Distributional Impacts of Centralized School Choice*

Jaewon Lee[1] and Suk Joon Son[2]

[1]Yale University
[2]The University of Tokyo

April 16, 2022

Abstract

Centralized school choice systems are being adopted widely, partly based on theoretical results that promise desirable outcomes such as matching stability and strategy-proofness. However, these theoretical results do not directly address many important distributional goals, such as racial integration and equity. Furthermore, even the theory-predicted outcomes may not arise if the applicants are not fully informed or rational. Using data from New York City's centralized high school choice system, we assess the impact of school choice on distributional outcomes and decompose the contributions of students' residential locations, preferences, informational frictions, and schools' admission policies. We also quantify matching stability. To these ends, we estimate a model of school applications that allows the applicants to have mistaken beliefs about admission chances and to consider only a limited set of schools. Exogenous variation, such as positions of schools in the school directory, along with rich information in students' rank-ordered lists of schools, identifies the model. The results show that, while school choice integrates races to a small degree and improves welfare across races, these gains and the stability of the school assignments are compromised by deviations from fully informed behavior. Schools' screening policies contribute to racial segregation and tend to place Asian and White students in their preferred schools.

*We are indebted to Steven Berry, Philip Haile, and Yusuke Narita for their constant guidance and support. We thank Jason Abaluck, Victor Aguirregabiria, Joseph Altonji, Ian Ball, Marianne Bernatzky Koehli, Barbara Biasi, Ali Bray, Jin-wook Chang, Yeon-koo Che, Andrew Chesher, Susan Dynarski, Chao Fu, Kenneth Gillingham, Charles Hodgson, John Eric Humphries, Mitsuru Igami, Gerald Jaynes, Alvin Klevorick, Changhyun Kwak, Koohyun Kwon, Soonwoo Kwon, Mariana Laverde, Costas Meghir, Luis Carvalho Monteiro, Christopher Neilson, Cormac O'Dea, Jonathan Pierot-Hawkins, Jeff Qiu, Katja Seim, Nicholas Snashall-Woodhams, Jaehee Song, Jintaek Song, Eduardo Souza-Rodrigues, Allen Vong, Stephanie Weber, and Seth Zimmerman for helpful discussions, advice, and encouragement. We are grateful of the support from the New York City Department of Education, particularly from Derek Li, Joshua Smith, and Stewart Burns Wade.

1 Introduction

School choice policies aim to mitigate the effect of students' residential locations on educational opportunities by allowing students access to schools beyond their neighborhoods. One version of these policies is to use a centralized assignment mechanism, often motivated by results from economic theory, such as those that guarantee stability, efficiency, or strategy-proofness.[1] However, the impacts of these mechanisms are still debated. The theoretical results depend on the assumption that the applicants make well-informed and rational choices.[2] They also assume ideal versions of the mechanisms that often differ from those implemented in practice.[3] Furthermore, while many policymakers regard distributive goals such as racial integration and equity to be among the primary objectives regarding school allocations, the theoretical results do not directly address these goals. Therefore, the impact of centralized school choice on a variety of outcomes is an empirical question.

In this paper, we use administrative data from the New York City (NYC) Department of Education (DOE) to examine the impacts of its centralized public high school choice procedure in 2016-2017. We develop a model of students' application behavior that allows for two types of optimization frictions: applicants may consider only a limited set of schools, and they may have incorrect beliefs about admission chances. We analyze the impact of school choice on integration and equity of welfare across different demographic groups. We further measure the contributions of different factors: students' residential locations, preferences, limited consideration sets, and schools' priority groups and rankings over the students. The model also allows us to measure the contribution of strategic reporting and quantify matching stability. We provide sufficient conditions that ensure the model is nonparametrically identified using the type of rank-ordered choice data typically available from centralized school choice systems.

NYC's high school assignment procedure allocates approximately 80,000 students to more than 700 school programs each year, forming a part of the largest centralized school choice system in the United States. The procedure matches students to schools based on the students' submitted rank-ordered lists of school programs and the school programs' priorities or rankings over the students. The assignment mechanism is a version of the Deferred Acceptance (DA) mechanism, which in theory guarantees matching stability and strategy-proofness

[1] See, e.g., Gale and Shapley (1962), Shapley and Scarf (1974), Ergin (2002), and Abdulkadiroğlu and Sönmez (2003). Such centralized mechanisms are used in, for example, New York City, Chicago, Boston, New Orleans, Paris, Spain, and Romania (Abdulkadiroglu et al., 2017; Fack et al., 2019).

[2] See, e.g., Hassidim et al. (2017), Li (2017), and Fack et al. (2019).

[3] See, e.g., Abdulkadiroğlu et al. (2005), Haeringer and Klijn (2009), and Calsamiglia et al. (2010).

under idealized assumptions.

However, theory promises little about important distributional outcomes such as diversity in schools and equity in education, which NYC DOE regards as its top priorities (NYC DOE, 2020a,b,c). Potentially reflecting this tension, students and others have raised concerns about the diversity in its schools and the equity across demographic groups.[4] Consistent with the concerns, we document that minority students tend to be matched to low-performing schools.

Moreover, even the theoretically predicted properties—matching stability and report truthfulness—may not hold in practice. These properties rely on the assumptions that applicants are fully informed and rational and that the canonical version of the DA mechanism is implemented. However, NYC's DA mechanism deviates from its canonical version,[5] and truthful reporting of preferences cannot be guaranteed to be weakly dominant. Therefore, there are situations where an applicant must assess admission chances in order to strategize optimally. We provide evidence that suggests students' reports are affected by admission chances in addition to their preferences. Evidence also suggests that applicants may not understand the properties of the mechanism and take admission chances into account even under situations where such behavior is weakly dominated. Furthermore, because more than 700 school programs are allocated through the mechanism, applicants are unlikely to be aware of every option. We find that students are significantly less likely to apply to the schools listed in the later pages of NYC's school directory, suggesting that they are not aware of all schools.

Imperfections in the applicants' knowledge or in matching mechanisms can have distributional consequences. Failure of strategy-proofness may undermine fairness; students with lower socioeconomic backgrounds may be less likely to understand the exact properties of NYC's DA mechanism. Therefore, they may have more difficulties in formulating an optimal report.[6] They may also be less informed about higher-quality schools.[7]

A model that does not allow for frictions in application behavior forces the researcher to interpret any observed behavior under school choice as optimal behavior in terms of the applicants' preferences, potentially biasing the results in favor of the use of school choice. Furthermore, a frictionless model attributes differences in the choice patterns across different demographic groups to differences in preferences when, in fact, they may be caused by differences in frictions. On the other hand, a model that allows for frictions enables the researcher

[4]See, e.g., https://gothamist.com/news/wheres-our-mayor-nyc-students-rally-against-school-segregation

[5]The students are constrained in the number of school programs that they can list, and there are reassignments following the initial assignment procedure.

[6]See, e.g., Pathak and Sönmez (2008), Sattin-Bajaj (2016), and Basteck and Mantovani (2018).

[7]See, e.g., Sattin-Bajaj (2016) and Corcoran et al. (2018).

to disentangle the contributions of different factors—including optimization frictions—to the observed outcomes and to provide guidance on possible policy interventions.

Therefore, my model allows each applicant to consider only a limited set of the school options and to have incorrect beliefs about equilibrium admission chances. An applicant may fail to consider some school because she is not aware of the school or feels she can never be admitted to the school. Consideration of a school is determined by a latent variable whose distribution depends on the observables. Even if the applicant considers a school, her beliefs about admission chances could still be incorrect. The beliefs about admission chances are determined by the applicant's expectations about her admission scores relative to the schools' cutoffs for admission.[8] Such specification allows for misunderstanding of the matching mechanism. We assume each student maximizes her expected utility with frictions in consideration and in assessments about admission chances.

Rich information in students' rank-ordered lists of schools and some exogenous variation, such as positioning of schools in the school directory, allows us to identify the model. For instance, while a lack of consideration may affect *which* schools are listed, it cannot affect *where* the listed schools will be ranked. Furthermore, while strategic behavior reflecting applicants' beliefs about admission chances may affect *where* the school will be ranked, it cannot affect *which* schools are listed unless the list length constraint binds. Identification is also aided by an assumption that certain observables, such as the page on which a school appears in NYC's school directory, can affect the consideration set but not preferences.[9] Another assumption that aids identification is that some students have a set of schools (e.g., noncompetitive schools close to home) that they will surely consider. We formalize these intuitive ideas with sufficient conditions for nonparametric identification, which demonstrate that the identification does not depend on functional form assumptions and clarify the role of each assumption and source of data variation.

Using the estimated model, we consider theoretical predictions under idealized implementations: stability of matching and optimality and truthfulness of students' reports. The estimated model allows us to quantify stability by measuring the prevalence of justified envy;[10] a stable matching does not have any case of justified envy. We find that limited consideration

[8]The model of beliefs closely follows that of Kapor et al. (2020) and is related to Ajayi and Sidibé (2017) and Luflade (2018).

[9]This is similar to the idea of Martin and Yurukoglu (2017) that use local channel positions as exogenous variation that shift channel viewership but are uncorrelated with the local political inclinations.

[10]We say that a student has *justified envy* for a school if the student and the school are not matched to each other, but both would prefer to be matched to one another than to (one of) the current assignments. This definition is consistent with the usage of the term in the literature.

results in significant amounts of justified envy. Students view approximately nine schools with justified envy on average; we estimate that Black and Hispanic students view approximately 10.4 schools with justified envy while Asian and White students view approximately 5.5 schools with justified envy. The estimated model also enables us to simulate students' subjectively optimal reports and estimate what fraction of the reports are truthfully ordered in terms of their preferences. Similarly, by comparing the subjectively optimal reports to the reports that are objectively optimal in terms of the equilibrium admission probabilities, the optimality of the reports can be evaluated.

We also quantify racial integration and equity in welfare. The differences in outcomes under school choice matching relative to counterfactual neighborhood school matching represent the impact of school choice. We then decompose the contributions of different factors—residential locations, student preferences, schools' priorities and rankings over the students, and optimization frictions. The results show that school choice slightly promotes racial integration relative to the neighborhood matching. For students of each race, the average proportion of the own-race students in the students' assigned schools decreases by approximately 0 to 10 percentage points. School choice also significantly improves welfare across all races; while only about 6% of the students would be matched to one of their top five preferred schools under the neighborhood matching, the proportion increases to approximately 35% under the school choice matching. However, these gains are compromised by deviations from fully informed behavior. If students considered all schools, students would be about twice as likely to be matched to their top five preferred schools. Schools' admission priorities and screening policies segregate races and tend to place Asian and White students to their preferred schools.

Related Literature My paper complements the studies that empirically examine the contributions of different factors to equity or segregation under centralized school choice procedures (Oosterbeek et al., 2019; Laverde, 2020; Sartain and Barrow, 2020). My paper disentangles the impacts of lack of information from preferences, given abundant evidence of informational frictions in NYC. Relatedly, Kessel and Olme (2018) focus on the impact of school priorities on segregation and Calsamiglia et al. (2020b) theoretically examine the impact of matching algorithms on segregation. Akbarpour et al. (2020) show that strategy-proof algorithms can neutralize the impacts of unequal outside options. There have been studies that examine the distributional impacts of school choice in other contexts (e.g., Epple and Romano, 1998; Hsieh and Urquiola, 2006; Bifulco and Ladd, 2007; Neilson, 2013; Altonji et al., 2015; Avery and Pathak, 2015; Hom, 2018).

My paper provides sufficient conditions for nonparametric identification of latent consideration and latent beliefs about admission chances, in addition to latent preferences, using data on observed choices. These results build on a broad literature that deals with some but not all of the issues discussed in my paper. Ajayi and Sidibe (2020) allow for beliefs about admission chances as well as limited consideration due to search costs. In their model, each applicant engages in a costly search process to expand her consideration set until the marginal benefit of search exceeds the marginal cost. They model the search technology to be homogeneous across observable student characteristics, while my paper allows it to be different across, for instance, ethnicities or neighborhoods. My paper further complements their paper by providing results on nonparametric identification. Agarwal and Somaini (2018) provide sufficient conditions for nonparametric identification of preferences while assuming full consideration and holding fixed a mode of beliefs. Kapor et al. (2020) estimate a model that allows for latent beliefs about admission chances in addition to latent preferences, using survey data on perceived admission chances and data on rank-ordered lists. My model of beliefs largely follows theirs, and my paper complements their work by providing results on identification that use data on observed choices rather than survey data. Relatedly, Luflade (2018), Calsamiglia et al. (2020a), and Ajayi and Sidibé (2017) estimate preferences and beliefs with the observed choice data without surveys. Some papers propose strategies for estimating preferences while allowing for mistaken beliefs under the Boston mechanism (He, 2017; Hwang, 2017) and while allowing for nontruthful behavior under the DA mechanism (Artemov et al., 2017; Fack et al., 2019).[11] My paper also relates to the broader literature on identification of discrete choice models.[12]

My paper quantifies the influences of different factors on student welfare and matching stability. Relatedly, Abdulkadiroğlu et al. (2017) compares student welfare under coordinated and uncoordinated assignment procedures in the NYC high school choice system. Luflade (2018) analyzes the value of information about admission chances on welfare. Other studies compare student welfare or matching stability under different school assignment procedures (e.g., Narita, 2016; Abdulkadiroglu et al., 2017; He, 2017; Hwang, 2017; Agarwal and Somaini, 2018; Luflade, 2018; Che and Tercieux, 2019; Kapor et al., 2020; Calsamiglia et al., 2020a). My

[11] There are also studies that only assume "weak" versions of the truthtelling assumption (e.g., Abdulkadiroğlu et al., 2017; Che and Tercieux, 2019).

[12] My paper relates to a strand of the discrete choice literature that discusses the identification of preferences and consideration sets (e.g., Goeree, 2008; Conlon and Mortimer, 2013; Gaynor et al., 2016; Hortaçsu et al., 2017; Abaluck and Adams, 2017) and of preferences and beliefs (e.g., Aguirregabiria (2021)). The approaches used in nonparametric identification results are further related to, for example, Thompson (1989), Bresnahan and Reiss (1991), Lewbel (2000), Berry et al. (2013), and Berry and Haile (2020).

paper attempts to prevent the influences of frictions in both consideration and the assessment of admission chances from being attributed to utilities.

We also contribute to the growing literature that documents frictions in centralized school choice. We first document evidence of nontruthful reporting even when such a strategy is weakly dominated, which is consistent with the findings in Artemov et al. (2017) and Fack et al. (2019). Such evidence complements the theoretical findings that the strategy-proofness of DA may be difficult for the boundedly rational agents to understand (Li, 2017; Ashlagi and Gonczarowski, 2018) and the related findings from surveys and experiments (Chen and Sönmez, 2006; Calsamiglia et al., 2010; Hassidim et al., 2017). We further document evidence that students may not be aware of all the available school options. Corcoran et al. (2018) provide evidence that information intervention affects application behavior in the NYC high school application procedure; others reach similar conclusions in other environments of school or college applications (e.g., Hastings and Weinstein, 2008; Hoxby and Turner, 2013; Ajayi et al., 2017; Dynarski et al., 2021).

2 Overview of New York City's High School Choice

This section gives an overview of the public high school choice in NYC. Section 2.1 provides the context of NYC's public high school choice. Section 2.2 explains the theoretical properties of the DA mechanism implemented in NYC and potential failures in practice.

2.1 The Context

The NYC public high school choice system matches approximately 80,000 students to more than 700 public high school programs each year. The system uses the following centralized procedure:

(1) Each applicant submits her rankings over the school programs. She can rank up to 12 school programs.

(2) Each school program ranks applicants according to the admission policies. The rankings can depend on the admission priority groups assigned to the students, screening based on students' past performances and other factors, and lotteries.

(3) NYC runs a student-proposing DA algorithm to assign students to school programs using the rankings of the students and the school programs.

Table 1: Characteristics of Students by Ethnicity

	Asian	Black	Hispanic	White	Total[a]
Proportion in the sample	16.2%	27.1%	40.7%	14.5%	98.5%
Female	49.7%	50.1%	49.1%	50.2%	49.6%
English Language Learner	8.6%	1.5%	13.5%	4.0%	7.9%
Subsidized lunch	68.3%	76.5%	81.2%	37.1%	70.9%
Neighborhood income[b]($)	58853.2	49797.2	47332.5	73893.4	54071.8
Home boroughs					
Bronx	5.4%	24.5%	36.2%	5.9%	23.1%
Brooklyn	28.5%	43.3%	20.1%	34.2%	30.0%
Manhattan	8.1%	8.5%	12.2%	13.5%	10.7%
Queens	53.8%	20.6%	26.9%	25.4%	29.6%
Staten Island	4.2%	3.2%	4.6%	21.0%	6.7%
Home language					
English	28.3%	91.6%	40.7%	69.0%	56.7%
Spanish	0.5%	0.6%	59.1%	1.0%	24.3%
Any Chinese	38.1%	0.1%	0.0%	0.3%	6.6%
Other	33.2%	7.7%	0.1%	29.7%	12.4%
State Reading Category					
High	46.3%	17.6%	17.2%	49.0%	27.2%
Middle	46.4%	66.7%	65.2%	46.0%	59.4%
Low	7.2%	15.7%	17.7%	4.9%	13.4%

Notes: Except for the *proportion in the sample*, all the percentage terms represent the proportions of the relevant categories within each ethnicity.

[a] 1.5% of students are multi-racial or Native American.

[b] based on the ZIP code of student's home address. Median household income from U.S. Census Bureau, 2013-2017 American Community Survey five-year estimates, in 2017 dollars.

See Appendix B for algorithmic rules of the DA and the details of implementation. The matching procedure in NYC creates incentives for the applicants to deviate from truthfully reporting their preferences, despite the well-known property of DA to be strategy-proof for the proposing side in its ideal implementation. This is discussed in Section 2.2.

Characteristics of the student sample are summarized in Table 1.[13] The district has many minority students and low-income students. Of the students in the sample, 40.7% of the students are Hispanic, 27.1% are Black, 16.2% are Asian, and 14.5% are White.[14] 71% of the students are eligible for free or reduced-price lunch.

[13] For discussions of the data and the sample, refer to Section 3.1.

[14] For a more detailed explanation of race and ethnicity, refer to the NYC DOE's survey on ethnicity and race identification: https://www.schools.nyc.gov/docs/default-source/default-document-library/pseform-english. We use race and ethnicity interchangeably in this paper.

Table 2: Characteristics of Schools by Borough

	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Graduation rate	0.69	0.74	0.79	0.79	0.78	0.75
	(0.14)	(0.14)	(0.16)	(0.16)	(0.10)	(0.15)
College/career rate	0.49	0.52	0.61	0.64	0.65	0.56
	(0.15)	(0.17)	(0.19)	(0.19)	(0.15)	(0.18)
Average grade 8 math (std.)	-0.50	-0.19	0.34	0.48	0.53	0.00
	(0.58)	(0.82)	(1.17)	(1.11)	(0.64)	(1.00)
Value-added (std.)	-0.10	0.02	0.05	0.10	-0.53	0.00
	(1.12)	(0.96)	(0.92)	(1.01)	(0.68)	(1.00)
Proportion White	0.03	0.08	0.10	0.11	0.43	0.08
	(0.04)	(0.12)	(0.15)	(0.11)	(0.21)	(0.13)
Proportion Black	0.28	0.54	0.26	0.28	0.17	0.34
	(0.12)	(0.28)	(0.15)	(0.26)	(0.12)	(0.24)
Proportion Asian	0.03	0.07	0.09	0.22	0.08	0.10
	(0.03)	(0.10)	(0.12)	(0.15)	(0.03)	(0.12)
Proportion Hispanic	0.65	0.30	0.52	0.35	0.28	0.45
	(0.13)	(0.22)	(0.21)	(0.21)	(0.12)	(0.24)
9th grade capacity	116.82	159.89	137.38	185.73	316.67	151.17
	(78.57)	(141.34)	(84.23)	(138.49)	(206.80)	(120.31)
Number of schools	111	113	101	79	9	413

Notes: The standard deviations in each respective borough or in NYC are given in parentheses. Standardized values are indicated by (std.). *College/career rate* indicates the proportion of students who graduated from high school four years after entering 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. *Value-added* is the measure of the school value-added by a school, which is provided by the NYC DOE and is based on the school's performance relative to a comparison group of similar students. All schools have equal weight regardless of their class sizes. The sample excludes the nine specialized high schools and schools with missing data.

The school characteristics are summarized in Table 2 by borough. Schools vary widely in their characteristics, within and across boroughs. For example, while on average Hispanic students comprise 65% of the student body in a school in the Bronx, they comprise only 28% in Staten Island. There is also wide within-borough variability; for instance, the standard deviation of the proportion of Hispanic students is as large as 22% within Brooklyn. The schools tend to be small; the average capacity of a 9th-grade class is around 150. While there are only nine schools in Staten Island, there are roughly around 100 schools in each of the other four boroughs.

2.2 Deferred Acceptance Algorithm: Theory and Practice

The DA algorithm has been gaining wider popularity[15] based partly on theoretical results that promise certain desirable properties. One such property is that the mechanism is strategy-proof for the applicants: truthfully reporting their preference rankings weakly dominates any other strategy. Another such property is matching stability. An important feature of matching stability is that the matching does not have any unmatched student-school pair such that each side prefers the other to (one of) the current assignment(s), i.e., the matching does not have any cases of justified envy.[16] However, these properties do not directly address distributional outcomes such as racial integration or the equity of assignments.

Even the two desirable outcomes promised by the theoretical results, namely, stability and truthtelling, may fail in practice. Survey- and experiment-based evidence shows that a fraction of applicants do not truthfully report even in DA mechanisms.[17] Complementing these results, theoretical studies have revealed that although Deferred Acceptance is strategy-proof, it is not "obviously strategy-proof" (Li, 2017) in generic cases in the sense that applicants with limited rationality may not understand its strategy-proofness (Ashlagi and Gonczarowski, 2018). The failure of strategy-proofness may undermine stability.[18] Stability can also fail when students consider only a limited set of schools. Furthermore, theoretically ideal versions of DA that guarantee strategy-proofness and stability are only occasionally implemented in practice.[19]

The matching procedure in NYC creates incentives for the applicants to deviate from truthfully reporting their preferences. This is because NYC's implementation of DA deviates from its canonical implementation in two respects. First, while the canonical implementation allows applicants to list arbitrarily many school programs, in NYC, applicants can list only up to 12 school programs. Students who wish to apply to more than 12 school programs must then decide which of these programs will be listed, which optimally depends not only on their preferences but also on their admission chances to the schools. Reflecting this, the 2017 NYC High School Directory states that "If you are applying to 'reach' programs, be

[15] DA is used in, for example, Boston, Chicago, Finland, Ghana, and Taiwan (Fack et al., 2019).

[16] Closely following the definition in Roth and Sotomayor (1992), a matching is *stable* if there does not exist: (1) any case of a *blocking pair*, i.e., an unmatched student-school pair where each side prefers the other to [one of] the current assignment[s] (which might be an empty seat or no school assignment), and (2) any case of *individual irrationality*, where a student [school] would prefer to remain unmatched [have one additional empty seat] than to be matched to [one of] the current assignment[s]. It follows that a student has justified envy if he is part of some blocking pair.

[17] See, e.g., Chen and Sönmez (2006), Calsamiglia et al. (2010), and Hassidim et al. (2017).

[18] See, e.g., Gale and Shapley (1962). When agents make only payoff-irrelevant deviations from truthful reporting, then the resulting matching can still be stable (Artemov et al., 2017; Fack et al., 2019).

[19] See, e.g., Abdulkadiroğlu et al. (2009), Haeringer and Klijn (2009), and Calsamiglia et al. (2010).

sure to include 'target' or 'likely-match' programs on your application." Second, while the canonical implementation conceives a single round of applications, in NYC, there is an after-market that follows the main round.[20] If a student believes that she can be matched to a school in this after-market, she may choose not to apply to this school in the main round.

In addition, given that there are more than 700 school programs in NYC, it is unlikely that the students are aware of every one of them. Corcoran et al. (2018) has found that providing information about high-performing schools in the local neighborhood altered the students' choices in NYC. Additionally, lower-income families may have differentially less information about high-performing schools due to differences in social networks or other reasons (Sattin-Bajaj, 2016).

In the next section, we document evidence of nontruthful behavior and frictions in information and in the assessment of admission chances.

3 Evidence of Frictions, Disparities, and Usage of Choice

This section presents some motivating descriptive evidence for the main analysis. Section 3.1 introduces the data used. Evidence in section 3.2 suggests a substantial lack of awareness of the schools. The evidence also suggests that students take admission chances into account when applying for the schools and misunderstand the properties of the matching algorithm. Section 3.3 documents the patterns of racial disparities and usage of school choice.

3.1 Data

The main source of data is the administrative data provided by the NYC DOE for the 2016–2017 academic year. The data include students' choices of rank-ordered lists, final school assignments, admission priorities at the school programs, and demographic information. The demographic information includes students' gender, race, English Language Learner status, language spoken at home, home address, subsidized lunch status, disability status, and performance on statewide seventh-grade English and math tests. As some demographic data are missing for the students who did not attend an NYC DOE public school at the time of application, we restrict the sample to be the eighth graders who were attending an NYC DOE public school at the time of application.[21] This sample includes the students who opted

[20] Until 2019, there was a second-round of DA for the schools with remaining seats (see, e.g., Narita (2016)). In 2020, the NYC DOE began to use waitlists, replacing the second-round DA.

[21] There are some ninth graders who participate in the process, but they constitute less than 5% of the total applicants, and they can apply to only a subset of the schools. For computational feasibility, for certain

out of the school choice process, who constitute approximately 5% of the sample. We also use publicly available school-level data provided by the NYC DOE.

3.2 Evidence of Frictions

Evidence shows that students face substantial frictions in learning about the school options and in making strategic decisions while going through the NYC high school application procedure. Table 3 suggests that students tend not to be aware of the schools listed later in the high school directory, which is their primary reference for the application process. It also suggests that students take admission chances into account in their portfolio decisions even when the list length constraint does not bind; in such a case, it is suboptimal to drop a school that the student prefers to the outside option unless the student has zero chance of admission. Each column in the table represents the estimates of a linear probability model that predicts whether a student has applied to a school. An observation is a student-school pair. The students who exhausted all the 12 slots were dropped. The columns labeled as *All* indicates that all such student-school pairs were used. *Surely Aware* uses only the student-school pairs for which the student is assumed to be aware of the school; these are the schools within a half mile from the student's home or within a quarter mile from the student's middle school. *Likely* uses only the student-school pairs for which the student is assumed to believe that he has a positive chance of assignment to the school upon application; these schools are those that did not fill their seats in the prior year and those such that the student is in the first priority group and the school accepted every such student in the prior year. *Surely Considered* uses the intersection of *Surely Aware* and *Likely*. The controls are the interactions of student ethnicity and other student characteristics (subsidized lunch status and indicators of borough) and school-level variables (including indicators of borough, average input math proficiency, attendance rates, and schools' ethnic composition) and student-school specific variables (polynomial in distance, indicators for the high school equaling the student's middle school, and indicators for school's borough being the same as the student's home borough).

Page rank denotes the within-borough rank of the schools in terms of the order in which they are listed in NYC's High School Directory, which is more than 600 pages long.[22] For example, a school with a page rank of five is the fifth school to be listed within its borough. In Table 3, we scaled this variable by 100 so that the fifth school has a value of 0.05. The

results, a random subsample of 10,000 or 20,000 students was used.

[22] According to Sattin-Bajaj et al. (2018), guidance counselors say that the printed directory is the main source of information for the applicants.

Table 3: Linear Probability Model: Regression of Application on Page Rank and Priorities

	Dependent variable: student applies to the school (percentage points)							
	Surely Considered	Surely Aware	Likely	All	Surely Considered	Surely Aware	Likely	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
100 page ranks	0.96	−0.60	−0.29***	−0.64***	−1.64*	−0.60	−0.420***	−0.475***
	(1.18)	(0.96)	(0.03)	(0.02)	(1.23)	(1.03)	(0.03)	(0.02)
Priority rank					−0.3	−0.7**	0.1***	−0.2***
					(0.6)	(0.3)	(0.01)	(0.004)
Mean application rate (p.p.)	15.5	17.0	0.9	1.5	16.5	18.2	0.9	1.6
%Δ apply when Δ page rank = 100	6.3%	-3.5%	-32.2%	-42.6%	-9.4%	-3.2%	-49.5%	-28.8%
%Δ apply when Δ priority rank = 1					-1.6%	-3.6%	11.2%	-15.2%
Controls	No	No	No	No	Yes	Yes	Yes	Yes
Observations	17,086	28,004	2,584,717	7,162,246	13,169	21,837	2,054,134	5,462,688
F Statistic	1.426	0.780	355.175***	2,192.422***	117.587***	175.700***	2,022.398***	5,184.701***

Notes: *p<0.1; **p<0.05; ***p<0.01. An observation is a student-school pair. Students who exhausted all 12 slots were dropped. *All* indicates that all student-school pairs after such drops were used. *Surely Aware* uses only the student-school pairs for which the student is assumed to be aware of the school. *Likely* uses only the student-school pairs for which the student is assumed to believe that she has a positive chance of assignment to the school upon application. *Surely Considered* uses the intersection of these two sets. The 100 page ranks variable indicates the within-borough rank of the schools in terms of the order in which they are listed in the school directory, scaled by 100, so that the first-listed school gets a value of 0.01. *Priority rank* denotes the rank of the admission priority group of the student for the school. Students of the first priority group are given the priority rank of 1.

schools are ordered alphabetically within each borough; therefore, it may be argued that page rank can only affect awareness and not preferences or admission chances. Table A.1 shows that page rank is largely uncorrelated with observable school characteristics. Priority rank denotes the rank of the admission priority group of the student for the school. Students in the first priority group are given the priority rank of one.[23]

The patterns of correlation between page ranks and applications in Table 3 show evidence for the joint hypothesis that (i) the page rank affects the application decision through the awareness channel, and (ii) applicants are indeed aware of the schools in the assumed Surely Aware sets. To see this, suppose that (ii) holds. Then, as the observations in (1), (2), (5), and (6) are the subsets of student-school pairs such that the school is in the Surely Aware set of the student, page rank cannot affect application through awareness in these columns. Thus, the near-zero association of page rank with the application rates in these columns suggests that page rank does not affect the application decision through any channel other than awareness. In contrast, an increase in page rank is associated with significant drops in application probabilities in columns (3), (4), (7), and (8), showing evidence that page rank does significantly affect application rates through awareness. On the other hand, if it is taken as given that page rank affects applications strictly through the awareness channel, these results are consistent with the hypothesis that the applicants are indeed aware of the schools in the assumed Surely Aware sets. The patterns of correlation between priority ranks and applications show evidence for the joint hypothesis that (i) the applicants take their admission priorities into account in their portfolio choice decisions (even when the list length constraint does not bind) exactly through the assessment of admission chances, and that (ii) applicants believe that they have positive admission chances for the Likely schools upon application. To see this, assume that (ii) holds. Note that the observations in columns (5) and (7) are the subsets of student-school pairs such that the school is in the Likely set of the student; therefore, for these columns, priority cannot affect applications by affecting the assessment of whether the applicants have any chance of admission. Thus, the near-zero association of priorities with application rates in these columns suggests that priorities do not affect the application decision through any channel other than admission probabilities. In contrast, lower priorities (higher priority ranks) are associated with drops in application probabilities in columns (6) and (8), showing evidence that priorities affect application rates by affecting the applicants' beliefs about admission probabilities. On the other hand, if it is taken as given that priorities affect applications strictly through admission probabilities

[23]If there are multiple school programs in a school, we selected the program that gives the student the most favorable priority rank.

Table 4: Regression of Submitted Rank on Priority Rank

	Dependent variable: Rank in submitted report					
	Listed 2	Listed 4	Listed 6	Listed 8	Listed 10	Listed 12
Priority rank	0.106**	0.153***	0.130***	0.108**	0.184***	0.241***
	(0.047)	(0.045)	(0.040)	(0.045)	(0.058)	(0.033)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	830	3,148	6,890	8,748	7,581	30,608
F Statistic	73.019***	154.843***	284.708***	345.411***	284.858***	1,124.645***

Notes: *p<0.1; **p<0.05; ***p<0.01. An observation is an applicant-school pair. *Listed k* selects applicants who listed k schools. *Priority rank* denotes the rank of the admission priority group of the student for the school. Students of the first priority group are given the priority rank of 1.

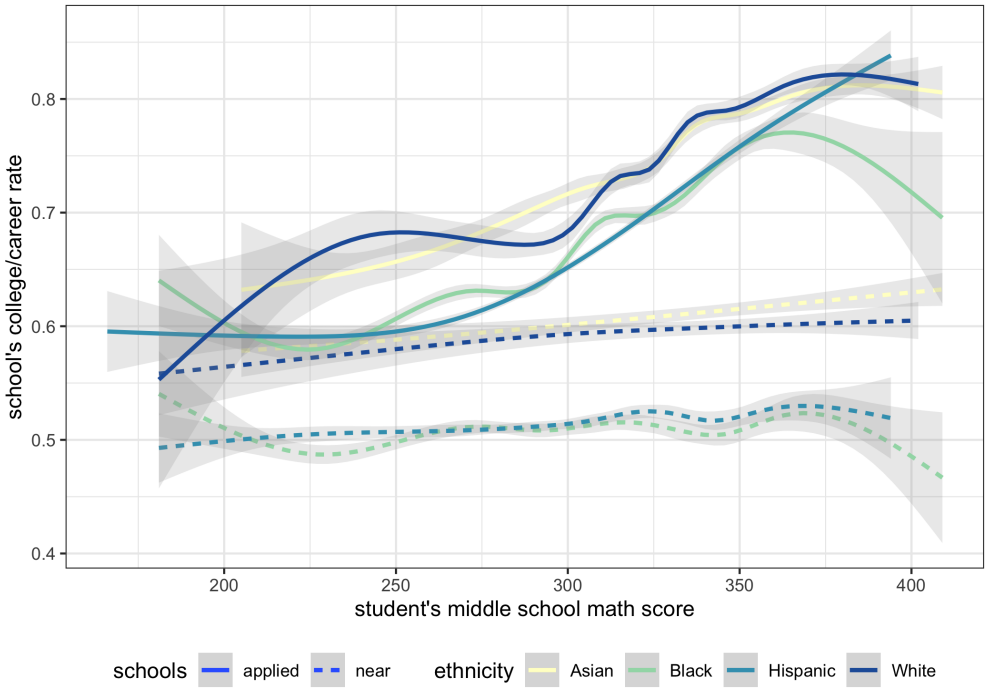
conditional on observables, these results can alternatively be interpreted as justifying the hypothesis that the applicants do believe that they have positive admission chances for the assumed Likely schools upon application.

Table 4 suggests that applicants take priorities into account in ranking the listed schools. Each column represents a linear regression of the submitted rank of a school on priority rank and controls using a sample of applicants who listed a given number of schools. The controls are the same as in Table 3 except that now page rank also constitute the controls. We see that priority group significantly predicts the rank at which a school is listed; an increase in the priority rank by one is associated with a 0.11 increase in the submitted rank for the students who list two schools and a 0.24 increase for the students who exhaust the list by filling in all the 12 slots. However, in a DA mechanism, the submitted rankings among the listed schools should optimally reflect only the preferences, not the admission chances. Therefore, to the extent that priorities are uncorrelated with the unobserved preferences conditional on the controls, the pattern in the table suggests that applicants make mistakes in strategizing due to misunderstandings about the properties of the DA mechanism.

3.3 Patterns of Disparities and Choice

Figure 1 document some patterns of racial disparities and usage of choice. *College/career rate* denotes the school's proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. *Student's middle school math score* is the applicant's performance in the New York State Mathematics test in middle school. The

Figure 1: Nearby and Applied Schools, by Ethnicity



Notes: *College/career rate* denotes the school's proportion of students who enrolled in college, a vocational program, or a public service program within six months of graduation. *Student's middle school math score* is the applicant's performance in middle school in the New York State Math test. The lines represent smoothed conditional means. The shaded regions represent 95% confidence intervals. The dashed lines indicate the schools within one mile from the applicant's home address. The solid lines represent the schools that the applicant has listed on the submitted rank-order report.

figure shows substantial racial disparities in the schools that applicants live close to (within one mile from home). These disparities do not converge even after controlling for applicants' performance in the mathematics tests. However, the patterns suggest that applicants use school choice to apply to higher-performing schools. Additionally, in terms of the schools that applicants apply to, the racial disparities are reduced, especially for high-performing applicants. Furthermore, high-performing applicants are more likely to apply more aggressively to high-performing schools. The pattern could be explained by differences in preferences, in awareness, or in assessments about admission chances. Figure A.1 document patterns of racial disparities in other school characteristics in terms of the schools that students live close to, apply to, and are assigned to.

4 Model of Students' Application Behavior

In this section, we lay out the model of school applications. Students are modeled as expected utility maximizers who are subject to two types of optimization frictions. First, they may consider only a limited set of school options. Second, they may have incorrect beliefs about the equilibrium assignment probabilities. In particular, these incorrect beliefs may reflect students' misunderstandings of the properties of the DA mechanism.[24]

A school is defined to be *considered* by an applicant if (1) he is aware of that school, and (2) he believes he has a positive chance of assignment to that school upon listing it.[25] The *consideration set* of applicant i, which is the set of schools considered by applicant i, is denoted by $\mathcal{C}_i$. Consideration of school j by applicant i is determined by a latent variable $c_{ij} \in (-\infty, \infty]$.[26] A school is considered if and only if $c_{ij} > 0$.

Formally, each applicant i solves

$$\max_{r \in \mathcal{R}(\mathcal{C}_i)} \sum_{j=0}^{J} p_{ij}^r v_{ij} \tag{4.1}$$

where r denotes the report, $j \in \{1, \cdots, J\} \equiv \mathcal{J}$ denotes a school that is matched through the application procedure, $j = 0$ denotes the outside option,[27] $p_{ij}^r \in [0, 1]$ denotes i's subjective belief about the probability of being assigned to j upon submitting report r, and v_{ij} is the utility that i derives from being assigned to j. An agent with a consideration set $\mathcal{C}_i$ chooses a report from $\mathcal{R}(\mathcal{C}_i)$, which denotes the set of all the ordered lists of schools in $\mathcal{C}_i$ with length at most 12, including an empty list denoted by $r = \emptyset$.[28] The empty list represents

[24] While truthtelling behavior is subsumed in the model as a special case, it is not assumed a priori.

[25] This definition differs from the typical definition of consideration in the discrete choice literature in that we also impose (2) in addition to (1). However, the imposition of (2) is natural in this setting where assignment is stochastic at the time of reporting. Furthermore, (the lack of) consideration may be interpreted to additionally capture some factors other than awareness and degenerate admission chances: fear of rejection, risk aversion, or the psychological cost of writing. In other words, the model of consideration intends to capture any reason other than preferences that might prevent a student from listing a school. We stress awareness and degenerate assignment probabilities, as evidence suggests these channels are significant.

[26] In Section 7, we will assume that there are certain schools that are surely considered by an applicant; such a school is denoted by $c_{ij} = \infty$ for notational convenience.

[27] The outside option is interpreted as the inclusive value of remaining unassigned in the main round of the application process.

[28] Formally, $\mathcal{R}(\mathcal{C}_i) \equiv \{\emptyset\} \cup \bigcup_{k=1}^{12} \left\{(j_1, \ldots, j_k) \in \mathcal{C}_i^k \middle| \, j_{l_1} \neq j_{l_2} \text{ for all } (l_1, l_2) \in \{1, \cdots, k\}^2 \text{ with } l_1 \neq l_2\right\}$. Note also that the students actually rank school programs rather than schools. However, for most of the observed characteristics, we do not observe them separately by programs—they are aggregated at the school level. Therefore, we simplify the analysis by modeling students as applying to schools as a whole rather than to programs.

nonparticipation in the first round of the application process. Although r is an ordered list, we occasionally abuse notations and treat r as if it were an unordered set. The solution to the maximization problem is denoted by r_i.

We model beliefs about assignment probabilities following Kapor et al. (2020), which is motivated by the cutoff and score representation of the matching algorithms (Agarwal and Somaini, 2018; Azevedo and Leshno, 2016). Under this representation, each student is assigned a score_{ij} for each school, which is determined as a function of the admission priority groups, admission lotteries, or the rank of the student based on schools' screening policies. One important aspect of DA is that the score_{ij} can never be a function of the student's submitted ranking of the school. Each school has a student-type-specific cutoff, i.e., $\mathsf{cutoff}_j(\mathsf{type}) =: \mathsf{cutoff}_{ij}$. In NYC, the type indicates whether the student has disabilities.[29] The representation states that each student is matched to her first school in the list for which her score_{ij} falls below cutoff_{ij}. That is,

$$i \text{ is matched to } j$$
$$\Leftrightarrow$$
$$j \text{ is the earliest-ranked school in } r_i \text{ for which } \mathsf{cutoff}_{ij} - \mathsf{score}_{ij} > 0.$$

We model beliefs about the assignment probabilities using this representation. Each student forms subjective assessments of his $\mathsf{cutoff}_{ij} - \mathsf{score}_{ij}$ for each school j. For student i, his assessment of $\mathsf{cutoff}_{ij} - \mathsf{score}_{ij} =: \mathsf{diff}_{ij}$ is represented by the student-specific random variable $\widetilde{\mathsf{diff}}_{ij}(k) := \widetilde{\mathsf{cutoff}}_{ij} - \widetilde{\mathsf{score}}_{ij}(k)$, where k denotes the rank of j in the report; the randomness represents the student's perceived uncertainty about the scores and the cutoffs. Note that the distribution of $\widetilde{\mathsf{score}}_{ij}(k)$ can depend on the rank k; although the rank cannot affect the scores in DA, we allow that students may not understand this property.[30] However, we do assume that applicants are monotone in their misunderstanding; they understand that ranking a school later can never improve their scores. Formally, we assume $k < k'$ implies $\widetilde{\mathsf{score}}_{ij}(k) \leq \widetilde{\mathsf{score}}_{ij}(k')$ for all (i, j) in any realization.

Using the scores-and-cutoffs representation, we model subjective beliefs as follows: if j is

[29] For school programs that have the *educational option* admission method, the type also depends on the applicant's reading category as determined by the English Language Arts (ELA) score in the middle school.

[30] If students correctly understood that the rank cannot affect the scores, they would always truthfully rank the schools among the ones that are listed. On the other hand, even in such a case, it is still possible that there exists some unlisted school that is considered and preferred to a listed school if the list length constraint binds.

listed in report r, then

$$
\begin{aligned}
& p_{ij}^r \\
& = \mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij'}(k_{j'}^r) < 0 \text{ for all } j' \text{ listed before } j)\mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k_j^r) > 0) \\
& = \Pi_{l=1}^{k-1}(1 - q_{ij_{r_l}l})q_{ijk}
\end{aligned}
\tag{4.2}
$$

where k_j^r denotes the rank of j in report r, q_{ijk} denotes $\mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0)$, and j_{r_l} denotes the school listed at the lth spot in r. If j is not listed in report r, then $p_{ij}^r = 0$.

The next section addresses the identification of the model.

5 Identifying Preferences, Consideration, and Beliefs

This section lays out an intuitive overview of the identification strategy, where we demonstrate how three channels—preferences, consideration, and beliefs—can be separated out. These ideas are formalized in Appendix C.1, where we develop sufficient conditions for non-parametric identification.

We first demonstrate that there is variation in the data that is affected only by preferences and consideration, and not by beliefs. Observation 1 shows that beliefs do not affect (1) the number of schools in an applicant's list or (2) whether a school is written on an applicant's list, given that the applicant's list contains strictly fewer than 12 schools.

Observation 1 (Variation reflecting only preferences and consideration)**.** *Suppose applicant i's list r_i has strictly fewer than 12 schools. Then, $j \in r_i$ if and only if both $c_{ij} > 0$ and $v_{ij} > 0$. Furthermore, r_i has strictly fewer than 12 schools if and only if $\{j \in \mathcal{J} | v_{ij} > 0, c_{ij} > 0\}$ has strictly fewer than 12 schools.*

The proof is given in Appendix C.4. Given that Observation 1 shows that there is data variation that is strictly affected by preferences and consideration, a natural question is whether there is also variation that can be used to disentangle preferences from consideration.

It is intuitive that such separation may be possible if (1) there were some schools that are "surely" considered by an applicant or if (2) there were shifters of consideration that were excluded from utilities. We define the *surely considered set* of applicant i as the set of schools that are surely considered by applicant i. It is denoted by $\mathcal{S}_i$, and $\mathcal{S}_i \subseteq \mathcal{C}_i$ with probability 1. The following observation, which follows as a corollary of Observation 1, is helpful in separating preferences and consideration using the surely considered sets.

Observation 2 (Variation almost only reflecting preferences)**.** *Suppose applicant i's list r_i has strictly fewer than 12 schools and that $j \in \mathcal{S}_i$. Then, $j \in r_i$ if and only if $v_{ij} > 0$.*

That is, if one focuses on the student-school pairs for which (1) the student does not exhaust all the slots in the report and (2) the student surely considers the school, then whether the school is on his list is solely determined by the preferences.

However, there is a problem in interpreting Observation 2 as a statement precisely about preferences; students who do not exhaust all the slots are not randomly selected. Rather, the selection is determined by both preferences and consideration, as discussed in Observation 1. Nonetheless, the selection problem is mild. First, if $(v_{i1}, \cdots, v_{iJ})$ are independent across j conditional on observables,[31] the selection problem vanishes through a more "careful" selection of the student-school pairs.[32] Second, while Observation 2 uses only the surely considered sets to disentangle preferences from consideration, we may also have some consideration shifters that are excluded from preferences. The conditions in Propositions C.1 and C.4 represent an idealized analogue of this situation,[33] and under such conditions the selection issue again vanishes. Third, if we were interested in the joint distribution of the utilities among only the surely considered schools, the selection problem almost vanishes in the sense that any point of the joint distribution function of the utilities can be bounded by an interval with a length of approximately .03 in expectation (see case (ii) of Proposition C.5).

Taken together, Observations 1 and 2 provide the basis for the separate identification of preferences and consideration. That is, intuitively, it may be possible to first identify preferences using Observation 2 and then identify consideration using Observation 1. Propositions C.1 and C.2 in Appendix C.1 formalize the intuition by providing sufficient conditions under which the distributions of preferences and consideration sets are nonparametrically identified.

To identify beliefs, we may use the remaining variation in the data. First, in Observations 1 and 2, we did not utilize the information in how the applicants ordered the schools; we used only the information of whether schools were listed. Second, we have not yet utilized the variation in the portfolio choices of applicants who had more than 12 considered schools that they preferred to the outside option. These aspects of data variation are affected by beliefs

[31] This is the case in the current version of empirical specification. Such a specification rules out, for example, the usage of a random coefficient model. However, given a wide set of observed student-level variables, such a specification may not be too restrictive. Pathak and Shi (2020) finds little gains in performance by allowing for random coefficients, given the allowed heterogeneity in coefficients along the observed students' characteristics.

[32] This is done by selecting (i, j) pairs such that $|r_i \backslash \{j\}| < 11$ where $|r_i \backslash \{j\}|$ denotes the number of schools in report r_i after excluding j from the report if it was listed. For more discussion, see Section 7.1.

[33] Proposition C.4 further assumes the presence of an observable utility shifter that are excluded from consideration.

in addition to preferences and consideration.

Observation 3 (Variation reflecting beliefs)**.**

(i) Suppose that the applicant has more than 12 schools that are acceptable and considered. Then, the identities of the schools in r_i are determined as a function of $\left(v_{ij}, c_{ij}, (p^r_{ij})_{r\in\mathcal{R}(\mathcal{J})}\right)_{j\in\mathcal{J}}$. In particular, the function is not constant in $(p^r_{ij})_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$.

(ii) Suppose that r_i contains at least two schools. Then, r_i is determined as a function of $\left(v_{ij}, c_{ij}, (p^r_{ij})_{r\in\mathcal{R}(\mathcal{J})}\right)_{j\in\mathcal{J}}$. In particular, the function is not constant in $(p^r_{ij})_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$.[34]

In a restricted setting, Proposition C.3 shows the conditions under which beliefs are nonparametrically identified. Because the nonparametric identifiability of beliefs is established in a substantially more restricted setting than that of preferences or consideration, in Appendix C.5, we discuss identifiability of beliefs with the parametric assumptions laid out in the next section.

6 Empirical Specification

Student Preferences We specify the utility v_{ij} for the empirical analysis as follows:

$$v_{ij} = x^v_j\beta + \xi^v_j + z^v_{ij}\alpha + \epsilon^v_{ij},$$

where $x^v_j = (x^v_{j1}, \cdots, x^v_{jK_x})$ denotes a vector of the observed school characteristics, ξ^v_j denotes the unobserved common taste shifter for j, and $z^v_{ij} = (z^v_{ij1}, \cdots, z^v_{ijK_z})$ denotes a vector of the observable variables that vary across i, possibly including interaction terms between $i-$ and $j-$level characteristics, (i,j)-specific terms, and also i-specific terms. The idiosyncratic taste shock is represented by $\epsilon^v_{ij} \sim_{i.i.d} N(0,1)$, and we assume that it is independent of $(x^v_j, \xi^v_j, z^v_{ij})$.[35] We also assume that ξ^v_j is independent of x^v_j.[36] The utilities v_{ij} are normalized in terms of both scale and location. The scale is normalized by setting the standard deviation of ϵ_{ij} equal to 1. The location is normalized by setting the value of the outside option to zero,

[34] From the construction of the maximization problem in Equation 4.1, report r_i and the identities in the report is a function of $(p^r_{ij})_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$. To see examples of nonconstancies of the functions with respect to $(p^r_{ij})_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$ under a simplified setting, see the cases in Proposition C.3 and the corresponding proof.

[35] The nonparametric identification results strongly suggest that we can allow richer forms of unobserved heterogeneity—for example, allowing for random coefficients, or allowing ϵ^v_{ij} to be correlated across j's. The current specification is done as a first pass.

[36] In relationship with this assumption, the paper does not currently interpret β as "causal." The assumption may be relaxed with instruments for potentially endogenous x_j's.

i.e., $v_{i0} = 0$. Thus, v_{ij} is interpreted as the utility of j relative to 0. As we allow $i-$specific terms in z_{ij}, the value of the outside option relative to all the inside options can vary across these student-level observables.

The observed school characteristics include average attendance and graduation rates, average math achievement in middle school, ethnic composition, and ninth grade enrollment. The observed student characteristics include subsidized lunch status, ethnicity, middle school math score, English proficiency, neighborhood income, and special education status. The observed student-school specific characteristics include the distance from the student's home to the school, and an indicator for whether school j is applicant i's middle school.

Consideration We specify the latent variable c_{ij} as

$$c_{ij} = \begin{cases} x_j^c \beta^c + z_{ij}^c \alpha^c + \epsilon_{ij}^c & \text{if } j \notin \mathcal{S}_i \\ +\infty & \text{if } j \in \mathcal{S}_i \end{cases}$$

where $\mathcal{S}_i$ denotes the surely considered set for applicant i, $x_j^c = (x_{j1}^c, \cdots, x_{jK_x^c}^c)$ denotes a vector of observed school characteristics, $z_{ij}^c = (z_{ij1}^c, \cdots, z_{ijK_z^c}^c)$ denotes a vector of the observable variables that vary across i, possibly including interaction terms between $i-$ and $j-$level characteristics, (i, j)-specific terms, and i-specific terms. The idiosyncratic taste shock is represented by $\epsilon_{ij}^c \sim_{i.i.d} N(0, 1)$,[37] and we assume that it is independent of (x_j, z_{ij}^c).[38] The scale is normalized by setting the variance of ϵ_{ij}^c equal to 1.

The observables (x_j^c, z_{ij}^c) always contain all the observables that enter utility, i.e., (x_j^v, z_{ij}^v). This reflects the possibility that any observable that shifts utility may also shift consideration. On the other hand, there may be variables that only enter (x_j^c, z_{ij}^c) but not (x_j^v, z_{ij}^v). In my specifications, these variables reflect the order in which the school appears in the school directory within its borough, the high school's distance from the applicant's middle school, and the applicants' admission probabilities at the schools.

More specifically, the page rank variable records the order in which the school appears in the NYC High School Directory (ranked within its borough), which is the main reference for the application process and is more than 600 pages long. The schools are ordered alphabetically within their respective boroughs in this directory. Because applicants may overlook the schools that are listed later, the page rank may shift consideration. However, because the

[37] With the independence assumption, the model becomes an alternative-specific consideration model (Swait and Ben-Akiva, 1987). For more discussion, see Abaluck and Adams (2017).

[38] The nonparametric identification results suggest that richer forms of unobserved heterogeneity can be allowed.

schools are ordered alphabetically, it is argued that page rank is excluded from the preferences. Section 3.2 discussed how Table 3 is consistent with the hypothesis. Table A.1 further shows that page rank is largely uncorrelated with observable school characteristics.

We allow a school's distance from an applicant's middle school to affect consideration, as the applicant may be more aware of the schools that are close to her middle school. We assume that the distance can affect consideration only within a two-mile boundary. While it is plausible that a student may prefer the schools that are close to her middle school because she expects her peers from the middle school to attend these schools, one can control for the number of students enrolling from the applicant's middle school.

We also allow a proxy of admission probability of a student at the school to affect her consideration as it likely affects her assessment about whether she has a positive probability of admission to the school. The probabilities are calculated conditional on the applicant ranking the school first in their list. Table 3 supports the hypothesis that admission priorities, which are correlated with admission chances, affect consideration but not preferences conditional on observable school characteristics. We assume that the proxies of the probabilities enter the consideration equation linearly.

Following the definition of the consideration set, the surely considered set is the intersection of (1) the set of schools the applicant is surely aware of and (2) the set of schools that she surely believes she has a positive chance of admission to upon application. My main specification assumes that the applicant is surely aware of schools within a .75 mile from her home or a quarter mile from her middle school. Additionally, we assume that the applicant surely believes she has some chance of admission to any school that did not fill its seats in the previous year or has a program such that she is in the first priority group and all of the students in the first priority group were admitted in the previous year. This specification results in approximately five surely considered schools per applicant on average. Note also that the surely considered sets are entirely determined by observables.

Beliefs Beliefs about the probabilities of assignments to schools are derived from the beliefs about the actual cutoffs and scores. A student's anticipation regarding the actual $\mathsf{diff}_{ij} \equiv \mathsf{cutoff}_{ij} - \mathsf{score}_{ij}$ is represented by the random variable $\widetilde{\mathsf{diff}}_{ij}(k)$. We parametrize the

distribution of $\widetilde{\mathsf{diff}}_{ij}(k)$ by

$$\begin{aligned}\widetilde{\mathsf{diff}}_{ij}(k) &= \mathsf{cutoff}_{ij} - \mathsf{score}_{ij} + \epsilon^b_{ijk} \\ &\equiv \underbrace{\mathsf{cutoff}_{ij} - \mathsf{score}_{ij} + \beta^{\mathsf{rank}}(k-1) + \mu(x^b_j, z^b_i) + \eta_{ij}}_{:=\delta^{\mathsf{diff}}_{ijk}} + \nu_{ij}\end{aligned}$$

where ϵ^b_{ijk} denotes the error in assessment. The error is affected by the pessimism bias $\mu(x^b_j, z^b_i)$, the idiosyncratic bias heterogeneity $\eta_{ij} \sim N(0, \sigma^2_\eta)$, and student i's doubt in assessment ν_{ij}. From the perspective of the student, his subjective assessment $\widetilde{\mathsf{diff}}_{ijk}$ follows distribution $N(\delta^{\mathsf{diff}}_{ijk}, \sigma^2_\nu)$. In other words, $\delta^{\mathsf{diff}}_{ijk}$ is the mean anticipated difference between the cutoff and score for agent i, and σ_ν represents the agent's level of doubt about his assessment.

The specification captures a variety of relevant scenarios. When $\beta^{\mathsf{rank}} = 0$, subjectively optimal lists are truthfully ordered in terms of utilities among the listed schools. This represents the correct understanding of a core property of DA: the scores do not depend on the submitted ranks, and therefore untruthful ordering within the listed schools is a weakly dominated strategy (Fack et al., 2019). However, this does not imply that students' reports are necessarily truthful (in a strong sense), as some unranked schools may be preferred to some of the ranked schools, even among those within the applicant's consideration set. On the other hand, if $\beta^{\mathsf{rank}} \neq 0$, the submitted rankings may not be truthfully ordered in terms of utilities even among the listed schools. When $\sigma^2_\nu = \infty$, the subjectively optimal lists are always truthful: students rank the schools truthfully among the considered schools that are preferred to the outside option until they exhaust all the 12 slots. This implies that ranked schools are always preferred to any considered but unranked school. Such completely truthful behavior may be objectively suboptimal when the list truncation binds. The model can approximate equilibrium beliefs when $\beta^{\mathsf{rank}} = 0$ and the distribution of $\mu(x^b_j, z^b_i) + \eta_{ij} + \nu_{ij}$ approximates the randomness in the actual (pre-realization) $\mathsf{cutoff}_{ij} - \mathsf{score}_{ij}$. In NYC, a student's score for a school program may depend on the student's admission priority group for the program, the school program's evaluation of the student based on factors such as past performance, and admission lotteries. Only certain school programs are allowed to screen students based on past performance.

With the specifications, it follows that

$$q_{ijk} \equiv \mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0) = \mathbb{P}_i\Big(\frac{\nu_{ijk}}{\sigma_\nu} > \frac{-\delta^{\mathsf{diff}}_{ijk}}{\sigma_\nu}\Big) = 1 - \Phi\Big(\frac{-\delta^{\mathsf{diff}}_{ijk}}{\sigma_\nu}\Big)$$

and this relationship and Equation 4.2 are used to express p^r_{ij} as a function of the belief

parameters and an unobservable η_{ij}.

7 Estimating Preferences, Consideration, and Beliefs

Section 7.1 describes the estimation procedure. Section 7.2 provides the summary of estimation results.

7.1 Estimation

The main results are estimated with the generalized method of moments. Two types of moment conditions are used: the first type is derived from a partial likelihood, and the second type is simulated moments. The first type of moment conditions is the scores of the (partial) likelihood of inclusion of school j in applicant i's report. The likelihood reflects the identifying information in Observations 1 and 2 or, more formally, that in Proposition and C.1 and C.2.[39] Accordingly, they give information about preferences and consideration but not about beliefs. In the likelihood, the sample consists only the (i, j) pairs such that $|r_i \backslash \{j\}| < 11$; with a slight abuse of notation, $|r_i \backslash \{j\}|$ denotes the number of schools in the report r_i after excluding j from the report if it was listed. That is, we select student-school pairs for which the $j-$excluded r_i has less than eleven schools. As $|r_i \backslash \{j\}| < 11$ implies $|r_i| < 12$, the condition selects applicants who have not exhausted the list, consistent with the statement in Observations 1 and 2. Selecting (i, j) pairs with $|r_i \backslash \{j\}| < 11$, rather than $|r_i| < 12$, resolves the potential selection problem discussed with regards to Observation 2; under the specification of the distribution of $(\epsilon_{ij})_{j \in \mathcal{J}}$ laid out in Section 6, $|r_i \backslash \{j\}| < 11$ is independent of ϵ_{ij} (Lemma D.1), so that the distribution of the unobservables are unaffected by such selection. The exact moment conditions are provided in Appendix D.1.

The second type of the moment conditions is derived from the students' ranking behavior and the identities of the schools in the full lists, which provide information about all the channels: preferences, consideration, and beliefs. These moment conditions either reflect the covariance of the observed characteristics with an indicator for a school being listed in the first top $k \in \{1, \cdots, 12\}$ slots by an applicant or with an indicator for a school being listed in the first top k slots by an applicant while another school not being included in these slots by the same applicant. These moment conditions use the identifying information in Observation

[39] We also weight (i, j) pairs for which i surely considers j, so that such pairs have a combined weight of 15% in the sample. In the current specification, such (i, j) pairs constitute only approximately 1% of the sample; we amplify their importance by weighting.

Table 5: Summary of Preference and Consideration

	Asian	Black	Hispanic	White
% Schools considered	14.2%	15.6%	13.2%	14.7%
% Schools considered outside the surely considered set	13.9%	14.8%	12.4%	14.3%
% Schools considered among schools preferred to outside option	33.0%	15.8%	12.7%	27.8%
% Students who exhaust 12 slots without any surely considered school	3.5%	3.9%	2.4%	0.5%
% Schools preferred to outside option	4.8%	10.6%	11.2%	3.9%
% Schools preferred to outside option among surely considered schools	14.5%	10.3%	11.2%	13.7%
% Schools preferred to outside option among considered schools	14.5%	10.9%	11.6%	8.2%
% Schools both considered and preferred to outside option	1.5%	1.7%	1.6%	1.2%

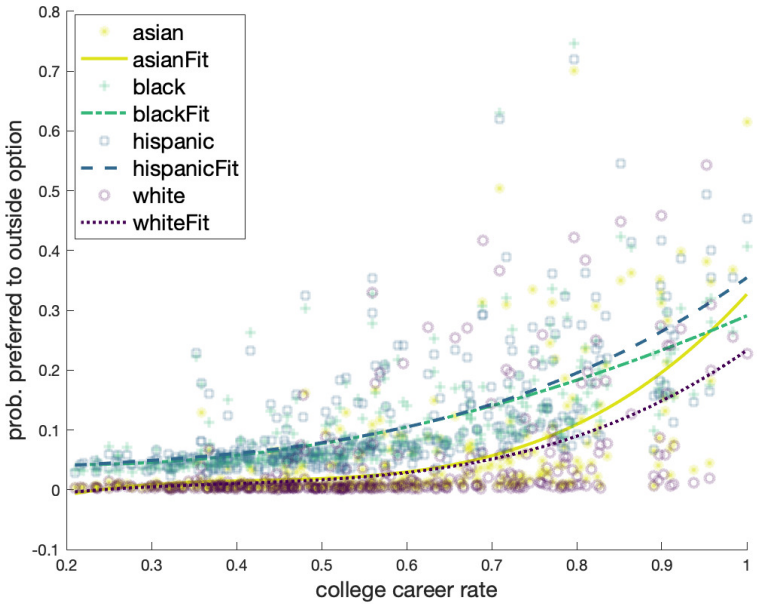
3 or, more formally, that in Proposition C.3. The exact moment conditions are provided in Appendix D.2. They are simulated moments, and we smooth these moments via importance sampling following Ackerberg (2009).

7.2 Estimates

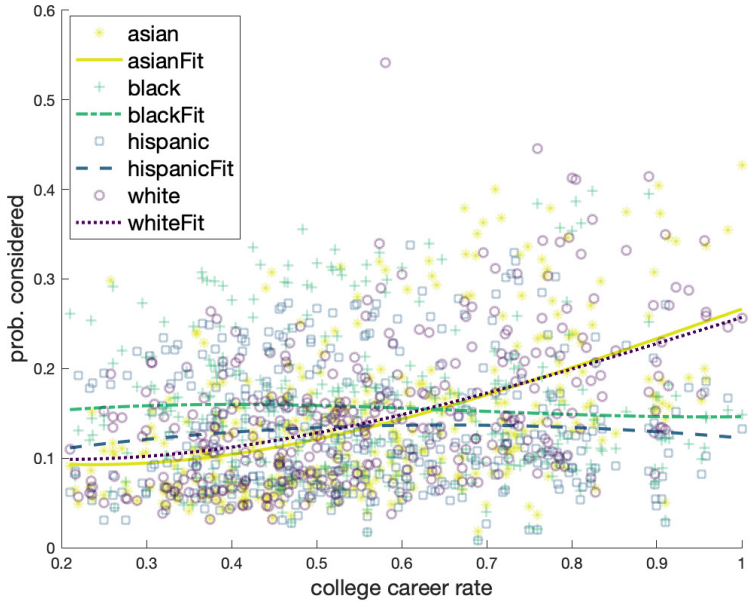
We present a summary of the key features of the estimated parameters in Table 5 (parameter estimates are presented in Tables A.3 and A.4). Regardless of ethnicity, the students are estimated to consider approximately 15% schools on average, which is about 65 schools in NYC. There is a notable difference across races in the proportion of schools considered among schools that are (counterfactually) preferred to the outside option—Asian and White students are much more likely to consider their preferred schools. This reflects both the fact that (1) White and Asian students tend to have fewer schools preferred to the outside option, potentially reflecting the fact that these students have better outside options, and (2) White and Asian students tend to live closer to higher-quality schools (see Figures 1 and A.1). Students who exhaust all the 12 slots almost always write some surely considered school on the list. Schools that are both considered and preferred to the outside option is around 1.5% across races.

Figure 2 further demonstrates a summary of the estimates. For each ethnicity, a point in

Figure 2: Performance of Preferred and Considered Schools by Ethnicity



(a) Probability of Being Preferred to the Outside Option



(b) Probability of Being Considered

Notes: For each ethnicity, each point in the scatter plot denote a school. Each line represents a cubic polynomial fit. *College/career rate* indicates the school's proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation.

the scatter plots represents a school in NYC. The lines represent cubic polynomial fits. In Figure 2a, compared to Asian and White students, Black and Hispanic students are more likely to prefer low-performing high schools to their outside options. This can be explained by Black and Hispanic students tending to have worse outside options and to live closer to lower-performing schools (see Figures 1 and A.1). On the other hand, the probability of high-performing schools being preferred to the outside options is similar across the races. Figure 2b shows that Asian and White students are more likely to consider high-performing schools, partly because they live closer to such schools; distance is an important determinant of consideration probabilities (see Table A.4). Later versions of the draft will include discussions about estimates of beliefs about admission chances.

8 The Impacts of Centralized School Choice in Practice

Section 8.1 presents the impact of NYC's high school choice on racial integration and the proportion of students matched to their preferred schools. It also decomposes the overall impact into the contributions of different factors. Section 8.2 quantifies the prevalence of justified envy. It also presents the steps necessary to quantify report optimality and truthfulness.

8.1 Distributional Outcomes and Decomposition

This section analyzes the impact of school choice on the distributional outcomes. The two distributional outcomes that we focus on are (1) racial integration and (2) the proportion of students matched to their top five preferred schools by each race. We further decompose the overall impact of school choice into the impacts of different factors—students' preferences, limited consideration sets, strategic mistakes, admission priority groups, and screening of the students by the schools.

To do this, we consider the actual and the counterfactual matchings as in Table 6. There are two counterfactual matchings without school choice: random matching and neighborhood matching. *Random* matching randomly allocates the students to the schools, respecting the capacity constraints of the schools. *Neighborhood* matching minimizes the total distance traveled by the students to the schools while respecting the capacity constraints of the schools.[40]

[40]Since some schools are dropped from the dataset due to missing data, and since we match every student in the dataset (including those who remain unmatched in the data) in random matching and neighborhood matching, schools do not have enough capacities to fit in every student. We proportionally expand the school capacities to have just enough seats for the students.

Table 6: Definition of the Matchings

A. Matchings without school choice

Matching	**Matching method**
Random	Random allocation of students to the schools
Neighborhood	Minimize total distance traveled by the students to the schools

B. Matchings with school choice

Matching	**Simulated?**	**Strategizing**	**Consideration Sets**	**Admission rankings**	**Preferences**
Actual	No	Actual	Actual	Actual	Actual
Estimated + Truthtelling	Yes	**Truthtell**	Estimated	Actual adm. priorities + Estimated screening	Estimated
Add Full Consideration	Yes	Truthtell	**All schools**	Actual adm. priorities + Estimated screening	Estimated
Add Random Screening	Yes	Truthtell	All schools	Actual adm. priorities + **Random screening**	Estimated
Add Random Admission Priorities (Preferences-Only Choice)	Yes	Truthtell	All schools	**Random adm. priorities** + Random screening	Estimated

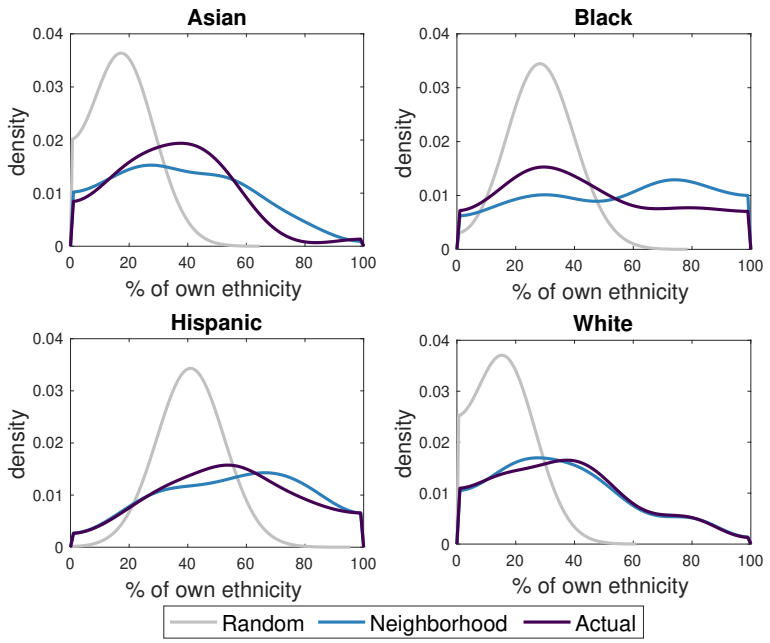
There are five matchings based on school choice: one is the actual matching, and the four others represent counterfactual matchings simulated under different scenarios. *Actual* matching is the actual school choice matching in 2017 from the main round. *Estimated + Truthtelling* matching represents the results from a simulated DA where the students have estimated preferences and consideration but are not strategic: students truthfully report their considered schools in the order of their preferences until they run out of the schools that are preferred to the outside option or reach the 12-school threshold. In this matching, schools rank students based on the actual admission priority groups and the estimated screening policies.[41] In the last three matchings, we shut off each factor's influence one by one and simulate the DA assignments. Compared to the Estimated + Truthtelling matching, *Add Full Consideration* matching turns off limited consideration by assuming that students consider every school. *Add Random Screening* matching turns off the schools' screening policies by forcing the screening schools to randomly screen students. *Add Random Admission Priority* matching turns off

[41] This matching is the current version of the most fully-fledged model; it remains to simulate the outcomes under the estimated beliefs. Actual matching and the Estimated + Truthtelling matching are different in two ways: (1) they reflect the modeling and estimation errors, and (2) we impose that students are truthtelling in the latter matching.

the effect of the existing admission priority groups by randomly allocating the students to the priority groups. Note that this matching purely reflects preferences without the influences of limited consideration, strategic behavior, nor admission priorities. In this regard, an alternative name for the matching is *Preference-Only Choice*.

8.1.1 Racial Integration

Figure 3: Percent of Own Ethnicity by Matching, Model-Free

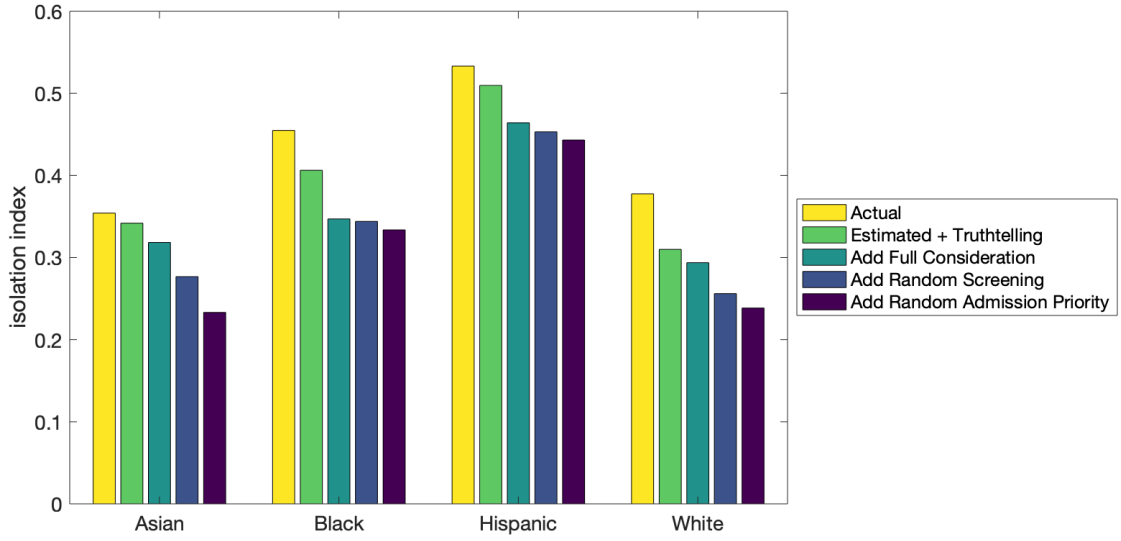


Notes: For each ethnicity and matching, the plot represents the kernel-smoothed density of the students with the same ethnicity in the students' assigned schools. See Table 6 and the discussions for the definitions of the matchings.

For each ethnicity, Figure 3 shows the density of the students matched to schools with different proportions of the students from their ethnicity. The figure shows that the main round of school choice tends to integrate ethnicities compared to the counterfactual neighborhood allocation; school choice tends to reduce the cases of students attending schools where their peers are mostly of their own ethnicity, with a possible exception for the Whites. However, the ethnicities are still substantially more isolated under the school choice allocation compared to the counterfactual case of random matching, which represents the scenario of full integration.

We then analyze which factors contribute to or hamper racial integration and by how much. Figure 4 shows the isolation index of each ethnicity by matching. For each ethnicity,

Figure 4: Isolation Indices by Matching



Notes: Each bar represents isolation index of an ethnic group in a matching. See Table 6 and the discussions for the definitions of the matchings.

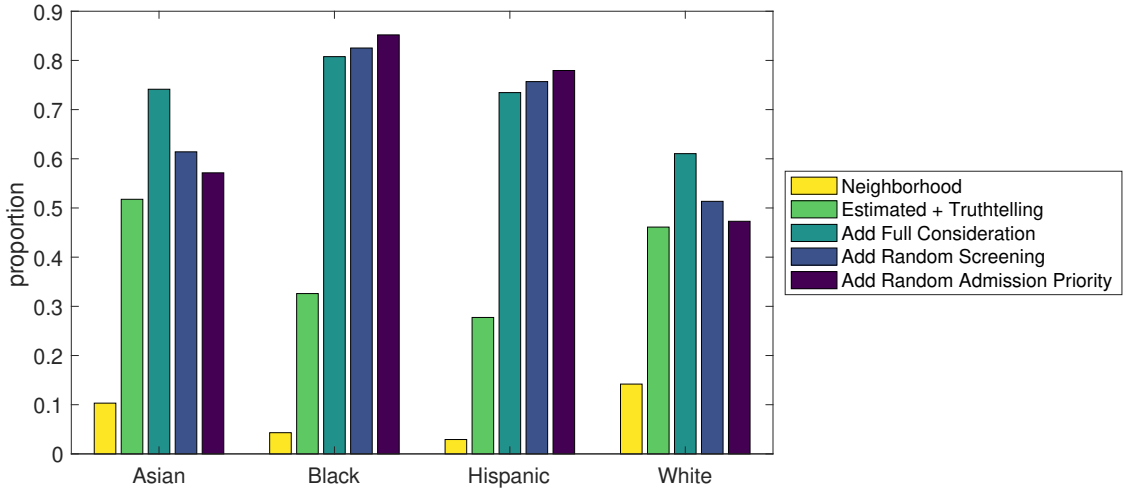
the isolation index denotes the average percentage of the students of the same ethnicity in a school; in other words, it is the mean of the corresponding distribution represented in Figure 3. Therefore, higher isolation indices of ethnicities indicate higher degrees of segregation. We observe that limitations in consideration tend to segregate; assuming full consideration, each ethnicity's isolation index decreases. Furthermore, we observe that the schools' ranking policies tend to segregate. Replacing the estimated screening policies by random screening, the isolation indices decrease; replacing the actual admission priorities with random priorities has a similar effect.

8.1.2 Welfare and Equity

Each bar in Figure 5 depicts the fraction of students who are matched to their top five preferred schools in terms of their utilities, for each ethnicity and matching. The sample includes both the schools that are considered and not considered. Viewing the Estimated + Truthtelling matching as an approximation to the current school choice matching,[42] we see that school choice tends to increase the proportion of students matched to their top five preferred schools

[42]Such approximation is only justified when truthtelling approximates the students' reporting strategy. A better approximation would be to use the assignments simulated using the estimated beliefs. This remains to be done.

Figure 5: Proportion Matched to Top Five Preferred Schools



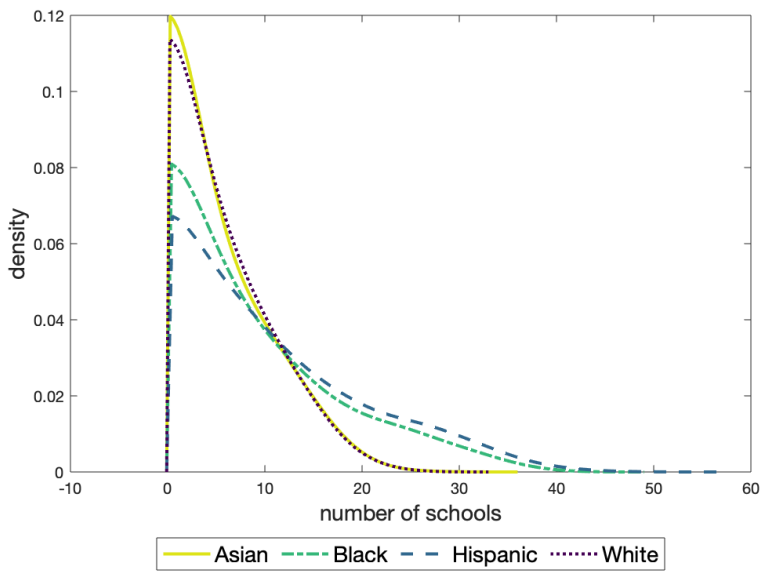
Notes: Each bar represents the fraction of the students matched to their top five preferred schools. The sample includes both the schools that are considered and those that are not. See Table 6 and the discussions for the definitions of the matchings.

compared to the neighborhood matching. The improvement is large: it increases such proportion from about 6.1% to 35.1% on average. We see that Asian and White students are more likely to be matched to their preferred schools compared to Black and Hispanic students in a neighborhood allocation, reflecting the disparities in the neighborhood schools' characteristics across races, as seen in Figures 1 and A.1. We also see that limited consideration substantially suppresses the proportion of students matched to their preferred schools, but the effect is larger for the Hispanic and Black students. The latter can be explained by the fact that Asian and White students tend to consider high-performing schools, as seen in Figures 2b and A.2. The results suggest that schools' screening policies tend to match Asian and White students to their preferred schools. This potentially reflects the fact that Asian and White students tend to have better performance in middle school (see, e.g., Table 1) so that they may be more likely to have higher admission scores for the schools that can screen students. We also see that Asian and White students tend to benefit from schools' admission priorities. This may reflect that a large proportion of the admission priorities are based on geographic proximity. Since Asian and White students live closer to higher-performing schools, they tend to be prioritized for admissions in these schools.

8.2 Empirical Assessments of the Theory-Predicted Outcomes

8.2.1 Matching Stability and Justified Envy

Figure 6: Number of Schools Viewed With Justified Envy



Notes: Each line represents the kernel density of the number of schools viewed with justified envy for students from each ethnic group.

To quantify matching stability, we count the cases of justified envy; a stable matching does not have any case of justified envy. A student views a school with *justified envy* if the student and the school are not matched to each other, but both would prefer to be matched to one another than to (one of) the current assignment(s).[43] In this definition, we interpret schools' coarse admission priorities over the students as their preferences. Unlike admission lottery or screening rankings, information on the admission priorities is available to the applicants at the time of application. Using the simulation results from Estimated + Truthtelling model, we show that the students have significant amounts of justified envy; on average, they view about 8.7 schools with justified envy.[44] There are also racial disparities. Figure 6 depicts the number of schools viewed with justified envy by ethnicity. We see that Black and Hispanic students tend to have more schools viewed with justified envy compared to Asian and White

[43]The definition is in line with the usage of the term in the literature.
[44]The interpretation implicitly assumes that students are truthtelling.

students;[45] on average, Black and Hispanic students view approximately 10.4 schools with justified envy while Asian and White students view 5.5 schools with such envy. This reflects the fact that Asian and White students tend to be matched to their preferred schools, as in Figure 5.

8.2.2 Reports' Optimality and Truthfulness

The following procedure enables quantification of the proportion of the reports that are truthful or optimal. One first simulates the reports based on the students' estimated preferences, consideration, and beliefs. Then, based on simulated preference and consideration, one can calculate the fraction of the reports that are truthfully ordered among the considered schools. Similarly, based on simulated preference and the equilibrium admission probabilities, one can quantify what proportion of the reports are optimal in terms of equilibrium admission probabilities. It is also possible to quantify these outcomes separately for different demographic groups.

9 Conclusion and Future Directions

In this paper, we use data on school applications and admissions from the NYC DOE to examine the impacts of its centralized public high school choice procedure in 2016-2017. We first analyze its impact on distributive outcomes. The results show that, compared to neighborhood allocation, school choice slightly improves racial integration and significantly increases the proportion of students matched to their preferred schools across all races. We further quantify the contributions of different factors. We find that admission priorities and screening policies tend to segregate races. They make it more likely for the Asian and White students to be matched to their preferred schools. We also find that limitations in consideration tend to segregate races and decrease welfare for all races. Furthermore, we estimate the prevalence of justified envy. We find that Black and Hispanic students are more likely to have justified envy than Asian and White students.

Viewed broadly, my findings provide support for the NYC DOE's recent policy initiatives.[46] Some NYC DOE schools have adopted a pilot policy to offer admission priority to students of lower socioeconomic status.[47] The NYC DOE also replaced the physical high

[45] More sensitivity checks are needed to confirm the robustness of this finding.

[46] For details, see https://www.schools.nyc.gov/docs/default-source/default-document-library/diversity-in-new-york-city-public-schools-english.

[47] Some schools are participating in an initiative to give admission priority to applicants who

school directory with an online version in an attempt to make it easier for the applicants to navigate through the schools and to provide more accurate and updated information.[48]

Methodologically, we develop and estimate a model of student application behavior that allows for two types of optimization frictions: applicants may consider only a limited set of school options and may have incorrect beliefs about admission chances. We provide sufficient conditions ensuring that the model is nonparametrically identified using the type of rank-ordered choice data typically available from centralized school choice systems.

The limitations of the current paper indicate future research directions. First, this paper currently models the student as taking the school characteristics as given from the year before the applications. Therefore, the counterfactual results presented here are best understood as short-run impacts. We plan to evaluate the long term impacts by allowing the school characteristics to change endogenously as new students are assigned to schools and by estimating production functions that map the input student characteristics to educational outcomes. Second, this paper treats the supply of schools as given. However, the current variety of highly differentiated schools in NYC likely depends on the employment of a large-scale school choice program. Therefore, an interesting research agenda would be to examine the response of the supply of schools due to the presence of school choice and its implications for welfare and the distributional outcomes.

are eligible for subsidized lunch, applicants in temporary housing, and English Language Learners. For more details, see https://www.schools.nyc.gov/enrollment/enrollment-help/meeting-student-needs/diversity-in-admissions.

[48]Further analysis is necessary to determine whether the results were as intended. For discussion, see https://www.thecity.nyc/education/2019/5/22/21211050/city-public-high-school-directory-takes-virtual-turn.

References

Abaluck, J. and Adams, A. (2017). What Do Consumers Consider Before They Choose? Identification From Asymmetric Demand Responses. Technical report, National Bureau of Economic Research.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The Welfare Effects of Coordinated Assignment: Evidence From the New York City High School Match. *American Economic Review*, 107(12):3635–89.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match. *American Economic Review*, 107(12):3635–3689.

Abdulkadiroglu, A., Che, Y.-K., Pathak, P. A., Roth, A. E., and Tercieux, O. (2017). Minimizing Justified Envy in School Choice: The Design of New Orleans' OneApp. Technical report, National Bureau of Economic Research.

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., and Roth, A. E. (2009). Strategy-Proofness Versus Efficiency in Matching With Indifferences: Redesigning the NYC High School Match. *American Economic Review*, 99(5):1954–1978.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, 93(3):729–747.

Ackerberg, D. A. (2009). A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation. *Quantitative Marketing and Economics*, 7(4):343–376.

Agarwal, N. and Somaini, P. (2018). Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica*, 86(2):391–444.

Aguirregabiria, V. (2021). Identification of firms' beliefs in structural models of market competition. *Canadian Journal of Economics/Revue canadienne d'économique*, 54(1):5–33.

Ajayi, K. and Sidibé, M. (2017). An empirical analysis of school choice under uncertainty.

Ajayi, K. and Sidibe, M. (2020). School choice under imperfect information. *Economic Research Initiatives at Duke (ERID) Working Paper*, (294).

Ajayi, K. F., Friedman, W. H., and Lucas, A. M. (2017). The Importance of Information Targeting for School Choice. *American Economic Review*, 107(5):638–643.

Akbarpour, M., Kapor, A., Neilson, C., van Dijk, W., and Zimmerman, S. (2020). Centralized school choice with unequal outside options.

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.

Artemov, G., Che, Y.-K., and He, Y. (2017). Strategic 'Mistakes': Implications for Market Design Research. *Work. Pap., Melbourne Univ., Melbourne, Aust.*

Ashlagi, I. and Gonczarowski, Y. A. (2018). Stable Matching Mechanisms Are Not Obviously Strategy-Proof. *Journal of Economic Theory*, 177:405–425.

Avery, C. and Pathak, P. (2015). The Distributional Consequences of Public School Choice. Technical Report w21525, National Bureau of Economic Research, Cambridge, MA.

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

Basteck, C. and Mantovani, M. (2018). Cognitive Ability and Games of School Choice. *Games and Economic Behavior*, 109:156–183.

Berry, S., Gandhi, A., and Haile, P. (2013). Connected Substitutes and Invertibility of Demand. *Econometrica*, 81(5):2087–2111.

Berry, S. and Haile, P. (2020). Nonparametric Identification of Differentiated Products Demand Using Micro Data. Technical Report w27704, National Bureau of Economic Research, Cambridge, MA.

Bifulco, R. and Ladd, H. F. (2007). School Choice, Racial Segregation, and Test-Score Gaps: Evidence From North Carolina's Charter School Program. *Journal of Policy Analysis and Management*, 26(1):31–56.

Bresnahan, T. F. and Reiss, P. C. (1991). Empirical Models of Discrete Games. *Journal of Econometrics*, 48(1-2):57–81.

Calsamiglia, C., Fu, C., and Güell, M. (2020a). Structural Estimation of a Model of School Choices: The Boston Mechanism Versus Its Alternatives. *Journal of Political Economy*, 128(2):642–680.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained School Choice: An Experimental Study. *The American Economic Review*, 100(4):1860–1874.

Calsamiglia, C., Martinez Mora, F., and Miralles, A. (2020b). School Choice Design, Risk Aversion, and Cardinal Segregation.

Che, Y.-K. and Tercieux, O. (2019). Efficiency and Stability in Large Matching Markets. *Journal of Political Economy*, 127(5):2301–2342.

Chen, Y. and Sönmez, T. (2006). School Choice: An Experimental Study. *Journal of Economic Theory*, 127(1):202–231.

Conlon, C. T. and Mortimer, J. H. (2013). Demand Estimation Under Incomplete Product Availability. *American Economic Journal: Microeconomics*, 5(4):1–30.

Corcoran, S. P., Jennings, J. L., Cohodes, S. R., and Sattin-Bajaj, C. (2018). Leveling the Playing Field for High School Choice: Results From a Field Experiment of Informational Interventions. Technical report, National Bureau of Economic Research.

Dynarski, S., Libassi, C., Michelmore, K., and Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, 111(6):1721–56.

Epple, D. and Romano, R. E. (1998). Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects. *American Economic Review*, pages 33–62.

Ergin, H. I. (2002). Efficient Resource Allocation on the Basis of Priorities. *Econometrica*, 70(6):2489–2497.

Fack, G., Grenet, J., and He, Y. (2019). Beyond Truth-Telling: Preference Estimation With Centralized School Choice and College Admissions. *American Economic Review*, 109(4):1486–1529.

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

Gaynor, M., Propper, C., and Seiler, S. (2016). Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service. *American Economic Review*, 106(11):3521–57.

Goeree, M. S. (2008). Limited Information and Advertising in the U.S. Personal Computer Industry. *Econometrica*, 76(5):1017–1074.

Haeringer, G. and Klijn, F. (2009). Constrained School Choice. *Journal of Economic Theory*, 144(5):1921–1947.

Hassidim, A., Marciano, D., Romm, A., and Shorrer, R. I. (2017). The Mechanism Is Truthful, Why Aren't You? *American Economic Review*, 107(5):220–224.

Hastings, J. S. and Weinstein, J. M. (2008). Information, School Choice, and Academic Achievement: Evidence From Two Experiments. *The Quarterly Journal of Economics*, 123(4):1373–1414.

He, Y. (2017). Gaming the Boston School Choice Mechanism in Beijing. Technical report, Toulouse School of Economics (TSE).

Hom, M. O. (2018). School Choice, Segregation and Access to Quality Schools: Evidence From Arizona. Technical report, Yale University.

Hortaçsu, A., Madanizadeh, S. A., and Puller, S. L. (2017). Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market. *American Economic Journal: Economic Policy*, 9(4):192–226.

Hoxby, C. and Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, 12:014.

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.

Hwang, S. I. M. (2017). How Does Heterogeneity in Beliefs Affect Students in the Boston Mechanism. *Unpublished working paper*.

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous Beliefs and School Choice Mechanisms. *American Economic Review*, 110(5):1274–1315.

Kessel, D. and Olme, E. (2018). School Choice Priority Structures and School Segregation. *Unpublished working paper*.

Laverde, M. (2020). Unequal Assignments to Public Schools and the Limits of School Choice.

Lewbel, A. (2000). Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables. *Journal of Econometrics*, 97(1):145–177.

Li, S. (2017). Obviously Strategy-Proof Mechanisms. *American Economic Review*, 107(11):3257–3287.

Luflade, M. (2018). The Value of Information in Centralized School Choice Systems. *Duke University*.

Martin, G. J. and Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99.

Narita, Y. (2016). Match or Mismatch: Learning and Inertia in School Choice.

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. pages 1–62.

NYC DOE (2020a). https://www.schools.nyc.gov/about-us/vision-and-mission.

NYC DOE (2020b). https://www.schools.nyc.gov/about-us/vision-and-mission/diversity-in-our-schools.

NYC DOE (2020c). https://www.schools.nyc.gov/about-us/vision-and-mission/equity-and-excellence.

Oosterbeek, H., Sóvágó, S., and Klaauw, B. (2019). Why Are Schools Segregated? Evidence From the Secondary-School Match in Amsterdam.

Pathak, P. A. and Shi, P. (2020). How Well Do Structural Demand Models Work? Counterfactual Predictions in School Choice. *Journal of Econometrics*.

Pathak, P. A. and Sönmez, T. (2008). Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism. *American Economic Review*, 98(4):1636–1652.

Roth, A. E. and Sotomayor, M. (1992). Two-Sided Matching. *Handbook of Game Theory with Economic Applications*, 1:485–541.

Sartain, L. and Barrow, L. (2020). The pathway to enrolling in a high-performance high school: Understanding barriers to access.

Sattin-Bajaj, C. (2016). *Unaccompanied Minors: Immigrant Youth, School Choice, and the Pursuit of Equity*. Harvard Education Press. Publication Title: Harvard Education Press.

Sattin-Bajaj, C., Jennings, J. L., Corcoran, S. P., Baker-Smith, E. C., and Hailey, C. (2018). Surviving at the Street Level: How Counselors' Implementation of School Choice Policy Shapes Students' High School Destinations. *Sociology of Education*, 91(1):46–71.

Shapley, L. and Scarf, H. (1974). On Cores and Indivisibility. *Journal of Mathematical Economics*, 1(1):23–37.

Swait, J. and Ben-Akiva, M. (1987). Incorporating Random Constraints in Discrete Models of Choice Set Generation. *Transportation Research Part B: Methodological*, 21(2):91–102.

Thompson, T. S. (1989). Identification of Semiparametric Discrete Choice Models.

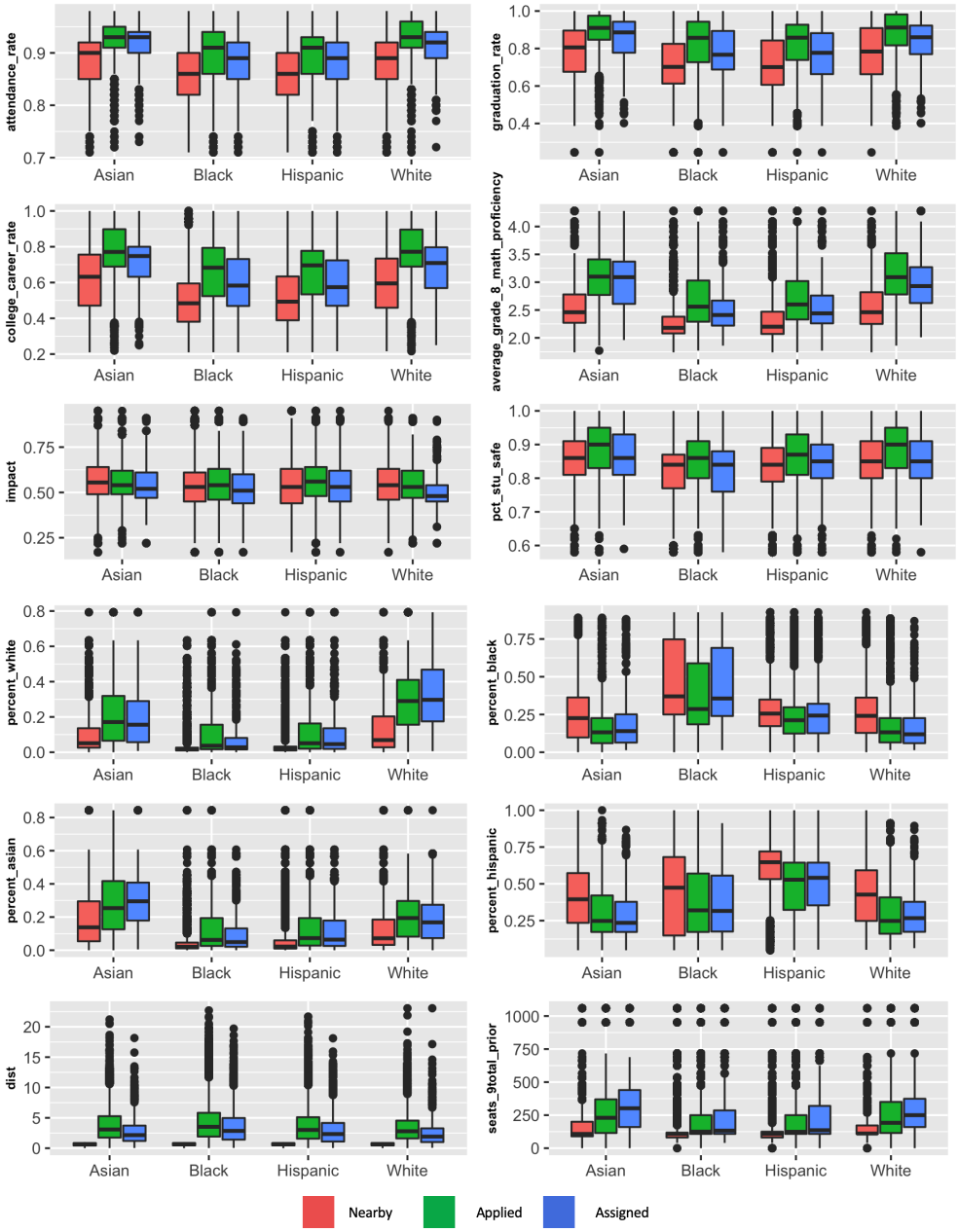
A Additional Tables and Figures

Table A.1: Regression of Page Rank on School Characteristics

	Dependent variable:
	Page rank
Constant	52.965
	(50.340)
Average grade 8 math proficiency (std.)	−5.129
	(3.738)
Graduation rate	41.384*
	(23.927)
Attendance rate	−54.318
	(60.453)
College/career rate	8.453
	(21.425)
Percent of students who feel safe	16.884
	(29.427)
9th grade seats	−0.007
	(0.015)
Percent Asian	−3.557
	(19.134)
Percent Black	1.555
	(8.762)
Percent White	−19.579
	(18.452)
Observations	352
R^2	0.037
F Statistic	1.456 (df = 9; 342)

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parentheses. Standardized values are indicated by (std.). *College/career rate* indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. Each school has equal weight regardless of class size. The sample excludes the nine specialized high schools and schools with missing data.

Figure A.1: Schools Nearby, Applied to, and Matched, by Ethnicity



Notes: Nearby schools are the schools within one mile from student's home. The applied and assigned schools are from the main round of applications. *pct_stu_safe* denotes the proportion of students who have reported that they feel safe in the school. *College_career_rate* indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation.

Table A.2: Regression of Submitted Rank on Page Rank and Priority Rank (Full Results)

	Dependent variable:										
	Rank in list										
	Listed 2	Listed 3	Listed 4	Listed 5	Listed 6	Listed 7	Listed 8	Listed 9	Listed 10	Listed 11	Listed 12
Page rank	−0.001	0.0004	0.0002	0.00001	0.001	0.001	0.003***	0.004***	0.003***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Priority rank	0.106**	0.089*	0.153***	0.161***	0.130***	0.232***	0.108**	0.273***	0.184***	0.269***	0.241***
	(0.047)	(0.048)	(0.045)	(0.042)	(0.040)	(0.044)	(0.045)	(0.054)	(0.058)	(0.058)	(0.033)
Observations	830	2,010	3,148	4,893	6,890	7,508	8,748	7,194	7,581	8,535	30,608
F Statistic	73.019***	110.040***	154.843***	211.750***	284.708***	302.162***	345.411***	284.378***	284.858***	323.827***	1,124.645***

Notes: *p<0.1; **p<0.05; ***p<0.01. An observation is an applicant-school pair. *Listed k* selects applicants who listed k schools in the lists. *Priority rank* denotes the rank of the admission priority group of the student for the school. Students of the first priority group are given the priority rank of 1.

Table A.3: Preference Parameter Estimates

	Asian		Black		Hispanic		White	
High is middle	2.50	(0.22)	2.07	(0.10)	2.39	(0.10)	2.23	(0.22)
Average grade 8 math (std.)	0.31	(0.18)	0.14	(0.03)	0.15	(0.04)	0.16	(0.07)
College/career rate	0.08	(0.47)	0.82	(0.11)	0.48	(0.14)	-0.15	(0.37)
9th grade seats (std.)	0.23	(0.03)	0.22	(0.02)	0.21	(0.03)	0.15	(0.02)
Distance to school	-0.08	(0.04)	-0.01	(0.01)	-0.03	(0.02)	-0.10	(0.02)
Proportion Asian	-0.51	(0.20)	-2.03	(0.13)	-1.58	(0.18)	-0.93	(0.21)
Proportion Black	-1.97	(0.41)	-1.76	(0.07)	-1.70	(0.16)	-2.06	(0.29)
Proportion Hispanic	-1.52	(0.37)	-1.68	(0.10)	-1.36	(0.10)	-1.61	(0.28)
Proportion White	-0.55	(0.18)	-1.46	(0.16)	-0.36	(0.24)	1.87	(0.36)
Standard deviation of ϵ_{ij}^{v}	1		1		1		1	
No. surely considered student-school pairs	2,030		7,693		11,505		1,966	
No. student-school pairs	524,051		826,826		1,286,758		508,981	

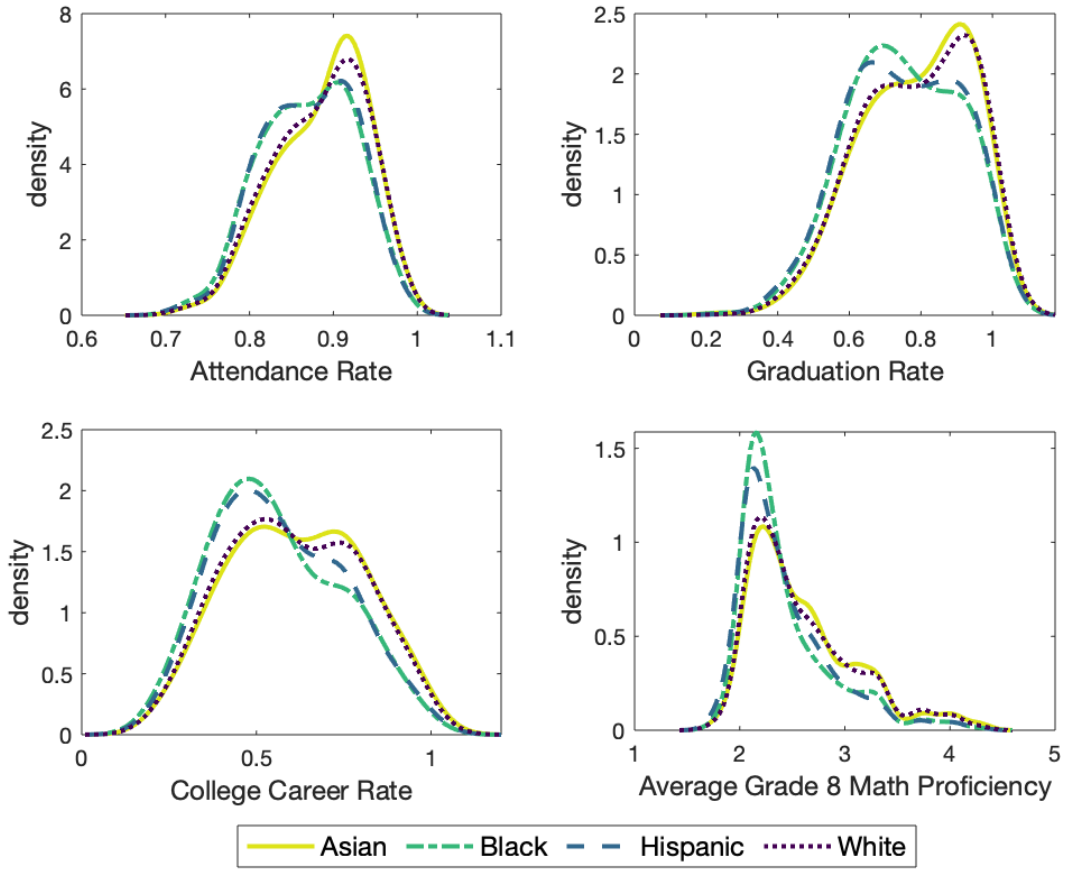
Notes: High is middle is an indicator variable reflecting that the middle school is the same school as the high school that the applicant is applying to. *College/career rate* indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. Standardized values are indicated by (std.).

Table A.4: Consideration Parameter Estimates

	Asian		Black		Hispanic		White	
Borough match	0.98	(0.31)	1.28	(0.07)	1.12	(0.09)	1.25	(0.19)
Close to middle school	3.23	(1.27)	7.42	(0.31)	2.98	(3.77)	1.12	(0.31)
Average grade 8 math (std.)	0.10	(0.24)	0.12	(0.04)	0.03	(0.04)	0.45	(0.14)
College/career rate	0.77	(0.75)	0.68	(0.17)	1.10	(0.19)	0.50	(0.74)
9th grade seats (std.)	0.02	(0.05)	-0.01	(0.02)	0.02	(0.03)	0.09	(0.03)
Distance to school	-0.15	(0.04)	-0.10	(0.01)	-0.10	(0.01)	-0.11	(0.02)
Page rank (std.)	0.00	(0.02)	-0.09	(0.01)	-0.07	(0.01)	0.01	(0.02)
Proxy probability of admission (std.)	-0.07	(0.04)	0.04	(0.01)	0.03	(0.01)	-0.07	(0.03)
Proportion Asian	-1.28	(0.46)	-2.02	(0.15)	-2.30	(0.16)	-3.24	(0.38)
Proportion Black	-1.03	(0.72)	-0.92	(0.12)	-1.99	(0.20)	-0.99	(0.64)
Proportion Hispanic	-0.54	(0.74)	-1.42	(0.15)	-1.16	(0.16)	-0.11	(0.68)
Proportion White	-1.45	(0.68)	-2.09	(0.14)	-2.55	(0.13)	-2.83	(0.36)
Standard deviation of ϵ_{ij}^{c}	1		1		1		1	
No. of surely considered student-school pairs	2,030		7,693		11,505		1,966	
Number of student-school pairs	524,051		826,826		1,286,758		508,981	

Notes: Borough match is an indicator variable reflecting that the student's home and the high school are located in the same borough. *College/career rate* indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation. Standardized values are indicated by (std.).

Figure A.2: Characteristics of Considered Schools by Ethnicity



Notes: College/career rate indicates the proportion of students who graduated from high school four years after they entered 9th grade and then enrolled in college, a vocational program, or a public service program within six months of graduation.

B Deferred Acceptance Mechanism in NYC

In the 2016–2017 school year, the DOE ran two rounds of DA assignments for the traditional (non-specialized) high schools and one round of such DA assignments for the nine specialized high schools. In my main analysis, we focus on the first round for the non-specialized high schools; approximately 70% of all the high school students in NYC attend one of these schools. This is the "main" round, in the sense that approximately 90% of the final matches are formed in this round.

Using the students' submitted rankings over the school programs and the programs' rankings over the students, the DA algorithm (Gale and Shapley (1962); Abdulkadiroğlu and Sönmez (2003)) matches the students to the school programs according to the following procedure.

- *Step* 1*:* Each applicant proposes to his first-ranked school program, if any. Each school program sorts the proposers according to its rankings and tentatively accepts all the highest-ranking proposers up to its capacity. It rejects any other proposers.

  Then for each $k \geq 2$,

- *Step k:* Each applicant who was not tentatively accepted by any program in Step $(k-1)$ proposes to his highest-ranked school program that has not previously rejected him, if any. Each school program sorts the new proposers and the applicants tentatively accepted previously according to its rankings and tentatively accepts all the highest-ranking applicants up to its capacity. All the other proposers are rejected.

The algorithm stops when there is no proposing student. Each student is assigned his final tentative assignment. In NYC high school match, the school programs have separate seats (capacities) for students with and without disabilities. Therefore, DA algorithms are run separately for the two student groups defined by their disabilities type.

C Identification: Details

C.1 Nonparametric Identification

In this section, we provide sufficient conditions for the nonparametric identification of the model. These results show that functional form assumptions are not necessary for identification and confirm the intuition in the stepwise-identification argument developed in Section

5. The main results are provided here, and Appendix C.2 provides additional results under stronger and weaker sets of assumptions. Taken together, these identification results clarify the role of each assumption and the source of data variation.

In stating the nonparametric identification results, we do not make any parametric assumption about utilities, latent consideration variables, and beliefs $(v_i, c_i, p_i) \equiv \big((v_{ij})_{j\in\mathcal{J}}, (c_{ij})_{j\in\mathcal{J}}, (p^r_{ij})_{r\in\mathcal{R}(\mathcal{J}),j\in\mathcal{J}}\big)$ as made in Section 6. Furthermore, we do not assume that the maximum allowed list length, denoted L, has to equal 12.

On the other hand, we do assume the following for every proposition. First, we assume that beliefs are generated by students making anticipations about differences in their scores and cutoffs, in the sense that Equation 4.2 holds. Second, we assume that perceived scores are increasing in rank as in Section 4. Third, we assume that the distribution of $v_i|z_i$ is continuous for every $z_i \in \text{supp}(z_i)$ and that $q_{ijk} \equiv \mathbb{P}_i(\widetilde{\mathsf{diff}}_{ij}(k) > 0) \in (0, 1)$ for every considered schools.

To discuss the results, we define two concepts: an extreme consideration shifter excluded from preferences and a special regressor with large support (Thompson, 1989; Lewbel, 2000).

Definition 1. Let $z_i \equiv (a_i, z_i^-)$. A $J-$dimensional random vector a_i is called an **extreme consideration shifter excluded from preferences** if $v_i \perp\!\!\!\perp a_i$ conditional on z_i^- and, for all z_i^- in its support, there exist some known $\bar{a}(z_i^-) \in \text{supp}(a_i|z_i^-)$ such that $\mathbb{P}\big(c_{ij} > 0|a_{ij} = \bar{a}_j(z_i^-)\big) = 1$.

In the empirical setting, the role of an extreme consideration shifter excluded from preferences is jointly played by surely considered sets[49] and the excluded consideration shifters, such as page rank and distance from middle school. However, they each play an imperfect role; surely considered sets only move certain schools' consideration probabilities for each student, and the excluded consideration shifters do not move consideration probabilities to 1, i.e., to the extreme.[50]

Definition 2. A random vector z_i^y is called a **special regressor for y_i with large support conditional on** x_i if $y_i = \tilde{y}_i - z_i^y$ with $\tilde{y}_i \perp\!\!\!\perp z_i^y$ conditional on x_i and $\text{supp}\left(z_i^y|x_i\right) = \mathbb{R}^K$ for all x_i in its support, where K is the dimension of y_i.

In the empirical setting, the role of a special regressor is played jointly[51] by any exogenous $(i, j)-$level observables, including distance to school, and the interactions between school

[49] The indicators for surely considered sets are determined as a function of only the observables; therefore, they are excluded from preferences conditional on observables.

[50] To complement the result, Proposition C.5 only assumes presence of surely considered sets.

[51] Results in Berry and Haile (2020) may be used to formally show how different variables can form an index that mimics the role of a special regressor. This is to be done in future work.

characteristics and the student-level observables.[52]

We first establish the nonparametric identifiability of preference. Proposition C.1 shows that the joint distribution of utilities is nonparametrically identified with a large-support special regressor for the utilities and an extreme consideration shifter.

Proposition C.1 (Identification of preferences)**.** *Suppose that we observe the following:*

(a) an extreme consideration shifter excluded from preferences, named a_i, and

(b) a special regressor for v_i, named z_i^v, with large support conditional on $z_i \backslash (z_i^v, a_i)$.

Then, the joint distribution of utilities conditional on observables, $\mathbb{P}\big(v_i \leq v \big| z_i\big)$, is identified for almost all $(v, z_i) \in \text{supp}(v_i, z_i)$.[53]

Intuitively, one can use the extreme consideration shifter to push the consideration probability of every school to 1, in which case the probability of listing schools becomes a sole function of the utilities. One can then use the special regressor to "trace out" the distribution of the utilities (Agarwal and Somaini, 2018). This distribution of the utilities is not conditioned on the value of the extreme consideration shifter, as it was assumed to be conditionally independent of the utilities. Note further that no assumption was made about allowed list length.

In my empirical model, the set of exogenous $(i, j)-$level observables that enter utilities, such as distance to school from home and the interactions between school characteristics and the student-level observables, play the role of the special regressor. Although they may not have large support in practice, it is not essential; with a special regressor with limited support, one can still obtain identification of the distribution of utilities on limited support. The role of an extreme consideration shifter excluded from preferences is jointly played (imperfectly) by surely considered sets and the excluded consideration shifters, such as page order and distance from middle school.

Now we turn to the identification of consideration. Proposition C.2 states that the distribution of consideration indicators $c_{ij}^* := \mathbb{1}(c_{ij} > 0)$ can be nonparametrically identified with a special regressor with large support, given that the distribution of utilities are already identified (potentially through Proposition C.1). It also assumes that the allowed list length L

[52]Note that most results—except case (ii) of Proposition C.4, which uses identification-at-infinity argument—can be extended to allow for limited support on the special regressor at the cost of identifying the distribution of the utilities or the latent variables for consideration on limited support.

[53]If the large support assumptions on the special regressors are weakened, then $\mathbb{P}\big(v_i \leq v \big| z_i\big)$ is also identified on a limited support.

equals the number of schools J, i.e., an applicant can list arbitrarily many schools. The joint distribution of consideration indicators is point-identified if the utilities v_i are independent of latent consideration variables c_i conditional on observables. It is partially identified if the conditional independence fails.

Proposition C.2 (Identification of consideration)**.** *Suppose that* $\mathbb{P}\big(v_i \leq v | z_i = z\big)$ *is identified for almost all* $(v, z) \in \text{supp}(v_i, z_i)$*. Suppose that we observe a special regressor for* c_i*, named* z_i^c*, with large support conditional on* $z_i \backslash z_i^c$*. Suppose also that* $L = J$*. Then,*

(i) if c_i *is independent of* v_i *conditional on* z_i*, the joint distribution of consideration indicators conditional on observables,* $\mathbb{P}\big(c_i^* \leq c^* \big| z_i\big)$*, is identified for almost all* $(c^*, z_i) \in \text{supp}(c_i^*, z_i)$*.*[54]

(ii) if c_i *is not independent of* v_i *conditional on* z_i*,* $\mathbb{P}\big((c_{ij}^*)_{j\in\mathcal{A}} \leq c^* \big| (v_{ij})_{j\in\mathcal{A}} > 0, z_i\big)$ *is identified for almost all* $(c^*, z_i) \in \text{supp}\left((c_{ij}^*)_{j\in\mathcal{A}}, z_i\right)$ *and for all* $\mathcal{A} \subseteq \mathcal{J}$*.*

Remark. In relation to Proposition C.1, it is allowed that $a_i = z_i^c$ or $z_i^c = z_i^v$.[55]

The intuition for part (i) is as follows. Given that an applicant can write an arbitrarily long list, whether to list a school is a function of only utilities and consideration.[56] However, knowing the distribution of the utilities already, the probability of schools being listed is informative only about consideration. The special regressor then traces out the distribution of c_i, the latent consideration variable. Since consideration indicator c_i^* is completely determined by c_i, the distribution of c_i^* is also traced out.

In my empirical model, the set of exogenous $(i, j)-$level observables that enter consideration equation, such as distance to high school from students' home or middle school and the interactions between school characteristics and the student-level observables, play the role of the special regressor. Again, the fact that they may not have large support in practice is not essential; a special regressor with limited support still enables the identification of the distribution of the latent consideration variable on limited support.

Now we turn to the identification of the beliefs about assignment probabilities. To present this result, we first define equivalent classes of beliefs. Two beliefs are behaviorally equivalent if they lead to the same reporting behavior conditional on any realization of the utilities and the consideration sets:

[54] If the large support assumptions on the special regressors are weakened, then $\mathbb{P}\big(c_i^* \leq c^* \big| z_i\big)$ are also identified on a limited support.

[55] On the other hand, it is never possible that $a_i = z_i^c = z_i^v$.

[56] Not allowing for truncated lists is a limitation of the result. Proposition C.4 present a result with potentially truncated lists with stronger data requirements.

Definition 3. Two beliefs $\{p_j^r\}_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$ and $\{p_j^{'r}\}_{j\in\mathcal{J},r\in\mathcal{R}(\mathcal{J})}$ are **behaviorally equivalent** if for all $v \in \mathbb{R}^J$ and $\mathcal{C}_i \subseteq \mathcal{J}$, $\arg\max_{r\in\mathcal{R}(\mathcal{C}_i)} v \cdot p^r = \arg\max_{r\in\mathcal{R}(\mathcal{C}_i)} v \cdot p^{'r}$.

where $(p^r) = (p_j^r)_{j\in\mathcal{J}}$ and similar for $(p^{'r})$. The notion of behavioral equivalence relates to the notion of normalization.

Here we state the identification result on beliefs, which holds under a restricted setting.

Proposition C.3 (Identification of beliefs)**.** *Suppose that* $\mathbb{P}\big(v_i \leq v, c_i^* \leq c^* | z_i = z\big)$ *is identified for every* $(v, c^*, z) \in \text{supp}(v_i, c_i^*, z_i)$*. Suppose that either (1)* $L = J = 2$*, or (2)* $L = 1$*. Suppose also that beliefs are constant given observables, i.e.* $p_{ij}^r = p_j^r(z_i)$ $\forall(i, j, r)$*. Then, beliefs* $\{p_j^r(z_i)\}_{j,r}$ *are identified up to behaviorally equivalent classes.*

C.2 Supplementary Propositions

Proposition C.4 (Identification of preferences and consideration with ideal data)**.** *Suppose that we observe* $z_i \equiv (z_i^v, z_i^c, z_i^-)$ *where* (z_i^v, z_i^c) *is a special regressor for* (v_i, c_i) *with large support conditional on* z_i^-*. Then,*

(i) if $L = J$*,* $\mathbb{P}\big(v_i \leq v, c_i^* \leq c^* | z_i = z\big)$ *is identified for every* $(v, c^*, z) \in \text{supp}(v_i, c_i^*, z_i)$*.*

(ii) if $L < J$*,* $\mathbb{P}\big(c_i^* \leq c^* | z_i = z\big)$ *is identified for every* $(c^*, z) \in \text{supp}(c_i^*, z_i)$ *and* $\mathbb{P}\big(v_i \leq v | z_i = z\big)$ *is identified for every* $(v, z) \in \text{supp}(v_i, z_i)$*.*

Proposition C.5 (Identification of preferences with surely considered sets)**.** *Suppose that we observe a special regressor for* v_i*, named* z_i^v*, with a large support conditional on* z_i^-*. Suppose also that* $\mathcal{S}_i \equiv \mathcal{S}(z_i)$ *is constant with respect to* z_i^v*. Then,*

(i) if $L = J$*,* $\mathbb{P}\big((v_{ij})_{j\in\mathcal{S}(z_i)} \leq v | z_i\big)$ *is identified for all* (v, z_i) *in its support.*

(ii) if $L < J$*,* $\mathbb{P}\big((v_{ij})_{j\in\mathcal{A}} \leq x | z\big)$ *is bounded within an interval of width* $\mathbb{P}(|r_i| = L, r_i \cap \mathcal{A} = \emptyset | z_i^v = x, z^-)$ *for all* $(x, z, \mathcal{A})$ *such that* $\mathcal{A} \subseteq \mathcal{S}(z)$ *with* $|\mathcal{A}| \leq L$*.*

C.3 Lemmas

These lemmas are used in the proofs of the observations and the propositions. We define that a school is *acceptable* if $v_{ij} > 0$ and *unacceptable* if $v_{ij} < 0$.

Lemma C.1. *Consider a list* r *that contains an unacceptable school before an acceptable school, and the lowest-ranked school is an acceptable school. Then, in any realization,* r *gives*

a weakly less payoff than an alternative list that switches the lowest-ranked school unacceptable school with the school that gives the maximum utility among the schools that follow this lowest-ranked unacceptable school.[57]

Lemma C.2 (Never write an unacceptable school)**.** *For any list r that contains a considered but unacceptable school, there is an alternative list that contains no unacceptable school and gives strictly higher expected utility.*

C.4 Proofs

Proof of Lemma C.1. By assumption, the list r has an unacceptable school before an acceptable school. Let j_- denote the lowest-ranked unacceptable school in the list. Then, by construction, (1) there are some schools that follow j_- and (2) these schools are all acceptable. Let the utility-maximum of these school be indicated by $j_{\max}$ (and there is always such a school). Then, the report r reads:

$$r = (\underbrace{\cdots}_{A}, j_-, \underbrace{\cdots}_{B}, j_{\max}, \underbrace{\cdots}_{C})$$

where A, B, and C denote the set of the schools in each respective position. Each of A, B, and C may or may not be empty.

Consider an alternative list r' that switches $j_{\max}$ with j_-, as in the statement:

$$r' = (\underbrace{\cdots}_{A}, j_{\max}, \underbrace{\cdots}_{B}, j_-, \underbrace{\cdots}_{C})$$

where the schools and the ordering within each A, B, and C is unaltered.

Representing an outcome in the relevant probability space by ω, we want to show that r' weakly dominates r for every ω, i.e., $v_{i\mu(i;r)}(\omega) \leq v_{i\mu(i;r')}(\omega)$ for all ω, where $\mu(i;r)$ is the assignment of i in the case that i reports r. To see this, suppose not: there is ω such that $v_{i\mu(i;r)}(r;\omega) > v_{i\mu(i;r')}(r';\omega)$. Then, it must be that the student get rejected at all the A schools under this ω regardless of submitting r or r', i.e.,

$$\tilde{\pi}_j(\omega) < \widetilde{\mathsf{score}}_{ij}(r(j);\omega) \equiv \widetilde{\mathsf{score}}_{ij}\big(r'(j);\omega\big) \ \forall j \in A$$

where $r(j)$ and $r'(j)$ denote the ranks of school j in r and r', respectively. This is because otherwise, he gets into the same school regardless of reporting r or r' and obtains the same

[57] The lemma is similar to what appears in the proof of Proposition 3 (ii) in He (2017).

utility. Note that it is impossible that he gets rejected in one report but not in the other report - his scores for any $j \in A$ under the two reports are exactly the same in the two reports as the submitted rank of any $j \in A$ in the two reports are the same. This is because score is restricted to depend only depends on certain aspects of the report - i.e., the rank.

Also, it must be that he gets rejected by j_- under r. Otherwise, conditioning on that the student is reject by all schools in A, this is the worst that can happen to him under r or r' because B and C can never have an unacceptable school by construction. Therefore, there is no way that j_- will strictly beat allocation under r'. Also, it must be that he gets rejected by $j_{\max}$ under r'; otherwise, this is the best that can happen to him under r or r' and so there is no way that allocation under r will strictly beat $j_{\max}$. Thus,

$$\tilde{\pi}_{j_-}(\omega) < \widetilde{\mathsf{score}}_{ij_-}\big(r(j_-);\omega\big)$$
$$\tilde{\pi}_{j_{\max}}(\omega) < \widetilde{\mathsf{score}}_{ij_{\max}}\big(r'(j_{\max});\omega\big)$$

Similarly, it must be that he fails to make the cutoffs (in either reports) by all schools in B. Otherwise, he gets same utility under the two reports. Note that he makes the cutoff in any of these schools in B by submitting r iff he does so in r'; the score for the school is the same under the two reports.

Further, it must be that he is rejected by $j_{\max}$ under r and j_- under r'. This follows from the assumption that perceived scores are monotonic in the submitted rank and the second step:

$$\tilde{\pi}_{j_-}(\omega) < \widetilde{\mathsf{score}}_{ij_-}\big(r(j_-);\omega\big) \leq \widetilde{\mathsf{score}}_{ij_-}\big(r'(j_-);\omega\big)$$
$$\tilde{\pi}_{j_{\max}}(\omega) < \widetilde{\mathsf{score}}_{ij_{\max}}\big(r'(j_{\max});\omega\big) \leq \widetilde{\mathsf{score}}_{ij_{\max}}\big(r(j_{\max});\omega\big)$$

By the same reasoning, it must be that he fails to make the cutoffs (in either reports) by all schools in C. Otherwise, he gets same utility under the two reports. Note that he makes the cutoffs in all of these schools in B by submitting r iff he does so in r'; the scores are the same under the two reports.

Then, they get rejected by all schools in either of the two reports, and is placed into outside option, in which they derive the same utility. This contradicts $v_{i\mu(i;r)}(r;\omega) > v_{i\mu(i;r')}(r';\omega)$ we started with. □

Proof of Lemma C.2. We first show that, for any r that contains an unacceptable school, there is an alternative list without any unacceptable school that gives weakly higher expected

utility.

Suppose that r has an unacceptable school at the very end. Then, it is straightforward to verify that dropping this school weakly increases expected utility. Repeat this process until the last school is an acceptable school. If the list is now composed of only the acceptable schools (or is empty), then such a list is an alternative list that we wanted to find.

If there are still some unacceptable schools in the list, then Lemma C.1 can be applied as there is some acceptable school after any unacceptable school. We further know that the new report found by the lemma must give weakly higher expected utility, as we've claimed that for any outcome ω, the new report must give utility weakly higher than the old report.

Apply the lemma to switch the lowest-ranked unacceptable school to a lower spot in the list. If this schools is now in the last spot, then drop this. If not, the schools that appear after this lowest-ranked unacceptable school are all acceptable, so that we can apply the lemma again. Continue to apply this lemma, this unacceptable the school gets moved to the last spot, in which case we can drop the unacceptable school and obtain even (weakly) higher expected utility.

If the resulting report is now filled with only acceptable schools (or is empty), we have found an alternative list that we wanted to find. If not, repeat the aforementioned process of moving the lowest-ranked unacceptable school down the list and then dropping it, until there is no unacceptable schools in the list. Every such process gives weakly higher expected utility, and therefore the resulting list gives weakly higher expected utility.

Note that the process above now has at least one occasion where an unacceptable but considered school is dropped from the last slot. By assumption, a student believes he has positive chance of matching to a considered school upon listing. Therefore, this drop strictly increases his expected utility.

□

Proof of Observation 1. Let L denote the maximum allowed length of the list. We show that the first statement holds.

To show that $j \in r_i$ implies both $j \in \mathcal{C}_i$ and $v_{ij} > 0$, we show the contrapositive. First, if $j \notin \mathcal{C}_i$, j cannot be on r_i by definition of consideration. Second, suppose that $v_{ij} < 0$ and $j \in \mathcal{C}_i$. By Lemma C.2, such a list with an unacceptable but considered school cannot be (subjectively) optimal.

Suppose now that $v_{ij} > 0$ and $j \in \mathcal{C}_i$, but $j \notin r_i$. Then one can strictly gain by adding j on the bottom of the list, which contradicts subjective optimality of r_i. The strict relation comes from $j \in \mathcal{C}_i$; a considered school has (subjectively) positive admission chance upon

listing. Addition of a school is possible since r_i has not exhausted all the available slots.

We now show that the second statement holds. The second statement is equivalent to the following statement: r_i has exactly L schools if and only if $\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}$ has L schools or more.

Suppose first that $|r_i| = L$ but $|\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}| < L$. Because all schools in r_i must be considered by definition, there must be some schools in r_i that is subjectively reachable but is unacceptable. By Lemma C.2, such a list cannot be subjectively optimal.

Suppose now that $|\{j \in \mathcal{J} | v_{ij} > 0, j \in \mathcal{C}_i\}| \geq L$ but $|r_i| < L$. Then, there must be some school $j \notin r_i$ such that $v_{ij} > 0$ and $j \in \mathcal{C}_i$. Adding j at the bottom of the list gives strictly higher payoff, contradicting that r_i is subjectively optimal.

□

Proof of Proposition C.1. I implicitly condition everything on $z_i \backslash (z_i^v, a_i)$. Take any $z^v \in \text{supp}(z_i^v)$ and the according $\bar{a} \equiv (\bar{a}_1(z^v), \cdots, \bar{a}_J(z^v))$. Note that $\mathbb{P}(c_i > 0 | \bar{a}) = 1$ implies $\mathbb{P}\Big(c_i > 0 \Big| z_i^v, \bar{a}\Big) = 1$ almost surely. Then, almost surely,

$$
\begin{aligned}
&\mathbb{P}\Big(j \notin r_i \ \forall j = 1, \cdots, J \Big| z_i^v = z^v, a_i = \bar{a}\Big) \\
&= \mathbb{P}\Big(c_{ij} < 0 \text{ or } v_{ij} < 0 \ \forall j = 1, \cdots, J \Big| z^v, \bar{a}\Big) && \text{by proof of Observation 1 with generalized } L \\
&= \mathbb{P}\Big(v_{ij} < 0 \ \forall j = 1, \cdots, J \Big| z^v, \bar{a}\Big) && \text{by } \mathbb{P}(c_i > 0 | z^v, \bar{a}) = 1 \\
&= \mathbb{P}\Big(v_i < 0 \Big| z^v, \bar{a}\Big) \\
&= \mathbb{P}\Big(v_i < 0 \Big| z^v\Big) && \text{by } v_i \perp\!\!\!\perp a_i | z_i^v \\
&= \mathbb{P}\Big(\tilde{v}_i < z^v\Big). && \text{by } \tilde{v}_i \perp\!\!\!\perp z_i^v
\end{aligned}
$$

As the first line is observed, the last line is identified almost surely for $z^v \in \mathbb{R}^J$ by the large support assumption on z_i^v. Then, by the independence assumptions on a_i and z_i^v, $\mathbb{P}(v_i > x | z^v, a) = \mathbb{P}(v_i > x | z^v) = \mathbb{P}(\tilde{v}_i > x + z^v)$. Therefore, $\mathbb{P}(v_i > x | z^v, a) = \mathbb{P}(v_i > x | z)$ is identified for almost every $(x, z) \in \text{supp}(v_i, z_i)$. □

Proof of Proposition C.2. I will implicitly condition everything on $z_i \backslash z_i^c$. I first prove (i). Take

any $z^c \in \text{supp}(z_i^c)$. Note that

$$\begin{aligned}
&\mathbb{P}\Big(j \in r_i \ \forall j = 1, \cdots, J \Big| z_i^c = z^c\Big) \\
&= \mathbb{P}\Big(c_i > 0, v_i > 0 \Big| z^c\Big) && \text{by proof of Observation 1 with generalized } L \\
&= \mathbb{P}\Big(c_i > 0 \Big| z^c\Big)\mathbb{P}\Big(v_i > 0 \Big| z^c\Big) && \text{by } c_i \perp\!\!\!\perp v_i | z_i^c \\
&= \mathbb{P}\Big(\tilde{c}_i > z^c\Big)\mathbb{P}\Big(v_i > 0 \Big| z^c\Big) && \text{by } \tilde{c}_i \perp\!\!\!\perp z_i^c
\end{aligned}$$

but the first line is observed and $\mathbb{P}\Big(v_i > 0 \Big| z^c\Big)$ is known on almost all $z^c \in \text{supp}(z_i^c)$ by assumption. Thus, $\mathbb{P}\Big(\tilde{c}_i > z^c\Big)$ is identified almost surely. By the assumptions on z_i^c, $\mathbb{P}\Big(c_i > x \Big| z^c\Big) = \mathbb{P}\Big(\tilde{c}_i > x + z^c\Big)$ and thus $\mathbb{P}\Big(c_i > x \Big| z^c\Big)$ is identified for almost all $(x, z^c) \in \text{supp}(c_i, z_i^c)$. The result follows from the definition of c_i^*, i.e. $c_{ij}^* = \mathbb{1}(c_{ij} > 0)$ for all (i, j).

The proof of (ii) follows analogously by noting that $\mathbb{P}\Big(j \in r_i \ \forall j \in \mathcal{A} \Big| z_i^c = z^c\Big) = \mathbb{P}\Big((\tilde{c}_{ij})_{j\in\mathcal{A}} > (z_j)_{j\in\mathcal{A}}^c\Big)\mathbb{P}\Big((v_{ij})_{j\in\mathcal{A}} > 0 \Big| z^c\Big)$ and that $\mathbb{P}\Big(j \in r_i \ \forall j \in \mathcal{A} \Big| z_i^c = z^c\Big)$ is observed while $\mathbb{P}\Big((v_{ij})_{j\in\mathcal{A}} > 0 \Big| z^c\Big)$ is assumed identified. □

Proof of Proposition C.3. Define $v_{ij}^* = v_{ij}\big(2 \cdot \mathbb{1}(v_{ij} > 0, c_{ij}^* = 1) - 1\big)$. Note first that the assumptions imply the distribution of $v_i^* \equiv (v_{ij})_{j\in\mathcal{J}}$ is known. Note also that $\arg\max_{r\in\mathcal{R}(\mathcal{C}_i)} v \cdot p^r = \arg\max_{r\in\mathcal{R}(\mathcal{J})} v^* \cdot p^r$. Therefore, two beliefs $p \equiv \{p_j^r\}_{j\in\mathcal{J}, r\in\mathcal{R}(\mathcal{J})}$ and $p' \equiv \{p_j'^r\}_{j\in\mathcal{J}, r\in\mathcal{R}(\mathcal{J})}$ are behaviorally equivalent if and only if for all $v \in \mathcal{R}^J$, $\arg\max_{r\in\mathcal{R}(\mathcal{J})} v \cdot p^r = \arg\max_{r\in\mathcal{R}(\mathcal{J})} v \cdot p'^r$. Let $C^r(p) \equiv \{v \in \mathbb{R}^J | r = \arg\max_{r\in\mathcal{R}(\mathcal{J})} v \cdot p^r\}$ for each $r \in \mathcal{R}(\mathcal{J})$. Then, two beliefs p and p' are behaviorally equivalent if and only if $C^r(p) = C^r(p')$ for all $r \in \mathcal{R}(\mathcal{J})$.

`Proof under assumption (1):` $L = J = 2$.

Implicitly condition on everything on z_i. From Observation 1, it is straightforward to verify that $\big(C^r(p)\big)_{r\in\mathcal{R}(\mathcal{J})}$ is pinned down by a single number $\delta \equiv \frac{p_1^{(1)} - p_1^{(2,1)}}{p_2^{(2)} - p_2^{(1,2)}}$.This can be checked by noting that $C^{\emptyset}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1, v_2 \leq 0\}$, $C^{(1)}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \geq 0, v_2 \leq 0\}$, $C^{(2)}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 \leq 0, v_2 \geq 0\}$, $C^{(1,2)}(p) = \{(v_1, v_2) \geq 0 | v_2/v_1 \leq \delta, \}$, and $C^{(2,1)}(p) = \{(v_1, v_2) \geq 0 | v_2/v_1 \geq \delta\}$. By assumption, everyone (in the subgroup defined by the observables) shares the common belief $p = \{p_j^r\}_{j\in\mathcal{J}, r\in\mathcal{R}(\mathcal{J})}$ and therefore $\mathbb{P}(\{v_{i2}/v_{i1} \geq \delta\} \cap \{v_i \geq 0\}) = \mathbb{P}(v_i \in C^{(2,1)}(p)) = \mathbb{P}(r_i = (2, 1))$. As $\mathbb{P}(v_i \leq v)$ is known, the left-hand side of the equation is calculable as a function of δ. On the other hand, the right-hand side is observable. Thus, belief is identified.

`Proof under assumption (2):` $L = 1$.

By assumption, everyone has the same belief, which I denote by p. Note that $C^{(j)}(p) = \{v \in \mathbb{R}^J | j = \arg\max_{k \in 0,1,\ldots,J} p_k^{(k)} v_k\} = \{v \in \mathbb{R}^J | j = \arg\max_{k \in 0,1,\ldots,J} \frac{p_k^{(k)}}{p_1^{(1)}} v_k\}$ for $j = 1, \ldots, J$ and $C^{\emptyset}(p) = \{(v_1, v_2) \in \mathbb{R}^2 | v_1, v_2 \leq 0\}$. Thus, the $C_r(p)'s$ are completely characterized by the vector $\tilde{p} \equiv (\tilde{p_2}, \cdots, \tilde{p}_J) \equiv (\frac{p_2}{p_1}, \ldots, \frac{p_J}{p_1})$. Therefore, belief is identified if $\tilde{p}$ is identified.

I now claim that one can use Corollary 1 of Berry et al. (2013), denoted BGH. In their notation, $x = \tilde{p}$, $\mathcal{X}^* = \mathcal{X} = \mathbb{R}_{++}^{J-1}$, and $\sigma(\tilde{p}) = (\sigma_2(\tilde{p}), \cdots, \sigma_J(\tilde{p})) : \mathcal{X} \subseteq \mathbb{R}^{J-1} \to \mathbb{R}^{J-1}$ where $\sigma_j(\tilde{p}) = \frac{\mathbb{P}(v_i \in C^{(j)}(\tilde{p}))}{\sum_{k=1}^{J} \mathbb{P}(v_i \in C^{(k)}(\tilde{p}))}$ for $j = 1, \ldots, J$. Note that the school $j = 1$ now plays the role of BGH's "outside option" (which is denoted $j = 0$ in their notation).[58] To see that the corollary applies, note first that $\mathcal{X}$ is a Cartesian product. Moreover, $\sigma_j(\tilde{p})$ is strictly decreasing in $\tilde{p}_k$ for all $j = \{1, \ldots, J\}$ and for all $k \neq 1, j$, as (1) $\sum_{k=1}^{J} \mathbb{P}(v_i \in C^{(k)}(\tilde{p}))$ is constant over $\tilde{p}$, and (2) $\mathbb{P}(v_i \in C^{(j)}(\tilde{p}))$ is strictly decreasing because v_i has full support. Thus, BGH's Corollary 1 applies and $\sigma(\tilde{p})$ is injective. □

Proof of Proposition C.4 . I first prove case (i). Take any $z \equiv (z^v, z^c, z^-)$ such that $z^v \in \mathbb{R}^J$, $z^c \in \mathbb{R}^J$, and $z^- \in \text{supp}(z_i^-)$. Then,

$$
\begin{aligned}
&\mathbb{P}\big(j \in r_i \ \forall j = 1, \cdots, J | z_i^v = z^v, z_i^c = z^c, z_i^- = z^-\big) \\
&= \mathbb{P}(\tilde{v}_i - z^v > 0, \tilde{c}_i - z^c > 0 | z^v, z^c, z^-) \\
&= \mathbb{P}(\tilde{v}_i > z^v, \tilde{c}_i > z^c | z^-) \\
&= \mathbb{P}(-\tilde{v}_i < -z^v, -\tilde{c}_i < -z^c | z^-)
\end{aligned}
$$

and since the first expression is observed for any $z^v \in \mathbb{R}^J$, $z^c \in \mathbb{R}^J$, and $z^- \in \text{supp}\left(z_i^-\right)$, the last expression is identified for any such (z^v, z^c, z^-). Thus, the joint distribution of $(-\tilde{v}_i, -\tilde{c}_i)$ conditional on z_i^-, and therefore the joint distribution of $(\tilde{v}_i, \tilde{c}_i)$ conditional on z_i^-, is identified on the support of z_i^-. As $v_i = \tilde{v}_i - z_i^v$ and $c_i = \tilde{c}_i - z_i^c$ with $(\tilde{v}_i, \tilde{c}_i) \perp\!\!\!\perp (z_i^v, z_i^c) | z_i^-$ and $z_i \equiv (z_i^v, z_i^c, z_i^-)$ is observed, the joint distribution of (v_i, c_i) conditional on z_i is identified for every z_i in its support.

[58] The outside option $j = 0$ as considered in my model is left out of the discussion here because their choice probability does not change according to p.

To show the first part of case (ii), note that

$$\begin{aligned}
&\mathbb{P}\big(r_i = \emptyset | z_i^v = z^v, z_i^c = z^c, z_i^- = z^-\big) \\
&= \mathbb{P}(v_{ij} \leq 0 \text{ or } c_{ij} \leq 0 \ \forall j \in \mathcal{J} | z^v, z^c, z^-) \\
&= \mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \text{ or } \tilde{c}_{ij} < z_{ij}^c \ \forall j \in \mathcal{J} | z^-)
\end{aligned}$$

Now, send all of the elements in z^c to negative infinity. By the dominated convergence theorem, the last expression converges to $\mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \ \forall j \in \mathcal{J}|z^-)$. Note that z_i^v is a special regressor for v_i with a large support. Use the special regressor similarly as before to identify the distribution of v_i. The second part of case (ii) follows similarly by sending all of the elements in z^v to negative infinity. □

Proof of Proposition C.5. Proof of part (i) follows by noting that

$$\begin{aligned}
&\mathbb{P}(r_i \text{ includes no school among } \mathcal{S}(z_i) \ |z_i = z) \\
&= \mathbb{P}((v_{ij})_{j \in \mathcal{S}(z_i)} \leq 0 | z_i = z) \\
&= \mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \ \forall j \in \mathcal{S}(z_i) | z_i^v = z^v, z_i^- = z^-) \\
&= \mathbb{P}((\tilde{v}_{ij})_{j \in \mathcal{S}(z_i)} \leq (z_j^v)_{j \in \mathcal{S}(z_i)} | z_i^- = z^-)
\end{aligned}$$

and using the independence of the special regressor to recover the distribution of $(v_{ij})_{j \in \mathcal{S}(z_i)}|z_i$.

I now show part (ii). Take $z_i = z$ and $\mathcal{A} \subseteq \mathcal{S}(z)$ with $|\mathcal{A}| \leq L$. Implicitly condition everything on z. Note that for any two events A and B, $\mathbb{P}(A|B)\mathbb{P}(B) \leq \mathbb{P}(A) \leq \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(B^c)$. Consider the events $A = \{v_{ij} \leq 0 \ \forall j \in \mathcal{A}\}$ and $B = \{|r_i| = L, r_i \cap \mathcal{A} = \emptyset\}^c$. One can verify that $\mathbb{P}(A|B) = \mathbb{P}(j \notin r_i \ \forall j \in \mathcal{A}|B)$ using Observation 2. Further, note that $\mathbb{P}(j \notin r_i \ \forall j \in \mathcal{A}|B)$ and $\mathbb{P}(B)$ is observable. Thus, $\mathbb{P}(A) \equiv \mathbb{P}(v_{ij} \leq 0 \ \forall j \in \mathcal{A}) = \mathbb{P}(\tilde{v}_{ij} \leq z_{ij}^v \ \forall j \in \mathcal{A})$ is bounded within an interval of length $\mathbb{P}(B^c)$. One can then use the special regressor similarly as before to bound $\mathbb{P}\big((v_{ij})_{j \in \mathcal{A}} \leq x | z\big)$. □

C.5 Parametric Identification of Beliefs

Because nonparametric point-identification results are unavailable for the subjective beliefs about probabilities of assignments, we give intuitive arguments about how each of the belief parameters are identified.

Doubt parameter $\sigma_\nu(z_i)$ can be identified by the degree of truthtelling behavior. When $\sigma_\nu(z_i) \to \infty$ so that the applicant becomes completely doubtful of his assessment of the differ-

ence between cutoffs and his scores, his conditional probability of assignment q_{ijk} approaches 0.5 for every school j regardless of the rank k at which he puts the school on the list, in which case it becomes optimal for him to truthtelly rank the considered schools that are preferred to the outside option, until the maximum allowed list length is reached. But we can detect truthtelling behavior in data; having identified the distribution of utilities and consideration from earlier steps, and since truthful report is a function of utilities and consideration, we know the counterfactual distribution of truthful reports.

Heterogeneity in bias among individuals, $\sigma_\eta(z_i)$, can be identified by unexplained variation in reporting behavior among individuals with similar characteristics. The variation can be generated by variance in v_{ij} and c_{ij}, but we already know those variances, and higher $\sigma_\eta(z_i)$ will amplify the variation.

Pessimism bias $\mu(x_j, z_i)$ can be identified by riskiness of schools in the portfolios of students with full lists. Higher pessimism for school j will tend to drive the conditional probability of acceptance to school j to zero, making it unlikely that the school is written in a full list.

$\beta^{\mathsf{rank}} = 0$ will result in (1) truthfully ordered list if the list constraint does not bind, and (2) strategically chosen portfolios when the list constraint does bind. That is, if the constraint binds, it is possible that there are some unlisted considered schools that gives higher utility than one of the listed schools. (This cannot happen if $\sigma_\nu(z_i) = \infty$.) Even in that case, however, among the chosen portfolio of the schools, schools will still be truthfully ranked.

D Estimation: Details

D.1 Likelihood of Inclusion and the Score Moments

D.1.1 Log-likelihoods

Here we derive the the formula of likelihoods of school inclusions and discuss why the true parameters maximize the likelihoods. The likelihoods that we consider are not standard in the sense that (1) they select students with $s_{ij} = 1$ and (2) one of the likelihoods is weighted. We show that the true parameters maximize the likelihoods despite being non-standard.

We first derive the formula of log-likelihood of inclusion of school j in the report of applicant i. The log-likelihood reflects the identifying information in Observations 1 and 2. It "selects" individuals with $s_{ij} := \mathbb{1}(|r_i \backslash \{j\}| < 11)$ for the reasons explained in Section 7.1;

from Lemma D.1 given that $(\epsilon^v_{ij}, \epsilon^c_{ij})_{j\in\mathcal{J}}$ is $i.i.d$ across j, $|r_i \backslash \{j\}| < 11$ is independent of $(\epsilon^v_{ij}, \epsilon^c_{ij})$ conditional on observables (x_j, z_i). Let $\iota_{ij} := \mathbb{1}(j \in r_i)$ denote the random variable indicating whether school j was included in the report r_i. Let $w_{ij} := \mathbb{1}(v_{ij} > 0)\mathbb{1}(c_{ij} > 0)$ and note that $w_{ij} = \iota_{ij}$ whenever $s_{ij} = 1$ following Observation 1. Let $f_{w|z,s}(\cdot|z', s'; \theta)$ denote the density of w_{ij} given $z_{ij} = z'$, $s_{ij} = s'$, and θ. Similarly define $f_{\iota|z,s}(\cdot|z', s'; \theta)$ and $f_{w|z}(\cdot|z'; \theta)$. We treat $(x_j)_j$ as nonrandom. Then,

$$
\begin{aligned}
&\log \Pi_i \Pi_{j:s_{ij}=1} f_{\iota|z,s}(\iota_{ij}|z_{ij}, 1; \theta) \qquad \text{(D.1)}\\
&= \log \Pi_i \Pi_{j:s_{ij}=1} f_{w|z,s}(w_{ij}|z_{ij}, 1; \theta)\\
&= \log \Pi_i \Pi_{j:s_{ij}=1} f_{w|z}(w_{ij}|z_{ij}; \theta)\\
&= \log \Pi_i \Pi_{j:s_{ij}=1, j\notin\mathcal{S}_i} \big(1 - \mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)\mathbb{P}(c_{ij} > 0|z_{ij}; \theta)\big)^{1-w_{ij}} \cdots\\
&\qquad \big(\mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)\mathbb{P}(c_{ij} > 0|z_{ij}; \theta^c)\big)^{w_{ij}} \cdots\\
&\qquad \Pi_{j:j\in\mathcal{S}_i} \big(1 - \mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)\big)^{1-w_{ij}} \mathbb{P}(v_{ij} > 0|z_{ij}; \theta^v)^{w_{ij}}\\
&= \sum_i \Bigg[\sum_{j:s_{ij}=1, j\notin\mathcal{S}_i} \Big[\big(1 - w_{ij}\big) \log \big(1 - \overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big) + w_{ij} \log \big(\overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big) \Big]\\
&\qquad + \sum_{j:s_{ij}=1, j\in\mathcal{S}_i} \Big[(1 - w_{ij}) \log \big(\Phi(-\psi^v_{ij})\big) + w_{ij} \log \big(\overline{\Phi}(-\psi^v_{ij})\big) \Big] \Bigg]
\end{aligned}
$$

where $\bar{\Phi}(\cdot) := 1 - \Phi(\cdot)$, $\psi^v_{ij} := v_{ij} - \epsilon^v_{ij}$, $\psi^c_{ij} := c_{ij} - \epsilon^c_{ij}$, θ^v denotes the preference parameters, and θ^c denotes the consideration parameters. For notational convenience, the dependence of ψ^v_{ij} on θ^v and the dependence of ψ^c_{ij} on θ^c are made implicit. The second equality comes from $s_{ij} = 1$ being independent of $(\epsilon^v_{ij}, \epsilon^c_{ij})$ and therefore also of (v_{ij}, c_{ij}) conditional on observables.

We now show that the population version of the log-likelihood is maximized by the true parameters θ_0. Define

$$
Q(\theta) := E_{\theta_0} \sum_{j:s_{ij}=1} \log f_{\iota|z,s}(\iota_{ij}|z_{ij}, 1; \theta) \equiv E_{\theta_0} \sum_{j:s_{ij}=1} \log f_{w|z}(w_{ij}|z_{ij}; \theta)
$$

where both w_{ij} and z_{ij} are random variables. This is the population version of the log-likelihood (Equation D.1) in the sense that

$$
\operatorname*{plim}_{n\to\infty} n^{-1} \log \Pi_i \Pi_{j:s_{ij}=1} f_{\iota|z,s}(\iota_{ij}|z_{ij}, 1; \theta) = Q(\theta)
$$

where n denotes the number of students in the sample and with the understanding that in the left-hand side (ι_{ij}, z_{ij}) are realized values.

Now we show $Q(\theta_0) \geq Q(\theta)$ for all θ. Note that

$$
\begin{aligned}
&E_{\theta_0}\Big[\frac{f_{w|z}(w_{ij}|z_{ij};\theta)}{f_{w|z}(w_{ij}|z_{ij};\theta_0)}\Big|z_{ij}\Big] \\
&= \frac{f_{w|z}(0|z_{ij};\theta)}{f_{w|z}(0|z_{ij};\theta_0)} f_{w|z}(0|z_{ij};\theta_0) + \frac{f_{w|z}(1|z_{ij};\theta)}{f_{w|z}(1|z_{ij};\theta_0)} f_{w|z}(1|z_{ij};\theta_0) \\
&= 1. \qquad \text{(D.2)}
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&Q(\theta) - Q(\theta_0) \\
&= E_{\theta_0} \sum_j \Big[s_{ij} \log f_{w|z}(w_{ij}|z_{ij};\theta) - \log f_{w|z}(w_{ij}|z_{ij};\theta_0) \Big] \\
&= \sum_j E_{\theta_0} \Big[s_{ij} \log \frac{f_{w|z}(w_{ij}|z_{ij};\theta)}{f_{w|z}(w_{ij}|z_{ij};\theta_0)} \Big] \\
&= \sum_j E_{\theta_0} \Big[E_{\theta_0} \Big[s_{ij} \log \frac{f_{w|z}(w_{ij}|z_{ij};\theta)}{f_{w|z}(w_{ij}|z_{ij};\theta_0)} \Big| s_{ij}, z_{ij} \Big] \Big] \\
&= \sum_j E_{\theta_0} \Big[s_{ij} E_{\theta_0} \Big[\log \frac{f_{w|z}(w_{ij}|z_{ij};\theta)}{f_{w|z}(w_{ij}|z_{ij};\theta_0)} \Big| s_{ij}, z_{ij} \Big] \Big] \\
&\leq \sum_j E_{\theta_0} \Big[s_{ij} \log E_{\theta_0} \Big[\frac{f_{w|z}(w_{ij}|z_{ij};\theta)}{f_{w|z}(w_{ij}|z_{ij};\theta_0)} \Big| s_{ij}, z_{ij} \Big] \Big] \\
&= \sum_j E_{\theta_0} \Big[s_{ij} \log E_{\theta_0} \Big[\frac{f_{w|z}(w_{ij}|z_{ij};\theta)}{f_{w|z}(w_{ij}|z_{ij};\theta_0)} \Big| z_{ij} \Big] \Big] \\
&= 0
\end{aligned}
$$

where the inequality holds by Jensen's inequality, the penultimate inequality holds from $(c_{ij}, v_{ij}) \perp\!\!\!\perp s_{ij}|z_{ij}$ and therefore $w_{ij} := \mathbb{1}(c_{ij} > 0)\mathbb{1}(v_{ij} > 0) \perp\!\!\!\perp s_{ij}|z_{ij}$, and the last equality holds from Equation D.2.

As the sample size of (i, j) pairs such that i surely considers j is small relatively those that do not have the sure-consideration relationship, in some specifications we weight the

sure-consideration pairs. The weighted log-likelihood is

$$w_{\text{NSC}} \sum_i \sum_{j:s_{ij}=1, j\notin\mathcal{S}_i} \Big[(1-\mathbb{1}(j\in r_i))\log\big(1-\overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big) + \mathbb{1}(j\in r_i)\log\big(\overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})\big)\Big]$$
$$+ w_{\text{SC}} \sum_i \sum_{j:s_{ij}=1, j\in\mathcal{S}_i} \Big[(1-\mathbb{1}(j\in r_i))\log\big(\Phi(-\psi^v_{ij})\big) + \mathbb{1}(j\in r_i)\log\big(\overline{\Phi}(-\psi^v_{ij})\big)\Big]$$

for some weights w_{NSC} and w_{SC} such that $w_{NSC}\sum_{j\notin\mathcal{S}_i} s_{ij} + w_{SC}\sum_{j\in\mathcal{S}_i} s_{ij} = \sum_{j\in\mathcal{J}} s_{ij}$. That is, the schools that are not surely considered are weighted by w_{NSC} and those that are surely considered are weighted by w_{SC}.

The true parameters maximize the population version of the weighted likelihood. To see this, suppose that (θ^v, θ^c) maximizes the weighted log-likelihood, where θ^v denotes the preference parameters and the θ^c denotes the consideration parameters. Let $\theta_0 := (\theta^v_0, \theta^c_0)$ denote the true parameter. Suppose that $\theta^v \neq \theta^v_0$. Then, θ^v_0, θ^c_0 gives larger value of the weighted likelihood than does (θ^v, θ^c), which is a contradiction. Thus, $\theta^v = \theta^v_0$. Now, suppose that $\theta^c \neq \theta^c_0$. Because (θ^v_0, θ^c_0) gives larger value for the first part, and doesn't have any implication for the second part, this is another contradiction. The "scores" of the weighted log-likelihood can be obtained analogously.

The corresponding likelihood scores for preference parameters are

$$\sum_i \sum_{j:s_{ij}=1} \Bigg[(1-\mathbb{1}(j\in\mathcal{S}_i))\Bigg((1-\mathbb{1}(j\in r_i))\frac{-\phi(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})}{1-\overline{\Phi}(-\psi^v_{ij})\overline{\Phi}(-\psi^c_{ij})} + \mathbb{1}(j\in r_i)\frac{\phi(-\psi^v_{ij})}{\overline{\Phi}(-\psi^v_{ij})}\Bigg)$$
$$+\mathbb{1}(j\in\mathcal{S}_i)\Bigg((1-\mathbb{1}(j\in r_i))\frac{-\phi(-\psi^v_{ij})}{\Phi(-\psi^v_{ij})} + \mathbb{1}(j\in r_i)\frac{\phi(-\psi^v_{ij})}{\overline{\Phi}(-\psi^v_{ij})}\Bigg)\Bigg]\frac{\partial\psi^v_{ij}(\theta^v)}{\partial\theta^v}$$

and the corresponding likelihood scores for consideration parameters are

$$\sum_i \sum_{j:s_{ij}=1} \Bigg[(1-\mathbb{1}(j\in\mathcal{S}_i))\Bigg((1-\mathbb{1}(j\in r_i))\frac{-\phi(-\psi^c_{ij})\overline{\Phi}(-\psi^v_{ij})}{1-\overline{\Phi}(-\psi^c_{ij})\overline{\Phi}(-\psi^v_{ij})} + \mathbb{1}(j\in r_i)\frac{\phi(-\psi^c_{ij})}{\overline{\Phi}(-\psi^c_{ij})}\Bigg)\Bigg]\frac{\partial\psi^c_{ij}(\theta^c)}{\partial\theta^c}.$$

D.2 Simulated Ordering Moments

Implicitly condition on $(x_j)_{j\in\mathcal{J}}$. Note that for any $f:\mathcal{R}\to\mathbb{R}^m$,

$$0 = \mathbb{E}\Big[f(r_i) - \mathbb{E}[f(r_i)|z_i]\Big|z_i\Big] = \mathbb{E}\Big[f(r_i) - \mathbb{E}[f(r(z_i, e_i;\theta_0))|z_i]\Big|z_i\Big]$$

where e_i denotes the vector of unobservables $(\epsilon_i^v, \epsilon_i^c, \eta_i)$, θ denotes the parameter vector, θ_0 denotes the true parameter vector, and $r(z_i, e_i; \theta)$ denotes the subjectively optimal report under (z_i, e_i, θ) which is uniquely defined with probability 1. Section E describes the procedure for simulating $r(z_i, e_i; \theta)$. It follows that

$$\begin{aligned}
&\mathbb{E}\Big[\Big(f(r_i) - \mathbb{E}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big]\Big)h(z_i)\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[f(r_i) - \mathbb{E}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big]\Big|z_i\Big]h(z_i)\Big] \\
&= 0
\end{aligned}$$

where $h(z_i)$ may be a $m-$dimensional vector.

The sample equivalent of this condition is

$$\frac{1}{I}\sum_i \Big(f(r_i) - \mathbb{E}^{\text{sim}}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big]\Big)h(z_i) = 0 \tag{D.3}$$

where in the brute-force version of simulation

$$\mathbb{E}^{\text{sim}}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big] = \frac{1}{S}\sum_s f\big(r(z_i, e_i^s; \theta_0)\big)$$

where the distribution of e_i^s is completely governed by θ_0 and not by z_i due to independence. We use a smoothed version of $\mathbb{E}^{\text{sim}}\big[f\big(r(z_i, e_i; \theta_0)\big)\big|z_i\big]$ following Ackerberg (2009).

The first type of simulated moments, given below, gives information about how individuals order the schools:

$$\mathbb{E}\Big[\frac{1}{J}\sum_j \Big(\mathbb{1}(j \in r_i^k) - \mathbb{P}\big(j \in r^k(z_i, e_i; \theta)\big)h_j(z_i)\Big] = 0 \quad \forall k = 1, \ldots, 12$$

where r_i^k is represents the report r_i truncated up to the kth slot, $r^k(\cdot)$ is the equivalent for the simulated report, and the set inclusion notation is used towards r_i^k and $r^k(\cdot)$ with a slight abuse. The condition uses $f(r_i) = \frac{1}{J}\big(\mathbb{1}(j \in r_i^k)\big)_{j\in\mathcal{J}}$ in the notation of equation D.3. The moment condition is implemented by

$$\frac{1}{IJ}\sum_i\sum_j \Big(\mathbb{1}(j \in r_i^k) - \mathbb{E}^{\text{sim}}\big[\mathbb{1}\big(j \in r^k(z_i, e_i; \theta)\big)\big|z_i\big]\Big)h_j(z_i)$$

and we use

$$h(z_i) = \big(1, z_{ij}, (z_{ij} - \bar{z}_i), (z_{ij} - \bar{z}_i)^2\big)_{j \in \mathcal{J}}.$$

Analogously, the second type of simulated moments give information about within-individual ordering:

$$\frac{1}{IJ(J-1)} \sum_i \sum_j \sum_{j' \neq j} \Big(\mathbb{1}(j \in r_i^k) \mathbb{1}(j' \notin r_i^k) - \mathbb{E}^{\text{sim}}\big[\mathbb{1}\big(j \in r_{i,\theta}^k\big) \mathbb{1}\big(j' \notin r_{i,\theta}^k\big) \big| z_i\big]\Big) \big(h(z_{ij}) - h(z_{ij'})\big)$$

where $r_{i,\theta}^k := r^k(z_i, e_i; \theta)$.

D.3 Lemmas

Lemma D.1. *The event $|r_i \setminus \{j\}| < 11$ is independent of $(\epsilon_{ij}^c, \epsilon_{ij}^v)$ conditional on observables.*

Proof. Fix the observables (x, z). We shall show that the event $|r_i \setminus \{j\}| < 11$ is the same as the event $\sum_{j' \neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11$. Being determined by only $(\epsilon_{ij'}^c, \epsilon_{ij'}^v)_{j' \neq j}$ (and $\xi_{j'}$), the latter is independent of $(\epsilon_{ij}^c, \epsilon_{ij}^v)$ as desired.

Note that

$$\begin{aligned} |r_i \setminus \{j\}| < 11 \quad &\text{iff} \quad |r_i \setminus \{j\}| < 11 \quad \text{and} \quad |r_i| < 12 \\ &\text{iff} \quad \sum_{j' \neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11 \quad \text{and} \quad |r_i| < 12 \\ &\text{iff} \quad \sum_{j' \neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11. \end{aligned}$$

The first equivalence holds because $|r_i \setminus \{j\}| < 11$ implies $|r_i| < 12$. The second equivalence holds due to the first statement of Observation 1 (since $|r_i| < 12$). For the last equivalence, "only if" holds trivially. The "if" holds due to the second statement of Observation 1; $\sum_{j' \neq j} \mathbb{1}\{c_{ij'} > 0, v_{ij'} > 0\} < 11$ implies $\sum_{j \in \mathcal{J}} \mathbb{1}\{c_{ij} > 0, v_{ij} > 0\} < 12$, which in turn implies $|r_i| < 12$ by Observation 1. □

E Simulating Subjectively Optimal Reports

Here we describe the procedure for obtaining the subjectively optimal reports[59]

$$r(z_i, e_i, \theta) = \arg\max_{r \in \mathcal{R}(\mathcal{C}_i)} \sum_{j=0}^{J} p_{ij}^r v_{ij} \tag{E.1}$$

where the distribution of $(\mathcal{C}_i, v_{ij}, p_{ij}^r)_{ij}$ depends on θ. Note that

$$\arg\max_{r \in \mathcal{R}(\mathcal{C}_i)} \sum_{j=0}^{J} p_{ij}^r v_{ij} = \arg\max_{r \in \mathcal{R}(\mathcal{J}_i^+)} \sum_{j=0}^{J} p_{ij}^r v_{ij}$$

where $\mathcal{J}_i^+ = \{j \in \mathcal{C}_i | v_{ij} > 0\}$, as students will never wish to list any school outside $\mathcal{J}_i^+$.

This problem is difficult to solve because as the size of a choice set, even after being reduced to $\mathcal{R}(\mathcal{J}_i^+)$, can be very large. For instance, with $|\mathcal{J}_i^+| = 20$, the choice set $\mathcal{R}(\mathcal{J}_i^+)$ is all possible ordered lists using the schools in $\mathcal{J}_i^+$ which has as many as $20!/(20-12)! \simeq 6.03 * 10^{13}$ elements. To make this problem solvable, we represent this problem as what resembles a finite-horizon dynamic programming problem, where a period is a slot in the list and a state is the set of schools already listed.

Let j_k represent the school listed in the kth spot. Note $p_{ij}^r = \Pi_{l=1}^{k-1}(1 - q_{ij_{r_l}l})q_{ijk}$. Let $K = \min\{12, |\mathcal{J}_i^+|\}$, which represents the last slot (or period) that the student optimally fills in. Each student solves the following problem:

$$\begin{aligned}
&\arg\max_{r \in \mathcal{R}(\mathcal{J}_i^+)} \sum_{j=0}^{J} p_{ij}^r v_{ij} \\
&= \max_{\{j_1, \cdots, j_K\} \subset \mathcal{J}_i^+} q_{ij_1 1} v_{ij_1} + (1 - q_{ij_1 1}) q_{ij_2 2} v_{ij_2} + \cdots + (1 - q_{ij_1 1}) \cdots (1 - q_{ij_{11} 11}) q_{ij_K} v_{ij_K} \\
&= \max_{\{j_1, \cdots, j_K\} \subset \mathcal{J}_i^+} q_{ij_1 1} v_{ij_1} + (1 - q_{ij_1 1}) \Big(q_{ij_2 2} v_{ij_2} + \cdots + (1 - q_{ij_2 2}) \cdots (1 - q_{ij_{11} 11}) q_{ij_K} v_{ij_K} \Big).
\end{aligned}$$

We solve the problem backwards from the last school the student puts in the list. Let $\mathcal{J}_k = \{j_1, \cdots, j_k\}$. Let

$$V_K^i(\{j_1, \cdots, j_{K-1}\}) = \max_{j \in \mathcal{J}_i^+ \setminus \mathcal{J}_{K-1}} q_{ijK} v_{ij}$$

[59] We ignore ties in optimal reports as they occur with probability zero.

and for $1 \leq k < K$, let

$$V_k^i(\{j_1, \cdots, j_{k-1}\}) = \max_{j \in \mathcal{J}_i^+ \setminus \mathcal{J}_{k-1}} q_{ijk} v_{ij} + (1 - q_{ijk}) V_{k+1}^i(\{j_1, \cdots, j_{k-1}, j\}).$$

Then,

$$V_1^i = \max_{j \in \mathcal{J}_i^+} q_{ij1} v_{ij} + (1 - q_{ij1}) V_2^i(\{j\}) = \max_{r \in \mathcal{R}(\mathcal{J}_i^+)} \sum_{j=0}^{J} p_{ij}^r v_{ij},$$

which shows that the original problem may be solved via the dynamic formulation.