# Distance to Schools and Equal Access in School Choice Systems

Mariana Laverde[*][†]

January 3, 2022

Most recent version

**Abstract**

This paper studies the limits of school choice policies in the presence of residential segregation. Using data from the Boston Public Schools choice system, I show that white prekindergarteners are assigned to higher-achieving schools than minority students, and that cross-race school achievement gaps under choice are no lower than would be generated by a neighborhood assignment rule. To understand why choice-based assignments do not reduce gaps in school achievement, I use data on applicants' rank-order choices to estimate preferences over schools, and consider a series of counterfactual assignments. I find that half of the gap in school achievement between white and Black or Hispanic students is explained by minorities' longer travel distance to high-performing schools. Differences in demand parameters explain a smaller fraction of the gap, while algorithm rules have no effect.

[*]I am grateful to Scott Ashworth, Kerwin Charles, Steven Durlauf, Seth Zimmerman for continuous advice and support. I also thank Santiago Caicedo, Michael Dinerstein, Margaux Luflade, Derek Neal, Luis Martinez, Peng Shi, Charlie Murry, Peter Ganong and the participants of the Harris Ph.D. Workshop and the Harris Student Applied Working Group, for helpful comments and discussions. I thank Lisa Harvey and Apryl Clarkson from Boston Public Schools for providing data and guidance. I gratefully acknowledge support by the Successful Pathway from School to Work initiative of the University of Chicago, funded by the Hymen Milgrom Supporting Organization. The conclusions of this research do not necessarily reflect the opinions or official position of Boston Public Schools, or the Massachusetts Department of Education. All errors are my own. This paper previously circulated under the name "Unequal Assignments to Public Schools and the Limits of School Choice".

[†]Department of Economics, Boston College. Email: mariana.laverde@bc.edu. Maloney Hall, Chestnut Hill, MA 02467, USA

# 1 Introduction

Since the late 1980s, many cities across the United States have adopted centralized school choice systems.[1] These systems allow families a choice among public schools, as opposed to neighborhood assignments where school districts assign students to schools based on proximity to residences. Neighborhood assignments replicate residential segregation and can sustain educational inequality across racial and income groups. Proponents of choice argue that by decoupling residences and schools, choice systems can reduce school segregation, improve student-school match quality, and equalize access to high-performing schools. In doing so, these systems generate competitive pressures that drive ineffective schools to improve. Among these objectives, school districts that implement school choice emphasize that a guiding principle of a student assignment plan includes creating equitable access to high-quality schools.[2]

This paper asks how effectively choice systems reduce cross-racial gaps in access to high-achieving schools. Using assignment data from Boston Public Schools (BPS), I begin by showing that under Boston's choice system, white prekindergartners are assigned to higher-achieving schools than Black and Hispanic students. Moreover, I document that cross-race school achievement gaps under choice are no lower than would be generated by an assignment based on proximity between residences and schools.[3]

An effective policy response to the above depends on an understanding of why the effects

[1] According to the non-profit *Education Commission of the States*, 47 states plus the District of Columbia have passed laws to allow or mandate a version of school choice. School districts that have implemented open enrollment include New York, Boston, Cambridge, Charlotte, and New Haven.

[2] Boston Public Schools superintendent expressed this in the proposal for the 1988 choice plan: "My overall goal is to create a student assignment plan that provides all Boston students with high-quality desegregated education" (Boston Desegregation Project 1988). Other example includes the Charlotte-Mecklenburg School District (https://www.cms.k12.nc.us/boe/Pages/2010%20GuidingPrinciplesforStudentAssignment.aspx)

[3] I generate a neighborhood assignment matching students to schools in order of proximity while taking into account school capacities. Specifically, I run a DA algorithm where preferences and priorities are fully determined by distance.

of choice are limited. Cross-race differences in choice-based assignments stem from either differences in demand for high-achieving schools, or from assignment rules that generate different probabilities of assignment conditional on parents' preferences. Thinking about demand, a key component is the distance between a school and the family's residence. A large body of evidence shows that parents value proximity to home when choosing a school (see Agarwal and Somaini 2019 for a summary). Parents may care about distance because their schedules are not flexible to allow them to get their children to schools that are farther, or might worry about longer commutes in school buses. This implies that the benefit of attending high-achieving schools may be upset for families who live far to these schools. I estimate how differences in distance to high-achieving schools contribute to cross-race gaps in access to these schools.

Turning to rules, those that link assignments to a student's residential location may contribute to the gap. Two assignment rules in Boston do this. First, students are prioritized for assignment based on proximity to schools. This means, that students who live closer to high-achieving schools are more likely to get assigned to these schools.[4] If white families live closer to high-achieving schools they have higher priority at these schools. Giving priority to students based on proximity to schools is common across school choice systems. Examples include the cities of New York and Barcelona (Abdulkadiroğlu et al. 2005, Calsamiglia and Güell 2018). The second rule limits the menu of schools a student can apply to based on her residential location.[5] If the menus of Black, Hispanic and white students have a different share of high-achieving seats, this restriction can mechanically contribute to differences in

[4]Although this priority mechanically increases the probability that a student in the walk-zone of a school—defined as a one mile radius—has a higher probability of being assigned to it relative to students who lives farther, Dur et al. (2018) show that under the design of Boston's algorithm this rule did not importantly increase the fraction of walk-zone students relative to an assignment where the proximity priority is abolished. This is explained by the precedence order between seats with a proximity priority and seats without it.

[5]BPS has had this type of restrictions since the early 1990s, and modified these menus in 2014, after the end of my study period.

access to high-performing schools.

To disentangle the contribution of differences in distance related costs, preferences for location-independent school attributes, and assignment rules, I first estimate a model of school demand using data on the rankings submitted by all first-round applicants to prekindergarten between 2010 and 2013. Under some identification assumptions detailed in Section 4, the demand model allows me to separately identify parental preferences for proximity and the average valuation for each school net of travel costs. In a second step, I use the preference parameters estimated in step one to generate counterfactual assignments that help me quantify the contribution of each mechanism. Under these counterfactual assignments, I vary the distance to schools, parents' demand parameters, and the assignment rules, to quantify the change in the gap under each.

To estimate the contribution of travel costs I study a counterfactual change in residential location. Here I ask, how would the ranking and subsequent assignment of a single Black or Hispanic student change if he faced the menu of distances that a typical white student faces? To answer this question, I randomly assign to each Black or Hispanic student a counterfactual residential location from the distribution of white students' locations, and generate counterfactual assignments in the new location using the parameters of the model and the assignment algorithm. By changing the location of a single student, I am able to evaluate the effects of a location change under the assumption that there are no changes to the schools' demographic composition, that might affect parental preferences. This counterfactual parallels the Moving to Opportunity (MTO) experiment that relocated families from high-poverty neighborhoods to low-poverty communities in the late 1990's.[6] While the papers that study the MTO experiment study medium and long-term consequences of the relocation on health, income and labor outcomes; results from the counterfactual I propose capture the immediate effect of reducing the cost of accessing high-performing public education.

[6] Papers that study the impacts of this experiment include Ludwig et al. (2013), Chetty et al. (2016), Katz et al. (2001), Kling et al. (2007), Clampet-Lundquist and Massey (2008)

In this context, changing the residential location of a student doesn't only change the distance menu. Students who are relocated may select schools from a different choice menu, and have a proximity priority at a different set of schools. To disentangle the effect of travel costs and assignment rules, in a second counterfactual I generate assignments assuming that there are no restrictions on choice menus, and later consider the case where proximity priorities are eliminated. The results from these counterfactuals pin down the effect of algorithm rules, and in combination with the results from the location change counterfactual, pin down the effect of travel costs.

To estimate the effect of heterogeneity in the demand for location-independent school attributes, I simulate assignments assuming a change in parental demand parameters. I generate assignments where Black and Hispanic students take white students' demand parameters, while the original residential location of each student is unchanged. Results from this counterfactual highlight how differences in demand for any location-independent school attribute impact the observed gap. Differences across races in these parameters may capture any dimension of heterogeneity in parental preferences, including the racial composition of students, teachers and staff, the languages taught at each school, or the schools' teaching and discipline methods and curriculum choices.

I find that after a change in residential location, the gap in school achievement between minority students and white students was reduced by around 50%, and a change in demand parameters explains about 30% for Hispanic families and 40% for Black families. Eliminating proximity priorities and choice menu restrictions has a negligible impact on the distribution of school achievement by race. This suggests that the effect of the residential location change is fully explained by changes in travel costs to high-achieving schools and not by location-specific assignment rules. These results are robust if I consider a random utility function that depends linearly or non-linearly on distance, and if I allow the valuation of the outside option to be neighborhood-specific.

The salience of travel costs on the resulting school choice assignments has important policy implications. It shows that even though choice systems give all families the option to sort into the best schools in the city, if families face different costs of accessing those schools, gaps in access to high-quality schools will not vanish. A weakened demand for higher-quality schools explained by differences in distance will lower competitive pressures, limiting the potential for improvement. Although the effects of distance may be weaker for high-schoolers and other older students who plausibly face lower transportation costs, barriers to access high-achieving schools in the earlier years may be critical for longer-term outcomes. Moreover, since under school choice, students are grandfathered into subsequent grades in the same school, students are likely to attend the same school from prekindergarten through the end of elementary school, amplifying the potential inequities.

**Related Literature.** This paper relates to several literatures. The first strand examines the effectiveness of school choice in generating system-wide improvements in school productivity. One side of the debate argues that by fostering school competition, choice systems boost school effectiveness (Friedman 1982, Chubb and Moe 1990, Hoxby 2003, Campos and Kearns 2021). In this group of papers, a recent evaluation of the Zones of Choice (ZOC) program in Los Angeles, finds that the program generated improvements in school effectiveness associated with competition, closing achievement and college enrollment gaps between ZOC neighborhoods and the rest of the city (Campos and Kearns 2021). On the other side, choice systems may not generate system-wide improvements if parents do not rank schools on the basis of school effectiveness (Abdulkadiroğlu et al. 2020, Hastings et al. 2009, Barseghyan et al. 2014, Borghans et al. 2015). Abdulkadiroğlu et al. (2020) finds that conditional on peer composition, parental valuation of schools in New York City (NYC) is unrelated to school effectiveness. Adding to this evidence, this paper shows that if effective schools are concentrated in some areas of the city, some parents may rank nearby schools of lower quality instead of effective ones.[7] This is consistent with findings by Angrist et al. (2021) that the

[7] Neilson 2013 and Allende 2019 find that horizontal differentiation across schools explained by distance contributes to reduced competition in Chile and Peru, under systems with public and private supply and choice in the form of vouchers.

effect of traveling to an out of walk-zone school under Boston's and NYC's choice systems does not result in test-score gains for Black and Hispanic 6th and 9th graders.

Moreover, this paper is related to the literature that studies the impact of school choice policies on student sorting. Most of the papers in this strand of literature focus on studying the effects of voucher policies on the composition of the student body by achievement and income, in both the public and private sectors (Epple and Romano 1998, Epple et al. 2004, Hsieh and Urquiola 2006, Altonji et al. 2015). I study an open enrollment plan, and the mechanisms that explain the observed sorting by race into schools by achievement levels.

This paper also contributes to the literature on neighborhood effects. I show that choice systems alone may not be sufficient to equalize opportunity for residents of impoverished neighborhoods. Growing up in low-opportunity areas has been found to be related to adult earnings and educational achievement (Chetty et al. 2014, Chetty et al. 2016, Chetty and Hendren 2018, Chetty et al. 2018), and some of these effects may be explained by the provision of public education in these areas (Biasi 2019, Laliberte 2018). This paper shows a first-order effect of location in the access to high-performing public education. I show that for many families the travel costs offset the benefits of attending high-achieving schools. The salience of families' perceived cost of attending a distant school and its effect on school demand is consistent with results that show substantial spatial variation of place-based effects for geographies as small as census tracts.

This paper adds some evidence to the literature that studies the impact of transportation costs on school choice policies, and how school busing can hep reduce these costs. During my study period BPS offered a very generous transportation service, that guaranteed school busing to all families assigned to a school farther than a mile from their homes, and capped every student travel time at an hour. At the time, BPS spent about 10% of it's budget on transportation, which constituted the highest per-student transportation cost in the US (Bertsimas et al. 2020)[8]. Trajkovski et al. 2021 find school buses are effective reducing travel

[8]Motivated by the high costs associated with the generous transportation system, in 2016 BPS redesigned

costs to families in NYC and can increase access to higher-performing schools under choice systems. Despite the potential for reducing travel costs, this paper shows there are limits to the impact of school buses in equalizing travel costs to high-achieving schools.

Finally, my analysis adds to a recent series of studies that leverage ranking data from centralized school assignments to study school demand and the properties of these assignments (See Agarwal and Somaini 2019 for a summary). Some of these papers study parental demand for schools under mechanisms that provide incentives to misrepresent preferences, and evaluate the welfare implications of such mechanisms. Others use rankings to study the determinants of parental demand and its implications for choice systems. In the closest paper to mine, Son (2020) quantifies the contribution of students' residential location, parental preferences, admission policies and optimization frictions on racial integration and the proportion of students assigned to their top five schools using data from the NYC high-school match. I concentrate on access to school achievement as opposed to school segregation, and focus on prekindergarteners for whom schooling investments are likely to have lasting effects.

My analysis focuses on studying differences in average school achievement. Average achievement is a bundled measure of the academic ability of the students a school enrolls, and the capacity of a school to generate improvements in student outcomes. In this paper, I am not able to speak of differences in effectiveness as opposed to peer composition. Nevertheless, schools that enroll high-achieving peers have been found to be more effective (Abdulkadiroğlu et al. 2020). This suggests that Black and Hispanic families experience longer commutes to effective schools and, just as for high-achieving schools, this will translate into inequities in the access to effective schools. This can reconcile evidence showing that parents place value on effective schools -albeit mediated by peer-composition- (Abdulkadiroğlu et al. 2020), and Black and Hispanic students don't experience gains in achievement from being assigned to schools out of their walkzone in Boston (Angrist et al. (2021)).

the bus routes to minimize the costs to the district, while maintaining student travel times (Bertsimas et al. 2020)

The rest of the paper is organized as follows. Section 2 discusses the institutional context and the data. Section 3 summarizes the main observed differences in application behavior, and discusses and presents evidence on the mechanisms. Section 4 presents the model used to recover demand parameters, discusses the assumptions, the estimation, and analyzes the results. Section 5 describes the methodology and assumptions made to run the counterfactual exercises and the results. I conclude in Section 6.

# 2 Elementary School Choice in Boston

## 2.1 The Assignment Mechanism

Parents who wish to apply for a prekindergarten seat in a school within BPS are required to submit to the school district a ranking of programs and schools ordered by preference. A school typically offers a couple of general education programs, as well as programs for language learners.[9] Students can rank any number of programs with the condition that they are housed in a school the student is eligible for. Eligibility is determined by the student's residential location. During the study period, Boston was divided into three zones: the north, east and west zones (Figure 2a). Students were eligible for any general education program in their residence zone, plus any within a mile of their home. Geographic restrictions that determine eligibility for language programs are similar to those of general education programs, nevertheless these restrictions are not always binding (Pathak and Shi 2013a). I assume, as Pathak and Shi (2013a) do, that English language learner (ELL) students can apply to any program across the city. There are also a handful of city-wide schools that accept applications from students all over the city. I refer to the set of schools a parent can apply to as the parents' choice-menu. Figure 2b shows a partition of the city that groups

[9] At the prekindergarten level general education programs are typically referred to as inclusion programs. I exclude from my analysis students applying to substantially separate programs since assignments for these students don't always follow the assignment rules and allow for exceptions when needed.

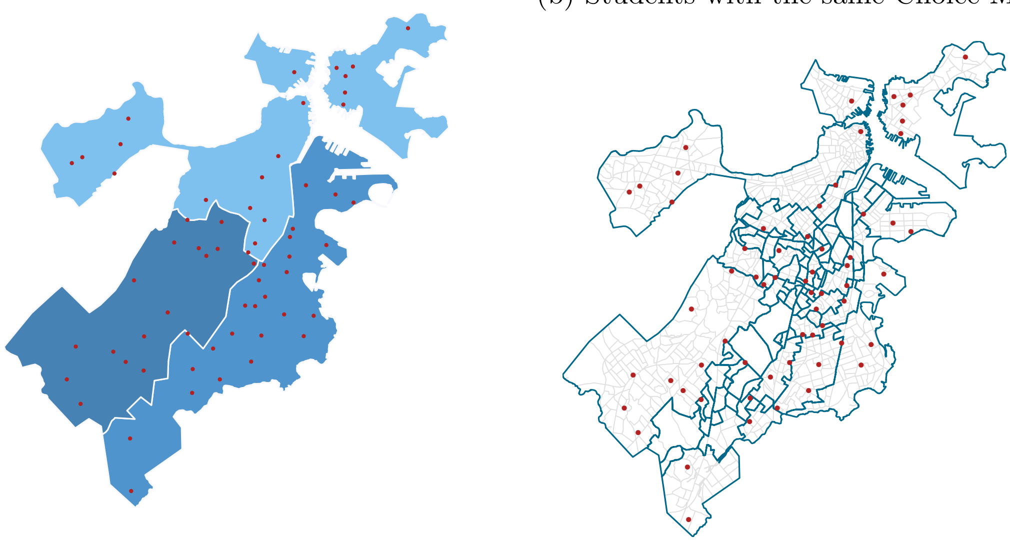
families with the same choice-menu.

Although parents in Boston apply to programs within schools, I make the simplifying assumption that parents rank schools. I transform school-program rankings into school rankings eliminating instances where different programs in the same school are ranked, and keeping the first time a school appears in the ranking (a similar assumption is made in Abdulkadiroğlu et al. 2020). Going forward I refer to parental preferences for schools.

Figure 1: Zoning and Choice Menus

(a) Zones

(b) Students with the same Choice Menu



Note: Red points are schools with a prekindergarten program in 2010. Choice menus are built using data on school and geocode coordinates.

Students are assigned to schools following a priority structure defined by the school district that is common across schools. Under this priority structure, students who have a sibling at a school have a higher priority at that school than students who do not have a sibling at the school. Also, students who live within a mile of a school—called the walk-zone of a school—have priority at that school over students that live farther away. Overall, students who both have a sibling and live in the walk-zone have the highest priority. These are followed by students who have a sibling, and then those who live in the walk-zone. The remaining students have the lowest priority. Ties within each group are broken with a random number assigned to each applicant. This guarantees that priorities generate a strict ordering of students.[10] School districts also determine school capacities, that is the number of seats available at each program. Preferences, priorities, and capacities feed the assignment algorithm that is a version of Gale and Shapley (1962) student-proposing DA algorithm (Balinski and Sönmez 1999; Abdulkadiroğlu and Sönmez 2003).

The DA algorithm guarantees that parents do not have incentives to misrepresent their true preferences when submitting rankings (Dubins and Freedman 1981, Roth 1982). This holds under the assumption that students are allowed to rank all desirable schools. Instances where school authorities restrict the length of submitted rankings may not generate truthful reports, even under the DA algorithm (Haeringer and Klijn 2009, Calsamiglia et al. 2010). BPS is one of a few districts that does not restrict the length of the submitted rankings. These properties make Boston a good setting for studying parental school demand.

Students assigned to a school farther than a mile from their homes are eligible for free bus transportation to and from school. The pick-up and drop-off location is set by the district to a location within the mile of a student's home. BPS estimates that the majority of riders are in elementary school, and attend a school with high populations of low income families. Among prekindergarten students, around half opted in for school transportation.

## 2.2 Data

I use two main data sources. First, data from BPS that covers the universe of first-round applicants to prekindergarten between the years 2010 and 2013. For each applicant I observe the rank-ordered list submitted, the school assigned or an indicator for whether the

[10] This priority structure is typically used in half of the seats in each school, while the remaining seats ignore walk-zone priorities all together. A more detailed description of the algorithm is given in the Appendix D. Dur et al. (2018) and Sönmez et al. (2019) discuss this design and it's properties.

student was unassigned, and the priority that generated the assignment.[11] I also observe the residential location[12] and demographic information of the student including their race.[13] First-round applicants represent over 80% of admitted students (Pathak and Shi 2017); the rest apply in the second round and are assigned after first-round applicants.

Second, I use yearly data on school characteristics from the Massachusetts Department of Education (DOE). From this source, I measure school achievement using the fraction of third-grade students at each school scoring advanced or proficient in the Massachusetts Comprehensive Assessment System (MCAS) math test. Most of the schools that offer a prekindergarten program also offer a third-grade program, and only a few offer up to first grade. For these I do not have measures of school achievement.[14]

Using the location of each school and the geocode of residence of each student, I measure the distance between students and schools in one of two ways: first, as the walking distance between the geocode's centroid and the school, and alternatively as the linear distance between the two points. The former is obtained using Google maps travel estimates. Using these locations, I also generate the walk-zone priority status for each student-school pair and the choice-menu of each student, recreating the procedure used by BPS.[15]

Ideally, I would have the sibling priority status of every student at every school. Nevertheless, I only observe the sibling priority status of student $i$ at school $j$, if $i$ was assigned to $j$ with

[11] A student will be unassigned if he is rejected from every school on his submitted rank list. Students who are unassigned in the first round can reapply in the second round or search for options outside the school district

[12] Residential locations are coded by the school district at the geocode level. Geocodes partition the city in 868 polygons of average area of 0.1 sq. miles. The assignment algorithm is built using such geocodes, hence that level of aggregation does not represent any loss of information for purposes of the assignment algorithm.

[13] I remove from my sample students with an invalid geocode that represent around 2% of the sample

[14] 5 schools in each year offer up to first grade

[15] Student $i$ is in the walk-zone of school $j$ if a one-mile radius from school $j$ intersects the geocode of residence of $i$. Similarly, I define the choice-menu of each student using data on the zone in which each school and geocode lies.

Table 1: Student Descriptive Statistics

| | All | Black | Hispanic | White | Asian | Other |
|---|---|---|---|---|---|---|
| *Applicants* | 8,869 | 22.9 | 42.8 | 22.8 | 7.8 | 3.6 |
| Tract Income | 55,551 | 43,705 | 49,873 | 76,753 | 55,166 | 63,660 |
| | (25,429) | (19,205) | (21,711) | (24,850) | (22,875) | (27,363) |
| *Applications* | | | | | | |
| Size of Choice Menu | 24.8 | 26.0 | 24.8 | 23.5 | 25.0 | 24.4 |
| | (2.4) | (2.2) | (2.4) | (1.9) | (1.9) | (2.3) |
| Distance in Choice Menu | 2.6 | 2.4 | 2.7 | 2.7 | 2.6 | 2.5 |
| | (0.8) | (0.7) | (0.9) | (0.8) | (0.8) | (0.8) |
| Maximum distance in Choice Menu | 5.6 | 5.5 | 5.8 | 5.3 | 5.9 | 5.3 |
| | (1.3) | (1.1) | (1.3) | (1.5) | (1.2) | (1.4) |
| Length of Submitted List | 5.0 | 5.5 | 5.0 | 4.8 | 4.1 | 5.7 |
| | (3.1) | (3.4) | (3.0) | (2.8) | (2.7) | (3.6) |
| Share English Language Learners | 37.5 | 19.4 | 58.2 | 11.4 | 64.7 | 11.7 |
| *Assignments* | | | | | | |
| Assigned Rank | 1.8 | 1.9 | 1.7 | 1.7 | 1.6 | 2.3 |
| | (2.2) | (2.1) | (1.8) | (2.7) | (1.6) | (3.3) |
| Distance to Assigned School | 1.2 | 1.3 | 1.3 | 1.0 | 1.1 | 1.2 |
| | (1.3) | (1.3) | (1.3) | (1.0) | (1.1) | (1.2) |
| Share Assigned with Sibling Priority | 36.0 | 31.3 | 34.4 | 43.8 | 40.0 | 33.9 |
| Share Assigned with Walk-Zone Priority | 48.4 | 47.4 | 46.6 | 53.5 | 48.1 | 49.1 |
| Share Unassigned | 26.1 | 23.0 | 24.2 | 33.2 | 22.7 | 30.8 |

Note: The first row of the table shows the total number of applicants and the share in each group. The second row shows the average tract-level household income taken from the 5 year 2010 ACS (I match the geocode of each applicant to a census tract by overlaying both geographies and keeping the tract with the largest share of each geocode's area). For the rest of the statistics, I show the mean and below the standard deviations in parenthesis with the exception of variables marked as shares, in which case I show a fraction. The average size of the choice menu and length of submitted list are measured in number of schools. I show linear distances measured in miles. The distance between a student and a school is the linear distance between the coordinates of each school and the centroid of the geocode of residence of each student. The length of the submitted list and the rank of the assigned school are computed under rankings transformed from school-program based to school based, the numbers under these transformations are smaller than the ones obtained under the school-program rankings. The share of students assigned with a sibling or walk-zone priorities are expressed as a fraction of all assigned students. If a students is assigned with a sibling and a walk-zone priority then it is included in both categories.

this priority. Throughout the analysis, I assume that all students that are not assigned with a sibling priority do not have a sibling at any school, and that students assigned with a sibling priority at $j$ do not have a sibling at other schools. Using data on the priorities that generated each assignment, I find evidence in support of this assumption. I find that in most schools every student who applied with a sibling priority was admitted. This means that for the set of schools each student ranks, I am able to observe a sibling status when it indeed exists, with the exception of students who have a sibling priority at multiple schools or those who rank the sibling school sufficiently low and are assigned to a school ranked higher. In the first case, I'm only able to account for the sibling status at the sibling's school ranked higher.[16]

**Students.** The sample has $8,869$ applicants to prekindergarten between 2010 and 2013. Close to half of the applicants to prekindergarten in Boston are Hispanic, while Black and white students are around one-fifth of the sample each. Asian and other minority families make-up around 10% of the applicant pool. This composition is in contrast to Boston's resident makeup, where white residents account for about half of the population.

Families can choose from a set of 25 schools on average. This contrasts with other school choice settings, such as NYC's high-school system, where families choose from about 700 options (Son 2020). Out of these options, families typically rank five options. Black students submit longer lists while white students submit shorter lists, potentially reflecting that outside options of white families are ranked higher among public schools.[17] Students who are unassigned after running the assignment algorithm may apply in a subsequent round. Since prekindergarten attendance is not mandatory, there are applicants who are not assigned to

[16] If the following conditions are satisfied a school did not reject a student with a sibling priority: First, if there are fewer assigned students than available seats then no student was rejected. Second, if a school accepted a student with either the walk-zone priority or with no priority then that school did not reject anyone with a sibling priority. Otherwise the resulting match would not be stable. The number of schools that do not satisfy either of these in 2010 is 3, in 2011 is 2 and in 2012 is 6. For these schools I cannot rule out that they rejected a student with a sibling priority.

[17] Son (2020) documents something similar in the case of NYC's high-school choice system.

any school and who need to search for options outside of the public school district. About a quarter of the students that apply in the first round are unassigned, and out of all unassigned students near 75% do not enroll in any public school.

Figure A.2 shows the spatial distribution of students by race. Although there are clear sorting patterns, students of all races can be found across the city. One way to quantify this is to zoom close to each school and see the distribution of residents in a close buffer by race. If I consider a 1.2 mile radius around every school, I find that on average there are several hundred students of each race who can apply to each school; and for all schools there are students of all races. Similarly, looking at applications I find that the average school has a couple hundred applicants from each race, and every school has applicants of all races (Table A.3).[18]

Table 2: Descriptive Statistics: Schools

| | Mean | StDev | Min | Max |
|---|---|---|---|---|
| *Capacity* | 30.9 | 15.7 | 6.0 | 108.0 |
| *Achievement* | | | | |
| % Scoring Advanced-Proficient Math | 46.1 | 19.2 | 2.0 | 90.0 |
| % Scoring Advanced-Proficient English | 37.8 | 16.0 | 10.0 | 86.0 |
| *Demographics* | | | | |
| % Black Students | 32.0 | 19.3 | 2.1 | 79.7 |
| % Hispanic Students | 44.2 | 19.3 | 14.3 | 91.1 |
| % White Students | 14.6 | 14.7 | 0.0 | 65.8 |
| % Low Income K Students | 67.5 | 19.8 | 7.7 | 100.0 |
| Observations | 258 (67 distinct schools) | | | |

Note: I do not observe achievement data for all schools in all years. There are a total of 17 missing observations (school-year pairs) of schools that do not offer third grade or for which data is restricted due to a small set of test takers.

**Schools.** Between 2010 and 2013, there were a total of 67 public schools that offered a

[18] I chose a 1.2 miles buffer because this is the average linear distance students travel to their assigned school. Choosing instead a one-mile buffer gives similar results.

prekindergarten program and not all schools had prekindergarten seats in all years. There is substantial variation in students' demographic characteristics and school achievement among these schools. While on average the share of third-grade students scoring advanced or proficient in math is 46%, the school with the lowest achievement had 2% of students scoring advanced or proficient, while for the highest-performing school the fraction was 90%. On average, schools have 32% Black students and 15% white students. Since both white and Black students represent about 20% of all applicants, this reflects the fact that schools with few white students ($< 10\%$) are about four times more likely than those with few Black students. Each school has on average 70% low-income students, and the school with the lowest fraction of low-income students has 8%.

# 3 The Gap in School Achievement and the Possible Explanations

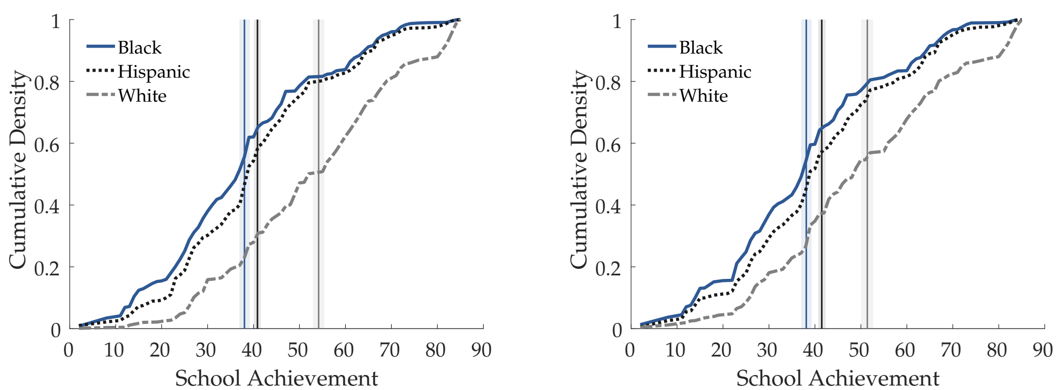
In this section I describe the two main facts that motivate the paper. Then, I discuss the mechanisms that can explain why in a choice setting we do not see a more equitable access to high-achieving schools. Finally I provide some evidence on the relevance of each mechanism.

## 3.1 The Racial Gap in School Achievement

Between 2010 and 2013, white prekindergarteners in Boston were assigned to schools that had higher average achievement and a smaller fraction of low-income students and minority students than their Black and Hispanic peers. While the average white student was assigned to a school where more than half of students scored advanced or proficient, these measures were close to 40% for Black and Hispanic students (Figure 3a). In terms of demographics, the average white student was assigned to a school with nearly 50% low-income kindergarten students, and for Hispanic and Black students the percentage is closer to 70%.[19]

Figure 3: Distribution of School Achievement under School Choice and Neighborhood Assignments

(a) School Choice DA Assignment (b) Neighborhood Assignment




Note: On the left, I plot the distribution of school achievement for the students assigned to prekindergarten between 2010 and 2013 by BPS. The measure of school achievement is the fraction of third-grade students in each school who scored advanced or proficient in the MCAS math test. On the right, I plot the distribution of school achievement under a counterfactual assignment where the same set of students are assigned to the school closest to their home, respecting school capacities.

Moreover, cross-race differences in school achievement under the choice system are not lower than those generated under a neighborhood assignment rule. Comparing the distribution of school characteristics generated by an assignment rule that uses parents' stated preferences with a neighborhood assignment serves as a good benchmark. The latter shows how these gaps would look if a neighborhood assignment were implemented under the current residential choices in Boston. I generate this alternative assignment running the DA algorithm with the set of all students assigned via the choice system, and redefining their preferences and priorities to be determined exclusively by proximity: students prefer schools closer to home, and schools prioritize students that live closer to schools. Under the proposed neighborhood

[19]If instead I considered the achievement and demographic characteristics of each school one year prior to the assignments these numbers don't change much. The gap in school achievement is 0.4 pp larger for Black students and 1.5 pp larger for Hispanic students. Data from a year before assignments are measures of the characteristics of schools that are observable to parents when they apply for admission at prekindergarten.

assignment, the distribution of school achievement is similar to that obtained under the choice rule (Figure 3b). Furthermore, I cannot reject the null hypothesis that the mean achievement is equal across assignments for Black and Hispanic students.[20]

Under this hypothetical experiment, I find that the choice and neighborhood assignments are different for around 80% of students. Out of these, white students sort into higher achieving schools while this is not the case for Black and Hispanic students. For all groups, the resulting hypothetical assignment does not show higher integration under choice, measured by the fraction of same race peers (Table A.1). Both these facts are in contrast with findings by Angrist et al. 2021 on the effects of choice in Boston for 6th and 9th graders. They find choice assignments are associated with integration and exposure to higher-achieving peers, specially for Black students.

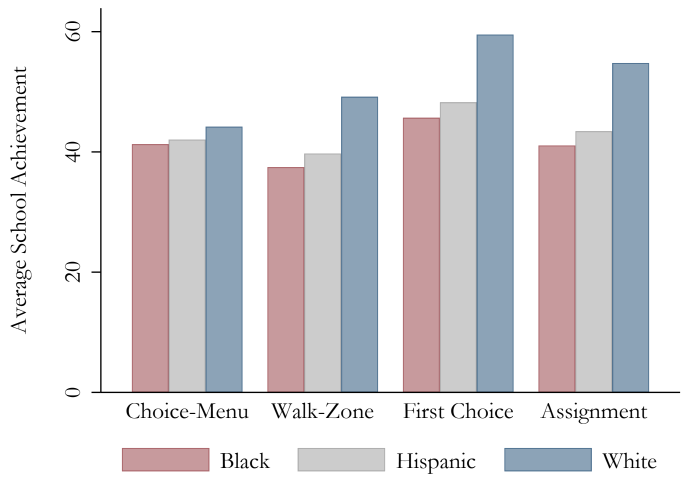
## 3.2 The Mechanisms

Studying how location effectively matters in choice-based settings is a first order concern to evaluate the equity consequences of choice-based policies. Even in a choice-based system where the link between residences and schools is weakened, the residential location of families may play a crucial role in their school assignment. If parents value proximity, the benefit of attending a high-achieving school may be upset by high travel costs. Also, assignment rules that constrain geographically the choices of families, or that prioritize students based on proximity to schools, can generate inequities even in the absence of travel costs.

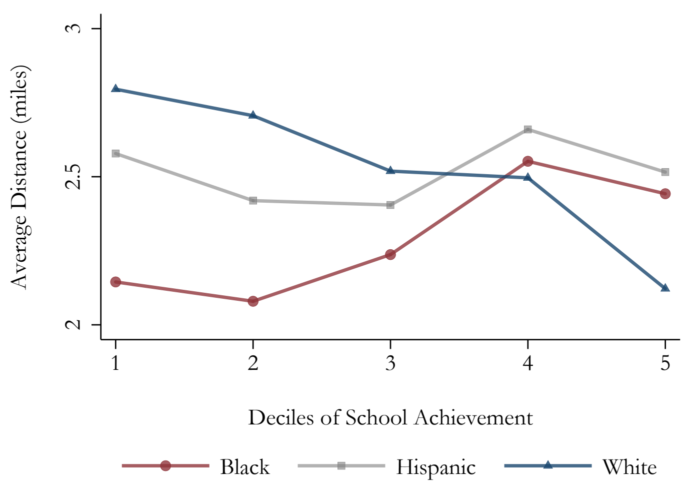
[20] Two tail p-values are 0.4 and 0.2 for Black and Hispanic students, respectively.

Figure 4: Reduced form Evidence

(a) Average School Achievement




(b) Distance to Schools by Race and Achievement




Note: Panel (a) shows the average school achievement weighted by capacity at the schools in the choice-menu and walk-zone of applicants by race. Also, the average school achievement for the schools ranked first and the schools assigned to students by race. In panel (b) I plot the average linear distance between schools and students of each race, by school achievement deciles. The positive slope for Black and Hispanic students show that, conditional on their location these families trade-off proximity and achievement. Also, white students live on average closer to schools in the top deciles of achievement.

Differences in school assignments may also be responding to differences in parental demand for location-independent school attributes. The average valuation parents place on schools, net of travel costs, aggregate the relative valuation of each school's amenities and disamenities. This includes school characteristics that are likely valued by all parents, like appropriate infrastructure, student safety, a low student-teacher ratio, and high school achievement. It also includes school characteristics that can be amenities for some families and disamenities for others, like the racial composition of students and teachers, the languages taught at each school, or the schools' teaching methods and curriculum choices. Heterogeneity in school demand is generated if parents disagree in the valuation of any single school characteristic.[21]

An analysis of the characteristics of the schools in the choice-menu of students reveals that choice-menu restrictions are unlikely to be an important contributor to the school achievement gap. The average school achievement at the schools in the choice-menu of Black, white and Hispanic students is similar (Figure 4a). On the contrary, Figure 4b shows that high-achieving schools tend to be closer to white families and this may impact assignments via walk-zone priorities or a higher school demand explained by lower commuting costs.[22]. Moreover, the figure shows that while Black and Hispanic families make trade-offs between proximity and school achievement white families do not; the schools closest to them are on average higher-achieving than those farther.[23] Having Black and Hispanic families bear larger commuting distances to high-achieving schools may impact their demand for these schools. Consistent with this, Figure 4a shows that the schools ranked first by white families are higher-achieving than those of Black and Hispanic families, and this gap resembles that of assignments. Off course, cross-race differences in school demand can also stem from heterogeneity in parental demand for location-independent school attributes. A reduced form analysis of rankings is insufficient to disentangle the contribution of preference heterogeneity

[21] Allers 2019 in The Washington Post describes how issues of representation and discrimination can be at the heart of Black families' choices, and sometimes conflict with academic attributes of schools.

[22] Related to this Walters (2018) finds suggestive evidence that in Boston charter middle schools tends to locate in lower-achieving areas of the city.

[23] See Table A.2 for regression results on these relationship

and travel costs. In the next section I discuss the model and identification assumptions used to estimate the contribution of each.

# 4 Estimating Parent Preferences

In this section, I present the model and assumptions used to recover parental preferences for schools. At the end of the section I discuss the estimated parameters and the fit of the model.

## 4.1 Model

I model preferences using a random utility model where $i \in \mathcal{I}$ index students and $j \in \mathcal{J}$ index schools. To capture rich heterogeneity in preferences I estimate separate models for 6 subgroups of students defined by the intersection of students' covariates. This strategy follows that of Abdulkadiroğlu et al. 2020 in a school choice setting and Hastings et al. 2017 and Langer 2016 in other settings. The covariate cells are defined as the intersection of the students' race and students' census tract income, where students are grouped together if their census tract income is above or below the median tract income among applicants.[24] For each covariate cell $c$ we use data on individual choices to estimate the model

$$u_{ij} = \beta_c d_{ij} + \gamma_c l_{ij} + \delta_{cj} + \epsilon_{ij} \tag{1}$$

where student $i$ is in cell $c$. The variable $d_{ij}$ denotes the walking distance from $i$'s residence to school $j$ and $l_{ij}$ captures whether a student is a language learner and school $j$ offers a program

[24] I do not include the student's English language learner status when defining clusters, and instead I capture parental preferences for same language programs in the utility model. By doing this I capture within cluster heterogeneity in school-student match effects and guarantee that preferences of English language learners are not bulked together. Otherwise I would be assuming that the value for Spanish speakers of a Cape-Verdean program is the same as the value of a program in Spanish.

in the student's first language. The parameter $\beta_c$ summarizes preferences for proximity for parents in cell $c$, and $\gamma_c$ captures parents' preferences for having a same language program. The parameter $\delta_{cj}$ summarize the location-independent attractiveness of school $j$. This includes parents' assessment of school characteristics that are observable and unobservable to the econometrician. Finally, $\epsilon_{ij}$ represents $i$'s idiosyncratic taste for school $j$. I assume the $\epsilon_{ij}$ are independent and distributed type-1 extreme value with scale parameter $\lambda_c$.

**Truth-telling.** I assume that submitted rankings are truthful. This means that parents rank all acceptable schools in true preference order. A school is acceptable if it is preferred to the outside option, which is the best option parents can find if unassigned in the first round. This assumption is motivated by the algorithm's incentive compatibility and the property that there are no restrictions on the number of schools parents can rank. Having restrictions over the length of submitted lists, even under the DA mechanism, can generate reports that are not truthful (Haeringer and Klijn 2009, Calsamiglia et al. 2010, Luflade 2018). Boston's choice system satisfies both properties.

Truth-telling can be violated if admission outcomes are largely predictable, or participants think they are. In this case, parents may misrepresent preferences by not ranking schools that are desirable but where parents perceive a low probability of admission. This is more likely to happen in settings where an applicant knows her own priority and the distribution of priorities before applying. For example, a college choice system where priorities are determined by a test score and historical cutoffs are observable to applicants. In the case of Boston, although parents can observe the category where they lie in the priority ladder, meaning, they know their sibling and walk-zone status, they do not observe the random number that determines their actual priority ranking, nor do they observe historical cutoffs to predict the fraction of admitted students with a sibling or walk-zone priority at each school. Moreover, even if parents were able to predict these probabilities with some level of accuracy, analysis of the admissions data reveals that there are only two programs (school-program combinations) that did not admit any students without a priority during the period 2010 to 2012, meaning that the probability of being accepted without a sibling and walk-zone status was not zero for the overwhelming majority of programs.

Under the presence of search costs, if parents are highly optimistic about their admission chances they may stop adding schools that are desirable to the bottom of their ranking. Arteaga et al. 2021 find evidence of this behavior under strategy-proof mechanisms in Chile and New Haven. This may be a concern specially if parents are skipping schools that are farther from their residence. In that case, if few families in a cluster live near a set of schools, we would estimate these $\delta_{cj}$ with bias and perhaps noise.[25] The data shows that on average parents in Boston do consider schools that are farther, reducing concerns for bias. While on average the farthest school in the menu of families is at about 5.5 miles from their residence, parents' farthest ranked school is on average 3.8 miles away (Figure A.3). Moreover, when I look at the 3.8 mile radius around each school, I find there are typically a couple of hundred students from each cluster in these areas, and there are school and cluster combinations for which these numbers are as low as 6 and 10 students (Table A.3). This ameliorates concerns about noise for the majority of schools. Confidence intervals for both the estimated parameters and the counterfactual assignments are estimated.

**Consideration Set.** I assume that students consider all schools in their choice set. This means, families can process information about all the schools they are eligible for and can rank all those options. The assumption is motivated by the relative small size of choice sets in this setting, where families have an average of 25 schools to choose from. This is in contrast with assumptions made in Son 2020, where families are asked to choose from around 430 high school programs in New York City. The author estimates that in this context families are aware of about 65 programs.

Consistent with the assumptions on consideration and truth-telling, if $R_i = (R_{i1}, \cdots, R_{il_i})$ is

[25] The mean utilities of white families for schools near Black and Hispanic families would be biased downward, the same for the mean utilities of Black and Hispanic families for schools near the residence of white families.

the rank-ordered list submitted by $i$ and $\mathcal{J}_i$ is the choice-menu of $i$ then

$$R_{i1} = \underset{j \in \mathcal{J}_i}{\arg\max} \quad u_{ij} \tag{2}$$

$$R_{ik} = \underset{j \in \mathcal{J}_i \setminus \{R_{im} : m < k\}}{\arg\max} \quad u_{ij} \tag{3}$$

Moreover, if $u_{i0}$ is the utility of the outside option then,

$$u_{ij} > u_{i0} \quad \forall \quad j \in R_i \tag{4}$$

$$u_{i0} > u_{ij} \quad \forall \quad j \in \mathcal{J}_i \setminus R_i \tag{5}$$

The utility $u_{i0}$ represents the expected utility at the time of the application of the best accessible alternative if unassigned in the first round. In practice, this includes options outside of the school district and undersubscribed schools within the district. Recall that out of all students unassigned in the first round, about 75% do not end up enrolling into any program within the school district. This means that of all the students for whom we get to observe their outside option, a majority have outside options that lie outside the school district. The remaining 25% do enroll in a school within the district after applying in a second round. Their choices in a second round may be additionally based on information acquired between rounds one and two about the availability of outside options. Concretely, parochial and other private options typically announce admission decisions simultaneously to BPS. If parents overestimated the probability of admission into their outside options, they will need to reconsider their choices within BPS. Second round applicants may also include parents who overestimated their probabilities of assignment into schools within the district in the first round. I do not model these dynamic considerations; instead, I interpret the parameters of the model as a summary of parents' preferences and expectations in the first round of applications.[26]

**Identification.** The parameters of the model $\{\beta_c, \gamma_c, (\delta_{cj})_j\}_c$ are identified modulo the scale parameters $\lambda_c$. This means that, unless we are willing to assume $\lambda_c$ is common across races,

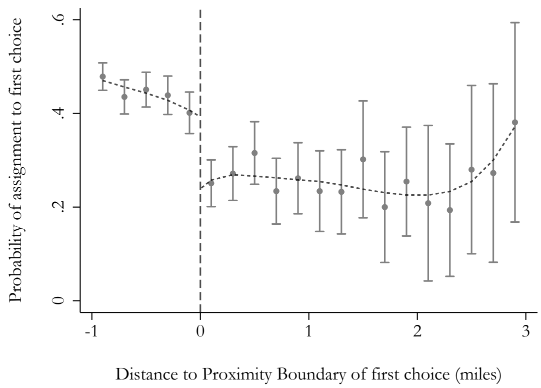
[26] Kapor et al. (2020) estimate interim beliefs in a similar setting

the cell-specific parameters of the model can't be compared. To guarantee identification of the school-specific mean utilities, I normalize the utility of the outside option to zero. This means, the $\delta_{cj}$ are estimated as deviations from that of the outside good.

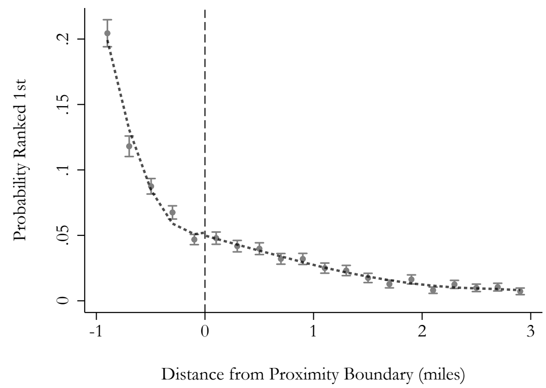
Two distinct sources of variation identify school mean utilities and preferences for proximity. Rankings of students who are equidistant from any pair of schools generate the variation used to identify school mean utilities, while parents who rank schools farther away on top of schools closer to their residence help identify parents' preferences for proximity. Identification of the distance parameter relies on the assumption that $\epsilon_{ij}$ is conditionally independent of $d_{ij}$. I assume families in cell $c$ may sort into neighborhoods following desirable observable and unobservable school characteristics captured by $\delta_{cj}$ and $\gamma_c$. The identification assumption is violated if families systematically choose their residence according to other unobserved variables correlated with their valuation of schools. In that case, the distance parameter will be biased away from zero driving the conclusion that students care about distance more than they really do.

The geographic discontinuities generated by the assignment algorithm provide good variation to study how predominant sorting related to school valuation is in this context. School walk-zone boundaries generate a sharp discontinuity in the probability of assignment since students that are in the walk-zone have a higher priority than students who live outside. Figure 5a shows the discontinuity in the probability of getting assigned to the first ranked school at the proximity boundary. This means that families who choose their residence near a school they find desirable benefit from choosing their residence 0.9 miles from the school relative to 1.1 miles from it. If families are sorting on these boundaries, we may see parents who live less than a mile from the desired school rank it in the first position more often than parents who live slightly more than a mile from it. In Figure 5, I plot the probability of ranking a school in the first position as the distance to the proximity boundary of that school changes. The zero in the x-axis represents the one-mile proximity threshold. That is, at zero a student lives at exactly one mile from the school in question. To the left of zero, students live within the walk-zone. The downward trend shows that parents value proximity

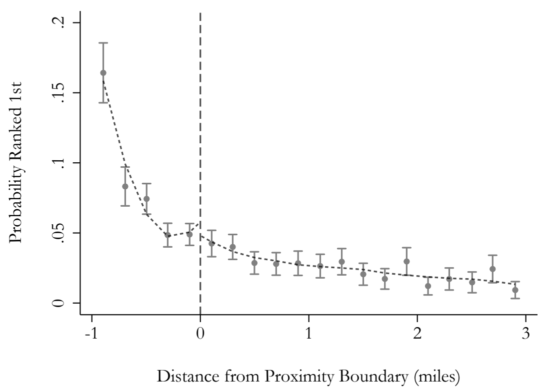
Figure 5: Proximity Priority and Ranking Behaviour



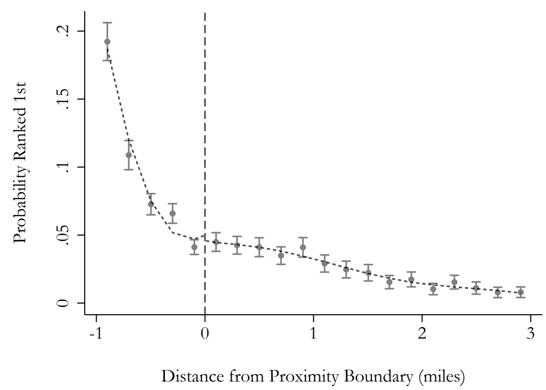

(a) Assignment to First Choice




(b) All Students - First Rank




(c) Black students - First Rank

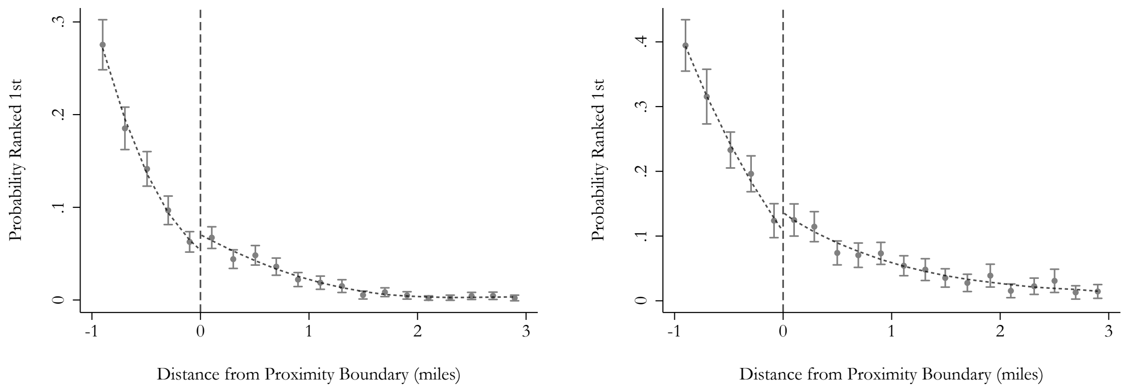



(d) Hispanic students - First Rank

Note: The graphs show bin-scatter plots with equally sized distance bins in each side of the boundary. For every student-school pair, I construct the linear distance between that student and the proximity boundary of the school. Negative values indicate that a student lives within the proximity priority zone. In panel (a) I restrict to every student's first ranked school and in the vertical axis I plot the probability of getting assigned to that first choice. In panels (b) - (f), the vertical axis is the probability that a family ranks a school as their first choice. Range plots show 95% confidence intervals, while the dashed line represents a local linear fit estimated on each side of the boundary. Competitive schools in panel (f) are the five schools that are ranked in the first position more often. Results are similar if I instead consider the schools who accepted the least number of students without any priority.

while a discontinuity at zero is evidence of sorting. The plots show no evidence of sorting on these boundaries for students in any group, and for a sample restricted to competitive schools.

Figure 5: Proximity Priority and Ranking Behaviour (Continued)




(e) White students - First Rank

(f) Competitive Schools - First Rank

**Estimation and Inference.** I estimate preferences for the sub-sample of Black, Hispanic and white students. I do not estimate preference parameters for Asian students and other racial minorities due to a small sample size.[27] In consequence there are a total of 6 covariate cells, with students spread across the city.

I estimate utility parameters for each cell by maximum likelihood. Details about the likelihood function are shown in Appendix C. Bootstrapped standard errors are obtained by sampling the data by student with replacement, keeping the application profile submitted by each student and re-estimating the model in each of 100 samples.

## 4.2 Parameter Estimates

Tables A.5 and A.6 show the estimated parameters for all clusters. Negative signs for the distance parameters ($\beta_c$) show that parents value proximity, as many papers have found (see a summary in Agarwal and Somaini (2019)), also, parents of students whose first language is not English value having a programs in the students' first language. School mean utilities $\delta_{cj}$, summarize the cluster-specific average attractiveness of a school after discounting the effect

[27] For these groups I use the submitted rankings instead of simulated rankings in the counterfactuals.

of distance. School mean utilities have a positive correlation with school achievement and the share of white students, and a negative correlation with the fraction of Black students and low income students (Table A.7). Nevertheless, the effect of the racial composition of a school on school mean utilities is stronger than that of achievement for most clusters, suggesting demographics is a big component of parental preferences for schools and a plausible source of preference heterogeneity (Table A.8).

To assess how preferences for proximity compare across groups, I simulate rankings after varying the distance between schools and applicants. Figure A.4 shows the average number of positions in the ranking a school would lose, and the share of applicants lost if the distance between a school and an applicant was increased by 0.1, 0.5, 1 and 2 miles. I find that in both the intensive and extensive margins, the rankings of white students are more sensitive to increases in distance. These results are influenced not only by parents' preference for proximity, but also by the availability of substitutes near families' residences, and the length of the lists submitted. Shorter lists submitted by white students explain in part the larger responses in both the extensive and intensive margins.

**Fit.** To evaluate the fit of the model I compare the characteristics of an assignment carried out using the rankings submitted by parents to BPS with assignments based on rankings simulated using the demand model.

To simulate rankings using the parameters of the model I assume that families rank every school that is preferred to the outside option in preference order. Then, the position of the outside option determines the length of the simulated rankings. I will make this assumption through-out the counterfactuals as well. The parameters of the model closely predict the distribution of school achievement, share of low-income students and, share of Black students at the schools assigned to students in each group, as well as the distribution of distance (Figure A.5). In some sense this is not surprising since the estimation ensured we would approximate the distribution of distance, and this alone heavily influences the characteristics of the school a student ends up being matched to. The model also approximates fairly well the fraction of students that are assigned to their first, second, and $n$-th choice, as well as the fraction of unassigned students in each group. These statistics, and specially the fraction of unassigned students, depend importantly on the length of the submitted rankings since there are remaining seats after the first round is done (both in our simulated version of market and in the market we observe). This suggests that the estimated values of the parameters—relative to the value of the outside option—captures well the trade-offs involved in choosing the number of schools a parent ranks.

# 5 Counterfactual Assignments

In this section, I describe how and under what assumptions the counterfactual assignments are generated, and then discuss the results. I simulate counterfactual assignments taking all the students that applied for a seat in 2011, and all the schools open for admission in that year. The counterfactuals will be used to estimate the contribution of the mechanisms described in Section 3.

## 5.1 Changing the location of a student

To estimate how much of the cross-race gap in school achievement can be attributed to the location of students, I evaluate how the submitted rankings and subsequent assignments of minority students would change if their residential location were randomly drawn from the set of white students' locations. After drawing a new residential location for a single minority student, I use the demand model to generate the ranking that the student would have submitted at that new location. Demand parameters do not change; nevertheless, the change in distance to all schools will shift travel costs. Also, choice-menu restrictions may limit and/or expand parents' available choices. I further assume that the length of the simulated rankings is determined by the position of the outside option—in other words, parents rank every school preferred to the outside option.

I consider the relocation of one student at a time. Changing the residential location of a single student guarantees that schools are unchanged across counterfactuals, and in consequence preference parameters are the same. If the locations of all students changed simultaneously we would expect, for instance, the demographic composition of schools to change. This means that I estimate the average impact of relocating a single minority student as opposed to the effect of relocating every minority student at the same time.

To build counterfactual locations, I randomly pair minority students and white students. In each counterfactual, the minority student will take the white student's residential location, choice-menu and, walk-zone priority. I consider three distinct assumptions to handle sibling priorities after a relocation. First, I assume students with a sibling lose any sibling priority they previously held. This is the case for a family that relocates and searches for a new school for both siblings. Second, I assume that families keep the sibling priority they had, meaning that the older sibling holds her seat and the youngest searches for one in the new location. Third, I use the parameters of the model to predict the school where the older sibling would have been assigned had they lived in the new location when the older sibling was applying for schools. Finally, I assume every student with a sibling has a sibling priority at their first-ranked school in the new location. This represents the extreme case where the older sibling is always assigned to the first ranked school, and hence the younger sibling is in a high priority group at that school. For each assumption, I generate assignments for all minority students at both their original location and their counterfactual locations and the corresponding distributions of school achievement. I simulate 20 counterfactual assignments for each minority student, one for each draw of $\epsilon_{ij}$.

Having no sibling priority means minority students in the counterfactual are at a disadvantage relative to white students, for whom at least a fraction are in a high priority group in the new location. Despite being at a disadvantage, such a relocation translated into increased access to high achieving schools that reduced the gap by close to 50% for both Black and Hispanic students (Table B.9). Assuming a student loses any sibling priority gives the same results to assuming they keep their original sibling priority. This is because in most cases the school where they held the priority is sufficiently far from their new residence.

A counterfactual where parents hold a sibling priority at a school near their residence contributes to a larger reduction in the gap. To predict the place where an older sibling would have been placed in the new residential location, I start by noting that running the assignment algorithm from the hypothetical situation where no student has a sibling priority predicts well the distribution of school achievement of the sibling's school, for those students with a sibling. This is because the residential location of parents strongly predicts the school assignments. Figure B.8 shows the distribution of school achievement of the sibling's school for the students in my sample, and overlayed is the distribution of school achievement obtained for these same families after running the assignment algorithm assuming no one has a sibling. These graphs show that if we re-started in a world where no one has a sibling priority, we would rapidly converge to the assignments we see today, taking as input only families' locations. I use this observation and generate a counterfactual sibling priority at the school a student would have been assigned after a relocation if no one had a sibling priority.[28]

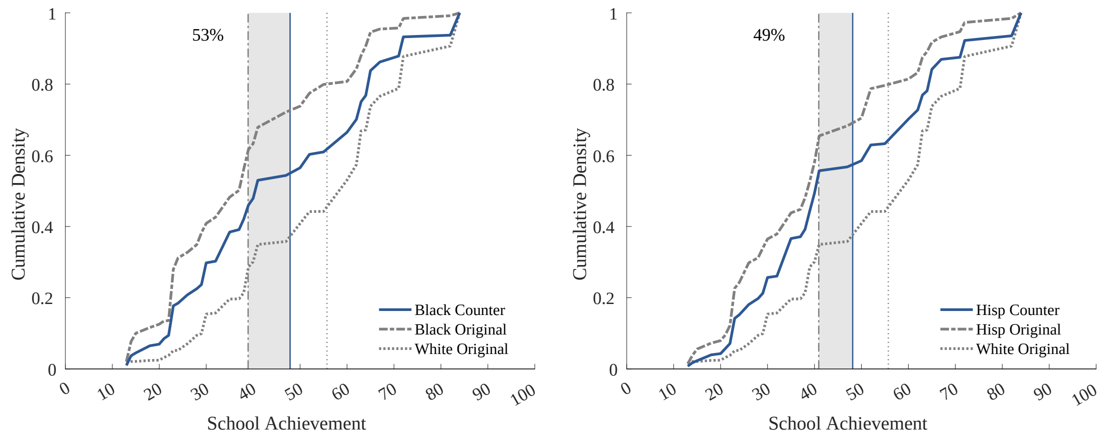
Figure 6 shows the results from a counterfactual assignment after predicting the sibling school for minority students at the new locations. I plot the distribution of school achievement for the schools assigned to white, Black and Hispanic students in their original residential locations, and for Black and Hispanic students in their counterfactual locations. In this case, the gap for Black and Hispanic students reduced by an additional 4 pp and 3 pp, relative to the counterfactual where families loose any sibling priority. If I assumed that students with a sibling have the priority at their first ranked school after a location change, I find the gap shrinks by and additional 4 pp for Black students to 57%, and and additional 5 pp for

[28]After running the assignments under the no-siblings assumptions I get that around 15% of students with a sibling end up unassigned. To predict the school where these students would have had a sibling, I choose the school in their ranking where they were closer to be assigned—that is, the school where the applicant was closest to the last assigned student in the priority ranking. The distance is measured by counting the number of students between both.

Hispanic students to 54% (Table B.9).

Figure 6: Change location of a student: Achievement at the Assigned School

(a) Black students (b) Hispanic students



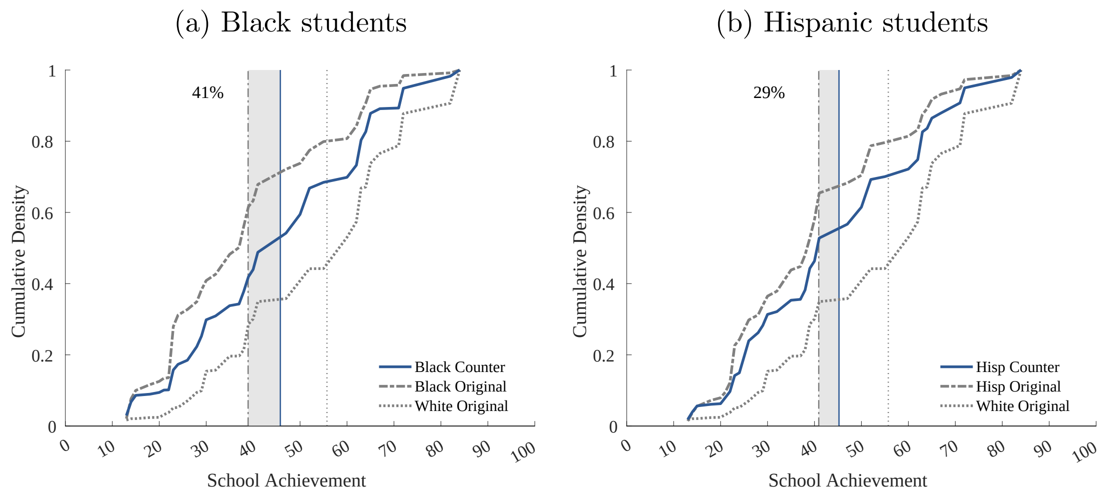

Note: Distribution of achievement in schools assigned to Black and Hispanic students under a counterfactual assignment where they are randomly assigned to a new residence drawn from the distribution of whites' residences, and the sibling's school in the new location is predicted using the parameters of the model for the students with a sibling. This is compared to the distribution for Black, Hispanic and white students in their original location.

In summary, changes to location contribute to increased access into high-achieving schools for minority students, even when relocation comes at the cost of losing any sibling priority. In this case, location is estimated to contribute to about half of the gap in access to high-achieving schools. Other assumptions about sibling status have a negligible added effect. In the absence of free busing, the original gap in access to quality education is expected to be larger (Trajkovski et al. 2021), as well as the effect of a location change in that case. Notice that the potential bias due to having overly confident families goes in the same direction of the effect of distance. In that case, the effect of distance is expected to be even larger.

## 5.2 Changing demand parameters

To study the contribution of demand heterogeneity to the gap in school achievement, I evaluate how the submitted rankings and subsequent assignments of minority students change if the demand parameters of minority families where those of white parents. In this counterfactual the residential location, walk-zone, and sibling priorities of every student are unchanged. For consistency with the previous counterfactual, I change the demand parameters of one student at a time. The counterfactual ranking obtained describes how parents of a minority student would rank schools in their original residential location if their demand parameters were those of white parents.

Figure 7: Change in demand parameters of a student: Achievement at the Assigned School




Note: Distribution of achievement in schools assigned to Black and Hispanic students under a counterfactual assignment where these students have the demand parameters of white students. This is compared to the original distribution of school achievement for Black, Hispanic and white students.

Under the proposed change, Black and Hispanic students are assigned to schools with higher average achievement. Figure 7 shows the distributions of school achievement for white and minority students under the original setting and for minority students in the counterfactual. The gap reduced by 41% and 29% for Black and Hispanic students, respectively (Table B.9). Notice that the potential bias due to having overly confident families goes in the opposite direction of the effect of a change in the demand parameters. In that case, the expected impact of preferences would be smaller.

### 5.3 Eliminate Choice Menu Restrictions and Walk-zone Priorities

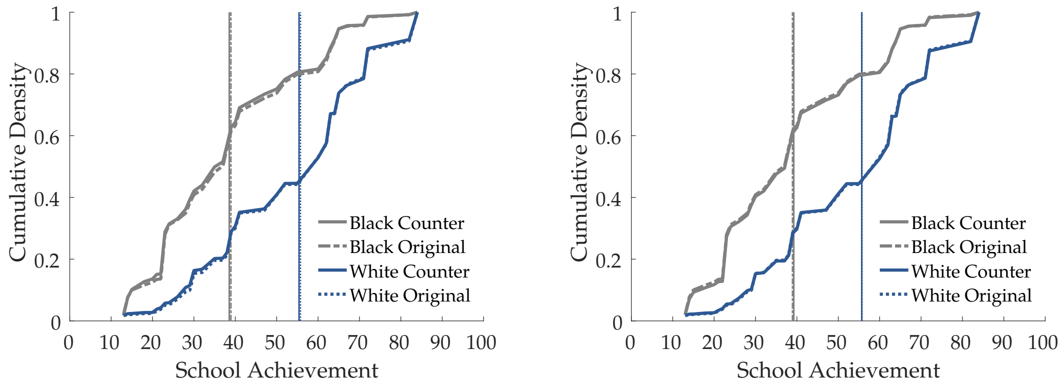
The effect of location on school assignments is sizeable. When a student changes locations not only do her travel costs change, but also, her choice menu and the set of schools where she has a walk-zone priority change. To disentangle the effect of the last two from that of travel costs, I run two additional counterfactual assignments. In the first, I eliminate choice-menu restrictions and allow parents to rank schools from across the city. Under this setting, parents of a minority student can rank the same schools they would have ranked under the location change counterfactual. Then, the only reason why these rankings wouldn't coincide is differences in travel costs to these schools from both locations. In the second counterfactual, I eliminate walk-zone priorities and run the assignment algorithm assuming no one has this priority. Eliminating priorities doesn't change parental rankings, but it does change assignments via priorities. This counterfactual captures the effect of a change in the probability of assignment into a school, that is explained solely by the walk-zone priority. Notice that since these counterfactuals are about algorithm rules, I do not isolate the effect of the counterfactual into a single student; instead, I change assignment rules for all students simultaneously, including white students.

Changing the rules of the assignment has a very small impact on the gap in access to high achieving schools. Figure 8 shows the distribution and average school achievement for Black students after eliminating choice-menu restrictions on the left, and the walk-zone priority on the right. Results for Hispanic students are sown in Figure B.9. After walk-zone priorities are eliminated the gap for Black families reduces by 2% while for Hispanic families the effect on the gap is statistically zero. Eliminating choice menu restrictions increases the gap for both Black and Hispanic families, although the effect is small (2% and 4% in each case) (Table B.9).

These results imply that these assignment rules have a limited impact on the cross-race gap in school achievement. In consequence, the results suggest that the effect of location is explained by changes in the distance to high-achieving schools and not by assignment rules

Figure 8: Eliminate location-specific rules - Black Students

(a) Choice-Menu (b) Walk-Zones




Note: Distribution of achievement in schools assigned to Black and white students under a counterfactual assignment where choice menu restrictions are eliminated (on the left), and walk-zone priorities are abolished (on the right).

that are location-specific. Moreover, these results suggest that changing these rules is not an effective policy to increasing access into high achieving schools for all students.

## 5.4 Change in the School Match After a Location Change

So far we have established that if a minority family faced the menu of distances that a typical white family does, they would access high-achieving schools at a higher rate. But, these higher-achieving schools may not be preferred by minority families to the schools assigned to them in their original locations for a variety of reasons. Using the parameters of the model I can assess whether Black and Hispanic students are assigned to schools with higher mean value after a location change. To do this, I compare the location-independent value of the school assigned under the original setting and the counterfactual. Let $\mu(i) \in \mathcal{J}$ be the school assigned to $i$ under the original setting and $\tilde{\mu}(i) \in \mathcal{J}$ be the school assigned to $i$ under the counterfactual. If $G_r$ is the set of students in racial group $r = \{B, H\}$, then the average change in school mean value for the students in $G_r$ expressed in miles is

$$\sum_{i\in G_r} \frac{\delta_{c\tilde{\mu}(i)} - \delta_{c\mu(i)}}{|\beta_c| \cdot |G_r|}$$

After a location change, Black and Hispanic students are matched to schools that have on average higher mean-value. The average change in school value for Black and Hispanic students is equivalent to reducing students' travel distance by 0.3 miles. These results imply that the costs generated by distance not only restrict access to quality but result in matches that are on average of lower mean value to these families.

## 5.5 Alternative Specifications of the Random Utility Model

I run counterfactuals using parameters from a set of alternative specifications for the random utility model. I consider two alternative modelling choices. First, a model of utility where preferences for distance are not linear but quadratic. This model would capture the possibility that the first miles travelled are marginally more costly. This would happen if families would only consider taking the train or the school bus if their travel exceeds some distance, and the marginal cost of each mile travelled is different across transportation modes.[29] Second, I consider the case where the value of the outside option is location specific. This would happen if some locations in the city have more affordable or higher quality private child care options.

To model the location specific outside options, I consider a partition of the city into 12 neighborhoods. This partition builds on neighborhood boundaries defined by the city, and in consequence is a partition of the city into areas that have a similar amenity supply. I assume that the residents of neighborhood $n$ that belong to the same cluster share a common value for the outside option. To guarantee the model is identified, I normalize to zero the value of the outside option in a neighborhood that intersects the three zones of choice that partition the city. I estimate the value of the outside option in the other neighborhoods

[29] Only families who are assigned to schools farther than a mile are offered free busing, so it is likely these non-linearities in preferences are present

relative to the value of the outside option in the reference neighborhood. Then, if $n(i)$ is $i$'s neighborhood, $u_{i0_{n(i)}} = 0$ if $n(i)$ is the reference neighborhood and $u_{i0_{n(i)}} = \kappa_{n(i)}$ otherwise, where $\kappa_{n(i)}$ is now a parameter of the model. Identification of outside option values follows from the fact that these are connected strict substitutes of the reference outside option (Berry et al. 2013).

I generate counterfactuals for a total of 4 specifications of the random utility model. These are the combination of the assumptions above. Namely having or not linearity with respect to distance, and having location-specific outside options, or not. The results presented in the section before correspond to the model where outside options are not location specific and the utility is linear in distance. Results from all specifications are highly consistent (Table B.10).

# 6 Conclusion

Among other objectives, choice-based systems are intended to increase equity and foster diversity by allowing students to sort into their preferred schools, while weakening the link between residential locations and school assignments. I document that in Boston, Black and Hispanic students are assigned to schools that are lower-achieving than the schools assigned to white students. I show that both cross-race differences in distance to high-achieving schools, and heterogeneity in demand for non-location school attributes contribute to this gap, while assignment rules that are location specific don't.

The salience of travel costs shows a first-order channel for why neighborhoods matter, highlighting how the effective provision of public goods can be affected by geography at very granular levels. In some way these results are not surprising. Most, if not all, of the papers that study school demand agree that distance is a key factor in parental choices. This paper takes this observation one step further and quantifies how much this cost limits the effectiveness of school choice policies in equalizing access to high-achieving schools. The results show that even in a generous choice environment where parents face minimal restrictions to their choices and free transportation is provided, distance can contribute greatly to inequity and that the design of the assignment algorithm can do little to break structural place-based inequities. This finding is not only relevant for the pre-kindergarten population. Not only we know that early investments can have lasting impacts on adult outcomes, but also, choice systems are typically designed to grandfather students into subsequent grades within a school. Then, even if travel costs are lower for older children, early assignments are held for several years after. In consequence, inequities in pre-kindergarten extend well after that period.

My model and counterfactuals cannot be used to predict how would access to high-achieving schools change if school locations changed, or alternatively if the achievement at a school increased. A thought experiment would be to consider an intervention that raises the test scores of a school in one year. How would this experiment change the access of incoming prekindergarteners to high-achieving peers in equilibrium? The main difficulty with learning about the outcomes of this experiment has to do with predicting families' housing choices and their demand for the hypothetical school.[30] Both these objects will determine the equilibrium distribution of distance to high-achieving schools by race. Most likely the new distribution will reflect some inequities in access to high-achieving schools, as school quality is capitalized in housing prices and low-income families are able to move close to those schools. In consequence, there is no reason to believe that an exogenous change in the achievement of one school will lead to a qualitatively different distribution of distance to high-achieving schools by race than the one observed in the data. Instead of taking this route, I concentrate on isolating the effect of distance under the conditions in the observed market.

Finally, although this paper suggests distance can contribute to inequities in access to effective schools, quantifying this remains an important exercise that this paper does not carry-out. If the effect of distance is sizable, this would help reconcile evidence showing that

[30] Perhaps such an intervention has effects on teacher sorting as well, but let's abstract from that.

families value effective schools -albeit mediated by demand for peer-quality- (Abdulkadiroğlu et al. 2020), but Black and Hispanic students in Boston do not experience test-score gains as a result of being assigned to schools out of their walkzone (Angrist et al. 2021).

# Acknowledgements

I gratefully acknowledge support by the Successful Pathway from School to Work initiative of the University of Chicago, funded by the Hymen Milgrom Supporting Organization.

# References

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005). The new york city high school match. *American Economic Review*, 95(2):364–367.

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review*, 110(5):1502–39.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American economic review*, 93(3):729–747.

Agarwal, N. and Somaini, P. J. (2019). Revealed preference analysis of school choice models. Technical report, National Bureau of Economic Research.

Allende, C. (2019). Competition under social interactions and the design of education policies. *Job Market Paper*.

Allers, K. S. (2019). The tough choices black parents face when choosing a school for their children. *The Washington Post*.

Altonji, J. G., Huang, C.-I., and Taber, C. R. (2015). Estimating the cream skimming effect of school choice. *Journal of Political Economy*, 123(2):266–324.

Angrist, J., Gray-Lobe, G., Idoux, C., and Pathak, P. A. (2021). Is busing worth the trip? school travel effects in boston and new york. *Unpublished manuscript*.

Arteaga, F., Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2021). Smart matching platforms and heterogeneous beliefs in centralized school choice. Technical report, National Bureau of Economic Research.

Balinski, M. and Sönmez, T. (1999). A tale of two mechanisms: student placement. *Journal of Economic theory*, 84(1):73–94.

Barseghyan, L., Clark, D., and Coate, S. (2014). Public school choice: an economic analysis. Technical report, National Bureau of Economic Research.

Berry, S., Gandhi, A., and Haile, P. (2013). Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111.

Bertsimas, D., Delarue, A., Eger, W., Hanlon, J., and Martin, S. (2020). Bus routing optimization helps boston public schools design better policies. *INFORMS Journal on Applied Analytics*, 50(1):37–49.

Biasi, B. (2019). School finance equalization increases intergenerational mobility: Evidence from a simulated-instruments approach. Technical report, National Bureau of Economic Research.

Borghans, L., Golsteyn, B. H., and Zolitz, U. (2015). Parental preferences for primary school characteristics. *The BE Journal of Economic Analysis & Policy*, 15(1):85–117.

Boston Desegregation Project (1988). (130). *Northeastern University Library, Archives and Special Collections*, 9(33).

Calsamiglia, C. and Güell, M. (2018). Priorities in school choice: The case of the boston mechanism in barcelona. *Journal of Public Economics*, 163:20–36.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained school choice: An experimental study. *American Economic Review*, 100(4):1860–74.

Campos, C. and Kearns, C. (2021). The impacts of neighborhood school choice: Evidence from los angeles' zones of choice. *Available at SSRN 3830628*.

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., and Porter, S. R. (2018). The opportunity atlas: Mapping the childhood roots of social mobility. Technical report, National Bureau of Economic Research.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Chubb, J. E. and Moe, T. M. (1990). America's public schools: Choice is a panacea. *The Brookings Review*, 8(3):4–12.

Clampet-Lundquist, S. and Massey, D. S. (2008). Neighborhood effects on economic self-sufficiency: A reconsideration of the moving to opportunity experiment. *American Journal of Sociology*, 114(1):107–143.

Dubins, L. E. and Freedman, D. A. (1981). Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494.

Dur, U., Kominers, S. D., Pathak, P. A., and Sonmez, T. (2018). Reserve design: Unintended consequences and the demise of boston's walk zones. *Journal of Political Economy*, 126(6):2457–2479.

Epple, D., Figlio, D., and Romano, R. (2004). Competition between private and public schools: testing stratification and pricing predictions. *Journal of public Economics*, 88(7-8):1215–1245.

Epple, D. and Romano, R. E. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, pages 33–62.

Friedman, M. (1982). Capitalism and freedom. 1962. *Whether government control over the provision of education through public schools is also efficient is a quite different matter.*

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic theory*, 144(5):1921–1947.

Hastings, J., Hortaçsu, A., and Syverson, C. (2017). Sales force and competition in financial product markets: the case of mexico's social security privatization. *Econometrica*, 85(6):1723–1761.

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145:1–46.

Hoxby, C. M. (2003). School choice and school productivity. could school choice be a tide that lifts all boats? In *The economics of school choice*, pages 287–342. University of Chicago Press.

Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from chile's voucher program. *Journal of public Economics*, 90(8-9):1477–1503.

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315.

Katz, L. F., Kling, J. R., and Liebman, J. B. (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics*, 116(2):607–654.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Laliberte, J.-W. P. (2018). Long-term contextual effects in education: Schools and neighborhoods. *V manuscript*.

Langer, A. (2016). (dis)incentives for demographic price discrimination in the new vehicle market. *Working Paper*.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., and Sanbonmatsu, L. (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review*, 103(3):226–31.

Luflade, M. (2018). The value of information in centralized school choice systems. *Unpublished manuscript*.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. *Documento de trabajo). Yale University. Recuperado de http://economics. sas. upenn. edu/system/files/event_papers/Neilson_2013_JMP_current. pdf*.

Pathak, P. A. and Shi, P. (2013a). Simulating alternative school choice options in boston-main report.

Pathak, P. A. and Shi, P. (2013b). Simulating alternative school choice options in boston-technical appendix.

Pathak, P. A. and Shi, P. (2017). How well do structural demand models work? counterfactual predictions in school choice. Technical report, National Bureau of Economic Research.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628.

Son, S. J. (2020). Distributional impacts of centralized school choice.

Sönmez, T., Yenmez, M. B., et al. (2019). Affirmative action with overlapping reserves. Technical report, Boston College Department of Economics.

Trajkovski, S., Zabel, J., and Schwartz, A. E. (2021). Do school buses make school choice work? *Regional Science and Urban Economics*, 86:103607.

Walters, C. R. (2018). The demand for effective charter schools. *Journal of Political Economy*, 126(6):2179–2223.

# Appendix

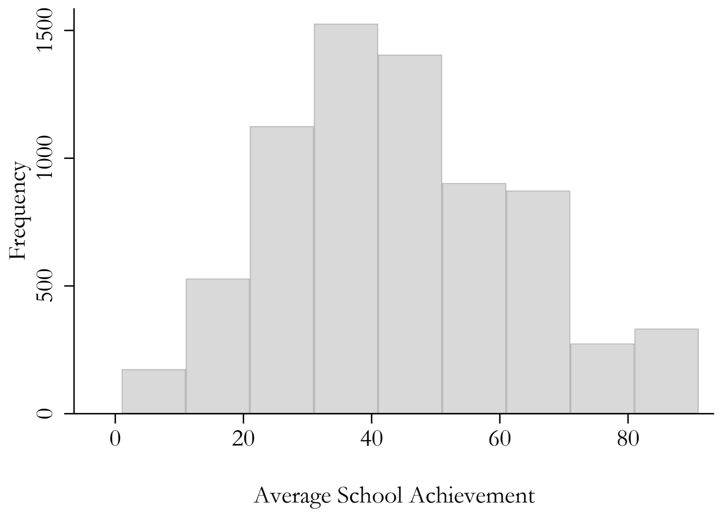
# A Supplementary Tables and Figures

Table A.1: Hypothetical Neighborhood Assignments and DA Assignments

| | All | Black | Hispanic | White |
|---|---|---|---|---|
| % Students assigned to the same school | 21.1 | 17.2 | 19.1 | 28.9 |
| *Students assigned to a different school* | | | | |
| Achievement - DA | 44.5 | 39.4 | 42.6 | 53.8 |
| Achievement - Neighborhood | 43.9 | 40.0 | 43.1 | 50.6 |
| Distance - DA | 1.5 | 1.6 | 1.5 | 1.3 |
| Distance - Neighborhood | 0.4 | 0.3 | 0.3 | 0.4 |
| % Low income in Kindergarten - DA | 68.0 | 71.5 | 70.8 | 56.5 |
| % Low Income in Kindergarten - Neighborhood | 68.6 | 71.7 | 70.7 | 58.7 |
| % Same race - DA | 45.2 | 43.0 | 54.2 | 26.4 |
| % Same race - Neighborhood | 41.3 | 42.7 | 48.1 | 22.7 |

Note: The first line in the table shows the fraction of students in each group who are assigned to the same school under BPS's assignment and the hypothetical neighborhood assignment described. Below, I show average school characteristics of the assigned school restricting the sample to students who are assigned to a different school under both rules.

Figure A.1: Histogram of School Achievement




Note: Histogram of school achievement weighted by school capacity for the years 2010 to 2012. School Achievement is measured as the fraction of third-grade students scoring advanced of proficient in the MCAS math tests.

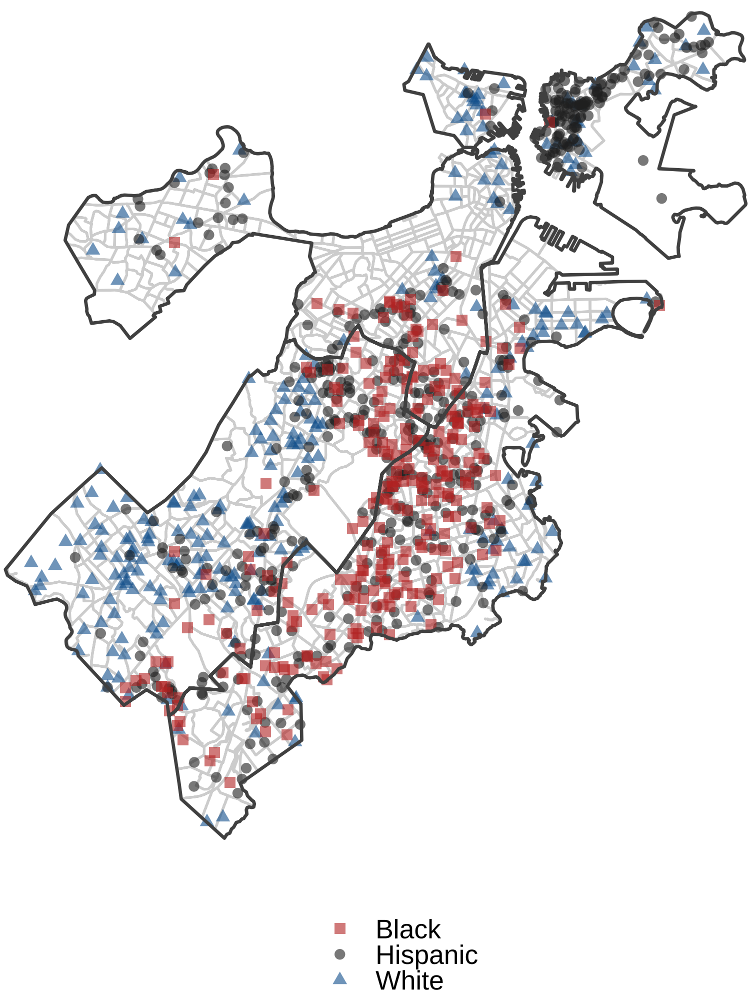
Table A.2: Relation Between Distance to Schools and School Achievement

| | School Achievement | | |
|---|---|---|---|
| | Black | Hispanic | White |
| Distance | 1.08 | 0.06 | −1.66 |
| | (0.05) | (0.04) | (0.05) |
| Observations | 46,647 | 82,853 | 41,894 |

Standard errors in parentheses

Note: Each column shows a regression between school achievement and distance. Each observation is a student-school pair, for schools in the choice-menu of every student. Standard errors in parenthesis.

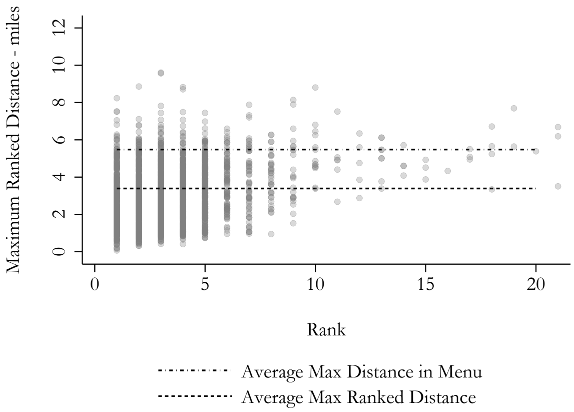
Figure A.2: Spatial Distribution of Applicants by Race



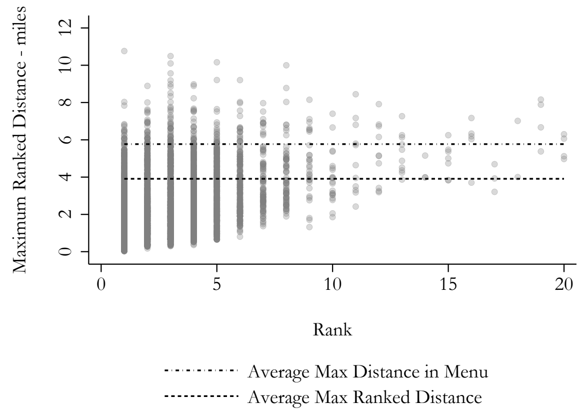

Note: Each point represents 10 students from the 2010-2012 pooled data, located randomly at the census tract level.

Figure A.3: Farthest Ranked School and Farthest School in Choice Menu

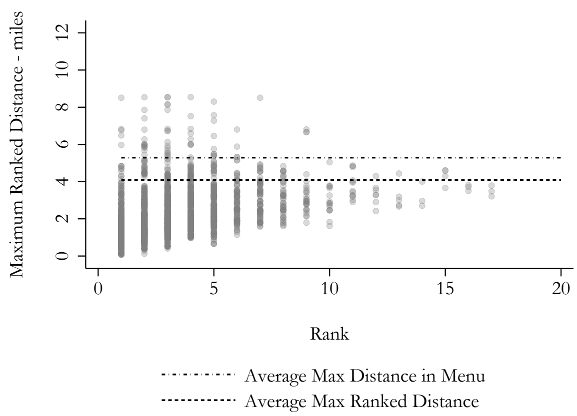
(a) Black Students




(b) Hispanic Students




(c) White Students




Note: Scatter plot of the maximum distance in each ranking and the position of that school in the ranking. The lines show the average of the maximum distance in the rankings, and the average of the maximum distance in the choice-menu for each group.

Table A.3: Number of applicants per school and potential applicants near each school

| | Mean | St. Dev | Min | Max |
|---|---|---|---|---|
| *Potential applicants that are within 3.8 miles of each school* | | | | |
| Cluster 1 - Black Q1 | 442.4 | 294.9 | 20 | 1,274 |
| Cluster 2 - Black Q2 | 175.2 | 108.9 | 6 | 496 |
| Cluster 3 - Hispanic Q1 | 618.9 | 297 | 116.0 | 1,515 |
| Cluster 4 - Hispanic Q2 | 390.6 | 207 | 31.0 | 1,052 |
| Cluster 5 - White Q1 | 60.5 | 32.4 | 10 | 144 |
| Cluster 6 - White Q2 | 463.9 | 307.9 | 43 | 1,224 |
| *Applicants per school* | | | | |
| Cluster 1 - Black Q1 | 113.0 | 77.2 | 17 | 373 |
| Cluster 2 - Black Q2 | 50.3 | 38.9 | 7 | 164 |
| Cluster 3 - Hispanic Q1 | 171.0 | 128.7 | 13 | 686 |
| Cluster 4 - Hispanic Q2 | 107.7 | 83.0 | 8 | 418 |
| Cluster 5 - White Q1 | 17.5 | 16.8 | 0 | 71 |
| Cluster 6 - White Q2 | 124.0 | 167.9 | 0 | 686 |

Note: The first block shows statistics by race on the number of applicants per school. In the second block I show statistics by race on the number of students that live within 1.2 miles of each school, measured using linear distance.

Table A.4: Characteristics of schools with missing school achievement

| | Not Missing Achievement | Missing Achievement |
|---|---|---|
| % Black | 32.3 | 28.8 |
| | (19.4) | (18.3) |
| % Hispanic | 43.7 | 48.8 |
| | (19.5) | (16.9) |
| % White | 14.6 | 14.5 |
| | (15.0) | (12.2) |
| % Low Income in K | 67.4 | 68.7 |
| | (19.9) | (19.4) |
| Observations | 235 | 23 |

Note: Statistics of school year observations where I do not observe school achievement.

# Table A.5: Preference Parameters

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| **Distance** | -0.46 | -0.44 | -0.60 | -0.65 | -0.78 | -0.90 |
| | (0.02) | (0.03) | (0.01) | (0.02) | (0.06) | (0.02) |
| **ELL** | 0.42 | 0.40 | 0.66 | 0.73 | 0.99 | 1.32 |
| | (0.13) | (0.20) | (0.04) | (0.06) | (0.23) | (0.29) |
| **sch 1** | 0.81 | 1.16 | 1.09 | 1.68 | 2.66 | 2.71 |
| | (0.13) | (0.14) | (0.16) | (0.10) | (0.38) | (0.09) |
| **sch 2** | 0.65 | 0.96 | 0.89 | 1.89 | 3.01 | 3.24 |
| | (0.15) | (0.17) | (0.15) | (0.09) | (0.37) | (0.09) |
| **sch 3** | -1.14 | -1.84 | -1.40 | -1.03 | -2.42 | -2.19 |
| | (0.10) | (0.37) | (0.11) | (0.20) | (7.14) | (2.79) |
| **sch 4** | -1.07 | -0.65 | 0.55 | 0.96 | -0.20 | 0.16 |
| | (0.09) | (0.15) | (0.06) | (0.07) | (0.34) | (0.10) |
| **sch 5** | 0.24 | 0.97 | -0.08 | -0.05 | -13.90 | -0.66 |
| | (0.13) | (0.26) | (0.13) | (0.25) | (0.00) | (1.38) |
| **sch 6** | -0.98 | -0.82 | -1.54 | -1.31 | -1.89 | -1.70 |
| | (0.09) | (0.21) | (0.11) | (0.19) | (6.20) | (0.40) |
| **sch 7** | 0.75 | 0.36 | 0.58 | 0.56 | 1.87 | 2.53 |
| | (0.26) | (0.37) | (0.08) | (0.18) | (0.24) | (0.29) |
| **sch 8** | -0.21 | -0.22 | -0.86 | -0.72 | -15.94 | -1.39 |
| | (0.09) | (0.11) | (0.14) | (0.13) | (0.00) | (0.40) |
| **sch 9** | -0.55 | -0.05 | -0.15 | 0.08 | 0.60 | 0.55 |
| | (0.12) | (0.15) | (0.13) | (0.09) | (0.34) | (0.09) |
| **sch 10** | -0.02 | 0.13 | 0.09 | 0.09 | -0.23 | 0.48 |
| | (0.15) | (0.17) | (0.12) | (0.09) | (0.49) | (0.08) |
| **sch 11** | -0.08 | 0.12 | 0.33 | 0.62 | 0.83 | 1.34 |
| | (0.14) | (0.18) | (0.13) | (0.08) | (0.36) | (0.07) |
| **sch 12** | -1.11 | -1.25 | -1.48 | -0.90 | -1.62 | -1.38 |
| | (0.10) | (0.17) | (0.12) | (0.15) | (4.98) | (0.32) |
| **sch 13** | -0.15 | -0.09 | -0.21 | 0.50 | -0.17 | -0.36 |
| | (0.12) | (0.15) | (0.10) | (0.10) | (0.53) | (0.22) |
| **sch 14** | -0.83 | -0.13 | -1.18 | -1.08 | -1.66 | -0.62 |
| | (0.10) | (0.19) | (0.11) | (0.17) | (4.98) | (0.22) |
| **sch 15** | -0.40 | -0.42 | -0.60 | -0.44 | -0.24 | -0.22 |
| | (0.11) | (0.20) | (0.09) | (0.17) | (0.44) | (0.19) |
| **sch 16** | -0.48 | -0.57 | -0.62 | -0.24 | -0.80 | -0.99 |
| | (0.08) | (0.14) | (0.08) | (0.11) | (0.46) | (0.26) |
| **sch 17** | -0.46 | -0.35 | -0.67 | 0.25 | 0.27 | 0.42 |
| | (0.09) | (0.19) | (0.09) | (0.11) | (0.29) | (0.13) |
| **sch 18** | -0.69 | -0.73 | -1.39 | -1.10 | -0.89 | -2.56 |
| | (0.10) | (0.13) | (0.13) | (0.18) | (2.56) | (1.62) |
| **sch 19** | 0.33 | 0.37 | 0.89 | 0.70 | 1.78 | 1.04 |
| | (0.16) | (0.27) | (0.11) | (0.17) | (0.20) | (0.20) |
| **sch 20** | -0.24 | -0.61 | 0.36 | 0.59 | 1.58 | 1.67 |
| | (0.21) | (1.41) | (0.14) | (0.20) | (0.29) | (0.20) |
| **sch 21** | 0.22 | -0.25 | 0.38 | 0.24 | 1.12 | 0.52 |
| | (0.17) | (0.38) | (0.12) | (0.19) | (0.29) | (0.28) |
| **sch 22** | 1.12 | 0.90 | 1.34 | 1.41 | 2.13 | 2.67 |
| | (0.14) | (0.16) | (0.11) | (0.12) | (0.58) | (0.26) |
| **sch 23** | 0.14 | 0.21 | 0.17 | 0.23 | -0.94 | -0.14 |
| | (0.11) | (0.15) | (0.12) | (0.11) | (4.13) | (0.23) |
| **sch 23** | -0.27 | 0.36 | -0.22 | 0.51 | 0.75 | 1.77 |
| | (0.10) | (0.14) | (0.09) | (0.08) | (0.41) | (0.08) |
| **Race** | B | B | H | H | w | w |
| **Income** | Q1 | Q2 | Q1 | Q2 | Q1 | Q2 |
| **No. student-school pairs** | 37,033 | 15,724 | 59,181 | 35,016 | 6,641 | 40,959 |

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| **sch 26** | -0.26 | -0.25 | -0.14 | 0.11 | -0.05 | 0.01 |
| | (0.10) | (0.22) | (0.10) | (0.12) | (0.43) | (0.12) |
| **sch 27** | -0.92 | -0.78 | -0.83 | -0.88 | -2.52 | -1.36 |
| | (0.10) | (0.24) | (0.07) | (0.14) | (6.46) | (0.28) |
| **sch 28** | -1.38 | -1.99 | -1.45 | -1.86 | -2.19 | -1.74 |
| | (0.10) | (0.24) | (0.13) | (0.24) | (6.60) | (0.37) |
| **sch 29** | -0.82 | -1.07 | -0.96 | -0.97 | -3.10 | -2.39 |
| | (0.09) | (0.18) | (0.08) | (0.13) | (6.22) | (0.40) |
| **sch 30** | -1.20 | -0.42 | -0.30 | 0.25 | 0.81 | 2.08 |
| | (0.17) | (0.24) | (0.09) | (0.13) | (0.32) | (0.11) |
| **sch 31** | -0.81 | -0.96 | -0.20 | -0.14 | -1.80 | -0.12 |
| | (0.12) | (0.35) | (0.12) | (0.17) | (5.51) | (0.16) |
| **sch 32** | -0.53 | 0.11 | -0.09 | 0.87 | 1.17 | 1.99 |
| | (0.14) | (0.18) | (0.08) | (0.09) | (0.30) | (0.09) |
| **sch 33** | -0.45 | -0.35 | 0.04 | 0.03 | 0.80 | 1.21 |
| | (0.22) | (0.33) | (0.12) | (0.19) | (0.41) | (0.12) |
| **sch 34** | 0.49 | 0.40 | 0.58 | 0.76 | 1.21 | 2.18 |
| | (0.12) | (0.20) | (0.10) | (0.12) | (0.42) | (0.11) |
| **sch 35** | -1.39 | -1.33 | -1.71 | -1.28 | -0.98 | -1.58 |
| | (0.17) | (0.28) | (0.33) | (0.43) | (5.76) | (3.66) |
| **sch 36** | 0.51 | 0.34 | -0.02 | 0.28 | -0.07 | -0.79 |
| | (0.07) | (0.11) | (0.07) | (0.09) | (0.35) | (0.21) |
| **sch 37** | -0.86 | -0.43 | -0.71 | 0.31 | 0.70 | 1.71 |
| | (0.14) | (0.19) | (0.13) | (0.10) | (0.50) | (0.08) |
| **sch 38** | -0.08 | -0.41 | 0.53 | 0.16 | 0.41 | -0.02 |
| | (0.22) | (0.32) | (0.08) | (0.14) | (0.22) | (0.27) |
| **sch 39** | 0.33 | 0.61 | 0.89 | 1.57 | 2.66 | 2.62 |
| | (0.15) | (0.17) | (0.13) | (0.09) | (0.37) | (0.10) |
| **sch 40** | -0.81 | -0.86 | -1.17 | -1.11 | -0.86 | -14.89 |
| | (0.10) | (0.22) | (0.09) | (0.18) | (0.63) | (0.00) |
| **sch 41** | -0.30 | -0.24 | -0.58 | -0.01 | 0.24 | 0.08 |
| | (0.08) | (0.14) | (0.08) | (0.11) | (0.34) | (0.15) |
| **sch 42** | -0.68 | -1.03 | 0.04 | 0.39 | 1.18 | 0.58 |
| | (0.32) | (0.44) | (0.08) | (0.16) | (0.18) | (0.22) |
| **sch 43** | -1.24 | -0.62 | -1.18 | -0.23 | 0.05 | 1.23 |
| | (0.13) | (0.23) | (0.12) | (0.12) | (0.47) | (0.12) |
| **sch 44** | 0.22 | 0.65 | 0.03 | 0.41 | 1.91 | 3.07 |
| | (0.16) | (0.21) | (0.10) | (0.15) | (0.24) | (0.11) |
| **sch 45** | -0.96 | -0.58 | -0.93 | -0.21 | -0.10 | -0.41 |
| | (0.12) | (0.15) | (0.12) | (0.12) | (0.42) | (0.15) |
| **sch 46** | -1.27 | -0.79 | -1.10 | 0.05 | 0.33 | 1.40 |
| | (0.10) | (0.17) | (0.11) | (0.12) | (0.26) | (0.13) |
| **sch 47** | 0.39 | 0.93 | 0.26 | 1.30 | 1.95 | 3.31 |
| | (0.08) | (0.12) | (0.08) | (0.09) | (0.24) | (0.13) |
| **sch 48** | 0.57 | 0.45 | -0.02 | -0.02 | -0.61 | -2.28 |
| | (0.08) | (0.11) | (0.10) | (0.10) | (0.49) | (2.00) |
| **sch 49** | 0.52 | 0.29 | 1.02 | 1.03 | 1.25 | 1.05 |
| | (0.18) | (0.27) | (0.08) | (0.13) | (0.21) | (0.20) |
| **sch 50** | -0.48 | -1.10 | -0.51 | -0.54 | -1.19 | -1.35 |
| | (0.07) | (0.18) | (0.06) | (0.10) | (2.45) | (0.21) |
| **sch 51** | -0.53 | -0.45 | -0.45 | -0.14 | -0.29 | -0.51 |
| | (0.10) | (0.18) | (0.09) | (0.12) | (0.32) | (0.20) |
| **Race** | B | B | H | H | w | w |
| **Income** | Q1 | Q2 | Q1 | Q2 | Q1 | Q2 |
| **No. student-school pairs** | 37,033 | 15,724 | 59,181 | 35,016 | 6,641 | 40,959 |

Note: Coefficients from regressions between the standardized $\delta_j^r$ and school characteristics by race. Standard errors in parenthesis.

Table A.6: Preference Parameters

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
|---|---|---|---|---|---|---|
| **sch 52** | -0.51 | -0.10 | -0.62 | 0.28 | 0.08 | 1.20 |
| | (0.11) | (0.16) | (0.10) | (0.12) | (0.46) | (0.10) |
| **sch 53** | -0.02 | -0.63 | 0.59 | 0.33 | 0.69 | 0.51 |
| | (0.23) | (0.41) | (0.07) | (0.16) | (0.19) | (0.23) |
| **sch 54** | -0.35 | -0.81 | 0.52 | 0.15 | 0.53 | 0.10 |
| | (0.25) | (0.41) | (0.07) | (0.15) | (0.19) | (0.28) |
| **sch 55** | -0.58 | -0.26 | -0.15 | -0.04 | -0.60 | -0.19 |
| | (0.14) | (0.18) | (0.10) | (0.08) | (0.50) | (0.10) |
| **sch 56** | -1.06 | -0.40 | -1.09 | -0.19 | 0.21 | 0.76 |
| | (0.20) | (0.19) | (0.20) | (0.12) | (0.57) | (0.10) |
| **sch 57** | -0.81 | -0.95 | -0.45 | -0.55 | -1.23 | -2.24 |
| | (0.11) | (0.28) | (0.10) | (0.18) | (0.48) | (1.97) |
| **sch 58** | -1.04 | -1.02 | -1.55 | -1.82 | -2.06 | -0.62 |
| | (0.10) | (0.28) | (0.13) | (0.31) | (6.72) | (0.23) |
| **sch 59** | -0.58 | -0.51 | -0.82 | -0.12 | 0.74 | 0.43 |
| | (0.14) | (0.24) | (0.15) | (0.15) | (0.46) | (0.16) |
| **sch 60** | 0.03 | -0.02 | -0.07 | 0.76 | 1.79 | 2.88 |
| | (0.13) | (0.22) | (0.12) | (0.14) | (0.32) | (0.14) |
| **sch 61** | 0.48 | 0.47 | -0.19 | 0.24 | 0.10 | 0.18 |
| | (0.08) | (0.10) | (0.09) | (0.10) | (0.44) | (0.19) |
| **sch 62** | 0.07 | -0.06 | 0.29 | 0.10 | 0.30 | -0.12 |
| | (0.19) | (0.35) | (0.12) | (0.21) | (0.32) | (0.25) |
| **sch 63** | 0.49 | 0.46 | 0.82 | 0.88 | 2.19 | 1.91 |
| | (0.16) | (0.30) | (0.09) | (0.16) | (0.19) | (0.17) |
| **sch 64** | -0.21 | 0.13 | -0.34 | 0.37 | 1.18 | 1.06 |
| | (0.12) | (0.16) | (0.10) | (0.11) | (0.33) | (0.12) |
| **sch 65** | 0.04 | -0.12 | 0.11 | 0.05 | 0.15 | -1.01 |
| | (0.10) | (0.23) | (0.07) | (0.14) | (0.29) | (0.24) |
| **sch 66** | 0.33 | 0.41 | -0.05 | 0.37 | 0.91 | 2.15 |
| | (0.11) | (0.21) | (0.09) | (0.14) | (0.31) | (0.09) |
| **sch 67** | 0.10 | 0.50 | 0.25 | 0.92 | 1.55 | 2.15 |
| | (0.10) | (0.14) | (0.09) | (0.08) | (0.39) | (0.08) |
| **sch 68** | -0.32 | -0.35 | -0.20 | -0.29 | -0.20 | -1.60 |
| | (0.11) | (0.24) | (0.07) | (0.16) | (0.33) | (0.34) |
| **sch 68** | -0.36 | -0.43 | -0.75 | -0.65 | -2.52 | -2.85 |
| | (0.09) | (0.14) | (0.12) | (0.13) | (7.09) | (3.74) |
| **Race** | B | B | H | H | w | w |
| **Income** | Q1 | Q2 | Q1 | Q2 | Q1 | Q2 |
| **No. student-school pairs** | 37,033 | 15,724 | 59,181 | 35,016 | 6,641 | 40,959 |

Note: Coefficients from regressions between the standardized $\delta_j^r$ and school characteristics by race. Standard errors in parenthesis.

Table A.7: School Mean Utilities and School Characteristics - Independent Regressions

| | **Standardized** $\delta_j^r/\beta$ | | | | | |
|---|---|---|---|---|---|---|
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
| % Scored Advanced-Proficient Math | 1.02 | 1.75 | 1.50 | 2.16 | 1.50 | 2.51 |
| | (0.32) | (0.31) | (0.32) | (0.31) | (0.28) | (0.29) |
| | [235] | [235] | [235] | [235] | [235] | [235] |
| % Scored Advanced-Proficient English | 1.09 | 2.43 | 1.34 | 2.64 | 1.80 | 3.20 |
| | (0.39) | (0.36) | (0.39) | (0.37) | (0.33) | (0.33) |
| | [236] | [236] | [236] | [236] | [236] | [236] |
| % of Black Students | -1.31 | -1.42 | -3.38 | -3.38 | -2.50 | -2.45 |
| | (0.31) | (0.30) | (0.24) | (0.24) | (0.24) | (0.26) |
| | [258] | [258] | [258] | [258] | [258] | [258] |
| % of Hispanic Students | -0.38 | -1.01 | 1.66 | 0.39 | 0.34 | -0.45 |
| | (0.32) | (0.31) | (0.30) | (0.32) | (0.28) | (0.30) |
| | [258] | [258] | [258] | [258] | [258] | [258] |
| % of White Students | 2.21 | 3.25 | 2.76 | 4.42 | 2.95 | 4.03 |
| | (0.39) | (0.36) | (0.38) | (0.31) | (0.33) | (0.31) |
| | [258] | [258] | [258] | [258] | [258] | [258] |
| % Low-Income Students in Kindergarten | -0.90 | -1.63 | -1.02 | -1.83 | -1.11 | -1.61 |
| | (0.29) | (0.27) | (0.29) | (0.27) | (0.26) | (0.27) |
| | [256] | [256] | [256] | [256] | [256] | [256] |
| Race | B | B | H | H | w | w |
| Income | Q1 | Q2 | Q1 | Q2 | Q1 | Q2 |

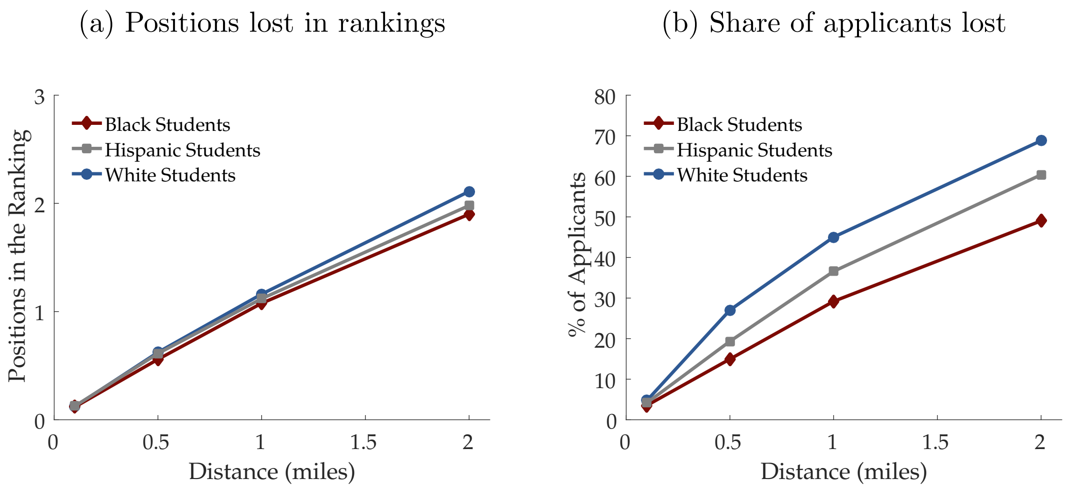
Note: Each coefficient is from an independent regression where the dependent variable is the standardized ratio $\delta_{jr}/\beta_c$. Standard errors in parenthesis and sample size shown below in square brackets.

Table A.8: School Mean Utilities and School Characteristics - Pooled Regressions

| | **Standardized** $\delta_j^r/\beta$ | | | | | |
|---|---|---|---|---|---|---|
| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 |
| % Scored Advanced-Proficient Math | 0.10 | 0.67 | 0.02 | 0.21 | 0.30 | 1.05 |
| | (0.37) | (0.33) | (0.29) | (0.26) | (0.27) | (0.28) |
| % of Black Students | 0.30 | 2.10 | -3.67 | -1.20 | 1.32 | -0.85 |
| | (1.14) | (1.02) | (0.90) | (0.80) | (0.83) | (0.86) |
| % of White Students | 2.26 | 4.04 | 2.11 | 4.12 | 2.88 | 6.67 |
| | (1.38) | (1.24) | (1.09) | (0.96) | (1.00) | (1.04) |
| % of Black Students Squared | -1.01 | -2.82 | 1.05 | -1.24 | -4.12 | 0.17 |
| | (1.51) | (1.35) | (1.19) | (1.05) | (1.09) | (1.14) |
| % of White Students Squared | -1.01 | -3.43 | -3.15 | -3.22 | -3.02 | -8.20 |
| | (2.42) | (2.16) | (1.91) | (1.69) | (1.75) | (1.82) |
| % Low-Income Students in Kindergarten | -0.07 | -0.43 | -0.23 | -0.35 | -0.07 | -0.15 |
| | (0.35) | (0.31) | (0.28) | (0.24) | (0.25) | (0.26) |
| Constant | -0.28 | -0.76 | 0.99 | 0.24 | -0.17 | -0.72 |
| | (0.41) | (0.37) | (0.33) | (0.29) | (0.30) | (0.31) |
| Observations | 233 | 233 | 233 | 233 | 233 | 233 |
| Race | B | B | H | H | w | w |
| Income | Q1 | Q2 | Q1 | Q2 | Q1 | Q2 |
| $R^2$ | 0.12 | 0.29 | 0.45 | 0.60 | 0.42 | 0.51 |

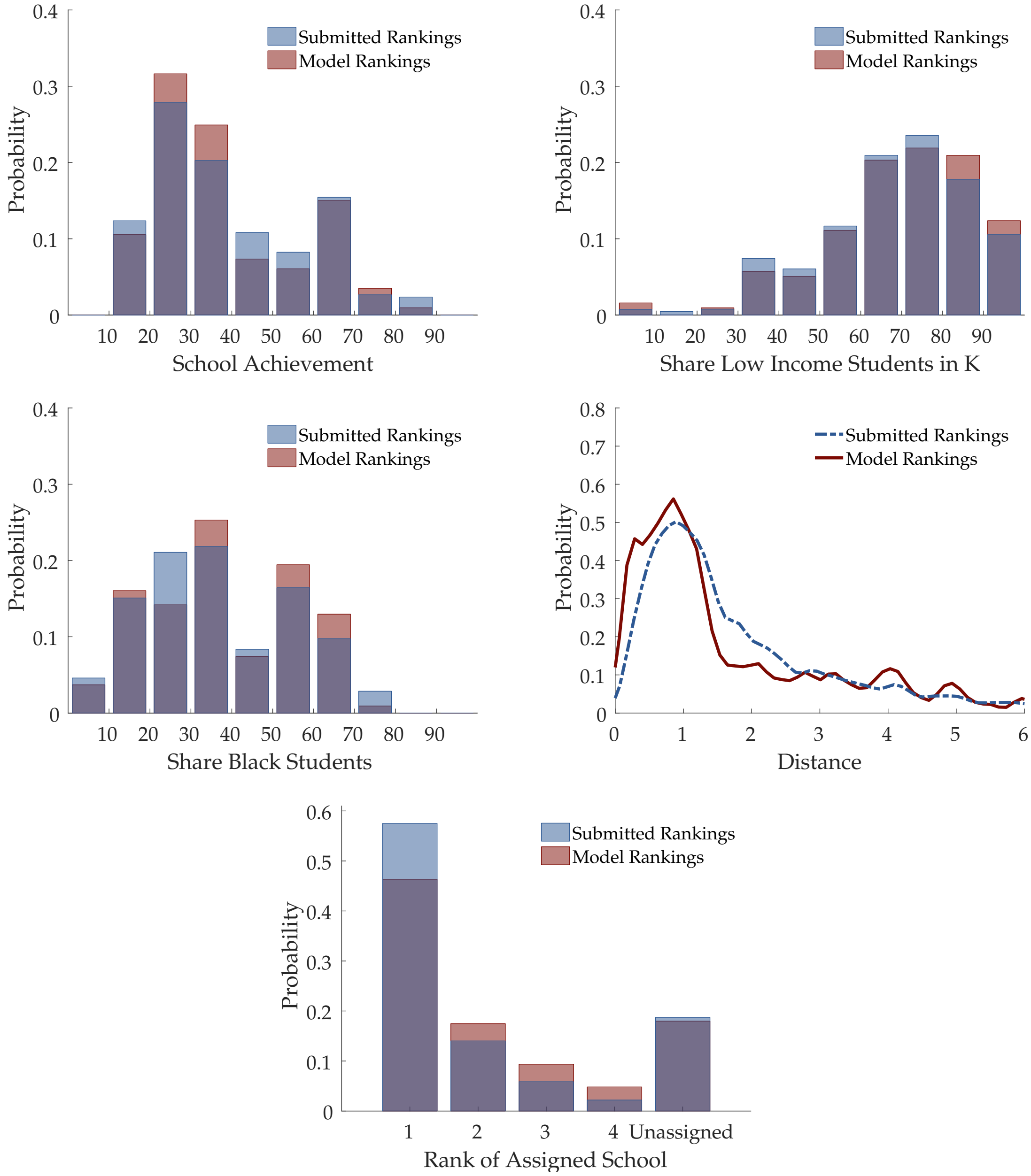
Note: Coefficients from regressions between the standardized $\delta_j^r$ and school characteristics by race. Standard errors in parenthesis.

Figure A.4: Intensive and extensive margin effect of distance increase on applications



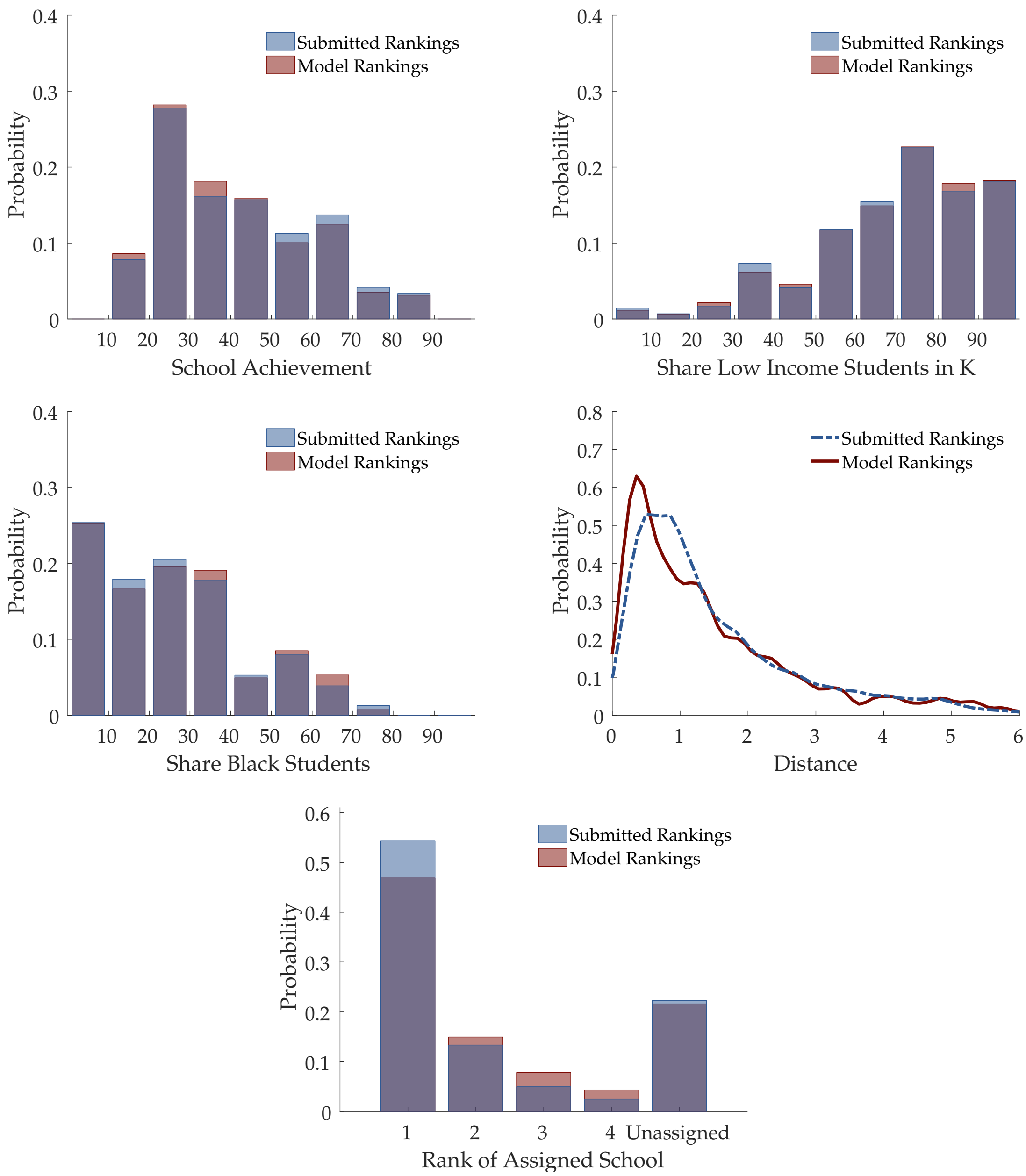

Note: The graphs show the intensive and extensive margin effects on rankings of an increase in the distance between schools and students. In panel (a) I plot the average number of positions that a school would lose if the distance between the school and the students increased by 0.1, 0.5, 1 and 2 miles. Here the average is taken across schools and students. In panel (b) I plot the average share of applicants lost if the distance between the school and the students increased by 0.1, 0.5, 1 and 2 miles, where the average is taken across schools. To model these changes I simulate rankings using the parameters of the model and random draws of $\epsilon_{ij}$ after I increase the distance between each school and all students.

Figure A.5: Fit of Estimated Parameters -Black Students



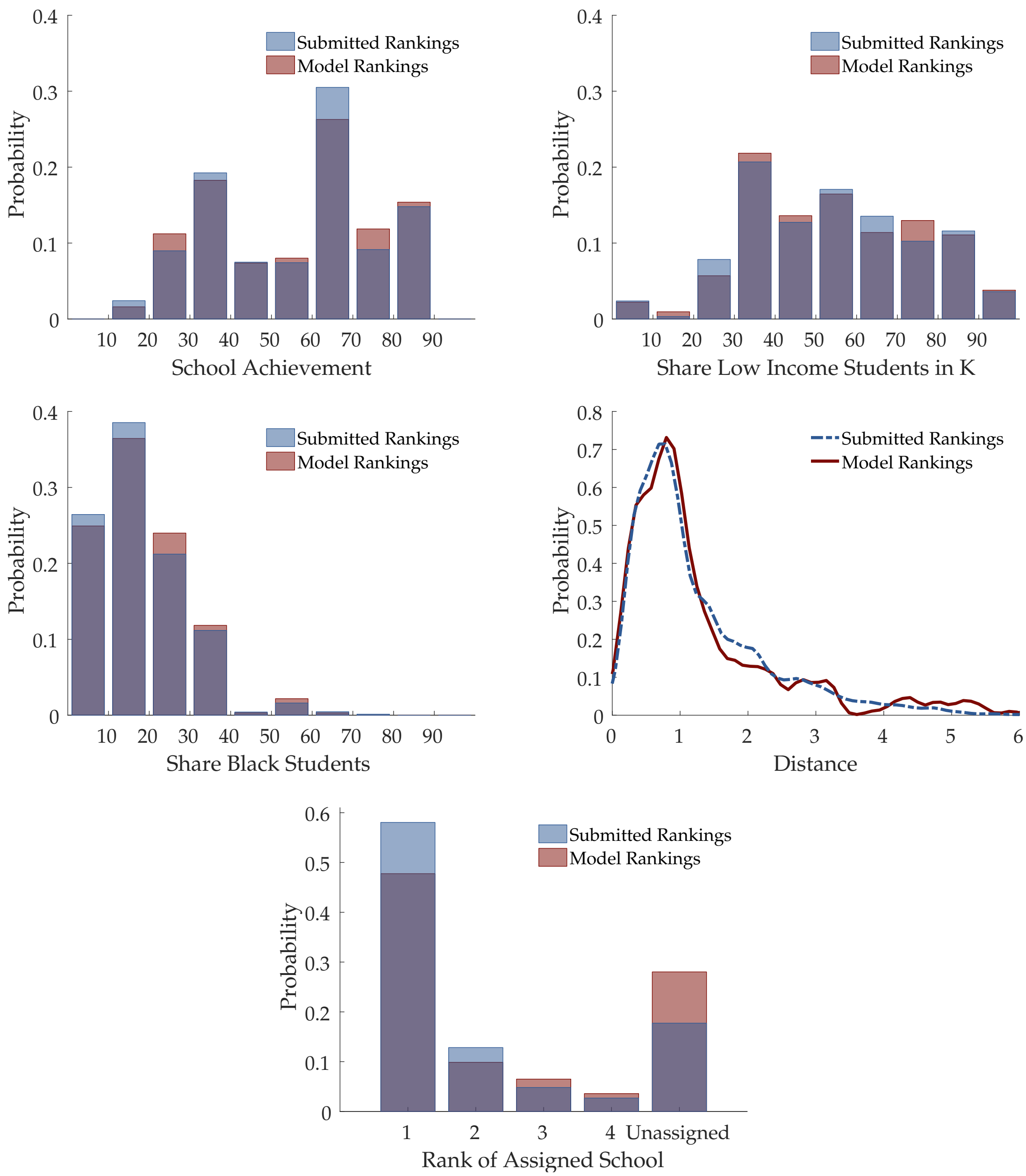

Note: Results for submitted rankings are obtained after running the DA using the rankings submitted by parents to BPS. Simulated rankings are obtained from rankings generated using demand parameters and realizations of $\epsilon$.

Figure A.6: Fit of Estimated Parameters - Hispanic Students




Note: Results for submitted rankings are obtained after running the DA using the rankings submitted by parents to BPS. Simulated rankings are obtained from rankings generated using demand parameters and realizations of $\epsilon$.

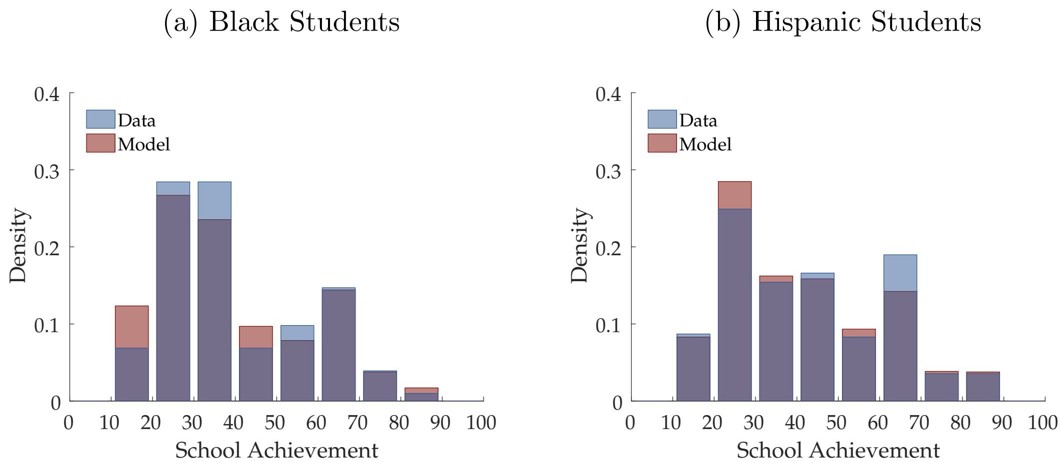
Figure A.7: Fit of Estimated Parameters - White Students




Note: Results for submitted rankings are obtained after running the DA using the rankings submitted by parents to BPS. Simulated rankings are obtained from rankings generated using demand parameters and realizations of $\epsilon$.

# B Counterfactual Analysis

Figure B.8: Distribution of School Achievement of Siblings' School




Note: In blue, the histogram of school achievement at the schools where Black and Hispanic students have a sibling priority. In red, the histogram of school achievement for the predicted sibling's school for the same set of students in their original residential locations. The prediction is made running the DA algorithm with simulated rankings, assuming no student has a sibling priority. After running this assignment, I say that school $j$ is the prediction of the sibling's school for students $i$, if $i$ is assigned to school $j$. Intuitively, this would have been the assignment the older sibling would have gotten, assuming the family's preference parameters, information, and residence did not change.

Table B.9: Summary Counterfactuals: Main Specification

| | **School Achievement** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Original | Preferences | Location 1 | Location 2 | Location 3 | Location 4 | Walkzone Priority | Choice Menu |
| Black Students | 38.9 | 45.8 | 47.2 | 47.1 | 47.9 | 48.5 | 39.3 | 38.5 |
| | (0.2) | (0.8) | (0.3) | (0.3) | (0.3) | (0.3) | (0.2) | (0.3) |
| Hispanic Students | 40.9 | 45.2 | 47.8 | 47.9 | 48.1 | 48.9 | 41.0 | 40.4 |
| | (0.1) | (0.6) | (0.2) | (0.2) | (0.2) | (0.2) | (0.1) | (0.1) |
| | **% Gap Reduction** | | | | | | | |
| | Original | Preferences | Location 1 | Location 2 | Location 3 | Location 4 | Walkzone Priority | Choice Menu |
| Black Students | | 40.8 | 49.1 | 48.6 | 53.2 | 56.9 | 1.9 | -2.4 |
| | | (4.6) | (1.7) | (1.7) | (1.7) | (1.7) | (0.4) | (0.6) |
| Hispanic Students | | 29.1 | 46.6 | 46.8 | 48.8 | 53.8 | 0.4 | -3.5 |
| | | (3.9) | (1.5) | (1.5) | (1.5) | (1.6) | (0.2) | (0.3) |

Note: Counterfactual results from main random utility specification. Bootstrapped standard errors are obtained by running counterfactuals with each of 100 vectors of preference parameters.

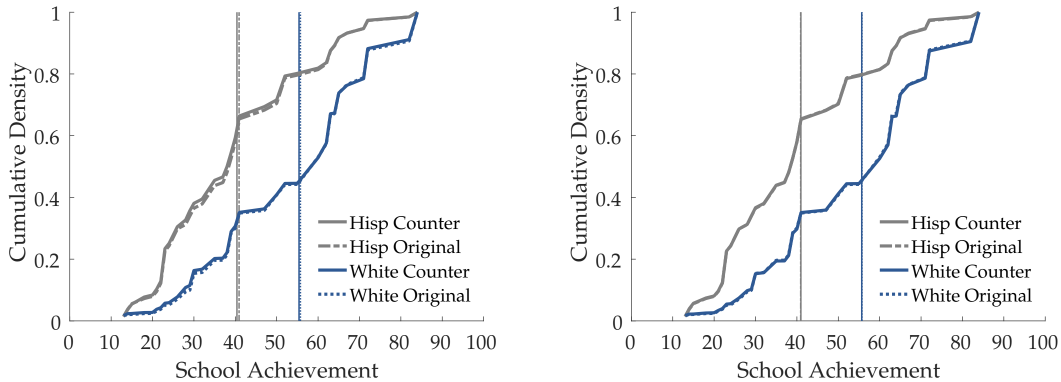
Table B.10: Summary Counterfactuals: Alternative Specifications

| | **% Gap Reduction** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Linear in distance + Neighborhood-level outside option** | | | | | | | |
| | Original | Preferences | Location 1 | Location 2 | Location 3 | Location 4 | Walkzone Priority | Choice Menu |
| Black Students | | 40.6 | 46.4 | 46.0 | 51.3 | 54.8 | 2.5 | -3.4 |
| | | (4.2) | (1.8) | (1.7) | (1.9) | (1.9) | (0.4) | (0.8) |
| Hispanic Students | | 29.5 | 44.8 | 45.0 | 47.5 | 52.3 | 0.4 | -3.1 |
| | | (4.0) | (1.8) | (1.8) | (1.8) | (2.0) | (0.2) | (0.4) |
| | **Quadratic in distance + Common outside option** | | | | | | | |
| | Original | Preferences | Location 1 | Location 2 | Location 3 | Location 4 | Walkzone Priority | Choice Menu |
| Black Students | | 39.7 | 49.6 | 49.2 | 53.7 | 57.0 | 1.8 | -1.9 |
| | | (4.6) | (1.8) | (1.7) | (1.8) | (1.8) | (0.3) | (0.7) |
| Hispanic Students | | 28.4 | 48.9 | 49.2 | 50.9 | 55.8 | 0.3 | -2.7 |
| | | (3.1) | (1.4) | (1.4) | (1.4) | (1.6) | (0.2) | (0.4) |
| | **Quadratic in distance + Neighborhood-level outside option** | | | | | | | |
| | Original | Preferences | Location 1 | Location 2 | Location 3 | Location 4 | Walkzone Priority | Choice Menu |
| Black Students | | 39.8 | 47.5 | 47.1 | 52.2 | 55.3 | 1.8 | -2.7 |
| | | (4.3) | (1.7) | (1.7) | (1.8) | (1.8) | (0.4) | (0.8) |
| Hispanic Students | | 28.8 | 47.3 | 47.6 | 50.1 | 54.5 | 0.1 | -2.5 |
| | | (3.0) | (1.7) | (1.7) | (1.7) | (2.0) | (0.2) | (0.4) |

Note: Counterfactual results from three alternative random utility specification. Bootstrapped standard errors are obtained by running counterfactuals with each of 100 vectors of preference parameters.

Figure B.9: Eliminate location-specific rules - Hispanic Students

(a) Choice-Menu (b) Walk-Zones




Note: Distribution of achievement in schools assigned to Hispanic and white students under a counterfactual assignment where choice-menu restrictions are eliminated (on the left), and walk-zone priorities are abolished (on the right).

# C Maximum Likelihood Function

Let $R_i = (R_{i1}, \cdots, R_{il_i})$ be the rank-order list submitted and $\mathcal{J}_i$ the choice set of $i$. The conditional likelihood of $R_i$ is

$$\mathcal{L}(R_i|\beta_c, \gamma_c, \delta_{\mathbf{cj}}) = \left[ \prod_{k=1}^{l_i} \frac{\exp(u_{iR_{ik}})}{1 + \sum_{j \in \mathcal{J}_i \setminus \{R_{im}: m<k\}} \exp(u_{iR_{ij}})} \right] \left[ \frac{1}{\sum_{j \in \mathcal{J}_i \setminus \{R_{im}: m<l_i\}} \exp(u_{iR_{ij}})} \right] \tag{6}$$

By maximum likelihood, I find the values of $\{\beta_c, \gamma_c, (\delta_{cj})_j\}_c$ that maximize

$$\prod_{i \in c(X_i)} \mathcal{L}(R_i|\beta_c, \gamma_c, \delta_{\mathbf{cj}})$$

for each $c$.

# D Assignment Algorithm

With the exception of a couple of schools, half of the seats at each school are assigned using the priority order explained in the main text. This includes sibling and walk-zone priorities. For the second half of seats, the priority does not include any walk-zone considerations. In consequence, students with a sibling have the first priority and the rest have the second priority. Ties between groups are broken using a unique random number drawn for each student.

Now, since a student may be eligible for seats in both halves at each school, a precedence order across halves is established. This is, the rule that determines whether a student is first considered for the first or second half of the seats at a school. A student with a walk-zone priority will be considered for the walk-half first while a student outside the walk-zone is considered for the second half first. The DA algorithm, described below, is ran over school halves.

- *Step 1:* Applicants are sorted in priority order in their first ranked schools and students in excess of capacity are rejected. Those who are not rejected are provisionally admitted.

- *Step k:* For students rejected in step $k - 1$, their next preferred option is considered. Each school ranks by priority order the set of provisionally admitted students jointly with those new students who are being considered in $k$. The program provisionally admits those with the highest priority and rejects students in excess of capacity. The algorithm stops when every rank list has been exhausted or when there are no rejections.

More details about the assignment algorithm can be found in Pathak and Shi 2013a and Pathak and Shi 2013b.