}$$

The second line of (12) describes non-degenerate lottery risk at screened schools. Lotteries create risk at screened schools because students who list lottery schools ahead of screened schools need not qualify for lottery-based admission; this happens with probability $1 - MID^0_{\theta s}$. There is "less risk" generated by lotteries than general risk because $\lambda_s(\theta, L)$ is more likely than $\psi_s(\theta, T)$ to equal zero or one, especially for screened schools.[14] Still, lottery risk may be enough to evaluate both screened and lottery schools in a match where applicants list schools of both types. This is worth highlighting because evidence on screened school effects generated by lottery risk comes partly from applicants with tie-breakers far from cutoffs. We explore this idea in the empirical application, below.

3.3 Estimating the Local Score

A sample analog of the theoretical local DA score described by Theorem 1 is shown here to converge uniformly to the corresponding local score for a finite market, in an asymptotic sequence that increases market size with a shrinking bandwidth. Our empirical application establishes the relevance of this asymptotic result by showing that applicant characteristics are balanced by offer status conditional on estimates of the local propensity score.

The sequence used to study the estimated score increases the size of a random sample of N applicants. We refer to sampled applicants by the order in which they're sampled, that is, by $i \in \{1, 2, ..., N\}$. The applicant sample is augmented with information on applicant type and large-market school capacities, $\{q_s\}$, which give the proportion of the market that can be seated at s. Each applicant is associated with an individual tie-breaker distribution, $F^i_v(r)$, as described above. We observe a realized tie-breaker value for each applicant, but not the underlying distribution.

Now, fix the number of seats at school s in each sampled finite market to be the integer part of Nq_s and run DA with these applicants and schools. We consider the limiting behavior of an estimator that uses the resulting $MID^v_{\theta s}$, τ_s, and marginal priorities generated by this single realization. Also, given a bandwidth $\delta_N > 0$, we determine $t_{is}(\delta_N)$ for each i and s. This is used to compute

$$\hat{m}_{Ns}(\theta, T) = |\{v > 0 : MID^v_{\theta s} = \tau_b \text{ and } t_{ib}(\delta_N) = c \text{ for some } b \in B^v_{\theta s}\}|.$$

Our propensity score estimator is constructed from plugging these ingredients into the formula in Theorem 1. If $t_{is}(\delta_N) = n$ or $t_{ib}(\delta_N) = a$ for some $b \in B_{\theta s}$, then

$$\hat{\psi}_{Ns}(\theta, T; \delta_N) = 0.$$

[14]The third line of (12) describes lottery risk at lottery schools.

Otherwise,

$$\hat{\psi}_{Ns}(\theta, T; \delta_N) = \begin{cases} 0.5^{\hat{m}_{Ns}(\theta,T)}(1 - MID^0_{\theta s}) & \text{if } t_{is}(\delta_N) = a \\ 0.5^{\hat{m}_{Ns}(\theta,T)} \max\left\{0, \tau_s - MID^0_{\theta s}\right\} & \text{if } t_{is}(\delta_N) = c,\ v(s) = 0 \\ 0.5^{1+\hat{m}_{Ns}(\theta,T)}(1 - MID^0_{\theta s}) & \text{if } t_{is}(\delta_N) = c,\ v(s) \neq 0. \end{cases}$$

Note that $\tau_s,\ MID^0_{\theta s}$, and $\hat{m}_{Ns}(\theta, T)$ in this expression are sample quantities.

As a theoretical benchmark for the large-sample performance of $\hat{\psi}_{Ns}(\theta, T; \delta_N)$, we define the true local score for the finite market of size N. This is

$$\psi_{Ns}(\theta, T) = \lim_{\delta \to 0} E_N[D_i(s) | \theta_i = \theta, T_i(\delta) = T],$$

where E_N is the expectation induced by the set of running variable distributions $\{F^i_v(r); i = 1, 2, ..., N\}$ for applicants in the finite market. This quantity fixes the distribution of types and the vector of proportional school capacities, as well as market size. $\psi_{Ns}(\theta, T)$ is the limit of the average of $D_i(s)$ across infinitely many tie-breaker draws in ever-narrowing windows near cutoffs in a match governed by these parameters. Because tie-breaker distributions are assumed to have continuous density in the neighborhood of any cutoff, the population average assignment rate is well-defined for any positive δ.

We're interested in the gap between the estimator $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ and the true local score $\psi_{Ns}(\theta, T)$ as N grows and δ_N shrinks. We can show that $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ described above converges uniformly to $\psi_{Ns}(\theta, T)$ in such a sequence. This result uses a regularity condition:

Assumption 2. *(Rich support) Let s^1_i identify applicant i's first choice school. In the continuum market, for every school s and every priority ρ held by a positive mass of applicants at s, the proportion of applicants with $s^1_i = s$ and $\rho_{is} = \rho$ is also positive.*

This says that for each priority group at school s represented among applicants in the continuum, some applicants list s first.

Uniform convergence of $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ is formalized below:

Theorem 2 (Consistency of the DA Local Score)**.** *In the asymptotic sequence described above and maintaining Assumptions 1 and 2, the estimated local propensity score $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ is a consistent estimator of $\psi_{Ns}(\theta, T)$ in the following sense: For any δ_N such that $\delta_N \to 0$ and $N\delta_N \to \infty$ as $N \to \infty$,*

$$\sup_{\theta \in \Theta, s \in S, T \in \{n,c,a\}^S} |\hat{\psi}_{Ns}(\theta, T; \delta_N) - \psi_{Ns}(\theta, T)| \xrightarrow{p} 0,$$

as $N \to \infty$.

Proof. The proof uses lemmas established in the appendix. The first lemma shows that the vector of DA cutoffs computed for the sampled market, $\hat{\xi}_N$, converges to the vector of cutoffs in the continuum, that is,

$$\hat{\xi}_N \xrightarrow{a.s.} \xi,$$

where ξ denotes the vector of continuum cutoffs. This result implies that the estimated score converges to the large-market local score as market size grows and bandwidth shrinks. Specifically, for all $\theta \in \Theta, s \in S$, and $T \in \{n, c, a\}^S$, we have

$$\hat{\psi}_{Ns}(\theta, T; \delta_N) \xrightarrow{a.s.} \psi_s(\theta, T)$$

as $N \to \infty$ and $\delta_N \to 0$.

The second lemma shows that the true finite market score with a fixed bandwidth, defined as $\psi_{Ns}(\theta, T; \delta_N) \equiv E_N[D_i(s)|\theta_i = \theta, T_i(\delta_N) = T]$, also converges to $\psi_s(\theta, T)$ as market size grows and bandwidth shrinks. That is, for all $\theta \in \Theta, s \in S$, $T \in \{n, c, a\}^S$, and δ_N such that $\delta_N \to 0$ and $N\delta_N \to \infty$ as $N \to \infty$,

$$\psi_{Ns}(\theta, T; \delta_N) \xrightarrow{p} \psi_s(\theta, T)$$

as $N \to \infty$.

Finally, the definitions of $\psi_{Ns}(\theta, T; \delta_N)$ and $\psi_{Ns}(\theta, T)$ imply that $|\psi_{Ns}(\theta, T; \delta_N) - \psi_{Ns}(\theta, T)| \xrightarrow{a.s.} 0$ as $\delta_N \to 0$. Combining these results shows that for all $\theta \in \Theta, s \in S$, and T, as $N \to \infty$ and $\delta_N \to 0$ with $N\delta_N \to \infty$, we have

$$\begin{aligned}
&|\hat{\psi}_{Ns}(\theta, T; \delta_N) - \psi_{Ns}(\theta, T)| \\
=&|\hat{\psi}_{Ns}(\theta, T; \delta_N) - \psi_{Ns}(\theta, T; \delta_N) + \psi_{Ns}(\theta, T; \delta_N) - \psi_{Ns}(\theta, T)| \\
\leq&|\hat{\psi}_{Ns}(\theta, T; \delta_N) - \psi_{Ns}(\theta, T; \delta_N)| + |\psi_{Ns}(\theta, T; \delta_N) - \psi_{Ns}(\theta, T)| \\
\xrightarrow{p}&|\psi_s(\theta, T) - \psi_s(\theta, T)| + 0 \\
=&0.
\end{aligned}$$

This yields the theorem since Θ, S, and $\{n, c, a\}^S$ are finite. □

Theorem 2 justifies our use of the formula in Theorem 1 to eliminate OVB in empirical work estimating school attendance effects.

4 A Brief Report on NYC Report Cards

Since the 2003-04 school year, the NYC Department of Education (DOE) has used DA to assign rising ninth graders to high schools. Each applicant for a ninth grade seat can list up to twelve programs. All traditional public high schools participate in the match, but charter schools and NYC's exam schools have separate admissions procedures.[15]

The NYC match is structured like the match described in Section 3: lottery schools use a common randomly assigned tie-breaker, while screened schools use a variety of non-lottery tie-breaking variables. Screened school tie-breakers are mostly distinct, with one for each school

[15] Some special needs students are also matched separately. The centralized NYC high school match is detailed in Abdulkadiroğlu *et al.* (2005, 2009). Abdulkadiroğlu *et al.* (2014) describe NYC exam school admissions.

or program, though some screened programs share a tie-breaker. In any case, our theoretical framework accommodates all of NYC's many tie-breaking protocols.[16]

Many high schools in the match host multiple programs, and each program can set its own priorities. Within priority groups, programs ration seats in one of two ways. Three types of programs rank applicants systematically. These include *screened* programs, which order applicants based on academic grades and previous attendance records, *audition* programs that order applicants using interviews or other qualitative assessments, and half of the seats in *educational option* (Ed-Opt) programs. Our analysis refers to programs of all three types as "screened" since all use some sort of non-lottery tie-breaker.

Programs that are not screened break ties using a common lottery tie-breaker. The group of lottery programs includes *unscreened* programs that admit students randomly, *limited unscreened* programs that admit randomly but give priority to students who attend information sessions, and the unscreened half of capacity at Ed-Opt programs. Our analysis uses Theorem 1 to compute propensity scores for programs rather than schools since programs are the unit of assignment.[17] But since the match yields a single offer, we can sum program propensity scores to produce school-level scores and then sum again for groups of schools. The score for attendance at any screened Grade A school, for example, is the sum of the scores for all screened Grade A schools in the match. For our purposes, a lottery school is a school hosting any lottery program; other schools are screened.[18]

In 2007, the NYC DOE launched a school accountability system that graded schools from A to F. This mirrors similar accountability systems in Florida and other states. NYC's school grades were determined by achievement levels and, especially, achievement growth, as well as by survey- and attendance-based features of the school environment. Growth looked at credit accumulation, Regents test completion and pass rates; performance measures were derived mostly from four- and six-year graduation rates. Some schools were ungraded. Figure 3 reproduces a sample letter-graded school progress report.[19]

The 2007 grading system was controversial. Proponents applauded the integration of multiple measures of school quality while opponents objected to high-stakes consequences of low school grades, such as school closure or consolidation. Rockoff and Turner (2011) provide a partial validation of the system by showing that low grades seem to have sparked school improvement. In 2014, the DOE replaced the 2007 scheme with school quality measures that place less weight on test scores and more on curriculum characteristics and subjective assessments of teaching

[16]NYC screened school tie-breakers are defined for all applicants who list a school on their preference form, and are reported as an integer reflecting raw tie-breaker order in this group. We scale these so as to lie in $(0, 1]$ by transforming raw tie-breaking realizations R_{iv} into $[R_{iv} - \min_j R_{jv} + 1]/[\max_j R_{jv} - \min_j R_{jv} + 1]$ for each tie-breaker v. This transformation produces a positive cutoff at s when only one applicant is seated at s and a cutoff of 1 when all applicants who list s are seated there.

[17]Seats for Ed-Opt programs are split into halves, one of which screens applicants using a single non-lottery tie-breaker while the other uses the common lottery number. These groups are further subdivided as described in the appendix.

[18]Some NYC high schools sort applicants on a coarse screening tie-breaker that allows ties, while breaking these ties using the common lottery number. Schools of this type are treated as lottery schools, adding priority groups defined by values of the screened tie-breaker.

[19]Walcott (January 2012) details the NYC grading methodology used in this period.

quality. The relative merits of the old and new systems continue to be debated.

We showcase the use of general tie-breaking for impact evaluation by estimating effects of being assigned to a Grade A school. This analysis uses application data from the 2011-12, 2012-13 and 2013-14 school years. Our sample includes first-time applicants seeking 9th grade seats, who submit preferences over programs in the main round of the NYC high school match. Data include school capacities and priorities, lottery numbers, and screened school tie-breakers, information that allows us to replicate the match. Our replication is nevertheless imperfect, perhaps due to mistakes in the recording of tie-breakers. Among other problems, we see occasional gaps in the lists that show how schools order their applicants, while some applicants appear to share a position at a given school. Our online appendix details the manner in which these and other problems related to match replication are addressed.

Students at Grade A schools have higher average SAT scores and higher graduation rates than do students at other schools. Differences in graduation rates across schools feature in media accounts of socioeconomic differences in NYC high school match results (see, e.g., Harris and Fessenden (2017) and Disare (2017)). Grade A students are also more likely than students attending other schools to be deemed "college- and career-prepared" or "college-ready".[20] These and other school characteristics are documented in Table 1. Achievement gaps between screened and lottery Grade A schools are especially large. This likely reflects selection bias induced by test-based screening.

Screened Grade A schools have a majority white or Asian student body, the only group of schools described in the table to do so. These schools are also over-represented in Manhattan, a borough that includes most of New York's wealthiest neighborhoods (though average family income is higher on Staten Island). Variables like spending and teacher experience are broadly similar across school types, while screened Grade A schools are somewhat larger than the rest.

Table 2 describes the nearly 153,000 eighth graders with non-missing baseline (application-year) covariates applying for ninth grade seats in fall 2012, 2013 and 2014. Roughly 130,000 list a Grade A school on their application form and a little over a third of these are offered a Grade A seat. The difference between total 9th grade enrollment (about 183,000) and the number of match participants is accounted for by groups of special education students outside the main match, direct-to-charter enrollment, and a few schools that straddle 9th grade. Applicants in the match have baseline (6th grade) scores above the overall district mean (baseline scores are standardized to the population of test-takers). As can be seen by comparing columns 2 and 3 in Table 2, however, the average characteristics of Grade A applicants are generally like those of the entire applicant population.

The statistics in column 4 of Table 2 show that applicants *enrolled* in a Grade A school (among schools participating in the match) are somewhat less likely to be black and have higher baseline scores than the total applicant pool. Here too, these gaps likely reflect selection bias at screened Grade A schools. Most of those attending a Grade A school were offered a seat there, and most ranked a Grade A school first. Grade A students are about twice as likely to go to a lottery school as to a screened school.

[20] These composite variables are determined as a function of Regents and AP scores, course grades, vocational or arts certification, and college admissions tests.

Enthusiasm for Grade A schools is far from universal: just under half of all applicants in the match list a Grade A school first. Around 31,000 Grade A applicants have non-degenerate risk of Grade A assignment, that is, an estimated $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ strictly between 0 and 1, conditional on which there's variation in offer status. Applicants at risk of Grade A assignment, described in column 5 of Table 2, have baseline scores and demographic characteristics much like those of the sample enrolled at a Grade A school. The ratio of screened to lottery enrollment among those with Grade A risk is also similar to the corresponding ratio in the sample of enrolled students (compare 33.4/16.6 in the former group to 71.9/28.5 in the latter).

All Grade A schools in the match have applicants exposed to non-degenerate assignment risk in at least one cohort. The y-axis in Figure 4 shows the number of applicants to each Grade A school added to the added-risk sample by consideration of general risk instead of lottery risk. This is plotted against the number of applicants subject only to lottery risk on the x-axis. Applicants are said to have lottery risk when their estimated $\lambda_s(\theta, L)$, the formula for risk when non-lottery tie-breakers are treated as priorities, is strictly between 0 and 1. Orange and blue circles plot numbers of applicants at risk for each lottery and screened Grade A school, respectively, where circle sizes are scaled by school capacity.

Figure 4 shows that the sample size gains yielded by consideration of general risk are both broad (that is, appearing for many schools) and large. Moreover, as can be seen from the points plotted near the y-axis, many schools with no applicants subject to lottery risk (mostly screened schools) have applicants with general risk. At the same time, blue circles away from the y-axis show that many screened schools have applicants subject to lottery risk.

The balancing property of propensity score conditioning is documented in Table 3, which reports raw and score-controlled differences in covariate means for applicants who do and don't receive Grade A offers. Score-controlled differences are estimated in the following setup. Let D_{1i} be a dummy indicating Grade A school offers and let $d_{1i}(x)$ be a dummy indicating $\hat{p}_{1i} = x$, where x indexes values the score might take. Likewise, let D_{0i} indicate offers at ungraded schools and let $d_{0i}(x)$ be a dummy indicating $\hat{p}_{0i} = x$. Estimated propensity scores for Grade A and ungraded schools offers, denoted $\hat{p}_{1i}$ and $\hat{p}_{0i}$, are computed by summing estimated scores for Grade A and ungraded schools, respectively. We control for ungraded school offers to ensure that estimated Grade A effects compare schools with high and low grades, omitting the ungraded.[21]

Let W_i be any applicant covariate measured before assignment, including features of θ_i. Balance tests are estimates of parameter γ_1 in

$$W_i = \gamma_1 D_{1i} + \gamma_0 D_{0i} + \sum_x \alpha_1(x) d_{1i}(x) + \sum_x \alpha_0(x) d_{0i}(x) + h(\mathcal{R}_i) + \nu_i, \tag{13}$$

with local linear control for the running variable parameterized as

$$h(\mathcal{R}_i) = \sum_{s \in S \setminus S_0} \omega_{1s} a_{is} + k_{is}[\omega_{2s} + \omega_{3s}(R_{iv(s)} - \tau_s) + \omega_{4s}(R_{iv(s)} - \tau_s)\mathbf{1}(R_{iv(s)} > \tau_s)], \tag{14}$$

where $\mathcal{R}_i \equiv [R_{i0}, ..., R_{iV}]'$ is the vector of tie-breakers, $S \setminus S_0$ is the set of screened programs, a_{is} indicates whether applicant i applied to program s, and $k_{is} = a_{is} \times \mathbf{1}(\xi_s - \delta_s < R_{iv(s)} < \xi_s + \delta_s)$

[21] Ungraded schools are mostly new or have insufficient data to determine a grade.

indicates applicants in a bandwidth of size δ_s around cutoff ξ_s. Parameters in (13) and (14) vary by application cohort. The sample is limited to applicants with non-degenerate Grade A offer risk. Bandwidths are estimated as suggested by Imbens and Kalyanaraman (2012), separately for each program, for the set of applicants in the relevant marginal priority group.[22]

As can be seen in column 2 of Table 3, applicants offered a Grade A seat are much more likely to have listed a Grade A school first, and listed more Grade A schools than did other applicants. Minority and free-lunch-eligible applicants are less likely to be offered a Grade A seat, while those offered a Grade A seat have much higher baselines scores, with gaps in the range of 0.3 and 0.4 standard deviations in favor of those offered. These raw differences notwithstanding, our theoretical results suggest that estimates of γ_1 in equation (13) should be close to zero. This is borne out by the estimates reported in column 4 of the table, which shows small, mostly insignificant differences in covariates by offer status when estimated using equation (13).

Table 3 also documents balance for applicants subject to lottery risk. In particular, column 5 reports estimates of γ_1 from a version of equation (13) estimated using the lottery risk propensity score, $\lambda_s(\theta, L)$, to define risk sets. Conditioning on lottery risk reduces the estimation sample to those with an estimated $\lambda_s(\theta, L)$ strictly between 0 and 1. Here too, we see only small differences in covariate means, most of which are not significantly different from zero. This is perhaps less impressive than the balance documented in column 4 since lottery risk is randomly assigned. Still, the estimates in columns 4 and 5 of Table 3 establish the empirical relevance of both the large-market framework and the notion of limiting local risk underlying the theoretical results in Section 3.

The encouraging balance results in Table 3 are especially noteworthy in view of Figure 1, which shows that an IK bandwidth is insufficiently narrow to drive the propensity score for qualification at Townsend Harris to the theoretical limit of one-half. Running variable control via local linear regression mitigates this approximation error. Our local linear regression estimation strategy, which combines saturated control for the propensity score with linear running variable control can be seen as a "doubly robust" score-based estimator of the sort suggested by Robins (2000) and Okui *et al.* (2012), the latter in an IV context. Even if the local score is poorly approximated, running variable controls minimizes omitted variable bias from non-lottery tie-breakers. At the same time, the theoretical score tells us which tie-breakers are important and for whom.

Causal effects of school attendance on test scores are estimated by 2SLS, using offer dummies as instruments for years of exposure to schools of a particular type. Exposure variables are denoted C_{1i} and C_{0i} for Grade A and ungraded schools, respectively. Effects on graduation outcomes are estimated by replacing years of exposure with dummies for ninth grade enrollment. The causal effects of interest are 2SLS estimates of parameter β_1 in

$$Y_i = \beta_1 C_{1i} + \beta_0 C_{0i} + \sum_x \alpha_{21}(x) d_{1i}(x) + \sum_x \alpha_{20}(x) d_{0i}(x) + g(\mathcal{R}_i) + \eta_i, \tag{15}$$

[22]Bandwidths are also computed separately for each outcome variable; we use the smallest of these for each program.

with associated first stage equations,

$$C_{1i} = \gamma_{11}D_{1i} + \gamma_{10}D_{0i} + \sum_x \alpha_{11}(x)d_{1i}(x) + \sum_x \alpha_{10}(x)d_{0i}(x) + h_1(\mathcal{R}_i) + \nu_{1i} \tag{16}$$

$$C_{0i} = \gamma_{01}D_{1i} + \gamma_{00}D_{0i} + \sum_x \alpha_{01}(x)d_{1i}(x) + \sum_x \alpha_{00}(x)d_{0i}(x) + h_0(\mathcal{R}_i) + \nu_{0i}.$$

Running variable control functions in these equations, denoted $h_1(\mathcal{R}_i), h_2(\mathcal{R}_i)$, and $g(\mathcal{R}_i)$, are analogous to (14). Risk set dummies $d_{1i}(x)$ and $d_{0i}(x)$ are included as in equation (13). Reported results are from specifications that also control for baseline math and English scores; free lunch, special education, and English language learner indicators, and for gender and race dummies (estimates without these controls are similar, though less precise). The three applicant cohorts in our sample are stacked, so all parameters except $\beta_1, \beta_0, \gamma_{11}, \gamma_{10}, \gamma_{01}$, and γ_{00} are interacted with cohort.

Theorems 1 and 2 imply that grade A and ungraded school offers are locally and asymptotically independent of potential outcomes conditional on estimates of the relevant local score. Given an exclusion restriction, the conditional random assignment of school offers supports a causal interpretation of the 2SLS estimates, $\hat{\beta}_1$ and $\hat{\beta}_0$, as capturing the effect of attendance at different sorts of schools. The exclusion restriction in this case means that Grade A and ungraded school offers have no effect on outcomes other than by boosting time spent at Grade A and ungraded schools.

This exclusion restriction fails when Grade A and ungraded school offers change school quality by moving applicants between schools of different quality *within* the Grade A or another sector. In other words, Grade A and ungraded school offers might change the type of school attended on margins other than a school's grade. We therefore explore multi-sector models that distinguish causal effects of attendance at different types of Grade A schools. Estimates of these multi-sector models are discussed following the discussion of overall Grade A effects.

OLS estimates of Grade A effects, reported as a benchmark in the second column of Table 4, show Grade A attendance is associated with higher SAT scores, graduation rates, and college and career readiness. The OLS estimates in Table 4 are constructed by fitting equation (15), without propensity score controls or instrumenting, in a sample that includes all participants in the high school match without regard to offer risk, though limited to applicants with the relevant outcomes. OLS estimates of SAT gains are around 6 points on a base of 473-4. Graduation gains are similarly modest at around 4 points, but effects on college and career readiness are substantial, running 8-10 points on a base rate around 40.

First stage effects of Grade A offers on Grade A enrollment, computed by estimating equation (16) and reported in Panel A of Table 4, show that offers of a Grade A seat boost Grade A enrollment by 1.8-1.9 years between the time of application and SAT test-taking. Grade A offers boost the likelihood of any Grade A enrollment by about 65 percentage points. This can be compared with Grade A enrollment rates around 18 percent among those not offered a Grade A seat in the match.

In contrast with the OLS estimates in column 2, the 2SLS estimates in column 4 of Table 4 suggest that most of the SAT gains associated with Grade A attendance reflect selection bias. The 2SLS estimate of an effect on SAT math is only around 2.4, though marginally significantly different from zero with a standard error of 0.8. The corresponding 2SLS estimate of reading gains is even smaller and not significantly different from zero, though estimated with similar precision.The 2SLS estimate for graduation shows a marginally significant gain of 3 points. The estimated standard error of 0.013 associated with the graduation estimate seems especially noteworthy, as this means our research design has the power to uncover even modest improvements.[23]

The strongest Grade A effects appear for indicators of college and career preparedness and college readiness. These two indicators, detailed in our online appendix, measure various sorts of achievement and certification milestones, and contribute to the determination of school grades. OLS estimates show gains around 10 points on a base of 42 for college and career preparedness. Similarly, OLS estimates suggest Grade A schools boost college readiness by around 8 points on a base of 37. 2SLS estimates are remarkably close to these and estimated with a level of precision about like that of the graduation estimates.

As explaining in Section 3.2, lottery risk alone can be used to identify Grade A effects for applicants in the NYC match, including those attending screened schools. Estimates identified by lottery risk, reported in column 5 of Table 4, are computed using a 2SLS set-up that replaces general risk controls with saturated control for $\lambda_s(\theta, L)$ in the sample with non-degenerate lottery risk. These "lottery-only" results are remarkably similar to the 2SLS estimates generated by combining all sources of risk. This finding suggests that applicants subject to screened-school risk are not particularly unusual, at least as far as Grade A treatment effects go. On the other hand, the move from lottery to general risk yields a valuable precision gain. For example, the standard error on the graduation effect falls from 0.018 using only lottery risk to 0.013 using general risk.

NYC education policy discussions often revolve around access to screened schools. This longstanding policy interest, along with concerns about within-sector changes in school quality that might violate our 2SLS exclusion restrictions, motivates an analysis that distinguishes screened from lottery Grade A effects.

The multi-sector estimates reported in Table 5 are from models that include separate endogenous variables for screened Grade A schools and for lottery Grade A schools, along with a third endogenous variable for the ungraded sector. Instruments for this just-identified set-up are two dummies indicating each sort of Grade A offer, as well as a dummy indicating ungraded school offers. 2SLS models include separate saturated propensity score controls for each Grade A school sector offer, as well as for the risk of an ungraded school offer. These multi-sector estimates are computed in a sample limited to applicants at risk of assignment to either a screened or lottery Grade A school.

OLS estimates again provide an interesting benchmark. As can be seen in the first two columns of Table 5, screened Grade A students appear to reap a large SAT advantage even after controlling for baseline achievement and other covariates. In particular, OLS estimates of Grade A effects for schools in the screened sector are on the order of 13-16 points. At the same

[23] Appendix Table B1 shows little difference in follow-up rates between applicants who are and aren't offered a Grade A seat. The 2SLS estimates in Table 4 are therefore unlikely to be compromised by differential attrition associated with Grade A offers.

time, Grade A lottery schools appear to generate achievement gains around only one point. By contrast, 2SLS estimates of multi-sector models, reported in columns 3 and 4 of Table 5, show equally modest SAT effects for Grade A schools in both sectors. These range from 2-3 points for math, with smaller estimates for reading that are not significantly different from zero. This suggests that OLS estimates of the screened school advantage are driven in part by selection bias.

2SLS estimates also suggest that both screened and unscreened Grade A schools boost graduation rates somewhat, showing marginally significant gains ranging from 0.047 for screened schools to 0.025 for lottery schools. But estimated Grade A effects on college and career preparedness and college readiness in column 3 and 4 are consistently larger than estimated effects on other outcomes. This may in part reflect the fact that Grade A schools are more likely than other schools to offer the sort of advanced courses that go into the college and career-related composites. Across all outcomes, the 2SLS estimates for screened and lottery schools in columns 3 and 4 are similar.

Paralleling the estimates reported in the last column of Table 4, the estimates in columns 5 and 6 of Table 5 use lottery risk alone to identify Grade A effects. Lottery risk is again defined by $\lambda_s(\theta, L)$. In this case, however, $\lambda_s(\theta, L)$ is computed separately for offers of seats at screened and unscreened Grade A schools. Importantly, the lottery-risk analysis generates estimates of screened school effects for screened school applicants with tie-breaker values away from screened-school cutoffs. The experiment implicit in this scenario arises from disqualification at more preferred lottery schools for applicants to screened schools. We can also compare the statistical precision of estimation strategies that exploit lottery and general risk, for effects of enrollment at Grade A schools with different admissions regimes.

Perhaps surprisingly, lottery variation alone is sufficient to capture a reasonably precise screened school effect, with standard errors below 3 points for the estimated effects on SAT scores reported in column 5. Although the 2SLS estimates in this case are not significantly different from zero, they're not far from the corresponding general risk effects. It's also worth noting that standard errors below 3 are small enough to allow detection of SAT gains under one-tenth of a standard deviation (the standard deviation of an SAT score is around 100). It seems fair to say, therefore, that SAT effects identified using lottery risk alone are informative.

The graduation effects identified by lottery risk are small and not significantly different from zero for either type of school. But these estimates are not statistically distinguishable from the corresponding effects identified by general risk. Lottery-risk estimates of effects on college and career preparedness are larger than the corresponding estimates identified by general risk (compare, for example, 0.17 in column 5 to 0.09 in column 3). Again, however, differences by risk source for both composite outcomes are not statistically significantly different from zero.

Estimates that distinguish screened from lottery schools highlight the research value of screened school tie-breaking. For example, lottery risk alone generates an SAT math effect of lottery school attendance with a standard error around 0.89 (shown in column 6). This standard error falls to 0.76 (in column 4) when the lottery school effect is estimated using general risk, a precision gain equivalent to that yielded by increasing sample size by one-third. By contrast, the corresponding precision gain for estimates of screened school effects is dramatic:

a standard error of 2.85 using lottery risk falls to 1.24 using general risk, a gain that otherwise requires around five times as much data. Similar precision gains from the use of general risk to identify screened schools effects are seen for other outcomes in the table. Although lottery risk can be used to generate useful estimates of screened school effects, the most powerful tool for evaluation of NYC schools exploits all sources of risk.

5 Summary and Next Steps

The spread of centralized school matching opens new horizons for impact evaluation. The research potential of such markets is extended here by marrying the conditional random assignment generated by lottery tie-breaking with RD variation at screened schools. The propensity score for DA with mixed multiple tie-breakers combines all sources of quasi-experimental variation in such a match. Our analysis also shows how markets with general tie-breakers can be used to study treatment effects at screened schools for applicants with tie-breakers away from screened-school cutoffs. This addresses concerns, often raised in an RD context, that causal effects identified for applicants local to cutoffs need not be relevant for the general population.

Our analysis of NYC school report cards suggests Grade A schools generate some gains for their students, boosting Math SAT scores and graduation rates by a few points. OLS estimates, by contrast, show considerably larger effects of Grade A attendance. Grade A screened schools enroll some of the city's highest achievers, but large OLS achievement gains from attendance at Grade A screened schools appear to be an artifact of selection bias. Concerns about access to such schools (expressed, for example, in Harris and Fessenden (2017)) may therefore be overblown. On the other hand, Grade A attendance convincingly increases the district's composite indicators of college and college and career preparedness and college readiness. These results probably reflect the greater availability of the advanced courses that contribute to the composites in Grade A schools.

On the methodological side, the NYC analysis illustrates the possible precision gains yielded by research strategies that exploit general risk. Because different risk sources affect screened and lottery school access differently, the exclusion restriction in this context turns in part on a common effects assumption. It's therefore worth asking whether screened and lottery schools should indeed be treated as having the same effect. Our analysis supports the idea that lottery and screened Grade A schools can be pooled and treated as a homogeneous sector with a common average causal effect.

In on ongoing work, Angrist, Pathak, Rokkanen and Zarate (2017) deploy the methods developed here to study of Chicago's exam schools. Further afield, our theoretical framework may be applicable to an analysis of causal effects of medical residency assignments. The US National Residency Matching Program assigns medical school graduates to hospitals using a version of DA with non-lottery tie-breakers (Roth and Peranson, 1999). This match can be leveraged to answer questions about the effects of alternative medical internships, such as the value of experience in high-volume or university-affiliated hospitals. Our framework may also be useful to study the effects of resources allocated by some auction mechanisms.

Our provisional agenda for further research also includes an investigation of econometric implementation strategies, such as bandwidth selection. Zajonc (2012) and Papay *et al.* (2011) propose procedures for joint bandwidth selection in RD designs with multiple running variables. Multivariate procedures may have better properties than our one-size-fits-all approach. The relative statistical performance of 2SLS and semiparametric estimators likewise warrants investigation, as does the development of propensity score estimators that compute the score by simulation. Finally, inference on treatment effects in the application reported here relies on conventional large sample reasoning of the sort widely applied in empirical RD applications. It seems natural to consider permutation or randomization inference along the lines suggested by Canay and Kamat (2017) and Cattaneo *et al.* (2015, 2017), and optimal inference and estimation strategies such as those recently introduced by Armstrong and Kolesár (2018) and Imbens and Wager (2018).

References

ABDULKADIROĞLU, A., ANGRIST, J. and PATHAK, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, **82(1)**, 137–196.

—, ANGRIST, J. D., NARITA, Y., PATHAK, P. A., and ZARATE, R. A. (2017a). Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago's Exam Schools. *American Economic Review, Papers and Proceedings.*

—, —, — and — (2017b). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica*, **85(5)**, 1373–1432.

—, CHE, Y.-K. and YASUDA, Y. (2015). Expanding "Choice" in School Choice. *American Economic Journal: Microeconomics*, **7(1)**, 1–42.

—, PATHAK, P. A. and ROTH, A. E. (2005). The New York City High School Match. *American Economic Review, Papers and Proceedings*, **95**, 364–367.

—, — and ROTH, A. E. (2009). Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match. *American Economic Review*, **99(5)**, 1954–1978.

AJAYI, K. (2014). Does School Quality Improve Student Performance? New Evidence from Ghana. *Working Paper, Boston University.*

ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

ANGRIST, J. D., PATHAK, P. A., ROKKANEN, M. and ZARATE, R. (2017). Choice and Consequence: Understanding Mismatch at Chicago's Exam Schools. *Working Paper, MIT.*

ARMSTRONG, T. B. and KOLESÁR, M. (2018). Optimal Inference in a Class of Regression Models. *Econometrica*, **86** (2), 655–683.

AZEVEDO, E. and LESHNO, J. (2016). A Supply and Demand Framework for Two-Sided Matching Markets. *Journal of Political Economy*, **124(5)**, 1235–1268.

BARROW, L., SARTAIN, L. and DE LA TORRE, M. (2016). The Role of Selective High Schools in Equalizing Educational Outcomes: Heterogeneous Effects by Neighborhood Socioeconomic Status. *FRB of Chicago Working Paper No. 2016-17.*

BEUERMANN, D., JACKSON, C. K. and SIERRA, R. (2016). Privately Managed Public Secondary Schools and Academic Achievement in Trinidad and Tobago: Evidence from Rule-Based Student Assignments. *Working Paper.*

CANAY, I. A. and KAMAT, V. (2017). Approximate Permutation Tests and Induced Order Statistics in the Regression Discontinuity Design. *Review of Economic Studies*, **85** (3), 1577–1608.

Cattaneo, M. D., Frandsen, B. R. and Titiunik, R. (2015). Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the US Senate. *Journal of Causal Inference*, **3(1)**, 1–24.

—, Titiunik, R. and Vazquez-Bare, G. (2017). Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality. *Journal of Policy Analysis and Management*, **36(3)**, 643–681.

—, —, — and Keele, L. (2016). Interpreting Regression Discontinuity Designs with Multiple Cutoffs. *Journal of Politics*, **78(4)**, 1229–1248.

Chen, Y. and Kesten, O. (2017). Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy*, **125** (1), 99–139.

Disare, M. (2017). City to Eliminate High School Admissions Method that Favored Families with Time and Resources. *Chalkbeat, June 6*.

Dobbie, W. and Fryer, R. G. (2014). Exam High Schools and Academic Achievement: Evidence from New York City. *American Economic Journal: Applied Economics*, **6(3)**, 58–75.

Dur, U., Pathak, P., Song, F. and Sönmez, T. (2018). Deduction dilemmas: The taiwan assignment mechanism, nBER Working Paper 25024.

Ergin, H. and Sönmez, T. (2006). Games of School Choice under the Boston Mechanism. *Journal of Public Economics*, **90**, 215–237.

Fort, M., Ichino, A. and Zanella, G. (2016). Cognitive and Non-cognitive Costs of Daycare 0-2 for Girls. *Working Paper*.

Frandsen, B. R. (2017). Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable is Discrete. In *Regression Discontinuity Designs: Theory and Applications*, Emerald Publishing Limited, pp. 281–315.

Frölich, M. (2007). Regression Discontinuity Design with Covariates (Unpublished Appendix). *IZA Discussion Paper No. 3024*.

Hahn, J., Todd, P. and Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, **69(1)**, 201–209.

Harris, E. and Fessenden, F. (2017). The Broken Promises of Choice in New York City Schools. *New York Times, May 5*.

Hastings, J., Neilson, C. and Zimmerman, S. (2013). Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile. *NBER Working Paper 19241*.

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

— and Wager, S. (2018). Optimized Regression Discontinuity Designs. *Review of Economics and Statistics*.

Jackson, K. (2010). Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago. *Economic Journal*, **120(549)**, 1399–1429.

— (2012). Single-sex Schools, Student Achievement, and Course Selection: Evidence from Rule-based Student Assignments in Trinidad and Tobago. *Journal of Public Economics*, **96(1-2)**, 173–187.

Kirkeboen, L., Leuven, E. and Mogstad, M. (2016). Field of Study, Earnings, and Self-Selection. *Quarterly Journal of Economics*, **131**, 1057–1111.

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

Lucas, A. and Mbiti, I. (2014). Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya. *American Economic Journal: Applied Economics*, **6(3)**, 234–263.

Okui, R., Small, D. S., Tan, Z. and Robins, J. M. (2012). Double Robust Instrumental Variables Regression. *Statistica Sinica*, **22(1)**, 173–205.

Papay, J. P., Willett, J. B. and Murnane, R. J. (2011). Extending the Regression-Discontinuity Approach to Multiple Assignment Variables. *Journal of Econometrics*, **161(2)**, 203–207.

Pathak, P. A. and Sönmez, T. (2013). School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation. *American Economic Review*, **103(1)**, 80–106.

Pop-Eleches, C. and Urquiola, M. (2013). Going to a Better School: Effects and Behavioral Responses. *American Economic Review*, **103(4)**, 1289–1324.

Robins, J. (2000). Robust Estimation in Sequentially Ignoring Missing Data and Causal Inference Models. *Proceedings of the American Statistical Association Section on Bayesian Statistical Science*.

Rockoff, J. and Turner, L. (2011). Short Run Impacts of Accountability of School Quality. *American Economic Journal: Economic Policy*, **2(4)**, 119–147.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrica*, pp. 41–55.

Roth, A. E. and Peranson, E. (1999). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, **89**, 748–780.

Sekhon, J. S. and Titiunik, R. (2017). On Interpreting the Regression Discontinuity Design as a Local Experiment. In *Regression Discontinuity Designs: Theory and Applications*, Emerald Publishing Limited, pp. 1–28.

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

van Handel, R. (2016). APC 550: Probability in High Dimension, Lecture Notes. *Princeton University. Retrieved from https://web.math.princeton.edu/ětexttildelow rvan/APC550.pdf on December 21, 2016.*

Walcott, D. (January 2012). NYC Department of Education: Progress Reports for New York City Public Schools.

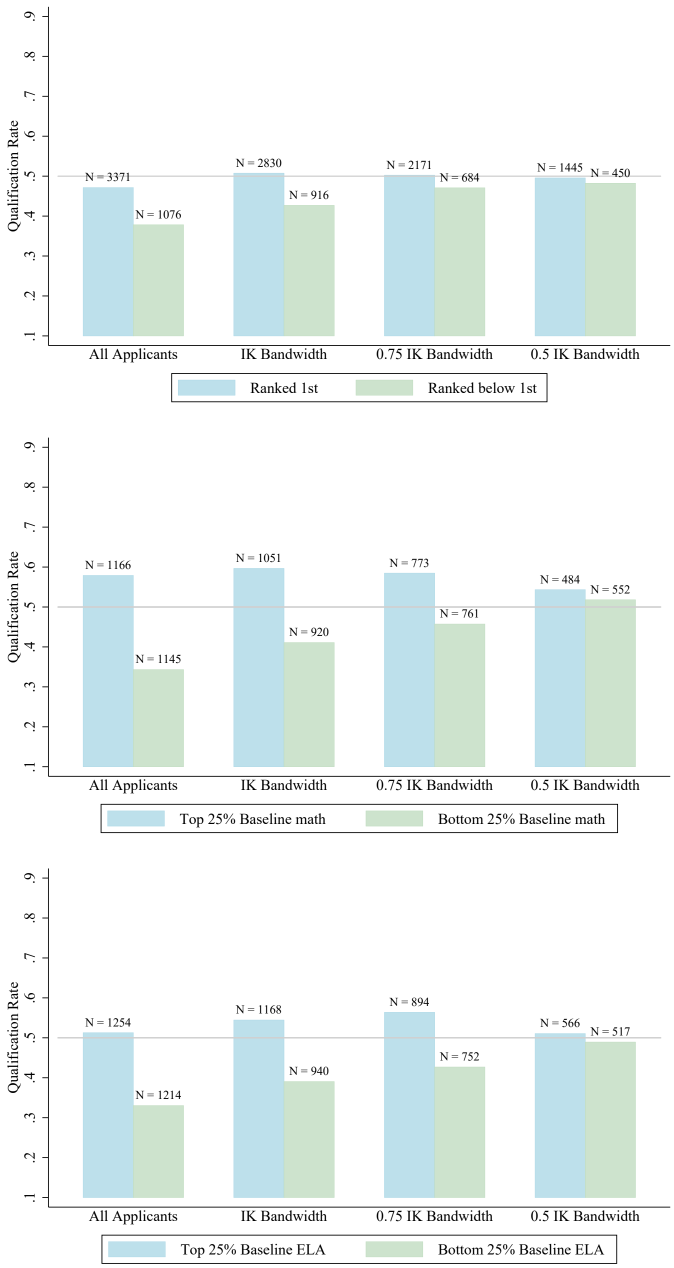
Wellner, J. A. (1981). A Glivenko-Cantelli Theorem for Empirical Measures of Independent but Non-Identically Distributed Random Variables. *Stochastic Processes and Their Applications*, **11(3)**, 309–312.

Wong, V. C., Steiner, P. M. and Cook, T. D. (2013). Analyzing Regression-Discontinuity Designs with Multiple Assignment Variables: A Comparative Study of Four Estimation Methods. *Journal of Educational and Behavioral Statistics*, **38** (2), 107–141.

Zajonc, T. (2012). Regression Discontinuity Design with Multiple Forcing Variables. *Essays on Causal Inference for Public Policy*, pp. 45–81.

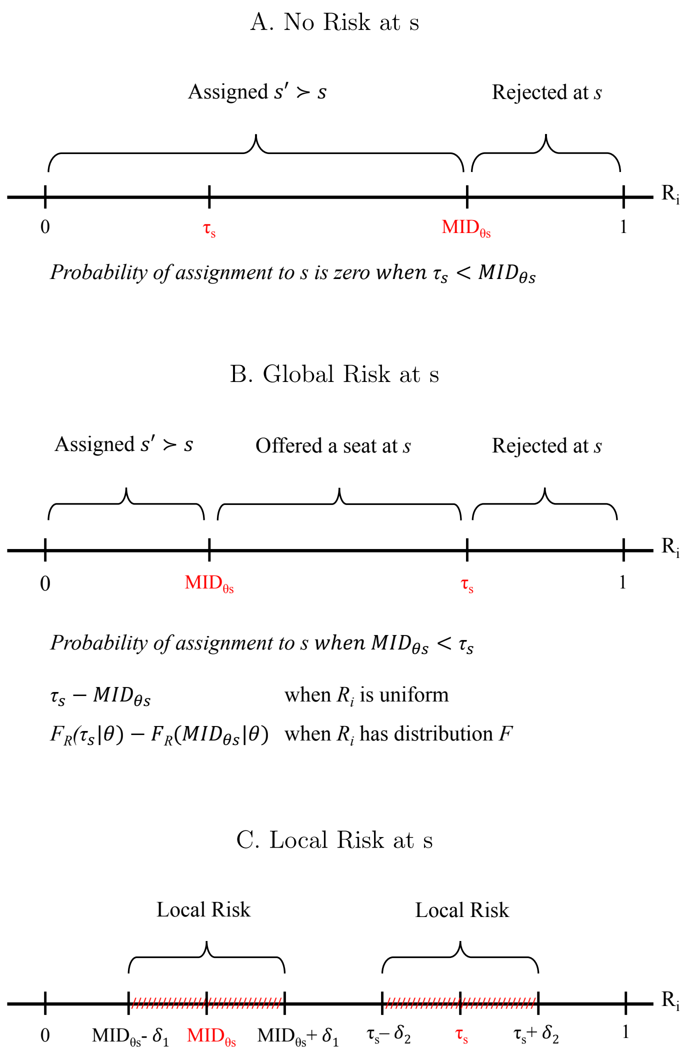
Zimmerman, S. D. (forthcoming). Elite Colleges and Upward Mobility to Top Jobs and Top Incomes. *American Economic Review.*

Figure 1: Qualification Rates Near the TH Cutoff



Notes: Bars show pooled qualification rates for applicants to Townsend Harris. The first panel shows qualification rates separately for applicants ranking the program as first choice and for applicants ranking it second or lower. The second and third panel show qualification rates separately for groups of applicants with baseline math and ELA scores in the upper and lower quartiles of the applicant score distribution. Qualification is defined as clearing the relevant program cutoff. The figure aggregates data for the cohorts 2011/12, 2012/13 and 2013/14. Bandwidths are 0.055 and 0.388, respectively, for a running variable between 0-1. Baseline scores are from 7th grade.

Figure 2: Visualizing Risk under Serial Dictatorship



Notes: This figure illustrates risk under serial dictatorship. R_i is the tie-breaker. $MID_{\theta s}$ is the most forgiving cutoff at schools preferred to s and τ_s is the cutoff at s.

Figure 3: 2011/12 Progress Report for East Side Community School

NYC Department of Education
Dennis M. Walcott, Chancellor

Progress Report 2011-12

East Side Community School

PRINCIPAL:	Mark Federman
DBN:	01M450
ENROLLMENT:	357
SCHOOL TYPE:	High School
PEER INDEX*:	2.38

*See p. 7 for more details on Peer Index.

PROGRESS REPORT

A — OVERALL SCORE **73.6** out of 100 — OVERALL PERCENTILE **75**

This school's overall score is greater than or equal to that of 75 percent of high schools.

For high schools, grades are based on cut scores determined prior to the release of the Progress Report. Further, schools with a four year graduation rate in the top third citywide cannot receive a grade lower than a C. Schools in their first year, without a graduating class or in phase out receive a report with no grade or score.

Progress Report Grades - High School

GRADE	SCORE RANGE	% OF SCHOOLS
A	70.0 or higher	35% of schools
B	58.0 - 69.9	37% of schools
C	47.0 - 57.9	20% of schools
D	40.0 - 46.9	5% of schools
F	39.9 or lower	3% of schools

QUALITY REVIEW

WD

Well Developed (2007-08)

The rating is based on three major categories of school performance: instruction that prepares students for college and careers, school organization and management, and quality of the learning environment.

A school that receives a Well Developed rating earned the highest grade for highly effective teaching and learning practice, strategic school management, and an excellent quality learning environment. For more information, see:

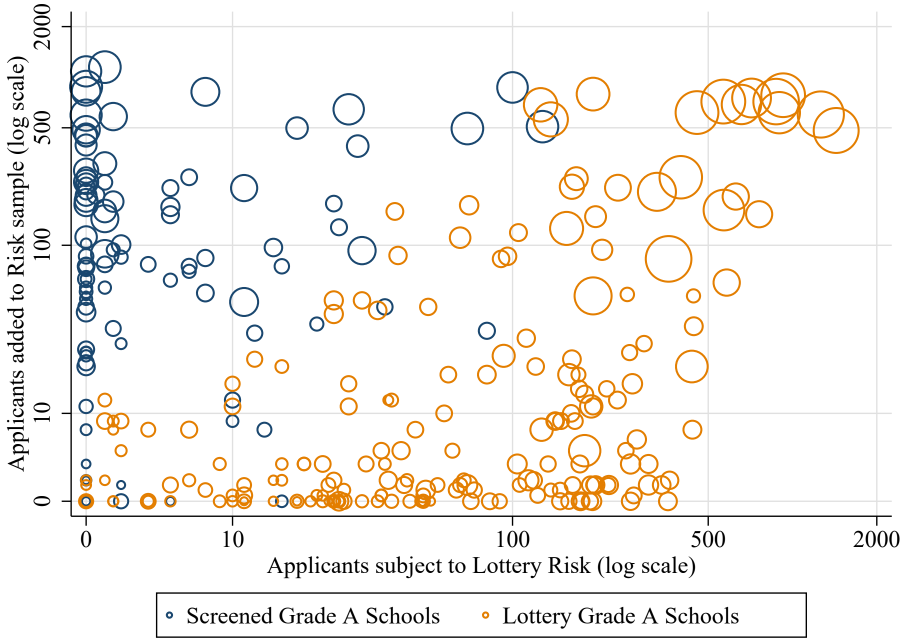
http://schools.nyc.gov/Accountability/tools/review

Overview Each school's Progress Report (1) measures student year-to-year progress, (2) compares the school to peer schools, and (3) rewards success in moving all children forward, especially children with the greatest needs.

CATEGORY	SCORE	GRADE	DESCRIPTION
Student Progress	**33.0** out of 55	B	Student Progress measures the annual progress students make toward meeting the state's graduation requirements by earning course credits and passing state Regents exams.
Student Performance	**14.4** out of 20	A	Student Performance measures how many students graduated within 4 and 6 years of starting high school, and the types of diplomas they earned.
School Environment	**10.9** out of 15	A	School Environment measures student attendance and a survey of the school community rating academic expectations, safety and respect, communication, and engagement.
College and Career Readiness	**7.3** out of 10	A	College and Career Readiness measures how well students are prepared for life after high school on the basis of passing advanced courses, meeting English and math standards, and enrolling in a post-secondary institution.
Closing the Achievement Gap	**8.0** (16 max)		Schools receive additional credit for exceptional graduation and college/career readiness outcomes of students with disabilities, English Language Learners, and students who enter high school at a low performance level.
Overall Score	**73.6** out of 100	A	The overall grade is based on the total of all scores above. Category scores may not add up to total score because of rounding.

Notes: This figure shows the 2011/12 progress report for East Side Community School. Source: *www.crpe.org*

Figure 4: Sample size gains relative to lottery risk



Notes: This figure plots increases in the number of applicants with non-degenerate risk of assignment at individual schools, ordered by the number of applicants who have risk when screened school admission is treated as determined solely by priorities. The number of applicants added measures the number of additional students at risk when risk is determined by running variable variation in a bandwidth around screened school cutoffs as well as by lottery risk. Circle sizes plot school capacity. Declines in risk are not shown.

Table 1. New York High School Characteristics

	Grade A schools			Grade B-F	Ungraded
	All	Screened	Lottery	Schools	Schools
	(1)	(2)	(3)	(4)	(5)
Panel A. Average Performance Levels					
SAT Math (200-800)	531	606	481	464	440
SAT Reading (200-800)	522	587	479	465	449
Graduation	0.75	0.89	0.68	0.59	0.38
College- and career-prepared	0.65	0.84	0.54	0.39	0.27
College-ready	0.59	0.82	0.45	0.34	0.24
Panel B. School Characteristics					
White or Asian students	0.43	0.59	0.32	0.26	0.15
Special Education	0.12	0.06	0.16	0.17	0.27
Free or Reduced Price Lunch	0.68	0.55	0.76	0.77	0.75
In Manhattan	0.27	0.49	0.12	0.16	0.28
Number of grade 9 students	420	430	414	413	86
Number of grade 12 students	374	413	348	351	53
High school size	1596	1700	1527	1509	426
Inexperienced teachers	0.11	0.10	0.12	0.11	0.28
Advanced degree teachers	0.53	0.59	0.49	0.50	0.30
New school	0.00	0.00	0.01	0.00	0.21
School-year observations	355	119	236	694	715

Notes. This table reports weighted average characteristics of school-year observations. Specialized and charter high schools admit applicants in a separate match and are considered screened and lottery schools, respectively. Panel A reports outcomes for cohorts enrolled in ninth grade in 2012-13, 2013-14 and 2014-15, and Panel B school characteristics in 2012-13, 2013-14 and 2014-15 by type of school. A screened school is any school without lottery programs. Graduation outcomes condition on ninth grade enrollment in the year following the match and are available for the first and second cohort only. Inexperienced teachers have 3 or fewer years of experience and advanced degree teachers a Masters or higher degree.

Table 2. Student Characteristics

	Ninth grade students		Eighth grade applicants in main match				
	All	Any Grade A	All	Listed Match A	Enrolled in Match A	At Risk at Match A	
						General	Lottery
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Demographics							
Black	30.7	20.1	29.1	29.3	22.9	22.5	26.9
Hispanic	40.2	33.7	38.9	39.3	38.2	40.1	48.5
Female	49.2	53.1	51.5	52.5	54.0	51.3	49.2
Special education	19.0	5.6	7.6	7.3	6.4	6.0	8.4
English language learners	7.5	4.4	6.0	5.7	5.2	5.0	7.2
Free lunch	78.6	70	77.3	77.2	73.6	75.6	80.2
Baseline scores							
Math (standardized)	0.056	0.528	0.207	0.233	0.334	0.333	0.033
English (standardized)	0.022	0.466	0.168	0.196	0.288	0.274	0.003
Offer rates							
Grade A school		81.8	29.4	34.6	87.2	47.2	42.6
Grade A screened school		28.5	9.9	11.7	26.5	12.9	1.8
Grade A lottery school		53.3	19.5	22.9	60.6	34.3	40.8
Listed Grade A first		82.6	47.3	55.6	84.4	78.1	77.3
9th grade enrollment							
Grade A school	30.9	100	32.7	37.6	100	49.7	43.2
Grade A screened school	11.6	39.8	13.2	15.0	28.4	16.6	3.3
Grade A lottery school	19.4	60.5	19.7	22.8	71.9	33.4	40.0
Students	182,249	48,985	153,107	130,160	40,301	30,760	18,814
Schools	603	174	569	567	159	532	512
School-year observations	1672	352	1584	1562	319	1420	1354

Notes. This table describes the population of NYC students. Column 1 and 2 show statistics for students enrolled in ninth grade in the 2012-13, 2013-14 and 2014-15 school years with non-missing demographics and baseline test scores. Columns 3 to 7 show statistics for ninth grade applicants, who participated in the NYC high school match one year earlier. A match A school is a grade A school that participates in the main NYC high school match. Students are said to have risk when they have a propensity score strictly between zero and one and they're in a score cell with variation in Grade A school offers. Baseline scores are from sixth grade and demographics from eighth grade.

Table 3. Statistical Tests for Balance

	All Applicants		Grade A Applicants at Risk		
	Non-offered mean		Non-offered mean	General risk	Lottery risk
	(1)	(2)	(3)	(4)	(5)
		Panel A. Application Covariates			
Grade A listed first	0.393	0.483***	0.761	0.009*	0.015**
		(0.002)		(0.005)	(0.006)
# of screened Grade A schools listed	1.11	0.534***	1.20	0.005	-0.032
		(0.011)		(0.006)	(0.022)
# of lottery Grade A schools listed	1.69	0.227***	2.10	-0.014	-0.023
		(0.008)		(0.017)	(0.024)
		Panel B. Baseline Covariates			
Black	0.339	-0.131***	0.230	-0.002	-0.000
		(0.003)		(0.006)	(0.008)
Hispanic	0.405	-0.055***	0.402	0.004	0.002
		(0.003)		(0.007)	(0.009)
Female	0.527	0.003	0.517	0.003	-0.012
		(0.003)		(0.008)	(0.009)
Special education	0.078	-0.018***	0.063	-0.005	-0.007
		(0.001)		(0.004)	(0.005)
English language learners	0.075	-0.017***	0.064	0.003	0.001
		(0.001)		(0.004)	(0.005)
Free lunch	0.846	-0.091***	0.818	-0.006	-0.006
		(0.002)		(0.006)	(0.007)
Baseline scores					
Math (standardized)	0.110	0.379***	0.285	0.005	-0.011
		(0.005)		(0.010)	(0.013)
English (standardized)	0.081	0.349***	0.220	0.007	0.012
		(0.006)		(0.012)	(0.013)
N		129,720		30,673	18,743

Notes. This table reports balance statistics, computed by regressing covariates on dummies indicating a Grade A offer and an ungraded school offer, controlling for saturated Grade A and ungraded school propensity scores (columns (4) and (5)), and running variable controls (column (4)). Lottery risk in column (5) is computed by treating screened-school tie-breakers as priorities. The sample is limited to applicants with non-missing demographics and baseline test scores. Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4. Grade A School 2SLS Estimates

	All applicants		Applicants at risk		
				2SLS	
	Non-Enrolled Mean	OLS	Non-Offered Mean	General risk	Lottery risk
	(1)	(2)	(3)	(4)	(5)
			Panel A. First Stage Estimates		
SAT outcomes			0.448	1.79***	1.87***
(years of exposure)				(0.023)	(0.027)
Binary outcomes			0.178	0.636***	0.661***
(ever enrolled)				(0.008)	(0.010)
			Panel B. Second Stage Estimates		
SAT Math	474	6.48***	515	2.26**	2.17**
(200-800)	(106)	(0.152)	(109)	(0.718)	(0.880)
SAT Reading	473	5.42***	510	0.726	0.770
(200-800)	(93)	(0.138)	(93)	(0.658)	(0.801)
N		124,989		22,899	12,752
Graduated	0.697	0.037***	0.793	0.030**	0.022
		(0.003)		(0.013)	(0.018)
College- and Career-prepared	0.422	0.104***	0.587	0.096***	0.137***
		(0.003)		(0.015)	(0.020)
College-ready	0.367	0.075***	0.542	0.056***	0.053**
		(0.003)		(0.014)	(0.018)
N		121,074		19,150	11,200

Notes. This table reports estimates of the effects of Grade A high school enrollment. 2SLS estimates are from models with dummies for Grade A and ungraded schools treated as endogenous, limiting the sample to students with Grade A assignment risk. OLS estimates are from models that omit propensity score controls and include all students in the three match cohorts. All models include controls for baseline math and English scores, free lunch status, SPED and ELL status, gender, and race/ethnicity indicators. Estimates in column 4 are from models that include running variable controls. Non-offered means are for the general risk sample. Robust standard errors are in parenthesis for estimates and standard deviations for non-offered means. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5. Multi-Sector Grade A 2SLS Estimates

		OLS		Applicants at risk			
				General risk		Lottery risk	
		Screened Grade A	Lottery Grade A	Screened Grade A	Lottery Grade A	Screened Grade A	Lottery Grade A
		(1)	(2)	(3)	(4)	(5)	(6)
SAT Math		15.6***	1.27***	2.98**	2.42**	2.01	2.12**
(200-800)		(0.228)	(0.168)	(1.24)	(0.759)	(2.85)	(0.887)
	p-value			0.662		0.970	
SAT Reading		13.2***	0.979***	1.61	0.653	-0.810	0.702
(200-800)		(0.209)	(0.153)	(1.14)	(0.698)	(2.68)	(0.807)
	p-value			0.420		0.566	
	N	124,989		24,899		13,062	
Graduated		0.040***	0.036***	0.047**	0.025*	-0.017	0.025
		(0.003)	(0.003)	(0.018)	(0.015)	(0.056)	(0.019)
	p-value			0.266		0.461	
College- and Career-prepared		0.143***	0.087***	0.085***	0.102***	0.166**	0.138***
		(0.004)	(0.003)	(0.021)	(0.016)	(0.070)	(0.020)
	p-value			0.458		0.688	
College-ready		0.143***	0.044***	0.105***	0.042**	0.019	0.051**
		(0.004)	(0.003)	(0.021)	(0.015)	(0.067)	(0.019)
	p-value			0.005		0.638	
	N	121,074		20,542		11,448	

Notes. This table reports 2SLS second stage estimates of models that separately identify Grade A effects at screened and lottery schools, treating both as well as ungraded schools as endogenous. The sample is limited to students with either Grade A lottery or Grade A screened assignment risk. OLS models omit propensity score controls and include all students in the three match cohorts. All models include baseline covariate controls. (3) and (4) include running variable controls. Robust standard errors in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

A Illustrating Lottery and Screening Risk

Table A1 illustrates Theorem 1 for three applicants, highlighting the distinction between lottery and general risk, that is, the risk created by the combination of lottery and screened school tie-breaking. The first example shows how lotteries create risk at both lottery and screened schools. Example 1 concerns an applicant who lists three schools, only the first of which is a lottery school. At first choice school A, 80% of those in the marginal priority group are admitted, so $\tau_A = 0.8$ (lottery school cutoffs are shown in the first column). School A is listed first, so this applicant's $MID^0_{\theta A} = 0$ (reported in column 3), and the propensity score characterizing lottery risk (reported in column 4) is also 0.8. There is no other source of risk at A for this applicant.

At second choice school B, $MID^0_{\theta B} = 0.8$ because $\tau_A = 0.8$. But this applicant faces no risk at B because her priority places her among the never seated, that is, $t_{iB}(\delta) = n$ for small small enough δ (type classifications for lottery risk appear in column 2). Finally, at third choice school C, this applicant is always seated, that is, $t_{iC}(\delta) = a$. Since the applicant's $MID^0_{\theta C}$ is determined by the cutoff at A, the probability of being assigned C is $1 - MID^0_{\theta C} = 0.2$. Here too, lottery risk captures all the action. This example shows how lotteries create screened school risk even for applicants with screened running variable values far from cutoffs. This presents an interesting contrast with the RD scenarios considered by Hahn *et al.* (2001) and related work, where RD methods identify treatment effects local to cutoffs.

The second example shows how non-lottery tie-breaking creates risk for applicants who face no lottery risk. At first choice school D, a screened school, the applicant is in the marginal priority group, and near the school D cutoff, that is $t_{iD}(\delta) = c$ for i such that $R_i \in [\tau_D - \delta, \tau_D + \delta]$, which generates a propensity score of 0.5 (type classification for the general risk scenario appears in column 5). In this case, however, the applicant faces no lottery risk because in the lottery-only scenario, he's in Θ^n_D (seen in column 2) with a running variable value assumed to be above the cutoff. Consequently, the lottery risk propensity score for assignment to school D is zero. This applicant's second choice, school E, is a lottery school, but the applicant is never seated there since his priority is too low, placing him in Θ^n_E. The school E propensity score is therefore zero no matter how risk is calculated. The applicant's third choice, school F, is also a lottery school, where the applicant has high enough priority to be in Θ^a_E. The applicant's lottery risk at this school therefore reflects his certainty of finding a seat at F.

When the risk generated by screening at school D is taken into account, we see that the second applicant's $m_F(\theta, T) = |\{D\}| = 1$. Note also that $MID^0_{\theta F} = 0$ because the set of more preferred schools contains one screened school that uses a tie-breaker other than $v(s)$ and one lottery school, at which the applicant was not competitive. As can be seen in column 7, the propensity score for general risk of assignment to F is therefore:

$$\lambda_F(\theta, T) = 0.5^{m_F(\theta,T)}(1 - MID^0_{\theta F}) = 0.5 \times (1 - 0) = 0.5.$$

Importantly, assignment risk at schools D and F emerge from screening in spite of the fact that lottery risk is degenerate at each of this applicant's three choices.

The third example shows how screening and lottery risk interact. First choice school G is a screened school at which Applicant 3 is in Θ^c_G (in the classification scheme for general risk).

Allowing for screened school risk, the school G propensity score is therefore 0.5. To isolate lottery risk at G, we've placed the applicant in Θ^a_G, assuming his running variable at this school is below the school G cutoff (as can be seen in column 2). When screening variables are treated as priorities, this applicant is surely seated at G, creating a propensity score for lottery risk at school G equal to 1.

Second choice school H is a lottery school. Looking only at lottery risk, the applicant's $MID^0_{\theta H} = 1$, since he's always seated at his first choice. The applicant's lottery propensity score is therefore zero at H and all lower-listed choices. The propensity score for overall risk at H is also zero because we've assumed the applicant fails to clear marginal priority at this school. By contrast, this applicant is in the marginal priority group at his third choice, lottery school I. The school I lottery cutoff is 0.6. At school I, we have $m_I(\theta, T) = |\{G\}| = 1$, and $MID^0_{\theta I} = 0$. The propensity score for assignment to I is therefore:

$$\lambda_I(\theta, T) = 0.5^{m_I(\theta,T)}(\tau_I - MID^0_{\theta I}) = 0.5 \times (0.6 - 0) = 0.3.$$

Fourth choice school J is a lottery school with cutoff $\tau_J = 0.8$. Note that $MID^0_{\theta J} = 0.6$ because $\tau_I = 0.6$ and because I, a lottery school where the applicant is conditionally seated, is listed ahead of J. Screening risk at G also implies $m_J(\theta, T) = |\{G\}| = 1$. The propensity score for general risk is therefore:

$$\lambda_J(\theta, T) = 0.5^{m_J(\theta,T)}(\tau_J - MID^0_{\theta J}) = 0.5 \times (0.8 - 0.6) = 0.1.$$

Finally, last-listed school K is screened. The applicant is assumed to clear marginal priority at K, so any risk there must be generated by rejection at higher listed choices. Note that $MID^0_{\theta K} = 0.8$ because $\tau_J = 0.8$ is the most forgiving cutoff at lottery schools listed ahead of K. Since $m_K(\theta, T) = |\{G\}| = 1$, the propensity score for general risk at K is therefore

$$\lambda_K(\theta, T) = 0.5^{m_K(\theta,T)}(1 - MID^0_{\theta K}) = 0.5 \times (1 - 0.8) = 0.1.$$

This example shows how the interaction between screened and lottery risk takes an applicant with no lottery risk and exposes him to risk at four out of the five schools he's listed.[24]

[24]The examples in Table A1 have the feature that risk sums to one, because everyone is seated somewhere. This need not be the case in real markets, where some applicants, typically those who list few schools or list only schools with very limited capacity, may remain unassigned.

Table A1. Propensity Score Anatomy

Schools			Determinants of lottery risk			Determinants of general risk		
Name	Tie-breaker	τ	Θ	MID^0	Propensity score	Θ	MID^0	Propensity score
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Applicant 1: Lotteries Create Risk at Lottery and Screened Schools								
A	Lottery	0.8	*c*	0	0.8	*c*	0	0.8
B	Screened		*n*	0.8	0	*n*	0.8	0
C	Screened		*a*	0.8	0.2	*a*	0.8	0.2
Applicant 2: Screening Creates Risk at Lottery and Screened Schools								
D	Screened		*n*	0	0	*c*	0	0.5
E	Lottery	0.8	*n*	0	0	*n*	0	0
F	Lottery	0.6	*a*	0	1	*a*	0	0.5
Applicant 3: Lotteries and Screening Interact to Create Risk								
G	Screened		*a*	0	1	*c*	0	0.5
H	Lottery	0.8	*n*	1	0	*n*	0	0
I	Lottery	0.6	*c*	1	0	*c*	0	0.3
J	Lottery	0.8	*c*	1	0	*c*	0.6	0.1
K	Screened		*a*	1	0	*a*	0.8	0.1

Notes. This table shows how the propensity score is determined for three applicants. Each example describes the risk faced by a single applicant at each of the schools ranked: applicant 1 ranks schools A, B, and C in that order. Lottery risk is determined by taking screened school tie-breakers as priorities. Applicants to screened schools therefore have status a or n at these schools according to whether they clear the relevant screened tie-breaker cutoff. Column 1 reports randomization cutoffs for schools that use a lottery tie-breaker. Columns 2 and 5 reports applicant status (always seated [a], never seated [n], and conditionally seated [c]) in lottery-only and general risk scenarios. Columns 3 and 6 report the most informative disqualification for the lottery tie-breaker. MID^0 in column 3 is computed treating screened-school tie-breakers as priorities. MID^0 in column 6 looks at screened as well as lottery risk.

B Running Variables Coded as Ranks

An empirical rank transformation of independent running variables can be dependent. This section shows that running variables transformed into ranks become independent as the number of students grows to infinity. The assumption of independent running variables therefore holds for a continuum market as long as raw running variables are independent. Let $(X_1, X_2, \ldots)$ be a sequence of independent random variables. Define the rank function as follows.

$$\text{rank}_N(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{X_i < t\}.$$

Let

$$F_N(t) = \frac{1}{N}\sum_{i=1}^{N} P(X_i < t).$$

Proposition 5. *For all k, we have*

$$\left|\text{rank}_N(X_k) - F_N(X_k)\right| \to 0 \quad \text{a.s.}$$

Thus the process $(\text{rank}_N(X_k) : k \in \mathbb{N})$ converges to the independent sequence $(F_N(X_k) : k \in \mathbb{N})$ uniformly in k on a set of measure 1.

Proof of Proposition 5. We prove Proposition 5 using a few lemmas below.

Lemma 1 (Hoeffding's maximal inequality; Lemma 5.1 in van Handel (2016))**.** *Let A be a finite subset of $\mathbb{R}^N$ and write $\|A\|_2 = \sup_{a\in A}\|a\|_2$, where $||\cdot||_2$ is the square root of the sum of squares ($||x||_2 \equiv \sqrt{x_1^2 + \ldots + x_m^2}$ for any vector $x \equiv (x_1, \ldots, x_m)$). Let $X_1, \ldots, X_N$ be independent, centered (mean zero) random variables supported on $[-1, 1]$. Then we have*

$$E\sup_{a\in A}\left\{\sum_{i=1}^{N} a_i X_i\right\} \leq \|A\|_2 \sqrt{2\log|A|},$$

where $|A|$ is the cardinality of set A.

Lemma 2. *The expected supremum of the rank process satisfies*

$$E\ \sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{1}\{X_i < t\} - P(X_i < t)\right)\right|\right\} \leq \sqrt{\frac{8\log(n+1)}{n}}.$$

Proof of Lemma 2. Let $f_t(s) = \mathbb{1}\{s < t\}$. For each k, construct an independent random variable Y_k with the same distribution as X_k. Note that (i) $E[f_t(Y_k)] = P(X_k < t)$ and (ii) the law of $f_t(X_k) - f_t(Y_k)$ is symmetric around 0. By Jensen's inequality, we have

$$E\ \sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\left(f_t(X_i) - P(X_i < t)\right)\right|\right\} \leq E\ \sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\left(f_t(X_i) - f_t(Y_i)\right)\right|\right\}.$$

By symmetry and independence of the summands, their joint distribution does not change if we multiply each by an independent random variable ε_k that is uniformly distributed on $\{\pm 1\}$. This gives us

$$\begin{aligned} E \sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} (f_t(X_i) - f_t(Y_i)) \right| \right\} &= E \sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i \big(f_t(X_i) - f_t(Y_i) \big) \right| \right\} \\ &\leq E \left[\sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i f_t(X_i) \right| \right\} + \sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} -\varepsilon_i f_t(Y_i) \right| \right\} \right] \\ &= 2 \cdot E \sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i f_t(X_i) \right| \right\} \end{aligned}$$

We then have

$$\begin{aligned} 2 \cdot E \sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i f_t(X_i) \right| \right\} &= 2 \cdot E \left[E \left[\sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i f_t(X_i) \right| \right\} \middle| X_1, \ldots, X_N \right] \right] \\ &= 2 \cdot E \left[E \left[\sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i f_t(X_{(i)}) \right| \right\} \middle| X_1, \ldots, X_N \right] \right]. \end{aligned}$$

Here $X_{(i)}$ refers to the i^{th} smallest element of $\{X_1, \ldots, X_N\}$; we use the fact that the inner sum is invariant to re-ordering. As $t \in \mathbb{R}$ varies, the vector

$$u_t(X_1, \ldots, X_N) = \big(\tfrac{1}{N} f_t(X_{(1)}), \ldots, \tfrac{1}{N} f_t(X_{(N)}) \big)$$

takes on at most $n+1$ values, and we always have $||u_t||_2 \leq 1/\sqrt{N}$. This follows from observing that nu_t takes values in the set of increasing binary sequences of length N. Applying Lemma 1 to the inner expectation then gives

$$2 \cdot E \left[E \left[\sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i f_t(X_{(i)}) \right| \right\} \middle| X_1, \ldots, X_N \right] \right] \leq \sqrt{\frac{8 \log(N+1)}{N}}.$$

□

Lemma 3. *Write*

$$h(X_1, \ldots, X_N) = \sup_{t\in\mathbb{R}} \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \left(\mathbb{1}\{X_i < t\} - P(X_i < t) \right) \right| \right\}.$$

Then $P\big(|h(X_1, \ldots, X_N) - E[h(X_1, \ldots, X_N)]| > \delta \big) \leq e^{-2N\delta^2}$.

Proof of Lemma 3. Note that varying $X_i(\omega)$ can change $h(\omega)$ by at most $1/n$, for all ω in the sample space. Lemma 3 then follows from McDiarmid's inequality as stated in van Handel (2016) (Theorem 3.11). □

Putting together Lemmas 2 and 3, we obtain

$$P\left(\sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}(\mathbb{1}\{X_i<t\}-P(X_i<t))\right|\right\}\geq\sqrt{\frac{8\log(N+1)}{N}}+\delta\right)$$

$$\leq P\left(\sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}(\mathbb{1}\{X_i<t\}-P(X_i<t))\right|\right\}\geq E\big[\sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}(\mathbb{1}\{X_i<t\}-P(X_i<t))\right|\right\}\big]+\delta\right)$$

$$\leq e^{-2N\delta^2}$$

Since the sequence on the right-hand side is summable for any fixed δ, we apply the Borel-Cantelli lemma to obtain

$$P\left(\limsup_{N\to\infty}\sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}(\mathbb{1}\{X_i<t\}-P(X_i<t))\right|\right\}\geq\delta\right)=0.$$

Taking $\delta\to 0$ gives

$$\limsup_{N\to\infty}\sup_{t\in\mathbb{R}}\left\{\left|\frac{1}{N}\sum_{i=1}^{N}(\mathbb{1}\{X_i<t\}-P(X_i<t))\right|\right\}=0\quad\text{a.s.}$$

This implies that

$$\left|\text{rank}_N(X_k)-F_N(X_k)\right|\leq\sup_{t\in\mathbb{R}}\left\{\left|\text{rank}_N(t)-F_N(t)\right|\right\}\downarrow 0\quad\text{a.s.},$$

completing the proof of Proposition 5. □

C Proofs

C.1 Proposition 1

When $t=n$, $R_i>\tau_A+\delta$ for any small enough $\delta>0$, so that $1(R_i\leq\tau_A)=0$. This implies $\lim_{\delta\to 0}E[1(R_i\leq\tau_A)|\theta_i=\theta,t_{iA}(\delta)=t,W_i=w]=0$. When $t=a$, $R_i<\tau_A-\delta$ for any small enough $\delta>0$, so that $1(R_i\leq\tau_A)=1$. This implies $\lim_{\delta\to 0}E[1(R_i\leq\tau_A)|\theta_i=\theta,t_{iA}(\delta)=t,W_i=w]=1$. Finally, suppose $t=c$ and recall that $F_R(\tau_A|\theta,w)$ is differentiable for every θ and w and that $F_R'(\tau_A|\theta,w)\neq 0$ by Assumption 1(a). We then have:

$$\begin{aligned}
&\lim_{\delta\to 0}E[1(R_i\leq\tau_A)|\theta_i=\theta,t_{iA}(\delta)=c,W_i=w]\\
&\qquad=\lim_{\delta\to 0}\frac{P(\tau_A-\delta\leq R_i\leq\tau_A|\theta_i=\theta,W_i=w)}{P(\tau_A-\delta\leq R_i\leq\tau_A+\delta|\theta_i=\theta,W_i=w)}\\
&\qquad=\lim_{\delta\to 0}\frac{F_R(\tau_A|\theta,w)-F_R(\tau_A-\delta|\theta,w)}{F_R(\tau_A+\delta|\theta,w)-F_R(\tau_A-\delta|\theta,w)}\\
&\qquad=\lim_{\delta\to 0}\frac{\{F_R(\tau_A|\theta,w)-F_R(\tau_A-\delta|\theta,w)\}/\delta}{\{F_R(\tau_A+\delta|\theta,w)-F_R(\tau_A|\theta,w)\}/\delta+\{F_R(\tau_A|\theta,w)-F_R(\tau_A-\delta|\theta,w)\}/\delta}\\
&\qquad=\frac{F_R'(\tau_A|\theta,w)}{2F_R'(\tau_A|\theta,w)}=0.5,
\end{aligned}$$

where the last equality uses $F_R'(\tau_A|\theta,w)\neq 0$. The second last equality holds because the limit of a fraction of functions is the same as the fraction of the limits of the functions as long as the denominator converges to a nonzero limit. This completes the proof.

C.2 Proposition 4

We prove Proposition 4 using a strategy similar to that used to prove Theorem 1 in Abdulkadiroğlu *et al.* (2017b). Note first that admissions cutoffs ξ in a continuum market are invariant to tie-breaking outcomes R_i, and bandwidth δ: DA in the continuum depends on R_i's only through $G(I_0)$, the fraction of applicants in set $I_0 = \{i \in I \mid \theta_i \in \Theta_0, r_{iv} \leq r_v \text{ for all } v\}$ with various choices of Θ_0 and r. In particular, $G(I_0)$ doesn't depend on running variable realizations in the continuum market since for the empirical CDF of each running variable conditional on each type, $\hat{F}_v(\cdot|\theta)$, we always have $\hat{F}_v(\cdot|\theta) = F_v(\cdot|\theta)$ for any v and θ by the Glivenko-Cantelli theorem for independent but non-identically distributed random variables (Wellner, 1981). $G(I_0)$ doesn't depend on reference tie-breaking number r and bandwidth δ either since they affect only the distribution of a single student i's tie-breaking number R_i, which has no effect on $G(I_0)$ or cutoffs. As a consequence of the constancy of cutoffs ξ, marginal priority ρ_s is also constant for every school s.

Now, consider the propensity score for school s. Applicants who don't rank s have $p_s(\theta) = 0$. Among those who do rank s, those of type $\theta \in \Theta_s^n$ have $\rho_{\theta s} > \rho_s$. Therefore $p_s(\theta) = 0$ for every $\theta \in \Theta_s^n \cup (\Theta \backslash \Theta_s)$.

Applicants of type $\theta \in \Theta_s^a \cup \Theta_s^c$ may be assigned $\tilde{s} \in B_{\theta s}$, where $\rho_{\theta\tilde{s}} = \rho_{\tilde{s}}$. For each v, the proportion of type θ applicants assigned some $\tilde{s} \in B_{\theta s}^v$ where $\rho_{\theta\tilde{s}} = \rho_{\tilde{s}}$ is $F_v(MID_{\theta s}^v|\theta)$. In other words, for each v, the probability of not being assigned any $\tilde{s} \in B_{\theta s}^v$ where $\rho_{\theta\tilde{s}} = \rho_{\tilde{s}}$ for a type θ applicant is $1 - F_v(MID_{\theta s}^v|\theta)$. Since tie-breakers are assumed to be distributed independently of one another, the probability of not being assigned any $\tilde{s} \in B_{\theta s}$ where $\rho_{\theta\tilde{s}} = \rho_{\tilde{s}}$ for a type θ applicant is $\Pi_v(1 - F_v(MID_{\theta s}^v|\theta))$. Every applicant of type $\theta \in \Theta_s^a$ who is not assigned a higher choice is assigned s because $\rho_{\theta s} < \rho_s$, and so

$$p_s(\theta) = \Pi_v(1 - F_v(MID_{\theta s}^v|\theta)) \text{ for all } \theta \in \Theta_s^a.$$

Finally, consider applicants of type $\theta \in \Theta_s^c$ who are not assigned a higher choice. The fraction of applicants $\theta \in \Theta_s^c$ who are not assigned a higher choice is $\Pi_v(1 - F_v(MID_{\theta s}^v|\theta))$. Also, the values of the tie-breaking variable $v(s)$ of these applicants are larger than $MID_{\theta s}^{v(s)}$. If $\tau_s < MID_{\theta s}^{v(s)}$, then no such applicant is assigned s. If $\tau_s \geq MID_{\theta s}^{v(s)}$, then the ratio of applicants that are assigned s within this set is given by $\frac{F_{v(s)}(\tau_s|\theta) - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)}{1 - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)}$. Hence, conditional on $\theta \in \Theta_s^c$ and not being assigned a choice higher than s, the probability of being assigned s is given by $\max\{0, \frac{F_{v(s)}(\tau_s|\theta) - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)}{1 - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)}\}$. Therefore,

$$p_s(\theta) = \prod_{v \neq v(s)} (1 - F_v(MID_{\theta s}^v|\theta)) \times \max\left\{0, F_{v(s)}(\tau_s|\theta) - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)\right\} \text{ for all } \theta \in \Theta_s^c.$$

C.3 Theorem 1

For each $\delta > 0$, let

$$\psi_s(\theta, T, \delta, w) \equiv E[D_i(s)|\theta_i = \theta, T_i(\delta) = T, W_i = w]$$

be assignment risk for an applicant with $\theta_i = \theta, T_i(\delta) = T$, and characteristics $W_i = w$. Our proofs use a lemma that describes this assignment risk in a continuum market with the general tie-breaking structure in Section 3. Given this continuum market, for each tie-breaker $v = 2, ..., V+1$, let $e(v)$ denote the event that $\theta_i = \theta, R_{iu} > MID^u_{\theta s}$ for $u = 1, ..., v-1, T_i(\delta) = T$, and $W_i = w$. $e(1)$ denotes the event that $\theta_i = \theta, T_i(\delta) = T$, and $W_i = w$. Also, let

$$\Phi_\delta(v) \equiv \left(\frac{F_v(MID^v_{\theta s}|e(v)) - F_v(MID^v_{\theta s} - \delta|e(v))}{F_v(MID^v_{\theta s} + \delta|e(v)) - F_v(MID^v_{\theta s} - \delta|e(v))}\right)^{1\{t_b(\delta)=c \text{ for some } b\in B^v_{\theta s}\}} \text{ for } v > 0$$

$$\Phi_\delta \equiv (1 - MID^0_{\theta s})\Pi_{v=1}^V \Phi_\delta(v)$$

$$\Phi'_\delta \equiv \max\{0, \frac{F_{v(s)}(\tau_s|e(V+1)) - \max\{F_{v(s)}(MID^{v(s)}_{\theta s}|e(V+1)), F_{v(s)}(\tau_s - \delta|e(V+1))\}}{F_{v(s)}(\tau_s + \delta|e(V+1)) - \max\{F_{v(s)}(MID^{v(s)}_{\theta s}|e(V+1)), F_{v(s)}(\tau_s - \delta|e(V+1))\}}\} \text{ if } v(s) > 0$$

$$\Phi'_\delta \equiv \max\{0, \frac{\tau_s - MID^0_{\theta s}}{1 - MID^0_{\theta s}}\} \text{ if } v(s) = 0$$

Lemma 4. *In the general tie-breaking setting of Section 3, for any fixed $\delta > 0$, we have:*

$$\psi_s(\theta, T, \delta, w) = \begin{cases} 0 & \textit{if } t_s(\delta) = n \textit{ or } t_b(\delta) = a \textit{ for some } b \in B_{\theta s}, \\ \Phi_\delta & \textit{otherwise and } t_s(\delta) = a, \\ \Phi_\delta \times \Phi'_\delta & \textit{otherwise and } t_s(\delta) = c. \end{cases}$$

Proof of Lemma 4. We start verifying the first line in $\psi_s(\theta, T, \delta, w)$. Applicants who don't list s have $\psi_s(\theta, T, \delta, w) = 0$. Among those who list s, those of $t_s(\delta) = n$ have $\theta \in \Theta^n_s$ or, if $v(s) \neq 0$, $\theta \in \Theta^c_s$ and $R_{iv(s)} > \tau_s + \delta$. If $\theta \in \Theta^n_s$, then $\rho_{\theta s} > \rho_s$ so that $\psi_s(\theta, T, \delta, w) = 0$. Even if $\theta \notin \Theta^n_s$, as long as $\theta \in \Theta^c_s$ and $R_{iv(s)} > \tau_s + \delta$, student i never clear the cutoff at school s so $\psi_s(\theta, T, \delta, w) = 0$.

To show the remaining cases, take as given that it is not the case that $t_s(\delta) = n$ or $t_b(\delta) = a$ for some $b \in B_{\theta s}$. First note that other applicants of $t_b(\delta) \neq a$ for all $b \in B_{\theta s}$ and $t_s(\delta) = a$ or c may be assigned $b \in B_{\theta s}$, where $\rho_{\theta b} = \rho_b$. Since the (aggregate) distribution of tie-breaking variables for type θ students is $\hat{F}_v(\cdot|\theta) = F_v(\cdot|\theta)$, conditional on $T_i(\delta) = T$, the proportion of type θ applicants not being assigned any $b \in B_{\theta s}$ where $\rho_{\theta b} = \rho_b$ is $\Phi_\delta = (1 - MID^0_{\theta s})\Pi_v \Phi_\delta(v)$ since each $\Phi_\delta(v)$ is the probability of not being assigned to any $b \in B^v_{\theta s}$. To see why $\Phi_\delta(v)$ is the probability of not being assigned to any $b \in B^v_{\theta s}$, first note that if $t_b(\delta) \neq c$ for all $b \in B^v_{\theta s}$, then $t_b(\delta) = n$ for all $b \in B^v_{\theta s}$ so that applicants are never assigned to any $b \in B^v_{\theta s}$. Otherwise, i.e., if $t_b(\delta) = c$ for some $b \in B^v_{\theta s}$, then applicants are assigned to s if and only if their values of tie-breaker v clear the cutoff of the school that produces $MID^v_{\theta s}$, where applicants have $t_s(\delta) = c$. This event happens with probability

$$\frac{F_v(MID^v_{\theta s}|e(v)) - F_v(MID^v_{\theta s} - \delta|e(v))}{F_v(MID^v_{\theta s} + \delta|e(v)) - F_v(MID^v_{\theta s} - \delta|e(v))},$$

implying that $\Phi_\delta(v)$ is the probability of not being assigned to any $b \in B^v_{\theta s}$.

Given this fact, to see the second line, note that every applicant of type $t_s(\delta) = a$ who is not assigned a higher choice is assigned s for sure because $\rho_{\theta s} < \rho_s$ or $\rho_{\theta s} + R_{iv(s)} < \xi_s$. Therefore, we have

$$\psi_s(\theta, T, \delta, w) = \Phi_\delta.$$

Finally, consider applicants with $t_s(\delta) = c$. The fraction of these who are not assigned a higher choice is Φ_δ, as explained above. Also, for tie-breaker $v(s)$, the tie-breaking numbers of these applicants are larger (worse) than $MID_{\theta s}^{v(s)}$. If $\tau_s < MID_{\theta s}^{v(s)}$, then no such applicant is assigned s. If $\tau_s \geq MID_{\theta s}^{v(s)}$, then the ratio of applicants that are assigned s conditional on $\tau_s \geq MID_{\theta s}^{v(s)}$ is given by

$$\max\{0, \frac{F_{v(s)}(\tau_s|e(V+1)) - \max\{F_{v(s)}(MID_{\theta s}^{v(s)}|e(V+1)), F_{v(s)}(\tau_s - \delta|e(V+1))\}}{F_{v(s)}(\tau_s + \delta|e(V+1)) - \max\{F_{v(s)}(MID_{\theta s}^{v(s)}|e(V+1)), F_{v(s)}(\tau_s - \delta|e(V+1))\}}\} \text{ if } v(s) \neq 0$$

$$\max\{0, \frac{\tau_s - MID_{\theta s}^0}{1 - MID_{\theta s}^0}\} \text{ if } v(s) = 0$$

Hence, conditional on $t_s(\delta) = c$ and not being assigned a choice higher than s, the probability of being assigned s is given by Φ'_δ. Therefore, for students with $t_s(\delta) = c$, we have $\psi_s(\theta, T, \delta, w) = \Phi_\delta \times \Phi'_\delta$. $\square$

Lemma 5. *In the general tie-breaking setting of Section 3, for all s, θ, and sufficiently small $\delta > 0$, we have:*

$$\psi_s(\theta, T, \delta, w) = \begin{cases} 0 & \text{if } t_s(0) = n \text{ or } t_b(0) = a \text{ for some } b \in B_{\theta s}, \\ \Phi_\delta^* & \text{otherwise and } t_s(0) = a, \\ \Phi_\delta^* \times & \dfrac{F_{v(s)}(\tau_s|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))}{F_{v(s)}(\tau_s + \delta|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))} \\ & \text{otherwise and } t_s(0) = c \text{ and } v(s) \neq 0. \\ \Phi_\delta^* \times & \max\{0, \dfrac{\tau_s - MID_{\theta s}^0}{1 - MID_{\theta s}^0}\} \\ & \text{otherwise and } t_s(0) = c \text{ and } v(s) = 0. \end{cases} \tag{17}$$

where

$$\Phi_\delta^*(v) \equiv \left(\frac{F_v(MID_{\theta s}^v + \delta|e(v)) - F_v(MID_{\theta s}^v|e(v))}{F_v(MID_{\theta s}^v + \delta|e(v)) - F_v(MID_{\theta s}^v - \delta|e(v))} \right)^{1\{MID_{\theta s}^v = \tau_b \text{ and } t_b = c \text{ for some } b \in B_{\theta s}^v\}}$$

$$\Phi_\delta^* \equiv (1 - MID_{\theta s}^0)\Pi_{v=1}^V \Phi_\delta^*(v)$$

Proof of Lemma 5. The first line follows from Lemma 4 and the fact that $t_s(0) = n$ or $t_b(0) = a$ for some $b \in B_{\theta s}$ imply $t_s(\delta) = n$ or $t_b(\delta) = a$ for some $b \in B_{\theta s}$ for sufficiently small $\delta > 0$.

To get the remaining lines, first note that conditional on $t_s(0) \neq n$ and $t_b(0) \neq a$ for all $b \in B_{\theta s}$, we have $\Phi_\delta^*(v) = \Phi_\delta(v)$ and so $\Phi_\delta^* = \Phi_\delta$ holds for small enough δ. Φ_δ^* therefore provides the probability of not being assigned to a school preferred to s in the bottom three cases.

The second line is then by the fact that $t_s(0) = a$ implies $t_s(\delta) = a$ for small enough $\delta > 0$. The third line is by the fact that for small enough $\delta > 0$,

$$\begin{aligned} \Phi'_\delta &= \max\left\{0, \frac{F_{v(s)}(\tau_s|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))}{F_{v(s)}(\tau_s + \delta|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))}\right\} \\ &= \frac{F_{v(s)}(\tau_s|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))}{F_{v(s)}(\tau_s + \delta|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))}, \end{aligned}$$

where we invoke Assumption 1(b), which implies $MID^v_{\theta s} \neq \tau_s$. The last line directly follows from Lemma 4. $\square$

We use Lemma 5 to derive Theorem 1. We characterize $\lim_{\delta\to 0} \psi_s(\theta, T, \delta, w)$ and show that it coincides with $\psi_s(\theta, T)$ in the main text. In the first case in Lemma 5, $\psi_s(\theta, T, \delta, w)$ is constant (0) for any small enough δ. The constant value is also $\lim_{\delta\to 0} \psi_s(\theta, T, \delta, w)$ in this case.

To characterize $\lim_{\delta\to 0} \psi_s(\theta, T, \delta, w)$ in the remaining cases, note that by the differentiability of $F_v(\cdot|e(v))$ (recall Assumption 1), L'Hopital's rule implies:

$$\lim_{\delta\to 0} \frac{F_{v(s)}(\tau_s|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))}{F_{v(s)}(\tau_s + \delta|e(V+1)) - F_{v(s)}(\tau_s - \delta|e(V+1))} = \frac{F'_{v(s)}(\tau_s|e(V+1))}{2F'_{v(s)}(\tau_s|e(V+1))} = 0.5.$$

$$\lim_{\delta\to 0} \frac{F_v(MID^v_{\theta s} + \delta|e(v)) - F_v(MID^v_{\theta s}|e(v))}{F_v(MID^v_{\theta s} + \delta|e(v)) - F_v(MID^v_{\theta s} - \delta|e(v))} = \frac{F'_v(MID^v_{\theta s}|e(v))}{2F'_v(MID^v_{\theta s}|e(v))} = 0.5.$$

This implies $\lim_{\delta\to 0} \Phi^*_\delta(v) = 0.5^{1\{MID^v_{\theta s} = \tau_b \text{ and } t_b = c \text{ for some } b \in B^v_{\theta s}\}}$ since $1\{MID^v_{\theta s} = \tau_b$ and $t_b = c$ for some $b \in B^v_{\theta s}\}$ does not depend on δ. So

$$\lim_{\delta\to 0} \Phi^*_\delta = (1 - MID^0_{\theta s}) 0.5^{m_s(\theta, T)}.$$

where recall $m_s(\theta, T) = |\{v > 0 : MID^v_{\theta s} = \tau_b \text{ and } t_b = c \text{ for some } b \in B^v_{\theta s}\}|$.

Combining these limiting facts with the fact that the limit of a product of functions equals the product of the limits of the functions, we obtain the following: $\lim_{\delta\to 0} \psi_s(\theta, T, \delta, w) = 0$ if (a) $t_s = n$; or (b) $t_b = a$ for some $b \in B_{\theta s}$. Otherwise,

$$\lim_{\delta\to 0} \psi_s(\theta, T, \delta, w) = \begin{cases} 0.5^{m_s(\theta, T(0))}(1 - MID^0_{\theta s}) & \text{if } t_s(0) = a \\ 0.5^{m_s(\theta, T(0))} \max\left\{0, \tau_s - MID^0_{\theta s}\right\} & \text{if } t_s(0) = c \text{ and } v(s) = 0 \\ 0.5^{1 + m_s(\theta, T(0))}(1 - MID^0_{\theta s}) & \text{if } t_s(0) = c \text{ and } v(s) > 0. \end{cases} \quad (18)$$

This expression coincides with $\psi_s(\theta, T)$, completing the proof of Theorem 1.

C.4 Theorem 2

Here we prove the following lemmas used in the proof of Theorem 2 in the main text.

Lemma 6. *(Cutoff almost sure convergence)* $\hat{\xi}_N \xrightarrow{a.s.} \xi$ *where* ξ *denotes the vector of continuum market cutoffs.*

Lemma 7. *(Estimated local propensity score almost sure convergence) For all* $\theta \in \Theta, s \in S$, *and* $T \in \{a, c, n\}^S$, *we have* $\hat{\psi}_{Ns}(\theta, T; \delta_N) \xrightarrow{a.s.} \psi_s(\theta, T)$ *as* $N \to \infty$ *and* $\delta_N \to 0$.

Lemma 8. *(True bandwidth-specific propensity score almost sure convergence) For all* $\theta \in \Theta, s \in S$, $T \in \{a, c, n\}^S$, *and* δ_N *such that* $\delta_N \to 0$ *and* $N\delta_N \to \infty$ *as* $N \to \infty$, *we have* $\psi_{Ns}(\theta, T; \delta_N) \xrightarrow{a.s.} \psi_s(\theta, T)$ *as* $N \to \infty$.

Proof of Lemma 6

Lemma 6 is proved using a strategy similar to that used to prove Lemma 3 in Abdulkadiroğlu *et al.* (2017b). Using the Extended Continuous Mapping Theorem (Theorem 19.1 in van der Vaart (2000)), we first show deterministic convergence of cutoffs to ensure the continuous mapping result is applicable.

Modify the definition of G to describe the distribution of running variables as well as types: For any set of applicant types $\Theta_0 \subset \Theta$ and for any numbers $r_0, r_1 \in [0,1]^V$ with $r_{0v} < r_{1v}$ for all v, define the set of applicants of types in Θ_0 with random numbers worse than r_0 and better than r_1 as

$$I(\Theta_0, r_0, r_1) = \{i \in I \mid \theta_i \in \Theta_0, r_{v0} < r_{iv} \leq r_{v1} \text{ for all } v\}.$$

In a continuum market,

$$G(I(\Theta_0, r_0, r_1)) = E[1\{i \in I(\Theta_0, r_0, r_1)\}],$$

where the expectation is assumed to exist and taken over the running variable distributions. In a finite market with N applicants,

$$G(I(\Theta_0, r_0, r_1)) = \frac{|I(\Theta_0, r_0, r_1)|}{N}.$$

Let $\mathcal{G}$ be the set of possible G's defined above. For any two distributions G and G', the supnorm metric is defined by

$$d(G, G') = \sup_{\Theta_0 \subset \Theta, r_0, r_1 \in [0,1]^{V+1}} |G(I(\Theta_0, r_0, r_1)) - G'(I(\Theta_0, r_0, r_1))|.$$

The notation is otherwise as in the text.

Consider a deterministic sequence of economies described by a sequence of distributions $\{g_N\}$ over applicants, together with associated school capacities, so that for all N, $g_N \in \mathcal{G}$ is a potential realization produced by randomly drawing N applicants and their running variables from G. Assume that $g_N \to G$ in metric space $(\mathcal{G}, d)$. Let ξ_N denote the admissions cutoffs in g_N. Note the ξ_N is constant because this is the cutoff for a particular realized market g_N.

The proof first shows deterministic convergence of cutoffs for any convergent subsequence of g_N. Let $\{\tilde{g}_N\}$ be any subsequence of realized economies $\{g_N\}$. The corresponding cutoffs are denoted by $\{\tilde{\xi}_N\}$. Let $\tilde{\xi} \equiv (\tilde{\xi}_s)$ be the limit of $\tilde{\xi}_N$. The following two claims establish that $\tilde{\xi}_N \to \xi$, the cutoff associated with G.

Claim 1. *$\tilde{\xi}_s \geq \xi_s$ for every $s \in S$.*

Proof of Claim 1. This is proved by contradiction in three steps. Suppose to the contrary that $\tilde{\xi}_s < \xi_s$ for some s. Let $S' \subset S$ be the set of schools the cutoffs of which are strictly lower under $\tilde{\xi}$. For any $s \in S'$, define $I_N^s = \{i \in I | \tilde{\xi}_{Ns} < \rho_{is} + r_{is} \leq \xi_s$ and i lists s first$\}$ where I is the set of applicants in G, which contains the set of applicants in g_N for all N. In other words, I_N^s are the set of applicants listing school s first who have an applicant position in between $\tilde{\xi}_{Ns}$ and ξ_s.

Step (a): We first show that for our subsequence, when the market is large enough, there must be some applicants who are in I_N^s. That is, there exists N_0 such that for any $N > N_0$, we have $\tilde{g}_N(I_N^s) > 0$ for all $s \in S'$.

To see this, we begin by showing that for all $s \in S'$, there exists N_0 such that for any $N > N_0$, we have $G(I_N^s) > 0$. Suppose, to the contrary, that there exists $s \in S'$ such that for all N_0, there exists $N > N_0$ such that $G(I_N^s) = 0$. When we consider the subsequence of realized economies $\{\tilde{g}_N\}$, we find that

$$
\begin{aligned}
&\tilde{g}_N(\{i \in Q_s(\xi_N) \text{ such that } \pi_{\theta_i s}(r_i) \leq \xi_s\}) \\
&\quad = \tilde{g}_N(\{i \in Q_s(\xi_N) \text{ such that } \pi_{\theta_i s}(r_i) \leq \tilde{\xi}_{Ns}\}) \qquad (19) \\
&\qquad + \tilde{g}_N(\{i \in Q_s(\xi_N) \text{ such that } \tilde{\xi}_{Ns} < \pi_{\theta_i s}(r_i) \leq \xi_s\}) \\
&\quad = \tilde{g}_N(\{i \in Q_s(\xi_N) \text{ such that } \pi_{\theta_i s}(r_i) \leq \tilde{\xi}_{Ns}\}) \qquad (20) \\
&\quad \leq q_s \qquad (21)
\end{aligned}
$$

where $\pi_{\theta_i s}(r_i) \equiv \rho_{\theta_i s} + r_{is}$. Expression (20) follows from Assumptions 1 and 2 by the following reason. (20) does not hold, i.e., $\tilde{g}_N(\{i \in Q_s(\xi_N) \text{ such that } \tilde{\xi}_{Ns} < \pi_{\theta_i s}(r_i) \leq \xi_s\}) > 0$ only if $G(\{i \in I | \tilde{\xi}_{Ns} < \pi_{\theta_i s}(r_i) \leq \xi_s\}) > 0$. This and Assumptions 1 and 2 imply $G(\{i \in I | \tilde{\xi}_{Ns} < \pi_{\theta_i s}(r_i) \leq \xi_s \text{ and } i \text{ lists } s \text{ first}\}) \equiv G(I_N^s) > 0$, a contradiction to $G(I_N^s) = 0$.

Since $\tilde{g}_N$ is realized as N iid samples from G, $\tilde{g}_N(\{i \in I | \tilde{\xi}_{Ns} < \pi_{\theta_i s}(r_i) \leq \xi_s\}) = 0$. Expression (21) follows by our definition of DA, which can never assign more applicants to a school than its capacity for each of the N samples. We obtain our contradiction since $\tilde{\xi}_{Ns}$ violates the definition of DA cutoffs at s in $\tilde{g}_N$ since expression (21) means it is possible to increase the cutoff $\tilde{\xi}_{Ns}$ to ξ_s without violating the capacity constraint.

Given that we've just shown that for each $s \in S'$, $G(I_N^s) > 0$ for some N, it is possible to find an N such that $G(I_N^s) > \epsilon > 0$ for some $\epsilon > 0$. Since $g_N \to G$ and so $\tilde{g}_N \to G$, there exists N_0 such that for all $N > N_0$, we have $\tilde{g}_N(I_N^s) > G(I_N^s) - \epsilon > 0$. Since the number of schools is finite, such N_0 can be taken uniformly over all $s \in S$. This completes the argument for Step (a).

Step (a) allows us to find some N_0 such that for any $N > N_0$, $\tilde{g}_N(I_N^s) > 0$ for all $s' \in S'$. Let $\tilde{s}_N \in S$ and t be such that $\tilde{\xi}_{Ns}^{t-1} \geq \xi_s$ for all $s \in S$ and $\tilde{\xi}_{N\tilde{s}_N}^t < \xi_{\tilde{s}_N}$, where $\tilde{\xi}_{Ns}^t$ is school s's tentative cutoff at round t of the DA algorithm. That is, $\tilde{s}_N$ is one of the first schools the cutoff of which falls strictly below $\xi_{\tilde{s}_N}$ under the DA algorithm in $\tilde{g}_N$, which happens in round t of the DA algorithm. Such $\tilde{s}_N$ and t exist since the choice of N guarantees $\tilde{g}_N(I_N^s) > 0$ and so $\tilde{\xi}_{Ns} < \xi_s$ for all $s \in S'$.

Step (b): We next show that there exist infinitely many values of N such that the associated $\tilde{s}_N$ is in S' and $\tilde{g}_N(I_N^s) > 0$ for all $s \in S'$. It is because otherwise, by Step (a), there exists N_0 such that for all $N > N_0$, we have $\tilde{s}_N \notin S'$. Since there are only finitely many schools, $\{\tilde{s}_N\}$ has a subsequence $\{\tilde{s}_m\}$ such that $\tilde{s}_m$ is the same school outside S' for all m. By definition of $\tilde{s}_N$, $\tilde{\xi}_{m\tilde{s}_m} \leq \tilde{\xi}^t_{m\tilde{s}_m} < \xi_{\tilde{s}_m}$ for all m and so $\tilde{\xi}_{\tilde{s}_m} < \xi_{\tilde{s}_m}$, a contradiction to $\tilde{s}_m \notin S'$. Therefore, we have our desired conclusion of Step (b).

Fix some N such that the associated $\tilde{s}_N$ is in S' and $\tilde{g}_N(I_N^s) > 0$ for all $s \in S'$. Step (b) guarantees that such N exists. Let $\tilde{A}_{N\tilde{s}_N}$ and $A_{\tilde{s}_N}$ be the sets of applicants assigned $\tilde{s}_N$ under $\tilde{g}_N$ and G, respectively. All applicants in $I_N^{\tilde{s}_N}$ are assigned $\tilde{s}_N$ in G and rejected by $\tilde{s}_N$ in $\tilde{g}_N$. Since these applicants list $\tilde{s}_N$ first, there must exist a positive measure (with respect to $\tilde{g}_N$) of applicants outside $I_N^{\tilde{s}_N}$ who are assigned $\tilde{s}_N$ in $\tilde{g}_N$ and some other school in G; denote the set of them by $\tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}$. $\tilde{g}_N(\tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}) > 0$ since otherwise, for any N such that Step (b) applies,

$$\tilde{g}_N(\tilde{A}_{N\tilde{s}_N}) \leq \tilde{g}_N(A_{\tilde{s}_N} \setminus I_N^{\tilde{s}_N}) = \tilde{g}_N(A_{\tilde{s}_N}) - \tilde{g}_N(I_N^{\tilde{s}_N}),$$

which by Step (a) converges to something strictly smaller than $G(A_{\tilde{s}_N})$ since $\tilde{g}_N(A_{\tilde{s}_N}) \to G(A_{\tilde{s}_N})$ and $\tilde{g}_N(I_N^{\tilde{s}_N}) > 0$ for all large enough N by Step (a). Note that $G(A_{\tilde{s}_N})$ is weakly smaller than $q_{\tilde{s}_N}$. This implies that for large enough N, $\tilde{g}_N(\tilde{A}_{N\tilde{s}_N}) < q_{\tilde{s}_N}$, a contradiction to $\tilde{A}_{N\tilde{s}_N}$'s being the set of applicants assigned $\tilde{s}_N$ at a cutoff strictly smaller than the largest possible value $K+1$. For each $i \in \tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}$, let s_i be the school to which i is assigned under G.

Step (c): To complete the argument for Claim 1, we show that some $i \in \tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}$ must have been rejected by s_i in some step $\tilde{t} \leq t-1$ of the DA algorithm in $\tilde{g}_N$. That is, there exists $i \in \tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}$ and $\tilde{t} \leq t-1$ such that $\pi_{is_i} > \tilde{\xi}^{\tilde{t}}_{Ns_i}$. Suppose to the contrary that for all $i \in \tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}$ and $\tilde{t} \leq t-1$, we have $\pi_{is_i} \leq \tilde{\xi}^{\tilde{t}}_{Ns_i}$. Each such applicant i must prefer s_i to $\tilde{s}_N$ because i is assigned $s_i \neq \tilde{s}_N$ under G though $\pi_{i\tilde{s}_N} \leq \tilde{\xi}_{N\tilde{s}_N} < \xi_{\tilde{s}_N}$, where the first inequality holds because i is assigned $\tilde{s}_N$ in $\tilde{G}_N$ while the second inequality does because $\tilde{s}_N \in S'$. This implies none of $\tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}$ is rejected by s_i, applies for $\tilde{s}$, and contributes to decreasing $\tilde{\xi}^t_{N\tilde{s}_N}$ at least until step t and so $\tilde{\xi}^t_{N\tilde{s}_N} < \xi_{\tilde{s}_N}$ cannot be the case, a contradiction. Therefore, we have our desired conclusion of Step (c).

Claim 1 can now be established by showing that Step (c) implies there are $i \in \tilde{A}_{N\tilde{s}_N} \setminus A_{\tilde{s}_N}$ and $\tilde{t} \leq t-1$ such that $\pi_{is_i} > \tilde{\xi}^{\tilde{t}}_{Ns_i} \geq \tilde{\xi}_{Ns_i}$, where the last inequality is implies by the fact that in every market, for all $s \in S$ and $t \geq 0$, we have $\xi^{t+1}_s \leq \xi^t_s$. Also, they are assigned s_i in G so that $\pi_{is_i} \leq \xi_{s_i}$. These imply $\xi_{s_i} > \tilde{\xi}^{\tilde{t}}_{Ns_i} \geq \tilde{\xi}_{Ns_i}$. That is, the cutoff of s_i falls below ξ_{si} in step $\tilde{t} \leq t-1 < t$ of the DA algorithm in $\tilde{g}_N$. This contradicts the definition of $\tilde{s}_N$ and t. Therefore $\tilde{\xi}_s \geq \xi_s$ for all $s \in S$, as desired. $\square$

Claim 2. *By a similar argument, $\tilde{\xi}_s \leq \xi_s$ for every $s \in S$.*

Since $\tilde{\xi}_s \geq \xi_s$ and $\tilde{\xi}_s \leq \xi_s$ for all s, it must be the case that $\tilde{\xi}_N \to \xi$. The following claim uses this to show that $\xi_N \to \xi$.

Claim 3. *If $\tilde{\xi}_N \to \xi$ for every convergent subsequence $\{\tilde{\xi}_N\}$ of $\{\xi_N\}$, then $\xi_N \to \xi$.*

Proof of Claim 3. Since $\{\xi_N\}$ is bounded in $[0, K+1]^{|S|}$, it has a convergent subsequence by the Bolzano-Weierstrass theorem. Suppose to the contrary that for every convergent subsequence $\{\tilde{\xi}_N\}$, we have $\tilde{\xi}_N \to \xi$, but $\xi_N \not\to \xi$. Then there exists $\epsilon > 0$ such that for all $k > 0$, there exists $N_k > k$ such that $||\xi_{N_k} - \xi|| \geq \epsilon$. Then the subsequence $\{\xi_{N_k}\}_k \subset \{\xi_N\}$ has a convergent subsequence that does not converge to ξ (since $||\xi_{N_k} - \xi|| \geq \epsilon$ for all k), which contradicts the supposition that every convergent subsequence of $\{\xi_N\}$ converges to ξ. $\square$

The last step in the proof of Lemma 6 relates this fact to stochastic convergence.

Claim 4. *$\xi_N \to \xi$ implies $\hat{\xi}_N \xrightarrow{a.s.} \xi$*

Proof of Claim 4. This proof is based on two off-the-shelf asymptotic results from statistics. First, let G_N be the distribution over $I(\Theta_0, r_0, r_1)$'s generated by randomly drawing N applicants from G. Note that G_N is random since it involves randomly drawing N applicants. $G_N \overset{a.s.}{\to} G$ by the Glivenko-Cantelli theorem for independent but non-identically distributed random variables (Wellner, 1981). Next, since $G_N \overset{a.s.}{\to} G$ and $\xi_N \to \xi$, the Extended Continuous Mapping Theorem (Theorem 18.11 in van der Vaart (2000)) implies that $\hat{\xi}_N \xrightarrow{a.s.} \xi$, completing the proof of Lemma 6. $\square$

Proof of Lemma 7

$\hat{\psi}_{Ns}(\theta, T; \delta_N)$ is almost everywhere continuous in finite sample cutoffs $\hat{\xi}_N$, finite sample MIDs ($MID^v_{\theta s}$), and bandwidth δ_N. Since every $MID^v_{\theta s}$ is almost everywhere continuous in finite sample cutoffs $\hat{\xi}_N$, $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ is almost everywhere continuous in finite sample cutoffs $\hat{\xi}_N$ and bandwidth δ_N. Recall $\delta_N \to 0$ by assumption while $\hat{\xi}_N \xrightarrow{a.s.} \xi$ by Lemma 6. Therefore, by the continuous mapping theorem, as $N \to \infty$, $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ almost surely converges to $\hat{\psi}_{Ns}(\theta, T; \delta_N)$ with ξ replacing $\hat{\xi}_N$, which is $\psi_s(\theta, T)$.

Proof of Lemma 8

We use the following fact, which is implied by Example 19.29 in van der Vaart (2000).

Lemma 9. *Let X be a random variable distributed according to some CDF F over $[0, 1]$. Let $F(\cdot|X \in [x-\delta, x+\delta])$ be the conditional version of F conditional on X being in a small window $[x-\delta, x+\delta]$ where $x \in [0, 1]$ and $\delta \in (0, 1]$. Let $X_1, ..., X_N$ be iid draws from F. Let $\hat{F}_N$ be the empirical CDF of $X_1, ..., X_N$. Let $\hat{F}_N(\cdot|X \in [x-\delta, x+\delta])$ be the conditional version of $\hat{F}_N$ conditional on a subset of draws falling in $[x-\delta, x+\delta]$, i.e., $\{X_i | i = 1, ..., n, X_i \in [x-\delta, x+\delta]\}$. Suppose (δ_N) is a sequence with $\delta_N \downarrow 0$ and $\delta_N \times N \to \infty$. Then $\hat{F}_N(\cdot|X \in [x-\delta_N, x+\delta_N])$ uniformly converges to $F(\cdot|X \in [x-\delta_N, x+\delta_N])$, i.e.,*

$$\sup_{x' \in [0,1]} |\hat{F}_N(x'|X \in [x-\delta_N, x+\delta_N]) - F(x'|X \in [x-\delta_N, x+\delta_N])| \to_p 0 \text{ as } N \to \infty \text{ and } \delta_N \to 0.$$

Proof of Lemma 9. We first prove the statement for $x \in (0,1)$. Let P be the probability measure of X and $\hat{P}_N$ be the empirical measure of $X_1, ..., X_N$. Note that

$$
\begin{aligned}
&\sup_{x' \in [0,1]} |\hat{F}_N(x'|X \in [x-\delta_N, x+\delta_N]) - F(x'|X \in [x-\delta_N, x+\delta_N])| \\
=& \sup_{t \in [-1,1]} |\hat{F}_N(x+t\delta_N|X \in [x-\delta_N, x+\delta_N]) - F(x+t\delta_N|X \in [x-\delta_N, x+\delta_N])| \\
=& \sup_{t \in [-1,1]} |\frac{\hat{P}_N[x-\delta_N, x+t\delta_N]}{\hat{P}_N[x-\delta_N, x+\delta_N]} - \frac{P_X[x-\delta_N, x+t\delta_N]}{P_X[x-\delta_N, x+\delta_N]}| \\
=& \frac{1}{\hat{P}_N[x-\delta_N, x+\delta_N]P_X[x-\delta_N, x+\delta_N]} \\
&\times \sup_{t \in [-1,1]} |\hat{P}_N[x-\delta_N, x+t\delta_N]P_X[x-\delta_N, x+\delta_N] - \hat{P}_N[x-\delta_N, x+\delta_N]P_X[x-\delta_N, x+t\delta_N]| \\
=& \frac{1}{\hat{P}_N[x-\delta_N, x+\delta_N]P_X[x-\delta_N, x+\delta_N]} \\
&\times \sup_{t \in [-1,1]} |\hat{P}_N[x-\delta_N, x+t\delta_N](P_X[x-\delta_N, x+\delta_N] - \hat{P}_N[x-\delta_N, x+\delta_N]) \\
&\quad + \hat{P}_N[x-\delta_N, x+\delta_N](\hat{P}_N[x-\delta_N, x+t\delta_N] - P_X[x-\delta_N, x+t\delta_N])| \\
\le& \frac{1}{\hat{P}_N[x-\delta_N, x+\delta_N]P_X[x-\delta_N, x+\delta_N]} \\
&\times \{ \sup_{t \in [-1,1]} \hat{P}_N[x-\delta_N, x+t\delta_N]|\hat{P}_N[x-\delta_N, x+\delta_N] - P_X[x-\delta_N, x+\delta_N]| \\
&\quad + \sup_{t \in [-1,1]} \hat{P}_N[x-\delta_N, x+\delta_N]|\hat{P}_N[x-\delta_N, x+t\delta_N] - P_X[x-\delta_N, x+t\delta_N]|\} \\
=& \frac{1}{P_X[x-\delta_N, x+\delta_N]} \\
&\times \{|\hat{P}_N[x-\delta_N, x+\delta_N] - P_X[x-\delta_N, x+\delta_N]| + \sup_{t \in [-1,1]} |\hat{P}_N[x-\delta_N, x+t\delta_N] - P_X[x-\delta_N, x+t\delta_N]|\} \\
=& \frac{A_N}{P_X[x-\delta_N, x+\delta_N]},
\end{aligned}
$$

where

$$
A_N = |\hat{P}_N[x-\delta_N, x+\delta_N] - P_X[x-\delta_N, x+\delta_N]| + \sup_{t \in [-1,1]} |\hat{P}_N[x-\delta_N, x+t\delta_N] - P_X[x-\delta_N, x+t\delta_N]|.
$$

The above inequality holds by the triangle inequality and the second last equality holds because $\sup_{t \in [-1,1]} \hat{P}_N[x-\delta_N, x+t\delta_N] = \hat{P}_N[x-\delta_N, x+\delta_N]$.

We show that $A_N/P_X[x-\delta_N, x+\delta_N] \xrightarrow{p} 0$. Example 19.29 in van der Vaart (2000) implies that the sequence of processes $\{\sqrt{n/\delta_N}(\hat{P}_N[x-\delta_N, x+t\delta_N] - P_X[x-\delta_N, x+t\delta_N]) : t \in [-1,1]\}$ converges in distribution to a Gaussian process in the space of bounded functions on $[-1,1]$ as $N \to \infty$. We denote this Gaussian process by $\{\mathbb{G}_t : t \in [-1,1]\}$. We then use the continuous mapping theorem to obtain

$$
\sqrt{n/\delta_N}A_N \xrightarrow{d} |\mathbb{G}_1| + \sup_{t \in [-1,1]} |\mathbb{G}_t|
$$

as $N \to \infty$. Since $\{\mathbb{G}_t : t \in [-1,1]\}$ has bounded sample paths, it follows that $|\mathbb{G}_1| < \infty$ and $\sup_{t\in[-1,1]} |\mathbb{G}_t| < \infty$ for sure. By the continuous mapping theorem, under the condition that $N\delta_N \to \infty$,

$$\begin{aligned}
(1/\delta_N)A_N &= (1/\sqrt{N\delta_N}) \times \sqrt{n/\delta_N}A_N \\
&\xrightarrow{d} 0 \times (|\mathbb{G}_1| + \sup_{t\in[-1,1]} |\mathbb{G}_t|) \\
&= 0.
\end{aligned}$$

This implies that $(1/\delta_N)A_N \xrightarrow{p} 0$, because for any $\epsilon > 0$,

$$\begin{aligned}
\Pr(|(1/\delta_N)A_N| > \epsilon) &= \Pr((1/\delta_N)A_N < -\epsilon) + \Pr((1/\delta_N)A_N > \epsilon) \\
&\leq \Pr((1/\delta_N)A_N \leq -\epsilon) + 1 - \Pr((1/\delta_N)A_N \leq \epsilon) \\
&\to \Pr(0 \leq -\epsilon) + 1 - \Pr(0 \leq \epsilon) \\
&= 0,
\end{aligned}$$

where the convergence holds since $(1/\delta_N)A_N \xrightarrow{d} 0$. To show that $A_N/P_X[x-\delta_N, x+\delta_N] \xrightarrow{p} 0$, it is therefore enough to show that $\lim_{N\to\infty}(1/\delta_N)P_X[x-\delta_N, x+\delta_N] > 0$. We have

$$\begin{aligned}
(1/\delta_N)P_X[x-\delta_N, x+\delta_N] &= (1/\delta_N)(F_X(x+\delta_N) - F_X(x-\delta_N)) \\
&= (1/\delta_N)(2f(x)\delta_N + o(\delta_N)) \\
&= 2f(x) + o(1) \\
&\to 2f(x) \\
&> 0,
\end{aligned}$$

where we use Taylor's theorem for the second equality and the assumption of $f(x) > 0$ for the last inequality.

We next prove the statement for $x = 0$. Note that

$$
\begin{aligned}
&\sup_{x' \in [0,1]} |\hat{F}_N(x'|X \in [-\delta_N, \delta_N]) - F(x'|X \in [-\delta_N, \delta_N])| \\
=& \sup_{t \in [0,1]} |\hat{F}_N(t\delta_N|X \in [0, \delta_N]) - F(t\delta_N|X \in [0, \delta_N])| \\
=& \sup_{t \in [0,1]} |\frac{\hat{F}_N(t\delta_N)}{\hat{F}_N(\delta_N)} - \frac{F_X(t\delta_N)}{F_X(\delta_N)}| \\
=& \frac{1}{\hat{F}_N(\delta_N)F_X(\delta_N)} \sup_{t \in [0,1]} |\hat{F}_N(t\delta_N)F_X(\delta_N) - \hat{F}_N(\delta_N)F_X(t\delta_N)| \\
=& \frac{1}{\hat{F}_N(\delta_N)F_X(\delta_N)} \sup_{t \in [0,1]} |\hat{F}_N(t\delta_N)(F_X(\delta_N) - \hat{F}_N(\delta_N)) + \hat{F}_N(\delta_N)(\hat{F}_N(t\delta_N) - F_X(t\delta_N))| \\
\leq& \frac{1}{\hat{F}_N(\delta_N)F_X(\delta_N)} \{ \sup_{t \in [0,1]} \hat{F}_N(t\delta_N)|\hat{F}_N(\delta_N) - F_X(\delta_N)| + \sup_{t \in [0,1]} \hat{F}_N(\delta_N)|\hat{F}_N(t\delta_N) - F_X(t\delta_N)| \} \\
=& \frac{1}{F_X(\delta_N)} \{ |\hat{F}_N(\delta_N) - F_X(\delta_N)| + \sup_{t \in [0,1]} |\hat{F}_N(t\delta_N) - F_X(t\delta_N)| \} \\
=& \frac{A_N^0}{F_X(\delta_N)},
\end{aligned}
$$

where $A_N^0 = |\hat{F}_N(\delta_N) - F_X(\delta_N)| + \sup_{t \in [0,1]} |\hat{F}_N(t\delta_N) - F_X(t\delta_N)|$. By the argument used in the above proof for $x \in (0, 1)$, we have $(1/\delta_N)A_N^0 \xrightarrow{p} 0$. It also follows that

$$
\begin{aligned}
(1/\delta_N)F_X(\delta_N) &= (1/\delta_N)(f(0)\delta_N + o(\delta_N)) \\
&= f(0) + o(1) \\
&\to f(0) \\
&> 0.
\end{aligned}
$$

Thus, $\frac{A_N^0}{F_X(\delta_N)} \xrightarrow{p} 0$, and hence $\sup_{x' \in [0,1]} |\hat{F}_N(x'|X \in [-\delta_N, \delta_N]) - F(x'|X \in [-\delta_N, \delta_N])| \xrightarrow{p} 0$. The proof for $x = 1$ follows from the same argument. $\square$

Consider any deterministic sequence of economies $\{g_N\}$ such that $g_N \in \mathcal{G}$ for all N and $g_N \to G$ in the $(\mathcal{G}, d)$ metric space. Let $p_{Ns}(\theta, T; \delta_N)$ be the (finite-market, deterministic) bandwidth-specific propensity score for a particular g_N.

For Lemma 8, it is enough to show deterministic convergence of this finite-market score, that is, $p_{Ns}(\theta, T; \delta_N) \to \psi_s(\theta, T)$ as $g_N \to G$. To see this, let G_N be the distribution over $I(\Theta_0, r_0, r_1)$'s induced by randomly drawing N applicants from G. Note that G_N is random and that $G_N \overset{a.s.}{\to} G$ by Wellner (1981)'s Glivenko-Cantelli theorem for independent but non-identically distributed random variables. $G_N \overset{a.s.}{\to} G$ and $p_{Ns}(\theta, T; \delta_N) \to \psi_s(\theta, T)$ allow us to apply the Extended Continuous Mapping Theorem (Theorem 18.11 in van der Vaart (2000)) to obtain $\tilde{p}_{Ns}(\theta, T; \delta_N) \xrightarrow{a.s.} \psi_s(\theta, T)$ where $\tilde{p}_{Ns}(\theta, T; \delta_N)$ is the random version of $p_{Ns}(\theta, T; \delta_N)$ defined for G_N.

We prove convergence of $p_{Ns}(\theta, T; \delta_N) \to \psi_s(\theta, T)$ as follows. For simplicity, consider the case where every school use a common non-random running variable, i.e., $v(s) = v(s') > 0$ for all schools s and s'. Let $\tilde{\xi}_{Ns}$ and $\tilde{\xi}_{Ns'}$ be the random cutoffs at s and s', respectively, in g_N, and

$$\begin{aligned}
\tau_{\theta s} &\equiv \xi_s - \rho_{\theta s}, \\
\tau_{\theta s_-} &\equiv \max_{s' \succ_\theta s} \{\xi_{s'} - \rho_{\theta s'}\}, \\
\tilde{\tau}_{N\theta s} &\equiv \tilde{\xi}_{Ns} - \rho_{\theta s}, \\
\tilde{\tau}_{N\theta s_-} &\equiv \max_{s' \succ_\theta s} \{\tilde{\xi}_{Ns'} - \rho_{\theta s'}\}.
\end{aligned}$$

where $\tau_{\theta s_-} = 0$ and $\tilde{\tau}_{N\theta s_-} = 0$ when there is no school s' such that $s' \succ_\theta s$. We can express $\psi_s(\theta, T)$ and $p_{Ns}(\theta, T; \delta_N)$ as follows:

$$\psi_s(\theta, T) = \begin{cases} 0 & \text{if } t_{s_2} = a \text{ or } t_s = n \\ 0.5 & \text{if } t_{s_2} = c \text{ or } t_s = c \\ 1 & \text{if } t_{s_2} = n \text{ and } t_s = a \end{cases}$$

$$p_{Ns}(\theta, T, \delta_N) = P_N(\tilde{\tau}_{N\theta s} \geq R_i > \tilde{\tau}_{N\theta s_-} | T_i(\delta_N) = T, \theta_i = \theta)$$

where $s_1 \equiv \arg\max_{s' \succ_\theta s} \{\tilde{\xi}_{Ns'} - \rho_{\theta s'}\}$ and $s_2 \equiv \arg\max_{s' \succ_\theta s} \{\xi_{s'} - \rho_{\theta s'}\}$. P_N is the probability induced by randomly drawing running variables given g_N, and R_i is a random (not realized) running variable for applicant i.

By Lemma 6, with probability 1, for all $\epsilon_1 > 0$, there exists N_1 such that for all $N > N_1$,

$$|\tilde{\xi}_{Ns'} - \xi_{s'}| < \epsilon_1 \text{ for all } s',$$

which implies that with probability 1,

$$\begin{aligned}
&|\tilde{\tau}_{N\theta s_-} - \tau_{\theta s_-}| \\
=&|\{\tilde{\xi}_{Ns_1} - \rho_{\theta s_1}\} - \{\xi_{s_2} - \rho_{\theta s_2}\}| \\
<&\begin{cases} |\{\tilde{\xi}_{Ns_1} - \rho_{\theta s_1}\} - (\{\tilde{\xi}_{Ns_2} - \rho_{\theta s_2}\} + \epsilon_1)| \text{ if } \xi_{s_2} - \rho_{\theta s_2} \geq \tilde{\xi}_{Ns_1} - \rho_{\theta s_1} \\ |\{\tilde{\xi}_{Ns_1} - \rho_{\theta s_1}\} - (\{\tilde{\xi}_{Ns_2} - \rho_{\theta s_2}\} - \epsilon_1)| \text{ if } \xi_{s_2} - \rho_{\theta s_2} < \tilde{\xi}_{Ns_1} - \rho_{\theta s_1} \end{cases} \\
=&\epsilon_1
\end{aligned}$$

where in the first equality, $s_1 \equiv \arg\max_{s' \succ_\theta s} \{\tilde{\xi}_{Ns'} - \rho_{\theta s'}\}$ and $s_2 \equiv \arg\max_{s' \succ_\theta s} \{\xi_{s'} - \rho_{\theta s'}\}$. The inequality is by $|\tilde{\xi}_{Ns'} - \xi_{s'}| < \epsilon_1$ for all s'.

For all $\epsilon > 0$, the above argument when we set $\epsilon_1 < \epsilon/2$ implies that there exists N_0 such that for all $N > N_0$,

$$\begin{aligned}
p_{Ns}(\theta, T, \delta_N) &= P_N(\tilde{\tau}_{N\theta s} \geq R_i > \tilde{\tau}_{N\theta s_-} | T_i(\delta_N) = T, \theta_i = \theta) \\
&\in (\psi_s(\theta, T) - \epsilon, \psi_s(\theta, T) + \epsilon),
\end{aligned}$$

where the last inclusion holds by the following reason. Suppose

$$\tau_{\theta s} < \tau_{\theta s_-} \text{ or } \tau_{\theta s} < r_{v(s)} \text{ or } r_{v(s_2)} < \tau_{\theta s_-}.$$

Lemma 6 guarantees that for large enough N, we have

$$\tau_{N\theta s} < \tau_{N\theta s_-} \text{ or } \tau_{N\theta s} < r_{v(s)} \text{ or } r_{v(s_2)} < \tau_{N\theta s_-}.$$

For small enough δ_N, therefore, it is never true that $\tilde{\tau}_{N\theta s} \geq R_i > \tilde{\tau}_{N\theta s_-}$, implying $p_{Ns}(\theta, T, \delta_N) = 0 = \psi_s(\theta, T)$.

Consider another case where

$$\tau_{\theta s} > r_{v(s)} > \tau_{\theta s_-}.$$

Lemma 6 guarantees that for large enough N, we have

$$\tau_{N\theta s} > r_{v(s)} > \tau_{N\theta s_-}.$$

For small enough δ_N, therefore, it is always true that $\tilde{\tau}_{N\theta s} \geq R_i > \tilde{\tau}_{N\theta s_-}$, implying $p_{Ns}(\theta, T, \delta_N) = 1 = \psi_s(\theta, T)$.

Finally, suppose

$$\tau_{\theta s} \geq \tau_{\theta s_-} \text{ and } (\tau_{\theta s} = r_{v(s)} \text{ or } r_{v(s_2)} = \tau_{\theta s_2}).$$

In this case, for any $\epsilon > 0$, for large enough N,

$$\begin{aligned} p_{Ns}(\theta, T, \delta_N) &= P_N(\tilde{\tau}_{N\theta s} \geq R_i > \tilde{\tau}_{N\theta s_-} | T_i(\delta_N) = T, \theta_i = \theta) \\ &\in (P(\tau_{\theta s} \geq R_i > \tau_{\theta s_-} | T_i(\delta_N) = T, \theta_i = \theta) - \epsilon, P(\tau_{\theta s} \geq R_i > \tau_{\theta s_-} | T_i(\delta_N) = T, \theta_i = \theta) + \epsilon), \end{aligned}$$

where the inclusion is by Lemmas 6 and 9. The interval shrinks to $0.5 = \psi_s(\theta, T)$.

Therefore,

$$p_{Ns}(\theta, T, \delta_N) \to \psi_s(\theta, T),$$

completing the proof of Lemma 8.

D Empirical Appendix

D.1 Additional Results

Table B1 reports estimates for differential attrition, computed by estimating the same models as in the statistical tests for balance presented in table 3. Under general risk, applicants who receive Grade A school offers have a slightly likelihood to have taken the SAT. Decomposing Grade A schools into screened and unscreened schools, applicants who receive Grade A lottery school offers are 2.4 percent more likely to have SAT scores, while offers to Grade A screened schools do not correspond to a statistically significant difference in the likelihood of having follow-up SAT scores. A modest difference that seems unlikely to bias the 2SLS Grade A estimates reported in tables 4 and 5.

Table B2 reports estimates of the effect of enrollment in an ungraded high school that correspond to the models presented in Table 4. As discussed above, estimates that consider general risk add applicants and schools to the analysis sample. Also for estimates of ungraded school effects, the move from lottery risk to general risk yields a valuable precision gain, as can be seen by comparing results reported in columns 4 and 5. For instance, the associated standard error falls from 0.054 when the graduation effect is estimated using only lottery risk to 0.034 when estimated by exploiting general risk in Grade A assignment. While the OLS estimates yield a small positive effect on SAT and a strong negative effects on graduation outcomes, the 2SLS estimates do not suggest any statistically significant effect of ungraded school attendance.

D.2 Bandwidth Computation and Robustness Checks

Bandwidths are estimated as suggested by Imbens and Kalyanaraman (2012), separately for each screened program s, where $v(s) \neq 0$ and $\exists$ applicants with $R_v(s) > \tau_s$. The bandwidth is estimated for the set of applicants who are in the relevant marginal priority group and are assigned $R_v(s)$ by s. Bandwidths are also computed separately for each outcome variable. We then use the minimum bandwidth across SAT and graduation outcomes.

To check the robustness of this procedure, we computed several checks that impose restrictions on the distribution of running variables and the bandwidth size. Table B3 reports 2SLS estimates of the effects of Grade A high school enrollment, computed by estimating the same models as in table 4, imposing several restrictions on the distribution of running variables. Estimates in column 3 are from a model that ignores general risk that is generated at screened programs with four or more duplicate running variable values in the estimated bandwidth, resulting in modest changes in sample and effect size. Estimates in column 4 are from a model that ignores general risk that is generated at screened programs with a gap in running variable positions within the estimated bandwidth of four or more, which leads to a somewhat larger drop in sample size. Columns 5 and 6 combine these restrictions, reducing the sample of applicants with general risk by 17 and 31 percent, respectively. The restrictions however leave effect size and precision mostly unchanged.

Table B4 reports 2SLS estimates for several variations of the bandwidth size and computation method. Halving the bandwidth size at screened programs reduces the sample of applicants with general risk by about 20 percent, while doubling of the bandwidth size results in an about 20 percent larger sample. The estimates in column 5 are from regressions that use sixth grade baseline scores instead of SAT and graduation outcomes when computing the bandwidth. Again, the variations leave the magnitude of Grade A school effects and precision mostly unchanged.

Table B1. Differential Attrition

	General Risk				Lottery Risk
	Non-offered	Grade A School			Any
	mean	Any	Screened	Lottery	Grade A
	(1)	(2)	(3)	(4)	(5)
Took SAT exam	0.761	0.017**	-0.002	0.017**	0.021**
		(0.007)	(0.012)	(0.007)	(0.008)
N		30,673	10,486	25,841	18,743
Has binary outcomes	0.635	0.003	0.003	0.002	0.003
(Enrolled in ninth grade)		(0.002)	(0.002)	(0.002)	(0.002)
N		30,673	10,486	25,841	18,743

Notes. This table reports differential attrition estimates, computed by regressing covariates on dummies indicating a Grade A offer and an ungraded school offer (columns (2)-(5)), controlling for saturated Grade A and ungraded school propensity scores and running variable controls (columns (2)-(4)). Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

Table B2. Ungraded School 2SLS Estimates

	All applicants		Applicants at risk		
				2SLS	
	Non-Enrolled Mean	OLS	Non-Offered Mean	General risk	Lottery risk
	(1)	(2)	(3)	(4)	(5)
			Panel A. First Stages		
SAT			0.308	1.75***	1.67***
(years of exposure)				(0.056)	(0.075)
Binary outcomes			0.138	0.597***	0.547***
(ever enrolled)				(0.021)	(0.028)
			Panel B. Second Stages		
SAT Math	470	0.603**	515	2.76	2.02
(200-800)	(102)	(0.191)	(109)	(1.87)	(2.54)
SAT Reading	470	0.761***	510	1.81	1.15
(200-800)	(91)	(0.178)	(93)	(1.79)	(2.41)
N		124,989		22,899	12,752
Graduated	0.611	-0.229***	0.793	0.062*	0.035
		(0.003)		(0.034)	(0.054)
College- and Career Prepared	0.365	-0.114***	0.587	0.051	0.120**
		(0.003)		(0.037)	(0.057)
College-ready	0.321	-0.073***	0.542	0.015	0.001
		(0.003)		(0.036)	(0.054)
N		121,074		19,150	11,200

Notes. This table reports the estimates of the effect of enrollment in an ungraded high school produced by the models that generate the Grade A enrollment effects reported in Table 4. Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

Table B3. Grade A Effects with Running Variable Restrictions

	Lottery risk	General risk				
		No RV Restriction	4+ Duplicates in BW	4+ Gap in BW	4+ Gap or 4+ Duplicates in BW	3+ Gap or 3+ Duplicates in BW
	(1)	(2)	(3)	(4)	(5)	(6)
SAT Math	2.17**	2.26**	1.97**	2.91***	2.68***	1.37*
(200-800)	(0.880)	(0.718)	(0.728)	(0.750)	(0.755)	(0.814)
SAT Reading	0.770	0.726	0.566	0.919	0.690	0.464
(200-800)	(0.801)	(0.658)	(0.669)	(0.687)	(0.695)	(0.741)
N	12752	22899	21740	20291	19273	16147
Graduation	0.022	0.030**	0.028**	0.026*	0.025*	0.024
	(0.018)	(0.013)	(0.014)	(0.014)	(0.015)	(0.017)
N	11200	19150	18056	16926	15999	13160

Notes. This table reports 2SLS estimates of the effects of Grade A high school enrollment as in table 4. Estimates in column 2 correspond to the estimates in column 4 in table 4 and impose no restriction on the distribution of running variables. Estimates in column 3 are from a model that excludes general risk that is created at screened programs with four or more duplicate running variable values in the bandwidth. Estimates in column 4 are from a model that excludes general risk that is created at screened programs with a gap in running variable ranks in the bandwidth of four ranks or larger. Estimates in column 5 are from a model that combines the two restrictions from columns 3 and 4, excluding general risk that is created at screened programs with either duplicates or a gap. Estimates in column 6 are from a model that applies a stricter version of the restriction in column 5. Estimates in columns 2 to 6 are from models that include running variable controls. Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

Table B4. Grade A Effects with Alternative Bandwidths

	Lottery risk	General Risk			
		Benchmark	Half Bandwidth Size	Double Bandwidth Size	Using Baseline Scores for BW Computation
	(1)	(2)	(3)	(4)	(5)
SAT Math	2.17**	2.26**	1.88**	1.88**	2.25**
(200-800)	(0.880)	(0.718)	(0.805)	(0.657)	(0.734)
SAT Reading	0.770	0.726	0.883	-0.087	1.17*
(200-800)	(0.801)	(0.658)	(0.733)	(0.607)	(0.670)
N	12752	22899	17975	27966	22005
Graduation	0.022	0.030**	0.028*	0.020*	0.034**
	(0.018)	(0.013)	(0.016)	(0.012)	(0.014)
N	11200	19150	15251	23110	18398

Notes. This table reports OLS and 2SLS estimates of the effects of Grade A high school enrollment as in table 4. Estimates in column 2 correspond to the estimates in column 4 in table 4 and impose no restriction on the distribution of running variables and estimation of bandwidths. Estimates in column 3 are from a model that halfes the size of the estimated bandwidth at screened programs. Estimates in column 4 are from a model that doubles the size of the estimated bandwidth. Estimates in column 5 are from a model where 6th grade baseline math and english test scores were used instead of outcomes to compute the IK bandwidth. Estimates in columns 2 to 5 are from models that include running variable controls. Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%