

NBER WORKING PAPER SERIES

IMPACT EVALUATION IN MATCHING MARKETS WITH GENERAL TIE-BREAKING

Atila Abdulkadiroglu
Joshua D. Angrist
Yusuke Narita
Parag A. Pathak

Working Paper 24172
<http://www.nber.org/papers/w24172>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2017

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. Tim Armstrong, Eduardo Azevedo and Whitney Newey provided helpful feedback. We're especially indebted to Adrian Blattner and Ignacio Rodriguez 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. Abdulkadiroglu 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Atila Abdulkadiroglu, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Impact Evaluation in Matching Markets with General Tie-Breaking
Atila Abdulkadiroglu, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak
NBER Working Paper No. 24172
December 2017
JEL No. C26,D47,I20

ABSTRACT

Many centralized matching schemes incorporate a mix of random lottery and non-lottery tie-breaking. A leading example is the New York City public school district, which uses criteria like test scores and interviews to generate applicant rankings for some schools, combined with lottery tie-breaking at other schools. We develop methods that identify causal effects of assignment in such settings. Our approach generalizes the standard regression discontinuity design to allow for many running variables and treatments, some of which are randomly assigned. We show that lottery variation generates assignment risk at non-lottery programs for applicants away from non-lottery cutoffs, while non-lottery variation randomizes applicants near cutoffs regardless of lottery risk. These methods are applied to evaluate New York City's school progress assessments, which give schools letter grades as a summary measure of quality. Our estimates reveal that although Grade A schools boost achievement, these gains emerge only for students who attend lottery schools. Attendance at a coveted Grade A screened school, including some of the highest performing in the district, generates no measurable effects. Evaluation methods that fail to take advantage of both lottery and non-lottery variation miss this difference in impact.

Atila Abdulkadiroglu
Department of Economics
Duke University
213 Social Sciences Building
Durham, NC 27708
and NBER
atila.abdulkadiroglu@duke.edu

Joshua D. Angrist
Department of Economics, E52-436
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and IZA
and also NBER
angrist@mit.edu

Yusuke Narita
Yale University
37 Hillhouse Avenue
Room 38
New Haven, CT 06511
yusuke.narita@yale.edu

Parag A. Pathak
Department of Economics, E52-426
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
ppathak@mit.edu

1 Introduction

Large school districts are increasingly turning to sophisticated matching mechanisms to allocate their seats (see, e.g., Pathak (2017)). In addition to producing fair and transparent admissions decisions, these mechanisms offer a unique resource for research and accountability: the data they generate can be used to construct unbiased estimates of school and sector 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, the tie-breaker becomes a source of quasi-experimental variation in school assignment.

Many districts break ties with a single random “lottery number.” As we’ve shown in Abdulkadiroğlu et al. (2017a), lottery tie-breaking assigns students to schools as if 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 high dimensional 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 school seats are offered to applicants conditional on their 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 (Rosenbaum and Rubin, 1983). Because the DA propensity score for a market with lottery tie-breaking depends on only a few school-level cutoffs, the score has a distribution much coarser than the distribution of applicant types that it conditions on.

We turn here to the problem of crafting research designs from a broader class of centralized mechanisms in which the tie-breaking variable is non-random and potentially correlated with unobserved potential outcomes. Such non-random tie-breaking fails to satisfy the equal-treatment-of-equals (ETE) property that’s key to the lottery-based research design detailed in Abdulkadiroğlu et al. (2017a).¹ Non-random tie-breaking—such as used for school assignment in Boston, Chicago, and New York City—raises important challenges for causal inference in matching markets. 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 hurdles 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 for

¹Non-lottery tie-breaking embedded in centralized assignment schemes has been used in studies of schools in Chile (Hastings et al., 2013; Zimmerman, 2016), Italy (Fort et al., 2016), Ghana (Ajayi, 2014), Kenya (Lucas and Mbiti, 2014), Norway (Kirkeboen et al., 2016), Romania (Pop-Eleches and Urquiola, 2013), Trinidad and Tobago (Jackson, 2010, 2012; Beuermann et al., 2016), and the U.S. (Abdulkadiroğlu et al., 2014; Dobbie and Fryer, 2014; Barrow et al., 2016). This earlier work fails to exploit the full extent of quasi-experimental variation induced by centralized assignment.

each applicant type. This means that the propensity score under general tie-breaking may be no coarser than the underlying type distribution. Moreover, with the distribution of tie-breakers unknown, we cannot easily estimate the propensity score by simulation. We solve these problems by integrating the non-parametric RD model introduced by Hahn et al. (2001) with the large-market model of matching markets used to study random tie-breaking in Abdulkadiroğlu et al. (2017a).² Local RD strategies are shown here as driving the propensity score to a constant value of one-half at cutoffs. This framework extends the domain of non-parametric RD identification strategies to assignment schemes with multiple running variables, some of which may be randomly assigned, and multiple cutoffs.

The utility of impact evaluation in matching markets with general tie-breaking is illustrated here through an investigation of school value-added in New York City (NYC) public schools. Specifically, we exploit school assignment variation generated by the NYC high school match, which assigns students to seats in a DA match that integrates distinct screened school tie-breaking with a common lottery tie-breaker. Screened schools are schools that use an entrance exam or other criteria to rank applicants instead of a lottery. The quasi-experimental assignment variation generated by this system is used to answer questions about school quality. Our results show that attendance at one of New York City’s “Grade A schools,” whether screened or lottery, has a modest effect on achievement, though not on graduation or other outcomes. Perhaps surprisingly, achievement gains emerge only for students who attend lottery schools; attendance at a coveted Grade A screened school, including some of the highest performing in the district, generates no measurable effects. Evaluation methods that fail to take advantage of both lottery and non-lottery variation miss this important difference in impact.

The next section introduces notation for DA and explains the large market sequence used here to derive an analytic propensity score. Section 3 presents results for markets with a single non-random tie-breaker, generalizing the formula for the DA propensity score derived in Abdulkadiroğlu et al. (2017a). Section 4 extends these results to markets using multiple non-identically distributed tie-breakers. This section also introduces the idea of a local or limiting propensity score defined in a neighborhood of school cutoffs and develops a theoretical framework that covers the mix of lottery and non-lottery tie-breakers used in the NYC school match. Section 5 applies these results to NYC’s school progress assessments. We conclude with a brief summary and an agenda for further work.

2 DA with General Tie-breaking

A school choice problem is defined by a set of applicants, schools, and school capacities. Applicants have strict preferences over schools while schools have priorities over applicants. For example, schools may prioritize applicants who live nearby or with currently enrolled siblings. Let $s = 0, 1, \dots, S$ 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 unit continuum, with applicants indexed by values in the unit interval. Seating is constrained by a capacity

²Large-market results for the special case of serial dictatorship with a single non-random tie-breaker are sketched in Abdulkadiroğlu et al. (2017b). See Lee and Lemieux (2010) for a survey of RD methods.

vector, $\mathbf{q} = (q_0, q_1, q_2, \dots, q_S)$, where q_s is defined as the proportion of I that can be seated at school s . We assume $q_0 = 1$.

Applicant i 's preferences over schools constitute a partial ordering of schools, \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 $\rho_{is} \in \{1, \dots, K, \infty\}$ 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 $\boldsymbol{\rho}_i = (\rho_{i1}, \dots, \rho_{iS})$ 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_θ and priorities ρ_θ . Θ denotes the set of possible types. An *assignment* $\boldsymbol{\mu} = \{\mu_i | i \in I\}$ specifies each applicant's assigned school or assignment to the outside option. The proportion of applicants assigned to school s is at most q_s .

A *mechanism* is a set of rules determining $\boldsymbol{\mu}$ as a function of preferences, priorities, and a set of tie-breaking variables. Tie-breakers can be random lottery numbers, in which case they're distributed independently of type, or variables like entrance exam scores, that depend on type. The distribution of tie-breakers, which we also refer to as *running variables* because they play a role like that of running variables in RD-style research designs, induces a distribution of assignments.

The mechanisms of interest to us are stochastic, generating a probability or "risk" of assignment for each applicant to each school. In practice, a match is executed by a computer program that can be run repeatedly (though may in practice be run only once). Assignment risk is produced by repeatedly drawing tie-breakers from the running variable distributions for each applicant and re-running the match program, fixing everything else. Randomness in a match is therefore induced by randomness in running variables in these repeated executions.

Consider first a market with a single running variable common to all schools, denoted R_i for applicant i . Although R_i is not necessarily a lottery number, we assume (without loss of generality) that it's distributed over $[0, 1]$, with cumulative distribution function F_R^i . Imagine, for example, a non-lottery tie-breaker such as a test score; this can be seen as scaled "percent right." This common support notwithstanding, running variables 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.

By the law of iterated expectations, the probability a type θ applicant has a running variable below any value r is $F_R(r|\theta) = E[F_R^i(r)|\theta_i = \theta]$, where $F_R^i(r)$ is F_R^i evaluated at r . To be concrete, imagine again 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 yields $F_R(r|\theta_0) \neq F_R(r|\theta_1)$.

A widely-seen special case is when R_i is a lottery number drawn independently from the same known distribution for all applicants, typically, a uniform distribution over $[0, 1]$. In this case, $F_R(r|\theta) = F_R^i(r) = r$ for any $r \in [0, 1]$ and for all i and θ .

Defining DA

Most school districts and college admissions using centralized assignment employ a version of the DA mechanism introduced by Gale and Shapley (1962). In school choice markets with a common tie-breaker, DA works like this:

Each applicant applies to his most preferred school. Each school ranks these applicants, 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 applies to his next most preferred school. Each school ranks these new applicants *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 applications (some applicants may remain unassigned).

A large body of theoretical work establishes the attractive properties of this mechanism, showing that it's sure to converge and is strategy-proof, and that the resulting allocations are stable in a sense detailed in Roth and Sotomayor (1990).

Let r_i be the realization of single tie-breaker R_i in a particular match. For any set of applicant types $\Theta_0 \subset \Theta$ and for any number $r_0 \in [0, 1]$, define the set of applicants in Θ_0 with tie-breaking numbers less than r_0 to be

$$I(\Theta_0, r_0) = \{i \in I \mid \theta_i \in \Theta_0, r_i \leq r_0\}.$$

We use the shorthand notation $I_0 = I(\Theta_0, r_0)$ to denote this set and denote the fraction of applicants in I_0 by $G(I_0)$.

We define DA using the notation above, nesting the finite-market and continuum cases. First, combine applicants' priority status and realized tie-breaking variable into a single number for each school, called *applicant rank* at school s :

$$\pi_{is} = \rho_{is} + r_i.$$

Since the difference between any two priorities is at least 1 and tie-breaking variables are between 0 and 1, rank is lexicographic in priority and the running variable.

DA can be described by a vector of admissions cutoffs that evolves through "rounds" (Abdulkadiroğlu et al., 2015; Azevedo and Leshno, 2016). We index these rounds by t and write the vector of admissions cutoffs in round t as $\mathbf{c}^t = (c_1^t, \dots, c_S^t)$. Eligible applicants with rank π_{is} smaller than c_s^t *qualify* for provisional admission at s in round t . Since no applicant is seated initially, cutoffs begin as the largest possible value of an eligible applicant's rank, $K + 1$, that is, $c_s^1 = K + 1$ for all s . In each round, each applicant *demand*s his most preferred school from among those at which he qualifies for provisional seating. Formally, the demand for seats at school s conditional on \mathbf{c}^t is defined as:

$$Q_s(\mathbf{c}^t) = \{i \in I \mid \pi_{is} \leq c_s^t \text{ and } s \succ_i \tilde{s} \text{ for all } \tilde{s} \in S \text{ such that } \pi_{i\tilde{s}} \leq c_{\tilde{s}}^t\}.$$

In other words, school s is demanded by applicants with rank below the cutoff at s , who prefer s to any other school at which they're also below the relevant cutoff.

Cutoffs evolve according to demand in each round as follows:

$$c_s^{t+1} = \begin{cases} K + 1 & \text{if } G(Q_s(\mathbf{c}^t)) < q_s, \\ \max \{x \in [0, K + 1] \mid G(\{i \in Q_s(\mathbf{c}^t) \text{ such that } \pi_{is} \leq x\}) \leq q_s\} & \text{otherwise;} \end{cases}$$

This expresses the idea that in each DA round, seats at s are provisionally offered to all who demand a seat, including applicants provisionally offered a seat in the previous round and applicants rejected from higher ranked choices, up to school capacity. The cutoff at s in round t is defined as the highest rank that can be provisionally seated at s in round t . If demand falls below capacity, everyone who demands a seat gets one and the cutoff is set to the highest possible rank.

The vector of cutoffs determining a DA allocation is defined as the limit

$$c_s = \lim_{t \rightarrow \infty} c_s^t; \quad s = 1, \dots, S.$$

The set of applicants that are assigned school s under DA is the demand for seats at these limiting cutoffs, that is, the set $\{i \in Q_s(\mathbf{c})\}$ where $\mathbf{c} = (c_1, \dots, c_S)$. Since $c_s \leq K + 1$, an ineligible applicant is never assigned to school s . Azevedo and Leshno (2016) and Abdulkadiroğlu et al. (2017a) show that the cutoff definition of DA is valid in the sense that: (a) the necessary limits exist for every economy, finite or continuum; (b) for every finite economy, the allocation upon convergence matches that produced by DA as described by Gale and Shapley (1962) and the many studies building on their work. In finite markets, cutoffs converge in a finite number of rounds but depend on realizations of the running variable. As a result, cutoffs in finite markets are stochastic. In our theoretical continuum market, however, cutoffs are fixed.

3 DA with a Single Non-lottery Tie-breaker

Suppose we'd like to estimate the causal effect of attendance at a particular school or group of schools relative to alternative schools. This task is complicated by the fact that school assignments reflect applicant preferences, school priorities, and running variables, all of which are likely to be correlated with outcomes like achievement or high school graduation. Conditional on applicant type, however, we need only worry about OVB from the running variable. We therefore start by exploiting the fact that OVB induced by correlation between type and the offer of a seat at school s is controlled by conditioning on a scalar function of type, the propensity score.

3.1 Characterizing the DA Propensity Score

The propensity score is the conditional probability of assignment to s , written

$$p_s(\theta) = E[D_i(s) \mid \theta_i = \theta],$$

where $D_i(s)$ indicates the offer of a seat at school s to applicant i .

Our characterization of $p_s(\theta)$ uses the notion of a *marginal priority* group at school s . The marginal priority group consists of applicants for whom seats are rationed by running variables

when a school is over-subscribed. Formally, marginal priority, ρ_s , is the integer part of the cutoff, c_s . Conditional on being rejected by all more preferred schools and on applying to school s , an applicant is assigned s with certainty if $\rho_{is} < \rho_s$, that is, if he clears 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 a school-specific *randomization cutoff*. The *randomization cutoff* at school s , denoted τ_s , is the decimal part of the cutoff at s , that is, $\tau_s = c_s - \rho_s$.

These observations motivate a partition of the set of applicant types according to marginal priorities at s . Let Θ_s denote the set of applicant types who rank s and divide Θ_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)*

The set Θ_s^n contains applicant types who have worse-than-marginal priority at s . No one in this group is assigned to s . Θ_s^a contains applicant types that 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 ranked more highly. Finally, Θ_s^c is the subset of Θ_s that is marginal at s , that is, the marginal priority group at s . These conditionally seated applicants are assigned s when they're not assigned a higher choice *and* have a running variable value that clears the randomization cutoff at s .

A second key component of our score formulation captures the fact that failure to qualify at schools other than s may truncate the distribution of running variables in the marginal priority group for school s . To characterize the distribution of running variables among those at risk of assignment to s , we introduce notation for the set of schools ranked above s . Applicants of type θ view the following set of schools as better than s :

$$B_{\theta s} = \{s' \in S \mid s' \succ_{\theta} s\}.$$

Type θ 's *most informative disqualification* (MID) tells us how the running variable distribution among applicants to s is truncated by offers at the schools these applicants prefer to s . For each type and school, $MID_{\theta s}$ is defined as a function of cutoffs at schools in $B_{\theta s}$:

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

$MID_{\theta s}$ is zero when type θ applicants have worse-than-marginal priority at all higher ranked schools: when no applicants for s can be seated at a more preferred school, there's no running variable truncation among those at risk of assignment to s . On the other hand, when type θ applicants are always seat in at least one school in $B_{\theta s}$ truncation is complete, and $MID_{\theta s} = 1$.

The third line in the definition of $MID_{\theta s}$ reflects the fact that, among applicants for whom $\rho_{\theta b} = \rho_b$ for some $b \in B_{\theta s}$, anyone who fails to clear 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 fail to qualify at schools with cutoffs below 0.5. Consequently, to keep track of the truncation induced by disqualification at all schools an applicant of type θ prefers to s , we need to record only the most forgiving cutoff that such applicants fail to clear.

This machinery allows us to describe offers of a seat at s as determined by bounds on the running variable. Applicant i in Θ_s^a is seated at s provided $r_i > MID_{\theta s}$, that is, when the applicant fails to clear the relevant MID. Applicant i in Θ_s^c is seated at s if and only if $\tau_s > r_i > MID_{\theta s}$, that is, when his lottery number clears the randomization cutoff at s but not his MID. The DA propensity score is determined by the probability of these events.

In real-world matching markets, evaluation of cutoff-clearing probabilities is complicated by the fact that cutoffs are stochastic and correlated with individual running variables. We therefore turn to a large market model that expands the number of applicants and capacities to infinity while keeping the set of types fixed. Abdulkadiroğlu et al. (2017a) presents evidence on the performance of this large-market approximation, showing that estimates of large market scores effectively balance conditioning variables. In the large market model, DA cutoffs, and hence marginal priority groups and $MID_{\theta s}$, are fixed. The probability of events like $\tau_s > r_i > MID_{\theta s}$ then depends only on the distribution $F_R(r|\theta)$ evaluated at $MID_{\theta s}$ and τ_s . Abdulkadiroğlu et al. (2017a) uses this property to derive the DA score for markets with a single lottery tie-breaker, that is, for markets in which $F_R(r|\theta) = r$.

Theorem 1, below, extends this result to produce a score formula for DA using a common running variable with an unspecified tie-breaker distribution.

Theorem 1. *Consider a continuum economy populated by applicants of type $\theta \in \Theta$ to be assigned to schools indexed by $s \in S$. These schools use a single tie-breaker, R_i , distributed according to $F_R(r|\theta)$ for type θ . For all s and θ in this economy, we have:*

$$p_s(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_s^n \\ 1 - F_R(MID_{\theta s}|\theta) & \text{if } \theta \in \Theta_s^a \\ (1 - F_R(MID_{\theta s}|\theta)) \times \max \left\{ 0, \frac{F_R(\tau_s|\theta) - F_R(MID_{\theta s}|\theta)}{1 - F_R(MID_{\theta s}|\theta)} \right\} & \text{if } \theta \in \Theta_s^c \end{cases}$$

where $p_s(\theta) = 0$ when $MID_{\theta s} = 1$ and $\theta \in \Theta_s^c$.

Theorem 1 is a special case of Theorem 2, discussed below and proved in Appendix A.2.2. Because cutoffs in the large market are fixed, the risk of finding a seat at s for applicants in Θ_s^a is determined solely by the probability of drawing r_i above the relevant MID, while risk for applicants in Θ_s^c is determined by the probability of drawing r_i below the cutoff at s , conditional on failing to clear the relevant MID. The max function in the third line of the score arises from the fact that when $MID_{\theta s}$ exceeds the cutoff at s , failure to clear MID implies disqualification at s .

Some school districts use a version of DA without priorities; Boston and Chicago exam schools are leading examples. Without priorities, DA simplifies to a mechanism known as *serial dictatorship* (SD), where the serial order comes from an exam score. Abdulkadiroğlu et al. (2017b) gives the large market score for SD with single non-lottery tie-breaker. This result is a special case of Theorem 1, stated here as a corollary:

Corollary 1 (Abdulkadiroğlu et al. (2017b)). *Consider a continuum economy with no priorities, populated by applicants of type $\theta \in \Theta$, to be assigned to schools indexed by $s \in S$. These schools use a single tie-breaker, R_i , distributed according to $F_R(r|\theta)$ for type θ . For all s and θ in this economy, we have:*

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

Without priorities, Θ_s^n and Θ_s^a are empty. The probability of assignment at s is therefore determined solely by draws from the truncated distribution of running variables remaining after eliminating applicants seated at schools they’ve ranked more highly. This is the probability on the third line of Theorem 1.

Many applications of DA use a single uniformly distributed tie-breaker. This produces the special case of Theorem 1 derived in Abdulkadiroğlu et al. (2017a). Uniform tie-breaking implies $F_R(r) = r$, yielding the following result as an additional corollary:

Corollary 2 (Abdulkadiroğlu et al. (2017a)). *Consider a continuum economy populated by applicants of type $\theta \in \Theta$, to be assigned to schools indexed by $s \in S$. These schools use a single tie-breaker distributed uniformly for all types, θ . For all s and θ in this economy, we have:*

$$p_s(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_s^n \\ 1 - MID_{\theta_s} & \text{if } \theta \in \Theta_s^a \\ (1 - MID_{\theta_s}) \times \max \left\{ 0, \frac{\tau_s - MID_{\theta_s}}{1 - MID_{\theta_s}} \right\} & \text{if } \theta \in \Theta_s^c \end{cases}$$

where $p_s(\theta) = 0$ when $MID_{\theta_s} = 1$ and $\theta \in \Theta_s^c$.

Estimates of the DA propensity score in this case require only an estimate of MID_{θ_s} and a partition of types into those that are never, always, and conditionally seated. This scenario contrasts with the scores in Theorem 1 and Corollary 1, which depend on distributions $F_R(\tau_s|\theta)$ and $F_R(MID_{\theta_s}|\theta)$. We tackle the problem of estimating these unknown distributions below.

4 Mixed Multiple Tie-Breaking

Many urban school districts incorporate an assortment of tie-breaking rules in a partially or completely centralized DA match. Some districts centralize particular sectors. For example, selective exam schools in Boston, Chicago, and NYC use test scores to rank their applicants in an exam-sector match. In addition to test-based tie-breaking at exam schools, NYC runs a centralized traditional-sector match that incorporates a variety of tie-breakers for its other diverse “screened schools” (Abdulkadiroğlu et al., 2005, 2009). These schools rank their applicants using tie-breakers derived from interviews, auditions, or GPA in earlier grades, as well as test scores.

The NYC match also includes many schools that use a common lottery tie-breaker. NYC’s combination of screened and lottery tie-breaking motivates our interest in the DA propensity score for markets with mixed multiple tie-breakers. We show here how a complex menu of tie-breaking rules generates research designs that compare a wide range of schools. These results

are then applied to assess the NYC Department of Education (DOE) school assessment system, which grades schools A-F as a measure of their quality. Our analysis estimates causal effects of attending a Grade A school on achievement and high school graduation, thereby revealing whether highly rated schools indeed improve outcomes for their students.

Formal analysis of multiple tie-breaking requires additional notation to keep track of different tie-breakers. Let $v \in \{0, 1, \dots, V\}$ index running variables and let $\{S_v : v \in \{0, 1, \dots, V\}\}$ be a partition of schools such that S_v is the set of schools using tie-breaker v . As a result, 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 running variable at schools in S_v . For any v and students $i \neq j$, R_{iv} and R_{jv} are assumed to be independent, though not necessarily identically distributed. Likewise, for $v \neq v'$, running variables R_{iv} and $R_{iv'}$ are initially assumed to be independent (an assumption relaxed shortly, below). Define function $v(s)$ such that $s \in S_{v(s)}$; that is, $v(s)$ is the index of the running variable used at school s . We adopt the convention that $v = 0$ identifies the lottery tie-breaker and therefore S_0 constitutes the set of lottery schools. The algorithm that implements DA in this multiple tie-breaker setting is described in Appendix A.1.³

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

$$F_v(r|\theta) = E[F_v^i(r)|\theta_i = \theta]. \quad (1)$$

This is the fraction of type θ applicants whose running variable for the set of schools in S_v is less than or equal to r . We again assume running variables have support on $[0, 1]$. This allows us to work with randomization cutoffs in $[0, 1]$. Importantly, however, the distribution of normalized R_{iv} remains dependent on type. In other words, normalization to the unit interval does not simplify the conditional distribution defined in (1) to a distribution common to all types θ .

With multiple tie-breakers, any truncation of running variable distributions by qualification at higher-ranked choices is specific to R_{iv} . We therefore define running-variable-specific MIDs for schools in set S_v , denoted MID^v for $v \geq 0$. To this end, partition B_{θ_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:

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

$MID_{\theta_s}^v$ is a key determinant of the propensity score in large markets with mixed multiple tie-breakers:

Theorem 2. *Consider a continuum economy populated by applicants of type $\theta \in \Theta$, to be assigned to schools indexed by $s \in S$. These schools use a set of tie-breakers distributed independently of*

³Real-world tie-breakers, including those in New York City, often take the form of 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. This dependence vanishes as the number of applicants being ranked grows.

one another according to $F_v(r|\theta)$ for type θ and $v \in \{0, 1, \dots, V\}$. For all s and θ in this economy,

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

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

Theorem 2 says that the risk of being seated at s for applicants in the always-seated set is given by the risk of failing to find a seat at any more preferred school. Since running variables are assumed independent, this probability is described by a product term, $\prod_v (1 - F_v(MID_{\theta_s}^v|\theta))$. For applicants in Θ_s^c , assignment risk multiplies the probability of not being seated at a more preferred school by the conditional probability of drawing a school s running variable that clears the cutoff at s , conditional on failing to clear the relevant MID at s , denoted $MID_{\theta_s}^{v(s)}$. Because running variables are assumed to be independent, it's only $R_{iv(s)}$ that is truncated by failing to win a seat at more preferred schools.⁴

This result generalizes Theorem 1 to cover a market with multiple tie-breakers, not all uniformly distributed or independent of type. Theorem 2 covers mixed multiple tie-breaking because one or more of the tie-breakers indexed by v might be uniformly distributed, as in Denver's unified district-charter school match analyzed by Abdulkadiroğlu et al. (2017a), while allowing for a mix of lottery and screened schools employing a variety of non-random tie-breakers, as in the NYC match.

4.1 Local-to-cutoff Propensity Scores

The propensity score identifies the largest sample of applicants and schools for which there is within-type variation in treatment assignment (these applicants have a score strictly between 0 and 1). Moreover, as first shown by Rosenbaum and Rubin (1983), control for the propensity score as a function of covariates, X , eliminates X as a source of OVB. In evaluation research using the propensity score to control for applicant type, score conditioning therefore eliminates bias from correlation between type and potential outcomes. Score conditioning accomplishes this while also solving a dimensionality problem: since applicants can rank many schools, there are almost as many types as applicants. In markets with a single random tie-breaker, however, the score given by Corollary 2 takes on far fewer values. Propensity score control therefore yields a practical empirical research design.

Theorem 2 leaves us with three empirical challenges not encountered in earlier work using the score to eliminate OVB in school research. First, with non-random tie-breakers like test scores, the conditional running variable distributions, $F_v(\cdot|\theta)$, are likely to depend on θ , so the score in Theorem 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 $MID_{\theta_s}^v$. Second, $F_v(\cdot|\theta)$ is typically

⁴The cross-running variable independence assumption is relaxed in the derivation of a local propensity score, which defines large-market risk for applicants near cutoffs.

unknown. This precludes straightforward computation of the propensity score by repeatedly sampling from $F_v(.|\theta)$ (i.e., we cannot easily compute what Abdulkadiroğlu et al. (2017a) call a “simulated score”).

Finally, while control for the propensity score eliminates confounding from type, assignments are a function of running variables as well as type, and screened-school running variables are likely to be correlated with potential outcomes. In other words, centralized assignment with screened schools fails to satisfy the ETE property satisfied by lottery tie-breaking (ETE says applicants i and j are equally likely to be offered a seat at school s when $\theta_i = \theta_j$).

Our solution to these problems builds on Hahn et al. (2001) and a host of related work developing non-parametric RD identification strategies. These strategies eliminate running variable confounding by looking at applicants with running variable values near cutoffs. This also turns out to solve the problem of unknown and high-dimensional $F_v(.|\theta)$. Specifically, our analysis of the assignment risk generated by applications to screened schools focuses on subsets of applicants with running variable values near these schools’ cutoffs. This yields a *local propensity score*. By contrast, assignment risk at lottery schools is global and not limited to variation near cutoffs. Perhaps surprisingly, however, assignment risk at screened schools also generates assignment risk at lottery schools, even for applicants who clear marginal priority at the latter.

Understanding Screened School Risk

We illustrate these ideas in a simple market with three schools, A, B, and C. A is a screened school, while B and C are lottery schools. Applicants differ in their preferences over B and C, but all rank A first, a prestigious selective school. Suppose applicants who prefer B over C tend to have better outcomes. We’d like to eliminate this difference in preferences as a source of OVB in comparisons meant to reveal causal effects of attendance at A. Likewise, students with higher test scores are more likely to get into school A, another source of OVB in any analysis of the effects of being offered a seat at A.

The running variable for applicant i at A is R_i . Since school A is ranked first, $MID_{\theta A} = 0$ for all θ (there is no truncation by qualification at higher-ranked schools). The SD propensity score in Corollary 1 then simplifies to

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

Although elementary, this score depends on the unknown distribution $F_R(\tau_A|\theta)$ and is therefore an unrestricted function of θ . This global dependence on type notwithstanding, the following proposition shows that for sets of applicants in a δ -neighborhood of a fixed running variable value, the score in this case converges to either 0, 0.5, or 1 as δ goes to zero. This result opens the door to a non-parametric RD-style analysis of offer variation around screened school cutoffs.

The fact that the score can be simplified in this way solves two problems raised by an evaluation of school attendance. First, for any school s , conditioning on a local score makes assignment risk locally independent of applicant type. Second, and perhaps less obviously, conditioning on the score eliminates OVB from running variables that are correlated with potential outcomes. To make this conditional independence statement precise, let W_i denote any applicant characteristic other than type and let W_{0i} be the potential value of W_i that is revealed when $D_i(s) = 0$,

while W_{1i} is the potential value revealed when $D_i(s) = 1$. These two potential values might be the same, as for covariates (race is unchanged by school assignment) or for test scores when assignment has no effect on achievement. In cases where they differ, as for outcomes affected by treatment, only one is seen in a given assignment realization. We say that W_i is *fixed under re-randomization* if $W_{0i} = W_{1i}$ for all i . Proposition 1 below shows that local-to-cutoff comparisons make assignment risk independent of any re-randomization-invariant W_i as well as of type.

Proposition 1. *Let $F_R(\cdot|\theta, w) = E[F_R^i(\cdot)|\theta_i = \theta, W_i = w]$ and assume that $F_R(\cdot|\theta, w)$ is differentiable at τ_s with $F'_R(\tau_s|\theta, w) \neq 0$ for every θ and any w . Also, define*

$$\lambda_s(\theta, w; r, \delta) = E[1(R_i \leq \tau_s)|\theta_i = \theta, W_i = w, R_i \in (r - \delta, r + \delta)].$$

Then, for every θ and any w ,

$$p_s(\theta, w; r) \equiv \lim_{\delta \rightarrow 0} \lambda_s(\theta, w; r, \delta) = \begin{cases} 1 & \text{if } r < \tau_s \\ 0.5 & \text{if } r = \tau_s \\ 0 & \text{if } r > \tau_s. \end{cases} \quad (2)$$

Proof. If $r < \tau_s$, then for any $\delta \in (0, \tau_s - r)$ and for all θ and w , we have

$$R_i \in (r - \delta, r + \delta) \Rightarrow R_i < \tau_s,$$

which implies

$$p_s(\theta, w; r) = \lim_{\delta \rightarrow 0} \lambda_s(\theta, w; r, \delta) = 1.$$

Similarly, if $r > \tau_s$, then for any $\delta \in (0, r - \tau_s)$, we have

$$R_i \in (r - \delta, r + \delta) \Rightarrow R_i > \tau_s,$$

which implies

$$p_s(\theta, r) = 0.$$

Finally, suppose $r = \tau_s$. We have

$$\begin{aligned} p_s(\theta, w; r) &= \lim_{\delta \rightarrow 0} \frac{P(r - \delta < R_i \leq \tau_s | \theta_i = \theta, W_i = w)}{P(r - \delta < R_i < r + \delta | \theta_i = \theta, W_i = w)} \\ &= \lim_{\delta \rightarrow 0} \frac{F_R(\tau_s | \theta, w) - F_R(\tau_s - \delta | \theta, w)}{F_R(\tau_s + \delta | \theta, w) - F_R(\tau_s - \delta | \theta, w)} \\ &= \lim_{\delta \rightarrow 0} \frac{\{F_R(\tau_s | \theta, w) - F_R(\tau_s - \delta | \theta, w)\} / \delta}{\{F_R(\tau_s + \delta | \theta, w) - F_R(\tau_s | \theta, w)\} / \delta + \{F_R(\tau_s | \theta, w) - F_R(\tau_s - \delta | \theta, w)\} / \delta} \\ &= \frac{F'_R(\tau_s | \theta, w)}{2F'_R(\tau_s | \theta, w)} = 0.5, \end{aligned}$$

where the last line uses $F'_R(\tau_s | \theta, w) \neq 0$. This completes the proof. \square

Although we've used the shorthand $p_s(\theta, w; r)$ for the limiting score, the right hand side of equation (2) is unrelated to an applicant's type or running variable (though it varies with the choice of r). Proposition 1 therefore generates a local conditional independence relation, a result formalized below:

Corollary 3 (Local Conditional Independence). *Let $D_i(s) = 1$ if and only if $1(R_i \leq \tau_s)$. Then*

$$\lim_{\delta \rightarrow 0} P[D_i(s) = 1 | \theta_i = \theta, W_i = w, R_i \in (r - \delta, r + \delta), p_s(\theta, w; r) = p] = p.$$

Proof. Write $\lambda_s(\theta_i, W_i; r, \delta)$ and $p_s(\theta_i, W_i; r)$ for $\lambda_s(\theta, w; r, \delta)$ and $p_s(\theta, w; r)$ with unspecified conditioning arguments. For any realized $p_s(\theta_i, W_i; r)$, we have that

$$\begin{aligned} \lim_{\delta \rightarrow 0} E[D_i(s) | \theta_i, W_i, R_i \in (r - \delta, r + \delta); p_s(\theta_i, W_i; r)] \\ = \lim_{\delta \rightarrow 0} \lambda_s(\theta_i, W_i; r, \delta) \\ = p_s(\theta_i, W_i; r). \end{aligned}$$

The first equality follows from the fact that $p_s(\theta, w; r)$ is written as a function of θ, w , and the event $R_i \in (r - \delta, r + \delta)$, which are also in the conditioning argument. The second uses the definition of $p_s(\theta_i, W_i; r) = p$. \square

Proposition 1 and Corollary 3 imply that in this simple market, comparisons of applicants just above and just below the cutoff at s identify an average causal effect of being seated at s . This is a version of an identification result presented as Theorem 2 in Hahn et al. (2001). Here, identification follows from the conditional independence of assignments and potential outcomes implied by the corollary. Intuitively, this is because conditional independence means that applicants who are and are not treated have the same average counterfactual outcomes (see Rubin (1977) for a formal statement). The Hahn et al. (2001) identification proof relies on continuity of conditional expectations, which can be shown to be implied by our regularity conditions. As in Hahn et al. (2001) and related theoretical work, this a limiting result, which holds only for observations right at the cutoff. Even so, a growing body of empirical work suggests we can eliminate OVB using empirical strategies that look at data close to cutoffs. In practice, closeness can be determined using the data-driven bandwidths derived in Imbens and Kalyanaraman (2012).

4.2 Assignment Risk with General Tie-Breaking

An analysis of general multiple tie-breaker markets requires additional notation. For any vector of screened-school running variable values $\mathbf{r} = (r_v)_{v=1, \dots, V} \in \mathbb{R}^V$, let

$$N(\mathbf{r}, \delta) = \{y \in \mathbb{R}^V : y_v \in (r_v - \delta, r_v + \delta) \text{ for all } v = 1, \dots, V\}$$

denote a δ -neighborhood of \mathbf{r} for $\delta > 0$. The probability of an offer from school s for applicants with screened school running variables in $N(\mathbf{r}, \delta)$ is

$$p_s(\theta, \mathbf{r}, \delta) = E[D_i(s) | \theta_i = \theta, \mathbf{R}_i \in N(\mathbf{r}, \delta)],$$

where $\mathbf{R}_i \equiv (R_{iv})_{v=1, \dots, V}$. Note that s may be a lottery school or a screened school. Let

$$p_s(\theta, \mathbf{r}) = \lim_{\delta \rightarrow 0} p_s(\theta, \mathbf{r}, \delta)$$

be the local propensity score for assignment to school s . The notation $p_s(\theta, \mathbf{r})$ reflects the fact that this general limit score depends (coarsely) on type as well as on \mathbf{r} .

With multiple running variables, we invoke the following regularity condition, which says that conditional running variable distributions are differentiable and that DA cutoffs are distinct:

Assumption 1. (a) $F_v(\cdot | \theta_i = \theta, \mathbf{R}_i \in N(\mathbf{r}, \delta))$ is differentiable and strictly increasing for every v, \mathbf{r} , and $\delta > 0$, and (b) $\tau_s \neq \tau_{s'}$ for any $s \neq s'$.

In this setting, the local propensity score for assignment of type θ applicants to school s is determined in part by the number of screened schools preferred to s at which these applicants are in the marginal priority group with a tie-breaker close to randomization cutoffs. As in the simple market described above, “close” in the limit means running variables right at cutoffs. With multiple tie-breakers, we must count the number of such close encounters using running-variable-specific ranks. For any $\mathbf{r} \in \mathbb{R}^V$, let

$$m_{\theta s}(\mathbf{r}) = |\{b \in \cup_{v \geq 1} B_{\theta s}^v : \rho_{\theta b} = \rho_b \text{ and } r_{v(b)} = \tau_b\}|$$

be the number of preferred-to- s screened schools b at which θ is marginal and where $r_{v(b)}$ equals the cutoff for b . $m_{\theta s}(\mathbf{r})$ is the number of schools where type θ applicants with running variables equal to the vector \mathbf{r} face non-degenerate RD risk in the limit. As expressed in equation (2) for the single-school case, applicants are offered seats at each such school with probability one-half.

The local propensity score depends on the position of applicants’ screened-school tie-breakers relative to screened-school cutoffs, as well as on applicant type. We therefore extend the partition of Θ_s as follows. Never seated applicants are classified in this case according to:

$$\Theta_s^n = \left\{ (\theta, \mathbf{r}) \mid \rho_{\theta s} > \rho_s \text{ or, if } v(s) \neq 0, (\rho_{\theta s} = \rho_s \text{ and } r_{v(s)} > \tau_s) \right\}, \quad (3)$$

where the first condition refers to applicants to lottery and screened schools whose priority $\rho_{\theta s}$ is worse than marginal priority ρ_s , and the second refers to screened- s applicants who (i) have marginal priority at s , and (ii) whose running variable $r_{v(s)}$ exceeds the relevant cutoff. Always seated applicants are classified according to:

$$\Theta_s^a = \left\{ (\theta, \mathbf{r}) \mid \rho_{\theta s} < \rho_s \text{ or, if } v(s) \neq 0, (\rho_{\theta s} = \rho_s \text{ and } r_{v(s)} < \tau_s) \right\}, \quad (4)$$

where the first condition refers to applicants to lottery and screened schools who clear marginal priority ρ_s , and the second refers to screened- s applicants who (i) have marginal priority at s , and (ii) whose running variable $r_{v(s)}$ is below the relevant cutoff. Finally, conditionally seated applicants are classified according to:

$$\Theta_s^c = \left\{ (\theta, \mathbf{r}) \mid \rho_{\theta s} = \rho_s \text{ and, if } v(s) \neq 0 \text{ then } r_{v(s)} = \tau_s \right\}, \quad (5)$$

where the first condition, which holds for both lottery and screened schools, refers to applicants with marginal priority at s , and the second, which is only relevant when s is screened, identifies cases where the running variable equals the relevant cutoff.

These preliminaries lead to a general theorem characterizing local assignment risk in markets with mixed multiple tie-breakers. Note also that, as in the single-school example, risk is determined by the choice of evaluation point, \mathbf{r} . Applicants in the general case are “local” to a vector of hypothesized cutoff values, \mathbf{r} , which includes, but is not limited to, a value for the school- s tie-breaker, $r_{v(s)}$.

Theorem 3. Consider a continuum economy populated by applicants of type $\theta \in \Theta$ to be assigned to schools indexed by $s \in S$ and suppose Assumption 1 holds. If $(\theta, \mathbf{r}) \in \Theta_s^n$ or $(\theta, \mathbf{r}) \in \Theta_b^a$ for some $b \in B_{\theta s}$, the local propensity score is

$$p_s(\theta, \mathbf{r}) = 0.$$

Otherwise,

$$p_s(\theta, \mathbf{r}) = \begin{cases} 0.5^{m_{\theta s}(\mathbf{r})}(1 - MID_{\theta s}^0) & \text{if } (\theta, \mathbf{r}) \in \Theta_s^a \\ 0.5^{m_{\theta s}(\mathbf{r})} \max\{0, \tau_s - MID_{\theta s}^0\} & \text{if } (\theta, \mathbf{r}) \in \Theta_s^c \text{ and } v(s) = 0 \\ 0.5^{1+m_{\theta s}(\mathbf{r})}(1 - MID_{\theta s}^0) & \text{if } (\theta, \mathbf{r}) \in \Theta_s^c \text{ and } v(s) \neq 0. \end{cases} \quad (6)$$

This theorem starts with a scenario where applicants to s are either sure to do better or are never seated at s and therefore face zero risk at s . In this case, we need not worry about whether s is a screened or lottery school. In other scenarios, where applicants do not do better than 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 (each screened school qualification occurs with probability 0.5 for marginal applicants). These sources of risk combine to produce the first line of (6). 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 produces the addition of 1 to the exponent in the third line of equation (6).

Theorem 3 also yields a general conditional independence relation, which can be written like that in Corollary 3, with the modification that the assignment variable, $D_i(s)$, is generated by the general market structure covered by the theorem and

$$\lim_{\delta \rightarrow 0} P[D_i(s) = 1 | \theta_i = \theta, W_i = w, \mathbf{R}_i \in N(\mathbf{r}, \delta), p_s(\theta, w; \mathbf{r}) = p] = p_s(\theta, \mathbf{r}). \quad (7)$$

The limiting score in this case depends on applicant type, though in a restricted manner. Conditional independence of school assignment and potential outcomes in turn allows us to eliminate OVB by conditioning on $p_s(\theta, \mathbf{r})$ for applicants in a neighborhood of \mathbf{r} .

Isolating Lottery Risk

An important implication of Theorem 3 is that application to lottery schools creates assignment risk at screened schools. This risk is revealed by the formula for screened school propensity scores computed for values of \mathbf{r} away from cutoffs. In this case, $m_{\theta s}(\mathbf{r}) = 0$, and applicants at all screened schools are in either Θ_s^a or Θ_s^n for any s that is screened. In other words, this scenario turns screened-school tie-breakers into priorities. Consequently, the risk of assignment to a screened school s can be simplified to

$$p_s(\theta, \mathbf{r}) = \begin{cases} 0 & \text{if } (\theta, \mathbf{r}) \in \Theta_s^n \text{ or } (\theta, \mathbf{r}) \in \Theta_s^a \text{ for some } b \in B_{\theta s} \\ (1 - MID_{\theta s}^0) & \text{if } (\theta, \mathbf{r}) \in \Theta_s^a \text{ and } (\theta, \mathbf{r}) \in \Theta_s^n \text{ for all } b \in \cup_{v \geq 1} B_{\theta s}^v. \end{cases} \quad (8)$$

Lottery risk creates risk at screened schools because students who rank lottery schools ahead of screened schools need not qualify for lottery-based admission; this risk is $1 - MID_{\theta s}^0$. In

this same no-screened-risk scenario, the risk of assignment to any lottery school simplifies to the expression in Corollary 2 of Theorem 1, with the proviso that the relevant MID_{θ_s} is $MID_{\theta_s}^0$, as determined by the common lottery tie-breaker in the more general market. Thus, lottery risk is enough to evaluate screened and unscreened schools in a market where many applicants rank schools of both types. We explore this idea in the empirical application detailed in the next section, while also documenting the efficiency and school-coverage gains generated by adding screened school risk to the econometric mix.

Table 1 illustrates Theorem 3 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 ranks 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 ranked first, so this applicant's $MID_{\theta_A}^0 = 0$ (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, $MID_{\theta_B}^0 = 0.8$ because $\tau_A = 0.8$. But this applicant faces no risk at B because her priority places her among the never seated, that is, her type is such that she's in Θ_B^n (type classifications for lottery risk appear in column 2). Finally, at third choice school C, this applicant is always seated, that is, she is in Θ_C^a . Since the applicant's $MID_{\theta_C}^0$ is determined by the cutoff at A, the probability of being assigned C is $1 - MID_{\theta_C}^0 = 0.2$. 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. This places the applicant in Θ_D^c and 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_D^n (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_E^n . 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_E^a . 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 $m_{\theta_F}(\mathbf{r}) = |\{D\}| = 1$. Note also that $MID_{\theta_F}^0 = 0$ 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:

$$p_F(\theta, \mathbf{r}) = 0.5^{m_{\theta_F}(\mathbf{r})}(1 - MID_{\theta_F}^0) = 0.5 \times (1 - 0) = 0.5.$$

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_G^c (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_G^a , 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 $MID_{\theta H}^0 = 1$, since he’s always seated at his first choice. The applicant’s lottery propensity score is therefore zero at H and all lower-ranked 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 $m_{\theta I}(\mathbf{r}) = |\{G\}| = 1$, and $MID_{\theta I}^0 = 0$. The propensity score for assignment to I is therefore:

$$p_I(\theta, \mathbf{r}) = 0.5^{m_{\theta I}(\mathbf{r})}(\tau_I - MID_{\theta I}^0) = 0.5 \times (0.6 - 0) = 0.3.$$

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

$$p_J(\theta, \mathbf{r}) = 0.5^{m_{\theta J}(\mathbf{r})}(\tau_J - MID_{\theta J}^0) = 0.5 \times (0.8 - 0.6) = 0.1.$$

Finally, last-ranked 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 ranked choices. Note that $MID_{\theta K}^0 = 0.8$ because $\tau_J = 0.8$ is the most forgiving cutoff at lottery schools ranked ahead of K. Since $m_{\theta K}(\mathbf{r}) = |\{G\}| = 1$, the propensity score for general risk at K is therefore

$$p_K(\theta, \mathbf{r}) = 0.5^{m_{\theta K}(\mathbf{r})}(1 - MID_{\theta K}^0) = 0.5 \times (1 - 0.8) = 0.1.$$

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 ranked.⁵

Screening in Serial Dictatorship

Some districts assign seats at some schools in a DA match with no priorities and a common tie-breaker with an unknown distribution. We call this match *screened serial dictatorship* (SSD). Abdulkadiroğlu et al. (2017b) show that the Chicago assignment mechanism for the district’s exam schools can be implemented as a SSD. Other SSD settings are discussed in Kirkeboen et

⁵The examples in Table 1 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 rank few schools or rank only schools with very limited capacity, may remain unassigned.

al. (2016); Hastings et al. (2013); Lucas and Mbiti (2014); Pop-Eleches and Urquiola (2013); Abdulkadiroğlu et al. (2014); Dobbie and Fryer (2014).

Under SSD, we can drop superscript v in MID_{θ_s} and write the single running variable placeholder as the scalar r , as in Section 3. The SSD propensity score, originally presented as Proposition 2 in Abdulkadiroğlu et al. (2017b) as part of an analysis of CPS exam schools, can be derived as an implication of Theorem 3.

Corollary 4. *Suppose Assumption 1 holds for a continuum market assigning seats by SSD. For all s and θ in this market,*

$$p_s(\theta, r) = \begin{cases} 0 & \text{if } \tau_s < MID_{\theta_s}, \\ 0.5 & \text{if } MID_{\theta_s} < \tau_s, \end{cases}$$

for $r = MID_{\theta_s}, \tau_s$, and

$$p_s(\theta, r) = \begin{cases} 1 & \text{if } r \in (MID_{\theta_s}, \tau_s), \\ 0 & \text{otherwise} \end{cases}$$

for other values of r .

4.3 Estimating the Local Propensity Score

We're interested in the asymptotic behavior of propensity score estimates based on Theorem 3. A sample analog of the local score is shown here to converge (almost surely) in market size to the propensity score defined in Theorem 3. This explains in part why we expect the sample analog of the local score to produce ignorable offers in real finite markets. Our empirical application validates this conjectured good performance by showing that applicant characteristics are balanced by offer status conditional on estimates of the local propensity score.

The asymptotic sequence for the estimated score works as follows: draw a random sample of n applicants, i_1, i_2, \dots, i_n , from a continuum economy with a unit mass of applicants $[0, 1]$, augmented with school capacities $\{q_s\}$, type distribution G , and running variable distributions $(F_v^i)_{i \in [0,1], v=0,1,\dots,V}$ with $F_0^i = U[0, 1]$ for every i . We think of i_1, i_2, \dots, i_n as a set of applicant identity numbers drawn from $U[0, 1]$. These are renumbered below, labeling $i = i_1, i_2, \dots, i_n$ by $i = 1, 2, \dots, n$. For each sampled applicant $i = 1, 2, \dots, n$ and each running variable $v \in \{0, 1, \dots, V\}$, draw a realization from F_v^i . This sampling scheme produces an empirical distribution of types and running variables, denoted G_n . That is, G_n is the sample analog of G .

Fix the proportion of seats at school s in the sampled economy to be q_s and run DA with these applicants and schools. Compute $MID_{\theta_s}^v$, τ_s , and partition Θ_s by observing cutoffs \hat{c}_n and school assignments in this single realization. Our estimator uses these quantities to construct a sample analogue of $p_s(\theta, \mathbf{r})$ in Theorem 3. Specifically, we partition Θ_s , the set of applicants who rank s , into sets of never, always, and conditionally seated applicants. The empirical partition looks at bandwidths as well as cutoffs. The never seated are:

$$\Theta_s^n(\delta) = \{(\theta, \mathbf{r}) \mid \rho_{\theta_s} > \rho_s \text{ or, if } v(s) \neq 0, (\rho_{\theta_s} = \rho_s \text{ and } r_{v(s)} > \tau_s + \delta)\},$$

which is the same as (3) except that the second condition refers to applicants who (i) are in the marginal priority group, and (ii) whose running variable $r_{v(s)}$ at screened s exceeds right-hand window boundary $\tau_s + \delta$. A data-determined bandwidth is computed using the methods in Imbens and Kalyanaraman (2012).

Likewise, construct the empirical set of always seated applicants by

$$\Theta_s^a(\delta) = \{(\theta, \mathbf{r}) \mid \rho_{\theta s} < \rho_s \text{ or, if } v(s) \neq 0, (\rho_{\theta s} = \rho_s \text{ and } r_{v(s)} \leq \tau_s + \delta)\},$$

which is the same as (4) except that the second condition refers to applicants who (i) are in the marginal priority group, and (ii) whose running variable at screened s fails to clear left-hand window boundary $\tau_s - \delta$.

Finally, the empirical set of conditionally seated applicants is

$$\Theta_s^c(\delta) = \{(\theta, \mathbf{r}) \mid \rho_{\theta s} = \rho_s \text{ and, if } v(s) \neq 0 \text{ then } r_{v(s)} \in (\tau_s - \delta, \tau_s + \delta)\},$$

which is the same as (5) except that the second condition refers to applicants to screened schools who (i) are in the marginal priority group, and (ii) whose running variable is in window $(\tau_s - \delta, \tau_s + \delta]$, that is, these applicants are near the school- s cutoff.

Our propensity score estimator is computed as follows. If $(\theta, \mathbf{r}) \in \Theta_s^n(\delta_n)$ or $(\theta, \mathbf{r}) \in \Theta_s^a(\delta_n)$ for some $b \in B_{\theta s}$, then

$$\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n) = 0.$$

Otherwise,

$$\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n) = \begin{cases} 0.5^{m_{\theta s}(\mathbf{r})}(1 - MID_{\theta s}^0) & \text{if } (\theta, \mathbf{r}) \in \Theta_s^a(\delta_n) \\ 0.5^{m_{\theta s}(\mathbf{r})} \max\{0, \tau_s - MID_{\theta s}^0\} & \text{if } (\theta, \mathbf{r}) \in \Theta_s^c(\delta_n), v(s) = 0 \\ 0.5^{1+m_{\theta s}(\mathbf{r})}(1 - MID_{\theta s}^0) & \text{if } (\theta, \mathbf{r}) \in \Theta_s^c(\delta_n), v(s) \neq 0 \end{cases}$$

where δ_n is the bandwidth for a sampled economy of size n .

We consider the gap between estimator $\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n)$ and theoretical local propensity score $p_s(\theta, \mathbf{r})$ as n grows and δ_n shrinks. This convergence result uses a regularity condition:

Assumption 2. (*Rich support*) For any $s \in S$ and priority $\rho \in \{1, \dots, K, \infty\}$ with $G(\{i \in I : \rho_{is} = \rho\}) > 0$, we have $G(\{i \in I : \rho_{is} = \rho, i \text{ ranks } s \text{ first}\}) > 0$.

This assumption says that in the continuum economy, every school is ranked first by at least some applicants in every non-empty priority group defined for that school.

In this setup, propensity scores estimated with data-dependent bandwidth converge almost surely to the limiting local score described in Theorem 3, a result formalized below:

Theorem 4. *In the asymptotic sequence described above and maintaining Assumptions 1 and 2, the estimated local propensity score $\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n)$ is a consistent estimator of the local propensity score $p_s(\theta, \mathbf{r})$ in the following sense: For all $\theta \in \Theta, s \in S, \mathbf{r} \in [0, 1]^V$ and δ_n such that $\delta_n \rightarrow 0$ as $n \rightarrow \infty$,*

$$|\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n) - p_s(\theta, \mathbf{r})| \xrightarrow{a.s.} 0,$$

as $n \rightarrow \infty$. Moreover, since θ has finite support, this convergence is uniform in θ .

Theorem 4 justifies our use of the formula in Theorem 3 to eliminate OVB in empirical work estimating school attendance effects.

5 A Brief Report on NYC Report Cards

Since the 2003-04 school year, the NYC DOE has used DA to assign rising ninth graders to high schools. Each applicant for a ninth grade seat can rank up to twelve programs. All traditional public high schools participate in the match, but charter schools and New York City’s exam schools have a separate admissions process.⁶

The NYC assignment mechanism uses mixed multiple tie-breakers in a setup like that described in Section 4.3: lottery schools use a common randomly assigned tie-breaker, while screened schools use a variety of tie-breaking variables. Our empirical analysis treats non-lottery tie-breakers as distinct, one for each screened school, though the theoretical discussion above allows screened schools to share tie-breakers.

Schools may host multiple programs and each program sets its own priorities, such as preference for applicants from the same borough. Within priority groups, programs ration seats in one of two ways. Three types of programs rank applicants systematically. These include *screened* programs, which rank students based on academics and previous attendance records, *audition* programs that use interviews or other qualitative assessments, and *educational option* programs that screen for half their seats using middle school test scores. Our analysis refers to programs and schools 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 educational option programs. Our analysis computes propensity scores for programs rather than schools since programs are the unit of assignment. But we aggregate program propensity scores to produce a school-level score when evaluating causal effects such as the effect of attendance at a Grade A School. For our purposes, a “screened school” is a school hosting any screened program.⁷

In 2007, the NYC DOE launched a school accountability system that graded schools from A to F. This mirrored similar accountability systems in Florida and other states. New York City’s school grades were determined by achievement levels and, especially, achievement growth, as well as by survey- and attendance-based measures of school environment. Growth looked at credit accumulation, Regents completion and pass rates; performance measures were derived mostly from four- and six-year graduation rates. Some schools were ungraded.⁸

The 2007 grading system was controversial. Proponents applauded the integration of multiple measures of quality while opponents objected to high-stakes consequences of low grades, such as school closure. Rockoff and Turner (2011) provide a partial validation of the system by showing that low grades appeared to spark school improvement. In 2014, a newly elected mayor had the

⁶The centralized NYC high school match is detailed in Abdulkadiroğlu et al. (2005, 2009). Abdulkadiroğlu et al. (2014) describe NYC exam school admissions.

⁷Because applicants receive at most one offer, the propensity score for a school is the sum of the scores for programs hosted by the school. Some NYC high schools sort applicants on a screening tie-breaker, breaking screened-running-variable ties using the common lottery number. Schools of this type are treated as lottery schools, after allowing for additional priority groups defined by the screened running variable.

⁸Walcott (2012) details the NYC grading methodology of this period.

2007 system replaced with school quality measures that place less weight on test scores and more on curriculum characteristics and subjective assessments of teaching quality. The relative merits of the old and new systems continue to be debated.

We showcase the use of mixed multiple tie-breakers for impact evaluation by estimating the effects of being assigned to a Grade A school on student achievement and graduation rates. This analysis uses application data from the 2012-13 and 2013-14 school years. Our sample includes first-time applicants to ninth grade, who submit a rank order list of programs in the main round of the NYC high school match. The available data include the inputs producing the match: school capacities and priorities, lottery numbers, and screened school tie-breakers. Our replication of the match is nevertheless imperfect. This may be due to clerical errors in the recording of tie-breakers. Among other problems, we see occasional gaps in the lists showing how schools rank their applicants and some applicants share a rank. Our online appendix details the manner in which these problems are addressed.

Students at Grade A schools have higher average Regents, PSAT, and SAT scores than do students at other schools. These school average characteristics are documented in Table 2, which also shows that Grade A schools have higher graduation rates than other schools in the district. This difference in graduation rates features in media accounts of socioeconomic characteristics of NYC match outcomes (see, e.g., Harris and Fessenden (2017) and Disare (2017)).

Screened Grade A schools are also somewhat larger than the district average size, but Grade A lottery schools are much smaller. Consistent with this, class sizes at screened Grade A schools tend to exceed those at other schools while lottery Grade A schools have classes that are smaller than those at screened Grade A schools. Students enrolled at Grade A schools are about as likely to be nonwhite as students who go elsewhere, and qualify for a free lunch and special education services at about the same rate. On the other hand, Grade A lottery schools have more minority students than both Grade A screened schools and the district as a whole. We also see an impressive gap in peer mean achievement between screened and lottery Grade A schools. Finally, teachers at Grade A lottery schools are, on average, about as experienced as those teaching elsewhere but are less likely to have an advanced degree (Master or higher).

In the two application cohorts described in Table 3, nearly 110,000 eighth graders applied in the high school match, and roughly 93,000 ranked a Grade A school. Grade A applicants mostly look like typical NYC eighth graders. On the other hand, while applicants subject to lottery risk of Grade A assignment have baseline (seventh grade) test scores around the district mean, applicants subject to general risk tend to have scored considerably higher than the district average. This reflects the fact that the group subject to general risk includes many screened school applicants, and most screened schools screen using test scores.

Figure 1 compares the number of applicants subject to lottery and general risk at each school with any Grade A risk. Specifically, Panel A shows numbers of applicants subject to lottery risk for each lottery school, where schools are ordered by capacity. Red arrows indicate the sample size gains yielded by moving from lottery to general risk among applicants to lottery schools. Perhaps not surprisingly, in the sample of lottery school applicants, the gains from exploiting general risk are mostly small. By contrast, the sample size gains yielded by moving from lottery to general risk at screened schools, documented in Panel B, are both broad (that is, appearing

for many schools) and large. Importantly, however, many screened schools also have applicants subject to lottery risk, reflecting scenarios like those described in Table 1.

The balancing property of propensity score conditioning is documented in Table 4, which reports Grade A school offer gaps in covariate means. Specifically, Table 4 reports estimates generated by the following setup. Let W_i be an applicant characteristic, drawn from a set that includes θ_i . Let D_{1i} be a dummy for a Grade A school offer with corresponding propensity score, \hat{p}_1 . Likewise, let D_{2i} indicate offers at an ungraded school with corresponding propensity score \hat{p}_2 . We control for ungraded school offers to ensure that Grade A effects compare sets of schools with high and low grades.

Using x_1 and x_2 to index values in the support of \hat{p}_1 and \hat{p}_2 , balance tests are estimates of parameter γ_1 in

$$W_i = \gamma_1 D_{1i} + \gamma_2 D_{2i} + \sum_{x_1} \alpha_1(x_1) \hat{p}_1(x_1) + \sum_{x_2} \alpha_2(x_2) \hat{p}_2(x_2) + h(r_i) + \nu_i, \quad (9)$$

with local linear control for the running variable according to

$$h(r_i) = \sum_{s \in S \setminus S_0} \omega_{1s} a_{is} + \omega_{2s} b_{is} + \omega_{3s} b_{is} \times (r_i - \tau_s) + \omega_{4s} b_{is} \times \mathbf{1}(r_i > \tau_s) \times (r_i - \tau_s), \quad (10)$$

where $S \setminus S_0$ is the set of screened programs, b_{is} is a bandwidth indicator defined as $b_{is} = \mathbf{1}(c_s - \delta_s < r_i < c_s + \delta_s)$, and a_{is} indicates whether applicant i applied to program s . Parameters $(\omega_{1s} - \omega_{4s})$ in (10) vary by application cohort. Bandwidths are estimated as suggested by Imbens and Kalyanaraman (2012), separately for each program, for the set of applicants who are in the relevant marginal priority group.⁹ Our theoretical results suggest that estimates of γ_1 in equation (9) should be close to zero. This is borne out by the results reported in columns 3 and 4 of Table 4, which show small, mostly insignificant differences in covariates by offer status. These can be compared with the large and often significant uncontrolled gaps reported in the first column. Moreover, as can be seen in the last two columns of the table, applicant and application characteristics are also balanced for assignment offers at lottery Grade A schools and screened Grade A schools.¹⁰

Causal effects of school attendance on test scores are measured as the effect of years of exposure to schools of a particular type. Exposure variables are denoted C_{1i} and C_{2i} 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

$$Y_i = \beta_1 C_{1i} + \beta_2 C_{2i} + \phi'_0 X_i + \sum_{x_1} \alpha_{01}(x_1) \hat{p}_1(x_1) + \sum_{x_2} \alpha_{02}(x_2) \hat{p}_2(x_2) + g(r_i) + \eta_i, \quad (11)$$

⁹Bandwidths are also computed separately for each outcome variable; the minimum across outcomes is used for all.

¹⁰Table A1 reports follow up differentials for each of our outcomes.

with associated first stage equations,

$$C_{1i} = \gamma_{11}D_{1i} + \gamma_{12}D_{2i} + \phi'_1 X_i + \sum_{x_1} \alpha_{11}(x_1)\hat{p}_1(x_1) + \sum_{x_2} \alpha_{12}(x_2)\hat{p}_2(x_2) + h_1(r_i) + \nu_{1i}$$

$$C_{2i} = \gamma_{21}D_{1i} + \gamma_{22}D_{2i} + \phi'_2 X_i + \sum_{x_1} \alpha_{21}(x_1)\hat{p}_1(x_1) + \sum_{x_2} \alpha_{22}(x_2)\hat{p}_2(x_2) + h_2(r_i) + \nu_{2i}.$$

Running variable controls in these equations, denoted $h_1(r_i)$, $h_2(r_i)$, and $g(r_i)$, are analogous to those in equation (10). The vector of baseline and demographic controls, X_i , includes baseline math and English scores; free lunch, special education, English language learner indicators; and gender and race dummies.

As a benchmark, the OLS estimates reported in the second column of Table 5, suggest Grade A attendance boost achievement and graduation rates. These estimates are constructed by fitting equation (11), without propensity score controls or instrumenting, in a sample that includes all participants in the high school match without regard to offer risk. By contrast, the 2SLS estimates reported in columns 3 and 4 suggest that any gains associated with Grade A attendance reflect selection bias. Compare, for example, the positive OLS estimate indicating a 0.13 increase in graduation rates to the negative 2SLS estimate of about -0.01 in column 4 of Table 5. This finding is noteworthy since graduation levels are a major determinant of school report card grades. Our results suggest that although Grade A schools indeed have higher graduation rates, they do not in and of themselves boost graduation. Higher English scores appear to be the only clear gains from Grade A school attendance; these are increased by 0.03 when estimated with covariates using general risk. This result is reasonably precise, with a standard error of 0.01. Looking across columns, however, the estimated gain in English achievement becomes smaller and less conclusively positive when estimated without covariates.

Estimates that exploit quasi-experimental variation around screened school cutoffs as well as lottery risk, reported in columns headed “general risk” in Table 5, have the advantage of adding applicants and schools to the analysis sample. The move from lottery risk to general risk yields a valuable precision gain, as can be seen by comparing results reported in columns 3 and 4 or columns 5 and 6. This increase in precision is driven both by larger sample sizes and larger first stages in the older cohort.¹¹ For instance, the associated standard error falls from 0.062 when the graduation effect is estimated using only lottery risk to 0.031 when estimated by exploiting general risk in Grade A assignment.

Because much of the policy discussion revolving around New York City’s school evaluation and assignment systems has focused on access to screened schools, we’re interested in distinguishing screened Grade A from lottery Grade A effects. The multi-sector estimates reported in Table 6 are from models that include separate endogenous variables for each type of Grade A school, along with a dummy for a single ungraded sector. Multi-sector models include separate propensity score controls for both sectors, as well as for the single ungraded sector. The results in column 3 distinguish the two types of school effects while using only lottery variation, that is, treating screened-school running variables as priorities. Perhaps surprisingly, lottery variation alone is

¹¹Table A2 reports first stage estimates by year. The first stages are larger for 2012-2013 than 2011-12 because our match rates are higher. The first stage is larger under general risk than lottery risk in 2012-13, and is similar in 2013-14.

sufficient to capture a reasonably precise screened school effect, with an estimated standard error for the graduation estimate of 0.053 (the estimate itself is a virtual zero at -0.011). The corresponding lottery school effect (reported in column 4 of Table 6) is 0.066, also not significantly different from zero. On the other hand, the estimates in column 4 show small but marginally significant gains from exposure to a lottery Grade A school. Again, it's worth emphasizing that this result is obtained while exploiting only lottery risk.

The research value of screened school tie-breaking emerges clearly in an analysis that distinguishes screened Grade A effects from lottery Grade A school effects. This value is apparent in a comparison of the estimates in columns 5 and 6 of Table 6, which were computed using general risk, with those in columns 3 and 4, using only lottery risk. Precision gains for estimates of the effects of lottery school exposure are modest, but not trivial for graduation outcomes (compare, for example, standard errors of 0.074 in column 6 with a standard error of 0.094 in column 4). The payoff to analyses exploiting general risk is most dramatic for estimates of the effects of screened school attendance. The standard error associated with Regents English, for example, falls from 0.016 in column 3 to 0.010 in column 5. As can be seen in column 7, this precision gain would otherwise require increasing the sample 2.5 times. The precision gain for estimates of screened school graduation effects are equally impressive, equivalent to a near tripling in sample size. The ratios in columns 7 and 8 show that in sample size terms, the gains from moving from lottery to general risk range from 1.5 to 2.9, depending on sector and outcome.

6 Summary and Next Steps

The spread of centralized matching markets opens new horizons for impact evaluation. The research potential of such markets is realized here by marrying the large-market DA propensity score generated by lottery tie-breaking with RD designs at screened schools. A by-product of our analysis is a novel view of RD as inducing a limiting propensity score that's constant at one-half. Our analysis also shows how markets with mixed multiple tie-breakers can be used to study treatment effects at screened schools for applicants away from screening cutoffs. This addresses a long-standing critique of traditional non-parametric RD estimates, which argues that effects local to cutoffs are potentially misleading for the general population. Mixed multiple tie-breaking produces variation away from cutoffs through the process of disqualification at more highly ranked choices.

Our analysis of NYC school progress report cards shows modest improvements on Regents exams for students attending a Grade A school. Importantly, however, the gains from attending a highly graded school appear only for schools that select applicants by lottery. While NYC's Grade A screened schools enroll some of the city's highest achievers, attendance at a Grade A screened school does not appear to boost student performance. Concerns about access to such schools (as expressed, for example, in Harris and Fessenden (2017)) would therefore seem to be misplaced. It's also noteworthy that attendance at a Grade A school fails to increase high school graduation rates or performance on PSAT and SAT tests.

Further afield, our theoretical framework may be applicable to the US National Residency Matching Program, which matches 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 value of alternative medical training programs. As we’ve discussed elsewhere, important non-DA mechanisms can also be brought under the DA umbrella with suitable modifications to the inputs. This includes the Boston or immediate acceptance mechanism or mechanisms used for exam school assignment systems used in major US cities and abroad (see, e.g., Abdulkadiroğlu et al. (2014), Balinski and Sönmez (1999), Dobbie and Fryer (2014), and Kirkeboen et al. (2016)). In ongoing work, Angrist et al. (2017) deploy the methods developed here in a study of Chicago’s exam schools. Our framework may also be useful for the evaluation of resources allocated in auction markets. Some auctions can be seen as a version of DA with non-lottery tie-breakers, where applicants are bidders and running variables are bids.¹²

Our provisional agenda for further research also includes work on econometric implementation strategies, such as bandwidth selection. More refined 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. In addition to development of a strategy to compute the mixed multiple tie-breaker score by simulation, an open question is how to use a simulated score to eliminate OVB in a match that does not satisfy ETE.

¹²Existing impact evaluations using RD in auctions include Narayanan and Kalyanam (2015) and Kawai and Nakabayashi (2014).

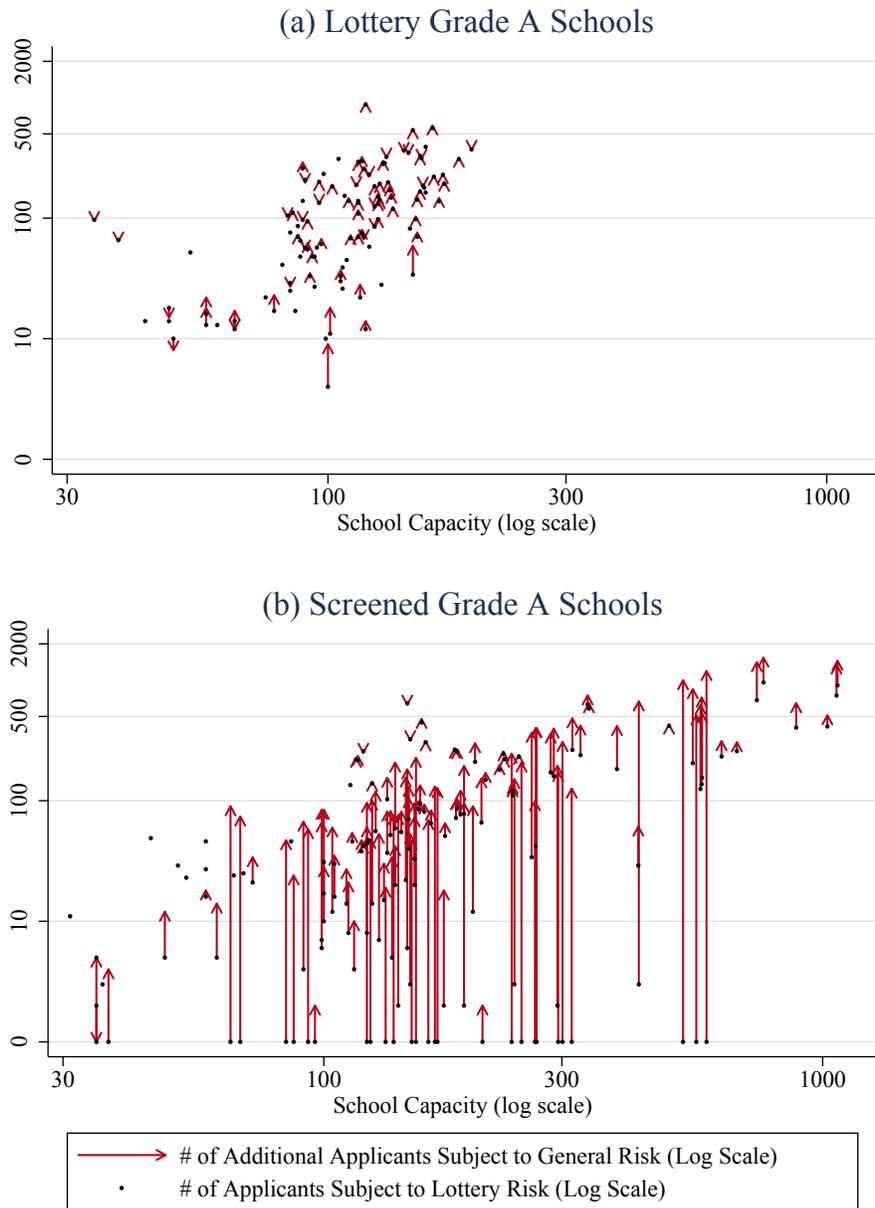
References

- Abdulkadiroğlu, Atila, Josh Angrist, and Parag Pathak**, “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools,” *Econometrica*, 2014, *82(1)*, 137–196.
- , **Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak**, “Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation,” *Econometrica*, 2017, *85(5)*, 1373–1432.
- , – , – , – , – , – , and **Roman A. Zarate**, “Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago’s Exam Schools,” *American Economic Review, Papers and Proceedings*, 2017.
- , **Parag A. Pathak, and Alvin E. Roth**, “The New York City High School Match,” *American Economic Review, Papers and Proceedings*, 2005, *95*, 364–367.
- , – , and – , “Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match,” *American Economic Review*, 2009, *99(5)*, 1954–1978.
- , **Yeon-Koo Che, and Yosuke Yasuda**, “Expanding “Choice” in School Choice,” *American Economic Journal: Microeconomics*, 2015, *7(1)*, 1–42.
- Ajayi, Kehinde**, “Does School Quality Improve Student Performance? New Evidence from Ghana,” 2014. Working paper, Boston University.
- Angrist, Joshua D., Parag A. Pathak, Miikka Rokkanen, and Roman Zarate**, “Choice and Consequence: Understanding Mismatch at Chicago’s Exam Schools,” 2017. Working paper, MIT.
- Azevedo, Eduardo and Jacob Leshno**, “A Supply and Demand Framework for Two-Sided Matching Markets,” *Journal of Political Economy*, 2016, *124(5)*, 1235–1268.
- Balinski, Michel and Tayfun Sönmez**, “A Tale of Two Mechanisms: Student Placement,” *Journal of Economic Theory*, 1999, *84*, 73–94.
- Barrow, Lisa, Lauren Sartain, and Marisa de la Torre**, “The Role of Selective High Schools in Equalizing Educational Outcomes: Heterogeneous Effects by Neighborhood Socioeconomic Status,” 2016. FRB of Chicago Working Paper No. WP-2016-17.
- Beuermann, Diether, C Kirabo Jackson, and Ricardo Sierra**, “Privately Managed Public Secondary Schools and Academic Achievement in Trinidad and Tobago: Evidence from Rule-Based Student Assignments,” 2016. Working Paper.
- Disare, Monica**, “City to eliminate high school admissions method that favored families with time and resources,” 2017. Chalkbeat, June 6.
- Dobbie, Will and Roland G. Fryer**, “Exam High Schools and Academic Achievement: Evidence from New York City,” *American Economic Journal: Applied Economics*, 2014, *6(3)*, 58–75.

- Fort, Margherita, Andrea Ichino, and Giulio Zanella**, “Cognitive and Non-cognitive Costs of Daycare 0-2 for Girls,” 2016. Working Paper.
- Gale, David and Lloyd S. Shapley**, “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, 1962, 69, 9–15.
- Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw**, “Identification and estimation of treatment effects with a regression-discontinuity design,” *Econometrica*, 2001, 69 (1), 201–209.
- Harris, Elizabeth and Ford Fessenden**, “The Broken Promises of Choice in New York City Schools,” 2017. New York Times, May 5.
- Hastings, Justine, Christopher Neilson, and Seth Zimmerman**, “Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile,” 2013. NBER Working paper 19241.
- Imbens, Guido and Karthik Kalyanaraman**, “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *Review of Economic Studies*, 2012, 79(3), 933–959.
- Jackson, Kirabo**, “Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago,” *Economic Journal*, 2010, 120(549), 1399–1429.
- , “Single-sex Schools, Student Achievement, and Course Selection: Evidence from Rule-based Student Assignments in Trinidad and Tobago,” *Journal of Public Economics*, 2012, 96(1-2), 173–187.
- Kawai, Kei and Jun Nakabayashi**, “Detecting large-scale collusion in procurement auctions,” *Available at SSRN 2467175*, 2014.
- Kirkeboen, Lars, Edwin Leuven, and Magne Mogstad**, “Field of Study, Earnings, and Self-Selection,” *Quarterly Journal of Economics*, 2016, 131, 1057–1111.
- Lee, David S. and Thomas Lemieux**, “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 2010, 48, 281–355.
- Lucas, Adrienne and Isaac Mbiti**, “Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya,” *American Economic Journal: Applied Economics*, 2014, 6(3), 234–263.
- Narayanan, Sridhar and Kirthi Kalyanam**, “Position effects in search advertising and their moderators: A regression discontinuity approach,” *Marketing Science*, 2015, 34 (3), 388–407.
- Pathak, Parag A.**, “What Really Matters in Designing School Choice Mechanisms,” in Bo Honore, Monika Piazzesi, Ariel Pakes, and Larry Samuelson, eds., *Advances in Economics and Econometrics, 11th World Congress of the Econometric Society*, Cambridge University Press, 2017.

- Pop-Eleches, Cristian and Miguel Urquiola**, “Going to a Better School: Effects and Behavioral Responses,” *American Economic Review*, 2013, *103*(4), 1289–1324.
- Rockoff, Jonah and Lesley Turner**, “Short Run Impacts of Accountability of School Quality,” *American Economic Journal: Economic Policy*, 2011, *2*(4), 119–147.
- Rosenbaum, Paul R. and Donald B. Rubin**, “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 1983, pp. 41–55.
- Roth, Alvin E. and Elliott Peranson**, “The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design,” *American Economic Review*, 1999, *89*, 748–780.
- **and Marilda A. O. Sotomayor**, *Two-sided Matching: a Study in Game-theoretic Modeling and Analysis*, Cambridge University Press: Econometric Society monographs, 1990.
- Rubin, Donald B.**, “Assignment to Treatment Group on the Basis of a Covariate,” *Journal of Educational Statistics*, 1977, *2*(1), 1–26.
- van der Vaart, A. W.**, *Asymptotic Statistics*, Cambridge University Press, 2000.
- Walcott, Dennis**, “NYC Department of Education: Progress Reports for New York City Public Schools,” 2012. January.
- Wellner, Jon A.**, “A Glivenko-Cantelli theorem for empirical measures of independent but non-identically distributed random variables,” *Stochastic Processes and Their Applications*, 1981, *11* (3), 309–312.
- Zimmerman, Seth D.**, “Making the One Percent: The Role of Elite Universities and Elite Peers,” Technical Report, National Bureau of Economic Research 2016.

Figure 1: 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 their capacity. Dots plot the number of students who have risk when screened school admission is treated as determined solely by priorities. Arrow length 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. The plot shows only schools with a capacity of at least 30 seats. Declines in risk of 2 or fewer students are not shown.

Table 1. 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)
<i>Example 1: Lotteries Create Risk at Lottery and Screened Schools</i>								
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
<i>Example 2: Screening Creates Risk at Lottery and Screened Schools</i>								
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
<i>Example 3: Lotteries and Screening Interact to Create Risk</i>								
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 in three examples. Each example describes the risk faced by a single applicant at each of the schools ranked: the applicant in example 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.

Table 2. New York School Characteristics

	Grade A schools			Other schools (4)
	All (1)	Screened (2)	Lottery (3)	
<i>Panel A: Average Performance Levels</i>				
Regents Math (standardized)	0.12	0.25	-0.18	-0.13
Regents English (standardized)	0.22	0.32	-0.03	-0.03
PSAT NMSQT (60-240)	113	118	103	108
SAT Reasoning (400-1600)	929	958	851	896
Graduation	0.74	0.76	0.69	0.63
Graduation with advanced regents	0.18	0.21	0.09	0.13
<i>Panel B: School Characteristics</i>				
Total per student expenditures	17423	17033	18362	17579
Expenditure on classroom instruction	7689	7522	8091	7946
Number of grade 9 students	1214	1529	455	1540
Number of grade 12 students	1052	1336	369	1280
Number of grade 9 to 12 students	4581	5793	1661	5565
Average class size grade 9	24.2	25.5	21.1	21.1
<i>Panel C: Peer Attributes</i>				
Fraction minority	0.68	0.60	0.88	0.71
Fraction female	0.52	0.53	0.51	0.50
Fraction special education	0.13	0.12	0.18	0.14
Seventh grade Math	0.15	0.29	-0.19	-0.03
Seventh grade English	0.13	0.26	-0.17	-0.04
<i>Panel D: Teacher Attributes</i>				
Inexperienced teachers	0.10	0.09	0.13	0.14
Advanced degree teachers	0.50	0.54	0.39	0.47
Teacher turnover rate	0.14	0.13	0.17	0.18

Notes. This table reports weighted average characteristics of school-year pairs as observed in data on school average Regents, SAT, and graduation outcomes (Panel A) and data from New York City school report cards (Panels B, C, and D), for 2011-12 and 2012-13. Grade A schools are defined based on the year before application. A screened school is any school with at least one screened program. A lottery school is any school without screened programs. School-year pairs are weighted by the number of students enrolled at each school, separately for each year. Inexperienced teachers have 3 or fewer years of experience. Advanced degree teachers have a Masters or higher degree.

Table 3. Applicant Characteristics

	Eighth grade		Grade A applicants	Score in (0,1)	
	students (1)	Applicants (2)		Lottery risk (4)	General risk (5)
Asian	0.154	0.170	0.163	0.110	0.189
Black	0.309	0.286	0.297	0.297	0.257
Hispanic	0.405	0.390	0.406	0.493	0.424
White	0.125	0.147	0.127	0.095	0.123
Female	0.487	0.512	0.525	0.503	0.521
Special education	0.186	0.068	0.067	0.077	0.061
English language learners	0.123	0.107	0.103	0.112	0.091
Free lunch	0.861	0.839	0.850	0.892	0.855
Year 2012/13	0.497	0.492	0.491	0.478	0.500
Brooklyn	0.316	0.312	0.301	0.304	0.290
The Bronx	0.215	0.210	0.228	0.321	0.244
Manhattan	0.134	0.130	0.139	0.086	0.098
Queens	0.277	0.287	0.281	0.236	0.322
Staten Island	0.059	0.061	0.051	0.052	0.046
Baseline scores					
Seventh grade Math	0.021	0.187	0.180	-0.016	0.258
Seventh grade English	0.016	0.187	0.190	0.003	0.254
Students	123687	109244	92630	17,353	25,052
Schools				111	134
School-year pairs				156	202

Notes. This table describes the population of New York City eighth graders in the 2011-12 and 2012-13 school years. Column 1 shows statistics for enrolled eighth graders. Column 2 shows statistics for first-time high school applicants who applied in the main round of the NYC high school match in 2012 or 2013. Column 3 reports statistics for applicants who ranked a Grade A school. Column 4 shows statistics for applicants with lottery risk of Grade A assignment; column 5 shows statistics for applicants with general risk of Grade A assignment. 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. New York had 142 Grade A schools in our analysis period. Counts of schools and school-year pairs in this table include only schools with at least 10 applicants at risk of Grade A assignment. Baseline seventh grade Math and English test scores are standardized by year to the population of test-takers. Boroughs refer to the applicants' borough of residence.

Table 4. Statistical Tests for Balance

	Non-offered mean (1)	No controls	Lottery risk	General risk		
		Grade A school (2)	Grade A school (3)	Grade A school (4)	Lottery Grade A school (5)	Screened Grade A school (6)
<i>Panel A: Application Covariates</i>						
# of schools ranked	6.656	-0.303*** (0.021)	0.028 (0.048)	-0.022 (0.035)	-0.057 (0.062)	0.028 (0.041)
# of Grade A schools ranked	3.087	0.799*** (0.013)	0.033 (0.031)	0.006 (0.021)	-0.035 (0.040)	0.038 (0.025)
Grade A ranked first	0.605	0.439*** (0.003)	-0.002 (0.006)	0.007 (0.005)	-0.005 (0.008)	0.014** (0.006)
# of screened Grade A schools ranked	1.008	0.240*** (0.009)	-0.007 (0.019)	-0.014 (0.015)	-0.033 (0.032)	0.008 (0.016)
# of lottery Grade A schools ranked	2.079	0.558*** (0.011)	0.040 (0.024)	0.020 (0.015)	-0.002 (0.025)	0.028 (0.020)
F-test (mvreg)			0.822	1.177	0.694	2.058
p-val			(0.511)	(0.319)	(0.596)	(0.068)
<i>Panel B: Baseline Covariates</i>						
Asian	0.177	0.003 (0.003)	-0.002 (0.006)	0.003 (0.006)	0.010 (0.006)	0.002 (0.008)
Black	0.272	-0.079*** (0.003)	-0.011 (0.008)	-0.015** (0.007)	-0.010 (0.012)	-0.006 (0.008)
Hispanic	0.426	0.008** (0.003)	0.012 (0.009)	0.011 (0.008)	0.004 (0.013)	0.002 (0.010)
White	0.117	0.066*** (0.002)	0.002 (0.006)	0.001 (0.005)	-0.005 (0.006)	0.001 (0.007)
Female	0.529	0.010** (0.003)	-0.000 (0.009)	-0.001 (0.008)	-0.002 (0.013)	0.003 (0.010)
Special education	0.060	-0.004** (0.002)	0.012** (0.005)	0.010** (0.004)	0.006 (0.007)	0.008 (0.005)
English language learners	0.085	0.004** (0.002)	0.005 (0.006)	0.004 (0.005)	-0.009 (0.008)	0.003 (0.006)
Free lunch	0.862	-0.058*** (0.003)	-0.009 (0.006)	-0.008 (0.005)	-0.008 (0.007)	-0.008 (0.007)
Baseline scores						
Seventh grade Math	0.209	0.192*** (0.006)	-0.003 (0.013)	0.011 (0.011)	0.028 (0.018)	0.003 (0.012)
Seventh grade English	0.221	0.167*** (0.006)	0.003 (0.013)	-0.000 (0.011)	0.026 (0.019)	-0.008 (0.011)
F-test (mvreg)			1.366	1.701	1.571	0.721
p-val			(0.189)	(0.074)	(0.109)	(0.690)
N		90914	16505	23778	9614	18243
Points of support (Grade A)			607	986	509	746

Notes. This table reports balance statistics, computed by regressing covariates on dummies indicating a Grade A offer and an ungraded school offer. The sample includes applicants for ninth grade seats in 2011-12 and 2012-13. Columns 3 and 4 report estimates that condition on saturated controls for Grade A and ungraded school propensity scores. The sample for columns 3-6 includes students who are at risk of Grade A assignment. Lottery risk in column 3 is computed by treating screened-school tie-breakers as priorities. Column 5 reports balance for students at risk of an offer at lottery schools; column 6 reports balance statistics for students at risk of an offer at screened schools. Both sets of estimates control for the relevant propensity score. Estimates in columns 4-6 are from models that include running variable controls. Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5. Grade A School Effects

	Non-offered mean (1)	OLS (2)	2SLS			
			No covariate controls		Baseline/demographic controls	
			Lottery risk (3)	General risk (4)	Lottery risk (5)	General risk (6)
Regents Math (standardized)	0.211	0.075*** (0.005)	-0.018 (0.043)	0.039 (0.037)	0.026 (0.037)	0.043 (0.033)
	N	95352	16161	23491	14440	21331
Regents English (standardized)	0.303	0.065*** (0.002)	0.006 (0.016)	0.017 (0.012)	0.027** (0.014)	0.032** (0.010)
	N	87633	14745	21684	13169	19690
PSAT NMSQT (60-240)	117	0.013 (0.093)	-1.186 (0.912)	-0.063 (0.714)	0.015 (0.685)	0.168 (0.563)
	N	54143	8156	12725	7678	12139
SAT Reasoning (400-1600)	961	-0.826** (0.353)	-3.937 (3.593)	3.202 (2.752)	2.342 (2.621)	3.700* (2.176)
	N	58270	9218	14594	8197	13225
Graduation	0.701	0.129*** (0.004)	0.017 (0.056)	-0.013 (0.035)	0.031 (0.062)	-0.001 (0.031)
	N	51086	9608	13012	7964	8791
Graduation with Advanced Regents	0.169	-0.000 (0.003)	-0.037 (0.038)	-0.009 (0.027)	-0.030 (0.040)	0.023 (0.025)
	N	51086	9608	13012	7964	8791

Notes. This table reports OLS and 2SLS estimates of the effects of Grade A high school enrollment on Regents Math, Regents English, SAT, PSAT, and graduation outcomes. Estimates in columns 3-6 are from models with dummies for Grade A and ungraded schools treated as endogenous. OLS estimates are from models that omit propensity score controls and include all students in the two match cohorts, though not limited to those in the match. The 2SLS sample is limited to students with Grade A assignment risk. Regents Math and English scores are standardized to the population of test takers in the year and semester tested. SAT Reasoning is the sum of SAT Math and Reading. PSAT NMSQT is the sum of PSAT Math, Critical Reading and Writing. SAT and PSAT are scores from the first time the test is taken. The estimates in columns 2, 5, and 6 include controls for seventh grade Math and English scores, free lunch status, SPED and ELL status, gender, and race/ethnicity indicators. Estimates in columns 4 and 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 6. Multi-Sector 2SLS Estimates

	Non-offered mean		Lottery risk		General risk		Equivalent sample size gain	
	Screened (1)	Lottery (2)	Screened (3)	Lottery (4)	Screened (5)	Lottery (6)	Screened (7)	Lottery (8)
Regents Math (Standardized)	0.319	0.012	-0.043 (0.041)	0.099** (0.046)	0.009 (0.034)	0.099** (0.045)	1.454	1.045
N			15960		23704			
Regents English (Standardized)	0.384	0.164	0.007 (0.016)	0.048** (0.016)	0.026** (0.010)	0.044** (0.015)	2.560	1.138
N			14653		21994			
PSAT NMSQT (60-240)	119	111	0.424 (0.699)	-0.628 (0.867)	0.270 (0.509)	-0.912 (0.811)	1.886	1.143
N			8744		13865			
SAT Reasoning (400-1600)	981	904	0.570 (2.913)	3.436 (3.001)	2.241 (2.072)	3.094 (2.832)	1.977	1.123
N			9299		15008			
Graduated	0.749	0.682	-0.011 (0.053)	0.066 (0.094)	-0.040 (0.031)	0.032 (0.074)	2.923	1.614
N			8655		12183			
Graduated with Advanced Regents	0.226	0.109	-0.031 (0.036)	-0.031 (0.061)	-0.019 (0.026)	-0.045 (0.048)	1.917	1.615
N			8655		12183			

Notes. This table reports 2SLS estimates of models that separately identify Grade A effects at screened and lottery schools, treating both as endogenous. These models also include an endogenous dummy for ungraded schools. The sample is limited to students with either Grade A lottery or Grade A screened assignment risk. All models include controls for seventh grade Math and English scores, free lunch status, SPED and ELL status, gender, and race/ethnicity indicators. Estimates in columns 5 and 6 are from models that include running variable controls. Robust standard errors in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A1. Differential Attrition

	No controls	Lottery risk	General risk		
	Grade A (1)	Grade A (2)	Grade A (3)	Grade A lottery (4)	Grade A screened (5)
Has Regents Math score	0.019*** (0.003)	0.008 (0.006)	0.003 (0.005)	0.001 (0.009)	-0.001 (0.006)
Has Regents English score	0.051*** (0.003)	-0.013* (0.008)	-0.010 (0.007)	-0.013 (0.011)	-0.011 (0.008)
Has PSAT NMSQT score	0.048*** (0.003)	0.012 (0.009)	0.004 (0.008)	0.031** (0.012)	-0.025** (0.009)
Has SAT Reasoning score	0.089*** (0.003)	0.017* (0.009)	0.007 (0.009)	0.028** (0.013)	-0.014 (0.010)

Notes. This table reports estimates of differential attrition for test score outcomes, computed by regressing the variables at left on dummies indicating Grade A and ungraded school offers using models similar to those used to assess balance. Samples include applicants in 2011-12 and 2012-13 who were enrolled in NYC in eighth grade (the grade before high school). The samples for columns 2-5 are limited to students who were at risk of Grade A assignment. Differential attrition estimates condition on saturated controls for Grade A and ungraded school propensity scores. Column 4 reports estimates for students at risk of an offer at lottery schools; column 5 reports estimates for students at risk at screened schools. Lottery risk is computed by treating screened-school tie-breakers as priorities. All models include controls for seventh grade Math and English scores, free lunch status, SPED and ELL status, gender, and race/ethnicity indicators. Estimates in columns 3-5 are from models that include running variable controls. Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

Table A2. First Stage Estimates for Grade A Schools

	2012/2013		2013/2014	
	Lottery risk	General risk	Lottery risk	General risk
	(1)	(2)	(3)	(4)
Regents Math (years of exposure)	0.280*** (0.018)	0.295*** (0.016)	0.565*** (0.013)	0.490*** (0.012)
N	6963	9849	7515	11521
Regents English (years of exposure)	0.713*** (0.039)	0.884*** (0.035)	1.477*** (0.031)	1.457*** (0.027)
N	6454	9205	6746	10518
PSAT NMSQT (years of exposure)	0.464*** (0.028)	0.565*** (0.024)	1.042*** (0.026)	1.040*** (0.024)
N	5932	8560	6421	9901
SAT Reasoning (years of exposure)	0.784*** (0.042)	0.999*** (0.038)	1.591*** (0.032)	1.573*** (0.028)
N	7964	10986	8580	12829
Ever enrolled (dummy)	0.295*** (0.014)	0.363*** (0.013)	0.615*** (0.011)	0.602*** (0.010)
N	7964	10986	8580	12829

Notes. This table reports first stage estimates of the effect of Grade A offers on Grade A enrollment, separately by application cohort and type of risk. First stage estimates differ across outcomes and years because different outcomes are seen in follow-up windows that differ by test and grade. The sample includes applicants for eighth grade seats in 2011-12 and 2012-13 facing the indicated type of risk. Estimates condition on saturated controls for Grade A and ungraded school propensity scores. The estimates in columns 1 and 3 control for propensity scores that treat screened-school tie-breakers as priorities. All models include controls for seventh grade Math and English scores, free lunch status, SPED and ELL status, gender, and race/ethnicity indicators. Estimates in columns 2 and 4 are from models that include running variable controls. Robust standard errors are in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%

A Theoretical Appendix

To simplify notation, we define the applicant rank for any $(r_0, \mathbf{r}) \in \mathbb{R}^{V+1}$ as follows:

$$\pi_{\theta_s}(r_0, \mathbf{r}) = \rho_{\theta_s} + r_{v(s)}.$$

When $\pi_{\theta_s}(r_0, \mathbf{r}) < c_s$, either applicant of type θ has priority below the marginal priority ($\rho_{\theta_s} < \rho_s$) or she has priority equal to the marginal priority and her running variable value is less than the randomization cutoff ($\rho_{\theta_s} = \rho_s$ and $r_{v(s)} < \tau_s$).

A.1 Defining DA: Generalization

The definition of DA extends to the case with multiple mixed tie-breaking. For any set of applicant types $\Theta_0 \subset \Theta$ and for any number $(r_0, \mathbf{r}) \in [0, 1]^{V+1}$, define the set of applicants in Θ_0 with tie-breaking numbers less than (r_0, \mathbf{r}) to be

$$I(\Theta_0, r_0, \mathbf{r}) = \{i \in I \mid \theta_i \in \Theta_0, r_{iv} \leq r_v \text{ for all } v\}.$$

We use the shorthand notation $I_0 = I(\Theta_0, r_0, \mathbf{r})$ to denote this set and let the fraction of applicants in I_0 be denoted by $G(I_0)$. As in the case with single tie-breaking combine applicants' priority status and tie-breaking variable realization into *applicant rank* at school s , $\pi_{\theta_i s}(\mathbf{r}_i)$. The remaining notation and definition of DA is the same as in the case with single tie-breaking (Section 2).

Let us confirm that DA always converges for the general tie-breaking case. Our general formulation defines the DA match as determined by cutoffs found in the limit of a sequence. Recall that these cutoffs evolve according to

$$c_s^{t+1} = \begin{cases} K + 1 & \text{if } F(Q_s(\mathbf{c}^t)) < q_s, \\ \max\{x \in [0, K + 1] \mid F(\{i \in Q_s(\mathbf{c}^t) \text{ such that } \pi_{\theta_i s}(\mathbf{r}_i) \leq x\}) \leq q_s\} & \text{otherwise,} \end{cases}$$

where $Q_s(\mathbf{c}^t)$ is the demand for seats at school s for a given vector of cutoffs \mathbf{c}^t and is defined as

$$Q_s(\mathbf{c}^t) = \{i \in I \mid \pi_{\theta_i s}(\mathbf{r}_i) \leq c_s^t \text{ and } s \succ_i \tilde{s} \text{ for all } \tilde{s} \in S \text{ such that } \pi_{i\tilde{s}} \leq c_{\tilde{s}}^t\}. \quad (12)$$

The following result confirms that these limiting cutoffs exist, i.e., that the sequence \mathbf{c}^t converges.

Lemma 1. *Consider any economy with general tie-breaking as defined in Section 4. Construct a sequence of cutoffs, c_s^t , for this economy as described above. Then, $\lim_{t \rightarrow \infty} c_s^t$ exists.*

Proof. See the proof of Lemma 2 in Abdulkadiroğlu et al. (2017a). □

Note that this result applies to the cutoffs for both finite and continuum economies. In finite markets, at convergence, these cutoffs produce the allocation we get from the usual definition of DA (e.g., as in Gale and Shapley (1962)). This can be seen by noting that

$$\max\{x \in [0, K + 1] \mid F(\{i \in Q_s(\mathbf{c}^t) \text{ such that } \pi_{\theta_i s}(\mathbf{r}_i) \leq x\}) \leq q_s\}$$

$$= \max\{x \in [0, K + 1] \mid |\{j \in Q_s(\mathbf{c}^t) : \pi_{js} \leq x\}| \leq k_s\},$$

implying that the tentative cutoff at school s in step t in our DA formulation, which is determined by the left hand side of this equality, is the same as that in Gale and Shapley (1962)'s DA formulation, which is determined by the right hand side of the equality. Our DA formulation and the Gale and Shapley (1962) formulation therefore produce the same cutoff at each step. This also implies that, in finite markets, our DA cutoffs are found in a finite number of iterations, since DA as described by Gale and Shapley (1962) converges in a finite number of steps.

A.2 Proofs

A.2.1 Preliminaries

Our proofs use a lemma that describes the bandwidth-specific propensity score in a continuum economy under the assumptions in Section 4. For a profile $\mathbf{r} = (r_v)_{v \in \{1, \dots, V\}} \in [0, 1]^V$ of reference values of tie-breaking variables and each $v = 1, \dots, V$, let $e(v)$ denote the event that $\theta_i = \theta$, $R_{iu} > MID_{\theta_s}^u$ for $u = 1, \dots, v - 1$, and $R_i \in N(\mathbf{r}, \delta)$. Also, let

$$\Phi_\delta(v) \equiv \frac{F_v(r_v + \delta|e(v)) - \max\{F_v(MID_{\theta_s}^v|e(v)), F_v(r_v - \delta|e(v))\}}{F_v(r_v + \delta|e(v)) - F_v(r_v - \delta|e(v))} \text{ for } v \geq 1$$

$$\Phi_\delta \equiv (1 - MID_{\theta_s}^0)\Pi_v \Phi_\delta(v)$$

$$\Phi'_\delta \equiv \max \left\{ 0, \frac{\min\{F_{v(s)}(\tau_s|e(v)), F_{v(s)}(r_{v(s)} + \delta|e(v))\} - \max\{F_{v(s)}(MID_{\theta_s}^{v(s)}|e(v)), F_{v(s)}(r_{v(s)} - \delta|e(v))\}}{F_{v(s)}(r_{v(s)} + \delta|e(v)) - \max\{F_{v(s)}(MID_{\theta_s}^{v(s)}|e(v)), F_{v(s)}(r_{v(s)} - \delta|e(v))\}} \right\}.$$

Lemma 2. *In the mixed tie-breaking setting of Section 4, consider a continuum economy populated by applicants of type $\theta \in \Theta$ to be assigned to schools indexed by $s \in S$. For any fixed $\delta > 0$, we have:*

$$p_s(\theta, \mathbf{r}, \delta) = \begin{cases} 0 & \text{if } \theta \in \Theta_s^n \text{ or } (MID_{\theta_s}^v = 1 \text{ or } r_v + \delta \leq MID_{\theta_s}^v \text{ for some } v), \\ \Phi_\delta & \text{if } \theta \in \Theta_s^a, MID_{\theta_s}^v \neq 1 \text{ and } r_v + \delta \geq MID_{\theta_s}^v \text{ for all } v, \\ \Phi_\delta \times \Phi'_\delta & \text{if } \theta \in \Theta_s^c, MID_{\theta_s}^v \neq 1, \text{ and } r_v + \delta \geq MID_{\theta_s}^v \text{ for all } v. \end{cases}$$

Proof. Note first that admissions cutoffs \mathbf{c} in a continuum economy are invariant to tie-breaking outcomes R_i , reference tie-breaking number \mathbf{r} , and bandwidth δ : DA in the continuum depends on (R_i) only through $F(I_0)$ for sets $I_0 = \{i \in I \mid \theta_i \in \Theta_0, r_{iv} \leq r_v \text{ for all } v\}$ with various choices of Θ_0 and \mathbf{r} . In particular, $F(I_0)$ doesn't depend on running variable realizations in the continuum economy 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 Wellner (1981)'s Glivenko-Cantelli theorem for independent but non-identically distributed random variables. $F(I_0)$ doesn't depend on reference tie-breaking number \mathbf{r} and bandwidth δ either since \mathbf{r} and δ affect only the distribution of a single student i 's tie-breaking number R_i , which has no effect on $F(I_0)$ or cutoffs. Likewise, marginal priority ρ_s is constant for every school s .

Now, consider the bandwidth-specific propensity score $p_s(\theta, \mathbf{r}, \delta) \equiv E[D_i(s)|\theta_i = \theta, \mathbf{R}_i \in N(\mathbf{r}, \delta)]$. We start with verifying the first line in the above characterization of $p_s(\theta, \mathbf{r}, \delta)$. Applicants who don't rank s have $p_s(\theta, \mathbf{r}, \delta) = 0$. Among those who do rank s , those of type $\theta \in \Theta_s^n$

have $\rho_{\theta s} > \rho_s$ so that $p_s(\theta, \mathbf{r}, \delta) = 0$. Even if $\theta \notin \Theta_s^n$, if $MID_{\theta s}^v = 1$ or $r_v + \delta \leq MID_{\theta s}^v$ for some v , student i with $\theta_i = \theta$ and $R_i \in N(\mathbf{r}, \delta)$ is always assigned some $b \in B_{\theta s}$ so that $p_s(\theta, \mathbf{r}, \delta) = 0$.

To show the remaining cases, first note that applicants of type $\theta \in \Theta_s^a \cup \Theta_s^c$ with $MID_{\theta s}^v \neq 1$ and $r_v + \delta \geq MID_{\theta s}^v$ for all v 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 $R_i \in N(\mathbf{r}, \delta)$, the proportion of type θ applicants not being assigned any $b \in B_{\theta s}$ where $\rho_{\theta b} = \rho_b$ is $\Phi_\delta = (1 - MID_{\theta s}^0)\Pi_v\Phi_\delta(v)$ since each $\Phi_\delta(v)$ is the probability of not being assigned to $b \in B_{\theta s}^v$.

To see the second line, note that 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 if $\theta \in \Theta_s^a$, $MID_{\theta s}^v \neq 1$ and $r_v + \delta \geq MID_{\theta s}^v$ for all v , we have

$$p_s(\theta, \mathbf{r}, \delta) = \Phi_\delta.$$

Finally, consider applicants of type $\theta \in \Theta_s^c$ with $MID_{\theta s}^v \neq 1$, and $r_v + \delta \geq MID_{\theta s}^v$ for all v . The fraction of applicants $\theta \in \Theta_s^c$ 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 within this set is given by

$$\frac{\min\{F_{v(s)}(\tau_s|e(v)), F_{v(s)}(r_{v(s)} + \delta|e(v))\} - \max\{F_{v(s)}(MID_{\theta s}^{v(s)}|e(v)), F_{v(s)}(r_{v(s)} - \delta|e(v))\}}{F_{v(s)}(r_{v(s)} + \delta|e(v)) - \max\{F_{v(s)}(MID_{\theta s}^{v(s)}|e(v)), F_{v(s)}(r_{v(s)} - \delta|e(v))\}}.$$

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 Φ'_δ . Therefore, for all $\theta \in \Theta_s^c$ with $MID_{\theta s}^v \neq 1$ and $r_v + \delta \geq MID_{\theta s}^v$ for all v , we have $p_s(\theta, \mathbf{r}, \delta) = \Phi_\delta \times \Phi'_\delta$. \square

A.2.2 Theorem 2

With $\delta = 1$, the conditioning event $R_i \in N(\mathbf{r}, \delta)$ becomes vacuous for any \mathbf{r} so that $p_s(\theta, \mathbf{r}, \delta) = p_s(\theta)$. Note

$$\begin{aligned} \Phi_1(v) &= \frac{F_v(r_v + 1|\theta) - \max\{F_v(MID_{\theta s}^v|\theta), F_v(r_v - 1|\theta)\}}{F_v(r_v + 1|\theta) - F_v(r_v - 1|\theta)} \\ &= 1 - F_v(MID_{\theta s}^v|\theta) \\ \Phi'_1 &= \max\left\{0, \frac{\min\{F_{v(s)}(\tau_s|e(v)), F_{v(s)}(r_{v(s)} + 1|e(v))\} - \max\{F_{v(s)}(MID_{\theta s}^{v(s)}|e(v)), F_{v(s)}(r_{v(s)} - 1|e(v))\}}{F_{v(s)}(r_{v(s)} + 1|e(v)) - \max\{F_{v(s)}(MID_{\theta s}^{v(s)}|e(v)), F_{v(s)}(r_{v(s)} - 1|e(v))\}}\right\} \\ &= \max\left\{0, \frac{F_{v(s)}(\tau_s|\theta) - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)}{1 - F_{v(s)}(MID_{\theta s}^{v(s)}|\theta)}\right\}, \end{aligned}$$

where the equalities for $\Phi_1(v)$ and Φ'_1 are by $F_v(r + 1|e(v)) = 1$ and $F_v(r - 1|e(v)) = 0$ for all $r \in [0, 1]$ and v .

Lemma 2 then implies Theorem 2 by the following reasoning:

$$\begin{aligned} &p_s(\theta) \\ &= p_s(\theta, \mathbf{r}, 1) \end{aligned}$$

$$\begin{aligned}
&= \begin{cases} 0 & \text{if } \theta \in \Theta_s^n \text{ or } (MID_{\theta_s}^v = 1 \text{ or } r_v + 1 \leq MID_{\theta_s}^v \text{ for some } v) \\ \prod_v (1 - F_v(MID_{\theta_s}^v | \theta)) & \text{if } \theta \in \Theta_s^a, MID_{\theta_s}^v \neq 1 \text{ and } r_v + 1 > MID_{\theta_s}^v \text{ for all } v, \\ \prod_v (1 - F_v(MID_{\theta_s}^v | \theta)) \times \max \left\{ 0, \frac{F_{v(s)}(\tau_s | \theta) - F_{v(s)}(MID_{\theta_s}^{v(s)} | \theta)}{1 - F_{v(s)}(MID_{\theta_s}^{v(s)} | \theta)} \right\} & \text{if } \theta \in \Theta_s^c, MID_{\theta_s}^v \neq 1, \text{ and } r_v + 1 > MID_{\theta_s}^v \text{ for all } v, \end{cases} \\
&= \begin{cases} 0 & \text{if } \theta \in \Theta_s^n, \\ \prod_v (1 - F_v(MID_{\theta_s}^v | \theta)) & \text{if } \theta \in \Theta_s^a, \\ \prod_v (1 - F_v(MID_{\theta_s}^v | \theta)) \times \max \left\{ 0, \frac{F_{v(s)}(\tau_s | \theta) - F_{v(s)}(MID_{\theta_s}^{v(s)} | \theta)}{1 - F_{v(s)}(MID_{\theta_s}^{v(s)} | \theta)} \right\} & \text{if } \theta \in \Theta_s^c, \end{cases} \\
&= \begin{cases} 0 & \text{if } \theta \in \Theta_s^n \\ \prod_v (1 - F_v(MID_{\theta_s}^v | \theta)) & \text{if } \theta \in \Theta_s^a \\ \prod_{v \neq 0} (1 - F_v(MID_{\theta_s}^v | \theta)) \times \max \{0, \tau_s - MID_{\theta_s}^0\} & \text{if } \theta \in \Theta_s^c \text{ and } v(s) = 0 \\ \prod_v (1 - F_v(MID_{\theta_s}^v | \theta)) \times \max \left\{ 0, \frac{F_{v(s)}(\tau_s | \theta) - F_{v(s)}(MID_{\theta_s}^{v(s)} | \theta)}{1 - F_{v(s)}(MID_{\theta_s}^{v(s)} | \theta)} \right\} & \text{if } \theta \in \Theta_s^c \text{ and } v(s) \neq 0 \end{cases}
\end{aligned}$$

where the second equality is by the expressions for $\Phi_1(v)$ and Φ'_1 established above. The second last equality is by the following reason: If $MID_{\theta_s}^v = 1$, we have $F_v(MID_{\theta_s}^v | \theta) = 1$ and so $p_s(\theta) = 0$, which is nested by the other two cases. Assuming $MID_{\theta_s}^v \neq 1$, it is true for sure that $MID_{\theta_s}^v < r_v + 1$ for all v (since $MID_{\theta_s}^v < 1 < r_v + 1$). The last equality is because the lottery number $v = 0$ uses the known distribution of $U[0, 1]$ and so $F_0(r | \theta) = r$ for all r and θ .

A.2.3 Theorem 3

Lemma 3. *In the mixed tie-breaking setting of Section 4, consider a continuum economy populated by applicants of type $\theta \in \Theta$ to be assigned to schools indexed by $s \in S$. For all s and θ , for sufficiently small $\delta > 0$:*

$$p_s(\theta, r, \delta) = \begin{cases} 0 & \text{if } \theta \in \Theta_s^n \text{ or } (MID_{\theta_s}^v = 1 \text{ or } r_v < MID_{\theta_s}^v \text{ for some } v) \\ \Phi^* & \text{if } \theta \in \Theta_s^a, MID_{\theta_s}^v \neq 1, \text{ and } r_v \geq MID_{\theta_s}^v \text{ for all } v \\ \Phi^* \times \max \left\{ 0, \frac{\tau_s - MID_{\theta_s}^0}{1 - MID_{\theta_s}^0} \right\} & \text{if } v(s) = 0, \theta \in \Theta_s^c, MID_{\theta_s}^v \neq 1, \text{ and } r_v \geq MID_{\theta_s}^v \text{ for all } v \\ \Phi^* \times \frac{F_{v(s)}(\tau_s | e(v)) - F_{v(s)}(\tau_s - \delta | e(v))}{F_{v(s)}(\tau_s + \delta | e(v)) - F_{v(s)}(\tau_s - \delta | e(v))} & \text{if } v(s) \neq 0, \theta \in \Theta_s^c, MID_{\theta_s}^v \neq 1, \text{ and } r_v \geq MID_{\theta_s}^v \text{ for all } v \end{cases} \quad (13)$$

where

$$\Phi^*(v) \equiv \left(\frac{F_v(MID_{\theta_s}^v + \delta | e(v)) - F_v(MID_{\theta_s}^v | e(v))}{F_v(MID_{\theta_s}^v + \delta | e(v)) - F_v(MID_{\theta_s}^v - \delta | e(v))} \right)^{1\{MID_{\theta_s}^v = r_v\}}$$

$$\Phi^* \equiv (1 - MID_{\theta_s}^0) \Pi_{v=1}^V \Phi^*(v)$$

Proof. The first line follows from the fact that $r_v < MID_{\theta_s}^v$ is equivalent to $r_v + \delta < MID_{\theta_s}^v$ for sufficiently small $\delta > 0$.

To get the remaining lines, note that conditional on $MID_{\theta_s}^v \neq 1$ and $r_v \geq MID_{\theta_s}^v$ for all v , Φ_δ is the same as Φ^* by the following reason: When $MID_{\theta_s}^v = r_v$, $\Phi_\delta(v)$ is the same as $\Phi^*(v)$ by substituting $MID_{\theta_s}^v = r_v$ into the definition of $\Phi_\delta(v)$; this is true for any fixed δ . Otherwise, $MID_{\theta_s}^v < r_v$ and so for sufficiently small δ ,

$$\Phi_\delta(v) \equiv \frac{F_v(r_v + \delta|\theta) - \max\{F_v(MID_{\theta_s}^v|\theta), F_v(r_v - \delta|\theta)\}}{F_v(r_v + \delta|\theta) - F_v(r_v - \delta|\theta)} = \frac{F_v(r_v + \delta|\theta) - F_v(r_v - \delta|\theta)}{F_v(r_v + \delta|\theta) - F_v(r_v - \delta|\theta)} = 1.$$

$\Phi^* = \Phi_\delta$ therefore holds and Φ^* provides the probability of not being assigned to a school preferred to s .

The second line is then by the equivalence between $r_v > MID_{\theta_s}^v$ and $r_v + \delta > MID_{\theta_s}^v$ for small enough $\delta > 0$.

To see the third line, use the same equivalence between $r_v > MID_{\theta_s}^v$ and $r_v + \delta > MID_{\theta_s}^v$ for small enough $\delta > 0$ and the fact that if $v(s) = 0$, then

$$\Phi'_\delta = \max \left\{ 0, \frac{F_0(\tau_s|e(v)) - F_0(MID_{\theta_s}^0|e(v))}{1 - F_0(MID_{\theta_s}^0|e(v))} \right\} = \max \left\{ 0, \frac{\tau_s - MID_{\theta_s}^0}{1 - MID_{\theta_s}^0} \right\},$$

where we use the fact that the lottery number $v = 0$ uses the known distribution of $U[0, 1]$ and so $F_0(r|\theta) = r$ for all r and θ .

Finally, to see the last line, again note the same equivalence between $r_v > MID_{\theta_s}^v$ and $r_v + \delta > MID_{\theta_s}^v$ for small enough $\delta > 0$. Also, if $v(s) \neq 0$, then for sufficiently small $\delta > 0$ and any $\theta \in \Theta_s^c$, which implies $\rho_{\theta_s} = \rho_s$ and $r_{v(s)} = \tau_s$, we have

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

where we invoke Assumption 1 (b), which implies $MID_{\theta_s}^v \neq \tau_s$. \square

In Lemma 3, in the first case, $p_s(\theta, \mathbf{r}, \delta)$ is constant (0) for any small enough δ . The constant value is also the limit $p_s(\theta, \mathbf{r})$ in this case.

To characterize $p_s(\theta, \mathbf{r})$ in the remaining cases, note that by the differentiability of $F_v(\cdot|e(v))$ (recall Assumption 1), L'Hopital's rule implies:

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

$$\lim_{\delta \rightarrow 0} \frac{F_v(MID_{\theta_s}^v + \delta|e(v)) - F_v(MID_{\theta_s}^v|e(v))}{F_v(MID_{\theta_s}^v + \delta|e(v)) - F_v(MID_{\theta_s}^v - \delta|e(v))} = \frac{F'_v(MID_{\theta_s}^v|e(v))}{2F'_v(MID_{\theta_s}^v|e(v))} = 0.5,$$

which implies $\lim_{\delta \rightarrow 0} \Phi^*(v) = 0.5^{1\{r_v = MID_{\theta_s}^v\}}$ and so

$$\lim_{\delta \rightarrow 0} \Phi^* = (1 - MID_{\theta_s}^0)0.5^{\sum_v 1\{r_v = MID_{\theta_s}^v\}} = (1 - MID_{\theta_s}^0)0.5^{n_{\theta_s}(\mathbf{r})}.$$

Combining these limiting facts with Lemma 3 and the fact that the limit of a product of functions equals the product of the limits, we obtain the following limit propensity score for the remaining cases:

$$\begin{aligned}
p_s(\theta, \mathbf{r}) = & \\
\left\{ \begin{array}{ll} 0 & \text{if } (\theta, \mathbf{r}) \in \Theta_s^n \text{ or } r_v < MID_{\theta_s}^v \text{ or } MID_{\theta_s}^v = 1 \text{ for some } v \geq 1 \\ 0.5^{m_{\theta_s}(\mathbf{r})}(1 - MID_{\theta_s}^0) & \text{if } (\theta, \mathbf{r}) \in \Theta_s^a \text{ and } r_v \geq MID_{\theta_s}^v \neq 1 \text{ for all } v \geq 1 \\ 0.5^{m_{\theta_s}(\mathbf{r})} \max \{0, \tau_s - MID_{\theta_s}^0\} & \text{if } (\theta, \mathbf{r}) \in \Theta_s^c \text{ and } r_v \geq MID_{\theta_s}^v \neq 1 \text{ for all } v \geq 1 \text{ and } v(s) = 0 \\ 0.5^{1+m_{\theta_s}(\mathbf{r})}(1 - MID_{\theta_s}^0) & \text{if } (\theta, \mathbf{r}) \in \Theta_s^c \text{ and } r_v \geq MID_{\theta_s}^v \neq 1 \text{ for all } v \geq 1 \text{ and } v(s) \neq 0 \end{array} \right. \quad (14)
\end{aligned}$$

A.2.4 Theorem 4

The proof uses a lemma proven below. This lemma shows that the vector of cutoffs computed for the sampled economy, $\hat{\mathbf{c}}_n$, converges to the vector of cutoffs in the continuum economy.

Lemma 4. (*Cutoff almost sure convergence*) $\hat{\mathbf{c}}_n \xrightarrow{a.s.} \mathbf{c}$ where \mathbf{c} denotes the vector of continuum economy cutoffs.

This result implies Theorem 4. Note first that $\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n)$ is almost everywhere continuous in finite sample cutoffs $\hat{\mathbf{c}}_n$, finite sample MIDs ($MID_{\theta_s}^v$), and bandwidth δ_n . Since every $MID_{\theta_s}^v$ is almost everywhere continuous in finite sample cutoffs $\hat{\mathbf{c}}_n$, $\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n)$ is almost everywhere continuous in finite sample cutoffs $\hat{\mathbf{c}}_n$ and bandwidth δ_n . Recall $\delta_n \rightarrow 0$ by assumption while $\hat{\mathbf{c}}_n \xrightarrow{a.s.} \mathbf{c}$ by Lemma 4. Therefore, by the continuous mapping theorem, as $n \rightarrow \infty$, $\hat{p}_{ns}(\theta, \mathbf{r}, \delta_n)$ almost surely converges to $\hat{p}_{ns}(\theta, \mathbf{r}, 0)$ with \mathbf{c} replacing $\hat{\mathbf{c}}_n$, which is $p_s(\theta, \mathbf{r})$.

Proof of Lemma 4

Lemma 4 is proved using a strategy similar to that used to prove Lemma 3 in Abdulkadiroğlu et al. (2017a). 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 $\mathbf{r}_0, \mathbf{r}_1 \in [0, 1]^V$ 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

$$I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1) = \{i \in I \mid \theta_i \in \Theta_0, r_{v0} < r_{iv} \leq r_{v1} \text{ for all } v\}.$$

In a continuum economy,

$$G(I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1)) = E[1\{i \in I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1)\}],$$

where the expectation is assumed to exist. In a finite economy with n applicants,

$$G(I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1)) = \frac{|I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1)|}{n}.$$

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

$$d(G, G') = \sup_{\Theta_0 \subset \Theta, \mathbf{r}_0, \mathbf{r}_1 \in [0, 1]^{V+1}} |G(I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1)) - G'(I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1))|.$$

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 $g_n \rightarrow G$ in metric space (\mathcal{G}, d) . Let \mathbf{c}_n denote the admissions cutoffs in g_n . Note the \mathbf{c}_n is constant because this is the cutoff for a particular realized economy 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 $\{\tilde{\mathbf{c}}_n\}$. Let $\tilde{\mathbf{c}} \equiv (\tilde{c}_s)$ be the limit of $\tilde{\mathbf{c}}_n$. The following two claims establish that $\tilde{\mathbf{c}}_n \rightarrow \tilde{\mathbf{c}}$, the cutoff associated with G .

Claim 1. $\tilde{c}_s \geq c_s$ for every $s \in S$.

Proof of Claim 1. This is proved by contradiction in three steps. Suppose to the contrary that $\tilde{c}_s < c_s$ for some s . Let $S' \subset S$ be the set of schools the cutoffs of which are strictly lower under $\tilde{\mathbf{c}}$. For any $s \in S'$, define $I_n^s = \{i \in I | \tilde{c}_{ns} < \pi_{\theta_{is}}(\mathbf{r}_i) \leq c_s \text{ and } i \text{ ranks } s \text{ first}\}$ 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 ranking school s first who have an applicant rank in between \tilde{c}_{ns} and c_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 such that for any $n > N$, 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 such that for any $n > N$, we have $G(I_n^s) > 0$. Suppose, to the contrary, that there exists $s \in S'$ such that for all N , there exists $n > N$ such that $G(I_n^s) = 0$. When we consider the subsequence of realized economies $\{\tilde{g}_n\}$, we find that

$$\begin{aligned} & \tilde{g}_n(\{i \in Q_s(\mathbf{c}_n) \text{ such that } \pi_{\theta_{is}}(\mathbf{r}_i) \leq c_s\}) \\ &= \tilde{g}_n(\{i \in Q_s(\mathbf{c}_n) \text{ such that } \pi_{\theta_{is}}(\mathbf{r}_i) \leq \tilde{c}_{ns}\}) \end{aligned} \tag{15}$$

$$\begin{aligned} & \quad + \tilde{g}_n(\{i \in Q_s(\mathbf{c}_n) \text{ such that } \tilde{c}_{ns} < \pi_{\theta_{is}}(\mathbf{r}_i) \leq c_s\}) \\ &= \tilde{g}_n(\{i \in Q_s(\mathbf{c}_n) \text{ such that } \pi_{\theta_{is}}(\mathbf{r}_i) \leq \tilde{c}_{ns}\}) \end{aligned} \tag{16}$$

$$\leq q_s. \tag{17}$$

Expression (16) follows from Assumptions 1 and 2 by the following reason. (16) does not hold, i.e., $\tilde{g}_n(\{i \in Q_s(\mathbf{c}_n) \text{ such that } \tilde{c}_{ns} < \pi_{\theta_{is}}(\mathbf{r}_i) \leq c_s\}) > 0$ only if $G(\{i \in I | \tilde{c}_{ns} < \pi_{\theta_{is}}(\mathbf{r}_i) \leq c_s\}) > 0$. This and Assumptions 1 and 2 imply $G(\{i \in I | \tilde{c}_{ns} < \pi_{\theta_{is}}(\mathbf{r}_i) \leq c_s \text{ and } i \text{ ranks } s \text{ first}\}) \equiv G(I_n^s) > 0$, a contradiction to $G(I_n^s) = 0$.

Since \tilde{g}_n is realized as n iid samples from G , $\tilde{g}_n(\{i \in I | \tilde{c}_{ns} < \pi_{\theta_{is}}(\mathbf{r}_i) \leq c_s\}) = 0$. Expression (17) 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{c}_{ns} violates the definition of DA cutoffs at s in \tilde{g}_n since expression (17) means it is possible to increase the cutoff \tilde{c}_{ns} to c_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 $g_n \rightarrow G$ and so $\tilde{g}_n \rightarrow G$, there exists N such that for all $n > N$, we have $\tilde{g}_n(I_n^s) > G(I_n^s) - \epsilon > 0$. Since the number of schools is finite, such N can be taken uniformly over all $s \in S$. This completes the argument for Step (a).

Step (a) allows us to find some N such that for any $n > N$, $\tilde{g}_n(I_n^s) > 0$ for all $s' \in S'$. Let $\tilde{s}_n \in S$ and t be such that $\tilde{c}_{ns}^{t-1} \geq c_s$ for all $s \in S$ and $\tilde{c}_{n\tilde{s}_n}^t < c_{\tilde{s}_n}$. That is, \tilde{s}_n is one of the first schools the cutoff of which falls strictly below $c_{\tilde{s}_n}$ 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 $\tilde{c}_{ns} < c_s$ 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 such that for all $n > N$, 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 , $\tilde{c}_{m\tilde{s}_m} \leq \tilde{c}_{m\tilde{s}_m}^t < c_{\tilde{s}_m}$ for all m and so $\tilde{c}_{\tilde{s}_m} < c_{\tilde{s}_m}$, 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 rank \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 $\tilde{g}_n(A_{\tilde{s}_n}) \rightarrow G(A_{\tilde{s}_n})$ 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 $\pi_{is_i} > \tilde{c}_{ns_i}^{\tilde{t}}$. 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 $\pi_{is_i} \leq \tilde{c}_{ns_i}^{\tilde{t}}$. 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 $\pi_{is_i} \leq \tilde{c}_{ns_i} < c_{\tilde{s}_n}$, 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 $\tilde{c}_{n\tilde{s}_n}^t$ at least until step t and so $\tilde{c}_{n\tilde{s}_n}^t < c_{\tilde{s}_n}$ 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 $\pi_{is_i} > \tilde{c}_{ns_i}^{\tilde{t}} \geq \tilde{c}_{ns_i}$, where the last inequality is implied by the fact that in every economy, for all $s \in S$ and $t \geq 0$, we have $c_s^{t+1} \leq c_s^t$. Also, they are assigned s_i in G so that $\pi_{is_i} \leq c_{s_i}$. These imply $c_{s_i} > \tilde{c}_{ns_i}^{\tilde{t}} \geq \tilde{c}_{ns_i}$. That is, the cutoff of s_i falls below c_{s_i} 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 $\tilde{c}_s \geq c_s$ for all $s \in S$, as desired. \square

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

Since $\tilde{c}_s \geq c_s$ and $\tilde{c}_s \leq c_s$ for all s , it must be the case that $\tilde{\mathbf{c}}_n \rightarrow \mathbf{c}$. The following claim uses this to show that $\mathbf{c}_n \rightarrow \mathbf{c}$.

Claim 3. *If $\tilde{\mathbf{c}}_n \rightarrow \mathbf{c}$ for every convergent subsequence $\{\tilde{\mathbf{c}}_n\}$ of $\{\mathbf{c}_n\}$, then $\mathbf{c}_n \rightarrow \mathbf{c}$.*

Proof of Claim 3. Since $\{\mathbf{c}_n\}$ 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 $\{\tilde{\mathbf{c}}_n\}$, we have $\tilde{\mathbf{c}}_n \rightarrow \mathbf{c}$, but $\mathbf{c}_n \not\rightarrow \mathbf{c}$. Then there exists $\epsilon > 0$ such that for all $k > 0$, there exists $n_k > k$ such that $\|\mathbf{c}_{n_k} - \mathbf{c}\| \geq \epsilon$. Then the subsequence $\{\mathbf{c}_{n_k}\}_k \subset \{\mathbf{c}_n\}$ has a convergent subsequence that does not converge to \mathbf{c} (since $\|\mathbf{c}_{n_k} - \mathbf{c}\| \geq \epsilon$ for all k), which contradicts the supposition that every convergent subsequence of $\{\mathbf{c}_n\}$ converges to \mathbf{c} . \square

The last step in the proof of Lemma 4 relates this fact to stochastic convergence.

Claim 4. *$\mathbf{c}_n \rightarrow \mathbf{c}$ implies $\hat{\mathbf{c}}_n \xrightarrow{a.s.} \mathbf{c}$*

Proof of Claim 4. This proof is based on two off-the-shelf asymptotic results from mathematical statistics. First, let G_n be the distribution over $I(\Theta_0, \mathbf{r}_0, \mathbf{r}_1)$'s generated by randomly drawing n applicants from G . Note that G_n is random since it involves randomly drawing n applicants. $G_n \xrightarrow{a.s.} G$ by the Glivenko-Cantelli theorem (Theorem 19.1 in van der Vaart (2000)). Next, since $G_n \xrightarrow{a.s.} G$ and $\mathbf{c}_n \rightarrow \mathbf{c}$, the Extended Continuous Mapping Theorem (Theorem 18.11 in van der Vaart (2000)) implies that $\hat{\mathbf{c}}_n \xrightarrow{a.s.} \mathbf{c}$, completing the proof of Lemma 4. \square