Approaches to Education Market Design

by

Yusuke Narita

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

© Yusuke Narita, MMXVI. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature redacted

Author .................................................................

Department of Economics

May 13, 2016

Signature redacted

Certified by .....................

Parag Pathak
Professor of Economics
Thesis Supervisor

Signature redacted

Accepted by .........

Ricardo Caballero
Ford International Professor of Economics
Chairman, Departmental Committee on Graduate Studies
Approaches to Education Market Design

by

Yusuke Narita

Submitted to the Department of Economics
on May 13, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of essays about how to improve education markets through analyzing data generated by such markets. In chapter 1, I start with looking at how families decide which school to attend in a school choice system. Though such systems are designed assuming that families make well-informed choices upfront, I use data from NYC’s high school choice system to show that families’ choices change after the initial match as they learn about schools. I develop an empirical model of evolving demand for schools under learning, switching costs, and demand responses to prior assignments. The estimates suggest that there are even more changes in underlying demand, undermining the welfare performance of the initial match. To alleviate the cost of demand changes, I investigate dynamic mechanisms that best accommodate choice changes. These mechanisms improve on the existing discretionary reapplication process. In addition, the gains from the mechanisms dramatically change depending on the extent of demand-side inertia caused by switching costs. Thus, the gains from a centralized market depend not only on its design but also on demand-side frictions (such as demand changes and inertia).

In chapter 2, I turn to education production after students start attending schools. In centralized school admissions systems, rationing at oversubscribed schools often involves lotteries on top of preferences of students and schools. This random assignment is extensively used by empirical researchers to identify the effect of getting in a school on outcomes such as test scores. I theoretically study whether a popular empirical research design extracts a random assignment as intended, providing a condition under which the research design successfully extracts a random assignment.

Chapter 3 (with Atila Abdulkadiroğlu, Josh Angrist and Parag Pathak) considers the complementary question of how best to use the lottery-generated variation for impact evaluation. We develop easily-implemented strategies that fully exploit the random assignment embedded. We apply these methods to find large achievement gains from charter school attendance in Denver. By analyzing test-score consequences, chapters 2 and 3 complement chapter 1’s analysis of welfare/happiness consequences of school attendance.

Thesis Supervisor: Parag Pathak
Title: Professor of Economics
Acknowledgments

I am most grateful to Parag Pathak for his tireless help from the very start of my Ph.D life. His way of doing research was a big hint for me to try to form a view over a broad spectrum of social issues and analytical tools in economics. Daron Acemoglu and Josh Angrist are two extreme examples of what it is like to be a social scientist, which I am still struggling with. My experience in graduate school would have been much less fun without their humor, charm, and growls.

Nikhil Agarwal provided me with some of the most sharp criticisms to this thesis; apparently I need a little more time to fully resolve the issues he raised. Joint work with Atila Abdulkadiroğlu, John Hatfield, and Toshiji Kawagoe taught me a lot. I also thank other MIT faculty members for their kind input and interest, especially Itai Ashlagi, David Autor, Esther Duflo, Glenn Ellison, Bob Gibbons, Jon Gruber, Jerry Hausman, Whitney Newey, James Poterba, Jean Tirole, and Juuso Toikka. Koichiro Ito and Kosuke Uetake gave much guidance at some corners of the graduate school life. Pre-Ph.D advisors and mentors — late Masa Aoki, Hide Ichimura, Michi Kandori, Yosuke Yasuda, and especially Fuhito Kojima — remained to be sources of occasional but important instruction.

There are too many friends I should acknowledge. Numerous meals, drinks and silly chats with Dana Chandler, David Jimenez-Gomez, Ankur Mani, Ameya Muley, Harry Di Pei, Giovanni Reggiani, Peng Shi, Nathan Zorzi, and especially Mark Hou were relieving. Relatively more serious conversations and always sloppy mock presentations with Jie Bai, Alex Bartik, Vivek Bhattacharya, Alonso Bucarey, Masao Fukui, Chishio Furukawa, Ludo Gazze, Arda Gitmez, Sally Hudson, Peter Hull, Ryota Iijima, Gaston Illanes, Tetsuya Kaji, Ye Ji Kee, Yuhei Miyauchi, Manisha Padi, Brendan Price, Elizabeth Setren, Mayara Silva, Kentaro Tomoeda, and Yufei Wu were helpful. Warm welcomes by Gabriel Carroll, Scott Kominers, and Mike Powell were memorable. Daniel Waldinger and Louis Chen tirelessly corrected my English. Louis has also been an incredibly generous roommate for years.

Finally, I would say even bigger thank you to innumerable other non-economics friends and family members.
## Contents

1 Match or Mismatch: Learning and Inertia in School Choice 11
   1.1 Introduction .................................................. 11
   1.2 A First Look at Evolving Choices ............................. 16
      1.2.1 Evolving School Choices in NYC ....................... 16
      1.2.2 Choice Frictions and Learning ....................... 18
   1.3 Uncovering Evolving Demand ................................. 20
      1.3.1 Dynamic School Choice under Learning and Switching Costs ... 20
      1.3.2 Identification ........................................... 27
      1.3.3 Estimation ................................................. 33
      1.3.4 Estimates and Fit ....................................... 34
   1.4 Welfare Consequences of Evolving Demand ................... 36
      1.4.1 Costs of Ignoring Demand Changes ..................... 36
      1.4.2 Evaluating Reapplication Processes ................... 37
   1.5 Conclusion and Future Directions ............................ 41
   1.6 Appendix ....................................................... 54
      1.6.1 Data .......................................................... 54
      1.6.2 A First Look at Evolving Choices: Details ............ 55
      1.6.3 Uncovering Evolving Demand: Details .................. 56
      1.6.4 Theory for the Counterfactual Analysis ................ 61

2 Natural Experiments and Strategy-proofness 87
   2.1 Introduction .................................................. 87
   2.2 Framework ...................................................... 90
      2.2.1 Generalized Deferred Acceptance Mechanisms ........... 90
      2.2.2 First Choice Empirical Research Design ............... 93
      2.2.3 Motivating Example ....................................... 94
      2.2.4 Strategy-proofness for Schools .......................... 95
   2.3 Results ......................................................... 96
      2.3.1 Sufficiency ................................................ 96
      2.3.2 Almost Necessity ......................................... 99
      2.3.3 Sufficiency Revisited .................................... 101
2.4 Discussion and Extensions .......................................................... 103
  2.4.1 Interpretation of Strategy-proofness for Schools ......................... 103
  2.4.2 Alternative Definition of Extracting a Random Assignment ............... 103
  2.4.3 Qualification Instrumental Variable Research Design ...................... 103
  2.4.4 Alternative Definitions of Research Designs ............................... 105
  2.4.5 Top Trading Cycles Mechanism .............................................. 107
2.5 Future Directions .......................................................................... 109
2.6 Proofs .......................................................................................... 112
  2.6.1 Proof of Theorem 1 .................................................................. 112
  2.6.2 Proof of Corollary 2.b .............................................................. 123
  2.6.3 Proof of Proposition 2 ............................................................... 124

3 Research Design Meets Market Design: Using Centralized Assignment
for Impact Evaluation ........................................................................... 127
  3.1 Introduction .................................................................................. 127
  3.2 Understanding the DA Propensity Score .......................................... 130
    3.2.1 Propensity Score Pooling ...................................................... 130
    3.2.2 Further Pooling in Large Markets ........................................... 132
  3.3 Score Theory ................................................................................ 134
    3.3.1 Setup .................................................................................... 134
    3.3.2 Characterizing the DA Propensity Score .................................... 137
    3.3.3 Estimating the DA Propensity Score ......................................... 141
    3.3.4 Identification ......................................................................... 143
    3.3.5 Estimation .............................................................................. 145
  3.4 School Effectiveness in Denver ........................................................ 146
    3.4.1 Computing the DA Propensity Score ........................................ 147
    3.4.2 DPS Data and Descriptive Statistics ........................................ 149
    3.4.3 Score-Based Balance .............................................................. 151
    3.4.4 Effects of Charter Enrollment ................................................ 154
    3.4.5 Alternative IV Strategies ....................................................... 155
    3.4.6 Charter School Effects with a Mixed Counterfactual .................... 157
  3.5 Summary and Directions for Further Work ........................................ 162
  3.6 Theoretical Appendix .................................................................... 177
    3.6.1 Defining DA: Details .............................................................. 177
    3.6.2 Conditional Independence of DA-generated Offers ...................... 178
    3.6.3 Proof of Theorem 2 ................................................................ 179
    3.6.4 Proof of Theorem 3 ................................................................ 180
    3.6.5 First Choice and Qualification Instruments: Details .................... 185
    3.6.6 Extension to a General Lottery Structure .................................... 186
    3.6.7 The Boston (Immediate Acceptance) Mechanism ....................... 187
3.7 Empirical Appendix ......................................................... 188
3.7.1 Data ................................................................. 188
Chapter 1

Match or Mismatch: Learning and Inertia in School Choice

1.1 Introduction

From public housing to entry-level labor markets to school choice, centralized matching markets are a prominent form of public policy. These markets are usually designed assuming that participants make well-informed choices upfront, anticipating what they will demand in the future. Little is known, however, about whether this assumption holds in practice. In this paper, I study how families’ demand for schools evolves using data from NYC’s high school choice system, which serves more than 90,000 families applying for 700 schools each year. I develop a framework to recover families’ evolving demand under learning and switching costs, where demand is also allowed to change in response to prior assignments. I use the estimated framework to quantify demand changes and their welfare consequences.

My analysis uses administrative panel data from NYC’s high school choice system, consisting of the centralized first-round in December and a discretionary reapplication process in April. The data records families’ participation decisions as well as rank-ordered school choices both in the first-round and the reapplication process, allowing me to trace the dynamics of school choices. In Section 1.2, I demonstrate that contrary to the premise of well-informed upfront choices, families’ choices change after the initial match as they learn about schools. About 7% of families reapply and at least 70% of these reapplicants reverse choice (preference) orders over schools between the first round and the reapplication process. Most choice reversals can be rationalized only by real demand changes and not by strategic behavior with unchanged demand.\footnote{I say an applicant exhibits choice reversals if the following holds: In the reapplication process, she attempts to switch from the first-round assignment $s$ to another school that is unranked or ranked below $s$ in the preference list she reports in the first-round market. I say an applicant exhibits surely nonstrategic choice reversals if she exhibits choice reversals and does not exhaust her first-round preference, i.e., rank 11} These choice changes appear to be mainly
caused by learning. Most families self-report that they change their choices because of new information about schools or their preferences about schools. Moreover, consistent with their self-reports, families' choices become more correlated with and responsive to school characteristics. In particular, compared with initial applications, reapplications rank schools that are closer to families' homes and that are academically no worse.

The above choice changes provide only a lower bound on the amount of changes in underlying demand. There may or may not be additional unobserved demand changes. For families who do not reapply, their behavior does not directly reveal demand changes. However, some families may experience demand changes but not reapply because of "switching costs" or "reapplication costs," in which case demand changes exist but are not directly observed. For example, the time cost of filling out a paper application form may constitute reapplication costs.

To distinguish these scenarios and recover underlying evolving demand, the core part of this paper in Section 1.3 develops a structural model of dynamic school choice. The model incorporates demand changes by learning and reapplication costs. I also let demand change in response to initial assignments in the first round. Specifically, I allow the utility of the first-round assignment to change by some positive or negative amount, which can differ across applicants. Such demand responses are likely if families obtain more information about their initial assignments or if they experience psychological endowment effects about them. I provide a strategy to distinguish these behavioral elements by exploiting three institutional features. First, because of capacity constraints, many applicants are initially assigned to a school other than the most preferred school. Second, the assignments to the most preferred and less preferred schools are partly random, thanks to admissions lotteries used in the first-round assignment mechanism. Finally, reapplicants make new choices that are rank-ordered. These features allow me to decompose observed behavior into the model components (demand changes, reapplication costs, and demand changes in response to prior assignments).

Intuitively, the identification logic is as follows. For simplicity, let me ignore demand responses to initial assignments and assume that there are only two preference ranks, the first choice and the lower choice. Many applicants are "lower-choice non-reapplicants," who are initially assigned to their old lower choice but do not reapply, due either to reapplication costs or demand changes. This fact allows me to measure the total effects of demand changes and reapplication costs. There are also "first-choice reapplicants," who are initially assigned to their old first choice but reapply, which must be because of demand changes. Other applicants assigned to their first choice may also experience demand changes but be locked in by reapplication costs. The fraction of reapplicants or fewer schools though she could have ranked up to 12 schools. I show that (1) 71% of reapplicants exhibit choice reversals, (2) about 80% of the choice reversals are surely nonstrategic, and (3) surely nonstrategic choice reversals are consistent with optimal behavior only if intrinsic preferences and demand change between the first round and the reapplication process.
among all applicants assigned to the first choice thus tells us the difference between the amounts of demand changes and reapplication costs.

Now suppose that admissions lotteries in the first-round mechanism guarantee that initial assignments are randomly assigned and applicants assigned to the first-choice and lower-choice are comparable people with similar demand changes and reapplication costs. I can compare the fractions of lower-choice non-reapplicants and first-choice reapplicants to measure the amount of reapplication costs. Heuristically,

\[
\text{(fraction of non-reapplicants among applicants assigned to the initial lower choice)} \\
- \text{(fraction of reapplicants among applicants assigned to the initial first choice)} \\
= (\text{demand changes} + \text{reapplication costs}) - (\text{demand changes} - \text{reapplication costs}) \\
= 2 \times \text{reapplication costs},
\]

which separates reapplication costs from demand changes. The precise implementation of this argument generates complications related to more than two preference ranks, admissions lotteries embedded in assignment mechanisms, and demand responses to initial assignments. Solutions for these challenges are detailed below.

I estimate the model and find a significant role of learning, reapplication costs, and demand responses to initial assignments. Crucially, as detailed in Section 1.4, the estimates suggest that there are substantially more changes in underlying demand than in observed choices. These hidden demand changes are masked by reapplication costs, which prevent families from reapplying and expressing demand changes.\(^2\)

As a result, the welfare cost of ignoring demand changes is large. To measure the welfare cost, I compare the real first-round assignment based on old demand with the counterfactual "frictionless benchmark." The frictionless benchmark is defined as what would have been produced by the same first-round assignment mechanism, had families made choices based on their new demand after learning. Since the two differ only in whether families’ choices are based on old or new demand, the difference between the two captures the welfare costs of ignoring demand changes by learning.\(^3\) The real and frictionless assignments turn out to be significantly different. Specifically, the two assignments give different allocations (schools) to a majority of families; the average welfare loss under the real first-round assignment compared with the frictionless benchmark is more than 1-mile-equivalent, when I measure it by new demand assumed to be quasi-linear in the distance between the family and the school locations.\(^4\) This magnitude corresponds

\(^2\)Demand responses to initial assignments also lower the reapplication rate. The estimates show that families tend to get to prefer initially assigned schools more, compared with other schools. This satisfaction with initial assignments lowers the reapplication rate.

\(^3\)Except applicants’ choices, every other input is unchanged between the real first-round assignment and the frictionless benchmark. For example, school capacities and their preferences or priorities over applicants are fixed.

\(^4\)The 1-mile-equivalent utility unit can be interpreted as corresponding to traveling 1 mile every school
to more than .15 standard deviations in the distribution of utilities from schools for each applicant. Demand changes thus undermine the welfare performance of the initial match that ignores demand changes.

The large welfare cost of ignoring demand changes motivates me to investigate ways to alleviate the cost by accommodating demand changes. As already explained, NYC runs a discretionary, human-driven reapplication process. It is also possible to run a centralized algorithm for the reapplication process. I build a dynamic version of the school-student assignment model, analyze centralized designs of the reapplication process, and show that the centralized reapplication processes are the “best possible” mechanisms to accommodate choice changes. I evaluate how well the discretionary and centralized reapplication processes alleviate the welfare cost of ignoring demand changes. I find that both types of reapplication processes produce welfare gains, but the centralized reapplication processes are more effective and produce gains more than twice as large as those from the discretionary process.

This evaluation of reapplication processes takes estimated reapplication costs as given. There are technological changes and school districts’ and social entrepreneurs’ initiatives that may ease reapplication costs (e.g., online systems for more easily making and updating school choices). To measure the potential effects of such demand-side changes or interventions, I finally investigate how the performance of reapplication processes depends on reapplication costs and the resulting demand-side inertia. I find that the gains from the centralized reapplication mechanisms change by several times depending on the extent of demand-side inertia, which governs how much demand changes are revealed in reapplications.

These findings show that learning causes significant demand changes, which in turn undermine the welfare performance of the initial match and result in the large welfare cost of ignoring demand changes. Dynamic reapplications processes, especially centralized ones, help alleviate the welfare loss by accommodating changing demand. In addition, the gains from the mechanisms substantially change depending on the extent of demand-side inertia caused by reapplication costs. Thus, in the dynamic real world, the gains from a centralized market depend not only on its design but also on demand-side frictions such as demand changes (arising from learning) and inertia (caused by reapplication costs). This sheds empirical light on the potential importance of demand-side interventions that attempt to alleviate these frictions (e.g., applications for more easily searching school characteristics, online systems for more easily updating school choices).

5 Many school districts and social entrepreneurs have been launching such initiatives. For example, see http://www.dnainfo.com/new-york/20131113/washington-heights/six-apps-launch-guide-families-through-high-school-admissions-process and http://izonenyc.org/initiatives/innovate-nyc-schools/#scdc. However, aside from a few field
Related Literature. This paper is a first empirical study on the welfare performance of a dynamic centralized matching market. For that purpose, I develop an empirical model of evolving demand for schools under learning, reapplication costs, and demand changes in response to prior assignments, which have been studied in labor and public economics and industrial organization. I propose a novel approach to identify and estimate the different model elements using institutional features of centralized school choice systems. Finally, I combine the estimated model with a theoretical analysis of dynamic centralized matching markets to conduct welfare analysis.

More specifically, this paper combines and contributes to four different strands of the literature. First, the descriptive analysis of evolving school choices uses a revealed preference idea and relates to the economics and psychology literatures on revealed preference changes in the field, e.g., see papers reviewed in Blundell (1988) and Varian (2006) as well as more recent papers including Abaluck and Gruber (2011), Echenique et al. (2011), and Choi et al. (2014). These papers mainly focus on econometric difficulties in the detection and interpretation of preference changes.6

Second, my analysis suggests that school choice changes appear to be primarily caused by frictions in initial choices and learning about schools. This suggested importance of frictions in school choices echoes existing studies such as Hastings and Weinstein (2008), Jochim et al. (2014), Andrabi et al. (2015), Wiswall and Zafar (2015), and Hastings et al. (2015).7 Similar findings are also present in non-education contexts (Fang et al., 2008; Kling et al., 2012; Handel and Kolstad, 2015). Unlike these studies, I study the dynamics of frictions and learning within the same applicant or family. In this respect, this paper relates to non-education papers like Farber and Gibbons (1996), Ketcham et al. (2012), and Ketcham et al. (2015) on learning.

Third, my empirical model is a model of dynamic school choice with frictions in initial choices, reapplication costs, and demand responses to initial assignments. My model is thus at the intersection of empirical school choice models, demand models with frictions,

6I am more interested in welfare consequences of demand changes in dynamic centralized markets. To my knowledge, few papers investigate any pervasive effects of changes during administrative time delays. An exception is Autor et al. (2015), who study the effect of administrative decision time on the labor force participation and earnings of disability insurance applicants. Also, existing studies on decentralized matching markets emphasize the importance of demand changes in their discussion of unravelling. See Roth (2002) for an overview.

7NYC’s former deputy director of high school enrollment also points to potential frictions in the school choice process: “Given how massive the New York City process is, (...) the process by which those choices are made remains complicated, and very much depends on expertise or the ability to spend an excessive amount of time understanding how it works. Many students still go without either.” (http://ny.chalkbeat.org/2015/08/07/why-high-school-admissions-actually-doesnt-work-for-many-city-students-and-how-it-could/ #.VdP8ZHj5q20)

Finally, this paper’s theoretical analysis of centralized dynamic reapplication processes relates to the literature on the design of school-student assignment mechanisms, initiated by Abdulkadiroglu and Sönmez (2003), building on the classic two-sided matching problem (Roth and Sotomayor, 1990). Their model is extended and applied by Abdulkadiroğlu et al. (2009) to analyze the NYC institution, but both papers consider only static models and avoid dynamic considerations. By contrast, I use a dynamic model to take demand changes into considerations. In this sense, this paper’s theoretical analysis shares some attributes with recent papers on dynamic aspects of matching market design, e.g., Ünver (2010), Pereyra (2013), Dur and Kesten (2014), Kennes et al. (2014), Kurino (2014), Anderson et al. (2015), Akbarpour et al. (2015), Baccara et al. (2015), Leshno (2015), and Kadam and Kotowski (2015). This paper differs, however, because these studies exclude demand changes or use additional structures (e.g., binary preferences) to analyze their applications, which makes it difficult to apply them to analyze this paper’s problem. More importantly, none of the above theoretical papers connects theory to data.

1.2 A First Look at Evolving Choices

1.2.1 Evolving School Choices in NYC

I start by documenting how families’ school choices evolve over time. My analysis uses administrative panel data from the public high school choice system in NYC for the 2004-5 school year. This system contains more than 700 high school programs of various types across Greater New York. Some schools are academically selective, while others put emphasis on the arts. 8th (and some 9th) graders living in NYC may apply to these

---

8My model is also related to demand models with learning (see Ching et al. 2013 for a review), but different in that these learning models usually consider forward-looking consumers who try to learn the quality of frequently-used products (e.g., detergents) by experimenting with multiple products. This learning-by-experimentation aspect seems secondary in my context of education, where it is not easy to switch from one school to another.
schools. Each year, about 90,000 families apply, and most of them are admitted by some school. This system has been organizing applications and selections via the following centralized procedure:

(1) Each applicant ranks up to 12 schools in the order of her preference.

(2) Each school ranks applicants using its preference or priority as well as lottery numbers.

(3) NYC runs a strategy-proof algorithm (the “deferred acceptance” algorithm) on applicants’ and schools’ preferences to make an initial assignment of applicants to schools.\(^9\)

On top of this initial match, the system has an additional reapplication process. After being informed of the initially assigned school, any applicant is allowed to reapply against the assigned school if she is not satisfied with it. A reapplicant needs to fill out a paper reapplication form with a written reason for reapplying and turn it in to the guidance counselor at her middle school. In the reapplication, she is asked to rank up to 3 other schools she currently prefers over the initial assignment. She can rank the same schools as in the initial application. The initial assignment is guaranteed, i.e., if her reapplication is rejected, she is assigned her initially assigned school.

The timeline of initial applications and reapplications is available in Figure 1-1. Applicants make initial applications during November and December. After the announcement of the initial assignment, some families file reapplications during April and May. For reapplicants, the time interval between their initial application dates and reapplication dates are of mean 153.8 days and standard deviation 7.7 days (Appendix Figure 1.6.8). Thanks to the reapplication process, for those who reapply, I observe their school choices at two different points in time, which allows me to investigate how their demand for schools evolves.\(^10\)

About 7% (6430 applicants) of 91289 applicants reapply (Table 1.1). NYC accepts and reassigns 21% of reapplications to other schools in a discretionary, human-driven decision-making process.

\(^9\)The details of the algorithm will be explained in Appendix 1.6.4.

\(^10\)After the initial application process described below, there is the “supplementary round” for students who are not matched in the initial match. I exclude the supplementary round from the analysis because the supplementary round lets students rank only schools they do not rank in the initial application process, which makes it impossible to observe any clear choice changes or reversals between the initial application process and the supplementary round. In addition, the separate system that NYC uses for allocating seats in selective “specialized programs” or exam schools, is also outside the scope of my analysis because the system does not provide information about dynamic choice changes. I also exclude from my analysis those applicants who enroll in schools other than their initial assignments through over-the-counter bargaining, because I observe little information about it. Finally, my description below is for school year 2004-5 and parts of it may not be applicable to the current institution. Nevertheless, NYC keeps using similar discretionary reapplication processes even in recent years. See http://insideschools.org/blog/item/1000904-kids-win-one-third-of-hs-appeals#.
reapplication process. To measure choice changes, I say an applicant *exhibits choice reversals* if she reappears against her initially assigned school $s$ by ranking another school that is ranked below $s$ or unranked in her initial application. Among those who reapply, 71% (4564 applicants) exhibit choice reversals. This number is about 5% of the whole population.

Crucially, most choice reversals can be rationalized only by intrinsic demand or preference changes. I say an applicant *exhibits surely nonstrategic choice reversals* if she exhibits choice reversals and ranks 11 or fewer schools in her initial application, even though she could have ranked up to 12 schools. Surely nonstrategic choice reversals are consistent with optimal behavior only if there are intrinsic demand changes between the first round and the reapplication process. To see this, suppose to the contrary that an applicant exhibits surely nonstrategic choice reversals but does not experience demand changes. Let $s$ be her initial assignment and $t$ be any other school that (1) she ranks in her reapplication, but (2) she ranks below $s$ or does not rank in her initial application. If she prefers $s$ to $t$, then she would be better off by dropping $t$ from her reapplication. If she prefers $t$ to $s$, then she would gain by ranking $t$ ahead of $s$ in her initial application: The deferred acceptance algorithm in the initial application process is known to be strategy-proof for applicants and guarantees this property (Abdulkadiroğlu and Sönmez, 2003). Thus, surely nonstrategic choice reversals can be rationalized only by real demand changes. Table 1.1 shows that about 80% of choice reversals are surely nonstrategic. This fact suggests that most choice reversals in the reapplication process reflect real demand changes rather than strategic behavior.

### 1.2.2 Choice Frictions and Learning

Characteristics of all applicants, reapplicants, reapplicants who exhibit choice reversals are in Table 1.2. Those who reapply (and exhibit choice reversals) look similar to the average applicant, though the former is slightly more likely to be a female 8th grader and have lower test scores. Why do these similar looking applicants reapply? There are many potential reasons, such as mistakes in initial applications, changes in the life situation (e.g., moving), changes in the information about schools, changes in intrinsic tastes and preferences, and peer effects related to which schools siblings, friends, and bullies are assigned.

To understand the relative importance of these factors, Figure 1-2 provides a break-

---

11 Note that I do not include choice reversals among other schools than the initially assigned school $s$. For example, there are cases where an applicant prefers $t(\neq s)$ to $u(\neq s)$ in the initial application, but prefers $u$ to $t$ in the reapplication. I ignore these cases to make my calculation conservative.

12 As another look at this fact, Appendix Figure 1.6.9b shows that many applicants reapply after being assigned to their top choices. Also, Appendix Figure 1.6.9a shows that the first choice market shares of schools change from the first round to the reapplication process, where the first choice market share of a school is the fraction of applicants who rank it first among all applicants who make a first choice.
down of reapplication reasons self-reported by reapplicants. Panel (a) shows that the vast majority of reapplicants claim that they reapply because of new information or learning about school characteristics or their preferences about school characteristics. For example, many families claim that they were not aware of how far away the initially assigned school is or how painful it is to travel to the initial assigned school. Only a small fraction (less than 5% each) ascribes their reapplications to other potentially important factors, such as mistakes in initial applications and moving after initial applications. This provides suggestive evidence that the main factor for evolving demand is frictions in the initial choice process and learning about schools.\textsuperscript{13}

Panel (b) provides a further breakdown of the largest category of new information. This further breakdown shows that a variety of observable and unobservable school characteristics matter. Nevertheless, only a tiny fraction of reapplicants claim that they reapply because they do or do not want to enter the same school as particular siblings or friends or bullies; as far as reapplication decisions are concerned, peer effects do not seem quantitatively important.

Consistent with their self reports, families’ choices become more correlated with and responsive to school characteristics in reapplications. Table 1.3 documents that, compared with initial applications, reapplications rank schools that are more than 20% closer to families’ homes. These distance reductions come without sacrificing academic achievement level, as shown in lower rows. This pattern holds across demographic groups (Appendix Table 1.6.11). In Table 1.4, to incorporate other horizontal characteristics, I run the following descriptive regression:

\[
y'_{s} = b'X_{s} + e'_{s},
\]

where \(y'_{s}\) is the first choice market share of school \(s\) in round \(t\), which is the first round or the reapplication process, where the \textit{first choice market share} of a school is the fraction of applicants who rank it first among all applicants who make a first choice. \(X_{s}\) is a vector of observable characteristics of school \(s\), which do not change from the first round to the reapplication process.

Table 1.4 shows \(R^{2}\)’s from the above regression for the first round and the reapplication process. Across various specifications of \(X_{s}\), \(R^{2}\) is always higher in the reapplication process; reapplications appear to be more attentive or responsive to observable school characteristics than initial applications do. I checked to ensure that this pattern is robust to many specifications with different school characteristics and their interactions. The \(R^{2}\) increase is almost always present across demographic groups defined by baseline test

\textsuperscript{13}Since the reapplication process is discretionary, some of the self-reported reapplication reasons may be contaminated by strategic reporting. However, there seems to be no clear reason to expect that strategic reporting overstates the new information category because other reasons, such as moving and mistakes, sound more legitimate.
scores and race (Appendix Table 1.6.10). Note that Tables 1.3 and 1.4 and Figure 1-2 use mutually exclusive aspects of the data: While Figure 1-2 classifies self-reported reasons for reapplications, Tables 1.3 and 1.4 correlate families' school choice behavior with observable school characteristics. Tables 1.3 and 1.4 thus provide another independent support for the possibility that families become more informed of observable school characteristics or their preferences about them as time goes by.

The above descriptive analysis documents that a significant fraction of families change their school choices mainly because of learning about schools. This analysis has several limitations, however. I cannot extrapolate the suggestive findings on learning (Tables 1.3 and 1.4 and Figure 1-2a) to the whole population since they are based on self-selecting reapplicants. More importantly, the descriptive analysis depends entirely on observed choice changes, but observed choice changes may underestimate changes in latent demand. In particular, for families who do not reapply, their behavior does not directly reveal demand changes. However, some families may experience demand changes but not reapply because of "switching costs" or reapplication costs, in which case demand changes exist but are not directly observed. For example, the time cost of filling out a paper application form may constitute a reapplication cost.

To distinguish these scenarios and recover underlying evolving demand, it is necessary to model how learning and the resulting demand changes do or do not come to the surface as observed choice changes in the presence of potential reapplication costs. I next integrate key pieces of the descriptive analysis into a structural model of dynamic school choice with learning and reapplication costs.

1.3 Uncovering Evolving Demand

1.3.1 Dynamic School Choice under Learning and Switching Costs

To recover underlying evolving demand that is not necessarily reflected in observed choice behavior, this section develops a structural model of dynamic school choice. The model incorporates demand changes by learning and reapplication costs. In addition, the model allows demand to change in response to initial assignments in the first round. Specifically, I allow the utility of the first-round assignment to change by a positive or negative amount, which can be heterogenous across applicants. Such demand responses are likely if families experience psychological endowment effects about their initial assignments or get more information about them.

\[1^{14}\] These $R^2$ increases may be trivial if initial applications shares are more dispersed. However, the standard deviation of market shares changes little between the two periods (0.0028 for initial application shares and 0.0027 for reapplication shares).
Demand Before and After Learning. The random utility of school $s$ for applicant $a$ in period 0 (the initial application process) is

$$U_{as}^0 = U_s^0 + \sum_{k=1}^K \beta_{ak}(1 + f_{ak})X_{ask} + \epsilon_{as}^0,$$

where $U_s^0$ is a school-specific effect, $X_{as} \equiv (X_{ask})_{k=1,...,K}$ is a vector of (interactions of) $a$’s and $s$’s observable characteristics (e.g., the distance between $a$’s and $s$’s locations), $\beta_a \equiv (\beta_{ak})_{k=1,...,K}$ is a vector of preference coefficients, and $\epsilon_{as}^0$ is an unobserved utility shock.

$f_a \equiv (f_{ak})_{k=1,...,K}$ is the only non-standard term and stands for frictions $a$ faces about how to value characteristics $X_{ask}$ in the initial application process. I interpret each $\beta_{ak}(1 + f_{ak})X_{ask}$ as $a$’s perceived valuation of $X_{ask}$ in $t = 0$. $f_{ak}$ can be positive or negative and heterogeneous across different characteristics. Two interpretations of this specification are possible. The first interpretation is that $f_{ak}$ is frictions about preferences $\beta_{ak}$ and an applicant may not know her preferences about characteristics $X_{ask}$, e.g., how painful it is to travel a certain distance. The alternative interpretation is that $f_{ak}$ is frictions about characteristics $X_{ask}$ and an applicant may not know $X_{ask}$, e.g., the distance to schools. Both interpretations and their combinations are consistent with the descriptive analysis and result in the same welfare implications below. I prefer the first interpretation; see “Discussions on Modeling Decisions” at the end of this section for an additional discussion. The modeling of the friction is motivated by Figure 1-2a and Tables 1.3 and 1.4, which suggest that applicants face frictions about observable school characteristics or their preferences about school characteristics in the initial application process. I assume that each applicant $a$’s initial preference $>_a s$ is based on perceived utilities $U_{as}^0$’s subject to frictions, i.e., $s >_a s'$ only if $U_{as}^0 > U_{as'}^0$.\footnote{I assume $s >_a s'$ for any ranked school $s$ and unranked school $s'$. See “Discussions on Modeling Decisions” at the end of this section for discussions about this truth-telling assumption.}

After the initial application, NYC runs the (applicant-proposing) deferred acceptance algorithm to give an initially assigned school $s_a^0$ to each applicant $a$. During and after the match-making process, applicants’ perceived utilities change. The random utility of school $s$ for applicant $a$ in period 1 (the reapplication process) is

$$\tilde{U}_{as}^1 = \frac{U_{s}^1 + \beta_aX_{as} + \epsilon_{as}^1}{\equiv U_{as}^1} + \gamma_a1\{s = s_a^0\}.$$ 

The first three terms are similar to those in initial utilities $U_{as}^0$ except that the frictions are normalized to zero, and school-specific effects $U_{s}^1 \equiv U_{s}^0 + U_{s}$ and unobserved utility \footnote{In reality, some of the frictions are likely to remain even in the reapplication process. I need to assume it away as a normalization, however, since the data contains only two periods.}
shocks $\epsilon_{as}^1 \equiv \epsilon_{as}^0 + \epsilon_{as}$ are subject to new unobserved shocks $U_s$ and $\epsilon_{as}$, respectively. I allow $U_s$ and $\epsilon_{as}^1$ to differ from $U_s^0$ and $\epsilon_{as}^0$, respectively, to accommodate the fact that demand changes are sometimes related to unobserved school characteristics such as how nice current students are (Figure 1-2b). As a result of this specification, unobserved utility shocks $\epsilon_{as}^0$ and $\epsilon_{as}^1$ are serially correlated for each applicant $a$. This is reasonable given the interpretation of unobserved utility shocks $\epsilon_{as}$ as the sum of unobserved utility components in period $t$ and that the unobserved determinants of utilities for an applicant are likely to be serially correlated.

The last term $\gamma_a 1\{s = s_a^0\}$, which is turned on if and only if school $s$ is applicant $a$’s initial assignment $s_a^0$, captures the possibility that an applicant’s utility from the initially assigned school may evolve differently than utilities from other schools do. For example, applicants may get more information about the initially assigned school than they would with other schools. Or they may begin to prefer the assigned school more because it admits them, or they get used to it (habit formation or endowment effects). I call $\gamma_a$ the initial assignment effect.

**Model of Reapplications 1: Rational Expectation.** Each applicant decides whether to reapply based on how preferable the initial assignment $s_a^0$ is with respect to new demand $\tilde{U}_{as}^1$’s. Recall that there are factors that may prevent applicants from reapplying, for example, the time cost of making and submitting a reapplication. In fact, each reapplicant needs to fill out a paper reapplication form with a written reason for reapplying, and turn it in to the guidance counselor at her middle school. I consider two models to incorporate such “reapplication costs” or “switching costs”. I call these models the rational expectation model and the naive free expectation model. The rational expectation model allows for school-specific reapplication acceptance probabilities, but assumes “rational” expectation about applicants’ expectations about reapplication acceptance probabilities. The naive free expectation model does not need the rational expectation assumption, but assumes that reapplicants form simplistic beliefs about how the reapplication process works.

The first rational expectation model consists of two layers. The first layer is about reapplication acceptance probabilities. In the reapplication process, each applicant $a$ who reapplies with new preference $(s_1, s_2, s_3)$ is re-assigned to at most one of schools $s_1, s_2, s_3$. Since there is no accurate algorithmic description of discretionary reapplication acceptance decisions by NYC, I suppose that the (mutually exclusive) re-assignment probabilities can be approximated by the following descriptive model. For each $i = 1, 2, 3,$

$$\Pr(a \text{ is re-assigned to } s_i) = \begin{cases} 
\Pr(b_0 + b_1X_{as_i} + b_2W_{as_i} + \xi_{as_i} \geq 0) & \text{if } s_i \neq \emptyset \\
0 & \text{if } s_i = \emptyset,
\end{cases}$$

22
where \( X_{as} \) is the characteristics of \( a \) and \( s \) used in the utility model, and \( \xi_{as} \sim iid \ EV(I) \) (logit) with usual variance normalization to \( \pi^2/6 \). (Results from a probit version are similar.) \( W_{as} \) contains additional factors that may affect reapplication acceptance decisions by NYC: a measure of how oversubscribed or popular \( s \) is (the number of applicants rejected by \( s \) in the initial application process), an indicator that \( a \) ranks \( s \) in the initial application, and another indicator that \( a \) is rejected by \( s \) in the initial application process. Let \( p_{as_i} \) be the estimate of \( \Pr(a \text{ is re-assigned to } s_i) \) I obtain by applying the above model to the reapplication acceptance data.

The second step consists of applicants’ reapplication decisions given acceptance probabilities in the first step. I assume that applicant \( a \) reapplyes if there is a combination of schools \((s_1, s_2, s_3)\) such that (a) for all \( i = 1, 2, 3, \) \( s_i \) is a school other than \( s^0_a \) or empty, (b) \( s_i \neq s_j \) or \( s_i = s_j = \emptyset \) for all \( i \neq j \), and (c)

\[
\frac{\sum_{i=1}^{3} p_{asi} U^1_{asi_i}}{\text{expected benefit from reapplying}} + (1 - \sum_{i=1}^{3} p_{asi_i}) U^1_{as0} - U^1_{as0} > \frac{c_a}{\text{cost}}
\]

\[
\Leftrightarrow \sum_{i=1}^{3} p_{asi} U^1_{asi_i} > \sum_{i=1}^{3} p_{asi_i} U^1_{as0} - \sum_{i=1}^{3} p_{asi_i} U^1_{as0} + c_a + \gamma_a \sum_{i=1}^{3} p_{asi_i} = c_a
\]

\[
\Leftrightarrow \sum_{i=1}^{3} p_{asi} U^1_{asi_i} > \sum_{i=1}^{3} p_{asi_i} U^1_{as0} + \bar{c}_a,
\]

where \( c_a \) is the reapplication cost.\(^{17}\) This model imposes the rational or sophisticated expectation assumption that each applicant believes that she is accepted by schools \( s_1, s_2, \) and \( s_3 \) with mutually exclusive probabilities \( p_{as1}, p_{as2}, \) and \( p_{as3} \), respectively. This assumption is unavoidable since there seems to be no way to identify subjective \( p_{as_i} \) for each \((a, s_i)\) pair. The condition for reapplying can be written as in the last line of (3), a discrete choice with switching costs \( \bar{c}_a \). I use “switching costs” to mean such combinations of reapplication costs and initial assignment effects, which I will separately identify.

If reapplying, applicant \( a \) ranks schools \((s_1, s_2, s_3)\) to maximize the expected benefit in the left hand side of (3), i.e., schools with largest \( p_{as} U^1_{as} \). I also assume that if applicant \( a \) reappplies but does not exhaust her new preference list, i.e., \( a \) ranks less than three schools in \( \succ^1_a \), any unranked school is less preferred to the guaranteed initial assignment \( s^0_a \) in \( U^1_{as} \).

The above reapplication acceptance model has limitations. For example, ideally, I would let \( b_i \) be heterogeneous across schools or applicants. However, this is infeasible because there are more than 700 schools, while only about 6000 applicants reapply, and each reapplicant ranks at most three schools in the reapplication. Instead, I include rich attributes of applicants and schools in \( X_{as} \) and \( W_{as} \). I also have to exclude the effects of \( s' \neq s \) on the probability that applicant \( a \)'s reapplication for school \( s \) is accepted. If accep-

\(^{17}\) A small fraction of applicants are not assigned to any school in the initial match. The above model is not well-defined for these unassigned applicants since \( s^0_a = \emptyset \) for them. For them, I assume the following model for the utility of the outside option \( \emptyset \): \( U^1_{as} = U^1_0 + \epsilon_{as} \) where \( U^1_0 \) is the outside-option-specific constant and \( \epsilon_{as} \) is an unobserved utility shock. I make the same assumption for the second model below.
tance probability $p_{as}$ depends not only on school $s_i$ but also on $s_j$, then the maximizer of the expected benefit of reapplying is not necessarily schools with largest $p_{as}^1 U_{as}^1$. To find the maximizer, I need to search over all possible combinations of up to 3 schools. Since there are more than 700 schools, the number of such combinations is prohibitively large, making estimation intractable. Due to these difficulties, I resort to the above simplified model.\footnote{In general, the above model can cause internal inconsistencies $\Sigma_{i=1}^3 p_{as_i} > 1$. In my data, however, reapplication acceptances are rare, and estimated $p_{as_i}$ is almost always less than 0.2. As a result, $\Sigma_{i=1}^3 p_{as_i} > 1$ never happens.}

**Model of Reapplications 2: Naive Free Expectation.** The main concern with the rational expectation model is that it imposes rational expectations. To deal with this issue, I also consider an alternative model that does not assume rational expectations, but instead assumes naive beliefs about how the reapplication process works. In the alternative model, applicant $a$ does not reapply if

$$\frac{p_a \left( \max_{s \neq s_0} \bar{U}_{as}^1 - \bar{U}_{as_0}^1 \right)}{\text{cost}} < c_a (> 0)$$

$$\Leftrightarrow \max_{s \neq s_0} U_{as}^1 - U_{as_0}^1 < c_a / p_a + \gamma_a \equiv \tilde{c}_a$$

$$\Leftrightarrow U_{as_0}^1 + \tilde{c}_a > U_{as}^1$$

for any $s \neq s_0^0$, (4)

where $c_a$ is the reapplication cost and $p_a$ is $a$’s subjective probability that $a$’s reapplication is accepted. I do not assume $p_a$ to be the same as the real reapplication acceptance probability.

This model imposes a simplifying assumption that the expected benefit from reapplying is expressed as $p_a$ times the utility difference between the initial assignment and the new most preferred school. If $s_0^0$ is $a$’s most preferred school in $\bar{U}_{as}$, i.e., $\max_s \bar{U}_{as}^1 = \bar{U}_{as_0}^1$, then the left hand side of the second line of (4) is negative and so $a$ never reapplies. Otherwise, $a$ reapplies when the expected benefit from doing so exceeds the reapplication cost $c_a$. Again, the condition can be written as in the last line of (4), a discrete choice with switching costs $\tilde{c}_a$.

For those who reapply, I observe new rank-ordered preference $\succ_{a}^1$ and assume that each reapplicant submits $\succ_{a}^1$ based on $\bar{U}_{as}$’s, i.e., $s \succ_{a}^1 s'$ only if $\bar{U}_{as} > \bar{U}_{as'}$. As in the rational expectation model, I also assume that if $a$ reappplies but does not exhaust her new preference list, i.e., applicant $a$ ranks less than three schools in $\succ_{a}^1$, any unranked school is less preferred to the guaranteed initial assignment $s_0^0$ in $\bar{U}_{as_0}^1$.

Comparing the two models, the rational expectation model allows for school-specific reapplication acceptance probabilities, while assuming rational expectation. The naive
free expectation model does not need the rational expectation assumption, but assumes that reapplicants form simplistic beliefs about how the reapplication process works. These two models are thus expected to be complementary and serve as robustness checks for each other. I estimate both models and show that the key results hold under both models. Before moving on to identification and estimation, however, I need to discuss other important modeling decisions.

Discussions of Modeling Assumptions

Choice Frictions and Learning. The key modeling decision about the evolving utility model is how to model frictions in initial choices. My model specifies them as \( \beta_{ak}(1 + f_{ak})X_{ask} \). Ideally, I would like to make frictions \( f_{ak} \) more flexible, for example, \( X_{ask} \) that is heterogeneous not only across applicants \( a \) and characteristics \( k \), but also across schools \( s \). Alternatively, an additive specification \( \beta_{ak}(X_{ask} + f_{ask}) \) may be another more flexible way to model frictions. In such more flexible models, \( \beta_{ak}(1 + f_{ask}) \) or \( \beta_{ak}(1 + \frac{f_{ask}}{X_{ask}}) \) (note that \( \beta_{ak}(X_{ask} + f_{ask}) = \beta_{ak}(1 + \frac{f_{ask}}{X_{ask}})X_{ask} \)) performs the same role as the taste coefficient on \( X_{ask} \) in usual discrete choice models with no frictions. However, it is unclear how I can identify the distribution of such coefficients that depend not only on \( a \) but also on \( s \). For example, consider a static discrete choice model with no frictions \( U_{as} = \beta_{as}d_{as} + \epsilon_{as} \) where \( d_{as} \) is the distance, which is always positive. For each \( a \), consider any preference coefficients \( (\beta_{as}) \) such that (1) each \( \beta_{as} \) is so large that the effect of \( \epsilon_{as} \) on choice probabilities is negligible and (2) \( \beta_{as1} > \beta_{as2} > \ldots \), where \( s_i \) is \( a \)'s observed \( i \)-th choice. Any such \( (\beta_{as}) \) can rationalize \( a \)'s observed choices, making identification impossible without a particular parametric assumption. To avoid the potential lack of identification, I make \( f_{ak} \) independent from \( s \).\(^{19}\) Even under this restriction, \( f_{ak} \) allows for rich heterogeneity across \( a \) and \( k \).

Yet another potential way to model frictions and learning is to introduce "consideration sets", i.e., subsets of schools applicants consider when they make initial choices. See Goeree (2008) for an existing empirical model of consideration sets. There is not enough variation in my data to allow for both frictions \( f_{ak} \) and consideration sets. Given a choice between frictions and consideration sets, I prefer frictions for several reasons. First, consistent with the friction specification, self-reported reasons for reapplications mention the initial lack of knowledge about school characteristics or their preferences about school characteristics more often than the initial lack of knowledge about the presence of particular schools. Second, for inferring a consideration-set-formation process, I need some variation that makes different schools more or less likely to enter consideration.

\(^{19}\)This discussion relates to recent attempts to identify and estimate discrete choice models with measurement errors, since frictions about \( X_{ask} \) or \( \beta_{ak} \) can be reinterpreted as measurement errors in \( X_{ask} \). See Hu (2015) for a review.
sets. However, it is unclear if time-series variation in my data is enough, since the contrast between initial applications and reapplications contains no variation across schools. For these reasons, this paper focuses on frictions $f_{ak}$, and I leave a consideration-set approach for future research.

**Deliberate Reapplication Decisions.** My reapplication model assumes that families make reapplication decisions by deliberately comparing their initial assignments with other schools according to their new demand. A potential concern is that reapplication decisions may be primarily driven by less systematic factors (e.g., inattention unrelated to initial assignments). However, there are several descriptive facts showing that, consistent with my model, families' reapplication behavior responds to the desirability of their initial assignments. For example, the more preferred school an applicant is initially assigned, the less likely she is to reapply (Figure 1-4); this correlation is always present across many subgroups defined by demographic characteristics and first-round application behavior (Appendix Figure 1.6.12). The next identification section will detail these facts. Appendix 1.6.3 further shows that this correlation is causal and structural. This suggests that many applicants, including those who do not reapply, compare their initial assignments and other schools.

In addition, the amount of choice reversals reapplicants exhibit is strongly correlated with the preference rank of their initially assigned school (Figure 1-3). These correlations are also consistent with my model. In my framework, the more preferred school an applicant is initially assigned, the smaller the expected benefit of reapplying is for her. Thus, those assigned to more preferred initial assignments need to experience larger demand changes to find it worth reapplying, compared with those assigned to less preferred initial assignments. As a result, conditional on reapplying, the amount of observed choice reversals, which reflect demand changes, should be decreasing in the preference rank of the initially assigned school, implying a pattern as in Figure 1-3.

These facts suggest that many families behave in ways consistent with deliberate reapplication decisions. If some people behave according to inattention unrelated to initial assignments, their behavior is likely to be absorbed by reapplication costs in my model. This misspecification concern is common in empirical studies on switching costs.

**Truthful Behavior in Initial Applications.** The model assumes that each applicant makes the initial preference $\succ_0^a$ as a non-strategic rank-ordered discrete choice based on old utilities $U_{as}^0$'s. This should be a reasonable assumption since (1) a majority of applicants (more than 70%) do not exhaust preference lists and rank 11 or fewer schools, and (2) the deferred acceptance algorithm used in the initial application process is strategy-proof for applicants and guarantees that the above truthful behavior is always optimal for any applicant who does not exhaust her preference list. Even for those who exhaust their preference lists, the deferred acceptance algorithm makes it always optimal for any
applicant to truthfully report her relative preference order over ranked schools.

The above discussion ignores the presence of the reapplication process. In principle, applicants may strategize in initial applications for switching to a more preferred school in the reapplication process. However, such strategic behavior is unlikely to benefit reapplicants: The reapplication acceptance rate is low (21%), and it is rare that reapplicants can switch to more preferred schools. Also, the reapplication process is a discretionary process with no algorithmic rule. It is thus unclear how to strategize in initial applications to benefit in the reapplication process. For example, one may suspect that in the reapplication process, it may be easier to be transferred to a school that is not ranked in initial applications, making it profitable to strategically drop some schools from initial applications. However, in the data, reapplication acceptances are more likely to be given to schools ranked in the first round.

Finally, there is an additional tractability consideration that forces me to ignore potential strategic behavior. With strategic behavior, I need to consider applicants' choices over combinations (lists) of schools, but the number of such combinations is prohibitively large in my setting with hundreds of schools. In light of these computational and conceptual reasons, I assume away strategic behavior in initial applications. Existing studies also use similar truth-telling assumptions (Hastings et al., 2008; Ajayi, 2013; Abdulkadiroglu et al., 2015b).

Outside Option. I do not explicitly model the outside option because it is unclear whether the data is informative about the outside option. Many applicants do not exhaust their initial preference lists, but it does not necessarily mean that all unranked schools, which are almost all of NYC schools, are less preferred over the outside option for them. A more reasonable interpretation seems to be that they are optimistic and expect that they will be assigned to one of the ranked schools for sure; see Robbins (2011) for an article that reports about such optimistic families. In this scenario, the data does not provide any information about the comparison between the outside option and NYC schools. I thus refrain from modeling the outside option.

1.3.2 Identification

My model allows for preferences ($\beta_a$), frictions in initial choices ($f_a$), reapplication costs ($c_a$), and initial assignment effects ($\gamma_a$). This flexibility may create an identification concern, because it is often difficult to separately identify heterogenous preferences and switching costs (Chamberlain, 1983; Heckman, 1991). In fact, no other model reviewed in the related literature seems to allow for all of the above model components simultaneously. This section explains which aspects of my data allow me to distinguish the model components. For brevity, I focus on the above key parameters and ignore school-specific effects $U^t_s$ and unobserved utility shocks $\epsilon^{t,as}$. 27
By identification results for standard discrete choice models with no frictions (Matzkin, 2007; Manski, 2009; Berry and Haile, 2015; Fox et al., 2012), the data on initial preferences $\gamma^0$ identifies the distribution of $\beta_{ak}(1 + f_{ak})$: It is because in initial utilities $U^0_{as}$, the composite term $\beta_{ak}(1 + f_{ak})$ has the same role as the preference coefficient in standard models. However, it is not possible to separate out preferences $\beta_{a}$ and frictions $f_{a}$ by the initial application data alone: insensitivity of choices to a certain characteristic (e.g., academic performance) may be because of weak preferences for it or frictions about it.

I need to use the data on reapplications to distinguish frictions $f_{a}$ and preferences $\beta_{a}$. Let me use the term demand changes to mean utility changes by frictions about characteristics $X_{as}$, i.e., $f_{ak}\beta_{ak}X_{ask}$'s. In combination with already-identified $\beta_{ak}(1 + f_{ak})$, identification of demand changes is enough to separate out frictions $f_{a}$ and preferences $\beta_{a}$. The identification of demand changes is aided by the increase in the correlation between characteristics $X_{as}$ and choices induced by $U^t_{as}$ from $t = 0$ to $t = 1$ (recall Tables 1.3 and 1.4). The difficulty is that reapplication behavior is subject to switching costs ($\hat{c}_{a}$ in the rational expectation model and $\hat{c}_{a}$ in the simplistic free expectation model). It is usually hard to separate demand changes from switching costs, because the low rate of choice changes, which we observe in most settings including mine, is usually consistent both with small demand changes and large switching costs (Chamberlain, 1983; Heckman, 1991). The main identification challenge is thus how to separately identify demand changes and switching costs. Moreover, I need to resolve an additional difficulty of how to decompose switching costs into reapplication costs $c_{a}$ and initial assignment effects $\gamma_{a}$.

To overcome these challenges, I exploit the following three institutional features of centralized school choice systems, including NYC's.

1. **Capacity constraints.** Due to capacity constraints, many applicants are initially assigned to a school other than the most preferred school.

2. **Partially random initial assignments.** In the first-round assignment mechanism, to determine initially assigned schools $s^0_{a}$, the algorithm uses student and school preferences, but school preferences are coarse or weak in that a school’s preference is indifferent among many students. NYC draws random lottery numbers to break ties or indifferences, and uses the resulting strict school preferences to compute initial

\[\text{Identification results for discrete choice models with similar complications appear in, e.g., Agarwal and Somaini (2015).}\]
assignments. This use of admissions lotteries makes initial assignments partially random.

(3) Rank-ordered reapplication preferences. Reapplicants make new, rank-ordered school choices in their reapplications, as explained in section 1.2.1.

Demand Changes and Switching Costs. Intuitively, the identification logic is as follows. For simplicity, let me ignore demand responses to initial assignments and focus on demand changes and reapplication costs. Assume that there are only two preference ranks, the first choice and the lower choice. Many applicants are "lower-choice non-reapplicants," who are initially assigned to their old lower choice but do not reapply, due either to reapplication costs or demand changes. This fact allows me to measure the total effects of demand changes and reapplication costs. There are also "first-choice reapplicants," who are initially assigned to their old first choice but reapply, which must be due to demand changes. Other applicants assigned to their first choice may also experience demand changes but are locked in by reapplication costs. The fraction of first-choice reapplicants in all applicants assigned to the first choice thus tells me the difference between the effects of demand changes and switching costs.

Now suppose that admissions lotteries in the first-round mechanism guarantee that initial assignments are randomly assigned and that applicants assigned to the first-choice and lower-choice are comparable people with similar demand changes and reapplication costs. I can compare the fractions of lower-choice non-reapplicants and first-choice reapplicants to measure the amount of reapplication costs. Heuristically,

\[
\text{(fraction of non-reapplicants among applicants assigned to the initial lower choice)} - \text{(fraction of reapplicants among applicants assigned to the initial first choice)} = \text{(demand changes + reapplication costs)} - \text{(demand changes - reapplication costs)} = 2 \times \text{reapplication costs},
\]

which separates reapplication costs from demand changes.

More precisely, thanks to institutional feature 1 (capacity constraints), I can compute moments like Figure 1-4a, where the solid black line relates the preference rank (with respect to initial \( >^0 \)) of the initially assigned school \( s^0_a \) to the conditional probability of reapplying. The line is upward-sloping, i.e., the more preferred school an applicant is assigned, the less likely she is to reapply.

This moment turns out to contain information to separate demand changes and switching costs. To illustrate this, let me start with the question of what the line should look like if there were no demand changes (\( \sum_{k=1}^{K} \beta_{ak} f_{ak} X_{ask} = 0 \)) and no switching costs (\( \tilde{c}_a = 0 \) or \( \tilde{c}_a = 0 \)). The answer is the dotted red line in Figure 1-4a: If an applicant is assigned the old first choice, it remains to be the new first choice by the no-demand-change assumption, and there is no reason for her to reapply. If an applicant is assigned an old lower choice,
by the no-demand-change assumption, the initially assigned school remains to be different from the new first choice. Thus, she should reapply as long as there is no switching cost, since reapplying gives her a positive probability of being re-assigned to a more preferred school without risking the initial assignment. (The initial assignment is guaranteed even if an applicant reapplications, as explained in Section 1.2.1.)

This hypothetical dotted line under no demand changes and no switching costs is different from the real solid line. The discrepancy between the two has to be caused by demand changes or switching costs.

To distinguish demand changes and switching costs, I use institutional feature 2 (partially random initial assignments). For simplicity, start by assuming that initial assignments are purely random. After illustrating the identification logic in this simplified case, I explain how to extend the identification logic to the case with real partially random initial assignments. Let me consider the question of what the black solid line should look like under no switching costs but with potential demand changes. To answer this, in Figure 1-4b, consider the solid short red arrow above “1” on the x axis. Let \( R \) be the length of the solid red arrow.

\[
R \equiv \Pr(a \text{ reapplications}|s_a^0 \text{ is } a's \text{ old 1st choice}) = \sum_{K \geq 2} \Pr(a's \text{ old } K-th \text{ choice}=a's \text{ new 1st choice}| s_a^0 \text{ is } a's \text{ old 1st choice})
\]

(by the assumption that there are no switching costs)

\[
= \sum_{K \geq 2} \Pr(a's \text{ old } K-th \text{ choice}=a's \text{ new 1st choice}| s_a^0 \text{ is } a's \text{ old } K-th \text{ choice})
\]

(by random assignment of \( s_a^0 \))

\[
= \sum_{K \geq 2} \Pr(a \text{ does not reapply}| s_a^0 \text{ is } a's \text{ old } K-th \text{ choice})
\]

(by the assumption that there are no switching costs)

\[
\equiv B,
\]

where \( B \) is defined as the sum of the lengths of the dotted blue arrows in the figure. However, Figure 1-4b shows that the sum of the lengths of the dotted blue arrows \( B \) (9.510)

\[22\]This reasoning assumes that the probability of a reapplication acceptance is positive for all possible preference ranks of the initial assignment. Appendix Figure 1.6.11 confirms that this is the case in the data. Estimated \( p_{ais} \) for the rational expectation model is also always positive.

\[23\]This step assumes that the preference rank of the randomly assigned initial assignment \( s_a^0 \) does not have direct effects on demand changes \( \beta_{akfak} \). Analogous assumptions are implicitly made in existing studies of switching costs reviewed in the related literature section in the following sense. In general, the separation of switching costs from heterogeneous preferences needs two assumptions: (i) The status quo alternative as the source of switching costs is (at least partly or conditionally) randomly assigned. (ii) The status quo does not have direct effects on any demand or preference changes. Existing studies use the contrast between active vs passive choice periods or variation in supply-side advertisements to defend (i). In addition, they need (ii) and usually assume that all demand changes are exogenous. By contrast, my approach as outlined above uses a really random assignment to guarantee (i). In addition, while I need to assume that the preference rank of the randomly assigned initial assignment \( s_a^0 \) does not have direct effects on \( \beta_{akfak} \), I allow for a certain direct effect of \( s_a^0 \) on new demand through initial assignment effects \( \gamma_a 1(s = s_a^0) \).
is larger than the length of the solid red arrow $R$ (0.025). This means that, regardless of the amount of demand changes, the no-switching-cost assumption leads to a contradiction with the data. In contrast, if there are switching costs ($\bar{c}_a > 0$), the model’s requirement changes to

$$R \lessgtr \sum_{K \geq 2} \Pr(a's\ old\ K-th\ choice=a's\ new\ 1st\ choice| s^0_a\ is\ a's\ old\ 1st\ choice)$$

$$\lessgtr \sum_{K \geq 2} \Pr(a\ does\ not\ reapply| s^0_a\ is\ a's\ old\ K-th\ choice)$$

$$\equiv B,$$

where previous equalities change to inequalities because additional people stop reapplying because of positive switching costs. The inequalities are consistent with Figure 1-4b. In this way, the discrepancy between the solid red arrow $R$ (the probability of reapplying conditional on being initially assigned the old first choice) and the dotted blue arrows $B$ (the sum of the probabilities of not reapplying conditional on being assigned to old non-first choices) has to be driven by switching costs and not by demand changes, telling us the amount of switching costs.

Figures 1-4c and 1-4d summarize the separate identification of switching costs and demand changes. Starting from the solid black line, move it up until the point where the solid red arrow and the dotted blue arrows are balanced. The new dotted blue line with triangle markers describes such a point. The difference between the solid black line and the dotted blue line with triangle markers has to be due to switching costs, while the remaining difference between the two dotted lines is due to demand changes. Figure 1-4d thus suggests both significant demand changes and significant switching costs.

The above discussion shows that the probability of reapplying conditional on the preference rank of the initial assignment reveals the amount of demand changes and switching costs. Appendix Figures 1.6.12a and 1.6.12b use this logic to suggest that different demographic groups face different amounts of demand changes and switching costs. Racial minorities and academically struggling families exhibit flatter gradients, a sign of larger demand changes by learning.24 Given this, I let the distribution of demand changes and switching costs as well as the other parameters heterogeneous across demographic groups in the estimation. In contrast, I find little heterogeneity among students with different initial assignments or different types of initial application behavior (Appendix Figures 1.6.12c and 1.6.12d). This may suggest that conditional on the preference rank of the initially assigned school, its identity may not matter for the amount of switching costs and demand changes.

---

24 This sheds additional light on heterogeneity of school choice behavior across demographic groups. See Hastings et al. (2008) and Nathanson et al. (2013) among others for related findings.
Handling Partial Randomization. The above identification logic is under the simplifying assumption that initial assignments are completely randomly assigned. In the NYC school choice system, however, initial assignments are not purely random since they are confounded by non-random preferences of applicants and their priorities at schools. Nevertheless, the above identification logic extends to the more complicated real case with partially random initial assignments.

NYC creates initial assignments via the deferred acceptance algorithm that uses many inputs such as preferences, priorities, lottery numbers, and capacities (see Section 1.2 and Appendix 1.6.4). Depending on these factors, different applicants have different assignment probabilities at schools. Yet, there is a way to find a set of applicants who share the same assignment probability at any school. Let me refer to a student's entire preference list and priorities at all schools as her type. The deferred acceptance algorithm treats students of the same type symmetrically in that everyone of a given type faces the same probability of assignment to any school. This is because the only information about a student the algorithm uses is her preference, priorities, and lottery number; conditional on type, therefore, all that remains to determine her assignment is her lottery number, which is independently and identically distributed across students.

This gives me a way to extend the identification analysis in Figure 1-4 to the case with partially random initial assignments: I can simply repeat the same analysis conditional on each type to separately identify switching costs and demand changes for that type. Switching costs and demand changes are thus allowed to be heterogenous across different types without sacrificing identification. As far as identification is concerned, partial randomization does not cause any serious problem. The empirical implementation of the identification argument in Figure 1-4d is also robust to the explicit consideration of imperfectly random initial assignments. Appendix Figure 1.6.13 provides a structural/causal version of Figure 1-4d that incorporates partially random assignment.

Initial Assignment Effects and Reapplication Costs. Now that I have explained how to identify $\beta_a$ (preferences), $f_a$ (frictions in initial choices), and switching costs, the final step is to decompose switching costs into reapplication costs $c_a$ and initial assignment effects $\gamma_a$. Conditional on reapplying, reapplication costs $c_a$ are sunk and do not affect new preferences reported in reapplications. On the other hand, initial assignment effects $\gamma_a$ remain to affect new preferences since they directly enter the utility of the initial assignment $s_a^0$. I use this fact and institutional feature 3 (rank-ordered reapplication preferences) to derive the following two restrictions on $\gamma_a$: (1) If applicant $a$ reappplies to switch from initial assignment $s_a^0$ to $s$, then $U_{s_a}^1 > U_{s_a}^0 + \gamma_a$. (2) If applicant $a$ reappplies but does not exhaust her new preference, i.e., rank only one or two schools, then $U_{s_a}^1 + \gamma_a > U_{s}^1$ for every unranked school $s$. These restrictions involve initial assignment effects $\gamma_a$ but not reapplication costs $c_a$, allowing me to separate out initial assignment effects $\gamma_a$. New preferences contain a lot of information about initial assignment effects $\gamma_a$ since
every reapplicant prefers some schools over the initial assignment $s^0_a$, while about 40% of reapplicants do not exhaust new preferences, implying that they prefer the initial assignment $s^0_a$ over unranked schools.

1.3.3 Estimation

In the estimation, characteristics $X_{as}$ include those frequently mentioned in reapplication reasons: the road distance between applicant $a$’s and school $s$’s locations, school $s$’s academic performance, type, size, and age. See Appendix 1.6.1 for the construction of these variables. For computational tractability and finite sample statistical precision, I need to impose distributional assumptions common to empirical discrete choice models. Let $g_a$ be applicant $a$’s demographic group defined by whether $a$ is white/asian or black/hispanic and whether $a$’s grade 7 reading grade category is high/middle or low (four groups in total). I assume that

- the period 0 coefficient $\beta_{ak}(1 + f_{ak})$ on $X_{as}$ is iid according to
  \[ \begin{cases} 
  \log N(\mu_{0k}^{g_a}, \sigma_{0k}^{g_a}) & \text{for negative distance or high academic performance} \\
  N(\mu_{1k}^{g_a}, \sigma_{1k}^{g_a}) & \text{for any other characteristic}, 
  \end{cases} \]

- the change in the coefficient due to learning is $\beta_{ak}f_{ak} \sim iid N(\mu_{1k}^{g_a}, \sigma_{1k}^{g_a})$,

- $U_s^0 \sim iid N(0, \sigma_0^{g_a})$,

- $U_s \sim iid N(0, \sigma_1^{g_a})$,

- $\gamma_a \sim iid N(\mu_{1k}^{g_a}, \sigma_{1k}^{g_a})$,

- $c_a \sim iid truncated N(\mu_{c}^{g_a}, \sigma_{c}^{g_a})$ in the rational expectation model, and

- $c_a/p_a \sim iid truncated N(\mu_{c}^{g_a}, \sigma_{c}^{g_a})$ in the naive free expectation model.

Note that motivated by heterogeneity across demographic groups in key descriptive moments (Appendix Figure 1.6.12 and Appendix Table 1.6.10), the whole parameter vector is allowed to be heterogeneous across demographic groups. I assume $\epsilon_{as}^0, \epsilon_{as} \sim iid EV(I)$ (logit) with usual variance normalization to $\pi^2/6$.

These assumptions make the model within each period a random coefficient logit model, which allows for flexible substitution patterns among schools (Train 2009 chapter 25). It is computationally prohibitive to separately estimate $U_s^0$ for each school $s$ since the sample contains a large number of schools. Instead, I adopt this random effects specification in the spirit of Rossi et al. (1996) and Abdulkadiroğlu et al. (2015b).

It is possible and more desirable to estimate the variance of $\epsilon_{as}$ rather than assuming it. Computational difficulties prevent me from reporting results from such an extension in this draft, however. I verified that the results reported below are robust (i.e., changes are less than 10%) to ignoring $\epsilon_{as}$ and using only the other parts of demand changes.
6). Appendix 1.6.3 derives a partly analytical joint likelihood function for a sequence of initial application preferences in \( t = 0 \) to reapplication decisions and preferences in \( t = 1 \). This likelihood function is parametrized by 
\[
\theta \equiv ((\mu^0_{0k}, \sigma^0_{0k}, \mu^0_{1k}, \sigma^0_{1k}), k=1, \ldots, K, \sigma^0_{0}, \sigma^0_{1}, \mu^0_{y}, \mu^0_{w}, \sigma^0_{y}, \sigma^0_{w})_{g \in G}
\]
where \( G \) is the set of the four demographic groups defined above.

I estimate parameter \( \theta \) using maximum simulated likelihood with 400 simulations using scrambled randomized Halton draws (Train 2009 chapter 9). The number of simulations appears to be enough for convergence because the estimates change little from 200 simulations to 400 simulations. I compute standard errors using the information identity with the Hessian being estimated by the outer product of the gradient of the simulated likelihood at the estimated parameter \( \hat{\theta} \) (Train 2009 chapter 9). I assume the utility function \( U^1_{as} \) to be quasi-linear in distance and will often measure results by distance-equivalent utilities in \( t = 1 \).

This estimation procedure uses admissions lotteries as follows. The likelihood involves initial assignments \( s^0_a \)'s as arguments, and I substitute realized initial assignments \( s^0_a \)'s in the data into the likelihood. As explained in the last identification section, these initial assignments \( s^0_a \)'s are conditionally randomly assigned by the first-round deferred acceptance algorithm with admissions lotteries; I provide empirical support for conditionally random assignment in Appendix 1.6.3. As a result, consistent with the identification argument using randomness in initial assignments, the likelihood-based estimation procedure uses the fact that initial assignments are conditionally random. Additional details of the estimation procedure and the construction of the estimation sample are in Appendix 1.6.3.

### 1.3.4 Estimates and Fit

The parameter vector \( \theta \) are high dimensional and not easy to directly interpret. I summarize key features of the estimated \( \hat{\theta} \) here. (Appendix Tables 1.6.15-1.6.18 show that many dimensions of \( \hat{\theta} \) are significant and exhibit sizable heterogeneity.) First, demand changes by learning are an important feature of my empirical model. Figure 1-5a plots the distribution of estimated demand changes due to learning about observable school characteristics, i.e., \( \sum^K_{k=1} \beta_{ak} f_{ak} X_{ask} \), against estimated overall new utilities (\( U^1_{as} \); the black distribution). This figure is based on the estimated rational expectation model. As can be expected from the largeness of NYC, overall utilities from schools are highly dispersed, and most and least preferred schools are different by more than 50 mile-equivalent units.

If there were no frictions and \( f_a = 0 \), the distribution of demand changes would be degenerate at the origin. This is far from the case, and there are large demand changes, as

---

27I do not simulate initial assignments since applicants make reapplication decisions and preferences conditional on realized initial assignments \( s^0_a \)'s in the data.

28Recall that \( \text{Var}(e^1_{as}) = \text{Var}(e^0_{as} + e_{as}) = 2 \text{Var}(e^0_{as}) \). When comparing period 0 and 1 utilities, I divide period 1 utilities by 2 so that the variance of the unobserved utility shock stays the same between \( t = 0 \) and \( t = 1 \).
shown in Figure 1-5a. Their magnitude is often equivalent to multiple miles, though as expected it is smaller compared with the dispersion of overall utilities.

Despite these demand changes, and despite the fact that a majority of applicants are initially assigned to a non-first-choice school, only a small fraction of applicants reapply. To explain this fact, Figure 1-5b plots the distribution of estimated initial assignment effects ($\hat{\gamma}_a$; the left blue distribution) and reapplication costs ($\hat{c}_a$; the left red distribution) by simulating the estimated rational expectation model. The figure compares them against estimated overall new utilities ($\hat{U}_{1a}$; the right black distribution). Estimated reapplication costs $\hat{c}_a$ and initial assignment effects $\hat{\gamma}_a$ are often significant and positive: Applicants not only face reapplication costs, but they also come to prefer their initially assigned schools more, compared with other schools. Consequently, unless the initially assigned school turns out to be much worse than a more preferred school, applicants do not find reapplying worthwhile, which explains the low observed reapplication rate. Similar patterns for the alternative naive free expectation model are reported in Appendix Figures 1.6.14a and 1.6.14b.

These estimation results confirm the suggestive descriptive evidence (in Sections 1.2 and 1.3.2) that there are demand changes by learning, but switching costs (combinations of reapplication costs and initial assignment effects) prevent many demand changes from translating into observed choice changes. In addition, since the structural estimates show that both reapplication costs and initial assignment effects lower the reapplication rate, I can quantify the relative contributions of learning, reapplication costs, and initial assignment effects to reapplication behavior.

As an evaluation of how well the estimated model fits the data, I simulate the estimated model and key moments 50 times, and compare the average simulated moments against the real ones. Figure 1-6 plots the real first choice market shares of schools in the initial application process against the simulated ones, where the first choice market share of a school is defined as the fraction of applicants who rank it first among all applicants who make a first choice. The real and simulated shares are highly correlated. On the reapplication behavior, Table 1.5 compares the most important moments — the fractions of reapplicants and those who exhibit choice reversals — between the data and the estimated model. Both models mimic the behavior of the data well in terms of these moments.

Finally, Table 1.6 compares the moment in Table 1.4 — changes in $R^2$'s from school-level regressions of schools' first choice market shares on observable school characteristics — between the data and the model. The model resembles the data in that reapplicants' choices become more correlated with and responsive to observable school characteristics from the first round to the reapplication process. These results suggest that the estimated models do decent jobs at matching key moments in the data.
1.4 Welfare Consequences of Evolving Demand

1.4.1 Costs of Ignoring Demand Changes

Estimates of the structural model reveal many more demand changes than are suggested by the low observed reapplication rates I showed in Section 1.2. To measure the amount of hidden demand changes, Table 1.5 shows the fractions of reapplicants as well as those with choice reversals in the counterfactual simulations of the estimated models without reapplication costs $c_a$ (while keeping the estimated initial assignment effects $\hat{\gamma}_a$ as they are). With no reapplication costs, the reapplication rate increases from 7% in the data to 30-40%. More importantly, a majority of the counterfactual reapplicants exhibit choice reversals, implying that the fraction of applicants with underlying demand changes (more than 20%) is several times larger than the fraction of those with choice reversals in the data (5%). These demand changes are masked by reapplication costs. This finding is consistent with the suggestive descriptive evidence of hidden demand changes in Figure 1-4.

As a result of these significant demand changes, the welfare cost of ignoring demand changes is large. To measure the welfare cost, I compare the real first-round assignment based on school choices induced by old demand $U_{as}^0$ with the counterfactual "frictionless benchmark." The frictionless benchmark is defined as what would have been produced by the same first-round deferred acceptance algorithm, had families made choices based on their new demand $U_{as}^1$. Since the two differ only in whether families make choices based on old or new demand, the difference between the two captures the welfare costs of ignoring demand changes by learning.

Table 1.7 summarizes welfare changes from the real first-round assignment to the frictionless benchmark. The real first-round assignment and the frictionless benchmark are

---

29 Initial assignment effects also lower the reapplication rate. As shown in Figure 1-5b, the estimates show that initial assignment effects are often positive and large, meaning that families tend to get to prefer initially assigned schools more, compared with other schools. This satisfaction with initial assignments lowers the reapplication rate by more than 20%, which is computed as the difference between the real reapplication rate and the counterfactual rate under no initial assignment effects $\gamma_a = 0$ (but with estimated reapplication costs $c_a$).

30 Except applicants’ choices, every other input is unchanged between the real first-round assignment and the frictionless benchmark. For example, school capacities and their preferences or priorities over applicants are fixed.

31 When simulating an assignment mechanism, I also simulate lottery numbers used by the mechanism to break ties in priorities. Recall the explanation in Section 1.3.2 about the use of lottery numbers in the NYC system, and see Appendix 1.6.3 for further details. Furthermore, it is sometimes the case that an applicant’s simulated preferences contain schools that the applicant does not rank in her real preference observed in the data. For such pairs of applicants and schools, the data does not provide priority information. In such cases, I simulate their priorities from the empirical distribution of priorities in the data. I do the same lottery number and priority simulation for the other counterfactual exercises. Finally, I assume that each assigned applicant experiences estimated initial assignment effect $\hat{\gamma}_a$ from her assignment. The results change little even if I exclude initial assignment effects. The same treatment of
significantly different. Specifically, the two assignments give different allocations (schools) to a majority of families. Also, the average welfare loss under the real first-round assignment compared with the frictionless benchmark is more than 1-mile-equivalent when I measure it by new demand $U_{as}^1$ after learning, which is assumed to be quasi-linear in the distance between the family and the school locations. The 1-mile-equivalent utility unit can be interpreted as corresponding to traveling 1 mile every school day during the high school years. This magnitude corresponds to more than .15 standard deviation in the distribution of utilities from all schools for each applicant. Significant learning and demand changes thus undermine the welfare performance of the initial match that ignores demand changes. This illustrates that demand-side choice frictions and learning significantly affect the welfare gains from a centralized market. This motivates me to investigate ways to accommodate demand changes, to which I turn next.

1.4.2 Evaluating Reapplication Processes

The large welfare costs of ignoring demand changes motivate me to investigate ways to alleviate the costs by accommodating demand changes. As already explained, NYC runs a discretionary, human-driven reapplication process, presumably to improve families' welfare by flexibly accommodating changing needs. It is also possible to run a centralized algorithm not only for the initial market but also for the reapplication process. This section evaluates how well the discretionary and counterfactual centralized reapplication processes alleviate the welfare cost of ignoring demand changes.

Theory

I start by introducing counterfactual centralized reapplication mechanisms. Appendix 1.6.4 builds a dynamic version of the school-student assignment model, analyzes two centralized designs of the reapplication process, and shows that they are the "best possible" mechanisms to accommodate choice changes. The counterfactual mechanisms are what I call the dynamic deferred acceptance mechanism and the deferred deferred acceptance mechanism. They are easily implemented by applying the deferred acceptance algorithm to observed choice data.

To define the mechanisms, take as given applicants' old preferences $\succ^0_A (\succ^0_a)$ in initial application, their new preferences $\succ^1_A (\succ^1_a)$ (which I construct below), and school preferences/priorities $\succ_S$ over applicants observed in the data. School preferences/priorities $\succ_S$ are strict preferences/priorities after tie-breaking by lottery numbers. The initial assignment effects applies to the other counterfactual exercises.

32I use new demand $U_{as}^1$ as my welfare measure since it is demand after learning at a point in time closer to enrollment periods; new demand is thus expected to be a better welfare measure than old demand $U_{as}^0$. In other words, in this and other counterfactual welfare analyses, I assume that frictions $f_a$ are welfare-irrelevant.
dynamic deferred acceptance mechanism determines an assignment as follows.\textsuperscript{33}

(1) Compute the initial match \( DA(\succ^0_A, \succ_s) \) in the initial application process, where function \( DA(\cdot, \cdot) \) maps each possible profile of applicant and school preferences into the assignment produced by the deferred acceptance algorithm under these input preferences.

(2) Give an initial match guarantee to each applicant by modifying each school \( s \)'s preference so that \( s \) most prefers applicants matched with \( s \) in step 1.

(3) Use modified school preferences/priorities \( \succ'_A \) and applicants' reapplication preferences \( \succ^1_A \) to get \( DA(\succ^{-1}_A, \succ'_s) \equiv \varphi^DA_{\text{dynamic}}(\succ^0_A, \succ^1_A, \succ_s) \) where \( \varphi^DA_{\text{dynamic}}(\succ^0_A, \succ^1_A, \succ_s) \) denotes the assignment under the dynamic deferred acceptance mechanism under input preferences \( (\succ^0_A, \succ^1_A, \succ_s) \).

On the other hand, the deferred deferred acceptance mechanism is the same as the dynamic deferred acceptance mechanism except that I make no modification to school preferences/priorities. That is, \( \varphi^DA_{\text{deferred}}(\succ^0_A, \succ^1_A, \succ_s) \equiv DA(\succ^1_A, \succ_s) \) where \( \varphi^DA_{\text{deferred}}(\succ^0_A, \succ^1_A, \succ_s) \) denotes the assignment under the deferred deferred acceptance mechanism under input preferences \( (\succ^0_A, \succ^1_A, \succ_s) \).

Appendix 1.6.4 provides a formal analysis of \( \varphi^DA_{\text{dynamic}} \) and \( \varphi^DA_{\text{deferred}} \), and shows that they are the "best possible" mechanisms to accommodate choice changes. Specifically, consider the following criteria of how well a mechanism accommodates choices changes and caters to new preferences (Appendix 1.6.4 formally defines these properties):

(I) "Fairness (stability)" with respect to \( (\succ^{-1}_A, \succ_s) \), i.e., no applicant-school pair "blocks" the outcome under that mechanism and has an incentive to jointly deviate from it to be matched with each other outside the mechanism.

(II) "Being less unfair (unstable)" than the initial match with respect to \( (\succ^{-1}_A, \succ_s) \), i.e., it is always the case that any applicant-school pair blocking the outcome under that mechanism also blocks the initial match.

(III) "Weak Pareto efficiency" with respect to \( \succ^1_A \), i.e., there is no other assignment that every applicant strictly prefers over the outcome under that mechanism.

(IV) "Always Pareto dominating the initial match" with respect to \( \succ^{-1}_A \).

(V) "Dynamic strategy-proofness," i.e., any preference manipulation by any applicant in any period is never strictly profitable with respect to that applicant's preference in that period.

\textsuperscript{33}Similar dynamic mechanisms have been discussed in existing studies (Pereyra, 2013; Coles et al., 2014; Kadam and Kotowski, 2015). The empirical and theoretical analysis in this paper does not appear to have an analog in their analyses.
In terms of these properties, \( \varphi_{\text{dynamic}}^{DA} \) and \( \varphi_{\text{deferred}}^{DA} \) are the best possible mechanisms, as the following result (shown in Appendix 1.6.4) implies. This result motivates using \( \varphi_{\text{dynamic}}^{DA} \) and \( \varphi_{\text{deferred}}^{DA} \) as the canonical centralized reapplication mechanisms.

**Proposition 1**

1.A) \( \varphi_{\text{dynamic}}^{DA} \) satisfies (II) being less unfair than the initial match, (III) weak Pareto efficiency, and (IV) always Pareto dominating the initial match (call this set of desiderata A), but not others.

1.B) \( \varphi_{\text{deferred}}^{DA} \) satisfies (I) fairness, (II) being less unfair than the initial match, (III) weak Pareto efficiency, and (V) dynamic strategy-proofness (call this set of desiderata B), but not others.

2) Consider any possible dynamic mechanism \( \varphi \). \( \varphi \) can satisfy only a subset of set A or B.

**Discretionary vs Centralized Reapplication Processes**

My counterfactual analysis studies how well the counterfactual centralized reapplication processes improve on the initial match and alleviate the welfare costs of ignoring demand changes. I also compare the centralized processes with the existing discretionary process. There are two ways to implement this particular evaluation. First, I can evaluate the reapplication processes in a descriptive way, without resorting to the model. An alternative, usual approach is to evaluate each reapplication process by simulating the estimated model. I start with the descriptive evaluation and then compare it with the model-based evaluation.

For empirically implementing and evaluating counterfactual centralized mechanisms \( \varphi_{\text{deferred}}^{DA} \) and \( \varphi_{\text{dynamic}}^{DA} \), I need to define \( \succ^1_a \), applicant preferences in the reapplication stage. In the descriptive counterfactual evaluation, I construct them as follows. If applicant \( a \) does not reapply in the data, then assume that \( \succ^1_a \) stays the same at \( \succ^0_a \) (the preference that applicant \( a \) submits in the initial application period). If \( a \) reapplies and reports a new reapplication preference, then define \( \succ^1_a \) as \( a \)'s reapplication preference followed by \( \succ^0_a \). For example, if \( \succ^0_a \) is \( (s_1, s_2) \) and \( a \) reapplies and ranks only \( s_3 \) in her reapplication preference, then \( \succ^1_a \) is \( (s_3, s_1, s_2) \). This construction of \( \succ^1_a \) ignores all unobserved demand changes and uses only observed reaplications and choice changes. Hidden demand changes (in Table 1.5) suggest that \( \succ^1_a \) may not be an appropriate welfare measure for applicants who do not reapply. However, this is probably not a big problem for this particular evaluation of reapplication processes, because reapplication processes mainly affect reapplicants, for whom \( \succ^1_a \) reflects demand changes and is expected to be a reasonable welfare measure.\(^{34}\)

I use \( \succ^0_A \) and \( \succ^1_A \) created above (as well as school capacities and priorities in the data)

\(^{34}\)The discretionary reapplication process affects only reapplicants. Centralized reapplication processes influence not only reapplicants but also others, but Appendix Figure 1.6.15 shows that most of applicants affected by centralized reapplication processes are reapplicants.
to simulate $\varphi^{DA}_{\text{dynamic}}, \varphi^{DA}_{\text{deferred}},$ and the discretionary reapplication process, and I compare them with the initial match with respect to $\succ^1_A$. Note that this procedure makes no use of the empirical model and is based only on objects directly observed in the data.

The comparison between the discretionary reapplication process and the deferred and dynamic deferred acceptance mechanisms is in Figure 1-7. Starting from the initial match as the common status quo, each line plots the distribution of preference rank improvements (with respect to $\succ^1_a$) of the finally assigned school under each of the three reapplication processes. Figure 1-7 shows that all reapplication processes produce welfare gains. More importantly, the centralized reapplication processes are more effective and produce gains more than twice as large as those from the discretionary reapplication process.

Table 1.8 summarizes this result and compares it with the results from an alternative evaluation based on the estimated structural models. In the structural counterfactual evaluation, I use the estimated models to simulate old and new utilities $U^0_{as}, U^1_{as}$, reapplication costs $c_a$, initial assignment effects $\gamma_a$, and associated initial application preferences $\succ^0_A$ (over up to 12 schools) induced by $U^0_{as}$. I then substitute the simulated initial application preferences (as well as school capacities and priorities in the data) into the deferred acceptance algorithm to obtain the initial match. I use this initial match and simulated $U^1_{as}, c_a, \gamma_a$ to simulate reapplication decisions. If applicant $a$ reapplies, let $\succ^1_a$ be the reapplication preference (over up to 3 schools) induced by $U^1_{as}$, followed by $\succ^0_a$ (as in the descriptive evaluation). Otherwise, define $\succ^1_a = \succ^0_a$. I then use $\succ^1_a$ and $\succ^0_a$ to compute the allocations under $\varphi^{DA}_{\text{deferred}}$ and $\varphi^{DA}_{\text{dynamic}}$. I compare the allocations produced by $\varphi^{DA}_{\text{deferred}}$ and $\varphi^{DA}_{\text{dynamic}}$ with the initial match with respect to simulated $U^1_{as}$ under each simulation, and take the average over 50 simulations.

The descriptive and structural evaluations show similar gains from the centralized reapplication processes, as shown in Table 1.8. This provides additional support for the finding that the centralized reapplication processes produce larger gains than those from the discretionary reapplication process. This also provides some confidence to the use of the estimated structural model for evaluating other counterfactual policies, for which no model-free or descriptive evaluation is possible. An example of a counterfactual policy that innately needs the estimated model is the frictionless benchmark (in Table 1.7) that accommodates all demand changes, many of which are unobserved (as shown in Table 1.5).

Reapplication Processes and Demand-side Inertia

The centralized reapplication processes are shown to accommodate observed reapplications and choice changes better than the discretionary reapplication process does. This evaluation of the centralized reapplication processes takes estimated reapplication costs as given. On the other hand, there are technological changes and school districts’ and social entrepreneurs’ initiatives that may ease reapplication costs (e.g., online systems for
more easily making and updating school choices).

To measure the potential effects of such demand-side interventions, I finally investigate the performance of the centralized reapplication processes relative to their "frictionless implementation." The frictionless implementation of the mechanisms is the hypothetical, possibly infeasible implementation that turns off reapplication costs $c_a$ so that applicants express demand changes with no barrier. The frictionless implementation is similar to the frictionless benchmark in Table 1.7. In particular, the frictionless implementation of $\varphi_{deferred}^DA$ is the same as the frictionless benchmark. The frictionless implementation is not subject to demand-side inertia due to reapplication costs, while the feasible centralized reapplication processes are; the difference between the two thus captures the welfare effect of demand-side inertia (given the market design fixed). For simulating and evaluating the frictionless implementation, I need to use the estimated empirical model. The use of the estimated model for evaluating the frictionless implementation is the same as that for the feasible implementation in Table 1.8, except that I set reapplication costs to $c_a = 0$ when simulating the model to evaluate the frictionless implementation.

The welfare effect of demand-side inertia is large, and the gains from the mechanisms change by several times depending on the extent of demand-side inertia caused by reapplication costs: Table 1.9 shows that starting from the counterfactual scenario with no switching costs, estimated switching costs dilute the gains from the mechanisms by more than 50%.\textsuperscript{35} Because of inertia and the resulting low participation, the reapplication processes reveal only a small fraction of the demand changes families experience as they learn. As a result, the centralized reapplication processes achieve no more than 30% of the welfare gains from their frictionless implementation, which responds to all the demand changes families would express if there were no demand-side inertia. This suggests that the gains from a dynamic centralized market depend to a large extent on demand-side inertia that prevents families from reapplying and expressing demand changes.

1.5 Conclusion and Future Directions

Centralized market-like institutions have become a widespread form of public policy. The success of these markets hinges on the assumption that participants make well-informed choices upfront. In this paper, I use data from NYC's school choice system to evaluate this assumption. I show that, contrary to the premise of well-informed upfront choices, families' choices change after the initial match as they learn about schools. To recover

\textsuperscript{35}In Tables 1.8 and 1.9, welfare calculations do not include reapplication costs $c_a$ (but include initial assignment effects $\gamma_a$). This choice is in order to make the comparison in Table 1.9 conservative and nontrivial: if welfare calculations include $c_a$, then turning off $c_a$ would trivially result in welfare improvements. Incorporating $c_a$ does not change the qualitative comparison among reapplication systems in Tables 1.8 and 1.9 since reapplicants experience reapplication costs under any reapplication system, and the welfare loss from reapplication costs is cancelled out when I compare different reapplication systems.
underlying evolving demand, I develop an empirical model of evolving demand for schools under learning, reapplication costs, and initial assignment effects. I exploit institutional features of centralized school choice systems, especially admissions lotteries, to separately identify these model components.

The estimates suggest that there are substantially more changes in underlying demand than in observed choices. These significant demand changes undermine the welfare performance of the initial match, and result in large welfare costs of ignoring demand changes. The large welfare costs of ignoring demand changes motivate me to investigate dynamic mechanisms, which I show can best accommodate choice changes in theory. I empirically find that these mechanisms significantly improve on the existing discretionary reapplication process and initial match. Also, the gains from the mechanisms change greatly depending on the extent of demand-side inertia caused by reapplication costs. Thus, demand-side frictions (such as learning, demand changes, and inertia) affect the gains from a centralized market as much as its design.

The suggested importance of frictions in participants' choices opens the door to many empirical and methodological questions. An implication of my results is the potential importance of demand-side technological changes or policy interventions that may alleviate demand-side frictions (e.g., online systems for more easily making and updating school choices, applications for more easily searching school characteristics). Little is known about the effects of such demand-side interventions, aside from what Hastings and Weinstein (2008) and Andrabi et al. (2015) report. Another related question is about the relationship between dynamic choice behavior and subsequent outcomes. For example, few papers study whether later, presumably better-informed school choices result in changes in academic achievement and other behavioral outcomes.

Methodologically, this paper contributes to understanding families' school choice dynamics. A variety of extensions are both possible and desirable. For example, extending the data and model to more than two periods would make it possible to study the speed of learning and the dynamics of inertia. Another potentially fruitful direction is to incorporate richer aspects of learning, e.g., latent consideration sets families use when making choices, families' anticipation and sophistication about future learning. Finally, while this paper focuses on welfare analysis based on assignments or offers, it is probably more desirable to study welfare from enrollment (rather than assignments as intermediate steps toward enrollment and educational experience). I leave these challenging directions for future research.
Figure 1-1: Timeline of the First-round and Reapplication Process

Notes: In this figure, the left black histogram plots the distribution of dates at which applicants file initial applications. The right red histogram does the same for reapplication dates conditional on those who reapply. See Section 1.2.1 for discussions about this figure.

Table 1.1: Evolving School Choices

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>All applicants in the first round</td>
<td>91,289</td>
<td>100%</td>
</tr>
<tr>
<td>Reapplicants in the aftermarket</td>
<td>6,430</td>
<td>7%</td>
</tr>
<tr>
<td>Reapplicants who exhibit preference reversals</td>
<td>4,564</td>
<td>5%</td>
</tr>
<tr>
<td>Reapplicants who exhibit surely nonstrategic preference reversals</td>
<td>3,464</td>
<td>4%</td>
</tr>
<tr>
<td>Lower bound on the fraction of</td>
<td></td>
<td>76%</td>
</tr>
<tr>
<td>nonstrategic preference reversals (4th row/3rd row)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table shows how many applicants reapply, exhibit any choice reversals, and exhibit surely nonstrategic choice reversals between the first-round and the reapplication process. I say an applicant exhibits choice reversals if she reapply against her initially assigned school by ranking another school that is unranked or ranked below s in her initial application. An applicant exhibits surely non-strategic choice reversals if she exhibits any choice reversals and does not exhaust her preference list (rank 11 or fewer schools). As explained in the main text, surely non-strategic choice reversals can be rationalized only by real demand changes. See Section 1.2.1 for discussions about this figure.
Table 1.2: Characteristics of Applicants and Reapplicants

<table>
<thead>
<tr>
<th></th>
<th>All applicants</th>
<th>All reapplicants</th>
<th>Reapplicants with preference reversals</th>
<th>Reapplicants with accepted reapplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average preference rank</td>
<td>4.0</td>
<td>5.7</td>
<td>4.3</td>
<td>5.5</td>
</tr>
<tr>
<td>of initial assignment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>50%</td>
<td>55%</td>
<td>56%</td>
<td>53%</td>
</tr>
<tr>
<td>8th graders</td>
<td>95%</td>
<td>98%</td>
<td>98%</td>
<td>97%</td>
</tr>
<tr>
<td>Top 2% of grade 7</td>
<td>2.1%</td>
<td>0.6%</td>
<td>0.3%</td>
<td>1.0%</td>
</tr>
<tr>
<td>English, Language, Arts test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Grade 7 reading grade category*

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Middle</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15%</td>
<td>72%</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>12%</td>
<td>77%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>78%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>19%</td>
<td>70%</td>
<td>11%</td>
</tr>
</tbody>
</table>

*Living area*

<table>
<thead>
<tr>
<th></th>
<th>Manhattan</th>
<th>Brooklyn</th>
<th>Queens</th>
<th>Bronx</th>
<th>Staten Island</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12%</td>
<td>31%</td>
<td>28%</td>
<td>22%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>11%</td>
<td>34%</td>
<td>31%</td>
<td>19%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>11%</td>
<td>32%</td>
<td>33%</td>
<td>20%</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>9%</td>
<td>32%</td>
<td>40%</td>
<td>12%</td>
<td>6%</td>
</tr>
</tbody>
</table>

*Home language*

<table>
<thead>
<tr>
<th></th>
<th>Chinese</th>
<th>Spanish</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>64%</td>
<td>91,289</td>
</tr>
<tr>
<td></td>
<td>8%</td>
<td>62%</td>
<td>6,430</td>
</tr>
<tr>
<td></td>
<td>8%</td>
<td>66%</td>
<td>4,564</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>56%</td>
<td>1,375</td>
</tr>
</tbody>
</table>

*Ethnicity*

<table>
<thead>
<tr>
<th></th>
<th>Asian</th>
<th>Hispanic</th>
<th>Black</th>
<th>White</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13%</td>
<td>39%</td>
<td>34%</td>
<td>14%</td>
<td>83,047</td>
</tr>
<tr>
<td></td>
<td>14%</td>
<td>41%</td>
<td>32%</td>
<td>13%</td>
<td>5,914</td>
</tr>
<tr>
<td></td>
<td>13%</td>
<td>43%</td>
<td>33%</td>
<td>10%</td>
<td>4,269</td>
</tr>
<tr>
<td></td>
<td>17%</td>
<td>39%</td>
<td>26%</td>
<td>18%</td>
<td>1,283</td>
</tr>
</tbody>
</table>

*Notes:* This table shows baseline characteristics of all applicants, applicants who reapply, applicants who reapply and exhibit choice reversals between their initial applications and reapplications, and applicants who reapply and are accepted. On “average preference rank of initial assignment,” I assume the preference rank of being unassigned is 13, the worst possible rank plus one. Sample sizes in lower rows are smaller than that in upper rows because some characteristics are missing for some applicants. The definitions of test-score-related variables are in Appendix 1.6.1. See Section 1.2.2 for discussions about this table.
Figure 1-2: Self-reported Reasons for Reapplication

(a) Self-reported Reasons for Reapplication

(b) Breakdown of “New Information”

Notes: Panel 1-2a classifies self-reported reasons for reapplying into main categories. Panel 1-2b focuses on the “New Information about Schools” category in Panel 1-2a and breaks it down into several subcategories. “Distance” is different from “Moving” in that the former does not refer to any address change. The construction of the categories are explained in Appendix 1.6.2. See Section 1.2.2 for discussions about this figure.

Table 1.3: Growing Response of Choices to Distance and Academic Achievement

<table>
<thead>
<tr>
<th>Distance (in miles)</th>
<th>Initial applications</th>
<th>Reapplications</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All applicants</td>
<td>Reapplicants</td>
<td>Reapplicants</td>
</tr>
<tr>
<td>1st choice school</td>
<td>4.5</td>
<td>4.9</td>
<td>3.9</td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>5.1</td>
<td>5.1</td>
<td>4.0</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>4.8</td>
<td>5.3</td>
<td>4.0</td>
</tr>
<tr>
<td>1st choice school</td>
<td>12.8%</td>
<td>11.3%</td>
<td>10.5%</td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>11.6%</td>
<td>10.1%</td>
<td>10.7%</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>15.1%</td>
<td>14.1%</td>
<td>10.7%</td>
</tr>
</tbody>
</table>

Notes: This table shows the average distance to and academic achievement level of ranked schools in initial applications and reapplications. Schools with low academic performance are “schools in need of improvement” in the official school brochure issued by NYC. Under the No Child Left Behind Act, New York State establishes annual performance goals in Mathematics and English Language Arts for all NYC public schools. Schools that do not meet these goals for two consecutive years are identified as schools in need of improvement. See Section 1.2.2 for discussions about this figure.
Table 1.4: Growing Response of Choices to School Characteristics

<table>
<thead>
<tr>
<th>Specification</th>
<th>Initial applications R²</th>
<th>Reapplications R²</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All applicants Reapplicants</td>
<td>Reapplicants Reapplicants</td>
<td></td>
</tr>
<tr>
<td>Specification 1</td>
<td>0.015 0.016</td>
<td>0.021 +31%</td>
<td></td>
</tr>
<tr>
<td>Location/borough dummies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification 2</td>
<td>0.027 0.038</td>
<td>0.054 +42%</td>
<td></td>
</tr>
<tr>
<td>Academic performance dummy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification 3</td>
<td>0.075 0.082</td>
<td>0.181 +121%</td>
<td></td>
</tr>
<tr>
<td>Program type dummies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification 4</td>
<td>0.268 0.229</td>
<td>0.315 +38%</td>
<td></td>
</tr>
<tr>
<td>Capacity dummies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (Schools)</td>
<td>750 750 750</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table shows R²'s from school-level regressions of schools' first choice market shares on various sets of observable school characteristics. The first choice market share of a school is defined as the fraction of applicants who rank it first among all applicants who make a first choice. Rows correspond to different sets of school characteristics included in regressions. Columns correspond to different samples in different periods used to compute market shares. The details of included characteristics are explained in Appendix 1.6.2. See Section 1.2.2 for discussions about this figure.

Figure 1-3: Falsification Check of the Empirical Model

(a) 
(No Pref Reversals) / (No Schools Ranked in New Pref) Conditional on Reapplying

(b) 
Number of Preference Reversals Conditional on Reapplying

Notes: This figure correlates two measures of the amount of choice reversals with the preference rank of the initially assigned school with respect to the initial preference (both are conditional on reapplicants). For each applicant who reapplies after being initially assigned to s, the number of choice reversals is defined as the number of schools t such that t is unranked or ranked below s in her initial application but ranked in her reapplication. See Section 1.3.1 for discussions about this figure.
Figure 1-4: Identification and Evidence of Demand Changes and Switching Costs

(a) Step 1
Hypothetical pattern under NO demand change & NO switching cost
Difference is due to demand change OR switching cost
Real pattern in data

(b) Step 2
Hypothetical pattern under NO demand change & NO switching cost
Real pattern in data

(c) Step 3
Hypothetical pattern under NO demand change & NO switching cost
Hypothetical pattern under NO switching cost
Real pattern in data

(d) Step 4
Hypothetical pattern under NO demand change & NO switching cost
Demand Change
Hypothetical pattern under NO switching cost
Switching Cost
Real pattern in data

Notes: These figures explain how I separately identify demand changes and switching costs from the solid black line observed in the data. The solid black line correlates the conditional probability of reapplying to the preference rank of the initially assigned school with respect to the initial preference. See Section 1.3.2 for the identification argument using these figures.
Figure 1-5: Summary of Estimates

(a) Demand Changes due to Learning

![Graph showing demand changes due to learning](image)

(b) Reapplication Costs & Initial Match Effects

![Graph showing reapplication costs and initial match effects](image)

Notes: Based on the rational expectation model in Section 1.3.1, Panel 1-5a plots the distributions of estimated overall new utilities ($\hat{U}_{as}^1$) and latent demand changes caused by frictions about observable school characteristics ($\sum_{k=1}^K \beta_{sk} f_{sk} X_{ask}$) for all (applicant $a$, school $s$) pairs. Panel 1-5b plots the distributions of estimated overall new utilities ($\hat{U}_{as}^1$), estimated reapplication costs ($\hat{c}_a$), and estimated initial assignment effects ($\hat{\gamma}_a$). Both panels are based on 50 simulations of the estimated model for each (applicant $a$, school $s$) pair. See Sections 1.3.1 and 1.3.3 for the details of the model and the estimation method, respectively. See Section 1.3.4 for discussions about this figure.
Figure 1-6: Fit (I) Initial Choices

Notes: This figure correlates the real first choice market share of each school in initial applications with the same share predicted by simulating the estimated model 50 times and averaging over them. The first choice market share of a school is defined as the fraction of applicants who rank it first among all applicants. See Section 1.3.4 for discussions about this figure.

Table 1.5: Fit (II) Reapplications, Choice Reversals, and Hidden Demand Changes

<table>
<thead>
<tr>
<th>Simulations of the estimated model</th>
<th>Model 1: Rational expectation</th>
<th>Model 2: Naive free expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>$c_a = 0$ &amp; estimated $Y_a$</td>
</tr>
<tr>
<td>Reapplicants</td>
<td>7.0%</td>
<td>30.8%</td>
</tr>
<tr>
<td></td>
<td>(0.15%)</td>
<td>(0.78%)</td>
</tr>
<tr>
<td>Reapplicants who exhibit choice reversals</td>
<td>5.0%</td>
<td>23.4%</td>
</tr>
<tr>
<td></td>
<td>(0.18%)</td>
<td>(1.15%)</td>
</tr>
<tr>
<td>2nd row/1st row</td>
<td>71.4%</td>
<td>76.1%</td>
</tr>
</tbody>
</table>

Notes: This table shows the fraction of applicants who reapply and exhibit choice reversals in the data and the estimated models with and without reapplication costs. I say an applicant exhibits choice reversals if she reappplies against her initially assigned school $s$ by ranking another school that is unranked or ranked below $s$ in her initial application. Each column for a model is based on simulating the corresponding estimated model 50 times and averaging over them. Simulation standard errors are in parentheses. See Sections 1.3.4 and 1.4.1 for discussions about this figure.
Table 1.6: Fit (III) Growing Correlation between Choices and School Characteristics

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Initial app. R²</th>
<th>Reapp. R²</th>
<th>Change</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reapplicants</td>
<td>Reapplicants</td>
<td>Reapplicants</td>
<td>Reapplicants</td>
<td>Reapplicants</td>
</tr>
<tr>
<td>Specification 1</td>
<td>0.018</td>
<td>0.020</td>
<td>+11.1%</td>
<td>+31.3%</td>
</tr>
<tr>
<td>(Location/borough dummies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification 2</td>
<td>0.039</td>
<td>0.041</td>
<td>+5.1%</td>
<td>+42.1%</td>
</tr>
<tr>
<td>(1+Academic performance dummy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification 3</td>
<td>0.096</td>
<td>0.250</td>
<td>+160.4%</td>
<td>+120.7%</td>
</tr>
<tr>
<td>(2+Program type dummies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification 4</td>
<td>0.587</td>
<td>0.520</td>
<td>-11.4%</td>
<td>+37.6%</td>
</tr>
<tr>
<td>(3+Capacity dummies)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (schools)</td>
<td>750</td>
<td>750</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Based on the rational expectation model, this table shows changes in $R^2$'s from school-level regressions of schools' first choice market shares on observable school characteristics. The first choice market share of a school is defined as the fraction of applicants who rank it first among all applicants who make a first choice. Rows correspond to different sets of school characteristics included in regressions. Columns correspond to different periods used to compute market shares. For the model, I compute predicted shares by simulating the estimated model 50 times and averaging over them. See Section 1.3.4 for discussions about this figure.

Table 1.7: Welfare Costs of Ignoring Demand Changes

<table>
<thead>
<tr>
<th>Real 1st-round assignment vs frictionless 1st-round assignment under new demand</th>
<th>Model 1: Rational expectation</th>
<th>Model 2: Naive free expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of winners</td>
<td>% of losers</td>
<td>Avg util change</td>
</tr>
<tr>
<td>+1.60 miles</td>
<td>47.5%</td>
<td>(0.01 miles)</td>
</tr>
<tr>
<td>(0.14%)</td>
<td>11.1%</td>
<td>+0.29 utility SD</td>
</tr>
<tr>
<td>(0.08 utility SD)</td>
<td>(0.11%)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table shows welfare changes from the real first-round assignment based on old demand to the counterfactual "frictionless benchmark" that is defined as what would have been produced by the same first-round assignment mechanism had families made choices based on their new demand after learning. "Winners" ("losers") are defined as applicants who be better (worse, respectively) off under the frictionless benchmark, compared with the real first-round assignment with respect to new demand $U_{a,n}^{1}$. Average utility changes are also measured by new demand $U_{a,n}^{1}$, which is assumed to be quasi-linear in the distance between the family and the school locations. "Utility SD" is measured by the standard deviation in the distribution of utilities from all schools for each applicant. Simulation standard errors over 50 simulations are in parentheses. See Section 1.4.1 for discussions about this figure.
Notes: In this figure, each line plots the distribution of improvements of the preference rank of the finally assigned school under each of the three ways to accommodate observed choice changes: the real discretionary reapplication process, the dynamic deferred acceptance mechanism, and the deferred deferred acceptance mechanism. Section 1.4.2 defines the latter two mechanisms. The common status quo for these comparisons is the initial assignment in the initial application process via the static deferred acceptance algorithm. The preference rank in April is defined with respect to new preference $>_{\text{new pref}}$ defined in Section 1.4.2. This distribution is conditional on applicants who get different assignments under the two mechanisms. The shaded area around a line indicates the 95% simulation confident interval over 200 simulations of lottery numbers used by the mechanisms to break ties in priorities. See Section 1.3.2 for details of the use of lottery numbers in the NYC system. There is no shaded area around the black line for the real reapplication process since I observed only one realization of the real reapplication process and it is not possible to simulate it. See Section 1.4.2 for discussions about this figure.
### Table 1.8: Centralized vs Discretionary Reapplication Processes (II)

<table>
<thead>
<tr>
<th>Model 1: Rational expectation</th>
<th>Model 2: Naive free expectation</th>
<th>Descriptive counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of winners</td>
<td>% of losers</td>
<td>Avg util change</td>
</tr>
<tr>
<td>Real discretionary aftermarket</td>
<td>N/A</td>
<td>+0.15 miles</td>
</tr>
<tr>
<td>Dynamic DA mech. With estimated reapplication costs</td>
<td>3.0%</td>
<td>0%</td>
</tr>
<tr>
<td>Dynamic DA mech. With estimated mech. C</td>
<td>3.0%</td>
<td>0%</td>
</tr>
<tr>
<td>Deferred DA mech. With estimated reapplication costs</td>
<td>5.4%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Deferred DA mech. With estimated mech. C</td>
<td>5.4%</td>
<td>3.4%</td>
</tr>
</tbody>
</table>

**Notes:** This table summarizes the effects of discretionary and centralized reapplication processes on family welfare in the form of the fractions of winners and losers, who would be better and worse off, respectively, by the introduction of each reapplication process compared with the initial match. The table also reports the average utility change in distance-equivalent utility units. I define winners, losers, and utility changes in terms of new utilities $U_{i}^{1}$ after learning (including initial assignment effects $\gamma_{i}$). “Utility SD” is measured by the standard deviation in the distribution of utilities from all schools for each applicant. Each row for a reapplication process is based on simulating the estimated model 50 times under the reapplication process and averaging over the simulations. Simulation standard errors are in parentheses. Simulation standard errors for the descriptive evaluation are negligible, as suggested by simulation standard errors in Figure 1-7. The last two columns are based on the descriptive counterfactual analysis detailed in Figure 1-7. See Section 1.4.2 for discussions about this table.
Table 1.9: Dynamic Reapplication Processes and Demand-side Inertia

<table>
<thead>
<tr>
<th></th>
<th>Model 1: Rational expectation</th>
<th>Model 2: Naive free expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% of winners</td>
<td>% of losers</td>
</tr>
<tr>
<td>With estimated reapp. costs</td>
<td>3.0%</td>
<td>0%</td>
</tr>
<tr>
<td>Dynamic DA mech.</td>
<td>(0.1%)</td>
<td>(0%)</td>
</tr>
<tr>
<td>(Infeasible)</td>
<td>(0.2%)</td>
<td>(0%)</td>
</tr>
<tr>
<td>Dilution by reapp. costs</td>
<td>-83%</td>
<td>N/A</td>
</tr>
<tr>
<td>With estimated reapp. costs</td>
<td>5.4%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Deferred DA mech.</td>
<td>(0.1%)</td>
<td>(0.04%)</td>
</tr>
<tr>
<td>(Infeasible)</td>
<td>(0.1%)</td>
<td>(0.1%)</td>
</tr>
<tr>
<td>Dilution by reapp. costs</td>
<td>-89%</td>
<td>-59%</td>
</tr>
<tr>
<td>(Infeasible)</td>
<td>(0.1%)</td>
<td>(0.1%)</td>
</tr>
</tbody>
</table>

Notes: This table summarizes the effects of discretionary and centralized reapplication processes on family welfare in the form of the fractions of winners and losers, who would be better and worse off, respectively, by the introduction of each reapplication process compared with the initial match. The table also reports the average utility change in distance-equivalent utility units. "With no reapplication costs" means making reapplication costs $c_a$ zero (while keeping initial assignment effects $y$ at the estimated value). I define winners, losers, and utility changes in terms of new utilities $U^1_{as}$ after learning (including initial assignment effects $y_a$). "Utility SD" is measured by the standard deviation in the distribution of utilities from all schools for each applicant. Each row for a reapplication process is based on simulating the estimated model 50 times under the reapplication process and averaging over the simulations. Simulation standard errors are in parentheses. See Section 1.4.2 for discussions about this table.
1.6 Appendix

1.6.1 Data

Applications, reapplications, and assignments. The data come from the NYC Department of Education (DOE) office. It contains information on students and their characteristics, schools/programs and their characteristics, the initial rank-order preference lists submitted by the students, students' priority statuses at schools, the initial assignment of students to schools, whether each student reapplyes, the new rank-order preference lists submitted by the students who reapply, and the dates of initial applications and reapplications. For applicants who reapply, the data also contain written reapplication reasons as self-reported by reapplicants.

 Applicant characteristics. The records from the DOE contain the street address, previous and current grade, gender, ethnicity, whether the student was enrolled in a public middle school, scores in middle school standardized tests, limited English proficiency status, and special education status. In addition, applicants are categorized into one of three categories based on their score on the seventh grade standardized reading test: top 16 percent (high), middle 68 percent (middle), and bottom 16 percent (low). This categorization is made mainly for admissions at "educational option" programs, which I explain below. Applicants are also categorized according to whether they are in the top 2% of the grade 7 English Language Arts (ELA) test.

 School/program characteristics. In the NYC high school system, there are three types of schools: schools that actively evaluate applicants and submit a ranking to the mechanism; schools that do not evaluate applicants, and instead order students by priorities, which are determined not at the school, but by the DOE; and schools at which a fraction of seats are reserved for students who are explicitly ranked by the school, while the rest are automatically categorized into priority groupings set by the DOE. "Screened" and "audition" schools are examples of the first type of school, at which staff review applicants based on criteria such as seventh grade academic performance, attendance, disciplinary actions, auditions, portfolio submissions, and interviews. "Unscreened" schools are examples of the second type of school. Priorities are based on geographic location, current middle school, or other criteria. Finally, the third class of schools, "educational option," are permitted to rank students for half of their positions, and are required to admit students according to priorities for the other half. Nearly half of all schools are educational option, and more than half of total district capacity is at schools that do not actively rank students. When priorities are used at unscreened and educational option programs, many students fall into the same priority class. Ties or indifferences within each priority group are broken by a single lottery commonly used by all schools.

 "Low academic performance" is a dummy for being categorized as one of the "schools in need of improvement" in the official school brochure issued by the DOE. Under the No
Child Left Behind Act, New York State establishes annual performance goals in Mathematics and English Language Arts for all NYC public schools. Schools that do not meet these goals for two consecutive years are identified as schools in need of improvement.

I say a school program is “new” if it was created in or after 2002. Otherwise, I call it “old.” Program capacities are not provided in the data. I have estimated program capacities from the assignment. Specifically, I define the capacity of a program as the number of students assigned to it in the main application process; this method should be justified because most schools reject at least some applicants in the application process and the capacity of any of these schools should be the same as the number of applicants assigned to it. I say a school is “tiny” if it is in the bottom 25% of the distribution of school capacities among schools in the sample. “Small,” “medium,” and “large” correspond to 25-50%, 50-75%, and 75-100% respectively.

Distance. ArcGIS (with the address-set in the Business Analyst toolbox version 10.0) is used to geocode student and school addresses and calculate the distance between them on the road network. An exact match was first used to determine if a student’s address can be geocoded precisely. If the results were unreliable, the student is assigned to the centroid of the zip-code. The vast majority of students were placed at the roof-top level. The OD Cost matrix tool in the Network Analyst toolbox was used to compute the distance by road for each student-school pair. The road network is also obtained from the Business Analyst toolbox.

1.6.2 A First Look at Evolving Choices: Details

Construction of Categories in Figure 1-2

For creating Figure 1-2, I randomly picked 10% of reapplication reasons and grouped them into categories mentioned in the Figure. To deal with multiple reasons reported by a single applicant, I first attached an equal weight of one to each reapplicant. When a single reapplicant refers to multiple reasons, I divide the reapplicant’s weight equally across all referred reasons. “Moving” contains all reasons that refer to address changes after the initial application process. “Distance” includes the other distance-related reasons that do not refer to address changes. “Mistake in Application” contains reasons that mention mistakes such as misspelling program codes in the initial application or not intending to apply for the initially assigned school. “Current students” contains reasons related to current students at the initially assigned school. For example, some families complain that current students are so scary that they do not want to go to the initially assigned school.
Construction of Regressors in Table 1.4

In Table 1.4, “Location (borough) dummies” are full dummies for being in each of Manhattan, Brooklyn, Queens, Bronx, and Staten Island. “Academic performance dummy” is the low academic performance dummy defined in Appendix 1.6.1. “Program type dummies” are full dummies for program types explained in Appendix 1.6.1. “Capacity dummies” are dummies for “tiny”, “small”, and “medium” defined in Appendix 1.6.1.

1.6.3 Uncovering Evolving Demand: Details

Identification: Details

Section 1.3.2 explains how to use admissions lotteries to separately identify switching costs and demand changes. Figure 1-4 converts the identification logic into suggestive descriptive evidence of significant demand changes and switching costs, but only under the simplifying assumption that initially assigned schools are completely randomly assigned. In the NYC school choice system, however, initial assignments are not purely random since initial assignments also depend on non-random preferences of applicants and their priorities at schools. This section extends the evidence of switching costs and demand changes in Figure 1-4 to the more complicated real case with partially random initial assignments.

As already explained in Section 1.3.2, as far as identification is concerned, partial randomization does not cause any serious problem: I can repeat the same analysis as in Figure 1-4 conditional on each applicant type (i.e., entire preference list and priorities at all schools) to separately identify switching costs and demand changes for that type. There is a problem, however, when I try to extend the empirical implementation of the identification argument in Figure 1-4 to partially random initial assignments. Since applicant type is a high-dimensional object (e.g., with 750 schools, the number of possible preferences is $750 \times 749 \times 748 \times ...$), few applicants usually share the same type in any reasonably sized data. This fact makes it infeasible to draw a conditional-on-full-type version of Figure 1-4. I therefore need to resort to a different strategy that conditions on something coarser to make initial assignments random.

Following the standard notation in econometrics of program evaluation, let $D_{as} = 1$ if applicant $a$ is assigned school $s$; $D_{as} = 0$ otherwise. $\succ_a^0 s'$ denotes applicant $a$’s old preference $a$ submits in her initial application. $s \succ_a^0 s'$ means applicant $a$ prefers school $s$ over $s'$ in her initial application. Let $\rho_{as}$ be $a$’s priority at $s$ (before tie-breaking by lotteries). Define $First_s \equiv \{a | s \succ_a^0 s' \text{ for all school } s' \neq s \text{ and there exists applicant } a' \text{ such that } \rho_{as} = \rho_{a's} \text{ and } D_{as} \neq D_{a's}\}$ as the set of applicants who rank school $s$ first and are in $s$’s “marginal priority group” where some applicants are assigned $s$ but others are not though all of them share the same priority at $s$. Since all applicants in $First_s$ rank school $s$ first and share the same priority at $s$, whether they get assigned to $s$ should be determined
solely by their lottery numbers. Therefore, assignments to school \( s \) within \( First_s \) can be thought as if being randomly assigned in a randomized controlled trial.

I use this strategy to create a structural or causal version of Figure 1-4 as follows. Let me pool \( First_s \) across all schools into a single sample \( \cup_s First_s \). Within \( \cup_s First_s \), consider the following regression (linear probability model) or its nonlinear logit or probit version:

\[
Y_a = \beta D_a + \sum_s \alpha_s X_{as} + \epsilon_a,
\]

where \( Y_a \equiv 1\{\text{applicant } a \text{ reappplies}\} \), \( D_a \equiv 1\{\text{applicant } a \text{ is assigned to her first choice school}\} \), and \( X_{as} \equiv 1\{\text{applicant } a \text{ ranks school } s \text{ first}\} \). By the above argument, \( D_a \) is asymptotically randomly assigned conditional on \( X_{as} \)'s (the identity of the first choice school) and thus estimated \( \hat{\beta} \) from the above regression is causally or structurally interpretable. In fact, Table 1.6.12 confirms that conditional on being in \( \cup_s First_s \) and \( X_{as} \)'s (the identity of the first choice school), baseline covariates are balanced between applicants who do and do not get first choice offers. Such covariate balance is lost, however, if I run the same regression with no controls. This suggests that \( D_a \) is indeed conditionally randomly assigned as intended.

Table 1.6.13 reports \( \hat{\beta} \) alongside the corresponding marginal effect estimates from the probit and logit versions of the above regression. The causal effect of being assigned to the first choice school on the probability of reapplying is precisely estimated and about 6%. The omitted variable bias appears to be small and the descriptive effect of being assigned to the first choice school on the probability of reapplying is about 7%.

Finally, let \( \hat{\alpha} \equiv \frac{\sum_s |First_s| \hat{\alpha}_s}{\sum_s |First_s|} \) be the weighted average of \( \hat{\alpha}_s \) across all schools with the weight being the number of applicants in \( First_s \), who are randomly assigned to or rejected by each school. Each \( \hat{\alpha}_s \) is the conditional probability of reapplying conditional on being assigned to a non-first-choice school and in \( First_s \). Recall that \( \hat{\beta} \) is a weighted average of the causal effects of being assigned to the first choice school on the probability of reapplying. Therefore, \( \hat{\alpha} + \hat{\beta} \) and \( \hat{\alpha} \) are (weighted averages of) the conditional probabilities of reapplying conditional on being assigned to the first choice school and a non-first-choice school, respectively, and in \( \cup_s First_s \), where applicants are randomly assigned between the first choice school and a non-first-choice school.

Figure 1.6.13 plots \( \hat{\alpha} + \hat{\beta} \) and \( \hat{\alpha} \) at \( x = 1 \) and \( x = lower \), respectively, focusing on applicants in \( \cup_s First_s \). Since applicants in the sample \( \cup_s First_s \) used for creating Figure 1.6.13 are purely randomly assigned between the two points on the \( x \) axis, Figure 1.6.13 can be interpreted as a causal or structural version of Figure 1-4. By the logic in Figure 1-4 and Section 1.3.2, Figure 1.6.13 structurally confirms that there are both significant switching costs and demand changes (at least for the subpopulation of applicants I focus on).
More formally, I can test the presence of switching costs and demand changes. If there were neither switching costs nor demand changes, all applicants assigned to a lower choice school would reapply while none of applicants assigned to the first choice school would reapply. This behavioral hypothesis corresponds to the statistical hypothesis that \( \alpha + \beta = 0 \) and \( \alpha = 1 \). Since \( \alpha \) and \( \beta \) are linear combinations of linear regressions coefficients \( \alpha_s \)'s and \( \beta_s \)'s, null hypotheses \( H_0 : \alpha + \beta = 0 \) and \( H_0 : \alpha = 1 \) are linear hypotheses. I can test these hypotheses by the usual Wald test.

Similarly, if there were no switching costs (but there may be demand changes), conditional on being in \( U_{First} \),

\[
\alpha + \beta = \Pr(a \text{ reapplies}|s_a^0 \text{ is } a's \text{ old } 1st \text{ choice})
\]
\[
= \sum_{K=2}^{K} \Pr(a \text{ does not reapply}|s_a^0 \text{ is } a's \text{ old } K-th \text{ choice})
\]
\[
= \sum_{K=2}^{K} \Pr(a \text{ does not reapply}|s_a^0 \text{ is } a's \text{ old } K-th \text{ choice})
\]
\[
\times \Pr(s_a^0 \text{ is } a's \text{ old } K-th \text{ choice}|s_a^0 \text{ is not } a's \text{ old } 1st \text{ choice})
\]
\[
= 1 - \alpha,
\]

where the inequality is because \( \Pr(a \text{ does not reapply}|s_a^0 \text{ is } a's \text{ old } K-th \text{ choice}) \geq 0 \) and \( \Pr(s_a^0 \text{ is } a's \text{ old } K-th \text{ choice}|s_a^0 \text{ is not } a's \text{ old } 1st \text{ choice}) \leq 1 \). Thus, the behavioral hypothesis of no switching costs corresponds to the statistical hypothesis that \( \alpha + \beta \geq 1 - \alpha \).

Table 1.6.14 shows that the Wald test rejects both the null hypothesis that there are no switching costs \( (\alpha + \beta = 1 - \alpha) \) and the null hypothesis that there are neither switching costs nor demand changes \( (\alpha + \beta = 0 \text{ or } \alpha = 1) \).

**Estimation: Details**

**Likelihood Derivation and Estimation Procedure**

In this section, I derive partly analytical likelihood functions from my empirical models (in Section 1.3.1) combined with the distributional assumptions in Section 1.3.3. The likelihood functions at the family level are computed for a sequence of choices from the initial application period \( t = 0 \) to the reapplication period \( t = 1 \), since due to initial assignment effects and reapplication costs, the likelihood of a choice in \( t = 1 \) depends on the choice and initial assignment in the previous period \( t = 0 \). Let \( s_{at}^{t} \) be applicant \( a \)'s \( t \)-th choice school in her period \( t \) preference \( \succeq_{a}^{t} \) \( (t = 0,1) \) and let \( \#_{a}^{t} \) be the number of schools \( a \) ranks in \( \succeq_{a}^{t} \).

In period 0 (the initial application period), every applicant \( a \) in the sample submits \( s_{at}^{0} \) and it implies that \( U_{asa}^{0} > U_{as}^{0} \) for every \( l = 1, \ldots, \#_{a}^{0} \) and every \( s \) with \( s_{at}^{0} \succeq_{a}^{0} s \), including unranked schools. By the formula for logit choice probabilities (Train 2009 chapter 3),
conditional on \((U^0 \equiv (U_s^0)_s, (\beta_{ak}(1 + f_{ak}))_{k=1,...,K})\), the likelihood of observing \( \succ^0_a \) is

\[
L^0_a \equiv \prod_{l=1}^{\#^0_a} \frac{\exp(U^0_{a_{sl}} + \sum_{k=1}^K \beta_{ak}(1 + f_{ak})X_{a_{sk},l})}{\sum_{s' \text{ with } s'_{al} \succ \succ^0_a} \exp(U^0_{s'} + \sum_{k=1}^K \beta_{ak}(1 + f_{ak})X_{ask})}.
\]

In period 1 (the reapplication process), even if applicant \( a \) does not reapply and does not submit new preference \( \succ^1_a \), she provides useful information. In particular, in the rational expectation model, applicant \( a \) reapplies if equation (3) in the main text holds, implying that conditional on \((U^1 \equiv (U^1_l)_s, (\beta_{ak}(1 + f_{ak}))_k, (\beta_{ak}f_{ak})_k, \epsilon^0_a \equiv (\epsilon^0_{as})_s, c_a/p_a, \gamma_a)\), the likelihood of observing \( a \)'s non-reapplication is

\[
L^1_{a, \text{no reap, rational}} \equiv 1 - \int 1 \{\text{equation (3) holds}\} dF_{\theta},
\]

where \( F_{\theta} \) is the distribution of utility function arguments, parametrized by \( \theta \). If applicant \( a \) reapplies and submits \( \succ^1_a \), then it implies that (I) equation (3) holds, (II) \( p_{as}^1 U_{as}^1 > p_{as} U_{as} \) for each \( l = 1,...,\#^1_a \) and every unranked \( s \neq s^0_a \), and (III) if \( a \) does not exhaust her reapplication preference, i.e., \( \#^1_a < 3 \) (recall any reapplicant can rank up to three schools in her reapplication preference), then \( U_{as}^1 + \gamma_a > U_{as}^1 \) for all \( s(\neq s^0_a) \) which \( a \) does not rank in \( \succ^1_a \). Thus, the likelihood of observing \( a \)'s reapplication and \( \succ^1_a \) is

\[
L^1_{a, \text{reap, rational}} = \frac{\prod_{l=1}^{\#^1_a} 1 \{p_{as}^1 U_{as}^1 > p_{as} U_{as} \} \text{ for all unranked } s \neq s^0_a}{\sum_{s \text{ unranked in } \succ^1_a} \exp(U^1_s + \sum_{k=1}^K \beta_{ak}X_{ask} + 1 \{s = s^0_a\} \gamma_a + \epsilon^0_{as})} \times [1(\#^1_a < 3)]
\]

where \( F'_{\theta} \) is the distribution of relevant utility function arguments conditional on (III). If \( a \) exhausts her reapplication preference, i.e., \( \#^1_a = 3 \), then \( F'_{\theta} \) is the same as the unconditional distribution.

In the naive free expectation model, not reapplying implies that \( U_{as}^1 + c_a/p_a + \gamma_a > U_{as}^1 \) for every \( s \neq s^0_a \). By the formula for logit choice probabilities, the likelihood of observing \( a \)'s non-reapplication is

\[
L^1_{a, \text{no reap, naive}} \equiv \frac{\exp(U^1_{s^0_a} + \sum_{k=1}^K \beta_{ak}X_{a_{sk}} + \epsilon^0_{as} + c_a/p_a + \gamma_a)}{\sum_s \exp(U^1_s + \sum_{k=1}^K \beta_{ak}X_{ask} + 1 \{s = s^0_a\} (c_a/p_a + \gamma_a) + \epsilon^0_{as})}.
\]

If applicant \( a \) reapplies and submits \( \succ^1_a \), then it implies that (i) it is not the case that \( U_{as}^1 + c_a/p_a + \gamma_a > U_{as}^1 \) for every \( s \neq s^0_a \), (ii) \( U_{as}^1 > U_{as}^1 \) for each \( l = 1,...,\#^1_a \) and
every $s$ with $s_{a1}^1 > \gamma_a$, and (iii) if $a$ does not exhaust her reapplication preference, i.e., $\#_a^1 < 3$ (recall any reapplicant can rank up to three schools in her reapplication preference), then $U_{a0} + \gamma_a > U_{as}$ for all $s(\neq s_a^0)$ which $a$ does not rank in $\gamma_a$. Conditional on $(U^1, (\beta_{ak}(1 + f_{ak}))_k, (\beta_{ak} f_{ak})_k, \epsilon_{a1}, c_a/pa, \gamma_a)$, the likelihood of observing $a$’s reapplication and $\gamma_a$ is

\[
L_a^{1, reap, naive} = \prod_{s=1}^{\#_a^1} \mathbb{1}\{U_{as} > U_{as}^1 \text{ for every } s \neq s_a^0\} \\
\times \frac{\exp(U_{a0}^1 + \sum_{k=1}^{K} \beta_{ak} X_{as}^2 k + \gamma_a + \epsilon_{a0}^0)}{\sum_{s \text{ unranked in } \gamma_a^1} \exp(U_{a0}^1 + \sum_{k=1}^{K} \beta_{ak} X_{as}^2 k + \mathbb{1}\{s = s_a^0\} \gamma_a + \epsilon_{a0}^0)} + (1 - \mathbb{1}\{\#_a^1 < 3\}).
\]

For each model, integrating over $(U^0, U^1, (\beta_{ak}(1 + f_{ak}))_k, (\beta_{ak} f_{ak})_k, \epsilon_{a1}, c_a/pa, \gamma_a)$ and all applicants gives the full, unconditional likelihood as follows:

\[
L(\theta) \equiv \Pi_a L_a^0(\mathbb{1}\{a \text{ reapplies}\}) L_a^{1, reap} + (1 - \mathbb{1}\{a \text{ reapplies}\}) L_a^{0, no reap} dF_\theta.
\]

For estimating $\theta$ under each model, I find $\hat{\theta}$ that maximizes the simulated version of the logarithm of the likelihood function. More precisely, for easing the computational burden, I take a two step procedure. The first step estimates the part of $\theta$ related to $U^0$ and $\beta_{ak}(1 + f_{ak})$ by maximizing the simulated version of the logarithm of the partial likelihood $\Pi_a L_a^0 dF_\theta$. Taking the estimated part of $\theta$ as given, the second step estimates the remaining part of $\theta$ by maximizing the simulated version of the logarithm of the remaining conditional likelihood for period 1, i.e., $\Pi_a L_a^{1, reap} + (1 - \mathbb{1}\{a \text{ reapplies}\}) L_a^{1, no reap} dF_\theta$. This two step approach is legitimate since (1) the estimation target of the first step $(U^0$ and $\beta_{ak}(1 + f_{ak}))$ is identified by period 0 preference $\gamma_a$ alone, and maximizing the period 0 partial likelihood gives us consistent estimates with conservative standard errors, and (2) the estimation target of the second step depends only on the period 1 conditional likelihood, and standard results for estimators based on maximizing a partial or conditional likelihood guarantee its consistency and asymptotic normality (Wong, 1986).

Finally, when simulating old utilities $U_{as}$’s in the second step, I do so conditional on applicant $a$’s observed initial application preference $\gamma_a$. This conditional simulation seems more appropriate than unconditional simulation for incorporating the fact that

\[\footnote{In particular, within each simulation, I first simulate $\beta_{ak}(1 + f_{ak})$ and $U_a^0$ unconditionally and take them as fixed. Then I simulate unobserved utility shocks $\epsilon_{as}^0$’s conditional both on initial application preference $\gamma_a$ and the simulated values of $\beta_{ak}(1 + f_{ak})$ and $U_a^0$ so that the resulting period 0 utilities $U_{as}$’s become distributed conditional on initial application preference $\gamma_a$.} \]
observed initial assignment \( s^0_\alpha \) (which is used in the estimation) is one of the twelve most preferred schools in initial application preference \( \succ^0_\alpha \) and underlying old utilities \( U^0_\alpha \)'s. If I simulate old utilities unconditionally, observed initial assignment \( s^0_\alpha \) is often not such a preferred school in simulated period 0 utilities (especially because the number of schools is large). I then need large reapplication costs and initial assignment effects to explain the low reapplication rate, which may result in overestimating reapplication costs and initial assignment effects.

### Estimation Sample Construction

**Schools.** 755 school programs are ranked by some students in the initial application process. Among them, I had to drop 5 schools that I could not geocode because I did not receive their address information. I also dropped applicants' choices associated with these dropped schools. This left 750 schools in the estimation and counterfactual sample.

**Applicants.** 91289 students make applications and submit preferences in the initial application process. Starting from them, I needed to focus on students who rank at least one in-sample school in the initial application process, since other students reveal no information about their preferences over in-sample schools. I excluded students in the top 2 percent of the grade 7 English language arts test score by the following reason: any students in the top 2 percent are guaranteed assignment to their first choice if they rank a program of a particular type, known as "educational option," as their first choice. A student who does not prefer an educational option program as her top choice may thus have an incentive to strategize and rank it as her top choice so that she receives it. If so, their stated preferences do not follow the truth-telling assumption necessary for identification. I also had to drop 9th graders, who participate in the mechanism for the second time and are potentially subject to different types of learning or effects of a prior assignment on utilities, and students from private schools, for whom I do not observe necessary demographic information. Finally, I dropped students for whom no address information was available, because I could not geocode them. As a result of this process, 76250 students remain in the estimation and counterfactual sample.

**Choices.** A tiny number of choices are recorded without preference ranks. I had to drop these choices.

### 1.6.4 Theory for the Counterfactual Analysis

**Model**

There are a finite set \( A \) of applicants and a finite set \( S \) of schools. Each school \( s \in S \) has a strict preference/priority relation \( \succ_s \) over the set of subsets of \( A \). In many settings including the NYC system, school priorities are coarse or weak and lotteries are used to
break ties or indifferences. In the theoretical analysis, I do not explicitly consider the tie-breaking process and take ex post strict priorities as given.

I assume that the preference relation of each school is responsive with capacity \( q_s \) (Roth and Sotomayor, 1990), i.e.,

1. For any \( a, \bar{a} \in A \), if \( \{a\} \succ_s \{\bar{a}\} \), then for any \( A' \subseteq A \smallsetminus \{a, \bar{a}\} \), \( A' \cup \{a\} \succ_s A' \cup \{\bar{a}\} \).
2. For any \( a \in A \), if \( \{a\} \succ_s \emptyset \), then for any \( A' \subseteq A \) such that \( |A'| < q_s \), \( A' \cup \{a\} \succ_s A' \), and
3. \( \emptyset \succ_s A' \) for any \( A' \subseteq A \) with \( |A'| > q_s \).

If a school’s preferences are responsive, then that school acts as if it has preferences over individual applicants and a quantity constraint, and the school takes the available highest-ranking applicants up to that quantity constraint. In addition, I assume that every applicant is acceptable to every school because I am primarily interested in the assignment of applicants to public schools as in the NYC system. The preference profile of all schools is denoted by \( \succ_s \equiv (\succ_s)_{s \in S} \). In this model, schools’ preferences can be interpreted as their intrinsic preferences or priorities exogenously given by law or the authorities. In NYC, the former interpretation is appropriate for some schools (e.g., “screened” schools) while the latter interpretation is suitable for other schools (e.g., “unscreened” schools). The model and the results are applicable to both interpretations.

On top of the above usual structure, the model specifies applicant preferences and allows them to evolve over time. In particular, there are two periods, 0 and 1. I interpret period 0 as the currently used application period (e.g., December) and period 1 as a later point in time that can potentially be used as an alternative application period (e.g., following April). In each period \( t = 0, 1 \), each applicant \( a \in A \) has a strict preference relation \( \succ_a^t \) over \( S \cup \{\emptyset\} \), where \( \emptyset \) denotes the outside option of the applicant. I distinguish \( \emptyset \) and \( \emptyset \), where \( \emptyset \) denotes an outside option while \( \emptyset \) is the empty set. \( \succ_a^1 \) may or may not be the same as \( \succ_a^0 \). \( \succ_a^1 \) is supposed to be a better measure of applicants’ welfare than \( \succ_a^0 \) since \( \succ_a^1 \) is at a later point in time and is likely to approximate better applicants’ preferences when they finally experience educational services at assigned schools. Unless otherwise noted, I impose no restriction on the relationship between \( \succ_a^0 \) and \( \succ_a^1 \). The weak preference relation associated with \( \succ_a^t \) is denoted by \( \succsim_a^t \) and so I write \( s \succsim_a^t \bar{s} \) if either \( s \succ_a^t \bar{s} \) or \( s = \bar{s} \). A preference profile of all applicants in period \( t \) is denoted by \( \succ_a^t \equiv (\succ_a^t)_{a \in A} \).

**Matching and Its Properties.** An outcome of the model is a matching, which is a vector \( \mu = (\mu_a)_{a \in A} \) that assigns each applicant \( a \) a seat at a school (or the outside option) \( \mu_a \in A \cup \{\emptyset\} \), and where each school \( s \in S \) is assigned at most \( q_s \) applicants. I denote by \( \mu_s \equiv \{a \in A : s = \mu_a\} \) the set of applicants who are assigned to school \( s \).
Let me introduce several desirable properties of a matching. I often use "w.r.t." to mean "with respect to." A matching \( \mu \) is individually rational w.r.t. \( \succ_A \) if \( \mu_a \succ_A A \) for every \( a \in A \). \( \mu \) is blocked by \( (a, s) \in A \times S \) w.r.t. \( (\succ_A, \succ_S) \) if \( s \succ_A a \) and there exists \( A' \subseteq \mu_a \cup a \) such that \( A' \succ_S \mu_a \). (I denote singleton set \( \{x\} \) by \( x \) when there is no room for confusion.) \( \mu \) is fair (stable) w.r.t. \( (\succ_A, \succ_S) \) if it is individually rational w.r.t. \( \succ_A \) and not blocked w.r.t. \( (\succ_A, \succ_S) \). A matching \( \mu \) is Pareto efficient for applicants w.r.t. \( \succ_A \) if there exists no matching \( \mu' \) such that \( \mu'_a \succ_A a \) for all \( a \in A \) and \( \mu'_a \succ_A a \) for at least one \( a \in A \). \( \mu \) is weakly Pareto efficient for applicants w.r.t. \( \succ_A \) if there exists no matching \( \mu' \) such that \( \mu'_a \succ_A a \) for all \( a \in A \).

**Mechanism and Its Properties.** Given the set of applicants \( A \) and schools \( S \), a (direct) mechanism is a function \( \varphi \) that maps each \( (\succ_A, \succ_S) \) into a matching. Since the domain contains both \( \succ_A \) and \( \succ_S \), this definition allows both for static and dynamic mechanisms. The fact that NYC runs the discretionary reapplication process after the initial match suggests that it is feasible to elicit preferences at two different points in time. For example, as explained below, the current initial match mechanism in NYC is static while some of the alternatives considered in this paper are dynamic; all of them can be described as examples of this definition. A mechanism \( \varphi \) is (weakly) Pareto efficient for applicants w.r.t. \( \succ_A \) if \( \varphi(\succ_A, \succ_S) \) is (weakly) Pareto efficient w.r.t. \( \succ_A \) for every \( (\succ_A, \succ_S) \). \( \varphi \) is fair (stable) w.r.t. \( (\succ_A, \succ_S) \) if \( \varphi(\succ_A, \succ_S) \) is fair w.r.t. \( (\succ_A, \succ_S) \) for every \( (\succ_A, \succ_S) \). I introduce the following algorithm to compute a matching.

**Gale and Shapley’s (Applicant-Proposing) Deferred Acceptance (DA) Algorithm.** Given any applicant preference profile \( \succ_A \), for example, \( \succ_A^0 \) or \( \succ_A^1 \). Given \( (\succ_A, \succ_S) \), the (applicant-proposing) deferred acceptance (DA) algorithm is defined as follows.

- **Step 1:** Each applicant \( a \in A \) applies to her most preferred acceptable school w.r.t. \( \succ_a \) (if any). Each school tentatively keeps the highest-ranking applicants up to its capacity, and rejects every other applicant.

In general, for any step \( t \geq 2 \),

- **Step \( t \):** Each applicant \( a \) who was not tentatively matched to any school in Step \( (t-1) \) applies to her most preferred acceptable school w.r.t. \( \succ_a \) that has not rejected her (if any). Each school tentatively keeps the highest-ranking applicants up to its capacity from the set of applicants previously tentatively matched to this school and the applicants newly applying, and rejects every other applicant.

The algorithm terminates at the first step at which no applicant applies to a school. Each applicant tentatively kept by a school at that step is allocated a seat in that school, resulting in a matching which I denote by \( DA(\succ_A, \succ_S) \). It is known that \( DA(\succ_A, \succ_S) \) is a fair and weakly Pareto efficient matching with respect to \( (\succ_A, \succ_S) \) for any \( (\succ_A, \succ_S) \).
As already explained in Section 1.2.1, the authorities in NYC ask applicants to submit their preferences around Nov and Dec. They then apply the DA algorithm to the submitted preferences to compute the matching. Recall that period 0 in my model corresponds to Nov or Dec. This leads me to interpret the current initial match mechanism as \( \varphi_{\text{initial}}^{DA}(\succ_A^0, \succ_A^1, \succ_S) = DA(\succ_A^0, \succ_S) \), i.e., the one that produces a matching by applying the DA algorithm to applicant preference in period 0, \( \succ_A^0 \). By the above properties of \( DA(\succ_A, \succ_S) \) with respect to its input preferences, \( \varphi_{\text{initial}}^{DA} \) has nice properties such as fairness and weak Pareto efficiency with respect to \( \succ_A^0 \). However, the empirical analysis in Sections 1.2 and 1.3 suggests that families’ preferences for schools change from \( t = 0 \) to \( t = 1 \). Under \( \succ_A^1 \), therefore, \( \varphi_{\text{initial}}^{DA} \) may not retain the desirable properties. In fact, I empirically show in Section 1.4 that \( \varphi_{\text{initial}}^{DA} \) produces non-negligible efficiency losses under \( \succ_A^1 \).

To improve \( \varphi_{\text{initial}}^{DA} \), Section 1.4.2 describes two alternatives mechanisms, the dynamic deferred acceptance mechanism \( \varphi_{\text{dynamic}}^{DA}(\succ_A^0, \succ_A^1, \succ_S) = DA(\succ_A^1, \succ_S) \), where \( \succ_S \) denotes modified school priorities defined in Section 1.4.2. Precisely speaking, \( \succ_S \) is defined as follows. For all \( s \),

- \( a \succ_s a' \) for any \( a \in DA_s(\succ_A^0, \succ_S) \) and any \( a' \notin DA_s(\succ_A^1, \succ_S) \) and
- \( a'' \succ_s a''' \) if and only if \( a'' \succ_s a''' \) for any \( a'', a''' \notin DA_s(\succ_A^0, \succ_S) \) or any \( a'', a''' \in DA_s(\succ_A^1, \succ_S) \)

The deferred deferred acceptance mechanism \( \varphi_{\text{deferred}}^{DA}(\succ_A^0, \succ_A^1, \succ_S) = DA(\succ_A^1, \succ_S) \). If \( \succ_A^0 = \succ_A^1 \), i.e., there is no preference change, then both alternatives reduce to the current one, i.e., \( \varphi_{\text{dynamic}}^{DA}(\succ_A^0, \succ_A^1, \succ_S) = \varphi_{\text{deferred}}^{DA}(\succ_A^0, \succ_A^1, \succ_S) = \varphi_{\text{initial}}^{DA}(\succ_A^0, \succ_A^1, \succ_S) \). Thus, neither of them performs worse than the current one even if there is no preference change.

The next section provides a formal characterization of the relationship among these mechanisms. In particular, I show that \( \varphi_{\text{dynamic}}^{DA} \) and \( \varphi_{\text{deferred}}^{DA} \) are the only “best possible” mechanisms in terms of efficiency and fairness with respect to \( \succ_A^1 \) and strategy-proofness. This result provides a theoretical foundation for considering these alternatives in the counterfactual analysis.

**Result**

In addition to the usual fairness and efficiency properties defined above, I introduce additional criteria to compare alternative reapplication mechanisms with \( \varphi_{\text{initial}}^{DA} \), the status quo initial match mechanism currently in place in many cities including NYC.

**Definition 1.** A mechanism \( \varphi \) always Pareto dominates another mechanism \( \varphi' \) w.r.t. \( \succ_A^1 \) if (1) for any \( (\succ_A^0, \succ_A^1, \succ_S) \) and any \( a, \varphi_a(\succ_A^0, \succ_A^1, \succ_S) \succ_a \varphi'_a(\succ_A^0, \succ_A^1, \succ_S) \) and (2) there are \( (\succ_A^0, \succ_A^1, \succ_S) \) and \( a \) such that \( \varphi_a(\succ_A^0, \succ_A^1, \succ_S) \succ_a \varphi'_a(\succ_A^0, \succ_A^1, \succ_S) \).

64
Definition 2. A mechanism $\varphi$ is less unfair than another mechanism $\varphi'$ w.r.t. $(>^t_A, >_s)$ if (1) for any $(>^0_A, >^1_A, >_s)$ and any $(a, s)$, if $(a, s)$ blocks $\varphi((>^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$, then $(a, s)$ also blocks $\varphi'(>(^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$, and (2) there are $(>^0_A, >^1_A, >_s)$ and $(a, s)$ such that $(a, s)$ blocks $\varphi'(>(^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$ but $(a, s)$ does not block $\varphi((>^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$.

In words, $\varphi$ is less unfair than $\varphi'$ if it is always the case that any applicant-school pair blocking the outcome under $\varphi$ also blocks that under $\varphi'$. By definition, if $\varphi$ is fair with respect to $(>^t_A, >_s)$ but $\varphi'$ is not, then $\varphi$ is less unfair than $\varphi'$ with respect to $(>^t_A, >_s)$.

Another possible definition of being less unfair is the following: (1) for any $(>^0_A, >^1_A, >_s)$ and any $a$, if there is $s$ such that $(a, s)$ blocks $\varphi((>^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$, then there is $s'$ such that $(a, s')$ blocks $\varphi'(>(^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$ ($s'$ may or may not be the same as $s$) and (2) there are $(>^0_A, >^1_A, >_s)$ and $a$ such that there is $s'$ such that $(a, s')$ blocks $\varphi'(>(^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$ but for any $s$, $(a, s)$ does not block $\varphi((>^0_A, >^1_A, >_s)$ w.r.t. $(>^t_A, >_s)$. The results below hold under either definition.

All properties defined so far are about fairness or efficiency with respect to stated preferences. To make sure that stated preferences reflect true ones, I would also like a mechanism to be incentive compatible.

Definition 3. A mechanism $\varphi$ is dynamically strategy-proof if the following holds for any $(>^0_A, >^1_A, >_s)$, any $a$, and any $(>^0_a, >^1_a)$: $\varphi_a((>^0_A, >^1_A, >_s) \succ_a^0 \varphi_a((>^0_a, >^1_a, >^0_a, >^1_a, >_s))$ and $\varphi_a((>^0_A, >^1_A, >_s) \succ_a^1 \varphi_a((>^0_A, >^1_a, >^0_a, >^1_a, >_s))$.

In words, under a dynamically strategy-proof mechanism, any preference manipulation by any applicant in any period is never strictly profitable with respect to that applicant’s preference in that period. I allow each applicant $a$ to manipulate both $>^0_a$ and $>^1_a$ in $t = 0$ but only $>^1_a$ in $t = 1$. This restriction is justified under the interpretation that $>^0_a$ is already reported and fixed in $t = 1$. When $\varphi$ is static, i.e., uses only $>^0_A$ or $>^1_A$ to compute the matching $\varphi((>^0_A, >^1_A, >_s)$, usual static strategy-proofness is the same as dynamic strategy-proofness.

In principle, I can define dynamic strategy-proofness in a more restrictive way, e.g., for any $(>^0_A, >^1_A, >_s)$ and any $a$ and any $(>^0_a, >^1_a)$, $\varphi_a((>^0_a, >^1_a, >^0_a, >^1_a, >^0_a, >^1_a, >_s) \succ_a^t \varphi_a((>^0_A, >^1_A, >_s))$ for both $t = 0, 1$. In the presence of potential preference reversals between $>^0_a$ and $>^1_a$, however, this definition is so restrictive that it is not satisfied even by static strategy-proof mechanisms such as $\varphi_{\text{initial}}$ since it is possible $\varphi_a((>^0_a, >^0_a, >^1_a, >^0_a, >^1_a, >^0_a, >^1_a, >_s) \succ_a^0 \varphi_a((>^0_A, >^1_A, >_s))$ but $\varphi_a((>^0_A, >^1_A, >_s) \succ_a^1 \varphi_a((>^0_a, >^0_a, >^1_a, >^0_a, >^1_a, >_s))$.

I regard the following five as basic desirable properties of a mechanism in my dynamic model with evolving preferences, where $>^1_a$ is the measure of applicants’ welfare. Note that period 1 is closer to enrollment periods and the empirical analysis in the main body suggests that $>^1_a$ appears to be subject to less severe information frictions about schools.
(I) Fairness with respect to $(\succ_A^1, \succ_S)$

(II) Being less unfair than $\varphi_{\text{initial}}^{DA}$ with respect to $(\succ_A^1, \succ_S)$

(III) Weak Pareto efficiency with respect to $\succ_A^1$

(IV) Always Pareto dominating $\varphi_{\text{initial}}^{DA}$ with respect to $\succ_A^1$

(V) Dynamic strategy-proofness

Properties (II) and (IV), which compare an alternative mechanism with $\varphi_{\text{initial}}^{DA}$, are important since many cities including NYC are currently using $\varphi_{\text{initial}}^{DA}$ as the status quo initial match. $\varphi_{\text{initial}}^{DA}$ satisfies only (V) among (I) to (V). My goal is thus to design reapplication mechanisms that improve on $\varphi_{\text{initial}}^{DA}$ and achieve (I)-(V). The following result shows that in terms of (I)-(V), $\varphi_{\text{dynamic}}^{DA}$ and $\varphi_{\text{deferred}}^{DA}$ are the best possible mechanisms I can design.

**Proposition 1.** 1.A) $\varphi_{\text{dynamic}}^{DA}$ satisfies (II) being less unfair than the initial match, (III) weak Pareto efficiency, and (IV) always Pareto dominating $\varphi_{\text{initial}}^{DA}$ (call this set of desiderata A), but not others.

1.B) $\varphi_{\text{deferred}}^{DA}$ satisfies (I) fairness, (II) being less unfair than $\varphi_{\text{initial}}^{DA}$, (III) weak Pareto efficiency, and (V) dynamic strategy-proofness (call this set of desiderata B), but not others.

2) Consider any possible mechanism $\varphi$. $\varphi$ can satisfy only a subset of set A or B.

The proof is in Appendix 1.6.4. Proposition 1 has several implications. First of all, recall that $\varphi_{\text{initial}}^{DA}$ does not satisfy any of desirable welfare properties (I)-(IV). Thus, (1.A) and (1.B) say each of $\varphi_{\text{dynamic}}^{DA}$ and $\varphi_{\text{deferred}}^{DA}$ is better than $\varphi_{\text{initial}}^{DA}$ and achieve some of welfare properties (I)-(IV) in the presence of preference changes. (1.A) and (1.B) also demonstrate certain tradeoffs between $\varphi_{\text{dynamic}}^{DA}$ and $\varphi_{\text{deferred}}^{DA}$. Part (2) shows that these tradeoffs are not resolvable, i.e., I cannot design a mechanism that is strictly better than any of the two mechanisms in terms of the above desiderata. In this sense, $\varphi_{\text{deferred}}^{DA}$ and $\varphi_{\text{dynamic}}^{DA}$ are the best possible alternatives I can obtain.

It may be useful to walk through what tradeoffs Proposition 1 embeds. A tradeoff Proposition 1 implies is that (IV) Pareto dominating $\varphi_{\text{initial}}^{DA}$ and (V) dynamic strategy-proofness are incompatible. This is a version of the classic efficiency-incentive tradeoff. There is another more important tradeoff. Proposition 1 also implies a tradeoff between (I) fairness and (IV) always Pareto dominating $\varphi_{\text{initial}}^{DA}$. This is a version of yet another classic tradeoff between efficiency and fairness.

To understand the implication of this efficiency-fairness tradeoff between (I) fairness and (IV) always Pareto dominating $\varphi_{\text{initial}}^{DA}$, let me consider a weaker version of always Pareto dominating $\varphi_{\text{initial}}^{DA}$. A mechanism $\varphi$ is *never dominated* by another mechanism $\varphi'$ w.r.t. $\succ_A^1$ if there is no $(\succ_A^0, \succ_A^1, \succ_S)$ such that $\varphi'(\succ_A^0, \succ_A^1, \succ_S)$ Pareto dominates $\varphi(\succ_A^0, \succ_A^1, \succ_S)$ w.r.t. $\succ_A^1$. If $\varphi$ dominates $\varphi'$ w.r.t. $\succ_A^1$, then $\varphi$ is never dominated by $\varphi'$ w.r.t. $\succ_A^1$. Thus $\varphi_{\text{dynamic}}^{DA}$ is never dominated by $\varphi_{\text{initial}}^{DA}$ w.r.t. $\succ_A^1$ since $\varphi_{\text{dynamic}}^{DA}$
always Pareto dominates \( \varphi_{\text{initial}}^{DA} \) w.r.t. \( \succ_A^1 \) (by Proposition 1.A). In contrast, \( \varphi_{\text{deferred}}^{DA} \) is sometimes dominated by \( \varphi_{\text{initial}}^{DA} \) w.r.t. \( \succ_A^1 \) (see Lemma 3 in Appendix 1.6.4). Thus, the most obvious reapplication mechanism \( \varphi_{\text{deferred}} \) (waiting until period 1 and applying the DA algorithm to applicant preferences in period 1) may make all applicants worse off with respect to their preferences in period 1. More generally, switching from \( \varphi_{\text{initial}}^{DA} \) to \( \varphi_{\text{deferred}}^{DA} \) produces "losers" who prefer \( \varphi_{\text{initial}}^{DA} \) while there is no such loser in the case of \( \varphi_{\text{dynamic}}^{DA} \). I empirically confirm and quantify this in Section 1.4.

Overall, Proposition 1 provides a theoretical basis for using \( \varphi_{\text{deferred}}^{DA} \) and \( \varphi_{\text{dynamic}}^{DA} \) as the best possible reapplication mechanisms in my counterfactual analysis in Section 1.4, which in turn empirically quantifies the effects of the two mechanisms.

**Proof of Proposition 1**

1.A) \( \varphi_{\text{dynamic}}^{DA} \) always Pareto dominates \( \varphi_{\text{initial}}^{DA} \) w.r.t. \( \succ_A^1 \) since its construction always guarantees any applicant \( a \) a seat at school \( DA_a(\succ_A^0, \succ_A^1, \succ_A^s) \) (if any) or a more preferred school w.r.t. \( \succ_A^1 \). It is weakly Pareto efficient w.r.t. \( \succ_A^1 \) since \( DA(\succ_A^0, \succ_A^1, \succ_A^s) \) is weakly Pareto efficient w.r.t. input preferences \( \succ_A^1 \) for any \( (\succ_A^0, \succ_A^1, \succ_A^s) \) (Abdulkadiroglu and Sonmez, 2003). It is less unfair than \( \varphi_{\text{initial}}^{DA} \) w.r.t. \( (\succ_A^0, \succ_A^1, \succ_A^s) \) by the following Lemma 1. It violates the other two properties by Lemma 2 and Corollary 1 proven below.

**Lemma 1.** \( \varphi_{\text{dynamic}}^{DA} \) is less unfair than \( \varphi_{\text{initial}}^{DA} \) w.r.t. \( (\succ_A^0, \succ_A^1, \succ_A^s) \).

**Proof.** Suppose \((a, s)\) blocks \( \varphi_{\text{dynamic}}^{DA}(\succ_A^0, \succ_A^1, \succ_A^s) \) w.r.t. \( (\succ_A^0, \succ_A^1, \succ_A^s) \), i.e., \( s \succ_a \varphi_{\text{dynamic},a}(\succ_A^0, \succ_A^1, \succ_A^s) \) and \( a \succ_s a' \) for some \( a' \in \varphi_{\text{dynamic},a}(\succ_A^0, \succ_A^1, \succ_A^s) \). (Note that as long as \( s \succ_a \varphi_{\text{dynamic},a}(\succ_A^0, \succ_A^1, \succ_A^s) \), it cannot be the case \( \vert \varphi_{\text{dynamic},a}(\succ_A^0, \succ_A^1, \succ_A^s) \vert < q_s \) by the construction of the DA algorithm.) It has to be the case

\[
(\varphi_{\text{dynamic},a'}(\succ_A^0, \succ_A^1, \succ_A^s) =) s = DA_a'(\succ_A^0, \succ_A^1, \succ_A^s)(= \varphi_{\text{initial},a'}(\succ_A^0, \succ_A^1, \succ_A^s)),
\]

since otherwise \( a \succ_s a' \) implies \( a' \succ_a a' \) and thus \((a, s)\) blocks \( \varphi_{\text{dynamic}}^{DA}(\succ_A^0, \succ_A^1, \succ_A^s) \equiv DA(\succ_A^0, \succ_A^1, \succ_A^s) \) w.r.t. \( (\succ_A^0, \succ_A^1, \succ_A^s) \), a contradiction to the fact that no \((a, s)\) blocks \( DA(\succ_A^0, \succ_A^1, \succ_A^s) \) w.r.t. \( (\succ_A^0, \succ_A^1, \succ_A^s) \) for any \((\succ_A^0, \succ_A^1, \succ_A^s)\). Then

\[
s \succ_a \varphi_{\text{dynamic},a}(\succ_A^0, \succ_A^1, \succ_A^s) \geq_a \varphi_{\text{initial},a}(\succ_A^0, \succ_A^1, \succ_A^s),
\]

where the second weak preference is because \( \varphi_{\text{dynamic}}^{DA} \) always Pareto dominates \( \varphi_{\text{initial}}^{DA} \) w.r.t. \( \succ_A^1 \). Combined with \( \varphi_{\text{initial},a}(\succ_A^0, \succ_A^1, \succ_A^s) \equiv DA_a(\succ_A^0, \succ_A^1, \succ_A^s) = s \), this means \((a, s)\) also blocks \( \varphi_{\text{initial}}^{DA}(\succ_A^0, \succ_A^1, \succ_A^s) \) w.r.t. \( (\succ_A^0, \succ_A^1, \succ_A^s) \). \( \square \)

1.B) \( \varphi_{\text{deferred}}^{DA} \equiv DA(\succ_A^1, \succ_A^s) \) is fair w.r.t. \( (\succ_A^1, \succ_A^s) \) and weakly Pareto efficient w.r.t. \( \succ_A^1 \) since \( DA(\succ_A^1, \succ_A^s) \) has these properties w.r.t. any input preferences (Abdulkadiroglu and Sonmez, 2003). Since \( \varphi_{\text{initial}}^{DA} \) is not fair w.r.t. \( (\succ_A^1, \succ_A^s) \), \( \varphi_{\text{deferred}}^{DA} \)’s fairness implies
that \( \phi_{deffered}^{DA} \) is less unfair than \( \phi_{initial}^{DA} \) w.r.t. \( (\succ_A, \succ_S) \). \( \phi_{deffered}^{DA} \) is dynamically strategy-proof since \( \phi_{deffered}^{DA} \) is static and statically strategy-proof while for static mechanisms, dynamic strategy-proofness is equivalent to static strategy-proofness. \( \phi_{deffered}^{DA} \) does not always Pareto dominate \( \phi_{initial}^{DA} \) w.r.t. \( \succ_1^A \) by Corollary 1 shown below.

2) This is implied by Lemma 2 and Corollary 1.

**Lemma 2.** There is no mechanism that always Pareto dominates \( \phi_{initial}^{DA} \) w.r.t. \( \succ_1^A \) and is dynamically strategy-proof.

**Proof.** Suppose to the contrary that there is a mechanism \( \phi \) that satisfies the above two properties. Since \( \phi \) always Pareto dominates \( \phi_{initial}^{DA} \) w.r.t. \( \succ_1^A \), there exist \( (\succ_0^A, \succ_1^A, \succ_S) \) and \( a \) such that \( \phi_a(\succ_0^A, \succ_1^A, \succ_S) \neq \phi_{initial,a}(\succ_0^A, \succ_1^A, \succ_S) \).

**Claim 1.** There exist \( (\succ_0^A, \succ_1^A, \succ_S) \) and \( a \) such that \( \phi_{initial,a}(\succ_0^A, \succ_1^A, \succ_S) >_{A} \phi_a(\succ_0^A, \succ_1^A, \succ_S) \).

**Proof.** Otherwise, for any \( (\succ_0^A, \succ_1^A, \succ_S) \) and any \( a \) with \( \phi_a(\succ_0^A, \succ_1^A, \succ_S) \neq \phi_{initial,a}(\succ_0^A, \succ_1^A, \succ_S) \), it is the case that \( \phi_a(\succ_0^A, \succ_1^A, \succ_S) \neq \phi_{initial,a}(\succ_0^A, \succ_1^A, \succ_S) \). This means that \( \phi \) always Pareto dominates \( \phi_{initial}^{DA} \) w.r.t. \( \succ_0^A \). Consider \( \succ_0^a \): \( \phi_a(\succ_0^A, \succ_1^A, \succ_S), \emptyset \). Since \( \phi_{initial,a}^{DA} \) is strategy-proof, \( \phi_{initial,a}^{DA}(\succ_0^A, \succ_0^a, \succ_1^A, \succ_S) = \emptyset \). Since \( \phi \) always Pareto dominates \( \phi_{initial,a}^{DA} \) w.r.t. \( \succ_0^A \), the CLAIM in Abdulkadiroglu et al. (2009) implies that \( \phi_a(\succ_0^A, \succ_0^a, \succ_1^A, \succ_S) = \emptyset \). This means that \( \succ_0^a \) is a profitable deviation for \( a \) w.r.t. \( \succ_0^a \) under \( \phi \) when the true preference profile is \( (\succ_0^a, \succ_0^a, \succ_1^A, \succ_S) \), a contradiction to the assumption that \( \phi \) is dynamically strategy-proof. Thus, there exist \( (\succ_0^A, \succ_1^A, \succ_S) \) and \( a \) such that \( \phi_{initial,a}^{DA}(\succ_0^A, \succ_1^A, \succ_S) >_{A} \phi_a(\succ_0^A, \succ_1^A, \succ_S) \).

For any \( a \) such that \( \phi_{initial,a}^{DA}(\succ_0^A, \succ_1^A, \succ_S) >_{A} \phi_a(\succ_0^A, \succ_1^A, \succ_S) \), who exists by Claim 1, consider the following preference: \( \succ_0^A, \phi_{initial,a}^{DA}(\succ_0^A, \succ_1^A, \succ_S), \emptyset \). Since \( \phi_a(\succ_0^A, \succ_1^A, \succ_S) = \phi_{initial,a}^{DA}(\succ_0^A, \succ_1^A, \succ_S) \), where the first equality is by the assumption that \( \phi \) always dominates \( \phi_{initial}^{DA} \), \( \succ_0^A \) is a profitable deviation for \( a \) w.r.t. \( \succ_0^A \) at \( (\succ_0^A, \succ_1^A, \succ_S) \), a contradiction.

**Definition 4.** A school preference profile \( \succ_S \) is **acyclic** if there exist no \( s_1, s_2 \in S \) and \( a_1, a_2, a_3 \in A \) such that

- \( a_1 \succ s_1, a_2 \succ s_1, a_3 \succ s_2, a_1 \) and
- there exist (possibly empty) disjoint sets of students \( A_{s_1}, A_{s_2} \subseteq A \setminus \{a_1, a_2, a_3\} \) such that \( |A_{s_1}| = q_{s_1} - 1, |A_{s_2}| = q_{s_2} - 1, a \succ s_1 a_2 \) for every \( a \in A_{s_1} \) and \( a' \succ s_2 a_1 \) for every \( a' \in A_{s_2} \).

**Lemma 3.** There is a mechanism that is fair w.r.t. \( (\succ_1^A, \succ_S) \) and never dominated by \( \phi_{initial}^{DA} \) if and only if \( \succ_S \) is acyclic.
Proof. Acyclicity is sufficient because acyclicity guarantees that $\varphi_{\text{deferred}}$ is Pareto efficient for applicants w.r.t. $\succsim_A$ (Ergin, 2002) and so is not dominated by $\varphi_{\text{initial}}$ w.r.t. $\succsim_A$. For the necessity part, suppose to the contrary that though $\succsim_s$ is cyclic, a mechanism that is fair w.r.t. $(\succsim_A, \succsim_s)$ is never dominated by $\varphi_{\text{initial}}$. By the definition of acyclicity, there exist $s_1, s_2 \in S$ and $a_1, a_2, a_3 \in A$ such that $a_1 \succsim_s a_2 \succsim_s a_3 \succsim_s a_1$ and there exist (possibly empty) disjoint sets of applicants $A_{s_1}, A_{s_2} \subseteq A \setminus \{a_1, a_2, a_3\}$ such that $|A_{s_1}| = q_{s_1} - 1, |A_{s_2}| = q_{s_2} - 1, a \succsim_s a_2$ for every $a \in A_{s_1}$ and $a' \succsim_s a_1$ for every $a' \in A_{s_2}$. Consider the following preference profile $\succsim_A$ of applicants (With 3 or more schools, it is easy to expand this example so that $a_2$ ranks some school in period 0):

\[
\begin{align*}
\succsim_{a_1} (t = 0, 1) &= s_2, s_1, \emptyset, \\
\succsim_{a_2} (t) &= \emptyset, \\
\succsim_{a_3} (t) &= s_1, s_2, \emptyset, \\
\succsim_{a_4} (t = 0, 1) &= s_1, s_2, \emptyset, \forall l \in A_{s_1}, \\
\succsim_{a_5} (t) &= s_2, s_1, \emptyset, \forall m \in A_{s_2}, \\
\succsim_{a_6} (t = 0, 1) &= s_2, s_1, \emptyset, \forall n \in A \setminus (A_{s_1} \cup A_{s_2} \cup \{a_1, a_2, a_3\}).
\end{align*}
\]

$\varphi_{\text{initial}}(\succsim_0, \succsim_A, \succsim_s)$ matches $\{a_3\} \cup A_{s_1}$ to $s_1$, $\{a_1\} \cup A_{s_2}$ to $s_2$, and leaves all the other applicants unmatched. $\varphi_{\text{deferred}}(\succsim_0, \succsim_A, \succsim_s)$ matches $\{a_1\} \cup A_{s_1}$ to $s_1$, $\{a_3\} \cup A_{s_2}$ to $s_2$, and leaves all the other students unmatched. However, since $s_2 \succsim_{a_1} s_1$ and $s_1 \succsim_{a_3} s_2$ for $t = 0, 1$, $\varphi_{\text{deferred}}(\succsim_0, \succsim_A, \succsim_s)$ is Pareto dominated by $\varphi_{\text{initial}}(\succsim_0, \succsim_A, \succsim_s)$ by the applicants w.r.t. both $\succsim_0$ and $\succsim_A$. Since any mechanism that is fair w.r.t. $(\succsim_A, \succsim_s)$ is (weakly) Pareto dominated by $\varphi_{\text{deferred}}(\succsim_0, \succsim_A, \succsim_s)$, this implies that any mechanism that is fair w.r.t. $(\succsim_A, \succsim_s)$ is Pareto dominated by $\varphi_{\text{initial}}$ w.r.t. both $\succsim_0$ and $\succsim_A$, a contradiction.

Corollary 1. There is no mechanism that is fair w.r.t. $(\succsim_A, \succsim_s)$ and Pareto dominates $\varphi_{\text{initial}}$ w.r.t. $\succsim_A$.

Extension

Proposition 1 is based on Definition 3 of dynamic strategy-proofness. Definition 3 may appear to be restrictive in that it allows applicants to jointly manipulate and commit to not only $\succsim_0$ but also $\succsim_A$ in period 0. The following alternative, weaker definition of dynamic strategy-proofness excludes such joint manipulations.

Definition 5. A mechanism $\varphi$ is weakly dynamically strategy-proof if the following holds for any $(\succsim_0, \succsim_A, \succsim_s)$ and any $a$: (0) for any $\succsim_0$, $\varphi_a(\succsim_0, \succsim_a, \succsim_s) \succeq_a \varphi_a(\succsim_0)$ and (1) for any $\succsim_0$, $\varphi_a(\succsim_0, \succsim_A, \succsim_s) \succeq_a \varphi_a(\succsim_A)$.
In words, under a weakly dynamically strategy-proof mechanism, any myopic or one-shot preference manipulation by any applicant in any period is never strictly profitable with respect to that applicant's preference in that period. Weak dynamic strategy-proofness thus requires a mechanism to be always immune to any myopic manipulation. Weak dynamic strategy-proofness is implied by dynamic strategy-proofness because the former allows for a smaller set of potential manipulations than the latter. Parts (1.A) and (1.B) of Proposition 1 remain to hold even under weak dynamic strategy-proofness (instead of dynamic strategy-proofness). It is open whether the remaining part (2) of Proposition 1 also remains correct. Nevertheless, the following partial result is true.

**Lemma 4.** There is no mechanism that always dominates $\varphi_{\text{initial}}^{DA}$ w.r.t. $\succ_A^1$, is weakly dynamically strategy-proof, and is weakly Pareto efficient w.r.t. $\succ_A^1$.

The proof is in Appendix 1.6.4. Lemma 4 and the proof of Proposition 1 imply that the only potentially necessary modification to Proposition 1 when using weak dynamic strategy-proofness is that there may exist a mechanism that always dominates $\varphi_{\text{initial}}^{DA}$ with respect to $\succ_A^1$, is weakly dynamically strategy-proof, but is not weakly Pareto efficient w.r.t. $\succ_A^1$ (and not fair with respect to $\succ_A^1$). I do not consider such a mechanism in the counterfactual analysis in Section 1.4 since even if it exists and I can construct it, the counterfactual analysis is mainly interested in efficiency gains over $\varphi_{\text{initial}}^{DA}$ and so prefers $\varphi_{\text{dynamic}}^{DA}$ (which is weakly Pareto efficient w.r.t. $\succ_A^1$) over it (which has to be not weakly Pareto efficient w.r.t. $\succ_A^1$).

**Proof of Lemma 4**

Suppose to the contrary that there is a mechanism $\varphi$ that satisfies the above three properties. Consider a problem with four applicants $a_1, a_2, a_3, a_4$ and four schools $s_1, s_2, s_3, s_4$ each of which has the capacity of one. Their preferences are as follows:

\[
\succ_{s_1}: \text{Anything with } a_3 \succ_{s_1} a_4 \succ_{s_1} a_1 \\
\succ_{s_2}: \text{Anything} \\
\succ_{s_3}: \text{Anything with } a_1 \succ_{s_3} a_3 \\
\succ_{s_4}: \text{Anything with } a_4 \succ_{s_4} a_2 \\
\succ^0_{a_1}: s_1, s_2, s_3, ... \\
\succ^0_{a_2}: s_4, s_2, ... \\
\succ^0_{a_3}: s_3, s_1, ... \\
\succ^0_{a_4}: s_1, s_4, ... \\
\succ^1_{a_1}: s_1, s_2, s_3, ... \\
\succ^1_{a_2}: s_3, s_4, s_2, ... \\
\succ^1_{a_3}: s_3, s_2, s_1, ...
\]
Note that $\varphi_{\text{initial}}^D(\succ^0_A, \succ^1_A, \succ_S) = (a_1, a_2, a_3, a_4)$ and this matching is Pareto efficient w.r.t. $\succ^1_A$. By the assumption that $\varphi$ always dominates $\varphi_{\text{initial}}^D$ w.r.t. $\succ^1_A$, $\varphi(\succ^0_A, \succ^1_A, \succ_S) = \varphi_{\text{initial}}^D(\succ^0_A, \succ^1_A, \succ_S) = (a_1, a_2, a_3, a_4)$. Thus $\varphi_{a_1}(\succ^0_A, \succ^1_A, \succ_S) = s_2$.

Now let me consider the following preference of $a_1$:

$\succ^0_{a_1} : s_3, \ldots$

Note that $\varphi_{\text{initial}}^D(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = (a_1, a_2, a_3, a_4)$. This matching is not weakly Pareto efficient w.r.t. $\succ^1_A$ since it is strongly Pareto dominated by $s_2$ and $s_3$. Thus $\varphi_{a_1}(\succ^0_A, \succ^1_A, \succ_S)$ has to be a matching that is weakly Pareto efficient w.r.t. $\succ^1_A$ and Pareto dominates $\varphi_{\text{initial}}^D(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$. I use the following fact.

**Lemma 5.** $\varphi_{a_1}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_1$.

**Proof.** By construction of $\succ^1_{a_1}$, $\varphi_{a_1}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_1$ or $s_2$ or $s_3$. It is thus enough to show $\varphi_{a_1}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) \neq s_2$ or $s_3$. If $\varphi_{a_1}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_3$ by the following reason: Suppose to the contrary that $\varphi_{a_1}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_3$. Inspections show that there is no matching $\mu$ such that $\mu_{a_1} = s_3$ and $\mu$ Pareto dominates $\varphi_{\text{initial}}^D(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$ with respect to $\succ^1_A$.

Suppose to the contrary that there is some matching $\mu$ such that $\mu_{a_1} = s_3$ and $\mu$ Pareto dominates $\varphi_{\text{initial}}^D(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$ with respect to $\succ^1_A$.

**Case 1:** If $\mu_{a_4} \succ^1_{a_4} \varphi_{\text{initial}, a_4}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_4$, then $\mu_{a_4} = s_3$, which is the only school $a_4$ strictly prefers to $s_4 = \varphi_{\text{initial}, a_4}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$. But this contradicts the assumption of $\mu_{a_1} = s_3$.

**Case 2:** If $\mu_{a_2} \succ^1_{a_2} \varphi_{\text{initial}, a_2}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_1$, then $\mu_{a_3} = s_2$, since $s_2$ and $s_3$ are the only schools $a_3$ strictly prefers to $s_1 = \varphi_{\text{initial}, a_3}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$ w.r.t. $\succ^1_A$ and $\mu_{a_1} = s_3$. Since $s_2 = \varphi_{\text{initial}, a_2}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$, it has to be the case $\mu_{a_2} = s_4$; $s_2$ and $s_4$ are the only schools $a_2$ strictly prefers to $s_2 = \varphi_{\text{initial}, a_2}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$ and $\mu_{a_1} = s_3$ by assumption. Then, since $s_4 = \varphi_{\text{initial}, a_4}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$, it has to be the case that $\mu_{a_4} \succ^1_{a_4} \varphi_{\text{initial}, a_4}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$. Thus this case reduces to Case 1, a contradiction.

**Case 3:** If $\mu_{a_2} \succ^1_{a_2} \varphi_{\text{initial}, a_2}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_2$, then $\mu_{a_2} = s_4$, since $s_3$ and $s_4$ are the only schools $a_2$ strictly prefers to $s_2 = \varphi_{\text{initial}, a_2}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$ w.r.t. $\succ^1_A$ and $\mu_{a_1} = s_3$. Then, since $\mu_{a_4} \succ^1_{a_4} \varphi_{\text{initial}, a_4}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S) = s_4$, then $\mu_{a_4} = s_3$, since $s_2$ and $s_3$ are the only schools $a_4$ strictly prefers to $s_3 = \varphi_{\text{initial}, a_4}(\succ^0_{a_1}, \succ^0_{-a_1}, \succ^1_A, \succ_S)$. Thus this case reduces to Case 2, a contradiction.
dominates $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$, which contradicts the assumption that $\varphi$ is weakly Pareto efficient and always dominates $\varphi_{\text{DA}}^{\text{initial}}$ w.r.t. $\succ_A$.

Next, $\varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) \neq s_2$ by the following reason: Suppose to the contrary that $\varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) = s_2$. Consider the following preference of $a_1$:

$$\succ_{a_1}^1: s_1, s_3, ...$$

Note that $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) = (a_1, a_2, a_3, a_4)$. This matching is not weakly Pareto efficient w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$ since it is strongly Pareto dominated by $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$. By the assumption that $\varphi$ is weakly Pareto efficient w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$ and always dominates $\varphi_{\text{DA}}^{\text{initial}}$ w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$, $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ has to be a matching that is weakly Pareto efficient w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$ (and so $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) \neq \varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$) and $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$ implies $\varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) = s_1$.

By the assumption that $\varphi$ is weakly Pareto efficient w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$ and always dominates $\varphi_{\text{DA}}^{\text{initial}}$ w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$, $\varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ has to be a matching that is weakly Pareto efficient w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$ (and so $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) \neq \varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$). This implies $\varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) = s_1$ by the following reason: Injections show that there is no matching $\mu$ such that $\mu_{a_1} = s_3$ and $\mu$ Pareto dominates $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$, thus $\varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) = s_3$ (the only match for $a_1$ than $s_1$ by the construction of $\succ_{a_1}^1$ and $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$) implies that $\varphi(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ does not Pareto dominates $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$, which contradicts the assumption that $\varphi$ Pareto dominates $\varphi_{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$. This implies $\varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S) = s_1$.

Thus this case reduces to Case 1, a contradiction.

---

Suppose to the contrary that there is some matching $\mu$ such that $\mu_{a_1} = s_3$ and $\mu$ Pareto dominates $\varphi_{\text{DA}}^{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $(\succ_{a_1}^1, \succ_{a_1}^1)$.

**Case 1:** If $\mu_{a_1} = s_3$, then $\mu_{a_4} = s_4$, since $s_2$ and $s_3$ are the only schools $a_3$ strictly prefers to $s_4 = \varphi_{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $\succ_{a_4}$. But this contradicts the assumption of $\mu_{a_1} = s_3$.

**Case 2:** If $\mu_{a_3} = s_3$, then $\mu_{a_4} = s_2$, since $s_2$ and $s_3$ are the only schools $a_3$ strictly prefers to $s_1 = \varphi_{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $\succ_{a_3}$ and $\mu_{a_1} = s_3$ by assumption. Since $s_2 = \varphi_{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$, it has to be the case that $\mu_{a_3} = s_2$ and $\mu_{a_4} = s_3$ by assumption. Then, since $s_1 = \varphi_{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$, it has to be the case that $\mu_{a_4} = s_4$. Thus this case reduces to Case 1, a contradiction.

**Case 3:** If $\mu_{a_2} = s_2$, then $\mu_{a_4} = s_4$, since $s_3$ and $s_4$ are the only schools $a_2$ strictly prefers to $s_2 = \varphi_{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$ w.r.t. $\succ_{a_2}$ and $\mu_{a_1} = s_3$ by assumption. Since $s_4 = \varphi_{\text{initial}}(\succ_{a_1}^0, \succ_{a_1}^0, \succ_{a_1}^1, \succ_S)$, it has to be the case that $\mu_{a_4} = s_4$. Thus this case reduces to Case 1, a contradiction.
\[ \varphi_{a_1}(\succ_{a_1}^0, \succ_{-a_1}^0, \succ_{a_1}^1, \succ_{-a_1}^1, \succ_s) = s_1 \succ_{a_1}^1 s_2 = \varphi_{a_1}(\succ_{a_1}^0, \succ_{-a_1}^0, \succ_{a_1}^1, \succ_s), \]

a contradiction to the assumption that \( \varphi \) is dynamically strategy-proof. Therefore it has
to be the case that \( \varphi_{a_1}(\succ_{a_1}^0, \succ_{-a_1}^0, \succ_{a_1}^1, \succ_s) \neq s_2. \)

Recall that by construction of \( \succ_{a_1}^1, \varphi_{a_1}(\succ_{a_1}^0, \succ_{-a_1}^0, \succ_{a_1}^1, \succ_s) = s_1 \) or \( s_2 \) or \( s_3. \) Thus the
above discussions imply \( \varphi_{a_1}(\succ_{a_1}^0, \succ_{-a_1}^0, \succ_{a_1}^1, \succ_s) = s_1. \)

Lemma 4 and the paragraph right after the preference descriptions show

\[ \varphi_{a_1}(\succ_{a_1}^0, \succ_{-a_1}^0, \succ_{a_1}^1, \succ_s) = s_1 \succ_{a_1}^t s_2 = \varphi_{a_1}(\succ_{a_1}^0, \succ_{a_1}^1, \succ_s) \text{ for } t = 0, 1, \]

which contradicts the assumption that \( \varphi \) is dynamically strategy-proof. Note that this
proof shows a stronger result than necessary in that \( \succ_{a_1}^0 \) is a profitable manipulation w.r.t.
both \( \succ_{a_1}^0 \) and \( \succ_{a_1}^1. \)
Figure 1.6.8: Timeline of the First-round and Reapplication Process: Details

Notes: This figure draws the histogram of the difference between the date at which the initial application is filed and the date at which the reapplication is filed (both are conditional on reapplicants). See Section 1.2.1 for discussions about this figure.

Figure 1.6.9: Evolving School Choices: Details

(a) Changing Market Shares of Schools

(b) Reapplications After Getting Top Choices

Notes: Panel 1.6.9a plots the first choice market share of each school in initial applications (the x axis) and reapplicants (the y axis), where the first choice market share of a school is defined as the fraction of applicants who rank it first among all applicants who make a first choice. I compute both the old and new shares conditional on reapplicants. Panel 1.6.9b correlates the number of reapplicants with the preference rank of the initially assigned school with respect to the initial preference. See Section 1.2.1 for discussions about this figure.
(a) Self-reported Reasons for Reapplication: Reapplicants with Choice Reversals

(b) Breakdown of "New Information"

Notes: Panel 1.6.10a classifies self-reported reasons for reapplication conditional on applicants who reapply and exhibit choice reversals. Panel 1.6.10b focuses on the "new information" category in Panel 1.6.10a and breaks it down into sub-categories. "Distance" is different from "Moving After Application" in that the former does not refer to any address change. See also Section 1.2.2 for discussions about this figure.

Table 1.6.10: Learning about Schools: Details

<table>
<thead>
<tr>
<th>Specification 1 (Location/borough dummies)</th>
<th>Grade 7 reading grade category</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Middle</td>
</tr>
<tr>
<td></td>
<td>+45%</td>
<td>+32%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specification 2 (1+Academic performance dummy)</th>
<th>Grade 7 reading grade category</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification 3 (2+Program type dummies)</td>
<td>Grade 7 reading grade category</td>
<td>Race</td>
</tr>
<tr>
<td>Specification 4 (3+Capacity dummies)</td>
<td>Grade 7 reading grade category</td>
<td>Race</td>
</tr>
</tbody>
</table>

Notes: This table shows the last column in Table 1.4 conditional on each demographic group. The details of included characteristics are in Appendix 1.6.2. See also Section 1.2.2 for discussions about this table.
### Table 1.6.11: Growing Response of Choices to Distance and Academic Achievement: Details

<table>
<thead>
<tr>
<th>Grade 7 reading grade category</th>
<th>High</th>
<th>Middle</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial applications</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st choice school</td>
<td>5.0</td>
<td>5.0</td>
<td>4.7</td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>5.2</td>
<td>5.2</td>
<td>4.8</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>5.3</td>
<td>5.3</td>
<td>5.0</td>
</tr>
<tr>
<td><strong>Distance Schools ranked above</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>3.4%</td>
<td>11.7%</td>
<td>17.1%</td>
</tr>
<tr>
<td><strong>Distance Schools ranked above</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>6.6%</td>
<td>14.6%</td>
<td>19.3%</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>4.4</td>
<td>4.9</td>
<td>5.9</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>White</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st choice school</td>
<td>4.3</td>
<td>4.6</td>
<td>4.6</td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>4.3</td>
<td>4.8</td>
<td>4.9</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>4.4</td>
<td>4.9</td>
<td>5.9</td>
</tr>
<tr>
<td><strong>Asian</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st choice school</td>
<td>4.2%</td>
<td>4.3%</td>
<td>4.3%</td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>3.4%</td>
<td>3.9%</td>
<td>4.3%</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>6.3%</td>
<td>6.9%</td>
<td>4.3%</td>
</tr>
<tr>
<td><strong>Hispanic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st choice school</td>
<td>4.8</td>
<td>5.5</td>
<td>4.4</td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>5.0</td>
<td>5.7</td>
<td>4.5</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>5.1</td>
<td>5.9</td>
<td>4.5</td>
</tr>
<tr>
<td><strong>Black</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st choice school</td>
<td>16.8%</td>
<td>11.0%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Schools ranked above initial assignment</td>
<td>14.9%</td>
<td>10.4%</td>
<td>10.7%</td>
</tr>
<tr>
<td>All ranked schools</td>
<td>19.0%</td>
<td>15.2%</td>
<td>10.7%</td>
</tr>
</tbody>
</table>

**Notes:** This table shows the average distance to and academic achievement level of ranked schools in initial applications and reapplications, conditional on each demographic group. *Schools with low academic performance* are "schools in need of improvement" in the official school brochure issued by NYC. Under the No Child Left Behind Act, New York State establishes annual performance goals in Mathematics and English Language Arts for all NYC public schools. Schools that do not meet these goals for two consecutive years are identified as schools in need of improvement. See Section 1.2.2 for discussions about this table.
Figure 1.6.11: Empirical Probability of Reapplication Acceptance

Notes: Panel 1.6.11 correlates the probability of reapplication acceptance (being accepted by some school) conditional on reapplying with the preference rank of the initially assigned school with respect to the initial preference. See Section 1.3.2 for discussions about this figure.
Figure 1.6.12: Robustness & Heterogeneity of Demand Changes & Switching Costs

(a) Heterogeneity across Baseline Test Scores

(b) Heterogeneity across Races

(c) Homogeneity across Old 1st Choice Schools

(d) Homogeneity across Initial Assignments

Notes: These figures correlate the conditional probability of reapplying to the preference rank of the initially assigned school with respect to the initial preference conditional on a group defined by a baseline characteristic. Combined with the logic in Figure 1-4 and Section 1.3.2, they suggest the (absence of) heterogeneity of demand changes and switching costs across groups. The construction of the groups are in Appendix 1.6.1. See Section 1.3.2 for discussions about these figures.
**Table 1.6.12: Covariate Balance Between Applicants with and without First Choice Offers**

<table>
<thead>
<tr>
<th>Covariate balance coefficients</th>
<th>Causal OLS</th>
<th>Descriptive OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female</strong></td>
<td>0.015</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.003)</td>
</tr>
<tr>
<td><strong>Living area</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manhattan</td>
<td>-0.002</td>
<td>-0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Brooklyn</td>
<td>-0.005</td>
<td>0.005**</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Queens</td>
<td>0.003</td>
<td>-0.047***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Bronx</td>
<td>0.004</td>
<td>0.007**</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Staten Island</td>
<td>0.000</td>
<td>0.041***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td><strong>Home language</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>-0.006</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Chinese</td>
<td>0.002</td>
<td>0.004***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>N (students)</td>
<td>11265</td>
<td>90978</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>-0.008*</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Black</td>
<td>0.007</td>
<td>-0.032***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>White</td>
<td>-0.008</td>
<td>0.052***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.002)</td>
</tr>
<tr>
<td><strong>1st choice school &amp;</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>marginal priority controls</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (students)</td>
<td>10,369</td>
<td>82,301</td>
</tr>
<tr>
<td>N (schools)</td>
<td>224</td>
<td>677</td>
</tr>
</tbody>
</table>

***: Significant at 1%, **: Significant at 5%, *: Significant at 10%

**Notes:** This table reports estimates of the causal effect of being assigned to the first choice school on baseline covariates based on the linear probability model described in the main text. "Descriptive OLS" is the regression of the reaplication dummy on the first choice assignment dummy with no control or sample restriction. Standard errors are in parentheses. See Section 1.3.2 and Appendix 1.6.3 for discussions about this table.
Table 1.6.13: Causal Effect of Being Assigned to the First Choice School on Reapplying

<table>
<thead>
<tr>
<th>Coefficient on offer from 1st choice school</th>
<th>Causal OLS</th>
<th>Causal probit marginal effect</th>
<th>Causal logit marginal effect</th>
<th>Descriptive OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.056***</td>
<td>-0.060***</td>
<td>-0.057***</td>
<td>-0.070***</td>
</tr>
<tr>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>1st choice school &amp; marginal priority controls</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (students)</td>
<td>11265</td>
<td>90978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (schools)</td>
<td>224</td>
<td>677</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***: Significant at 1%

Notes: This table reports estimates of the causal effect of being assigned to the first choice school on the probability of reapplying based on the linear probability, probit, and logit models described in the main text. “Descriptive OLS” is the regression of the reapplication dummy on the first choice assignment dummy with no control or sample restriction. Standard errors are in parentheses. See Section 1.3.2 and Appendix 1.6.3 for discussions about this table.

Figure 1.6.13: Evidence of Demand Changes and Switching Costs: Structural Version

Notes: This figure illustrates how to separately identify switching costs and demand changes from the solid black line observed in the data. The solid black line correlates the conditional probability of reapplying to the preference rank of the initially assigned school with respect to the initial preference conditional on applicants who are randomly assigned to their first and lower choice schools. As detailed in the main text, \( \hat{\alpha} + \hat{\beta} \) and \( \hat{\alpha} \) are estimates of the conditional probabilities of reapplying conditional on being assigned to the first choice school and a non-first-choice school, respectively, in \( \cup_F First_s \), where applicants are randomly assigned between the first choice school and a non-first-choice school. Combined with the logic in Figure 1-4 and Section 1.3.2, this suggests the presence of both switching costs and demand changes. See Section 1.3.2 and Appendix 1.6.3 for discussions about this figure.
Table 1.6.14: Tests of Switching Costs and Demand Changes

<table>
<thead>
<tr>
<th>Behavioral hypothesis</th>
<th>Statistical hypothesis</th>
<th>Wald test result</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$H_0$: No switching costs &amp; no demand changes</strong></td>
<td>$H_0: \alpha + \beta = 0$</td>
<td>Rejected</td>
<td>&lt;0.01%</td>
</tr>
<tr>
<td></td>
<td>$H_0: \alpha = 1$</td>
<td>Rejected</td>
<td>&lt;0.01%</td>
</tr>
<tr>
<td><strong>$H_0$: No switching costs</strong></td>
<td>$H_0: \alpha + \beta \geq 1 - \alpha$</td>
<td>Rejected</td>
<td>&lt;0.01%</td>
</tr>
</tbody>
</table>

*Notes:* This table reports results of Wald tests of the null hypothesis that there are no switching costs or the hypothesis that there are neither switching costs nor demand changes. See Appendix 1.6.3 for discussions about this table.
### Table 1.6.15: Estimates from the Rational Expectation Model: Preferences and Frictions

<table>
<thead>
<tr>
<th></th>
<th>Black or Hispanic</th>
<th>Low or Middle</th>
<th>White or Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Positive distance (10 miles, log-normal for $\beta_k(1 + f_{ak})$)</td>
<td>0.90*** (0.60)</td>
<td>4.52*** (1.77)</td>
<td>1.08*** (0.59)</td>
</tr>
<tr>
<td>School type unscreened dummy</td>
<td>-0.56** (1.27)</td>
<td>-2.52** (2.11)</td>
<td>-0.75 (1.02)</td>
</tr>
<tr>
<td>School type education option dummy</td>
<td>0.67*** (0.44)</td>
<td>0.21 (1.50)</td>
<td>0.13 (0.61)**</td>
</tr>
<tr>
<td>School type screened dummy</td>
<td>1.15*** (0.75)</td>
<td>1.43 (1.38)</td>
<td>0.06 (0.73)**</td>
</tr>
<tr>
<td>School type limited dummy</td>
<td>0.11 (0.83)**</td>
<td>0.48 (2.82**)</td>
<td>-0.55 (0.78)</td>
</tr>
<tr>
<td>School type audit dummy</td>
<td>-1.21*** (2.31)</td>
<td>-1.42 (0.08)</td>
<td>-1.15 (1.85)**</td>
</tr>
<tr>
<td>School type small new dummy</td>
<td>-3.57*** (1.87)**</td>
<td>0.13 (0.25)</td>
<td>-3.26*** (2.35)**</td>
</tr>
<tr>
<td>School type new old dummy</td>
<td>-2.57*** (0.90)</td>
<td>-1.21 (0.63)</td>
<td>-0.42 (1.22)</td>
</tr>
<tr>
<td>School type small old dummy</td>
<td>-1.68*** (0.27)</td>
<td>0.11 (0.32)</td>
<td>-0.86 (0.44)</td>
</tr>
<tr>
<td>School type medium old dummy</td>
<td>-1.35*** (0.37)**</td>
<td>0.68 (0.53)</td>
<td>-1.96*** (0.58)</td>
</tr>
<tr>
<td>Medium new school dummy</td>
<td>-0.72*** (0.14)</td>
<td>1.91 (0.07)</td>
<td>-0.28 (0.33)</td>
</tr>
<tr>
<td>Medium old school dummy</td>
<td>-0.81*** (0.00)</td>
<td>-0.68 (0.00)</td>
<td>-0.50*** (0.00)</td>
</tr>
</tbody>
</table>

Notes: Based on the rational expectation model, this table shows the estimates of the mean and standard deviation of $(\beta_k(1 + f_{ak}))_k$ (the coefficient on $X_{ask}$ in $t = 0$) and $(\beta_k f_{ak})_k$ (the coefficient in $t = 1$). Standard errors are in parentheses. See Sections 1.3.1 and 1.3.3 for the details of the model and the estimation method, respectively. Appendix 1.6.1 explains the construction of variables. See Section 1.3.4 for discussions about this table.
Table 1.6.16: Estimates from the Rational Expectation Model: Reapplication Costs and Initial Assignment Effects

<table>
<thead>
<tr>
<th>Grade 7 reading grade category: high</th>
<th>Black or Hispanic</th>
<th>White or Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reapplication cost $c_a$</td>
<td>Initial assignment effect $\gamma_a$</td>
<td>$c_a$</td>
</tr>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>1.11***</td>
<td>0.97</td>
<td>5.82*</td>
</tr>
<tr>
<td>(0.26)</td>
<td>(0.32)</td>
<td>(3.05)</td>
</tr>
</tbody>
</table>

Notes: Based on the rational expectation model, this table shows the estimates of the mean and standard deviation of reapplication costs $c_a$ and initial assignment effects $\gamma_a$. Standard errors are in parentheses. See Sections 1.3.1 and 1.3.3 for the details of the model and the estimation method, respectively. See Section 1.3.4 for discussions about this table.

Figure 1.6.14: Summary of Estimates: Naive Free Expectation Model

(a) Demand Changes due to Learning

(b) Reapplication Costs & Initial Match Effects

Notes: Based on the naive free expectation model in Section 1.3.1, Panel 1-5a plots the distributions of estimated overall new utilities ($\hat{U}_{1a}$) and latent demand changes due to frictions about observable school characteristics ($\sum_{k=1}^{K} \beta_{ak} f_{ak} X_{ask}$) for all (applicant $a$, school $s$) pairs. Panel 1-5b plots the distributions of estimated overall new utilities ($\hat{U}_{1a}$), estimated reapplication costs ($\hat{c}_a$), and estimated initial assignment effects ($\hat{\gamma}_a$). On $\hat{c}_a$, it plots values implied by the estimated value of $c_a/p_a$ and the rational expectation assumption that $p_a$ is equal to the empirical probability of reapplication acceptance. Both panels are based on 50 simulations of the estimated model for each (applicant $a$, school $s$) pair. See Sections 1.3.1 and 1.3.3 for the details of the model and the estimation method, respectively. See Section 1.3.4 for discussions about this figure.
Table 1.6.17: Estimates from the Naive Free Expectation Model: Preferences and Frictions

<table>
<thead>
<tr>
<th></th>
<th>Black or Hispanic</th>
<th>Low or Middle</th>
<th>High</th>
<th>White or Asian</th>
<th>Low or Middle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 7 reading grade category: high</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_k(1+f_{ak}) ) in ( t=0 )</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Negative distance (-10 miles)</td>
<td>0.90***</td>
<td>0.60***</td>
<td>3.34***</td>
<td>0.67</td>
<td>1.06***</td>
</tr>
<tr>
<td>log-normal for ( \beta_k(1+f_{ak}) )</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.30)</td>
<td>(0.48)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>School type unscreened dummy</td>
<td>-0.56*</td>
<td>1.27***</td>
<td>-2.69***</td>
<td>1.80*</td>
<td>-0.75</td>
</tr>
<tr>
<td>School type education option dummy</td>
<td>0.67***</td>
<td>0.44***</td>
<td>-1.36***</td>
<td>0.82</td>
<td>0.13</td>
</tr>
<tr>
<td>School type screened dummy</td>
<td>1.15***</td>
<td>0.77***</td>
<td>-0.12</td>
<td>0.49</td>
<td>0.06</td>
</tr>
<tr>
<td>School type limited dummy</td>
<td>0.11</td>
<td>0.83***</td>
<td>-0.67</td>
<td>1.65**</td>
<td>-0.55</td>
</tr>
<tr>
<td>High performance dummy</td>
<td>1.21***</td>
<td>2.31***</td>
<td>-1.20**</td>
<td>0.13</td>
<td>-1.15</td>
</tr>
<tr>
<td>School type audit dummy</td>
<td>-0.47***</td>
<td>1.28***</td>
<td>1.45***</td>
<td>0.73</td>
<td>-2.99***</td>
</tr>
<tr>
<td>(log-normal for ( \beta_k(1+f_{ak}) )</td>
<td>(0.07)</td>
<td>(0.11)</td>
<td>(0.50)</td>
<td>(0.44)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>Tiny new school dummy</td>
<td>-3.79***</td>
<td>1.87***</td>
<td>-0.52</td>
<td>0.47</td>
<td>-3.26***</td>
</tr>
<tr>
<td>Tiny old school dummy</td>
<td>-2.57***</td>
<td>0.00</td>
<td>3.21***</td>
<td>0.18</td>
<td>-2.05***</td>
</tr>
<tr>
<td>Small new school dummy</td>
<td>-1.68***</td>
<td>0.37</td>
<td>1.96***</td>
<td>0.15</td>
<td>-0.86*</td>
</tr>
<tr>
<td>Small old school dummy</td>
<td>-1.25***</td>
<td>0.37***</td>
<td>-1.99***</td>
<td>0.45</td>
<td>-1.06***</td>
</tr>
<tr>
<td>Medium new school dummy</td>
<td>-0.75***</td>
<td>0.01</td>
<td>2.31***</td>
<td>0.14</td>
<td>-0.28</td>
</tr>
<tr>
<td>Medium old school dummy</td>
<td>-0.81***</td>
<td>0.00</td>
<td>-0.79***</td>
<td>0.10</td>
<td>-0.30***</td>
</tr>
</tbody>
</table>

Notes: This table shows the estimates of the mean and standard deviation of \( (\beta_{ak}(1+f_{ak}))_k \) (the coefficient on \( X_{ask} \) in \( t=0 \)) and \( (\beta_{ak}f_{ak})_k \) (the coefficient in \( t=1 \)). Standard errors are in parentheses. See Sections 1.3.1 and 1.3.3 for the details of the model and the estimation method, respectively. Appendix 1.6.1 explains the construction of variables. See Section 1.3.4 for discussions about this table.
Table 1.6.18: Estimates from the Naive Free Expectation Model: Reapplication Costs and Initial Assignment Effects

<table>
<thead>
<tr>
<th>Grade 7 reading grade category</th>
<th>Black or Hispanic</th>
<th>White or asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Low or middle</td>
<td>High</td>
</tr>
<tr>
<td>Scaled reapplication cost</td>
<td>Initial assignment effect</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>SD</td>
<td>mean</td>
</tr>
<tr>
<td>(1.07)</td>
<td>(2.31)</td>
<td>(0.74)</td>
</tr>
</tbody>
</table>

N (students) : 5,021, 52,114, 5,271, 13,844

Notes: This table shows the estimates of the mean and standard deviation of scaled reapplication costs \( c_a/p_a \) and initial assignment effects \( \gamma_a \). Standard errors are in parentheses. See Sections 1.3.1 and 1.3.3 for the details of the model and the estimation method, respectively. See Section 1.3.4 for discussions about this table.

Figure 1.6.15: Centralized vs Discretionary Reapplication Processes (I): Details

(a) Dynamic DA Mechanism

(b) Deferred DA Mechanism

Notes: In Panel (a), the dotted line plots the distribution of the improvement of the preference rank of the finally assigned school under the dynamic deferred acceptance mechanism compared with the initial match. The preference rank is defined with respect to new preference \( \succ_{\frac{1}{2}} \) defined in the main text. This distribution is conditional on applicants who get different assignments under the two mechanisms. The shaded area below the dotted line plots the same distribution as the dotted line conditional on applicants who reapply. Panel (b) does the same for the deferred deferred acceptance mechanism. The shaded area around each dotted line indicates the 95% simulation confident interval over 200 simulations of lottery numbers used by the mechanisms to break ties in priorities.
Chapter 2

Natural Experiments and Strategy-proofness

2.1 Introduction

The spread of choice and quasi-market in public education is giving more families the option to attend a school other than their neighborhood default. As choice has proliferated, to accommodate heterogeneous preferences and respect various priorities based on family backgrounds, school assignment mechanisms have grown increasingly centralized and algorithmic. Centralized mechanisms solve the problem of matching the limited supply of school seats to the demand for them using algorithms inspired by and inspiring market design theory. As of this writing, centralized mechanisms are employed in Boston, Charlotte, Denver, New Orleans, Newark, New York City, San Francisco, Washington DC, and numerous Asian and European countries. Well-designed centralized assignment provides a transparent way to achieve a fair and efficient school seat allocation, while narrowing the scope for strategic behavior (Abdulkadiroğlu and Sönmez, 2003).

At the same time, centralized assignment generates valuable data for empirical research on education. In particular, when a school is oversubscribed, mechanisms often use random lotteries to ration limited school seats. This generates quasi-experimental variation in school assignment that opens the door to a variety of credible impact evaluations. Researchers have used such variation to study schools in the Bay Area (Bergman, 2014), Boston (Angrist et al., 2015b), Charlotte-Mecklenburg (Hastings et al., 2009; Deming, 2011; Deming et al., 2014), and New York (Bloom et al., 2010; Abdulkadiroğlu et al., 2013) among others.¹

¹Other studies use regression-discontinuity-style local tie-breaking to evaluate college majors in Norway (Kirkeboen et al., 2015) as well as popular selective schools in Chile (Hastings et al., 2013), Ghana (Ajayi, 2013), Kenya (Lucas and Mbiti, 2014a), Romania (Pop-Eleches and Urquiola, 2013), Trinidad and Tobago (Jackson, 2010, 2012), and the US (Abdulkadiroğlu et al., 2014a; Dobbie and Fryer, 2014). Narita (2015) uses lottery-generated randomization for identifying a structural econometric model of evolving demand for schools.
Despite the growing empirical use of centralized assignment, however, the above empirical work has only a limited foundation to support their empirical research design. Except empirically showing covariate balance, which is necessary but never sufficient for successful randomization, researchers rarely explain how their research design successfully extracts random components of assignment embedded in complex mechanisms that generate their data. As different mechanisms induce different treatment assignment processes (stratified randomized controlled trials), whether a research design works likely depends on which specific mechanism generates the data at hand. Although the theoretical market design literature has extensively analyzed mechanisms on the basis of welfare and strategic properties, it has so far little guidance to offer empirical researchers.\footnote{An exception is Abdulkadiroğlu et al. (2015a), which first raised this issue.}

This paper studies whether a widely used empirical research design successfully extracts conditionally random assignment of students to schools. This research design, which I refer to as the first choice research design, may be applied to any centralized mechanism that combines (1) applicants' rank-ordered preferences over schools, (2) their priority statuses (e.g., walk zone) at schools, and (3) lottery numbers for breaking ties in priority statuses, to generate an assignment of students to schools.

The first choice research design focuses on applicants who rank a given treatment school first, share the same priority status at the school, and are such that some of them are assigned the treatment school while others are not. Within this first choice sample, some applicants are assigned the treatment school while others are not, though all students rank it first and share the same priority; thus it appears that treatment assignment is determined solely by lottery numbers. Based on this idea, the first choice research design assumes that applicants are randomly assigned to the treatment school conditional on being in the first choice sample, and compares outcomes (e.g., test scores or crime rates) of those assigned to the treatment school against those not assigned within the first choice sample. The outcome difference between the two groups is interpreted as a causal effect of the treatment school. All of the above empirical examples use this research design or its extensions.

Despite its intuitive construction, I show that the first choice research design successfully extracts a random assignment only under a condition. In particular, for a class of mechanisms containing most of those used in practice, I prove the following result: The first choice research design extracts a random assignment (i.e., applicants in the first choice sample share the same assignment probability at the treatment school) for a data-generating mechanism if and almost only if the mechanism is ex post strategy-proof for schools. This result has important implications for applied research in that it justifies the first-choice research design under mechanisms that are known to be strategy-proof for schools, such as the Boston (immediate acceptance) mechanism (Ergin and Sönmez, 2006). At the same time, it formalizes a care against the research design for other widely-used
mechanisms that are not strategy-proof for schools; the deferred acceptance mechanism, another one used in Charlotte, and the top trading cycles mechanism are such examples (Roth and Sotomayor, 1990).

Some non-strategy-proof mechanisms such as deferred acceptance are known to be often approximately strategy-proof for schools in large markets with many students and schools (Roth and Peranson, 1999; Immorlica and Mahdian, 2005; Kojima and Pathak, 2009). This means that my sufficiency result also provides an asymptotic justification for the first choice research design even under mechanisms that are not strategy-proof in general. This may explain why the first choice design appears to work in empirical applications even under non-strategy-proof mechanisms. Viewed differently, the existing empirical support for the first choice research design in the form of covariate balance regressions may inform the empirical relevance of theoretical results on strategy-proofness in large markets.³

In these senses, my analysis highlights hidden two-way connections between two fields, the mostly theoretical literature on matching market design and the empirical and econometric literature on quasi-experimental education research design. These connections provide us with a formal basis for understanding when and why the first choice empirical research design does or does not work as intended.

Hidden theoretical support

in the form of strategy-proofness theorems

Matching market design

Quasi-experimental education research design

Hidden empirical support

in the form of covariate balance regressions

It is important to note that unlike with usual market design studies of strategy-proofness, I treat strategy-proofness not as a desideratum or incentive compatibility constraint but as an algorithmic property. This is because I am interested not in strategy-proofness itself but in its implications for empirical research. I am thus not concerned with the usual behavioral interpretation of strategy-proofness and do not need to assume that schools have intrinsic preferences, that school preferences are consistent with priori-

³ Abdulkadiroğlu et al. (2015a) develop a complementary framework based on an asymptotic approximation assuming the growing number of students and school seats. They use it to propose an improvement over the first choice research design and apply it to evaluate charter schools in Denver. They also confirm that the first choice research design always extracts a random assignment under many mechanisms in the limit of their particular large market sequence. In contrast, the framework in the current paper allows for general finite markets and is used to provide a condition (strategy-proofness for schools) for the first choice research design to extract a random assignment in finite sample. I also provide additional large market justifications for the research design in other large market sequences than Abdulkadiroğlu et al. (2015a)'s.
empirical implications of my theoretical result are free from any assumption on whether schools have preferences or are strategic in reality.

In the last section 2.4, I confirm that my main point is robust to a variety of modifications to the definitions of the research design and extracting a random assignment. I also analyze a widely used extension of the first choice research design, which I call the qualification instrumental variable research design. For this extended design, I obtain a more negative result that there is no mechanism within the class under which the qualification instrumental variable research design extracts a random assignment. This suggests that the qualification instrumental variable design is strictly more fragile than the first choice design.

2.2 Framework

I use a model of school-student assignment with coarse school priorities coupled with lotteries. There are a finite set $I$ of students and a finite set $S$ of schools. Each student $i \in I$ has a strict preference $\succ_i$ over $S \cup \{\emptyset\}$, where $\emptyset$ denotes the outside option of the student. $s$ is said to be acceptable for $i$ if $s \succ_i \emptyset$. A preference profile of all students is denoted by $\succ \equiv (\succ_i)_{i \in I}$. Each school $s$ has a capacity $c_s$. Schools also grant students various coarse priorities. $\rho_{is} \in \{1, \ldots, K\}$ denotes student $i$'s priority at school $s$, and $\rho_{is} < \rho_{js}$ means $s$ prioritizes $i$ over $j$. Priorities are coarse in the sense that it is possible that $\rho_{is} = \rho_{js}$ for some $i \neq j$. Let $\rho_s \equiv (\rho_{is})_{i \in I}$. Denote the type of student $i$ by $\theta_i = (\succ_i, (\rho_{is})_{s \in S})$. I call $X \equiv (I, S, \succ, (c_s)_{s \in S}, (\rho_s)_{s \in S})$ an assignment problem.

2.2.1 Generalized Deferred Acceptance Mechanisms

A (stochastic) mechanism maps each assignment problem into a distribution over matchings between students and schools. Mechanisms usually use lotteries to break indifferences or ties in priorities and then use the resulting strict priorities to create a matching. A random variable $R_{is}$ denotes a lottery number of student $i$ at school $s$. Assume that at each school, $R_{is}$ is iid across students according to $U[0, 1]$. On the correlation of lottery numbers across different schools, I consider two focal regimes. Under a single tie breaker (STB), each student has a single lottery number used by all schools, i.e., $R_{is} = R_{i\ell'}$ for all $i, s, \text{ and } s'$. Under a multiple tie breaker (MTB), each student has an independent lottery number at each school, i.e., $R_{is}$ and $R_{i\ell}$ are independent for all $i$ and $s \neq s'$.

Let $r_{is} \in [0, 1]$ denote $i$'s realized lottery number at school $s$ and let $R \equiv (R_{is})_{i \in I, s \in S}$, and

\footnote{In reality, the vast majority of school districts use STB, though some cities like Washington DC, New Orleans, and Amsterdam use MTB. It is possible but needs more messy notation to extend my analysis to any structure in between STB and MTB where some schools use a common lottery while others use independent ones.}

90
I define a class of mechanisms of my interest as follows.\textsuperscript{5}

**Definition 6.** A generalized deferred acceptance (gDA) mechanism $\varphi$ is a mechanism that can be expressed as the following procedure. Take any assignment problem as given.

1. Draw $r$ according to its lottery regime (STB or MTB).
2. For each $i$ and $s$, compute modified priority $\rho_{is}^\varphi \equiv f^\varphi(\rho_{is}) + g^\varphi(rank_{is})$ where $f^\varphi(\cdot): \mathbb{N} \rightarrow \mathbb{N}$ (N is the set of positive integers) is a strictly increasing function, $rank_{is}$ is the preference rank of $s$ in $>_i$, and $g^\varphi(\cdot): \mathbb{N} \rightarrow \mathbb{N}$ is a weakly increasing function. Define $s$’s ex post strict modified priority order $>_s^\varphi$ over students by $i >^\varphi_s i'$ if $\rho_{is}^\varphi + r_{is} < \rho_{is'}^\varphi + r_{is'}$.
3. Given $>_I$ and $>_s^\varphi$, run the following (student-proposing) deferred acceptance (DA) algorithm (Gale and Shapley, 1962).

- **Step 1:** Each student $i$ applies to her most preferred acceptable school (if any). Each school tentatively keeps the highest-ranking (with respect to $>_s^\varphi$) students up to its capacity, and rejects every other student.

- **Step $t$:** Each student $i$ who was not tentatively matched to any school in Step $t - 1$ applies to her most preferred acceptable school that has not rejected her (if any). Each school tentatively keeps the highest-ranking (with respect to $>_s^\varphi$) students up to its capacity from the set of students previously tentatively matched to this school and the students newly applying, and rejects every other student.

The algorithm terminates at the first step at which no student applies to a school. Each student tentatively kept by a school at that step is allocated a seat in that school, resulting in a matching.

gDA mechanisms are parametrized by the lottery regime (STB or MTB) and $(f^\varphi, g^\varphi)$. This gDA class includes most of the mechanisms used in practice as follows.

**Deferred Acceptance Mechanism**

Given an assignment problem and realized lottery numbers, the deferred acceptance (DA) mechanism (Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003) makes

\textsuperscript{5}Others also use similar classes of mechanisms. See, for example, Ergin and Sönmez (2006); Pathak and Sönmez (2008); Agarwal and Somaini (2015).
a matching through the DA algorithm in which schools' strict priorities are induced by \( \rho_{is} + r_{is} \). Since the DA mechanism makes no modification to priorities, it corresponds to the gDA mechanism with \((f^g(m) = m, g^g(n) = 0)\).

**Boston (Immediate Acceptance) Mechanism**

Given an assignment problem and realized lottery numbers, the **Boston (immediate acceptance) mechanism** (Abdulkadiroğlu and Sönmez, 2003; Ergin and Sönmez, 2006) is defined through the following immediate acceptance algorithm.

- **Step 1:** Each student \( i \) applies to her most preferred acceptable school (if any). Each school accepts its highest-priority (with respect to \( \rho_{is} + r_{is} \)) students up to its capacity and rejects every other student.

In general, for any step \( t \geq 2 \),

- **Step \( t \):** Each student who has not been accepted by any school applies to her most preferred acceptable school that has not rejected her (if any). Each school accepts its highest-priority (with respect to \( \rho_{is} + r_{is} \)) students up to its remaining capacity and rejects every other student.

The algorithm terminates at the first step in which no student applies to a school. Each student accepted by a school at some step of the algorithm is allocated a seat in that school. The immediate acceptance algorithm differs from the DA algorithm in that when a school accepts a student at a step, in the immediate acceptance algorithm, the student is guaranteed a seat at that school, while in the deferred acceptance algorithm, that student may be later displaced by another student with a better priority status.

The Boston mechanism can be seen as modifying priorities so that each school prioritizes students ranking it higher over students ranking it lower, and it is known that the Boston mechanism is a gDA mechanism with \((f^g(m) = m, g^g(n) = (K + 1)n)\) (Ergin and Sönmez, 2006). Under this \((f^g(m), g^g(n))\), \( i \succ_{r_s}^g i' \) for all \( i \) and \( i' \) with \( \text{rank}_{is} < \text{rank}_{i's} \) regardless of original priorities \( \rho_{is} \) and \( \rho_{i's} \) and lottery numbers \( r_{is} \) and \( r_{i's} \); \( i \succ_{r_s}^g i' \) for all \( i \) and \( i' \) with \( \text{rank}_{is} = \text{rank}_{i's} \) and \( \rho_{is} < \rho_{i's} \) regardless of lottery numbers \( r_{is} \) and \( r_{i's} \).

**Charlotte Mechanism**

The mechanism used in Charlotte is the same as the Boston mechanism except that each school respects the walk zone priority ahead of preference ranks so that every student is guaranteed a seat at her walk zone school (Hastings et al., 2009; Deming, 2011; Deming et al., 2014). Assume without loss of generality that \( \rho_{is} = 1 \) means \( i \) has walk zone priority at \( s \). The **Charlotte mechanism** is a gDA mechanism with \((f^g(m) = m + 1\{m > 1\}[K + (K + 1)|S|], g^g(n) = (K + 1)n)\). Under this \((f^g(m), g^g(n))\), \( i \succ_{r_s}^g i' \) for all \( i \) and
\( i' \) with \( \rho_{i's} = 1 \) and \( \rho_{i's} > 1 \) regardless of their preferences, other aspects of priorities, and lottery numbers; \( i \succ^{\varphi} i' \) for all \( i \) and \( i' \) with \( 1\{\rho_{i's} = 1\} = 1\{\rho_{i's} = 1\} \) and \( \text{rank}_{i's} < \text{rank}_{i's} \); \( i \succ_{\varphi} i' \) for all \( i \) and \( i' \) with \( 1\{\rho_{i's} = 1\} = 1\{\rho_{i's} = 1\} \), \( \text{rank}_{i's} = \text{rank}_{i's} \), and \( \rho_{i's} < \rho_{i's} \) regardless of their lottery numbers.

### 2.2.2 First Choice Empirical Research Design

Many empirical studies use data from gDA mechanisms to identify and estimate the causal effect of getting in a school on outcomes such as test scores, crime rates, college attendance, and earnings (Hastings et al. (2009), Bloom et al. (2010), Deming (2011), Abdulkadiroglu et al. (2013), Deming et al. (2014), Bergman (2014)). To describe their empirical strategy, fix any gDA \( \varphi \) and assignment problem \( X \) that generate the data at hand. Following the standard notation in econometrics, let \( D_{is}(r) = 1 \) if student \( i \) is assigned the treatment school \( s \) under (realized or counterfactual) lottery number \( r \); \( D_{is}(r) = 0 \) otherwise. Define \( \text{First}_{is}(r) = \{i \in I | \text{rank}_{i's} = 1 \text{ and there exists } i' \text{ such that } \text{rank}_{i's} = 1, \rho_{i's} = \rho_{i's}, \text{ and } D_{is}(r) \neq D_{i's}(r)\} \) as the set of students who rank \( s \) first and are in \( s \)'s "marginal priority group," where some students are assigned \( s \) but others are not though all of them share the same priority at \( s \). Let \( r_0 \) be the realized lottery numbers in the data.

The above papers' empirical strategy, which I call the **first choice research design**, compares the outcomes of students with \( D_{is}(r_0) = 1 \) against those with \( D_{is}(r_0) = 0 \) within \( \text{First}_{is}(r_0) \). The outcome difference between the two groups is then interpreted as the causal effect of being assigned \( s \) for those in \( \text{First}_{is}(r_0) \). The idea is that since all students in \( \text{First}_{is}(r_0) \) rank \( s \) first and share the same priority at \( s \), whether they get an offer from \( s \) should be determined solely by their lottery numbers and so independent from students’ covariates or choices correlated with outcomes. Therefore, offers from \( s \) within \( \text{First}_{is}(r_0) \) can be thought as if randomly assigned in a randomized controlled trial. Albeit intuitive, for the first choice research design to successfully identify a causal effect by this logic, assignments to \( s \) within \( \text{First}_{is}(r_0) \) have to be indeed random and not confounded.

---

\( ^6 \)Empirical researchers are often interested in the effect of a group of schools rather than an individual school. My analysis applies to such group-level treatments too. Also, when the effect of interest is that of attendance or enrollment rather than assignment (as in most of the above empirical examples), I would see attendance or enrollment as the endogenous treatment and use assignment as an instrument for the treatment. I would then use an instrumental variable method to estimate the effect of attendance or enrollment. My analysis is applicable to such instrumental variable settings. See footnote 11.

\( ^7 \)See also closely related papers such as Jackson (2010, 2012); Ajayi (2013); Hastings et al. (2013); Pop-Eleches and Urquiola (2013); Deming (2014); Dobbie and Fryer (2014); Lucas and Mbiti (2014a); Abdulkadiroglu et al. (2015a); Kirkeboen et al. (2015). Some of these papers use extensions of the first choice research design discussed in this section. See section 2.4.3 for details. Some of them leverage regression-discontinuity-type locally tie breaking rather than lotteries.

\( ^8 \)Since \( \text{rank}_{i's} = \text{rank}_{i's} = 1 \) holds and \( f^{\varphi}(\cdot) \) is strictly increasing by definition, \( \rho_{i's} = \rho_{i's} \) is equivalent to \( (f^{\varphi}(\rho_{i's}) + g^{\varphi}(1) \equiv \rho_{i's} = \rho_{i's} \equiv f^{\varphi}(\rho_{i's}) + g^{\varphi}(1)) \). Thus, I could replace \( \rho_{i's} = \rho_{i's} \) with \( \rho_{i's} = \rho_{i's} \) in the definition of \( \text{First}_{is}(r) \) without changing anything in the following analysis. Finally, note that it is possible \( \text{First}_{is}(r) = \emptyset \) for some or even all \( r \).
by non-random preferences or priorities. This requirement is formalized as the following concept.

**Definition 7.** The first choice research design extracts a random assignment under a gDA \( \varphi \) for school \( s \) at assignment problem \( X \) if given \( \varphi \) and \( X \), for all potential realized lottery numbers \( r \) and all \( j, k \in First_s(r) \),

\[
P(D_{js}(R) = 1) = P(D_{ks}(R) = 1).^9
\]

The first choice research design extracts a random assignment under a gDA \( \varphi \) if it does so for every school \( s \) at every assignment problem \( X \).

This condition says that conditional on being in \( First_s(r_0) \), offers from \( s \) are random and independent from students’ preferences and priorities summarized by \( \theta_i \). In the econometric terminology, this requires that the propensity score (Rosenbaum and Rubin, 1983) is constant within fixed set \( First_s(r_0) \) of students.\(^10\) Virtually all econometric program evaluation models (Heckman and Vytlacil, 2007; Manski, 2008; Angrist and Pischke, 2009) require this conditional independence for the first choice research design to identify a causal effect.\(^11\)

### 2.2.3 Motivating Example

Despite its intuitive construction, the first choice research design may fail to extract a random assignment. Consider the following example.

**Example 1.** There are applicants 1, 2, 3, and schools A and B with the following preferences and priorities:

\[\succ_1: A, B, \emptyset\]
\[\succ_2: A, \emptyset\]
\[\succ_3: B, A, \emptyset\]
\[\rho_A : 3, \{1, 2\}\]
\[\rho_B : 1, \{2, 3\}\]

\(^9\)An equivalent requirement is \( P(D_{ts}(R) = 1|\theta_t = \theta_k) = P(D_{ts}(R) = 1|\theta_t = \theta_j) \), where \( P(D_{ts}(R) = 1|\theta_t = \theta_k) \) means the probability of assignment to \( s \) for an arbitrary student of type \( \theta_k \), since any gDA guarantees that the assignment probability of a student depends only on her type and not on her identity.

\(^10\)My result is robust to changing the definition of extracting a random assignment. See section 2.4.2.

\(^11\)When assignment within \( First_s(r_0) \) is used as an instrument for an endogenous treatment such as enrollment, Definition 7 is interpreted as a conditional independence requirement for the instrument. For the instrument to identify a causal effect, it usually needs to satisfy not only conditional independence but also additional properties such as “exclusion” or “monotonicity.” See Heckman and Vytlacil (2007); Manski (2008); Angrist and Pischke (2009). In this case, Definition 7 becomes a key necessary condition. See also sections 2.4.3 and 2.4.4 for related discussions.
where \(\succsim_1: A, B, \emptyset\) means 1 prefers \(A\) over \(B\) and both schools are acceptable for 1. \(\rho_A: 3, \{1, 2\}\) means that \(A\) prioritizes 3 over 1 and 2 and is indifferent between 1 and 2. The capacity of each school is 1. The treatment school is \(A\).

The first choice research design does not extract a random assignment under the DA mechanism for \(A\) in Example 1. Under the DA mechanism, 1 is assigned to \(A\) when 1 has a better lottery number than 2 at \(A\). Otherwise, 3 is assigned to \(A\). Each of the two cases occurs with equal probability 0.5. Thus,

\[
P(D_{1A}(R) = 1) = 1/2 
eq 0 = P(D_{2A}(R) = 1),
\]

despite

\[
First_A(r) = \begin{cases} 
\{1, 2\} & \text{if } r_{1A} < r_{2A} \\
\emptyset & \text{otherwise}
\end{cases}
\]

Therefore, the first choice research design does not extract a random assignment under the DA mechanism.\(^{12}\) Imagine that school \(A\) has no achievement effect, and student 1 ranks more schools than student 2 because student 1 is more eager and higher achieving (regardless of whether to attend \(A\)). Conditional on \(First_A(r) = \{1, 2\}\), student 1 gets the seat at \(A\) but student 2 never does. Comparing 1 and 2 within \(First_A(r) = \{1, 2\}\) is thus likely mistakenly conclude \(A\) has a positive achievement effect. This raises the question of under what mechanisms the first choice research design extracts a random assignment as intended.

### 2.2.4 Strategy-proofness for Schools

The success and failure of the first choice research design turn out to be tightly connected to another seemingly unrelated property of mechanisms. So far I have been regarding priorities and lottery numbers as given public information. In this section, I depart from this assumption and imagine a hypothetical thought experiment in which schools have priorities and lottery numbers as their private information, and the priorities and lottery numbers represent their preferences; I will come back to the interpretation of this thought experiment at the end of this section. A gDA mechanism asks schools to report priorities and lottery numbers, which may or may not be the same as the true ones. The gDA mechanism then uses the reported priorities and lottery numbers to create a matching.

Given any \((I, S, \succsim_I, (c_s)_{s \in S})\), let \(\varphi(\rho, r) \equiv (\varphi_s(\rho, r))_{s \in S}\) be the matching produced by a gDA \(\varphi\) when the reported priorities and lottery numbers are \((\rho, r)\). I say \(s\)'s preference

\(^{12}\)It is possible to create a similar counterexample even when there are no priorities as long as ties are broken by MTB. Also, section 2.4.4 demonstrates the first choice research design would fail even if I modify it to the more refined version that pools applicants who rank the treatment school first and share the same priority at every school.
\( \succ_s \), which is defined over the set of subsets of I, is **responsive** with respect to \((c_s, \rho_s, r_s)\) (Roth and Sotomayor, 1990) if

1. For any \( i, i' \in I \), if \( \rho_{is} + r_{is} < \rho_{i's} + r_{i's} \), then for any \( I' \subseteq I \setminus \{i, i'\} \), \( I' \cup \{i\} \succ_s I' \cup \{i'\} \),
2. For any \( i \in I \) and any \( I' \subseteq I \) such that \( |I'| < c_s, I' \cup \{i\} \succ_s I' \), and
3. \( \emptyset \succ_s I' \) for any \( I' \subseteq I \) with \(|I'| > c_s \).

I use these concepts to define the following well-established property of a mechanism.

**Definition 8.** \( \varphi \) is **strategy-proof** for school \( s \) at assignment problem \( X \) if given \( X \), for all \((\rho^*, r^*)\), all \( \succ_s \) responsive with respect to \((c_s, \rho^*_s, r^*_s)\), and all \((\rho'_s, r'_s)\),

\[ \varphi_s(\rho^*, r^*) \gtrless_s \varphi_s((\rho'_s, r'_s), (\rho'^*_s, r'^*_s)), \]

where \( \gtrless_s \) is the weak preference associated with \( \succ_s \).\(^{13}\) \( \varphi \) is strategy-proof for schools if it is strategy-proof for every school \( s \) at every \( X \).

Though my setting has a stochastic nature due to lotteries, this usual definition of strategy-proofness is a non-stochastic concept. A standard behavioral interpretation of this concept is that no school ever has a preference manipulation that is profitable with respect to its true preference. It is crucial to note, however, that I am not concerned with this usual interpretation. As will become clearer in the next section, in contrast to usual studies on strategy-proofness, my analysis treats strategy-proofness not as a desideratum or an incentive compatibility requirement but as an algorithmic property. I am interested not in strategy-proofness itself but in its implications for empirical research. As a result, I do not need to assume or hope anything about school behavior. The following usual questions about strategy-proofness for schools are thus all irrelevant for interpreting my analysis: Do schools have intrinsic preferences? Are school preferences consistent with priorities? Do schools ever "game" the system? See section 2.4.1 for further discussions.

### 2.3 Results

#### 2.3.1 Sufficiency

Whether the first choice research design extracts a random assignment turns out to have a tight connection to strategy-proofness for schools. Most of all, strategy-proofness for schools is sufficient for the first choice research design to extract a random assignment.

\(^{13}\)The domain for \((\rho^*_s, r^*_s)\) and \((\rho'_s, r'_s)\) is \([1, \ldots, K]|I| \times [0, 1]|I|\), where \(|I|\) is the number of students. \( \rho^* \) may or may not be the same as \( \rho \) in \( X \). I do not require either \( r \) to be consistent with \( \varphi \)'s lottery structure. My analysis would not change even if I modified "all \( \succ_s \) responsive with respect to \((c_s, \rho^*_s, r^*_s)\), and" to "there exists \( \succ_s \) responsive with respect to \((c_s, \rho^*_s, r^*_s)\) such that for."
**Theorem 1.** The first choice research design extracts a random assignment under a gDA $\varphi$ if $\varphi$ is strategy-proof for schools.

The proof is in Appendix 2.6.1. Combined with existing results on strategy-proofness for schools, Theorem 1 provides affirmative results for the first choice research design under some of the gDA mechanisms. (I will see yet another key implication in section 2.3.3.)

**Corollary 2.** a) The first choice research design extracts a random assignment under the Boston mechanism with any lottery regime.

b) The first choice research design extracts a random assignment under the DA mechanism with STB when there are no priorities ($\rho_{is} = \rho_{js}$ for all $i, j, s$). This mechanism is often called random serial dictatorship.

**Proof.** (a) follows from Theorem 1 and Ergin and Sönmez (2006)'s Theorem 2 that the Boston mechanism is strategy-proof for schools. (b) follows from the proof of Theorem 1 and the fact that under the DA mechanism, truth-telling is optimal for any school $s$ when all the other schools report the same preference as $s$'s true preference, which corresponds to the case with STB and no priorities. See Appendix 2.6.2 for details.

The intuition for Theorem 1 is as follows. First of all, a sufficient condition for the first choice research design to extract a random assignment for school $s$ is that as in randomized controlled trials, any permutation or shuffle of lottery numbers $r_{is}$ within $First_s(r_0)$ translates into the corresponding permutation of offers or assignments $D_{is}(r)$. Let me call this the "Fisher property" after Ronald Fisher, the inventor of randomized experiments. Strategy-proofness for schools turns out to guarantee this Fisher property, as the following two-step argument illustrates.

For simplicity, consider the easiest case with no priority and MTB. (It will become clear in a few paragraphs why MTB is easier than STB.) The first step of the intuition is summarized as "a preference manipulation is a lottery permutation." It is a retrospectively obvious observation that with no priority and MTB, each school's strict priority is pinned down solely by lottery numbers at the school. Thus any preference manipulation by school $s$ in the hypothetical thought experiment (used for defining strategy-proofness) corresponds to a permutation of lottery numbers at $s$ in the real world. This step does not use strategy-proofness for schools.

The second step is summarized as "an unprofitable preference manipulation is an offer permutation." This second step is more subtle and claims that when a preference manipulation by school $s$ is unprofitable for $s$, the associated permutation of lottery numbers $r_{is}$ results in the corresponding permutation of assignments or offers $D_{is}(r)$. To get an idea of this, consider students $i_0, i_1 \in First_s(r_0)$ such that $s$ prefers/prioritizes $i_1$ over $i_0$ in

---

14I believe this fact is folk knowledge. A formal proof is implied by Hatfield et al. (2016)'s Lemma 1 and Proposition 7.
the true preference or the associated default lottery number. Assume that \( i_1 \) is assigned to \( s \) while \( i_0 \) is not under the true preference or the associated lottery number. Also, suppose that \( i_1 \) and \( i_0 \) are consecutive in \( s \)'s true preference or the associated default lottery number, i.e., there is no third student whom \( s \) prefers over \( i_0 \) but dis-prefers over \( i_1 \).

Now imagine a hypothetical preference manipulation by school \( s \) that switches \( i_1 \) and \( i_0 \) and claims to prefer \( i_0 \) over \( i_1 \). What does it mean that the data-generating gDA mechanism is strategy-proof for schools and the manipulation is unprofitable for \( s \)? It means that the mechanism would not let \( s \) newly get a student \( s \) prefers over \( i_1 \) (in \( s \)'s true preference) by virtue of the manipulation. Usually this in turn requires that the manipulation would not make any change in which students school \( s \) gets among students school \( s \) prefers over \( i_1 \). If so, the number of seats assigned to students school \( s \) prefers over \( i_1 \) is unchanged before and after the manipulation. Then, after \( s \)'s manipulation, the seat originally assigned to \( i_1 \) has to go to some student school \( s \) dis-prefers over \( i_1 \). This would require the gDA mechanism to assign the excess seat to \( i_0 \) (instead of \( i_1 \)), since \( s \) claims to prefer \( i_0 \) over \( i_1 \) in the manipulated preference and both \( i_0 \) and \( i_1 \) are in \( First_s(r_0) \) and thus rank \( s \) first.\(^{15}\)

This is the desired Fisher property: \( s \)'s preference manipulation of switching \( i_0 \) and \( i_1 \), which corresponds to permuting \( i_0 \) and \( i_1 \)'s lottery numbers at \( s \) \((r_{i_0s} \) and \( r_{i_1s} \)), results in the corresponding permutation of \( i_0 \) and \( i_1 \)'s offer statuses at \( s \) \((D_{i_0s}(r) \) and \( D_{i_1s}(r)) \). Therefore, if a gDA mechanism is strategy-proof for schools, any preference manipulation is unprofitable and so any lottery permutation within \( First_s(r_0) \) translates into the corresponding permutation of assignments. This enables lottery permutations to induce a successful randomized experiment within \( First_s(r_0) \).

This intuition is still far from a proof, however, since Theorem 1 allows for additional complications due to priorities and STB, which the above intuition ignores. In particular, when lotteries are correlated across schools as in STB, a lottery permutation affects multiple schools’ preferences/priorities simultaneously. This destroys the exact correspon-

\(^{15}\)More precisely, under any gDA mechanism \( \varphi \), given \( s \)'s preference manipulation \((\rho'_s, r'_s)\), it is the case that \( f^{\varphi}(\rho'_{i_0s}) + r'_{i_0s} < f^{\varphi}(\rho'_{i_1s}) + r'_{i_1s} \) since \( \rho'_{i_0s} + r'_{i_0s} < \rho'_{i_1s} + r'_{i_1s} \), which is because of the assumption that \( s \) claims to prefer \( i_0 \) over \( i_1 \) in \((\rho'_s, r'_s)\). Also, \( g^{\varphi}(rank_{i_0s}) = g^{\varphi}(rank_{i_1s}) \) by \( rank_{i_0s} = rank_{i_1s} = 1 \). (Recall \( i_0, i_1 \in First_s(r_0) \).) These jointly imply that

\[
\rho_{i_0s}^{\varphi} + r_{i_0s}^{\varphi} = f^{\varphi}(\rho'_{i_0s}) + g^{\varphi}(rank_{i_0s}) + r'_{i_0s} < f^{\varphi}(\rho'_{i_1s}) + g^{\varphi}(rank_{i_1s}) + r'_{i_1s} = \rho_{i_1s}^{\varphi} + r_{i_1s}^{\varphi},
\]

where \( \rho_{i_0s}^{\varphi} \) is the modified priority of student \( i \) at school \( s \) under \( s \)'s unilateral manipulation \((\rho'_s, r'_s)\). Thus, under any gDA mechanism \( \varphi \), \( s \) has to accept \( i_0 \) before accepting \( i_1 \) or anybody below \( i_1 \) (in \( \rho_{is} + r_{is} \)) under the manipulation \((\rho'_s, r'_s)\).

Note that this reasoning depends on the assumption that \( i_1 \) and \( i_0 \) rank \( s \) first. In fact, Section 2.4.3 shows that the intuition described here breaks down if I extend the first choice research design to another design that contains students who do not rank \( s \) first. For the extended research design, strategy-proofness for schools is shown to be no longer sufficient for extracting a random assignment even in the case with no priorities.
dence between a lottery permutation and a unilateral preference manipulation by a single school, and invalidates the above simple intuition. Nevertheless, the proof in Appendix 2.6.1 shows that the conclusion generally holds.

It may also help to illustrate Theorem 1 with a stylized example. Consider the Boston mechanism in Example 1. Imagine a thought experiment where schools have private preferences and the mechanism asks schools to report their preferences. First of all, school A is never matched with student 3 since 3 ranks A second and the seat at A is always filled by one of the two students who rank A first. A is thus matched with either 1 or 2. When A's true preference is such that $1 \succ_A 2$, A is matched with more preferred 1 by truth-telling. When A's true preference is with $2 \succ_A 1$, A is matched with more preferred 2 by truth-telling. Thus, there is no profitable preference manipulation for A; the Boston mechanism is strategy-proof for A in Example 1. This is a special case of the fact that the Boston mechanism is strategy-proof for schools (Ergin and Sönmez, 2006).

As is expected from strategy-proofness and Theorem 1, it is also true that the first choice research design extracts a random assignment for A in Example 1 under the Boston mechanism. Since only 1 and 2 rank A first with the same priority and only one of them with a better lottery number is assigned A under any r, $First_A(r) = \{1, 2\}$ for all r. Enumerating all lottery outcomes shows that 1 and 2 share the same assignment probability of 1/2 at A, i.e., $P(D_{1A}(R) = 1) = P(D_{2A}(R) = 1) = 1/2$. Thus the first choice research design extracts a random assignment.

2.3.2 Almost Necessity

Theorem 1 shows strategy-proofness for schools is sufficient for the first choice research design to extract a random assignment. It turns out to be not only sufficient but also nearly necessary. Consider the DA mechanism in Example 1. Imagine A's true preference is $3 \succ_A 1 \succ_A 2$ while B's is $1 \succ_B 2 \succ_B 3$. Under these true preferences, A is matched with 1. If A misreports $3 \succ'_A 2 \succ'_A 1$, however, A is matched with 3, the most preferred student with respect to $\succ_A$. Therefore, the DA mechanism is not strategy-proof for A in Example 1. This is a reconfirmation of the classic result that the DA mechanism is not strategy-proof for schools (Roth and Sotomayor, 1990).

Intuitively, A benefits from manipulating its preference and rejecting 1 by the following chain reactions of rejections and applications. After rejected by A, 1 next applies for B, which results in B's rejecting 3. 3, the most preferred student for A, then applies for and benefits A. The same chain reactions cause the first choice research design to fail. Depending on schools ranked below A, different applicants cause different chain reactions that have different effects on assignment probabilities at A. As I already saw in section 2.2.3, this can cause applicants in $First_A(r_0)$ to have different assignment probabilities.

---

16Since A's capacity is 1, I do not need to distinguish its preference over sets of students and its priority order over individual students.
Similarly, the first choice research design does not extract a random assignment either under the Charlotte mechanism, another mechanism that is not strategy-proof for schools and used in Hastings et al. (2009), Deming (2011), Deming et al. (2014): Suppose that 3's priority at A and 1's priority at B in Example 1 are walk zone priorities. In this case, the Charlotte mechanism coincides with the DA mechanism in Example 1. The Charlotte mechanism is thus manipulable by schools and the first choice research design fails under it in the same way as the DA mechanism. In addition, section 2.4.5 shows the same failure of the first choice design for yet another mechanism that is not strategy-proof for schools (the top trading cycles mechanism). In these senses, strategy-proofness for schools is almost necessary for the first choice research design to extract a random assignment.

**Empirical Illustration**

Denver Public Schools use the usual DA mechanism for unified public and charter school admissions. As I saw in section 2.3.2, the DA mechanism is not strategy-proof for schools and may not extract a random assignment via the first choice research design. To see whether the first choice research design extracts an exactly random assignment in Denver, I use the data from its DA mechanism as follows.

1. Given student preferences and school priorities and capacities as fixed, I simulate the DA mechanism by drawing counterfactual lottery numbers one million times. This gives us an approximate assignment probability \( \hat{P}(D_{is}(R) = 1) \) for each student \( i \) and school \( s \), i.e., the empirical frequency of \( i \)'s being assigned \( s \).\(^\text{18}\)

2. For each school \( s \) and each student \( i \) in \( \text{First}_s(r_0) \), I demean \( i \)'s assignment probability by subtracting the mean of assignment probabilities at \( s \) across all students in \( \text{First}_s(r_0) \). That is, I compute

\[
\hat{P}_{\text{demean}}(D_{is}(R) = 1) = \hat{P}(D_{is}(R) = 1) - \frac{\sum_{j \in \text{First}_s(r_0)} \hat{P}(D_{js}(R) = 1)}{\left| \text{First}_s(r_0) \right|}.
\]

\(^{17}\)In contrast, under the Boston mechanism analyzed in the last section, such chain reactions do not affect assignments to A. By its construction, under the Boston mechanism, each school is forced to prioritize students ranking it higher over students ranking it lower. As a result, chain reactions caused by student \( i \) at schools ranked below A involve only students who rank A lower than student \( i \) does. Since A rejects \( i \), A has to also reject any students ranking A lower than \( i \) and thus A never accept students involved in chain reactions \( i \) causes. Thus different chain reactions caused by different students have the same effect on assignments at A, that is, no effect at all. This is the reason why the Boston mechanism is strategy-proof for schools and the first choice research design extracts a random assignment under it.

\(^{18}\)In Denver, each school is divided into multiple sub-schools (called "buckets") with their own priorities and capacities. Buckets correspond to schools in my theoretical model. Below I use "schools" to mean buckets. See Abdulkadiroglu et al. (2015a) for more details of the Denver school admissions system.
(3) I plot this assignment probability deviation $\hat{P}_{demean}(D_{is}(R) = 1)$ across all schools $s$ and all students in $First_s(r_0)$.

Figure 2.5.1 is the resulting figure.\(^{19}\) If the 1st choice strategy extracts a random assignment, $\hat{P}(D_{is}(R) = 1) \approx \hat{P}(D_{js}(R) = 1)$ for all $s$ and all $i, j \in First_s(r_0)$ and so the assignment probability deviation $\hat{P}_{demean}(D_{is}(R) = 1)$ would be almost 0 (depending on simulation errors) for all $s$ and all $i$ in $First_s(r_0)$. As the figure shows, however, there are many visible assignment probability deviations that are far from 0. The mean is almost 0 (by construction) but the standard deviation is around 0.19. This provides an empirical illustration of the theoretical necessity of strategy-proofness for schools. I leave for future research whether these assignment probability deviations result in any serious bias in treatment effect estimates.

2.3.3 Sufficiency Revisited

I have demonstrated that the first choice research design successfully extracts a random assignment under a gDA mechanism if and almost only if the mechanism is strategy-proof for schools. This provides a justification for the research design under strategy-proof mechanisms such as the Boston mechanism. It also implies a caveat for it under non-strategy-proof mechanisms such as the DA, Charlotte, and top trading cycles mechanisms.

What is puzzling is that despite the above counterexample to the DA and Charlotte mechanisms, applications of the first choice research design under these mechanisms often find empirical support for successful randomization. In particular, they usually find that in $First_s(r_0)$, observable covariates of those with $D_{is}(r_0) = 1$ and those with $D_{is}(r_0) = 0$ are statistically balanced, a standard check of a necessary condition for successful randomization.\(^{20}\) How can I resolve the tension between their empirical validity findings and my theoretical invalidity result?

A potential resolution is already hinted by Theorem 1, the sufficiency of strategy-proofness for schools. Though the DA and Charlotte mechanisms are not strategy-proof for schools in general, subsequent work empirically and theoretically show that they are often approximately so in large markets with many students and schools (Roth and Peranson, 1999; Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Azevedo and Budish, 2013). The reason is that as the number of students and schools grows, chain reactions of rejections and applications at schools ranked below a manipulating school — which make the DA and Charlotte mechanisms manipulable in Example 1 — become less likely to go back to the manipulating school and benefit it.

\(^{19}\)Each student ranks only one school first and is in at most one school's $First_s(r_0)$. No student appears more than twice in Figure 2.5.1.

\(^{20}\)See for example Hastings et al. (2009)'s Table VI, Bloom et al. (2010)'s Table 2.1, Deming (2011)'s Table I, Abdulkadiroglu et al. (2013)'s Table 2, Deming et al. (2014)'s Table A2.
Unlike the small counterexample in Example 1, empirical work is only done with data with at least hundreds of students. Existing empirical settings may thus be subject to large market forces that make the DA and Charlotte mechanism almost non-manipulable by schools. If so, Theorem 1 guarantees the first choice research design approximately extracts a random assignment even under the DA and Charlotte mechanism. In fact, Figure 2.5.2, which plots assignment probabilities at A under various expansions of Example 1, reveals that as the market size grows, the discrepancy between 1 and 2’s assignment probabilities at A disappears, implying that breaks in randomization under the first choice research design become less and less relevant. (A computer program to implement this simulation is available upon request.)

This may explain why the first choice research design often works in empirical applications even under mechanisms that are not strategy-proof for schools. At the same time, the existing empirical support for the first choice research design under the DA and Charlotte mechanisms (recall footnote 20) can be reinterpreted as suggesting that large market forces emphasized by the above theoretical market design papers are empirically relevant.

My necessity and sufficiency findings provide a basis for understanding why the first choice research design could fail under some prominent mechanisms and why nevertheless existing empirical studies succeeded in implementing it without involving significant breaks in randomization. The reason is not that they extract an exactly random assignment, but rather that they implicitly resort to large market forces that make a broken random assignment approximately random. Theorem 1 thus provides a delicate asymptotic justification for the research design even under some mechanisms that are not strategy-proof for schools in general.

More precisely, Kojima and Pathak (2009)’s Theorem 1 and my Theorem 1 imply that for any “regular” sequence of random assignment problems indexed by the number of schools, the expected fraction of schools at which the first choice research design does not extract a random assignment goes to 0 as the number of schools goes to infinity.

The large market exercise in Figure 2.5.2b also suggests that strategy-proofness for schools (is nearly necessary but) may not be exactly necessary for the first choice research design to extract a random assignment. As is expected from Figure 2.5.2b, the first choice research design extracts a random assignment in the limit of expansions of Example 1 in Figure 2.5.2b, where there are infinitely many students of each type and infinitely many seats at each school (Abdulkadiroğlu et al., 2015a). However, the DA and Charlotte mechanisms are not strategy-proof for schools even in this limit by essentially the same logic as in section 2.3.2 (Azvedo, 2014). This also means that the current paper’s asymptotic justification of the first choice research design and Abdulkadiroğlu et al. (2015a)’s are independent and based on different logics.

One implication of this is that many empirical studies using the first choice design under non-strategy-proof mechanisms provide an unexpected set of settings where treatment assignment is only asymptotically random. For inference, therefore, ideally they should use recent econometric program evaluation methods based on asymptotically random treatment assignment, such as Canay et al. (2014) and Belloni et al. (2014).

102
2.4 Discussion and Extensions

2.4.1 Interpretation of Strategy-proofness for Schools

Section 2.3 demonstrates a tight connection between breaks in randomization under the first choice research design and room for potential profitable manipulations by schools in a hypothetical thought experiment where schools have and report preferences. "room for potential" and "in a hypothetical thought experiment" are key qualifications and they re-emphasize the above point is true regardless of whether schools take advantage of the room for manipulations or whether schools have and report preferences in reality. I treat strategy-proofness for schools not as a desideratum or an incentive compatibility constraint but as a purely algorithmic or physical property. As a result, the implications of I main result need no assumption on school behavior and incentives.

2.4.2 Alternative Definition of Extracting a Random Assignment

Definition 7, I main criterion of extracting a random assignment, requires that all students in realized fixed set $First_s(r_0)$ share the same assignment probability. A possible alternative definition treats $First_s(R)$ as random and requires that $P(D_{ia}(R) = 1|i \in First_s(R), \theta_i = \theta) = P(D_{ia}(R) = 1|i \in First_s(R))$ for all $i$ (for whom these probabilities are defined). $R$ is random lottery numbers rather than the realized ones.24

All of I arguments can be modified to accommodate this alternative definition. See footnote 37 for why Theorem 1 is correct even under the alternative definition. The discussion about the Boston mechanism under Example 1 in section 2.3.1 goes through even under the alternative definition since $P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta) = 1/2$ for $i = 1, 2$ and is independent from $\theta$. The analysis of the DA and Charlotte mechanisms in Example 1 in sections 2.2.3 and 2.3.2 also remains essentially the same since $P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_1) = 1 \neq 0 = P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_2)$.

To complement this section, section 2.4.4 will discuss additional robustness to changes in the definition of the first choice research design.

2.4.3 Qualification Instrumental Variable Research Design

So far, I have been focusing on the first choice research design. Several papers use an extension of the first choice research design, which I call the qualification instrumental variable (IV) research design (Pop-Eleches and Urquiola, 2013; Dobbie and Fryer, 2014; 24This alternative definition requires that the distribution of offer $D_{ia}(R)$ is independent from type $\theta_i$ or the propensity score as a confounder conditional on random event $i \in First_s(R)$. This flavor of independence conditional on a random event is reminiscent of Chamberlain (1980) and Rosenbaum (1984)'s conditional logit panel frameworks, where the treatment distribution is independent from individual heterogeneity conditional on the random empirical frequency of being treated in the past.
Lucas and Mbiti, 2014a; Angrist et al., 2015b). While the first choice research design tries to make offers random by focusing on a subset of students, the qualification IV research design considers all students and codes a random instrumental variable for non-random offers. Define the qualification IV by 

\[ Z_{is}(r) = \begin{cases} \rho_{is} + r_{is} \leq \max\{\rho_{js} + r_{js} | D_{js}(r) = 1\} & \text{if there is no } j \text{ with } D_{js}(r) = 1, \\ 0 & \text{otherwise}. \end{cases} \]

If \( Z_{is}(r) = 0 \) for all \( i \), then define \( Z_{is}(r) = 0 \) for all \( i \). The qualification IV for a student at a school is turned on if her priority rank at the school is better than that of some student assigned the school.

The qualification IV looks random conditional on \( \rho_{is} \) and is likely to be correlated with offer \( D_{is}(R) \) since \( i \) never get an offer when she is not qualified (\( Z_{is}(r) = 0 \)) but she can get an offer when she is qualified (\( Z_{is}(r) = 1 \)). Based on this idea, the qualification IV research design instruments for offer \( D_{is}(R) \) by qualification \( Z_{is}(R) \) conditional on \( \rho_{is} \). For this research design to identify a causal effect, the qualification IV needs to be random conditional on \( \rho_{is} \). (See references in footnote 11.)

**Definition 9.** The qualification IV research design extracts a random assignment under a gDA \( \varphi \) for school \( s \) at assignment problem \( X \) if given \( \varphi \) and \( X \), for all \( \rho \) and \( \theta \),

\[ P(Z_{is}(R) = 1 | \rho_{is} = \rho, \theta_i = \theta) = P(Z_{is}(R) = 1 | \rho_{is} = \rho). \]

The qualification IV research design extracts a random assignment under a gDA \( \varphi \) if it does so for every \( s \) at every \( X \).

The qualification IV research design nests the first choice research design in the sense that the former extracts a random assignment only if the latter does so. It is because the sample under the qualification IV research design, which is the set of all students, contains \( First_{is}(r) \) for all \( r \) and \( D_{is}(R) = Z_{is}(R) \) for all \( i \in First_{is}(r) \) and all \( r \). The random assignment of \( D_{is}(R) \) within \( First_{is}(r_0) \) is thus necessary for the random assignment of \( Z_{is}(R) \) for all students. Hence, the qualification IV research design needs to satisfy a stronger condition than strategy-proofness for schools to successfully extract a random assignment. It turns out that no gDA mechanism satisfies such a restrictive condition even in the simple case with no priorities.

**Proposition 2.** Even with no priorities \( \rho_{is} = \rho_{js} \) for all \( i, j, \) and \( s \), there is no gDA mechanism under which the qualification IV research design extracts a random assignment.\(^{25}\)

The proof of and intuition behind the Proposition are in Appendix 2.6.3.\(^{26}\) Proposition 2 sheds light on a contrast between the qualification IV and the first choice research design.\(^{25}\)The qualification IV research design does not extract a random assignment even under the top trading cycles mechanism. See Section 2.4.5 and Appendix 2.6.3.\(^{26}\)The proof also shows that (1) the negative result persists even if I use original priorities to define an alternative qualification IV as \( Z_{is}(r) = \begin{cases} \rho_{is} + r_{is} \leq \max\{\rho_{js} + r_{js} | D_{js}(r) = 1\} & \text{if there is no } j \text{ with } D_{js}(r) = 1, \\ 0 & \text{otherwise}. \end{cases} \) and (2) the qualification IV research design can fail even if I modify it to the more refined version that conditions on having the same priority at all schools.
designs. Unlike the first choice research design, strategy-proofness for schools is no longer sufficient for the qualification IV research design to extract a random assignment and it may extract an unintended broken random assignment not only under the DA mechanism but also under the Boston mechanism. In this sense, the first choice research design is a more robust research design than the qualification IV research design.

2.4.4 Alternative Definitions of Research Designs

Section 2.4.3 shows a potential trouble with the qualification IV \( Z_{1s}(r) \equiv 1\{\rho_{1s}^p + r_{1s} \leq \max\{\rho_{js}^p + r_{js}\mid D_{js}(r) = 1\}\} \), where \( \max\{\rho_{js}^p + r_{js}\mid D_{js}(r) = 1\} \) is a random priority cutoff that varies as the lottery outcome changes. One may expect a modification of the qualification IV solves the problem. For any constant \( \pi \in [0, K + 1]\), define the constant cutoff qualification IV by \( Z_{1s}^\pi(r) \equiv 1\{\rho_{1s}^p + r_{1s} \leq \pi\} \). In practice, I would define \( \pi \equiv \max\{\rho_{js}^p + r_{0js}\mid D_{js}(r_0) = 1\} \) where \( r_0 \) is the realized lottery number in the data. The constant cutoff qualification IV trivially extracts a random assignment since \( P(Z_{1s}^\pi(R) = 1|\rho_{1s}^p = \rho, \theta_i = \theta) = P(r_{1s} \leq \pi - \rho_{1s}^p|\rho_{1s}^p = \rho, \theta_i = \theta) = \pi - \rho \), which is independent from \( \theta \) conditional on \( \rho_{1s}^p = \rho \).

However, the constant cutoff qualification IV entails new problems other than randomness. First, when using \( \pi \equiv \max\{\rho_{js}^p + r_{0js}\mid D_{js}(r_0) = 1\} \), I define or select an instrument depending on the realized data. Such data-dependent model selection often makes standard inference invalid (Leamer, 1978). In addition, perhaps more importantly, the constant cutoff qualification IV may violate other requirements for a valid IV than random assignment or independence even in the simplest case with no priority. To see this, consider the following example.

Example 2. There are applicants 1, 2, 3, 4 and schools A and B with the following preferences and priorities:

\[
\succ_1, \succ_2: A, B, \emptyset \\
\succ_3, \succ_4: B, \emptyset \\
\rho_A, \rho_B : \{1, 2, 3, 4\}.
\]

The capacity of each school is 1. The treatment school is A. Without loss of generality, denote \( \rho_{1A}^p = \rho_{2A}^p = 0 \).

Consider two lottery outcomes at school A, \( r_A \equiv (r_{1A}, r_{2A}, r_{3A}, r_{4A}) \) with \( r_{1A} < r_{2A} < r_{3A}, r_{4A} \) and \( r_{2A} \leq 0.5 \), and \( r'_A \equiv (r'_{1A}, r'_{2A}, r'_{3A}, r'_{4A}) \) with \( r'_{3A}, r'_{4A} < r'_{2A} < r'_{1A} \) and \( r'_{2A} > 0.5 \). Clearly \( Z_{1A}^{0.5}(r_A) = Z_{2A}^{0.5}(r_A) = 1 \) and \( Z_{1A}^{0.5}(r'_A) = Z_{2A}^{0.5}(r'_A) = 0 \). On the other hand, under any \( g_{DA}, D_{1A}(r_A) = 1, D_{1A}(r'_A) = 0, D_{2A}(r_A) = 0, \) and \( D_{2A}(r'_A) = 1 \). This violates the monotonicity requirement for \( Z_{1A}^{0.5} \) as an instrument for \( D_{1A} \): Endogenous variables \( D_{1A} \) and \( D_{2A} \) move in the opposite directions in response to the same change in
the IV from $Z^{0.5}_{iA} = 1$ to $Z^{0.5}_{iA} = 0$. Monotonicity is required by many modern IV models with heterogeneous behavior and treatment effects (Heckman and Vytlacil, 2007; Manski, 2008; Angrist and Pischke, 2009). Therefore, while the constant cutoff qualification IV always extracts a random assignment, it may not be able to identify a causal effect due to monotonicity violations. Furthermore, since $First_A(r) = \{1, 2\}$ for all $r$, this monotonicity violation persists even if I restrict the sample to $First_A(r_0)$.

Example 2 also shows that yet another potential modification of the qualification IV does not work either. For any positive integer $m$, define the **constant rank qualification IV** by $Z^{m-\text{th}}_{iA}(r) \equiv 1\{r_i \leq m-\text{th}(\{r_j | j \in I\})\}$ where $m-\text{th}(\cdot)$ is the $m$-th order statistic. The constant rank qualification IV extracts a random assignment since $P(Z^{m-\text{th}}_{iA}(R) = 1 | \rho^-_{iA} = \rho, \theta = \theta) = m/|I|$ is independent from $\theta$. However, $Z^{2nd}_{1A}(r_A) = Z^{2nd}_{2A}(r_A) = 1(= Z^{0.5}_{1A}(r_A) = Z^{0.5}_{2A}(r_A))$ and $Z^{2nd}_{1A}(r'_A) = Z^{2nd}_{2A}(r'_A) = 0(= Z^{0.5}_{1A}(r'_A) = Z^{0.5}_{2A}(r'_A))$. Thus, by the same reason for $Z^{0.5}_{1A}$, $Z^{2nd}_{iA}$ also violates monotonicity regardless of whether I restrict the sample to $First_A(r_0)$.

Finally, going back to the original first choice and qualification IV research designs, they might fail even if I modify them to the more refined version that conditions on sharing the same priority at every school. Consider the following modification of Example 1.

**Example 3.** There are applicants 1, 2, 3, and schools A and B with the following preferences and priorities:

\[
\begin{align*}
\succ_1 & : A, B, \emptyset \\
\succ_2 & : A, \emptyset \\
\succ_3 & : B, A, \emptyset \\
\rho_A & : 3, \{1, 2\} \\
\rho_B & : \{1, 2, 3\}.
\end{align*}
\]

The indifferences in the school priorities are broken by STB. The capacity of each school is 1. The treatment school is A.

The only difference from Example 1 is $\rho_B$, which is now indifferent among all students. In Example 3, students 1 and 2 share the same priority at both A and B. However, the research designs do not extract a random assignment for them under the DA or Charlotte

---

27 Since both $\rho^-_{iA} = \rho^+_{2A}$ and $\rho_{1A} = \rho_{2A}$, the counterexample works even if I use original priorities to define an alternative constant cutoff qualification IV as $Z^{0.5}_{iA}(r) \equiv 1\{\rho^-_{iA} + r_i \leq 0.5\}$. Also, note that 1 and 2 share the same priority at all schools. Thus the constant cutoff qualification IV research design may fail to satisfy monotonicity even if I modify it to the more refined version that conditions on having the same priority at all schools.

28 Other possible definitions are $Z^{m-\text{th}}_{iA}(r) \equiv 1\{\rho^-_{iA} + r_i \leq m-\text{th}(\{\rho^+_{jA} + r_j | j \in I, \rho^+_{jA} = \rho^-_{jA}\})\}$ or $1\{\rho^-_{iA} + r_i \leq m-\text{th}(\{\rho^+_{jA} + r_j | j \in I, \rho^+_{jA} = \rho^-_{jA}\})\}$. The discussion below applies to these alternative definitions too.

29 All of the above points in this section 2.4.4 apply to the top trading cycles mechanism.
mechanism for $A$: Under the DA mechanism,

$$First_A(r) = \begin{cases} \emptyset & \text{if } r = (2, 1, 3) \\ \{1, 2\} & \text{otherwise.} \end{cases}$$

Nevertheless, enumerating all lottery outcomes shows that

$$P(D_{iA}(R) = 1|\theta_i = \theta_1) = P(Z_{iA}(R) = 1|\theta_i = \theta_1) = 1/2$$
$$\neq 1/3 = P(D_{iA}(R) = 1|\theta_i = \theta_2) = P(Z_{iA}(R) = 1|\theta_i = \theta_2).$$

Thus the research designs do not extract a random assignment under the DA mechanism even if they condition on sharing the same priority at every school. For the first choice research design, this point remains true even if using the alternative random assignment criterion in section 2.4.2 since

$$P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_1) = 3/5 \neq 2/5 = P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_2).$$

### 2.4.5 Top Trading Cycles Mechanism

Some cities such as New Orleans and San Francisco have used a mechanism outside the above generalized DA class. This mechanism, the top trading cycles mechanism, has also been advocated by matching market design researchers as a Pareto efficient mechanism that is also strategy-proof for students (Abdulkadiroglu and Sönmez, 2003).

**Definition 10.** The top trading cycles (TTC) mechanism creates a matching via the following procedure. Take any assignment problem as given.

1. Draw $r$ according to its lottery regime (STB or MTB).
2. Define $s$’s ex post strict priority order $\succ_{rs}$ over students by $i \succ_{rs} i'$ if $\rho_{is} + r_{is} < \rho_{i's} + r_{i's}$. 
3. Given $\succ_I$ and $(\succ_{rs})_{s \in S}$, run the following top trading cycles algorithm (Shapley and Scarf, 1974).

---

$^30$The probability calculations come from the following table.

<table>
<thead>
<tr>
<th>Lottery</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1, 3, 2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2, 1, 3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2, 3, 1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3, 1, 2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3, 2, 1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
• Step $t \geq 1$: Each student $i$ points to her most preferred acceptable school (if any). Students who do not point at any school are assigned to $\emptyset$. Each school $s$ points to its most preferred acceptable student. As there are a finite number of schools and students, there exists at least one cycle, i.e., a sequence of distinct schools and students $(i_1, s_1, i_2, s_2, \ldots, i_L, s_L)$ such that student $i_1$ points at school $s_1$, school $s_1$ points to student $i_2$, student $i_2$ points to school $s_2$, ..., student $i_L$ points to school $s_L$, and, finally, school $s_L$ points to student $i_1$. Every student $i_l$ ($l = 1, \ldots, L$) in any cycle is assigned to the school she is pointing at. Any student who has been assigned a school seat or the outside option as well as any school $s$ which has been assigned students such that the number of them is equal to its capacity $c_s$ is removed. If no student remains, the algorithm terminates. Otherwise, it proceeds to the next step.

This algorithm terminates in a finite number of steps as at least one student is matched with a school (or $\emptyset$) at each step and there are only a finite number of students.

It is possible to apply the first choice research design to data from the TTC mechanism. However, the first choice research design turns out not to extract a random assignment under the TTC mechanism.\textsuperscript{31} In Example 1, the TTC mechanism always assigns 1 to A under all $r$ (regardless of whether the lottery structure is STB or MTB); no randomization occurs. It implies that $First_A(r) = \{1, 2\}$ for all $r$, but $P(D_{is}(R) = 1|\theta_i = \theta_1) = 1 \neq 0 = P(D_{is}(R) = 1|\theta_i = \theta_2)$. Thus the first choice research design does not extract a random assignment under the TTC mechanism.\textsuperscript{32} This implies that the qualification IV research design also fails in this example.\textsuperscript{33}

This breakdown of the first choice research design seems related to the fact that the TTC mechanism is not strategy-proof for schools. In Example 1, imagine A’s true preference is $3 \succ_A 2 \succ_A 1$ while B’s is $1 \succ_B 2 \succ_B 3$. Under these true preferences, A is matched with 1. If A misreports $2 \succ'_A 1, 3$, however, A is matched with 2, who is more preferred to 1 with respect to $\succ_A$. Therefore, the TTC mechanism is not strategy-proof for A in Example 1. This is a reconfirmation of the well-known fact that the TTC mechanism is not strategy-proof for schools, and provides yet another support for the necessity of strategy-proofness for schools for randomization under the first choice research design.

\textsuperscript{31}It is possible to show that the more refined version of the first choice design which conditions on having the same priority at every school extracts a random assignment under the TTC mechanism. This makes a contrast with the DA and Charlotte mechanisms, under which even the more refined first choice design does not extract a random assignment (recall section 2.4.4).

\textsuperscript{32}This point remains the same under the alternative definition of extracting a random assignment in section 2.4.2: $P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_1) = 1 \neq 0 = P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_2)$.

\textsuperscript{33}In fact, the qualification IV research design under the TTC mechanism does not extract a random assignment even in the case without priorities. See Appendix 2.6.3.
2.5 Future Directions

This paper provides a first step toward deciphering axiomatic structures hidden in empirical research design uses of market design. This opens the door to several open questions. For example, while my framework assumes the use of random lottery numbers for tie-breaking, some existing empirical studies use data with regression-discontinuity-style local tie breaking by admissions test scores. Though I would expect the point of the current paper to be valid even in regression discontinuity situations, it is open how to extend this paper’s framework and results to a regression discontinuity setting. Also, my results point to the importance of strategy-proofness for schools within the gDA mechanism family. It is thus theoretically important to characterize gDA mechanisms that are strategy-proof for schools. I leave these directions for future research.

\footnote{The regression discontinuity extension involves technical difficulties. The framework in this paper studies finite assignment problems, with only finitely many applicants. In a corresponding regression discontinuity framework, the empirical distribution of the running variable has to be discrete, making a departure from most econometric models of regression discontinuity designs assuming continuous running variables. Incorporating discrete running variables would need a less standard model like Lee and Card (2008), and would make identification issues delicate. Also, related to the problem of identification under a discrete running variable, it is not trivial how to define the meaning of “extracting a random assignment” in such a regression discontinuity setting since, unlike the lottery framework in this paper, I usually don’t know the population distribution of the running variable.}
**Notes:** I simulate the DA mechanism by drawing counterfactual lottery numbers one million times. This gives us an approximate assignment probability. For each school $s$ and each student $i$ in $\text{First}_s(r_0)$, I demean $i$'s assignment probability by subtracting the mean of assignment probabilities at $s$ across all students in $\text{First}_s(r_0)$. I plot the demeaned assignment probability across all students in $\bigcup_{s \in S} \text{First}_s(r_0)$. 
Figure 2.5.2

(a) Increasing Students and Number of Schools

(b) Increasing Students and School Size

Notes: In Panel 2.5.2b, for each value of the x axis, I create another expansion of Example 1 with 2x schools $A_1, ..., A_x, B_1, ..., B_x$ with one seat each, and 3x students such that there are x students with each of the following three preferences:

$\succ_1: A_1, B_1, A_2, B_2, ..., A_x, B_x, \emptyset$

$\succ_2: A_1, ..., A_x, \emptyset$

$\succ_3: B_1, A_1, B_2, A_2, ..., B_x, A_x, \emptyset$

$\rho_{A_1}, ..., \rho_{A_x}: \{\text{students with } \succ_3\}, \{\text{students with } \succ_1 \text{ or } \succ_2\}$

$\rho_{B_1}, ..., \rho_{B_x}: \{\text{students with } \succ_1\}, \{\text{students with } \succ_2 \text{ or } \succ_3\}$.

In Panel 2.5.2a, for each value of the x axis, I create an expansion of Example 1 with x seats at each of schools A and B, and 3x students such that there are x students of each type=1, 2, 3. For each scenario, I approximate the assignment probabilities by simulating the DA mechanism 100000 times.
2.6 Proofs

2.6.1 Proof of Theorem 1

Main Proof

Proof. The proofs consists of lemmas and is structured as in Figure 2.6.3. Suppose that the first choice research design does not extract a random assignment under \( \varphi \) for some school \( s \) at some assignment problem \( X \). Fix \( \varphi, s, \) and \( X \) throughout. For each \( r \), define students \( i_0(r) \) and \( i_1(r) \) by

- \( i_0(r), i_1(r) \in First_s(r) \)
- \( D_{i_0(r)s}(r) = 0 \) and \( D_{i_1(r)s}(r) = 1 \)
- \( r_{i_0(r)s} \leq r_{is} \) for all \( i \in First_s(r) \) with \( D_{is}(r) = 0 \)
- \( r_{i_1(r)s} \geq r_{is} \) for all \( i \in First_s(r) \) with \( D_{is}(r) = 1 \)

If there are no two students satisfying the conditions, let \( i_0(r) = i_1(r) = 0 \). With this convention, \( i_0(r) \) and \( i_1(r) \) are uniquely well-defined for all \( r \).\(^{35}\) This proof will use the following equivalent representation of \( \varphi \).

Definition 11. Algorithm WAIT&GO\((r)\) operates in the following way.

1. Same as in Definition 6.
2. Same as in Definition 6.
3. Run the following sub-algorithm WAIT\((r)\): Remove \( i_0(r) \) and \( i_1(r) \) from \( X \) and run the DA algorithm on the remaining subproblem where schools' strict priorities are given by \( (>_{s,p}) \).
4. Next run the following sub-algorithm GO\((r)\): Let \( i_0(r) \) and \( i_1(r) \) apply for \( s \) and the remaining steps operate in the same way as the DA algorithm where schools' strict priorities are given by \( (>_{s,p}) \).

By McVitie and Wilson (1970)'s order irrelevance result, WAIT&GO\((r)\) and \( \varphi(r) \) (the simplified notation for \( \varphi(\rho, r) \)) produce the same matching for all \( r \). Let \( t - 1 \) be the last step at which WAIT\((r)\) stops and \( \mu_{t-1}(r) \equiv (\mu_{st-1}(r))_{s \in S} \) be the tentative matching at the end of step \( t - 1 \). Let \( t \) be the first step of GO\((r)\). (Note that \( t \) implicitly depends on \( r \).)

For each \( r \), define \( \sigma^*(r) = (\sigma^*_s(r))_{s \in S} \) as the following permutation of \( r \). If \( i_0(r) = i_1(r) = 0 \), then \( \sigma^*(r) = r \). Otherwise, if MTB is used, \( \sigma^*(r) \) is obtained by switching only \( i_0(r) \) and \( i_1(r) \) only in \( r_s \), i.e.,

\(^{35}\)If there are two \( i_0(r) \neq i_0(r) \) satisfying the conditions, then \( r_{i_0(r)s} > r_{i_0(r)s} \) and \( r_{i_0(r)s} < r_{i_0(r)s} \), a contradiction. The same logic holds for \( i_1(r) \) too.
\[
\bullet \sigma^*_{i_0(r)s}(r) = r_{i_1(r)s}
\]
\[
\bullet \sigma^*_{i_1(r)s}(r) = r_{i_0(r)s}
\]
\[
\bullet \sigma^*_{i_i(r)s}(r) = r_{i_i(r)s}
\]
\[
\bullet \sigma^*_{i_s}(r) = r_{i_i(r) s} \text{ for all } i \neq i_0(r), i_1(r)
\]
\[
\bullet \sigma^*_{i_s}(r) = r_{s'} \text{ for all } s' \neq s
\]

If STB is used, \(\sigma^*(r)\) is obtained by switching \(i_0(r)\) and \(i_1(r)\) in \(r_{s'}\) for all \(s'\), i.e., for all \(s'\)

\[
\bullet \sigma^*_{i_0(r)s'}(r) = r_{i_1(r)s'}
\]
\[
\bullet \sigma^*_{i_1(r)s'}(r) = r_{i_0(r)s'}
\]
\[
\bullet \sigma^*_{i_s}(r) = r_{i_s'} \text{ for all } i \neq i_0(r), i_1(r)
\]

For each \(r\) and each permutation \(\tilde{\sigma}_s(r) \neq \sigma^*_s(r)\) of \(r_s\) that switches two \(i'\) and \(i''\) who are consecutive within \(\text{First}_s(r)\), let \(\tilde{\sigma}(r) = (\tilde{\sigma}_s(r), r_{-s})\). If MTB is used, \(\tilde{\sigma}(r) = \times_{|S|} \tilde{\sigma}_s(r)\).

**Lemma 6.** (Existence of a “symmetry breaking” in a lottery number permutation) There exits \(r\) such that \(D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0\) or \(D_{i_0}(\tilde{\sigma}(r)) \neq D_{i_0}(\tilde{\sigma}(r))\) for some \(i \in \text{First}_s(r)\).

**Proof of Lemma 6.** It is enough to show that if there is no \(r\) such that there exists \(i \in \text{First}_s(r)\) such that \(D_{i_0}(\tilde{\sigma}(r)) \neq D_{i_0}(\tilde{\sigma}(r))\), then there exits \(r\) such that \(D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0\). Suppose to the contrary that for all \(r\), it is not the case \(D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0\).

**Step 1.A.** For all \(r\) with \(i_0(r) \neq \emptyset\) and \(i_1(r) \neq \emptyset\), \(D_{i_0(r)s}(\sigma^*(r)) = 1, D_{i_1(r)s}(\sigma^*(r)) = 0\), and \(D_{i_k}(\sigma^*(r)) = D_{i_k}(r)\) for all \(i \in \text{First}_s(r)\) with \(i \neq i_0(r)\) and \(i \neq i_1(r)\).

\[36\] I say \(i'\) and \(i''\) are consecutive within \(\text{First}_s(r)\) if there is no \(i''' \in \text{First}_s(r)\) such that \(r_{i'} < r_{i''} < r_{i''} \) or \(r_{i'} > r_{i''} > r_{i''} \).

113
Proof of Step 1.A. By the above assumption, it is enough to show that for all such \( r \), \( D_{i_1(r)}(\sigma^*(r)) = 0 \) and \( D_{i_0}(\sigma^*(r)) = D_{i_1}(r) \) for all \( i \in \text{First}_s(r) \) with \( i \neq i_0(r) \) and \( i \neq i_1(r) \). Let me consider \( \text{WAIT}(r) \) and \( \text{WAIT}(\sigma^*(r)) \). Since everything except \( i_0(r) \) and \( i_1(r) \)'s lottery numbers is the same between \( r \) and \( \sigma^*(r) \), both \( \text{WAIT}(r) \) and \( \text{WAIT}(\sigma^*(r)) \) produce the same tentative assignment \( \mu_{s-1}(r) = \mu_{s-1}(\sigma^*(r)) \equiv \mu_{s-1} \) for \( s \). Now start \( \text{GO}(r) \) and \( \text{GO}(\sigma^*(r)) \). Under \( r \), McVitie and Wilson (1970)'s order irrelevance result implies \( s \) rejects \( i_0(r) \) and tentatively accepts \( i_1(r) \), which implies \( \beta_{i_0(r)} \sigma + \gamma_{i_0(r)} > c_s-th\{\beta_{i_1(r)} + \gamma_{i_1(r)}\} \) where \( c_s-th\{\cdot\} \) is the \( s \)-th order statistic in the input set. Under \( \sigma^*(r) \), by definition of \( \sigma^*(r) \), \( \rho_{i_0(r)}^* + \rho_{i_1(r)}^* \sigma + \gamma_{i_1(r)} > c_s-th\{\beta_{i_0(r)} + \gamma_{i_0(r)}\} \) resulting in \( s \)'s rejecting \( i_1(r) \). Since any rejected student will never be accepted in the DA algorithm, this implies \( D_{i_1^*(r)}(\sigma^*(r)) = 0 \) and \( D_{i_0}(\sigma^*(r)) = D_{i_1}(r) \) for all \( i \in \text{First}_s(r) \) with \( i \neq i_0(r) \) and \( i \neq i_1(r) \) by the following reason. Suppose not. There exists \( i \in \mu_{s-1} \cap \text{First}_s(r) \setminus \{i_0(r), i_1(r)\} \) for whom, without loss of generality, \( D_{i_0}(\sigma^*(r)) = 0 \) and \( D_{i_1}(r) = 1 \). This implies that \( D_{i_0}(\sigma^*(r)) = D_{i_1}(r) = 0 \) since \( \rho_{i_0}^* + \gamma_{i_0} < \rho_{i_0}^* \sigma + \gamma_{i_0} < \rho_{i_1}^* + \gamma_{i_1} \) (the first inequality is by definition of \( i_1(r) \)) and so \( \rho_{i_0}^* + \sigma_{i_1}^* < \rho_{i_0}^* \sigma + \sigma_{i_1}^*(r) < \rho_{i_1}^* + \sigma_{i_1}^*(r) \). This is a contradiction to the assumption made above that for all \( r \), it is not the case \( D_{i_0}(\sigma^*(r)) = D_{i_1}(r) = 0 \). 

Step 1.B. For all \( r \) with \( i_0(r) = i_1(r) = \emptyset \), \( \varphi(\sigma^*(r)) = \varphi(r) \). For each \( r \) and each permutation \( \hat{\sigma}(r) \neq \sigma^*(r) \) defined above, \( D_{i_0}(\hat{\sigma}(r)) = D_{i_1}(r) \) for all \( i \in \text{First}_s(r) \).

Proof of Step 1.B. For all \( r \) with \( i_0(r) = i_1(r) = \emptyset \), \( \sigma^*(r) = r \) and it is trivial that \( \varphi(\sigma^*(r)) = \varphi(r) \). The second part is by assumption.

Step 1.C. For each \( r \), define \( o_s(r) \equiv |\{i \in \text{First}_s(r) | D_{i_0}(r) = 1\}| \). For each \( r \) and each permutation \( \sigma_s(r) \) of \( r_s \) that permutes lottery numbers only among members of \( \text{First}_s(r) \), let \( \sigma(r) \) be the following. If MTB is used, \( \sigma(r) = (\sigma_s(r), r_{-s}) \). If STB is used, \( \sigma(r) = \times_{|s|} \sigma_s(r) \). Then the following is true for all \( r \).

1. \( \text{First}_s(\sigma(r)) = \text{First}_s(r) \).
2. \( o_s(\sigma(r)) = o_s(r) \).
3. For all \( i \) and \( i' \) in \( \text{First}_s(\sigma(r)) \), if \( D_{i_0}(\sigma(r)) > D_{i'(r)}(\sigma(r)) \), then \( \sigma_{i_0}(r) < \sigma_{i'(r)}(r) \).

Proof of Step 1.C. Since any permutation can be expressed as a composition of contrapositions (permutations switching consecutive two elements), I can express any \( \sigma \) as a composition of \( \sigma^* \) and \( \hat{\sigma} \)'s defined in Steps 1.A and 1.B, respectively. Steps 1.A and 1.B imply (1) and (2). (3) follows from the fact that for all \( r \) and all \( i, i' \in \text{First}_s(r) \), \( \rho_{i_0}^* = \rho_{i'(r)}^* \) and another well-known property of the DA algorithm that for applicants who rank \( s \) first and share \( \rho_{i_0}^* \), \( D_{i_0} \) is monotonically decreasing in \( r_{i_0} \). \qed
For each $r$, let $\sigma(r)$ be a permutation of $r$ defined in Step 1.C. Let $\mathcal{R}$ be the set of all possible values of $r$. Partition $\mathcal{R}$ into $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_N$ such that within each $\mathcal{R}_n$, for all $r, r' \in \mathcal{R}_n$, $r$ can be obtained from $r'$ by such a permutation $r' = \sigma(r)$. This partition is well-defined by Step 1.C.1: Since Step 1.C.1 guarantees $\text{First}_s(r) = \text{First}_s(r')$, $r' = \sigma(r)$ for such a permutation $\sigma$ if and only if $r = \sigma(r')$ for such a (possibly different) permutation $\sigma$. Let $r_n$ be a generic element of $\mathcal{R}_n$. Note that $\text{First}_s(r_n)$ and $\sigma(r_n)$ are the same for all $r_n \in \mathcal{R}_n$ by Step 1.C.1 and 1.C.2, respectively. Step 1.C guarantees that conditional on each $\mathcal{R}_n$, $D_{ia}(R)$ is independent from $i$'s type for all $i \in \text{First}_s(r_n)$. That is, for all $n, i \in \text{First}_s(r_n)$, and $\theta$,

$$P(D_{ia}(R) = 1 | R \in \mathcal{R}_n, \theta_i = \theta) = \frac{o_s(r_n)}{|\text{First}_s(r_n)|}.$$ 

Therefore, for all $j \in \text{First}_s(r_0)$,

$$P(D_{ia}(R) = 1 | R \in \mathcal{R}_n, \theta_i = \theta_j) = \begin{cases} \frac{o_s(r_n)}{|\text{First}_s(r_n)|} & \text{if } \rho_{js} = \rho_{is}^o \text{ for any } i \in \text{First}_s(r_n) \\ 1 & \text{if } \rho_{js} < \rho_{is}^o \text{ for any } i \in \text{First}_s(r_n) \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{o_s(r_n)}{|\text{First}_s(r_n)|} & \text{if } \text{First}_s(r_0) = \text{First}_s(r_n) \\ 1 & \text{if } \rho_{js} < \rho_{is}^o \text{ for any } i \in \text{First}_s(r_n) \\ 0 & \text{otherwise} \end{cases}$$

$$= p_n$$

where the second last equality holds because $\rho_{js}^o = \rho_{is}^o$ for any $i \in \text{First}_s(r_n)$ if and only if $\text{First}_s(r_0) = \text{First}_s(r_n)$. $p_n$ is independent from $\theta_j$. This implies that for all $j \in \text{First}_s(r_0)$

$$P(D_{ia}(R) = 1 | \theta_i = \theta_j) = \Sigma_{n=1}^N P(D_{ia}(R) = 1 | R \in \mathcal{R}_n, \theta_i = \theta_j) \times P(R \in \mathcal{R}_n | \theta_i = \theta_j)$$

(by the law of total probability)

$$= \Sigma_{n=1}^N p_n \times P(R \in \mathcal{R}_n),$$

which is independent from $\theta_j$.

37 Under the alternative definition of extracting a random assignment in section 2.4.2, this part simplifies to the following. No other part of the proof depends on the definition of extracting random assignment.

$$P(D_{ia}(R) = 1 | i \in \text{First}_s(R), \theta_i = \theta) = \Sigma_{n=1}^N P(D_{ia}(R) = 1 | R \in \mathcal{R}_n, i \in \text{First}_s(R), \theta_i = \theta) \times P(R \in \mathcal{R}_n | i \in \text{First}_s(R), \theta_i = \theta)$$
assignment, a contradiction. This completes the proof of Lemma 6.

**Case 1.** There exists $r$ such that $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$. For each $r$, define $r_s^*$ as the following permutation of $r_s$. If $i_0(r) = i_1(r) = \emptyset$, let $r_s^* = r_s$. Otherwise,

- $r_i^* = r_i$ for all $i$ with $\rho_s^l = \rho_{i_0(r)s}^l = \rho_{i_1(r)s}^l$
- $r_{i_1(r)s}^* = \max\{r_{i_1(r)s}| \rho_s^l = \rho_{i_0(r)s}^l = \rho_{i_1(r)s}^l \text{ and } D_{i_1(r)s}(r) = 1\}$
- $r_{i_0(r)s}^* > r_{i_1(r)s}^*$ and there is no such $i$ that $\rho_s^l = \rho_{i_0(r)s}^l = \rho_{i_1(r)s}^l$ and $r_{i_0(r)s}^* > r_{i_1(r)s}^*$
- $r_i^* > r_j^*$ if and only if $r_i > r_j$ for all $i, j$ such that $\rho_s^l = \rho_{i_0(r)s}^l = \rho_{i_1(r)s}^l$ and $i \neq i_0(r), i \neq i_1(r), j \neq i_0(r), j \neq i_1(r)$.

For each $r$ and each $s' \neq s$, define $\bar{\sigma}_{s'}(r)$ as the following permutation of $r_{s'}$. (Note that $\bar{\sigma}_{s'}(r)$ implicitly depends on whole $r$.) If $i_0(r) = i_1(r) = \emptyset$ or MTB is used, then $\bar{\sigma}_{s'}(r) = r_{s'}$. Otherwise, $\bar{\sigma}_{s'}(r)$ is obtained by moving $i_1(r)$ to right above $i_0(r)$, i.e.,

- $\bar{\sigma}_{i_1(r)s'}(r) = \max\{r_{i_1}| r_{i_0} < r_{i_0(r)s'}\}$
- $\bar{\sigma}_{i_0(r)s'}(r) > \bar{\sigma}_{i_1(r)s'}(r)$ if and only if $r_{i_0} > r_{i_1(r)s'}$ for all $i, j$ such that $i, j \neq i_1(r)$.

**Lemma 7.** (Partial equivalence between $r$ and $(r_s^*, \bar{\sigma}_{s}(r)))$ For all $r$, $D_{i_0(r)s}(r_s^*, \bar{\sigma}_{s}(r)) = 0(= D_{i_1(r)s}(r_s^*, \bar{\sigma}_{s}(r)))$ and $D_{i_1(r)s}(r_s^*, \bar{\sigma}_{s}(r)) = 1(= D_{i_0(r)s}(r_s^*, \bar{\sigma}_{s}(r)))$.

**Proof of Lemma 7.** The following Steps 2.A and 2.B imply that for all $r$, $\varphi(r) = \varphi(r_s^*, \bar{\sigma}_{s}(r))$, which in turn implies Lemma 7. 

**Step 2.A.** For all $r$, $\varphi(r) = \varphi(r_s^*, \bar{\sigma}_{s}(r))$.

**Proof of Step 2.A.** If $i_0(r) = i_1(r) = \emptyset$ and so $r_s^* = r_s$, the above equality is trivial. Otherwise, $r_s^*$ is obtained from $r_s$ through a composition of two permutations. The first one permutes lottery numbers only among students in some set $I_1$ such that $\rho_s^{l_i} = \rho_s^{l_{i'}}$ for all $i', i'' \in I_1$ and there exists $i$ with $D_{i_0(r)s}(r) = 1$ such that $\max_{i' \in I_1}\{\rho_s^{l_{i'}} + r_{i'}\} \leq \rho_s^{l_i} + r_i$. The second permutation permutes lottery numbers only among students in some set $I_0$ such that $\rho_s^{l_i} = \rho_s^{l_{i'}}$ for all $i', i'' \in I_0$ and $\min_{i' \in I_0}\{\rho_s^{l_{i'}} + r_{i'}\} > \rho_s^{l_i} + r_i$ for all $i$ with $D_{i_0(r)s}(r) = 1$. The first (or second) permutation is a composition of special permutations $\delta$ satisfying the conditions in the Auxiliary Result (a) (or (b), respectively). Therefore the Auxiliary Result implies Step 2.A.

\[
\sum_{n=1}^{N} \frac{o_s(r_n)}{|\text{First}_s(r_n)|} \times \frac{|\{r \in R_0| i \in \text{First}_s(r)\}|}{|\{r \in R| i \in \text{First}_s(r)\}|},
\]

which is independent from $\theta_1 = \emptyset$ conditional on $i \in \text{First}_s(r)$.
Step 2.B. For all \( r \), \( \varphi(r^*, r^*) = \varphi(r^*, \tilde{s}_s(r)) \).

Proof of Step 2.B. If \( i_0(r) = i_1(r) = \emptyset \) and so \( \tilde{s}_s(r) = r_{-s} \), the above inequality is trivial. Otherwise, at the first step of the DA algorithm, students apply for schools in the same way both under \( (r^*_s, r_{-s}) \) and \( (r^*_s, \tilde{s}_s(r)) \). In particular, \( i_1(r) \) applies for \( s \) since \( i_1(r) \in \text{First}_s(r) \). Schools also tentatively accept students in the same way both under \( (r^*_s, r_{-s}) \) and \( (r^*_s, \tilde{s}_s(r)) \). In particular, \( ii(r) \) applies for \( s \) since \( ii(r) \in \text{First}_s(r) \). Schools also tentatively accept students in the same way both under \( (r^*_s, r_{-s}) \) and \( (r^*_s, \tilde{s}_s(r)) \). In particular, \( ii(r) \) applies for \( s \) since \( ii(r) \in \text{First}_s(r) \). As a result, since Step 2.A implies \( D_{i_1(r)}(r^*_s, r_{-s}) = 1 \), \( s \) tentatively accepts \( i_1(r) \) at the first step of the DA algorithm both under \( (r^*_s, r_{-s}) \) and \( (r^*_s, \tilde{s}_s(r)) \). Since \( a) \) \( s \) has the same preference \( r^*_s \) both under \( (r^*_s, r_{-s}) \) and \( (r^*_s, \tilde{s}_s(r)) \), \( b) \) the only possible difference between \( r^*_s \) and \( r^*_s \) is the position of \( i_1(r) \), and \( c) \) \( i_1(r) \) is tentatively kept by \( s \) and will never be rejected by \( s \) under \( (r^*_s, r_{-s}) \), the DA algorithm operates in the same way for the remaining steps, producing the same matching. \( \square \)

Lemma 8. (Partial equivalence between \( \sigma^*(r) \) and \( (\sigma^*_s(r^*_s, r_{-s}), \tilde{s}_s(r)) \)) For all \( r \) with \( D_{i_0(r)}(\sigma^*(r)) = D_{i_1(r)}(\sigma^*_s(r^*_s, r_{-s}, \tilde{s}_s(r)) = 0 \), it is the case \( D_{i_0(r)}(\sigma^*_s(r^*_s, r_{-s}), \tilde{s}_s(r)) = D_{i_1(r)}(\sigma^*_s(r^*_s, r_{-s}, \tilde{s}_s(r)) = 0 \).

Proof of Lemma 8. If \( i_0(r) = i_1(r) = \emptyset \) and so \( \sigma^*(r) = r = (\sigma^*_s(r^*_s, r_{-s}), \tilde{s}_s(r)) \), Lemma 8 is immediate. Otherwise, I first prove the following result.

Step 3.A. For all \( r \) with \( D_{i_0(r)}(\sigma^*(r)) = D_{i_1(r)}(\sigma^*(r)) = 0 \), it is the case \( \varphi(\sigma^*(r)) = \varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r)) \).

Proof of Step 3.A. Note that

a) \( \sigma^*_{i_0}(r) = r^*_{i_1}(r) \leq \min\{r^*_0(r), r^*_s(r), r^*_1(r)\} = \min\{\sigma^*_1(r) = r^*_1(r) = \sigma^*_1(r), r^*_1(r) \}

b) \( \rho^*_s + \sigma^*_s(r) > \rho^*_s + \sigma^*_s(r) \) for all \( j \) with \( \rho^*_s = \rho^*_s(r) \) and \( \sigma^*_s(r) > \sigma^*_s(r) \) and all \( i \) with \( D_{i_0}(\sigma^*(r)) = 1 \) since \( D_{i_0}(\sigma^*(r)) = 0 \) and \( i_0(r) \in \text{First}_s(r) \) and so \( i_0(r) \) ranks \( s \) first.

(a) and (b) imply that starting from \( \sigma^*(r) \), \( \sigma^*_s(r^*_s, r_{-s}) \) is obtained from \( \sigma^*_s(r) \) through a permutation that permutes lottery numbers only among students in some set \( I_0 \) such that \( \rho^*_s = \rho^*_s \) for all \( i ', i'' \in I_0 \) and \( \min_{i \in I_0} \{\rho^*_s + \sigma^*_s(r) \} > \rho^*_s + \sigma^*_s(r) \) for all \( i \) with \( D_{i_0}(\sigma^*(r)) = 1 \). This permutation is a composition of special permutations \( \delta 's \) that satisfy the conditions in the Auxiliary Result (b). Therefore the Auxiliary Result implies Step 8.A. \( \square \)
I now compare $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ and $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$. At the first step of the DA algorithm, students apply for schools in the same way both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$. In particular, $i_0(r)$ and $i_1(r)$ apply for $s$ since $i_0(r), i_1(r) \in First_s(r)$. Schools also tentatively accept students in the same way both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$. The other schools also do so since the only possible differences between $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ and $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ are the positions of $i_0(r)$ and $i_1(r)$, both of whom apply for $s$. If $s$ rejects both $i_0(r)$ and $i_1(r)$ at the first step, the proof is complete. Otherwise, $s$ tentatively accepts at least $i_0(r)$ since $\rho_{i_0(r)}^s + \rho_{i_1(r)}^s(r^*_s, r_{-s}) < \rho_{i_1(r)}^s + \sigma^*_s(r^*_s, r_{-s})$.

Since $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ and $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ are equivalent over $I \setminus \{i_0(r)\}$, the remaining steps of the DA algorithm operate in the same way both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ until the point where $s$ rejects $i_0(r)$. $s$ finally rejects $i_0(r)$ since it does so under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ by Step 3.A and $s$ has the same preference $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r))$. This implies Lemma 8. 

\textbf{Lemma 9.} (Existence of a profitable preference manipulation) There exist $(\rho^*, r^*)$, $s$ responsive with respect to $(\rho^*, r^*)$, and $(\rho^*, r^*)$ such that $\varphi_s((\rho^*, r^*)$, $(\rho^*, r^*)) > \varphi_s(\rho^*, r^*)$.

\textbf{Proof of Lemma 9.} Lemmas 6, 7, and 8 imply that there exists $r$ such that

- $D_{i_0(r)}(r^*_s, \sigma^*_s(r)) = 0$
- $D_{i_1(r)}(r^*_s, \sigma^*_s(r)) = 1$
- $D_{i_0(r)}(s\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r)) = D_{i_1(r)}(s\sigma^*_s(r^*_s, r_{-s}), \sigma^*_s(r)) = 0$

\textbf{Step 9.A. } $\varphi_s(r^*_s, \sigma^*_s(r)) = \mu_{st}(r^*_s, \sigma^*_s(r))$.

\textbf{Proof of Step 9.A.} Execute $\text{WAIT}(r^*_s, \sigma^*_s(r))$ and now start $\text{GO}(r^*_s, \sigma^*_s(r))$. $s$ rejects $i_0(r)$ and tentatively accepts $i_1(r)$ since $D_{i_0(r)}(r^*_s, \sigma^*_s(r)) = 0$ and $D_{i_1(r)}(r^*_s, \sigma^*_s(r)) = 1$. Suppose to the contrary $\varphi_s(r^*_s, \sigma^*_s(r)) \neq \mu_{st}(r^*_s, \sigma^*_s(r))$. Since $|\varphi_s(r^*_s, \sigma^*_s(r))| = |\mu_{st}(r^*_s, \sigma^*_s(r))| = c_s$ (because $s$ rejects $i_0(r)$ when choosing $\mu_{st}(r^*_s, \sigma^*_s(r))$ at step $t$), this implies there exists a student $i_2 \in \varphi_s(r^*_s, \sigma^*_s(r))$ such that $i_2 \notin \mu_{st}(r^*_s, \sigma^*_s(r))$. In addition, $i_2 \notin \mu_{st-1}(r^*_s, \sigma^*_s(r)) \cup i_1(r)$ has to be the case since otherwise (i.e., if $i_2 \notin \mu_{st}(r^*_s, \sigma^*_s(r))$ but $i_2 \in \mu_{st-1}(r^*_s, \sigma^*_s(r)) \cup i_1(r)$ and so $i_2$ applies for $s$ at step $t' < t$ in $\text{WAIT}(r^*_s, \sigma^*_s(r))$), $s$ rejects $i_2$ at step $t$ and so $i_2 \notin \varphi_s(r^*_s, \sigma^*_s(r))$, a contradiction. This means $i_2$ applies for $s$ at some step $t' > t$ and is tentatively accepted by $s$. This requires that $s$ rejects $i_1(r)$ before or at step $t'$ since by definition $i \succ r^*_s r^*_s i_1(r)$ for any $i \in \mu_{st}(r^*_s, \sigma^*_s(r)) \setminus i_1(r)$, which is because $s$ rejects $i_0(r)$ at step $t$ and there is no such $i$ that $\rho^*_s = \rho^*_s r^*_s \rho^*_s r^*_s i_1(r)$. This is a contradiction to the above fact $D_{i_1(r)}(r^*_s, \sigma^*_s(r)) = 1$. 

118
Step 9.B. There exists a step $t' > t$ in $\text{WAIT}&\text{GO}(\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r))$ such that $\mu_{st'}(\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r)) = \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$ where $i_2$ is a student with $i_2 \succ^s_{r^*_s} i_1(r)$

Proof of Step 9.B. Execute $\text{WAIT}(\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r))$ and now start $\text{GO}(\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r))$. \(s\) rejects $i_1(r)$ and tentatively accept $i_0(r)$ at step $t$ since $\mu_{st-1}$ is the same between $(r^*_s, \bar{s}_{-s}(r))$ and $(\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r))$ as well as $\rho_{i_1(r)s} + \sigma^*_{i_1(r)s}(r^*_s, r_{-s}) = \rho_{i_0(r)s} + r^*_i(r)s$. By $D_{i_0(r)s}(\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r)) = 0$ shown above, \(s\) rejects $i_0(r)$ at some later step $t' > t$ and instead tentatively accepts $i_2$ with $i_2 \succ^s_{r^*_s} i_1(r)$ which implies $i_2 \succ^s_{r^*_s} i_1(r)$ since

\[\rho_{i_0(r)s} + \sigma^*_{i_0(r)s}(r^*_s, r_{-s}) = \rho_{i_1(r)s} + r^*_i(r)s\]

and $\sigma^*_{i_2s}(r^*_s, r_{-s}) = r^*_{i_2s}$. Therefore, $\mu_{st'}(\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r)) = \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$ where $i_2$ is a student with $i_2 \succ^s_{r^*_s} i_1(r)$.

I am now ready to construct a profitable preference manipulation for \(s\) at \(X\). Let $\succ^s$ be any preference for \(s\) that is responsive with respect to $(c_s, \rho_s, r^*_s)$. Let $\rho'_s$ be a coarse priority order for \(s\) such that $\rho'_k \succ \rho'_j$ for all $k \notin \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$ and $j \in \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$ while $\rho'_k = \rho'_j$ if and only if $p_{ks} = p_{js}$ for all $j, k \notin \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$ or $j, k \in \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$. By Steps 9.A and 9.B, $\varphi_s((\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r))) = \mu_{st}(r^*_s, \bar{s}_{-s}(r))$ while $\varphi_s((\rho'_s, \rho_s), (\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r))) = \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$.

Also, $i_2 \succ^s_{r^*_s} i_1(r)$ established in Step 9.B implies $i_2 \succ^s_{r^*_s} i_1(r)$ as follows:

\[
i_2 \succ^s_{r^*_s} i_1(r)
\]

\[
\iff f^o(\rho_{i_2s}) + g^o(\text{rank}_{i_2s}) + r^*_{i_2s} < f^o(\rho_{i_1s}) + g^o(\text{rank}_{i_1s}) + r^*_{i_1s}
\]

\[
\Rightarrow f^o(\rho_{i_2s}) + r^*_{i_2s} < f^o(\rho_{i_1s}) + r^*_{i_1s}
\]

(since $\text{rank}_{i_1s} = 1 < \text{rank}_{i_2s}$ and $g^o(\cdot)$ is weakly increasing)

\[
\Rightarrow \rho_{i_2s} + r^*_{i_2s} < \rho_{i_1s} + r^*_{i_1s}
\]

(since $f^o(\cdot)$ is weakly increasing)

\[
\Rightarrow i_2 \succ^s_{r^*_s} i_1(r)
\]

Thus $\varphi_s((\rho'_s, \rho_s), (\sigma^*_{\bar{s}}(r^*_s, r_{-s}), \bar{s}_{-s}(r))) = \mu_{st}(r^*_s, \bar{s}_{-s}(r)) \cup i_2 \setminus i_1(r)$ since $\succ^s$ is responsive with respect to $(c_s, \rho_s, r^*_s)$; showing $(\rho'_s, \sigma^*_{\bar{s}}(r^*_s, r_{-s}))$ is a profitable manipulation for $s$ with respect to any $\succ^s$ responsive with respect to $(c_s, \rho_s, r^*_s)$; therefore $\varphi$ is not strategy-proof for $s$ at $X$.

Case 2. There exist $r$ and $i \in \text{First}_s(r)$ such that $D_{i,s}(\hat{s}(r)) \neq D_{i,s}(r)$ where $\hat{s}(r)$ is a permutation defined right before Lemma 6.

Lemma 10. $D_{i',s}(r) = D_{i'',s}(r) = 0$, where $i'$ and $i''$ are the two students whose lottery numbers are permuted in $\hat{s}(r)$.

Proof of Lemma 10. It is enough to show that if $D_{i',s}(r) = D_{i'',s}(r) = 0$ does not hold, then
Figure 2.6.3: Schematic structure of the proof (Case 1)

The first choice research design does not extracts a random assignment for $s$ under $\varphi$

\[ \downarrow \]

\[ (\rho, r) \xrightarrow{\text{"Symmetry breaking" in a lottery number permutation}} (\rho, \sigma^*(r)) \]

Partially equivalent (Lemma 6)

Partially equivalent (Lemma 7)

\[ (\rho, (r^*_s, \tilde{\sigma}_s(r))) \xrightarrow{\text{Profitable preference manipulation}} ((\rho'_{s}, \rho_{-s}), (\sigma'^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))) \]

\[ \downarrow \]

$\varphi$ is not strategy-proof for $s$

$D_{is}(\hat{\sigma}(r)) = D_{is}(r)$ for all $i \in First_s(r)$, a contradiction. For all $r$, the restriction on $\tilde{\sigma}_s(r)$ implies that $\rho'_{v's} = \rho'^{v}_{v's}$, both $i'$ and $i''$ rank $s$ first and $D_{iv's}(r) = D_{iv's}(r)$, which in turn imply that (a) there exists $i$ with $D_{is}(r) = 1$ such that $\max\{\rho'_{v's} + r_{v's}, \rho'^{v}_{v's} + r_{v's}\} \leq \rho'_{is} + r_{is}$ or (b) $\min\{\rho'_{v's} + r_{v's}, \rho'^{v}_{v's} + r_{v's}\} > \rho'_{is} + r_{is}$ for all $i$ with $D_{is}(r) = 1$. Thus, the Auxiliary Result implies $\varphi(\hat{\sigma}_s(r), r_{-s}) = \varphi(r)$ and so $D_{is}(\hat{\sigma}_s(r), r_{-s}) = D_{is}(r)$ for all $i \in First_s(r)$, completing the proof for the MTB case.

On the STB case, suppose to the contrary that $D_{iv's}(r) = D_{iv's}(r) = 1$. At the first step of the DA algorithm, students apply for schools in the same way both under $r$ and $\hat{\sigma}(r)$. In particular, $i'$ and $i''$ apply for $s$ since $i', i'' \in First_s(r)$. Schools also tentatively accept students in the same way both under $r$ and $\hat{\sigma}(r)$: The other schools than $s$ do so since the only possible differences between $\succ_r^{\tilde{\sigma}_s}$ and $\succ_{\hat{\sigma}(r)}^{\tilde{\sigma}_s}$ are the positions of $i'$ and $i''$, both of whom apply for $s$. $s$ accepts the same students including $i'$ and $i''$ since $D_{iv's}(r) = D_{iv's}(r) = 1$ and $\{\rho'_{v's} + r_{v's}\}_{j} | j$ applies for $s$ at the first step of the DA algorithm $\} = \{\rho'^{v}_{v's} + \tilde{\sigma}_js(r) | j \}$ applies for $s$ at the first step of the DA algorithm, which is because the same students apply for $s$ both under $r$ and $\hat{\sigma}(r)$, and $\rho'_{v's} + r_{v's} = \rho'^{v}_{v's} + \tilde{\sigma}_js(r)$ for all $j \neq i', i''$, $\rho'_{v's} + r_{v's} = \rho'^{v}_{v's} + \tilde{\sigma}_js(r)$, and $\rho'^{v}_{v's} + r_{v's} = \rho'^{v}_{v's} + \tilde{\sigma}_js(r)$. Since (a) the only possible differences between $\succ_r^{\tilde{\sigma}_s}$ and $\succ_{\hat{\sigma}(r)}^{\tilde{\sigma}_s}$ are the positions of $i'$ and $i''$, and (b) $i'$ and $i''$ are tentatively kept by $s$ and will never be rejected by $s$ under $r$, the DA algorithm operates in the same way for the remaining steps, producing the same matching and implying $D_{is}(\hat{\sigma}(r)) = D_{is}(r)$ for all $i \in First_s(r)$, a contradiction. \[\square\]
Case 2.a. There exist \( r \) and \( i \in \text{First}_s(r) \) such that \( D_{is}(\hat{\sigma}(r)) = 0 \neq 1 = D_{is}(r) \).

Let \( i^* \) be such that \( i^* \in \text{First}_s(r), D_{is}(r) = 1, \) and \( r_{i^*} \leq r_s \) for all \( i \in \text{First}_s(r) \) with \( D_{is}(r) = 1 \). Note that \( D_{i^*s}(\hat{\sigma}(r)) = 0 \neq 1 = D_{is}(r) \) since \( D_{is}(\hat{\sigma}(r)) = 0 \) and \( \rho_{i^*s}^\circ + \hat{\sigma}_{i^*s}(r) = \rho_{is}^\circ + \hat{\sigma}_{is}(r) = \rho_{is}^\circ + \hat{\sigma}_{is}(r) \). Without loss of generality, assume \( r_{i^*s} < r_{is} \) and so \( \hat{\sigma}_{i^*s}(r) < \hat{\sigma}_{is}(r) \). Let \( \hat{\sigma}^\#(r) \) be a further permutation of \( \hat{\sigma}(r) \) such that \( \hat{\sigma}^\#_{i^*s}(r) = \min\{\hat{\sigma}_{is}(r) | i \in \text{First}_s(r), D_{is}(r) = 0\} \), \( \hat{\sigma}^\#_{is}(r) = 2\text{nd-min}\{\hat{\sigma}_{is}(r) | i \in \text{First}_s(r), D_{is}(r) = 0\} \), where 2nd-min\{\} is the second minimum element, and \( \hat{\sigma}^\#_{j^*k^*}(r) > \hat{\sigma}^\#_{i^*s}(r) \) if and only if \( \hat{\sigma}_{j^*k^*}(r) > \hat{\sigma}_{is}(r) \) for all \( j, k \neq i^*, i'' \).

**Lemma 11.** \( D_{i^*s}(\hat{\sigma}^\#_{i''s}(r), \hat{\sigma}_{i''s}(r)) = D_{i^*s}(\hat{\sigma}^\#_{i's}(r), \hat{\sigma}_{i's}(r)) = 0 \)

*Proof of Lemma 11.* \( i^*, i'' \in \text{First}_s(r) \) implies \( \rho_{i''s}^\circ = \rho_{is}^\circ \). Also, by \( D_{i^*s}(\hat{\sigma}(r)) = 0 \neq 1 = D_{is}(r) \), Lemma 10, and \( D_{is}(\hat{\sigma}(r)) = 0 \neq 1 = D_{is}(r) \), it is the case min\{\} \( \rho_{i''s}^\circ + \hat{\sigma}_{i''s}(r), \rho_{is}^\circ + \hat{\sigma}_{is}(r) \} = \min\{\rho_{i''s}^\circ + \rho_{is}^\circ, \rho_{is}^\circ + \rho_{is}^\circ \} > \rho_{is}^\circ + \rho_{is}^\circ + \hat{\sigma}_{is}(r) \) for all \( i \) with \( D_{is}(\hat{\sigma}(r)) = 1 \). Thus \( D_{i^*s}(\hat{\sigma}(r)) = 0 \). By the Auxiliary Result (b) and the definition of \( \hat{\sigma}^\#(r), \varphi(\hat{\sigma}^\#(r), \hat{\sigma}_{i's}(r)) = \varphi(\hat{\sigma}(r)) \), which implies Lemma 11.

Until the end of Case 2.a, for all \( i \) with \( i \in \text{First}_s(r) \) and \( D_{is}(r) = 1 \), change \( i \)'s preference \( \succ_i \) to \( \succ_i' \) such that \( s \succ_i' \emptyset \succ_i' s' \) for all \( s' \neq s \). This does not change anything in the above analysis. Let \( \hat{\sigma}^{**}(r) \) be the permutation of \( \hat{\sigma}^\#(r) \) that switches \( i^* \) and \( i'' \), who are consecutive within \( \text{First}_s(r) \) under \( \hat{\sigma}^\#(r) \).

**Lemma 12.** \( D_{i^*s}(\hat{\sigma}^{**}(r), \hat{\sigma}_{i's}(r)) = 0 \) and \( D_{i^*s}(\hat{\sigma}^{**}(r), \hat{\sigma}_{i''s}(r)) = 1 \).

*Proof of Lemma 12.* By the Auxiliary Result (b) and the definition of \( \hat{\sigma}^\#(r), \varphi \) produces the same matching both under \( r \) and \( (\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) \). Since the only differences between \( (\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) \) and \( (\hat{\sigma}^{**}(r), \hat{\sigma}_{-s}(r)) \) are the positions of \( i^* \) and \( i'' \) in the priority order of \( s \) and the positions of \( i' \) and \( i'' \) in the priority order of \( s' \neq s \), both under \( (\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) \) and \( (\hat{\sigma}^{**}(r), \hat{\sigma}_{-s}(r)) \), the DA algorithm operates in the same way until \( i'' \) is rejected by \( s \), which happens for sure since \( D_{i^*s}(\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) = 0 \) (by Lemma 11) and \( \rho_{i''s}^\circ + \hat{\sigma}_{i''s}(r) > \rho_{is}^\circ + \hat{\sigma}_{is}(r) \). Since \( i'' \) has a weakly worse lottery number under \( \hat{\sigma}_{i's}(r) \) than under \( r_{i''} \) for all \( s' \neq s, i \) is less likely to crowd other applicants from other schools than \( s \) and the chain reactions of new rejections and applications caused by \( s \)'s rejection of \( i'' \) are less likely to go back to \( s \) under \( (\hat{\sigma}^{**}(r), \hat{\sigma}_{-s}(r)) \) than under \( (\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) \). Since the only other difference between \( (\hat{\sigma}^{**}(r), \hat{\sigma}_{-s}(r)) \) and \( (\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) \) is the school-s lottery numbers of \( i^* \) and \( i'' \), i.e., \( \hat{\sigma}^\#_{i's}(r) = \hat{\sigma}^{**}_{i's}(r) \neq \hat{\sigma}^{**}_{i''s}(r) = \hat{\sigma}^\#_{i''s}(r) \), and \( \rho_{i''s}^\circ = \rho_{i's}^\circ \), whenever \( i'' \) is rejected by \( s \) under \( (\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) \), \( i^* \) is rejected by \( s \) under \( (\hat{\sigma}^{**}(r), \hat{\sigma}_{-s}(r)) \). But \( i^* \) ranks only \( s \) in \( \succ_i \) and causes no additional rejections at other schools while \( i'' \) may rank other schools than \( s \) and may cause additional rejections at other schools. By these two factors, the set of applicants for \( s \) is weakly larger in the set inclusion sense under \( (\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) \) than under \( (\hat{\sigma}^{**}(r), \hat{\sigma}_{-s}(r)) \), i.e., \( \{ j | D_{js'}(\hat{\sigma}^{**}(r), \hat{\sigma}_{-s}(r)) = 0 \ \text{for all} \ s' \succ_j s \} \subset \{ j | D_{js'}(\hat{\sigma}^\#(r), \hat{\sigma}_{-s}(r)) = 0 \ \text{for all} \ s' \succ_j s \} \).
As a result it has to be the case that the cutoff at \( s \) is smaller (more strict) under \( (\hat{s}^\#(r), r_{-s}) \) than under \( (\hat{s}^{##}(r), \hat{s}_{-s}(r)) \) and \( \max\{\rho_{js}^s + \hat{s}^{##}(r) | D_{js}(\hat{s}^{##}(r), \hat{s}_{-s}(r)) = 1\} \geq \max\{\rho_{js}^s + \hat{s}^{##}(r) | D_{js}(\hat{s}^{##}(r), r_{-s}) = 1\} \geq \rho_{js}^s + r_{js} = \rho_{js}^s + \hat{s}^{##}(r) \), implying \( D_{js}(\hat{s}^{##}(r), \hat{s}_{-s}(r)) = 1 \) (by \( i'' \in First_s(r) \)). This also implies \( D_{js}(\hat{s}^{##}(r), \hat{s}_{-s}(r)) = 1 \) in Lemma 11.

The remaining part of the proof of Case 2.a is omitted since it is almost the same as Case 1 except \( i'' \) and \( i^* \) perform the roles of \( i_1(r) \) and \( i_0(r) \), respectively.

**Case 2.b.** There exist \( r \) and \( i \in First_s(r) \) such that \( D_{is}(\hat{s}(r)) = 1 \neq 0 = D_{is}(r) \).

The proof is almost the same as Case 2.a and is omitted.

**Auxiliary Result**

**Auxiliary Result.** The following holds for any gDA \( \varphi \) and any assignment problem.

a) For each \( r \), let \( \delta_s(r) \) be any permutation of \( r_s \) that switches only two students \( i' \) and \( i'' \) such that \( \rho_{js}^s = \rho_{is}^s \) and there exists \( i \) with \( D_{is}(r) = 1 \) such that \( \max\{\rho_{js}^s + r_{js}, \rho_{is}^s + r_{is}\} \leq \rho_{is}^s + r_{is} \). Then \( \varphi(r) = \varphi(\delta_s(r), r_{-s}) \), where \( \varphi(r) \) is a shorthand for \( \varphi(\rho, r) \).

b) For each \( r \), let \( \delta_s(r) \) be any permutation of \( r_s \) that switches only \( i' \) and \( i'' \) such that \( \rho_{js}^s = \rho_{is}^s \) and \( \min\{\rho_{js}^s + r_{js}, \rho_{is}^s + r_{is}\} > \rho_{is}^s + r_{is} \) for all \( i \) with \( D_{is}(r) = 1 \). Then \( \varphi(r) = \varphi(\delta_s(r), r_{-s}) \).

**Proof of the Auxiliary Result.** a) Under \( r \) or \( (\delta_s(r), r_{-s}) \), let \( t_0 \) be the step in the DA algorithm at which either \( i' \) or \( i'' \) or both first apply for \( s \). If there is no such a step \( t_0 \), then the DA algorithm works in the same way until its end both under \( r \) and \( (\delta_s(r), r_{-s}) \), completing the proof. Assume the existence of such a step \( t_0 \). Until step \( t_0 - 1 \), the DA algorithm operates in the same way both under \( r \) and \( (\delta_s(r), r_{-s}) \) since the only difference between the two situations is the positions of \( i' \) and \( i'' \) in \( r_s \). \( t_0 \) is thus common to \( r \) and \( (\delta_s(r), r_{-s}) \). Let \( I_{st_0} \) be the set of students who are kept by \( s \) from step \( t_0 - 1 \) or newly apply for \( s \) in step \( t_0 \). \( I_{st_0} \) is again the same between \( r \) and \( (\delta_s(r), r_{-s}) \). There are a few cases to consider.

**Case 1.** Both \( i' \) and \( i'' \) apply for \( s \) at step \( t_0 \). Under \( r \), \( s \) tentatively accepts both \( i' \) and \( i'' \) by the assumption that there exists \( i \) with \( D_{is}(r) = 1 \) such that \( \max\{\rho_{js}^s + r_{js}, \rho_{is}^s + r_{is}\} \leq \rho_{is}^s + r_{is} \). Under \( (\delta_s(r), r_{-s}) \), \( s \) again tentatively accepts both \( i' \) and \( i'' \). This is because \( \{\rho_{js}^s + r_{js}\}_{i \in I_{st_0}} = \{\rho_{is}^s + \delta_i(r_s)\}_{i \in I_{st_0}} \) (recall \( i', i'' \in I_{st_0} \) and \( I_{st_0} \) is the same between \( r \) and \( (\delta_s(r), r_{-s}) \) and the above fact that \( s \) tentatively accepts both \( i' \) and \( i'' \) under \( r \), which jointly imply that \( \max\{\rho_{js}^s + r_{js} + \delta_i(r_s), \rho_{is}^s + r_{is} + \delta_i(r_s)\} = \max\{\rho_{js}^s + r_{js}, \rho_{is}^s + r_{is}\} \leq c_{s-th}\{\rho_{js}^s + r_{js}\}_{i \in I_{st_0}} \) where \( c_{s-th} \) is the \( c_s \)-th order statistic. The DA algorithm also works in the same way for the remaining steps.

**Case 2.** Only one of \( i' \) or \( i'' \) applies for \( s \) at step \( t_0 \). Without loss of generality, suppose
only $i'$ applies for $s$ at step $t_0$. Under $r$, $s$ tentatively accepts $i'$ by the assumption that there exists $i$ with $D_{is}(r) = 1$ such that $\rho_{i's}^a + r_{i's} \leq \rho_{i's}^p + r_{i's}$.

**Case 2.a.** $r_{i's} > r_{i's'}$ (and so $\delta_{i's}(r) < \delta_{i's'}(r)$). Under $(\delta_{s}(r), r_{-s})$, $s$ also tentatively accepts $i'$ by the following reason. By the above fact that $s$ tentatively accepts $i'$ under $r$, it holds that $\rho_{i's}^a + r_{i's} \leq c_s-th(\{\rho_{i'a}^p + ri's\}_{i \in I_{st0}})$, implying $\rho_{i's}^a + \delta_{i's}(r) \leq c_s-th(\{\rho_{i'a}^p + \delta_{i's}(r)\}_{i \in I_{st0}})$.

**Case 2.b.** $r_{i's} < r_{i's'}$. Under $(\delta_{s}(r), r_{-s})$, $s$ also tentatively accepts $i'$ by the following reason. Suppose not. Then $c_s-th(\{\rho_{i'a}^p + ri's\}_{i \in I_{st0}}) \leq c_s-th(\{\rho_{i'a}^p + \delta_{i's}(r)\}_{i \in I_{st0}}) < \rho_{i's}^a + \delta_{i's}(r) = \rho_{i's'}^a + r_{i's'}$, where the last equality uses the assumption $\rho_{i's}^a = \rho_{i's'}^a$. Let's call $c_s-th(\{\rho_{i'a}^p + ri's\}_{i \in I_{st0}})$ the tentative cutoff for school $s$ at step $t_0$, which is common between $r$ and $(\delta_{s}(r), r_{-s})$, since $I_{st0}$ is the same between $r$ and $(\delta_{s}(r), r_{-s})$ and $r_{is} = \delta_{is}(r)$ for all $i \in I_{st} \setminus i'$. Since the tentative cutoff is monotonically decreasing in steps, the above inequality implies that for all $i$ with $D_{is}(r) = 1$, $\rho_{i'a}^p + ri's < \rho_{i's'}^a + r_{i's'}$, contradicting the assumption that there exists $i$ with $D_{is}(r) = 1$ such that $\rho_{i's}^a + r_{i's} \leq \rho_{i's}^a + r_{i's}$.

In all cases, the DA algorithm works in the same way at step $t_0$ under $r$ and $(\delta_{s}(r), r_{-s})$. The DA algorithm also works in the same way for the remaining steps by essentially the same reasoning as above Case 2.

b) The proof is similar to (a) and omitted. \(\square\)

### 2.6.2 Proof of Corollary 2.b

**Proof.** Consider a special case of the proof of Theorem 1 where I suppose that the first choice research design does not extract a random assignment under the DA mechanism with STB when there are no priorities, i.e., $\rho_{is} = \rho_{js}$ for all $i, j$, and $s$. By the STB lottery structure, $r_{is'} = r_{is''}$ for all $i, s'$, and $s''$. In this case, Case 2 never happen and only Case 1 is relevant. In Case 1, by the no-priority assumption, $r_{s}^* = \delta_{s'}(r)$ for all $s' \neq s$. This implies that under the preferences induced by $(\rho, (r_{s}^*, \delta_{-s}(r)))$, all schools share the same preference as $s$'s $r_{s}'$. The proof of Theorem 1 implies that when all the other schools than $s$ commonly report $(\rho_{s}, r_{s}^*)$, reporting $(\rho_{s}'$, $\sigma_{s}^*(r_{s}'$, $r_{-s}'))$ is a profitable preference manipulation for $s$ with respect to $r_{s}'$. This contradicts the fact that under the DA mechanism, truth-telling is optimal for $s$ when all the other schools report the same preference as $s$’s true preference. \(\square\)

Alternatively, it is possible to directly prove Corollary 2.b by the same logic as in the proof of Lemma 6.

---

\(38\)I believe this fact is folk knowledge. A formal proof is implied by Hatfield et al. (2016)'s Lemma 1 and Proposition 7.
2.6.3 Proof of Proposition 2

Proof. Consider applicants 1, 2, 3, 4, 5 and schools A, B, C with the following preferences and priorities:

\[
\succ_1: B, A, \emptyset \\
\succ_2: B, \emptyset \\
\succ_3: C, A, \emptyset \\
\succ_4, \succ_5: C, \emptyset \\
\rho_A, \rho_B, \rho_C : \{1, 2, 3, 4, 5\}
\]

The capacity of each school is 1. The treatment school is A. Note that this example involves no priorities.

Both students 1 and 3 rank A second and have the same priority at A. Thus, for all \(g_{DA}\), \(\rho_{iA}^e \equiv f^e(\rho_{1A}) + g^e(rank_{iA}) = f^e(\rho_{3A}) + g^e(rank_{3A}) \equiv \rho_{3A}^e\), which I denote by \(\rho\). Nevertheless, enumerating all lottery outcomes shows that under any \(g_{DA}\) (A computer program to implement this enumeration is available upon request),

\[
P(Z_{iA}(R) = 1|\rho_{iA}^e = \rho, \theta_i = \theta_1) = 2/3 \neq 5/6 = P(Z_{iA}(R) = 1|\rho_{iA}^e = \rho, \theta_i = \theta_3) \text{ under STB}
\]

\[
P(Z_{iA}(R) = 1|\rho_{iA}^e = \rho, \theta_i = \theta_1) = 2/3 \neq 3/4 = P(Z_{iA}(R) = 1|\rho_{iA}^e = \rho, \theta_i = \theta_3) \text{ under MTB}
\]

Therefore the qualification IV research design does not extract a random assignment under any \(g_{DA}\) even with no priorities.  

\(\square\)

The qualification IV research design fails in this example because 1 and 3 experience different levels of competition at B and C, respectively, before applying for A. Let us consider the following cases.

Case i. Neither 1 nor 3 applies for A, i.e., 1 and 3 are assigned B and C, respectively.

\(\rightarrow\) No student applies for A and A is undersubscribed. Both 1 and 3 are qualified at A.

Case ii. Only 1 applies for A.

\(\rightarrow\) 1 is always assigned A and qualified at A. 3 is qualified at A if 3 has a better lottery number than 1 at A.

\(^{39}\)Since both \(\rho_{1A}^e = \rho_{3A}^e\) and \(\rho_{1A} = \rho_{3A}\), the counterexample works even if I use original priorities to define an alternative qualification IV as \(Z''_{is}(r) \equiv 1\{\rho_{is} + \tau_{is} \leq \max\{\rho_{js} + \tau_{js}|D_{js}(r) = 1\}\}\). Also, note that 1 and 3 share the same priority at all schools. Thus the qualification IV research design can fail even if I modify it to the more refined version that conditions on having the same priority at all schools.
Case iii. Only 3 applies for A.
→ 3 is always assigned A and qualified at A. 1 is qualified at A if 1 has a better lottery number than 3 at A.

Case iv. Both 1 and 3 apply for A.
→ Only one of 1 and 3 with a better lottery number is assigned A and qualified at A.

I can ignore Case i since it does not cause any difference between 1’s and 3’s probabilities of qualification at A. Conditional on Case iv, 3 is more likely to be qualified since 3’s first choice (C) is more competitive than 1’s (B) and so 3 experiences less severe censoring of his lottery number before applying for A; consequently 3 is more likely to have a better lottery number than 1. Moreover, Case iii is more likely to happen than Case ii since 3’s first choice (C) is more competitive than 1’s (B) and so 3 is more easily rejected by the first choice and more likely to apply for A. In total, 3 is more likely to be qualified at A than 1 due to differential competitiveness at their first choice schools.

The above discussion illustrates a general point that students may have different qualification probabilities depending on which schools they rank higher than the treatment school. This does not matter for the first choice research design since in the first choice sample Firsts(r0), everybody ranks the same schools above the treatment school, that is, no school at all. In this sense there are more threats to the qualification IV design than to the first choice design.40

40Every point in this appendix 2.6.3 applies to the top trading cycles mechanism since the top trading cycles mechanism is equivalent to the DA mechanism in the no-priority example in the above proof.
Chapter 3


3.1 Introduction

Families in many large urban districts can now apply for seats at any public school in their district. The fact that some schools are more popular than others and the need to distinguish between students who have different priorities at a given school generates a matching problem. Introduced by Gale and Shapley (1962) and Shapley and Scarf (1974), matchmaking via market design allocates scarce resources, such as seats in public schools, in markets where prices cannot be called upon to perform this function. The market-design approach to school choice, pioneered by Abdulkadiroğlu and Sönmez (2003), is used in a long and growing list of public school districts in America, Europe, and Asia. Most of these cities match students to schools using a mechanism known as deferred acceptance (DA).

Two benefits of matchmaking schemes like DA are efficiency and fairness: the resulting match improves welfare and transparency relative to ad hoc alternatives, while lotteries ensure that students with the same preferences and priorities have the same chance of obtaining highly-sought-after seats. DA and related algorithms also have the virtue of narrowing the scope for strategic behavior that would otherwise give sophisticated families the opportunity to manipulate an assignment system at the expense of less-sophisticated participants (Abdulkadiroğlu et al., 2006; Pathak and Sönmez, 2008). No less important than these economic considerations is the fact that centralized assignment generates valuable data for empirical research on schools. In particular, when schools are oversubscribed, lottery-based rationing generates quasi-experimental variation in school assignment that can be used for credible evaluation of individual schools and of school reform models like charters.
Previous research exploiting the lotteries embedded in DA include studies of schools in Charlotte-Mecklenburg (Hastings et al., 2009; Deming, 2011; Deming et al., 2014) and New York (Bloom and Unterman, 2014; Abdulkadiroğlu et al., 2013). Causal effects in these studies are identified by compelling quasi-experimental variation, but the research designs deployed in this work fail to exploit the full power of the random assignment embedded in centralized assignment schemes. A major stumbling block in this context is the elaborate multi-stage nature of market-design matching. Market design weaves random assignment into an elaborate tapestry of information on student preferences and school priorities. In principle, all features of student preferences and school priorities can shape the probability of assignment to each school. Preferences and priorities are far from randomly assigned, of course. Families tend to prefer schools located in their neighborhoods, for example, while schools may grant priority to children poor enough to qualify for a subsidized lunch. It’s only conditional on preferences and priorities, therefore, that DA-generated assignments are independent of potential outcomes.

To eliminate the selection bias that arises from the dependence of assignments on preferences and priorities, research exploiting centralized assignment has focused either on offers of seats at students’ first choice schools, or relied on instrumental variables (IVs) indicating whether a student’s lottery number falls below the highest number offered a seat at all schools he’s ranked (we call this a qualification instrument). The first choice strategy conditions on the identity of the school ranked first, while qualification instruments condition on the set of schools ranked. These IV strategies are likely to produce estimates free of omitted variables bias. At the same time, both first-choice and qualification instruments discard much of the variation induced by DA.

This paper explains how to recover the full range of quasi-experimental variation embedded in centralized assignment. Specifically, we show how DA maps information on preferences, priorities, and school capacities into a conditional probability of random assignment, often referred to as the propensity score. As in other stratified randomized research designs, conditioning on the propensity score eliminates selection bias arising from the association between all conditioning variables and potential outcomes (Rosenbaum and Rubin, 1983). The payoff to propensity-score conditioning turns out to be substantial in our application: full stratification on preferences and priorities reduces degrees of freedom markedly, eliminating many schools and students from consideration, while score-based stratification leaves our research sample largely intact. The propensity score does more for us than reduce dimensionality, however. Because all applicants with score values strictly between zero and one contribute variation that can be used for evaluation, the propensity score identifies the maximal set of applicants for whom we have a randomized school-assignment experiment.

The propensity score generated by DA-type mechanisms does not have a general closed form solution. Our theoretical framework therefore revolves around an asymptotic “large market” approximation to the finite-market score. This DA propensity score is a function
of a few easily-computed sample statistics. We also construct propensity score estimates using simulation, that is, by drawing lottery numbers many times and computing the resulting average assignment rates across draws. Both the simulated and DA (analytic) propensity scores work well as far as covariate balance goes, but the approximate formula is, of course, much more quickly computed, and highlights specific sources of randomness and confounding in DA-based assignment schemes. In other words, the DA propensity score reveals the nature of the stratified experimental design embedded in a particular match.

Our test bed for the DA propensity score is an empirical analysis of charter school effects in the Denver Public School (DPS) district, a new and interesting setting for charter school impact evaluation. Because DPS assigns seats at traditional and charter schools in a unified match, the population attending DPS charters is less positively selected than in large urban districts with decentralized charter lotteries. This context makes DPS charter effects relevant for the ongoing debate over charter expansion. As far as we know, ours is the first charter evaluation to exploit an assignment scheme that simultaneously allocates seats in both the charter and traditional public school sectors.

The next section uses simple market design examples to explain the problem at hand. Following this explanation, Section 3.3 uses the theory of market design to characterize the propensity score for DA offers in large markets. Section 3.4 applies these results to estimate charter effects. Our empirical evaluation strategy uses an indicator for DA-generated charter offers as an instrument for charter school attendance in a two-stage least squares (2SLS) setup. This 2SLS procedure eliminates bias from non-random variation in preferences and priorities by controlling for the DA propensity score. This section also shows how to estimate effects for multiple sectors. Specifically, we look at DPS innovation schools, a popular alternative to the charter model. Finally, Section 3.5 summarizes our theoretical and empirical findings and outlines an agenda for further work. A theoretical appendix derives propensity scores for the Boston (Immediate Acceptance) mechanism and for DA with multiple tie-breaking.

---

1 Charter schools operate with considerably more independence than traditional public schools. Among other differences, many charters fit more instructional hours into a year by running longer school days and providing instruction on weekends and during the summer. Because few charter schools are unionized, they hire and fire teachers and administrative staff without regard to the collectively bargained seniority and tenure provisions that constrain such decisions in many public schools. About half of Denver charters implement versions of the No Excuses model of urban education. No Excuses charters run a long school day and year, emphasize discipline and comportment and traditional reading and math skills, and rely heavily on data and teacher feedback to improve instruction. For more background on the charter sector, see Abdulkadiroğlu et al. (2011) and Angrist et al. (2011).
3.2 Understanding the DA Propensity Score

We begin by reviewing the basic DA setup for school choice, showing how DA generates probabilities of school assignment that depend on preferences, priorities, and capacities.

A total of \( n \) students are to be assigned seats at schools of varying capacities. Students report their preferences by ranking schools on an application form or website, while schools rank students by placing them in priority groups. For example, a school may give the highest priority to students with already-enrolled siblings, second highest priority to those who live nearby, with the rest in a third priority group below these two. Each student is also randomly assigned a lottery number that distinguishes between those with the same preferences and priorities. DA assigns students to schools like this:

Each student applies to his most preferred school. Each school ranks all its applicants first by priority then by random number within priority groups and tentatively admits the highest-ranked applicants in this order up to its capacity. Other applicants are rejected.

Each rejected student applies to his next most preferred school. Each school ranks these new applicants together with applicants that it admitted tentatively in the previous round, first by priority and then by random number. From this pool, the school tentatively admits those it ranks highest up to capacity, rejecting the rest.

This algorithm terminates when there are no new applications (some students may remain unassigned).

DA produces a stable allocation in the following sense: any student who prefers another school to the one he has been assigned must be outranked at that school, either because everyone assigned there has higher priority, or because those who share the student’s priority at that school have higher lottery numbers. DA is also strategy-proof, meaning that families do as well as possible by submitting a truthful preference list (for example, there is nothing to be gained by ranking under-subscribed schools highly just because they are likely to yield seats). See Roth and Sotomayor (1990) for a review of these and related theoretical results.

3.2.1 Propensity Score Pooling

Within a given market structure (defined by the total number of applicants and the number of seats to be filled at each school), the probability that DA assigns student \( i \) a seat at school \( s \) depends on student \( i \)'s preferences and priority status at \( s \) and elsewhere. We refer to a student’s preferences and priorities as student type. For example, a student of one type might rank school \( b \) first, school \( a \) second, and have sibling priority at \( b \).
Suppose we’d like to estimate the causal effect of attending a particular school relative to other schools that students who rank this school might attend (our empirical work focuses on the causal effect of attendance at groups of schools, but the logic behind such comparisons is similar). Families exhibit a strong preference for schools in their neighborhood, so offer rates are high for nearby schools. At the same time, there are important differences in average achievement across Denver neighborhoods. Consequently, families who live in lower-income neighborhoods are more likely to be offered seats in schools attended by low achievers for reasons unrelated to the causal effects of school attendance. This is important in our context because there are more charters in poor areas. Likewise, selection bias can arise from failure to control for priorities: in some districts, for example, students poor enough to qualify for a subsidized lunch are granted priority for seats at schools of one sort or another.

DA treats students of the same type symmetrically in that everyone of a given type faces the same probability of assignment to each school. In other words, conditional on type, all that remains to determine school assignment is a random number that is independent of student characteristics and potential outcomes (this intuitive claim is proved in the next section). We can therefore eliminate selection bias in comparisons of those who are and aren’t offered seats at particular schools simply by conditioning on type. As a practical matter, however, we’d like to avoid full type conditioning, since this produces many small and even singleton or empty cells, reducing the sample available for impact analysis dramatically. The following simple example illustrates this point.

**Example 1.** Five students \{1, 2, 3, 4, 5\} apply to three schools \{a, b, c\}, each with one seat. Student 5 has the highest priority at c and student 2 has the highest priority at b, otherwise the students have the same priority at all schools. We’re interested in measuring the effect of an offer at school a. Student preferences are

\[
1 : a \succ b, \\
2 : a \succ b, \\
3 : a, \\
4 : c \succ a, \\
5 : c,
\]

where \(a \succ b\) means that a is preferred to b. Students 3 and 5 find only a single school acceptable.

Note that no two students here have the same preferences and priorities. Consequently, full-type stratification puts each student into a different stratum. This rules out research strategies that rely on full type conditioning to eliminate selection bias. But full type conditioning is unnecessary in this case because DA assigns students 1, 2, 3, and 4 to school a each with probability 0.25. This calculation reflects the fact that 5 beats 4 at c.
by virtue of his priority there, leaving 1, 2, 3, and 4 all applying to a in the second round and no one advantaged there. The impact of assignment to a can therefore be analyzed in a single stratum containing four students. As we show formally in Section 3, this stratification scheme is determined by the propensity score, the conditional probability of random assignment to a. Specifically, we can use a dummy indicating offers at a as an instrument for attendance at a in a sample that includes types 1-4: offers among these types occur with equal frequency and are therefore independent of potential outcomes and student characteristics. At the same time, the fact that offers of seats at school a boost enrollment there generates the relevant first stage.

3.2.2 Further Pooling in Large Markets

Under DA, the propensity score for assignment to school a is determined both by a student’s failure to win a seat at schools he ranks more highly than a and by the odds he wins a seat at a in competition with those who have also ranked a and similarly failed to find seats at schools they’ve ranked more highly. This two-part structure leads to a large-market approximation that generates pooling beyond that provided by the finite-market propensity score. We illustrate this point via a second simple example.

Example 2. Four students \{1, 2, 3, 4\} apply to three schools \{a, b, c\}, each with one seat. There are no school priorities and student preferences are

\[
1 : c, \\
2 : c \succ b \succ a, \\
3 : b \succ a, \\
4 : a.
\]

As in Example 1, each student is of a different type.

Let \(p_a(i)\) for \(i = 1, 2, 3, 4\) denote the probability that type \(i\) is assigned school a. With four students, \(p_a(i)\) comes from \(4! = 24\) possible lottery draws, all equally likely. Given this modest number of possibilities, \(p_a(i)\) is easily calculated by enumeration:

- Not having ranked a, type 1 is never assigned there, so \(p_a(1) = 0\).
- Type 2 is seated at a when schools he’s ranked ahead of a, schools b and c, are filled by others, and when he also beats type 4 in competition for a seat at a. This occurs for the two realizations of the form \((s, t, 2, 4)\) for \(s, t = 1, 3\). Therefore, \(p_a(2) = 2/24 = 1/12\).
- Type 3 is seated at a when the schools he’s ranked ahead of a—in this case, only b—are filled by others, while he also beats type 4 in competition for a seat at a. b can be filled by type 2 only when 2 loses to 1 in the lottery at c. Consequently, type
is seated at a only in a sequence of the form \((1, 2, 3, 4)\), which occurs only once. Therefore, \(p_a(3) = 1/24\).

- Finally, since type 4 gets the seat at a if and only if the seat does not go to type 2 or type 3, \(p_a(4) = 21/24\).

In this example, the propensity score differs for each student. But in larger markets with the same distribution of types, the score is smoother. To see this, consider a large market that replicates the structure of this example \(n\) times, so that \(n\) students of each type apply to 3 schools, each with \(n\) seats.\(^2\) With large \(n\), enumeration of assignment possibilities is a chore. We can, however, simulate the propensity score by repeatedly drawing lottery numbers.

The relationship between simulated probabilities of assignment and market size for Example 2, plotted in Figure 1, reveals that as the market grows, the distinction between types 2 and 3 disappears. In particular, Figure 1 shows that for large enough \(n\),

\[ p_a(2) = p_a(3) = 1/12; \quad p_a(1) = 0; \quad p_a(4) = 10/12 = 5/6, \]

with the probability of assignment at a for types 2 and 3 converging quickly. This convergence is a consequence of a result established in the next section, which shows that the large-market probabilities that types and 2 and 3 are seated at a are both determined by failure to win a seat at b. The fact that student 3 ranks c ahead of b is irrelevant.

A patient analyst can always approximate a finite market score by simulation, but our large-market results reveal why some schools and applicant types are subject to random assignment (even at schools that are under-subscribed), why applicants of different types share the same risk, and why for some applicants, assignment risk is degenerate (even at schools that are over-subscribed). A signal feature of the large market characterization is the role played by lottery qualification cutoffs at schools ranked ahead of school a in determining probabilities of assignment at a. This is illustrated by Example 2, which shows that, in the large-market limit, among schools that an applicant prefers to a, we need only be concerned with what happens at the school at which it’s easiest to qualify. In general, this most informative disqualification (MID) determines how distributions of lottery numbers for applicants of differing types are effectively truncated before entering the competition for seats at a. As we show below, the fact that the large market score depends on type only through a set of constructs like MID allows us to replace full type conditioning with something much smoother.

\(^2\)Many market-design analysts have found this sort of large-market approximation useful. Examples include Abdulkadiroğlu et al. (2015c); Azevedo and Leshno (2014); Budish (2011); Che and Kojima (2010); Kesten and Ünver (2015).
3.3 Score Theory

3.3.1 Setup

A general school choice problem, which we refer to as an economy, is defined by a set of students, schools, school capacities, student preferences over schools, and student priorities at schools. Let \( I \) denote a set of students, indexed by \( i \), and let \( s = 1, \ldots, S \) index schools. We consider markets with a finite number of schools, but with either finite \((n)\) or infinitely many students. As in Abdulkadiroğlu et al. (2015c) and Azevedo and Leshno (2014), the latter setting is referred to as a continuum economy. In a continuum economy, \( I = [0, 1] \) and school capacities are defined as the fraction of the continuum that can be seated at each school.

Student \( i \)'s preferences over schools constitute a partial ordering of schools, \( \succ_i \), where \( a \succ_i b \) means that \( i \) prefers school \( a \) to school \( b \). Each student is also granted a priority at every school. Let \( \rho_{is} \in \{1, \ldots, K, \infty\} \) denote student \( i \)'s priority at school \( s \), where \( \rho_{is} < \rho_{js} \) means school \( s \) prioritizes \( i \) over \( j \). For instance, \( \rho_{is} = 1 \) might encode the fact that student \( s \) has sibling priority at school \( s \), while \( \rho_{is} = 2 \) encodes neighborhood priority, and \( \rho_{is} = 3 \) for everyone else. We use \( \rho_{is} = \infty \) to indicate that \( i \) is ineligible for school \( s \). Many students share priorities at a given school, in which case \( \rho_{is} = \rho_{js} \) for some \( i \neq j \). Let \( \mathbf{p}_i = (\rho_{i1}, \ldots, \rho_{iS}) \) be the vector of student \( i \)'s priorities for each school. Student type is denoted by \( \theta_i = (\succ_i, \mathbf{p}_i) \). We say that a student of type \( \theta \) has preferences \( \succ_\theta \) and priorities \( \rho_\theta \). \( \Theta \) denotes the set of all possible types.

An economy is also characterized in part by a non-negative capacity vector, \( \mathbf{q} \), which is normalized by the total number of students, or by their measure when students are indexed continuously. In a finite economy, where the set \( I \) contains \( n \) students and each school \( s \) has \( k_s \) seats, capacity is defined by \( q_s = \frac{k_s}{n} \). In a continuum economy, \( q_s \) is the proportion of the set \( I \) that can be seated at school \( s \).

The analysis here is concerned with school assignment mechanisms that use lotteries to distinguish between students with the same preferences and priorities. Student \( i \)'s lottery number, \( r_i \), is the realization of a uniformly distributed random variable on \([0, 1]\), independent and identically distributed for all students. In particular, lottery draws are independent of type. In what follows, we consider a centralized assignment system relying on a single lottery number for each student. Extension to the less-common multiple tie-breaking case, in which a student may have different lottery numbers at different schools, is discussed in the theoretical appendix.

For any set of student types \( \Theta_0 \subset \Theta \) and for any number \( r_0 \in [0, 1] \), define the set of students in \( \Theta_0 \) with lottery number less than \( r_0 \) to be

\[
I(\Theta_0, r_0) = \{ i \in I \mid \theta_i \in \Theta_0, r_i \leq r_0 \}.
\]
We use the shorthand notation $I_0 = I(\Theta_0, r_0)$ for sets of applicants defined by type and lottery number. Also, when $r_0 = 1$, so that $I_0$ includes all lottery numbers, the second argument is omitted and $I_0 = \{ i \in I \mid \theta_i \in \Theta_0 \}$ for various choices of $\Theta_0$.

When discussing a continuum economy, we let $F(I_0)$ denote the fraction of students in $I_0$. Since lottery numbers are uniform and independent of type, this is given by

$$F(I_0) = E[1\{\theta_i \in \Theta_0\}] \times r_0,$$

where $E[1\{\theta_i \in \Theta_0\}]$ is the proportion of types in set $\Theta_0$. In a finite economy with $n$ students, the corresponding fraction is computed as

$$F(I_0) = \frac{|I_0|}{n}.$$

For a continuum economy, $F(I_0)$ is fixed, that is, non-stochastic. By contrast, $F(I_0)$ for a finite economy depends on the realized lottery draw. Either way, the student side of an economy is fully characterized by the distribution of types and lottery numbers, for which we sometimes use the shorthand notation, $F$. Note also that every finite economy has a continuum analog. This analog can be constructed by replicating the type distribution and the number of seats in the finite economy, while fixing the proportion of seats at school $s$ to be $q_s$.

**Defining DA**

We define DA using the notation outlined above, nesting the finite-market and continuum cases. First, combine priority status and lottery realization into a single number for each student and school, the student rank:

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

Since the difference between any two priorities is at least 1 and random numbers are between 0 and 1, student rank is lexicographic in priority and lottery numbers.

DA proceeds in a series of rounds. Denote the evolving vector of admissions cutoffs in round $t$ by $c^t = (c^t_1, \ldots, c^t_S)$. The demand for seats at school $s$ conditional on $c^t$ is defined as

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

In other words, school $s$ is demanded by students with rank below the school-$s$ cutoff, who prefer school $s$ to any other school for which they are also below the relevant cutoff.

The largest possible value of an eligible student’s rank is $K + 1$, so we can start with...
\[ c_s^1 = K + 1 \text{ for all } s. \] Cutoffs then evolve as follows:

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

where, because the argument for \( F \) can be written in the form \( \{i \in I_0 \mid \theta_i \in \Theta_0, r_i \leq r_0\} \), the expression is well-defined. This formalizes the idea that when the demand for seats at \( s \) falls below capacity at \( s \), the cutoff is \( K + 1 \). Otherwise, the cutoff at \( s \) is the largest value such that demand for seats at \( s \) is less than or equal to capacity at \( s \).

The final admissions cutoffs determined by DA for each school \( s \) are given by

\[ c_s = \lim_{t \to \infty} c_s^t. \]

The set of students that are assigned school \( s \) under DA is the demand for seats at the limiting cutoffs: \( \{i \in Q_s(c)\} \) where \( c = (c_1, \ldots, c_S) \). Since \( c_s \leq K + 1 \), an ineligible student is never assigned to school \( s \).

We write the final DA cutoffs as a limiting outcome to accommodate the continuum economy; in finite markets, DA converges in a finite number of rounds. Appendix 3.6.1 shows that this description of DA is valid in the sense that: (a) the necessary limits exist for every economy, finite or continuous; (b) for every finite economy, the allocation upon convergence matches that produced by DA as usually described (for example, by Gale and Shapley (1962) and the many market design studies building on their work).

**Defining the Propensity Score**

DA-generated offers depend on preferences and priorities as well as on lottery numbers. DA can therefore be seen as inducing a stratified randomized trial, where the “strata” are defined by type, \( \theta \). Because students of different types are likely to have different outcomes for reasons unrelated to their DA assignments, we’re interested in isolating the variation in offers determined by lottery numbers alone.

As in Rosenbaum and Rubin (1983)’s classic analysis of covariate conditioning, a key component of our effort to isolate random assignment within strata is the **propensity score**. The propensity score for a market of any size, denoted \( p_s(\theta) \), is the scalar function of type defined by

\[ p_s(\theta) = \Pr[D_i(s) = 1 | \theta_i = \theta], \]

where \( D_i(s) \) indicates whether student \( i \) is offered a seat at school \( s \). This function has domain given by the set of types who rank \( s \). We think of this as the group of applicants to \( s \); for the moment, the notation \( p_s(\theta) \) ignores the fact that the propensity score depends on market size.

Propensity score conditioning is motivated by a pair of conditional independence re-
results. We first have the fact that DA offers are randomly assigned conditional on student type. In other words, for any random variable $W_i$ that is independent of lottery numbers (this can be anything that is not a function of lottery numbers, including potential outcomes and student characteristics like free lunch status), the offer distribution satisfies

$$P[D_i(s) = 1 | W_i, \theta_i = \theta] = P[D_i(s) = 1 | \theta_i = \theta]. \quad (3.1)$$

Although unsurprising, this result provides a necessary foundation for everything that follows; Appendix 3.6.2 therefore presents a formal proof.

The examples in Section 3.2 show that full type conditioning, that is, conditioning on each value of $\theta$, reduces the sample available for impact evaluation and can eliminate schools and students from a causal analysis. It’s natural, therefore, to consider grouping and smoothing schemes that implicitly pool values of $\theta$. The propensity score theorem tells us how this pooling can be accomplished while still ensuring against omitted variables bias from any association between student type and potential outcomes. Given (3.1), it follows from Rosenbaum and Rubin’s (1983) propensity score theorem (and from our proof of (3.1)) that propensity score conditioning is enough to ensure that offers are independent of $W_i$. In other words,

$$P[D_i(s) = 1 | W_i, p_s(\theta_i) = p_s(\theta)] = P[D_i(s) = 1 | p_s(\theta_i) = p_s(\theta)] = p_s(\theta). \quad (3.2)$$

Equation (3.2) implies that propensity score conditioning eliminates the possibility of omitted variables bias due to the dependence of offers on type.\(^3\)

At first blush the conditional independence property described by Equation (3.2) might seem to be of little practical value: knowing nothing about the functional form of $p_s(\theta)$, we are left with as many possible score values as there are types. Our next step, therefore, is to derive an expression for $p_s(\theta)$ that exploits the structure of DA, showing how the score generated by DA indeed pools types.

### 3.3.2 Characterizing the DA Propensity Score

A key component in our characterization of $p_s(\theta)$ is the notion of a marginal priority group at school $s$. The marginal priority group consists of applicants for whom seats are allocated by lottery if the school is over-subscribed. Formally, the marginal priority, $\rho_s$, is the integer part of the cutoff, $c_s$. Conditional on being rejected by all more preferred schools and applying for school $s$, a student is assigned $s$ with certainty if his $\rho_i < \rho_s$, that is, if he clears marginal priority. Applicants with $\rho_i > \rho_s$ have no chance of finding

---

\(^3\)Rosenbaum and Rubin (1983) also show that the propensity score is the coarsest balancing score, which in this case means that no coarser function of type ensures conditional independence of $D_i(s)$ and $W_i$. Hahn (1998), Hirano et al. (2003), and Angrist and Hahn (2004) discuss the efficiency consequences of conditioning on the score.
a seat at $s$. Applicants for whom $\rho_{is} = \rho_s$ are marginal: these applicants are seated at $s$ when their lottery numbers fall below a school-specific lottery cutoff. The lottery cutoff at school $s$, denoted $\tau_s$, is the decimal part of the cutoff at $s$, that is, $\tau_s = c_s - \rho_s$.

These observations motivate a partition determined by marginal priorities at $s$. Let $\Theta_s$ denote the set of student types who rank $s$ and partition $\Theta_s$ according to

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

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

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

The set $\Theta_s^a$ contains applicant types who have worse-than-marginal priority at $s$. No one in this group is assigned to $s$. $\Theta_s^a$ contains applicant types that clear marginal priority at $s$. Some of these applicants may end up seated at a school they prefer to $s$, but they’re assigned $s$ for sure if they fail to find a seat at any school they’ve ranked more highly. Finally, $\Theta_s^c$ is the subset of $\Theta_s$ that is marginal at $s$. These applicants are assigned $s$ when they’re not assigned a higher choice and have a lottery number that clears the lottery cutoff at $s$.

A second key component of our score formulation reflects the fact that failure to qualify at schools other than $s$ may truncate the distribution of lottery numbers in the marginal priority group for $s$. To characterize the distribution of lottery numbers among those at risk of assignment at $s$, we first define the set of schools ranked above $s$. Specifically, applicants of type $\theta$ view the following set of schools as better than $s$:

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

An applicant’s most informative disqualification (MID) at $s$ is defined as a function of the cutoffs at schools in $B_{\theta s}$

$$MID_{\theta s} = \begin{cases} 
0 & \text{if } \rho_{\theta \tilde{s}} > \rho_s \text{ for all } \tilde{s} \in B_{\theta s}, \\
1 & \text{if } \rho_{\theta \tilde{s}} < \rho_s \text{ for some } \tilde{s} \in B_{\theta s}, \\
\max\{\tau_\tilde{s} \mid \tilde{s} \in B_{\theta s} \text{ and } \rho_{\theta \tilde{s}} = \rho_s\} & \text{if } \rho_{\theta \tilde{s}} = \rho_s \text{ for some } \tilde{s} \in B_{\theta s} \text{ and } \rho_{\theta \tilde{s}} > \rho_s \text{ otherwise}. 
\end{cases}$$

$MID_{\theta s}$ tells us how the lottery number distribution among applicants to $s$ is truncated by qualification at schools these applicants prefer to $s$. $MID_{\theta s}$ is zero when type $\theta$ students have worse-than-marginal priority at all higher ranked schools: when no $s$ applicants can be seated at a more preferred school, there’s no lottery number truncation among those at risk of assignment to $s$. On the other hand, when at least one school in $B_{\theta s}$ is undersubscribed, no one of type $\theta$ competes for a seat at $s$. Truncation is therefore complete, and $MID_{\theta s} = 1$.

The definition of $MID_{\theta s}$ also reflects the fact that, among applicants for whom $\rho_{\theta \tilde{s}} = \rho_s$ for some $\tilde{s} \in B_{\theta s}$, any student who fails to clear $\tau_\tilde{s}$ is surely disqualified at schools with
lower cutoffs. For example, applicants who fail to qualify at a school with a cutoff of 0.5 fail to qualify at schools with cutoffs below 0.5. Therefore, to keep track of the truncation induced by disqualification at all schools an applicant prefers to \( s \), we need to record only the most forgiving cutoff that an applicant fails to clear.

In finite markets, \( MID_\theta s \) varies from one lottery draw to another, but in a continuum economy, \( MID_\theta s \) is fixed. Consider the large-market analog of Example 2 in which \( n \) students of each of four types compete for the \( n \) seats at each of three schools. In this example, there’s a single priority group, so everyone is marginal. For large \( n \), we can think of realized lottery numbers as being distributed according to a continuous uniform distribution over [0, 1]. Types 2 and 3 rank different schools ahead of \( a \), i.e., \( B_{3a} = \{b\} \) while \( B_{2a} = \{b, c\} \). Nevertheless, because \( \tau_c = 0.5 < 0.75 = \tau_b \), we have that \( MID_{2a} = MID_{3a} = \tau_b = 0.75 \). To see where these cutoffs come from, note first that among the 2\( n \) type 1 and type 2 students who rank c first in this large market, those with lottery numbers lower (better) than 0.5 are assigned to \( c \) since it has a capacity of \( n \): \( \tau_c = 0.5 \). The remaining type 2 students (half of the original mass of type 2), all of whom have lottery numbers higher (worse) than 0.5, must compete with all type 3 students for seats at \( b \).

We therefore have 1.5\( n \) school-b hopefuls but only \( n \) seats at \( b \). All type 3 students with lottery numbers below 0.5 get seated at \( b \) (the type 2 students all have lottery numbers above 0.5), but this doesn’t fill \( b \). The remaining seats are therefore split equally between type 2 and 3 students in the upper half of the lottery distribution, implying that the highest lottery number seated at \( b \) is \( T_m = 0.75 \).

The following theorem uses the marginal priority and MID concepts to define an easily-computed DA propensity score that is a deterministic function of applicant type:

**Theorem 2.** Consider a continuum economy populated by applicants of type \( \theta \in \Theta \) to be assigned to schools indexed by \( s \in S \). For all \( s \) and \( \theta \) in this economy, we have:

\[
p_s(\theta) = \varphi_s(\theta) \equiv \begin{cases} 
0 & \text{if } \theta \in \Theta^n_s, \\
(1 - MID_{\theta s}) & \text{if } \theta \in \Theta^a_s, \\
(1 - MID_{\theta s}) \times \max \left\{ 0, \frac{\tau_s - MID_{\theta s}}{1 - MID_{\theta s}} \right\} & \text{if } \theta \in \Theta^c_s.
\end{cases}
\]  

(3.3)

where we also set \( \varphi_s(\theta) = 0 \) when \( MID_{\theta s} = 1 \) and \( \theta \in \Theta^c_s \).

The proof appears in Appendix 3.6.3.

The case without priorities offers a revealing simplification of this result. Without priorities, DA is the same as a random serial dictatorship (RSD), that is, a serial dictatorship with applicants ordered by lottery number (see, e.g., Abdulkadiroğlu and Sönmez 1998, Svensson 1999, Pathak and Sethuraman 2010).\(^4\) Theorem 2 therefore implies the following corollary, which gives the RSD propensity score:

\(^4\)Exam school seats are often assigned by a serial dictatorship based on admission test scores instead of random numbers (see, e.g., Abdulkadiroğlu et al. 2014b, Dobbie and Fryer 2014). A generalization
Corollary 3. Consider a continuum economy with no priorities populated by applicants of type \( \theta \in \Theta \), to be assigned to schools indexed by \( s \in S \). For all \( s \) and \( \theta \) in this economy, we have:

\[
\varphi_s(\theta) \equiv (1 - M ID_{\theta s}) \times \max \left\{ 0, \frac{\tau_s - M ID_{\theta s}}{1 - M ID_{\theta s}} \right\} = \max \{0, \tau_s - M ID_{\theta s}\}.
\]

Without priorities, \( \Theta^a_s \) and \( \Theta^b_s \) are empty. The probability of assignment at \( s \) is therefore determined solely by draws from the truncated distribution of lottery numbers remaining after eliminating applicants seated at schools they’ve ranked more highly. Applicants’ whose most informative disqualification exceeds the cutoff at school \( s \) cannot be seated at \( s \) because disqualification at a more preferred school implies disqualification at \( s \).

In a match with priorities, the DA propensity score also accounts for the fact that random assignment at \( s \) occurs partly as a consequence of not being seated a school preferred to \( s \). Applying these principles in the continuum allows us to describe the DA propensity score as follows:

i) Type \( \Theta^a_s \) applicants have a DA score of zero because these applicants have worse-than-marginal priority at \( s \).

ii) The probability of assignment at \( s \) is \( 1 - M ID_{\theta s} \) for applicants in \( \Theta^a_s \) because these applicants clear marginal priority at \( s \), but not at higher-ranked choices. Applicants who clear marginal priority at \( s \) are guaranteed a seat there if they don’t do better. Not doing better means failing to clear \( M ID_{\theta s} \), the most forgiving cutoff to which they’re exposed in the set of schools preferred to \( s \). Since lottery numbers are uniform, this happens occurs with probability \( 1 - M ID_{\theta s} \).

iii) Applicants in \( \Theta^c_s \) are marginal at \( s \) but fail to clear marginal priority at higher-ranked choices. For these applicants to be seated at \( s \) they must fail to be seated at a higher-ranked choice and win the competition for seats at \( s \). As for applicants in \( \Theta^a_s \), the proportion in \( \Theta^c_s \) left for consideration at \( s \) is \( 1 - M ID_{\theta s} \). Applicants in \( \Theta^c_s \) are marginal at \( s \), so their status at \( s \) is also determined by the lottery cutoff at \( s \). If the cutoff at \( s \), \( \tau_s \), falls below the truncation point, \( M ID_{\theta s} \), no one in this partition finds a seat at \( s \). On the other hand, when \( \tau_s \) exceeds \( M ID_{\theta s} \), seats are awarded by drawing from a continuous uniform distribution on \([M ID_{\theta s}, 1]\). The resulting assignment probability is therefore \((\tau_s - M ID_{\theta s})/(1 - M ID_{\theta s})\).

Applying Theorem 2 to the large-market version of Example 2 explains the convergence in type 2 and type 3 propensity scores seen in Figure 1. With no priorities, both types are in \( \Theta^a_s \). As we’ve seen, \( M ID_{2a} = M ID_{3a} = \tau_b = 0.75 \), that is, type 2 and 3 students of RSD, multi-category serial dictatorship, is used for Turkish college admissions (Balinski and Sönmez, 1999).
seated at $a$ must have lottery numbers above 0.75. It remains to compute the cutoff, $\tau_a$. Types 2 and 3 compete only with type 4 at $a$, and are surely beaten out there by type 4s with lottery numbers below 0.75. The remaining 0.25 seats are shared equally between types 2, 3, and 4, going to the best lottery numbers in $[0.75, 1]$, without regard to type. The lottery cutoff at $a$, $\tau_a$, is therefore $0.75 + 0.25/3 = \frac{5}{6}$. Plugging these into equation (3.3) gives the DA score for types 2 and 3:

$$\varphi_a(\theta) = (1 - MID_{\theta_a}) \times \max \left\{ 0, \frac{\tau_a - MID_{\theta_a}}{1 - MID_{\theta_a}} \right\}$$

$$= (1 - 0.75) \times \max \left\{ 0, \frac{5/6 - 0.75}{1 - 0.75} \right\}$$

$$= \frac{1}{12}.$$  

The score for type 4 is the remaining probability, $1 - (2 \times \frac{1}{12}) = \frac{5}{6}$.

The DA propensity score is a simple function of a small number of intermediate quantities, specifically, $MID_{\theta_s}$, $\tau_s$, and marginal priority status at $s$ and elsewhere. As in Example 2, it's common to find that different types have the same marginal priority status and $MID_{\theta_s}$, simplifying the score in a manner that facilitates empirical work. In stylized examples, we can easily compute continuum values for these parameters. In real markets with elaborate preferences and priorities, it's natural to use sample analogs for score estimation. As we show below, this generates a consistent estimator of the propensity score for finite markets.

### 3.3.3 Estimating the DA Propensity Score

We're interested in the limiting behavior of score estimates based on Theorem 2. The asymptotic sequence for our large-market analysis works as follows: randomly sample $n$ students and their lottery numbers from a continuum economy, described by type distribution $F$ and school capacities, $\{q_s\}$. Call the distribution of types and lottery numbers in this sample $F_n$. Fix the proportion of seats at school $s$ to be $q_s$ and run DA with these students and schools. Compute $MID_{\theta_s}$, $\tau_s$, and partition $\Theta_s$ by observing cutoffs $c_n$ and assignments in this single realization, then plug these quantities into equation (3.3). This generates an estimated propensity score, $\hat{p}_{ns}(\theta)$, constructed by treating a size-$n$ sample economy like its continuum analog. The actual propensity score for this finite economy, computed by repeatedly drawing lottery numbers for the sample of students described by $F_n$ and the set of schools with proportional capacities $\{q_s\}$, is denoted $p_{ns}(\theta)$. We consider the gap between $\hat{p}_{ns}(\theta)$ and $p_{ns}(\theta)$ as $n$ grows. The analysis here makes use of a regularity condition:
Assumption 1. (First choice support) For any \( s \in S \) and priority \( \rho \in \{1, \ldots, K\} \) with \( F(\{i \in I : \rho_{is} = \rho\}) > 0 \), we have \( F(\{i \in I : \rho_{is} = \rho, i \text{ ranks } s \text{ first}\}) > 0 \).

This says that in the continuum economy, every school is ranked first by at least some students in every priority group defined for that school.

In this setup, the propensity score estimated by applying Theorem 2 to data drawn from a single sample and lottery realization converges almost surely to the propensity score generated by repeatedly drawing lottery numbers. This result is presented as a theorem:

**Theorem 3.** In the asymptotic sequence described by \( F_n \) with proportional school capacities fixed at \( \{q_s\} \) and maintaining Assumption 1, the DA propensity score \( \hat{p}_{ns}(\theta) \) is a consistent estimator of \( p_{ns}(\theta) \) in the following sense: For all \( \theta \in \Theta \) and \( s \in S \),

\[
|\hat{p}_{ns}(\theta) - p_{ns}(\theta)| \xrightarrow{a.s.} 0.
\]

*Proof.* The proof uses intermediate results, given as lemmas in the theoretical appendix. The first lemma establishes that the vector of cutoffs computed for the sampled economy, \( \hat{c}_n \), converges to the vector of cutoffs in the continuum economy. That is,

\[
\hat{c}_n \xrightarrow{a.s.} c,
\]

where \( c \) denotes the continuum economy cutoffs. This result, together with the continuous mapping theorem, implies

\[
\hat{p}_{ns}(\theta) \xrightarrow{a.s.} \varphi_s(\theta).
\]

In other words, the propensity score estimated by applying Theorem 2 to a sampled finite economy converges to the DA propensity score for the corresponding continuum economy.

A second lemma establishes that for all \( \theta \in \Theta \) and \( s \in S \),

\[
p_{ns}(\theta) \xrightarrow{a.s.} \varphi_s(\theta).
\]

since \( \varphi_s \) is a continuous function of cutoffs. That is, the actual (re-randomization-based) propensity score in the sampled finite economy also converges to the propensity score in the continuum economy.\(^5\)

\(^5\)See also Azevedo and Leshno (2014), who provide convergence results for the cutoffs and conditional-on-type probabilities of assignment generated by a sequence of stable matchings, showing that the empirical assignment rates for types in a finite market converge to the continuum probability of assignment. The two lemmas in the appendix differ from Azevedo and Leshno (2014)'s results in that they use Assumption 1 and are proved using the extended continuous mapping theorem. The characterization of the DA propensity score in Theorem 2 does not appear to have an analog in the Azevedo and Leshno (2014) framework.
Combining these two results shows that for all $\theta \in \Theta$ and $s \in S$,

$$|\hat{p}_{ns}(\theta) - p_{ns}(\theta)| \xrightarrow{a.s.} |\varphi_s(\theta) - \varphi_s(\theta)| = 0,$$

completing the proof. Since both $\Theta$ and $S$ are finite, this also implies uniform convergence, i.e., $\sup_{\theta \in \Theta, s \in S} |\hat{p}_{ns}(\theta) - p_{ns}(\theta)| \xrightarrow{a.s.} 0$. \hfill $\square$

Theorem 3 justifies our use of the formula in Theorem 2 to control for student type in empirical work estimating school attendance effects. Specifically, the theorem explains why, as in Example 2, it may be enough to stratify on applicants’ most informative disqualification and marginal priority status instead of all possible values of $\theta$ when estimating the causal effects of school attendance. Not surprisingly, however, a number of implementation details associated with this strategy remain to be determined. These gaps are filled below.

### 3.3.4 Identification

Conditioning on estimates of the propensity score to control for type, we use DPS’s first-round charter offers to construct instrumental variables estimates of the effects of charter enrollment on achievement. How should the resulting IV estimates be interpreted? Our IV procedure identifies causal effects for applicants treated when DA produces a charter offer but not otherwise; in the local average treatment effects (LATE) framework of Imbens and Angrist (1994) and Angrist et al. (1996), these are charter-offer compliers. IV fails to reveal average causal effects for applicants who decline a first round DA charter offer and are assigned another type of school in round 2 (in the LATE framework, these are never-takers). Likewise, IV methods are not directly informative about the effects of charter enrollment on applicants not offered a charter seat in round 1, but who nevertheless find their way into a charter school in the second round (LATE always-takers).

To flesh out this interpretation and the assumptions on which it rests, let $C_i$ be a charter enrollment indicator and let $D_i$ indicate the offer of a charter seat. These variables indicate attendance and offers at any charter school, rather than at a specific school. Since DA produces a single offer, offers of seats at particular schools are mutually exclusive. We can therefore construct $D_i$ by summing individual charter offer dummies. Likewise, the propensity score for this variable, $p_D(\theta) \equiv E[D_i|\theta]$, is obtained by summing the scores for all charter schools to which $i$ has applied.

The population of charter-offer compliers (LATEs) is defined by potential treatment status. Potential treatment status (charter enrollment status) is indexed against the DA offer instrument, denoted $D_i$. In particular, we see potential treatment $C_{1i}$ when $D_i$ is switched on and potential treatment $C_{0i}$ otherwise (both of these are also assumed to
exist for all $i$). Observed treatment is therefore

$$C_i = C_{0i} + (C_{1i} - C_{0i})D_i.$$  

Compliers have $C_{1i} - C_{0i} = 1$, an event that happens when $C_{1i} = 1$ and $C_{0i} = 0$.

Causal effects are determined by potential outcomes, indexed against $C_i$. Initially, we allow for the fact that offers might have a direct effect on outcomes even knowing $C_i$. This possibility is expressed by writing potential outcomes as $Y_{1i}(d)$ and $Y_{0i}(d)$. This means that when $D_i = d$, we see $Y_{1i}(d)$ if $i$ is treated and we see $Y_{0i}(d)$ otherwise. All four of these potential outcomes are assumed to exist for all $i$.

Equation (3.1) implies that conditional on $\theta_i = \theta$, the offer variable, $D_i$, is independent of potential outcomes and assignments. In a manner analogous to the conditional independence of single-school offers described by equation (3.1), this can be expressed by writing:

$$\{Y_{1i}(1), Y_{1i}(0), Y_{0i}(1), Y_{0i}(0), C_{1i}, C_{0i}\} \perp D_i|\theta_i;$$  \hspace{1cm} (3.4)

where the vector $\{Y_{1i}(1), Y_{1i}(0), Y_{0i}(1), Y_{0i}(0), C_{1i}, C_{0i}\}$ plays the role of $W_i$. Likewise, as for single-school offers in equation (3.2), the propensity score theorem implies

$$\{Y_{1i}(1), Y_{1i}(0), Y_{0i}(1), Y_{0i}(0), C_{1i}, C_{0i}\} \perp D_i|p_D(\theta_i).$$  \hspace{1cm} (3.5)

The conditional independence conditions described by (3.4) and (3.5) allow us to estimate causal effects of charter offers, the treatment indicated by $D_i$. In practice, however, we’re interested in the effects of charter attendance, the treatment indicated by $C_i$.

Identification of average causal effects of charter school attendance requires an exclusion restriction. Specifically, we assume

$$Y_{ji}(1) = Y_{ji}(0) \equiv Y_{ji}; j = 0, 1.$$  

In other words, charter offers are assumed to be unrelated to outcomes for applicant $i$ once we know whether this applicant attended a charter school. The exclusion restriction allows us to replace the four double-index potential outcomes in (3.4) and (3.5) with two single-index potential outcomes, $Y_{1i}$ and $Y_{0i}$.

The case for the exclusion restriction is less immediate than that for conditional independence. We might worry, for example, that outcomes are affected by lottery numbers even for applicants whose charter status is unchanged by the lottery (that is, for always-takers and never-takers). Denver’s second round allocates any remaining school seats in an ad hoc school-by-school application process, unrelated to lottery numbers drawn in the first round. But lottery numbers can nevertheless affect the second round indirectly by changing opportunities. Consider, for example, a skittish charter applicant who chooses a popular non-charter option when $D_i = 0$ in round 1. Fearing the long charter school day
and having applied to charter schools only to satisfy his mother, this applicant also goes
non-charter if his \( D_i = 1 \). But in this case, having been offered a charter seat in round 1, he must settle for a less desirable and perhaps lower-quality non-charter option in round 2. This violates the exclusion restriction if \( Y_{oi}(1) \neq Y_{oi}(0) \). We must therefore either assume away within-sector differences in potential outcomes, or introduce a finer-grained parameterization of school sector effects. The latter approach is explored in Section 3.4.6, below.

In addition to the conditional independence and exclusion restrictions, we also assume that, conditional on the propensity score, charter offers cause charter enrollment for at least some students, and that charter offers can only make charter enrollment more likely, so that \( C_{1i} \geq C_{0i} \) for all \( i \). Given these assumptions, the conditional-on-score IV estimand is a conditional average causal affect for compliers, that is:

\[
E[Y_i|D_i = 1, p_{D}(\theta_i) = x] - E[Y_i|D_i = 0, p_{D}(\theta_i) = x] = E[Y_{1i} - Y_{0i}|p_{D}(\theta_i) = x, C_{1i} > C_{0i}],
\]

where \( p_{D}(\theta_i) \) is the charter-offer propensity score associated with applicant \( i \)'s type and \( x \) indexes values in the support of \( p_{D}(\theta) \).

### 3.3.5 Estimation

In view of the fact that (3.6) generates a distinct causal effect for each score value, it's natural to consider parsimonious models that use data from all propensity-score cells to estimate a single average causal effect. We accomplish this by estimating a 2SLS specification with first and second stage equations that can be written

\[
C_i = \sum_x \gamma(x) d_i(x) + \delta D_i + X_i^t \lambda + \nu_i, \tag{3.7}
\]

\[
Y_i = \sum_x \alpha(x) d_i(x) + \beta C_i + X_i^t \mu + \epsilon_i, \tag{3.8}
\]

where the \( d_i(x) \)'s are dummies indicating values of \( p_{D}(\theta_i) \), indexed by \( x \), and \( \gamma(x) \) and \( \alpha(x) \) are the associated “score effects” in the first and second stages. The coefficient \( \delta \) in (3.7) is the first-stage effect of charter offers on charter enrollment, while the coefficient \( \beta \) in (3.8) is the causal effect of interest. These first and second stage equations include baseline covariates, \( X_i \), to increase precision and adjust for any chance imbalances in student characteristics.

As a check on the 2SLS specification, we also report semiparametric estimates of

\[
E[Y_{1i} - Y_{0i}|C_{1i} > C_{0i}].
\]

In contrast with the additive 2SLS setup, the semiparametric procedure requires only correct specification of the propensity score to generate a single
average causal effect for all compliers. The semiparametric strategy is founded on Abadie (2003)'s observation that the conditional independence and exclusion restrictions imply:

\[
E[Y_0 | C_{1i} > C_{0i}] = \frac{1}{Pr(C_{1i} > C_{0i})} E \left[ \frac{C_i Y_i(D_i - p_D(\theta_i))}{(1 - p_D(\theta_i)) p_D(\theta_i)} \right],
\]

\[
E[Y_{1i} | C_{1i} > C_{0i}] = \frac{1}{Pr(C_{1i} > C_{0i})} E \left[ \frac{(1 - C_i) Y_i((1 - D_i) - (1 - p_D(\theta_i)))}{(1 - p_D(\theta_i)) p_D(\theta_i)} \right].
\]

Subtracting and rearranging, we have:

\[
E[Y_{1i} - Y_0 | C_{1i} > C_{0i}] = \frac{1}{Pr(C_{1i} > C_{0i})} E \left[ \frac{Y_i(D_i - p_D(\theta))}{(1 - p_D(\theta_i)) p_D(\theta_i)} \right]. \tag{3.9}
\]

The first stage in this case, \( P[C_{1i} > C_{0i}] \), is constructed using

\[
P[C_{1i} > C_{0i}] = E \left[ \frac{C_i(D_i - p_D(\theta_i))}{(1 - p_D(\theta_i)) p_D(\theta_i)} \right]. \tag{3.10}
\]

The semi-parametric IV estimator used here is the sample analog of the right hand side of (3.9) divided by the sample analog of (3.10).

### 3.4 School Effectiveness in Denver

Since the 2011 school year, DPS has used DA to assign students to most schools in the district, a process known as SchoolChoice. Denver school assignment involves two rounds, but only the first round uses DA. Our analysis therefore focuses on the initial round.

In the first round of SchoolChoice, parents can rank up to five schools of any type, including traditional public schools, magnet schools, innovation schools, and most charters. A neighborhood school is also ranked automatically (if a student has a neighborhood school, the district adds his neighborhood school to his choice list as the last choice). Schools ration seats using a mix of priorities and a single lottery number. Priorities vary across schools and typically involve siblings and neighborhoods. Seats may be reserved for a certain number of subsidized-lunch students and for children of school staff. Reserved seats are allocated by splitting schools and assigning the highest priority status to students in the reserved group at one of the sub-schools created by a split. Match participants can only qualify for seats in a single grade.

The DPS match distinguishes between groups of seats at a given school, known as “buckets.” Buckets in the same school have distinct priorities and capacities. DPS converts applicants' preferences over schools into preferences over buckets, splitting off separate

\footnote{Covariates are incorporated in the semiparametric estimation procedure by adding \( X_i \) to a logit model for \( E[D_i | X_i, p_D(\theta_i)] \) and using fitted values from this instead of estimates of \( p_D(\theta_i) \) in equations (3.9) and (3.10).}
sub-schools for each. The upshot for our purposes is that DPS’s version of DA assigns seats at the sub-schools determined by seat reservation policies and buckets rather than schools, while the relevant propensity score captures the probability of offers at sub-schools. The discussion that follows refers to propensity scores for schools, with the understanding that the fundamental unit of assignment is a bucket, from which assignment rates to schools have been constructed.\(^7\)

### 3.4.1 Computing the DA Propensity Score

The score estimates used as controls in equations (3.7) and (3.8) were constructed three ways. The first is a simulation-based benchmark: we ran DA for one million lottery draws and recorded the proportion of draws in which applicants of a given type in our fixed DPS sample were seated at each school.\(^8\) By a conventional law of large numbers, this simulated score converges to the actual finite-market score as the number of draws increases. In practice, of course, the number of replications is far smaller than the number of possible lottery draws, so the simulated score takes on more values than we’d expect to see for the actual score. For applicants with a simulated score strictly between zero and one, the simulated score takes on more than 1,100 distinct values (with fewer than 1,300 types in this sample). Because many simulated score values are exceedingly close to one another (or to 0 or 1) some of the estimators that control for the simulated score use values that have been rounded.

We’re particularly interested in taking advantage of the DA score defined in Theorem 2. This theoretical result is used for propensity score estimation in two ways. The first, which we label a “formula” calculation, applies equation (3.3) directly to the DPS data. Specifically, for each applicant type, school, and entry grade, we identified marginal priorities, and applicants were allocated by priority status to either \(6', 9',\) or \(W.\) The DA score, \(\varphi_s(\theta)\) is then estimated by computing the sample analog of \(MID_\theta\) and \(\tau_s\) in the DPS assignment data and plugging these into equation (3.3).

The bulk of our empirical work uses a second application of Theorem 2, which also starts with marginal priorities, MIDs, and cutoffs in the DPS data. This score estimate, however, is given by the empirical offer rate in cells defined by these variables. This score estimate, which we refer to as a “frequency” calculation, is closer to an estimated score of the sort discussed by Abadie and Imbens (2012) than is the formula score, which ignores realized assignment rates. The large-sample distribution theory in Abadie and Imbens (2012) suggests that conditioning on an estimated score based on realized assignment

\(^7\)DPS modifies traditional DA mechanism by recoding the lottery numbers of all siblings applying to the same school to be the best random number held by any of them. This modification (known as “family link”) changes the allocation of only about 0.6% of students from that generated by standard DA. Our analysis incorporates family link by defining distinct types for linked students.

\(^8\)Calsamiglia et al. (2014) and Agarwal and Somai (2015) simulate the Boston mechanism as part of an effort to estimate preferences in a structural model of latent preferences over schools.
rates may increase the efficiency of score-based estimates of average treatment effects.

Propensity scores for school offers tell us the number of applicants subject to random assignment at each DPS charter school. These counts, reported in columns 3-5 of Table 1 for the three different score estimators, range from none to over 300. The proportion of applicants subject to random assignment varies markedly from school to school. This can be seen by comparing the count of applicants subject to random assignment with the total applicant count in column 1. The randomized applicant count calculated using frequency and formula score estimates are close, but some differences emerge when a simulated score is used.

Column 5 of Table 1 also establishes the fact that at least some applicants were subject to random assignment at every charter except for the Denver Language School, which offered no seats. In other words, every school besides the Denver Language School had applicants with a simulated propensity score strictly in the unit interval. Three schools for which the simulated score shows very few randomized applicants (Pioneer, SOAR Oakland, Wyatt) have an empirical offer rate of zero, so the frequency version of the DA propensity score is zero for these schools (applicant counts based on intervals determined by DA frequency and formula scores appear in columns 3 and 4).

DA produces random assignment of seats for students ranking charters first for a much smaller set of schools. This can be seen in the last column of Table 1, which reports the number of applicants with a simulated score strictly between zero and one, who also ranked each school first. The reduced scope of first-choice randomization is important for our comparison of strategies using the DA propensity score with previously-employed IV strategies using first-choice instruments. First-choice instruments applied to the DPS charter sector necessarily ignore many schools. Note also that while some schools had only a handful of applicants subject to random assignment, over 1,400 students were randomized in the charter sector as a whole.

The number of applicants randomized at particular schools can be understood further using Theorem 2. Why did STRIVE Prep - GVR have 116 applicants randomized,

---

9The data analyzed here come from files containing the information used for first-round assignment of students applying in the 2011-12 school year for seats the following year (this information includes preference lists, priorities, random numbers, assignment status, and school capacities). School-level scores were constructed by summing scores for all component sub-schools used to implement seat reservation policies and to define buckets. Our empirical work also uses files with information on October enrollment and standardized scores from the Colorado School Assessment Program (CSAP) and the Transitional Colorado Assessment Program (TCAP) tests, given annually in grades 3-10. A data appendix describes these files and the extract we've created from them. For our purposes, “Charter schools” are schools identified as “charter” in DPS 2012-2013 SchoolChoice Enrollment Guide brochures and not identified as “intensive pathways” schools, which serve students who are much older than typical for their grade.

10The gap here is probably due to our treatment of family link. The Blair charter school, where the simulated score randomization count is farthest from the corresponding DA score counts, has more applicants with family link than any other school. Unlike our DA score calculation, which ignores family link, the simulated score accommodates family link by assigning a unique type to every student affected by a link.
even though Table 1 shows that no applicant with non-degenerate offer risk ranked this
school first? Random assignment at GVR is a consequence of the many GVR applicants
randomized by admissions offers at schools they’d ranked more highly. This and related
determinants of offer risk are detailed in Table 2, which explores the anatomy of the DA
propensity score for 6th grade applicants to four middle schools in the STRIVE network.
In particular, we see (in column 8 of the table) that all randomized GVR applicants were
randomized by virtue of having $MID_{θ_s}$ inside the unit interval, with no one randomized
at GVR’s own cutoff (column 7 counts applicants randomized at each school’s cutoff).

In contrast with STRIVE’s GVR school, few applicants were randomized at STRIVE’s
Highland, Lake, and Montbello campuses. This is a consequence of the fact that most
Highland, Lake, and Montbello applicants were likely to clear marginal priority at these
schools (having $ρ_{θ_s} < ρ_s$), while having values of $MID_{θ_s}$ mostly equal to zero or 1,
eliminating random assignment at schools ranked more highly. Interestingly, the Federal
and Westwood campuses are the only STRIVE schools to see applicants randomized
around the cutoff in the school’s own marginal priority group. We could therefore learn
more about the impact of attendance at Federal and Westwood by changing the cutoff
there (e.g., by changing capacity), whereas such a change would be of little consequence
for evaluations of the other schools.

Table 2 also documents the weak connection between applicant randomization counts
and a naive definition of over-subscription based on school capacity. In particular, columns
2 and 3 reveal that four out of six schools described in the table ultimately made fewer
offers than they had seats available (far fewer in the case of Montbello). Even so, as-
signment at these schools was far from certain: they contribute to our score-conditioned
charter school impact analysis.

A broad summary of DPS random assignment appears in Figure 2. Panel (a) captures
the information in columns 3 and 6 of Table 1 by plotting the number of first-choice
applicants subject to randomization as black dots, with the total randomized at each
school plotted as an arrow pointing up from these dots (schools are indexed on the x-axis
by their capacities). This representation highlights the dramatic gains in the number of
schools and the precision with which they can be studied as a payoff to our full-information
approach to the DA research design. These benefits are not limited to the charter sector,
a fact documented in Panel (b) of the figure, which plots the same comparisons for non-
charter schools in the DPS match.

### 3.4.2 DPS Data and Descriptive Statistics

The DPS population enrolled in grades 3-9 in the Fall of 2011 is roughly 60% Hispanic, a
fact reported in Table 3, along with other descriptive statistics. We focus on grades 3-9
in 2011 because outcome scores come from TCAP tests taken in grades 4-10 in the spring
of the 2012-13 school year. The high proportion Hispanic makes DPS an especially interesting and unusual urban district. Not surprisingly in view of this, almost 30 percent of DPS students have limited English proficiency. Consistent with the high poverty rates seen in many urban districts, three quarters of DPS students are poor enough to qualify for a subsidized lunch. Roughly 20% of the DPS students in our data are identified as gifted, a designation that qualifies them for differentiated instruction and other programs.

Nearly 11,000 of the roughly 40,000 students enrolled in grades 3-9 in Fall 2011 sought to change their school for the following year by participating in the assignment, which occurs in the spring. The sample participating in the assignment, described in column 2 of Table 3, contains fewer charter school students than appear in the total DPS population, but is otherwise demographically similar. It's also worth noting that our impact analysis is limited to students enrolled in DPS in the baseline (pre-assignment) year of 2011. The sample described in column 2 is therefore a subset of that described in column 1. The 2012 school assignment, which also determines the propensity score, includes the column 2 sample plus new entrants.

Column 3 of Table 3 shows that of the nearly 11,000 DPS-at-baseline students included in the assignment, almost 5,000 ranked at least one charter school. We refer to these students as charter applicants; the estimated charter attendance effects that follow are for subsets of this applicant group. DPS charter applicants have baseline achievement levels and demographic characteristics broadly similar to those seen district-wide. The most noteworthy feature of the charter applicant sample is a reduced proportion white, from about 19% in the centralized assignment to a little over 12% among charter applicants. It's also worth noting that charter applicants have baseline test scores close to the DPS average. This contrasts with the modest positive selection of charter applicants seen in Boston (reported in Abdulkadiroglu et al. 2011).

A little over 1,400 charter applicants have a frequency estimate of the probability of charter assignment between zero and one; the count of applicants subject to random assignment rises to about 1,500 when the score is estimated by simulation. Charter applicants subject to random assignment are described in columns 4 and 6 of Table 3. Although only about 30% of charter applicants were randomly assigned a charter seat, these students look much like the full charter applicant pool. The main difference is a higher proportion of applicants of randomized applicants originating at a charter school (that is, already enrolled at a charter at the time they applied for seats elsewhere). Columns 5 and 7, which report statistics for the subset of the randomized group that enrolls in a charter school, show slightly higher baseline scores among charter students.

---

11 Grade 3 is omitted from the outcome sample because 3rd graders have no baseline test.
3.4.3 Score-Based Balance

Conditional on the propensity score, applicants offered a charter seat should look much like those not offered a seat. Moreover, because offers are randomly assigned conditional on the score, we expect to see conditional balance in all applicant characteristics and not just for the variables that define an applicant’s type. We assess the balancing properties of the DA propensity score using simulated expectations. Specifically, drawing lottery numbers 400 times, we ran DA and computed the DA propensity score each time, and then computed average covariate differences by offer status. The balance analysis begins with uncontrolled differences in average applicant characteristics, followed by regression-adjusted differences that put applicant characteristics on the left-hand side of regression models like equation (3.7), omitting the covariate controls, $X_i$.

Uncontrolled comparisons by offer status, reported in columns 1 and 2 of Table 4, show large differences in average student characteristics, especially for variables related to preferences. For instance, across 400 lottery draws, those not offered a charter seat ranked an average of 1.4 charters, but this figure increases by almost half a school for applicants who were offered a charter seat. Likewise, while fewer than 30% of those not offered a charter seat had ranked a charter school first, the probability applicants ranked a charter first increases to over 0.9 (that is, 0.29+0.62) for those offered a charter seat. Column 2 also reveals important demographic differences by offer status; Hispanic applicants, for example, are substantially over-represented among those offered a charter seat.\(^{12}\)

Conditioning on frequency estimates of the DA propensity score reduces differences by offer status markedly. This can be seen in columns 3-5 of Table 4. The first set of conditional results, which come from regression models with linear control for the propensity score rather than dummies, show virtually no difference by offer status in the odds a charter is ranked first or that an applicant is Hispanic. Offer gaps in other application and demographic variables are also much reduced in this specification. Columns 4 and 5 of the table show that non-parametric control for the DA propensity score (implemented by dummying all score values in the unit interval, with an average of 39 values across simulations when rounded to nearest hundredth and an average 47 without rounding) reduces offer gaps even further. These results establish that a single DPS applicant cohort is large enough for the DA propensity score to eliminate selection bias.\(^{13}\)

Columns 6-8 of Table 4, which report estimated offer gaps conditional on a simulated propensity score, show that the simulated score does a better job of balancing treatment and control groups than does the DA score. Differences by offer status conditional on

---

\(^{12}\)Table 4 omits standard errors because the only source of uncertainty here is the modest simulation error arising from the fact that we’ve drawn lottery numbers 400 instead of infinitely many times.

\(^{13}\)Table B4 provides a computational proof of Theorem 3 by reporting offer gaps of the sort shown in Table 4 for scaled-up versions of the DPS economy. Doubling the number of applicants and seats at each school in the DPS market pushes conditional gaps down markedly, and multiples of 4 and 8 make these small gaps even smaller.
the simulated score, whether estimated linearly or with nonparametric controls, appear mostly in the third decimal place. This reflects the fact that simulation recovers the actual finite-market propensity score (up to simulation error), while the DA propensity score is an asymptotic approximation that should be expected to provide perfect treatment-control balance only in the limit. It’s worth noting, however, that the simulated score starts with 1,148 unique values. As a practical matter, the simulated score must be smoothed to accommodate non-parametric control. Rounding to the nearest hundredth leaves us with 51 points of support, close to the number of support points seen for the DA score. Rounding to the nearest ten-thousandth leaves 121 points of support. Finer rounding produces noticeably better balance for the number-of-schools-ranked variable.

Our exploration of score-based balance is rounded out with the results from a traditional balance analysis such as would be seen in analyses of a randomized trial. Specifically, Table 5 documents balance for the DPS match by reporting the usual t and F-statistics for offer gaps in covariate means. Again, we look at balance conditional on propensity scores for applicants with scores strictly between 0 and 1. As can be seen in Table 5a, application covariates are well-balanced by non-parametric control for either DA or simulated score estimates (linear control for the DA propensity score leaves a significant gap in the number of charter schools ranked).14

Table 5a also demonstrates that full control for type leaves us with a much smaller sample than does control for the propensity score: models with full type control are run on a sample of size 301, a sample size reported in the last column of the table. Likewise, the fact that saturated control for the simulated score requires some smoothing can be see in the second last column showing the reduced sample available for estimation of models that control fully for a simulated score rounded to the nearest ten-thousandth.

Not surprisingly, a few significant imbalances emerge in balance tests for the longer list of baseline covariates, reported in Table 5b. Here, the simulated score seems to balance characteristics somewhat more completely than does the DA score, but the F-statistics (reported at the bottom of the table) that jointly test balance of all baseline covariates fail to reject the null hypothesis of conditional balance for any specification reported.

Baseline score gaps as large as −0.1σ appear in some of the comparisons at the bottom of the table. The fact that these gaps are not mirrored in the comparisons in Table 4 suggests the differences in Table 5 are due to chance. Still, we can mitigate the effect of chance differences on 2SLS estimates of charter effects by adding baseline score controls (and other covariates) to our empirical models. The inclusion of these additional controls also has the salutary effect of making the 2SLS estimates of interest considerably more precise (covariates include dummies for grade tested, gender, origin school charter status, race, gifted status, bilingual status, subsidized lunch eligibility, special education, limited

14 Table 5 reports the results controlling for frequency estimates of the DA propensity score and the simulated propensity score. Balance results using formula estimates of the score appear in Appendix Table B3.
English proficient status, and baseline test scores; baseline score controls are responsible for most of the resulting precision gain). Finally, it's worth noting that the imbalance left after conditioning on the DA propensity score turns out to matter little for the 2SLS estimates we're ultimately after.

Modes of Inference

Econometric inference typically tries to quantify the uncertainty due to random sampling. What then, to make of the fact that the analysis reported here uses data on all DPS applicants from 2012? On one hand, we might imagine that the applicants we happen to be studying constitute a random sample from some larger population of possible applicants. At the same time, the statistical uncertainty in our empirical work can also be seen as a consequence of random assignment: we see only a single lottery draw for each applicant, one of many possibilities even when the sample of applicants is viewed as fixed.

In an effort to determine whether the distinction between sampling inference and randomization inference matters for our purposes, we computed randomization p-values by repeatedly drawing lottery numbers and calculating offer gaps in covariates conditional on the simulated propensity score. Regression conditioning on the simulated score produces near-perfect balance in Table 4 so this distribution is what we should expect to see under the null hypothesis of no difference by treatment assignment. Randomization p-values are therefore given by quantiles of the t-statistics in the distribution resulting from these repeated draws.

The p-values associated with conventional robust t-statistics for covariate balance turn out to be close to the corresponding randomization p-values. For the number of charter schools an applicant has ranked, for example, the conventional p-value for balance is 0.885 while the corresponding randomization p-value is 0.850. This is consistent with a classic result on the asymptotic equivalence of randomization and sampling tests for differences in means (see, e.g., 15.2 in Lehmann and Romano 2005).

Abadie et al. (2014) generalize results on the large-sample equivalence of randomization and sampling inference to cover regression estimates of treatment effects and tests for covariate balance of the sort reported here. If the regression function is linear and the regression of treatment on controls is linear, the usual robust covariance matrix associated with random sampling is asymptotically valid for the sampling distribution induced by random assignment. The treatment in our case is an offer dummy, while the controls are dummies or a linear model for the propensity score. The second of these requirements holds here when the controls fully saturate the propensity score (ignoring any additional covariates). The first requires constant offer effects given a saturated model for the score. The models estimated here don't quite satisfy these conditions (they're not fully saturated) but do not seem to be so far off that this matters for inference.

A related issue arises from the fact that the empirical strategy used here conditions on
estimates of the propensity score (the simulated score is also an estimate since it’s based on a finite number of draws). As noted by Hirano et al. (2003) and Abadie and Imbens (2012), conditioning on an estimated as opposed to a non-stochastic known score may affect sampling distributions of the resulting estimated causal effects. We therefore checked conventional large-sample p-values against randomization p-values for the reduced-form charter offer effects associated with the 2SLS estimates discussed in the next section. Robust asymptotic sampling formulas again generate p-values close to a randomization-inference benchmark, regardless of how the score behind these estimates was constructed. In view of these findings, we rely on the usual robust standard errors and test statistics for inference about 2SLS estimates of treatment effects.\textsuperscript{15}

3.4.4 Effects of Charter Enrollment

2SLS estimates of charter attendance effects are remarkably similar to the corresponding semiparametric estimates. This is apparent in Table 6, which compares 2SLS estimates of models with additive score controls to semiparametric estimates of average treatment effects constructed using three versions of the score. Compare, for example, frequency-score-controlled 2SLS estimates of effects on math and reading of 0.496 and 0.127 with semiparametric estimates around 0.44 and 0.11. At the same time, standard errors for the semiparametric estimates are higher than those for 2SLS (semiparametric precision is estimated using a Bayesian bootstrap that randomly reweights observations; see Shao and Tu (1995) for an introduction). Semiparametric estimates weighted using a simulated score are especially imprecise. The similarity of 2SLS and semiparametric estimates and the relative simplicity of 2SLS estimation leads us to report only 2SLS estimates in what follows.\textsuperscript{16}

A DA-generated charter offer boosts charter school attendance rates by about 0.4. These first stage estimates, shown in the first row of Table 7, are computed by estimating equation (3.7). The first stage of 0.4 reflects the fact that many charter applicants who are not offered a seat in the SchoolChoice first round ultimately find their way into a charter school by applying to schools directly in the second round (specifically, 43% of the charter applicants analyzed in Table 7 are always-takers who enroll in charters even without a first-round charter offer, while fewer than 20% of the analysis sample are never-takers.

\textsuperscript{15}Appendix Table B2 reports conditional-on-score estimates of attrition differentials by offer status. Here, we see marginally significant gaps on the order of 4-5 points when estimated conditional on the DA propensity score. Attrition differentials fall to a statistically insignificant 3 points when estimated conditional on a simulated score. The estimated charter attendance effects discussed below are similar when computed using either type of score control, so it seems unlikely that differential attrition is a source of bias in our 2SLS estimates.

\textsuperscript{16}2SLS also obviates the need for judgements regarding bootstrap methods or implementation. We found, for example, that a conventional nonparametric bootstrap for the semiparametric estimators requires trimming or tuning to eliminate the influence of occasional small first stage estimates.
who decline charter offers). First-stage estimates of around 0.68 computed without score controls, shown in column 4 of the table, are clearly biased upwards.\textsuperscript{17}

2SLS estimates of charter attendance effects on test scores, reported below the first-stage estimates in Table 7, show remarkably large gains in math, with smaller effects on reading. The math gains reported here are similar to those found for charter students in Boston (see, for example, Abdulkadiroğlu et al. 2011). Previous lottery-based studies of charter schools likewise report substantially larger gains in math than in reading. Here, we also see large and statistically significant gains in writing scores.

Importantly for our methodological agenda, the estimated charter attendance effects reported in Table 7 are largely invariant to whether the propensity score is estimated by simulation or by a frequency or formula calculation that uses Theorem 2. Compare, for example, math impact estimates of 0.496, 0.524, and 0.543 using frequency-, formula-, and simulation-based score controls, all estimated with similar precision. This alignment validates the use of Theorem 2 to control for applicant type.

Estimates that omit propensity score controls highlight the risk of selection bias in a naive 2SLS empirical strategy. This is documented in column 4 of Table 7, which shows that 2SLS estimates of math and writing effects constructed using DA offer instruments while omitting propensity score controls are too small by about half. A corresponding set of OLS estimates without propensity score controls, reported in column 5 of the table, also tends to underestimate the gains from charter attendance attendance. The results in column 6 show that adding score controls to the OLS model pulls the estimates up a little, but a substantial gap between between these and the corresponding set of 2SLS estimates remains.

3.4.5 Alternative IV Strategies

We're interested in comparing 2SLS estimates constructed using a DA offer dummy as an instrument while controlling for the DA propensity score with suitably-controlled estimates constructed using first-choice and qualification instruments. As noted in Section 3.3.4, we expect DA-offer instruments to yield a precision gain and to increase the number of schools represented in the estimation sample relative to these two previously-employed IV strategies.\textsuperscript{18}

\textsuperscript{17}The estimates reported in this table control for baseline test scores and the covariates described earlier. These extra controls are not necessary for consistent causal inference but their inclusion increases precision (Estimates without covariates appear in the appendix). Estimates here are for scores in grades 4-10. The pattern of results in an analysis that separates high schools from middle and elementary schools is similar. The sample used for IV estimation is limited to charter applicants with the relevant propensity score in the unit interval, for which score cells have offer variation in the data at hand (these restrictions amount to the same thing for the frequency score). The OLS estimation sample includes charter applicants, ignoring score- and cell-variation restrictions.

\textsuperscript{18}Studies using first-choice instruments to evaluate schools in districts with centralized assignment include Abdulkadiroğlu et al. (2013), Deming (2011), Deming et al. (2014), and Hastings et al. (2009).
Let \( R(\theta_i) \) be a variable that uniquely identifies the charter school that applicant \( i \) ranks first, along with his priority status at this school, defined for applicants whose first choice is indeed a charter school. \( R(\theta_i) \) ignores other schools that might have been ranked. The first-choice strategy is implemented by the following 2SLS setup:

\[
Y_i = \sum_x \alpha(x)d_i(x) + \beta C_i + \epsilon_i,
\]

\[
C_i = \sum_x \gamma(x)d_i(x) + \delta D_i^f + \nu_i,
\]

where the \( d_i(x) \)'s are dummies indicating values of \( R(\theta_i) \), indexed by \( x \), and \( \alpha(x) \) and \( \gamma(x) \) are the associated “risk set effects” in the first and second stages. The first-choice instrument, \( D_i^f \), is a dummy variable indicating \( i \)'s qualification at his or her first-choice school. In other words,

\[
D_i^f = 1[\pi_{is} \leq c_s \text{ for charter } s \text{ that } i \text{ has ranked first}].
\]

First choice qualification is the same as first choice offer since under DA, applicants who rank a first are offered a seat there if and only if they qualify at a.

The qualification strategy expands the sample to include all charter applicants, with the risk sets \( (R(\theta_i)) \) for qualification instruments identifying the set of all charter schools that \( i \) ranks, along with his or her priority status at each of these schools. In this case, \( R(\theta_i) \) ignores the order in which schools are ranked, coding only their identities, but priorities are associated with schools. The qualification instrument, \( D_i^q \), indicates qualification at any charter he or she has ranked. In other words,

\[
D_i^q = 1[\pi_{is} \leq c_s \text{ for at least one charter } s \text{ that } i \text{ has ranked}].
\]

In large markets, the instruments \( D_i^f \) and \( D_i^q \) are independent of type conditional on \( R(\theta_i) \); see Appendix 3.6.5 for details.

A primary source of inefficiency in the first-choice and qualification strategies is apparent in Panel A of Table 8. This panel reports two sorts of first stage estimates for each instrument: the first of these regresses a dummy indicating any charter offer—that is, our DA charter offer instrument, \( D_i^f \)—on each of the three instruments under consideration. A regression of \( D_i \) on itself necessarily produces a coefficient of one. By contrast, a first-choice offer boosts the probability of any charter offer by only around 0.77 in the sample of those who have ranked a charter first. This reflects the fact that, while anyone

---

First-choice instruments have also been used with decentralized assignment mechanisms (Abdulkadiroğlu et al. (2011), Cullen et al. (2006), Dobbie and Fryer (2011), and Hoxby et al. (2009)). Dobbie and Fryer (2014), Lucas and Mbiti (2014b), and Pop-Eleches and Urquiola (2013) use qualification instruments.

For example, an applicant who ranks A and B with marginal priority only at A is distinguished from an applicant who ranks A and B with marginal priority only at B.
receiving a first choice charter offer has surely been offered a charter seat, roughly 23% of the sample ranking a charter first is offered a charter seat at schools other than their first choice. The relationship between \( D_i^q \) and charter offers is even weaker, at around 0.48. This reflects the fact that for schools below the one ranked first, charter qualification is not sufficient for a charter offer.

The diminished impact of the two alternative instruments on charter offers translates into a weakened first stage for charter enrollment. The best case scenario, using all DA-generated offers (that is, \( D_i \)) as a source of quasi-experimental variation, produces a first stage of around 0.41. But first-choice offers boost charter enrollment by just 0.32, while qualification anywhere yields a charter enrollment gain of only 0.18. As always with comparisons of IV strategies, the size of the first stage is a primary determinant of relative precision.

At 0.071, the standard error of the DA-offer estimate is markedly lower than the standard error of 0.102 yielded by a first-choice strategy and well below the standard error of 0.149 generated by qualification instruments. In fact, the precision loss here is virtually the same as the decline in the intermediate first stages recorded in the first row of the table (compare 0.774 with 0.071/0.102 = 0.696 and 0.476 with 0.071/0.149 = 0.477). The loss here is substantial: columns 4 and 5 show the sample size increase needed to undo the damage done by a smaller first stage for each alternative instrument.\(^{20}\)

Only half as many schools are represented in the first-choice analysis sample as in the DA sample (At 24, the number of schools in the qualification sample is closer to the full complement of 30 schools available for study with DA offers). First-choice analyses lose schools because many lotteries fail to randomize first-choice applicants (as seen in Table 1). It’s therefore interesting to note that the first-choice estimate of effects on math and reading scores are noticeably larger than the estimates generated using DA offer and qualification instruments (compare the estimate of 0.5 using DA offers with estimates of 0.6 and 0.41 using first-choice and qualification instruments). This finding may reflect an advantage for those awarded a seat at their first choice school (Hastings et al. 2009; Deming 2011; Deming et al. 2014 find a general “first choice advantage” in analyses of school attendance effects.) By contrast, the DA offer instrument yields an estimand that is more representative of the full complement of charter schools in the match. In the same spirit, it’s worth noting that the first-choice and qualification IV samples include no 10th graders.

### 3.4.6 Charter School Effects with a Mixed Counterfactual

The 2SLS estimates in Tables 7 and 8 contrast charter outcomes with potential outcomes generated by attendance at a mix of traditional public schools and schools from other

---

\(^{20}\)The sample used to construct the estimates in columns 1-3 of Table 8 is limited to those who have variation in the instrument at hand conditional on the relevant risk sets controls.
non-charter sectors. We'd like to simplify this mix so as to produce something closer to a pure sector-to-sector comparison. Allowance for more than one treatment channel also addresses concerns about charter-offer-induced changes in counterfactual outcomes that might cause violations of the exclusion restriction.

Our first step in this effort is to describe the distribution of non-charter school choices for applicants who were and weren't offered a charter seat in the DPS assignment. We then identify the distribution of counterfactual (non-charter) school sectors for the group of charter-lottery compliers. Finally, we use the DA mechanism to jointly estimate causal effects of attendance at schools in different sectors, thereby making the non-charter counterfactual in our 2SLS estimates more homogeneous.

The analysis here builds on a categorical variable, \( W_i \), capturing school sector in which \( i \) is enrolled. Important DPS sectors besides the charter sector are traditional public schools, innovation schools, magnet schools, and alternative schools. Innovation and magnet schools are managed by DPS. Innovation schools design and implement innovative practices meant to improve student outcomes (for details and a descriptive evaluation of innovation schools, see Connors et al. 2013). Magnet schools serve students with particular styles of learning. Alternative schools serve mainly older students struggling with factors that may prevent them from succeeding in a traditional school environment. Smaller school sectors include a single charter middle school outside the centralized DPS assignment process (now closed) and a private school contracted to serve DPS students.

The distribution of enrollment sectors for students who do and don't receive a charter offer are described in the first two columns of Table 9. These columns show a charter enrollment rate of 87% in the group offered a charter seat, along with substantial but much smaller charter enrollment in the non-offered group.\(^{21}\) Perhaps surprisingly, only around 41% of those not offered a charter seat enroll in a traditional public schools, with the rest of the non-offered group distributed over a variety of school types. Innovation schools are the leading non-charter alternative to traditional public schools. Innovation schools operate under an innovation plan that waives some provisions of the relevant collective bargaining agreements (for a descriptive evaluation of these schools, see Connors et al. 2013).\(^{22}\)

The sector distribution for non-offered applicants with non-trivial charter risk appears in column 3 of Table 9, alongside the sum of the non-offered mean and a charter-offer treatment effect on enrollment in each sector in column 4. These extended first-stage estimates, computed by putting indicators \( 1(W_i = j) \) on the left-hand side of equation (3.7), control for the DA propensity score and therefore have a causal interpretation.

---

\(^{21}\)As noted in the discussion of the first stage estimates in Table 7, applicants unhappy with the offer they've receive in the first round of SchoolChoice can apply to schools individually in a second round.

\(^{22}\)Innovation waivers are subject to approval by the Denver Classroom Teachers Association (which organizes Denver public school teachers' bargaining unit), and they allow, for example, increased instruction time. DPS innovation schools appear to have much in common with Boston's pilot schools, a model examined in Abdulkadiroglu et al. (2011).
The number of applicants not offered a seat who end up in a charter school is higher for those with non-trivial charter offer risk than in the full applicant sample, as can be seen by comparing columns 3 and 1. The charter enrollment first stage that’s implicit in the column 4-vs-3 comparison matches the first stage in Table 7. First stages for other sectors show charter offers sharply reduce innovation school enrollment as well as reducing enrollment in traditional public schools.

The 2SLS estimates reported in Table 7 capture causal effect for charter lottery compliers. We describe the distribution of school sectors for compliers by defining potential sector enrollment variables, \( W_{i1} \) and \( W_{0i} \), indexed against charter offers, \( D_i \). Potential and observed sector variables are related by

\[
W_i = W_{0i} + (W_{i1} - W_{0i})D_i.
\]

In the population of charter-offer compliers, \( W_{i1} = \text{charter} \) for all \( i \): by definition, charter-offer compliers attend a charter school when the DPS assignment offers them the opportunity to do so. Here, we’re interested in \( E[1(W_{0i} = j)|C_{i1} > C_{0i}] \), that is, the sector distribution for charter-offer compliers in the scenario where they aren’t offered a charter seat. We refer to this distribution as describing enrollment destinies for compliers.

Enrollment destinies are marginal potential outcome distributions for compliers. As shown by Abadie (2002), these are identified by a simple 2SLS estimand. The details of our implementation of this identification strategy follow those in Angrist et al. (2015a), with the modification that instead of estimating marginal potential outcome densities for a continuous variable, the outcomes of interest here are Bernoulli.\(^{23}\)

Column 5 of Table 9 reveals that only about half of charter lottery compliers are destined to end up in a traditional public school if they aren’t offered a charter seat. The second most-likely counterfactual destiny for the younger applicant group is an innovation school, with nearly a third of non-offered compliers enrolling in one of these. The likelihood of an enrollment destiny outside the charter, traditional, and innovation sectors is much smaller.

**Isolating an Innovation School Effect**

The outsize role of innovation schools in counterfactual destinies motivates an empirical strategy that allows for distinct charter and innovation school treatment effects. By pulling innovation schools out of the non-charter counterfactual, we capture charter treatment effects driven mainly by the contrast between charter and traditional public schools. Comparisons with a more homogeneous counterfactual also mitigates bias that might arise

\(^{23}\)Briefly, our procedure puts \((1 - C_i)1(W_i = j)\) on the left hand side of a version of equation (3.8) with endogenous variable \(1 - C_i\). The coefficient on this endogenous variable is an estimate of \( E[1(W_{0i} = j)|C_{i1} > C_{0i}, X_i] \). The covariates and sample used here are the same as used to construct the 2SLS impact estimates reported in column 1 of Table 7.
from violations of the exclusion restriction (discussed in Section 3.3.4). And, of course, the innovation treatment effect is also of interest in its own right.

For the purposes of this discussion, we code the sector variable, $W_i$, as taking on the value 2 for innovation schools, the value 1 for charters, and 0 otherwise. The corresponding potential outcomes are $Y_{2i}, Y_{1i},$ and $Y_{0i}$. In principal, this leads to multiple heterogenous causal effects, $Y_{2i} - Y_{0i}$ and $Y_{1i} - Y_{0i}$, but the identification of multiple-treatment models with unrestricted heterogeneity raises issues that go beyond the scope of this paper.\(^\text{24}\)

Defining constant effects with the notation

\[
Y_{2i} - Y_{0i} = \beta_2, \\
Y_{1i} - Y_{0i} = \beta_1,
\]

our two-sector identification strategy can be motivated by the conditional independence assumption,

\[
Y_{0i} \perp Z_i | \theta_i, \tag{3.11}
\]

where $Z_i$ is a categorical variable that records DA-generated offers in each sector sector (charter, innovation, other).

The instruments here are indicators for charter and innovation-sector offers, $D^1_i = 1[Z_i = 1]$ and $D^2_i = 1[Z_i = 2]$. These dummy instruments are used in a 2SLS procedure with two endogenous variables, $C^1_i$ for charter school enrollment and $C^2_i$ for innovation school enrollment. Propensity score conditioning is justified by the fact that conditional independence relation (3.11) implies

\[
Y_{0i} \perp Z_i | p_1(\theta), p_2(\theta), \tag{3.12}
\]

where $p_1(\theta) = E[D^1_i | \theta]$ and $p_2(\theta) = E[D^2_i | \theta]$.

The 2SLS setup in this case consists of

\[
Y_i = \sum_x \alpha_1(x)d_1^i(x) + \sum_x \alpha_2(x)d_2^i(x) + \beta_1 C^1_i + \beta_2 C^2_i + \epsilon_i, \tag{3.13}
\]

\[
C^1_i = \sum_x \gamma_{11}(x)d_1^i(x) + \sum_x \gamma_{12}(x)d_2^i(x) + \delta_{11} D^1_i + \delta_{12} D^2_i + \nu_i, \tag{3.14}
\]

\[
C^2_i = \sum_x \gamma_{21}(x)d_1^i(x) + \sum_x \gamma_{22}(x)d_2^i(x) + \delta_{21} D^1_i + \delta_{22} D^2_i + \eta_i, \tag{3.15}
\]

where the dummy control variables, $d_1^i(x)$ and $d_2^i(x)$, saturate estimates of the scores for each treatment, $\hat{p}_1(\theta_i)$ and $\hat{p}_2(\theta_i)$, with corresponding score effects denoted by $\gamma$’s and $\alpha$’s in the first and second stage models. The sample used for this analysis is the union of

\(^{24}\)See Behaghel et al. (2013) and Blackwell (2015) for recent progress on this issue.
charter and innovation school applicants.

As noted by Imbens (2000) and Yang et al. (2014), the key conditional independence relation in this context (equation (3.12)) suggests we should choose a parameterization that fixes conditional probabilities of assignment for all treatment levels jointly. Joint score control replaces the additive score controls in equations (3.13), (3.14), and (3.15) with score controls of the form

\[ d_{i}^{12}(x^1, x^2) = 1[p_{1}(\theta_i) = x^1, p_{2}(\theta_i) = x^2], \]

where hats denote score estimates and the indices, \( x^1 \) and \( x^2 \), run independently over all values in the support for each score. This model generates far more score fixed effects than does equation (3.13).25 Fortunately, however, the algebra of 2SLS obviates the need for joint score control; additive control is enough.

To see why additive control is adequate for 2SLS models that exploit (3.12), note first that 2SLS estimates of the second stage equation, (3.13), can be obtained by first regressing each offer dummy on the full set of \( d_1^i(x) \) and \( d_2^i(x) \) and then using the residuals from this regression as instruments after dropping these controls from the model (see, e.g., Angrist and Pischke 2009). Note also that both sets of regressors in this auxiliary first-step model are functions of type. A regression of \( D_i^j \) on the full set of \( d_1^i(x) \) and \( d_2^i(x) \) therefore returns fitted values given by \( E[D_i^j|\theta] = p_j(\theta_i) \).

Suppose now that we replace additive controls, \( d_1^i(x) \) and \( d_2^i(x) \), with the full set of dummies, \( d_{12}^i(x^1, x^2) \), parameterizing the jointly-controlled model. Since the model here is saturated, a regression of \( D_i^j \) on the full set of \( d_{12}^i(x^1, x^2) \) dummies recovers the conditional expectation function of offers given both scores. By the law of iterated expectations, however, this is

\[ E[D_i^j|p_1(\theta_i), p_2(\theta_i)] = E\{E[D_i^j|\theta]|p_1(\theta_i), p_2(\theta_i)\} = p_j(\theta_i). \]

From this we conclude that the IV equivalent of 2SLS is the same in the additive and jointly-controlled models.26

As a benchmark, columns 1-2 of Table 10 compare charter-only and innovation-only estimates computed using DA (frequency) score controls. Each sample is limited to applicants to the relevant sector.27 A parallel set of single-sector estimates using simulated score controls appears in columns 5 and 6. The innovation first stage (the effect of an

---

25 When \( p_1(\theta) \) takes on \( k_1 \) values and \( p_2(\theta) \) takes on \( k_2 \) values, the additive model has \( k_1 + k_2 - 2 \) score parameters, while the joint model has \( k_1 k_2 - 1 \).

26 This conclusion holds in the population, but need not hold exactly in our data (because scores here are estimated by something more elaborate than a sample mean conditional on type) or for models that include additional covariates beyond saturated score controls.

27 Appendix Table B6 lists innovation schools and describes the random assignment pattern at these schools along the lines of Table 1 for charter schools. Covariate balance and differential attrition results for innovation schools are reported in Appendix Table B7.
innovation school offer on innovation school enrollment) is around 0.35. The pooled single-sector charter estimates in Table 10 are the same as those in Table 7. Not surprisingly in view of the substantially reduced number of applicants with non-trivial innovation offer risk (546 in column 2 and 613 in column 6 of Table 10), and the smaller innovation first stage, the innovation attendance effects are relatively imprecise. This imprecision notwithstanding, the innovation-only model generates a large negative and marginally significant effect on reading when estimated with the DA score.

2SLS estimates of equation (3.13) appear in columns 3 and 7 of Table 10. Charter school effects change little in this specification, but (insignificant) negative innovation estimates for math flip to positive when estimated using a model that also isolates charter treatment effects. The negative innovation school effects on reading seen in columns 2 and 6 also become smaller in the two-endogenous-variables models. Most interestingly, perhaps, the marginally significant positive charter school effect on reading (when estimating using DA score controls) also disappears. While charter students’ reading performance exceeds what we can expect to see were these students to enroll in a mix of traditional and (low-performing) innovation schools, the reading gap between charters and traditional public schools is a little smaller.

As the theoretical discussion above leads us to expect, the results of estimation with joint score controls, shown in columns 4 and 8 of Table 10, differ little from the estimates constructed using additive score controls reported in columns 3 and 7 (a marginally significant though still imprecisely estimate positive innovation effect on math scores emerges in column 4). Overall, it seems fair to say that the findings on charter effectiveness in Table 7 stand when charter effects are estimated using a procedure that removes the innovation sector from the charter enrollment counterfactual.

3.5 Summary and Directions for Further Work

We investigate empirical strategies that use the lottery number embedded in market design solutions to school matching problems as a research tool. The most important fruit of this inquiry is the DA propensity score, an easily-computed formula for the conditional probability of assignment to particular schools as a function of type. The DA propensity score reveals the nature of the experimental design generated as a by-product of market design and suggests directions in which match parameters might be modified so as to boost the research value of school assignment and other matching schemes. We also show how the DA score can be used to simultaneously evaluate attendance effects in multiple sectors or schools.

A score-based analysis of data from Denver’s unified school match reveals substantial gains from attendance at one of Denver’s many charter schools. The resulting charter effects are similar to those computed using single-school lottery strategies for Boston’s
charters reported in Abdulkadiroğlu et al. (2011). At the same time, as with previously reported results for Boston Pilot schools, Denver’s Innovation model does not appear to generate substantial achievement gains. Our analysis focuses on defining and estimating the DA propensity score, giving less attention to the problem of how best to use the score for estimation. Still, simple 2SLS procedures seem to work well, and the resulting estimates of DPS charter effects differ little from those generated by semiparametric alternatives.

The methods developed here should be broadly applicable to markets using the DA family of mechanisms for centralized assignment. At the same time, some markets and matches use mechanisms not covered by the DA framework. Most important on this list is the top trading cycles (TTC) mechanism (Shapley and Scarf, 1974; Abdulkadiroğlu and Sönmez, 2003), which allows students to trade priorities rather than treating priorities as fixed. We hope to report theoretical results on the TTC propensity score soon, along with results from an application to New Orleans Recovery School District, which has experimented with TTC matching (Abdulkadiroğlu et al., 2014c).

Finally, many matching problems, such as the selective exam schools analyzed by Jackson (2010); Dobbie and Fryer (2014); Abdulkadiroğlu et al. (2014b); Lucas and Mbiti (2014b); Pop-Eleches and Urquiola (2013) use non-randomly-assigned tie-breakers rather than a lottery. These schemes embed a regression discontinuity design inside a market design rather than embedding a randomized trial. The question of how best to define and exploit the DA propensity score for markets that combine regression-discontinuity tie-breaking with market design matchmaking is a natural next step on the market-design-meets-research-design agenda.
Figure 3.5.1: Propensity Scores and Market Size in Example 2

Notes: This figure plots finite-market propensity scores for expansions of Example 2 in Section 3.2.2. For each value of the x axis, we consider an expansion of the example with x students of each type. The propensity scores plotted here were computed by drawing lottery numbers 100,000 times.
Figure 3.5.2: Sample Size Gains from the Propensity Score Strategy

Notes: These figures compare the sample size under our DA propensity score strategy to that under the first choice strategy. Down arrows mean the two empirical strategies produce the same number of applicants subject to randomization at the corresponding schools. We say a student is subject to randomization at a school if the student has the DA propensity score (frequency) of assignment to that school that is neither 0 nor 1.
<table>
<thead>
<tr>
<th>School</th>
<th>Total applicants</th>
<th>Applicants offered seats</th>
<th>DA score (frequency)</th>
<th>DA score (formula)</th>
<th>Simulated score</th>
<th>Simulated score (first choice)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary and middle schools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cesar Chavez Academy Denver</td>
<td>62</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Denver Language School</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DSST: Cole</td>
<td>281</td>
<td>129</td>
<td>31</td>
<td>40</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>DSST: College View</td>
<td>299</td>
<td>130</td>
<td>47</td>
<td>67</td>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td>DSST: Green Valley Ranch</td>
<td>1014</td>
<td>146</td>
<td>324</td>
<td>344</td>
<td>357</td>
<td>291</td>
</tr>
<tr>
<td>DSST: Stapleton</td>
<td>849</td>
<td>156</td>
<td>180</td>
<td>189</td>
<td>221</td>
<td>137</td>
</tr>
<tr>
<td>Girls Athletic Leadership School</td>
<td>221</td>
<td>86</td>
<td>18</td>
<td>40</td>
<td>48</td>
<td>0</td>
</tr>
<tr>
<td>Highline Academy Charter School</td>
<td>159</td>
<td>26</td>
<td>69</td>
<td>78</td>
<td>84</td>
<td>50</td>
</tr>
<tr>
<td>KIPP Montbello College Prep</td>
<td>211</td>
<td>39</td>
<td>36</td>
<td>48</td>
<td>55</td>
<td>20</td>
</tr>
<tr>
<td>KIPP Sunshine Peak Academy</td>
<td>389</td>
<td>83</td>
<td>41</td>
<td>42</td>
<td>44</td>
<td>36</td>
</tr>
<tr>
<td>Odyssey Charter Elementary</td>
<td>215</td>
<td>6</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>14</td>
</tr>
<tr>
<td>Omar D. Blair Charter School</td>
<td>385</td>
<td>114</td>
<td>135</td>
<td>141</td>
<td>182</td>
<td>99</td>
</tr>
<tr>
<td>Pioneer Charter School</td>
<td>25</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>SIMS Fayola International Academy Denver</td>
<td>86</td>
<td>37</td>
<td>7</td>
<td>18</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>SOAR at Green Valley Ranch</td>
<td>85</td>
<td>9</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>37</td>
</tr>
<tr>
<td>SOAR Oakland</td>
<td>40</td>
<td>4</td>
<td>0</td>
<td>9</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>STRIVE Prep - Federal</td>
<td>621</td>
<td>138</td>
<td>170</td>
<td>172</td>
<td>175</td>
<td>131</td>
</tr>
<tr>
<td>STRIVE Prep - GVR</td>
<td>324</td>
<td>112</td>
<td>104</td>
<td>116</td>
<td>118</td>
<td>0</td>
</tr>
<tr>
<td>STRIVE Prep - Highland</td>
<td>263</td>
<td>112</td>
<td>2</td>
<td>21</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>STRIVE Prep - Lake</td>
<td>320</td>
<td>126</td>
<td>18</td>
<td>26</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>STRIVE Prep - Montbello</td>
<td>188</td>
<td>37</td>
<td>16</td>
<td>31</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>STRIVE Prep - Westwood</td>
<td>535</td>
<td>141</td>
<td>235</td>
<td>238</td>
<td>239</td>
<td>141</td>
</tr>
<tr>
<td>Venture Prep</td>
<td>100</td>
<td>50</td>
<td>12</td>
<td>17</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>Wyatt Edison Charter Elementary</td>
<td>48</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>High schools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSST: Green Valley Ranch</td>
<td>806</td>
<td>173</td>
<td>290</td>
<td>343</td>
<td>330</td>
<td>263</td>
</tr>
<tr>
<td>DSST: Stapleton</td>
<td>522</td>
<td>27</td>
<td>116</td>
<td>117</td>
<td>139</td>
<td>96</td>
</tr>
<tr>
<td>Southwest Early College</td>
<td>265</td>
<td>76</td>
<td>34</td>
<td>47</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>Venture Prep</td>
<td>140</td>
<td>39</td>
<td>28</td>
<td>42</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>KIPP Denver Collegiate High School</td>
<td>268</td>
<td>60</td>
<td>29</td>
<td>37</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td>SIMS Fayola International Academy Denver</td>
<td>71</td>
<td>15</td>
<td>6</td>
<td>22</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>STRIVE Prep - SMART</td>
<td>383</td>
<td>160</td>
<td>175</td>
<td>175</td>
<td>175</td>
<td>175</td>
</tr>
</tbody>
</table>

Notes: This table describes DPS charter applications. Column 1 reports the number of applicants ranking each school. Columns 3-6 count applicants with propensity score values strictly between zero and one according to different score computation methods. Column 6 shows the subset of applicants from column 5 who rank each school as their first choice.
<table>
<thead>
<tr>
<th>Campus</th>
<th>Eligible applicants</th>
<th>Capacity</th>
<th>Offers</th>
<th>$\Theta^u_s$</th>
<th>$\Theta^c_s$</th>
<th>$\Theta^a_s$</th>
<th>$\Theta^c_s$</th>
<th>$\Theta^a_s$</th>
<th>$\Theta^u_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVR</td>
<td>324</td>
<td>147</td>
<td>112</td>
<td>0</td>
<td>0</td>
<td>159</td>
<td>0</td>
<td>116</td>
<td>49</td>
</tr>
<tr>
<td>Lake</td>
<td>274</td>
<td>147</td>
<td>126</td>
<td>0</td>
<td>0</td>
<td>132</td>
<td>0</td>
<td>26</td>
<td>116</td>
</tr>
<tr>
<td>Highland</td>
<td>244</td>
<td>147</td>
<td>112</td>
<td>0</td>
<td>0</td>
<td>121</td>
<td>0</td>
<td>21</td>
<td>102</td>
</tr>
<tr>
<td>Montbello</td>
<td>188</td>
<td>147</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>128</td>
<td>0</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>Federal</td>
<td>574</td>
<td>138</td>
<td>138</td>
<td>78</td>
<td>284</td>
<td>3</td>
<td>171</td>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>Westwood</td>
<td>494</td>
<td>141</td>
<td>141</td>
<td>53</td>
<td>181</td>
<td>4</td>
<td>238</td>
<td>0</td>
<td>18</td>
</tr>
</tbody>
</table>

Notes: This table shows how formula scores are determined for STRIVE school seats in grade 6 (all 6th grade seats at these schools are assigned in a single bucket; ineligible applicants, who have a score of zero, are omitted). Column 3 records offers made to these applicants. Columns 4-6 show the number of applicants in partitions with a score of zero. Columns 7 and 8 show the number of applicants subject to random assignment. Column 9 shows the number of applicants with certain offers.
<table>
<thead>
<tr>
<th></th>
<th>Denver students</th>
<th>SchoolChoice applicants</th>
<th>Charter applicants</th>
<th>Propensity score in (0,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>DA score (frequency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Charter applicants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Charter students</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Simulated score</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Charter applicants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Charter students</td>
</tr>
<tr>
<td>Origin school is charter</td>
<td>0.133</td>
<td>0.080</td>
<td>0.130</td>
<td>0.259</td>
</tr>
<tr>
<td>Female</td>
<td>0.495</td>
<td>0.502</td>
<td>0.518</td>
<td>0.488</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.594</td>
<td>0.593</td>
<td>0.633</td>
<td>0.667</td>
</tr>
<tr>
<td>Black</td>
<td>0.141</td>
<td>0.143</td>
<td>0.169</td>
<td>0.181</td>
</tr>
<tr>
<td>White</td>
<td>0.192</td>
<td>0.187</td>
<td>0.124</td>
<td>0.084</td>
</tr>
<tr>
<td>Asian</td>
<td>0.034</td>
<td>0.034</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td>Gifted</td>
<td>0.171</td>
<td>0.213</td>
<td>0.192</td>
<td>0.159</td>
</tr>
<tr>
<td>Bilingual</td>
<td>0.039</td>
<td>0.026</td>
<td>0.033</td>
<td>0.038</td>
</tr>
<tr>
<td>Subsidized lunch</td>
<td>0.753</td>
<td>0.756</td>
<td>0.797</td>
<td>0.813</td>
</tr>
<tr>
<td>Limited English proficient</td>
<td>0.285</td>
<td>0.290</td>
<td>0.324</td>
<td>0.343</td>
</tr>
<tr>
<td>Special education</td>
<td>0.119</td>
<td>0.114</td>
<td>0.085</td>
<td>0.079</td>
</tr>
<tr>
<td>Baseline scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>0.000</td>
<td>0.015</td>
<td>0.021</td>
<td>0.037</td>
</tr>
<tr>
<td>Reading</td>
<td>0.000</td>
<td>0.016</td>
<td>0.005</td>
<td>-0.011</td>
</tr>
<tr>
<td>Writing</td>
<td>0.000</td>
<td>0.010</td>
<td>0.006</td>
<td>0.001</td>
</tr>
</tbody>
</table>

|                          |                |                         |                    | Simulated score         |
|                          |                |                         |                    | Charter applicants       |
|                          |                |                         |                    | Charter students         |
|                          |                |                         |                    | DA score (frequency)     |
|                          |                |                         |                    | Charter applicants       |
|                          |                |                         |                    | Charter students         |
| N                        | 40,143         | 10,898                  | 4,964              | 1,436                    |
|                          |                |                         |                    | 828                      |
|                          |                |                         |                    | 1,523                    |
|                          |                |                         |                    | 781                      |

Notes: This table describes the population of Denver 3rd-9th graders in 2011-2012, the baseline and application year. Statistics in column 1 are for charter and non-charter students. Column 2 describes the subset that submitted an application to the SchoolChoice system for a seat in grades 4-10 at another DPS school in 2012-2013. Column 3 reports values for applicants ranking any charter school. Columns 4-7 show statistics for charter applicants with propensity score values strictly between zero and one. Test scores are standardized to the population in column 1.
### Table 4: Expected balance

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Non-offered mean (1)</th>
<th>No controls (2)</th>
<th>Linear control (3)</th>
<th>Rounded (hundredths) (4)</th>
<th>Saturated (5)</th>
<th>Linear control (6)</th>
<th>Rounded (hundredths) (7)</th>
<th>Rounded (ten thousandths) (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Application covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of schools ranked</td>
<td>4.375</td>
<td>-0.341</td>
<td>0.107</td>
<td>0.065</td>
<td>0.052</td>
<td>0.012</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>Number of charter schools ranked</td>
<td>1.426</td>
<td>0.474</td>
<td>0.109</td>
<td>0.069</td>
<td>0.055</td>
<td>0.004</td>
<td>0.004</td>
<td>0.000</td>
</tr>
<tr>
<td>First school ranked is charter</td>
<td>0.290</td>
<td>0.616</td>
<td>0.004</td>
<td>0.004</td>
<td>0.001</td>
<td>-0.002</td>
<td>-0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>B. Baseline covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin school is charter</td>
<td>0.083</td>
<td>0.115</td>
<td>-0.017</td>
<td>-0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Female</td>
<td>0.521</td>
<td>-0.007</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.595</td>
<td>0.094</td>
<td>-0.010</td>
<td>-0.005</td>
<td>-0.005</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Black</td>
<td>0.182</td>
<td>-0.031</td>
<td>0.005</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Gifted</td>
<td>0.201</td>
<td>-0.022</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.005</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Bilingual</td>
<td>0.025</td>
<td>0.020</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Subsidized lunch</td>
<td>0.767</td>
<td>0.073</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Limited English proficient</td>
<td>0.290</td>
<td>0.084</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.005</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.001</td>
</tr>
<tr>
<td>Special education</td>
<td>0.087</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.003</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Baseline scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>0.017</td>
<td>0.010</td>
<td>-0.023</td>
<td>-0.019</td>
<td>-0.020</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td>Reading</td>
<td>0.034</td>
<td>-0.070</td>
<td>-0.016</td>
<td>-0.014</td>
<td>-0.014</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td>Writing</td>
<td>0.029</td>
<td>-0.056</td>
<td>-0.019</td>
<td>-0.016</td>
<td>-0.016</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
</tbody>
</table>

**Average risk set points of support**

|                      | 87 | 39 | 47 | 1,148 | 51 | 121 |

**Notes:** This table reports average covariate balance by charter offer status across 400 lottery draws, with DA rerun each time. Balance is estimated by regressing each covariate on an any-charter simulated offer dummy, controlling for the propensity score variables indicated in each column header. The table reports averages of these balance coefficients. The sample includes applicants for 2012-13 charter seats in grades 4-10 who were enrolled in Denver at baseline. The charter offer variable indicates an offer at any charter school, excluding alternative charters. Column 1 reports the baseline characteristics of charter applicants who did not receive a charter offer. The average risk set points of support reported at the bottom of the table count the average number of unique values found in the support of the relevant propensity score. Except for columns 3 and 6, this excludes values of zero and one. The estimates in columns 4, 7 and 8 use score values rounded as indicated in the column header; the estimates in column 5 control for every score value seen in the data.
<table>
<thead>
<tr>
<th>Application variable</th>
<th>No controls</th>
<th>Linear control</th>
<th>Simulated score</th>
<th>Full applicant type controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DA score (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Linear control</td>
<td>Nonparametric</td>
<td>Nonparametric</td>
<td>Nonparametric</td>
</tr>
<tr>
<td></td>
<td>Round (hundredths)</td>
<td>Saturated (ten thousandths)</td>
<td>Round (hundredths)</td>
<td>Saturated (ten thousandths)</td>
</tr>
<tr>
<td>Number of schools ranked</td>
<td>-0.341***</td>
<td>0.097</td>
<td>0.059</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.103)</td>
<td>(0.095)</td>
<td>(0.094)</td>
</tr>
<tr>
<td>Number of charter schools ranked</td>
<td>0.476***</td>
<td>0.143***</td>
<td>0.100**</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.052)</td>
<td>(0.047)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>First school ranked is charter</td>
<td>0.612***</td>
<td>0.012</td>
<td>0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.025)</td>
<td>(0.022)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>N</td>
<td>4,964</td>
<td>1,436</td>
<td>1,289</td>
<td>1,247</td>
</tr>
<tr>
<td>Risk set points of support</td>
<td>88</td>
<td>40</td>
<td>47</td>
<td>1,148</td>
</tr>
<tr>
<td>Robust F-test for joint significance</td>
<td>1190</td>
<td>2.70</td>
<td>1.70</td>
<td>1.09</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.044</td>
<td>0.165</td>
<td>0.352</td>
</tr>
</tbody>
</table>

Notes: This table reports coefficients from regressions of the application variables in each row on a dummy for charter offers. The sample includes applicants for 2012-13 charter seats in grades 4-10 who were enrolled in Denver at baseline. Columns 1-7 are from regressions like those used to construct expected balance in Table 4, except that the tests reported here use realized DA offers, with test statistics and standard errors computed in the usual way. Column 8 reports the balance test generated by a regression with saturated controls for applicant type (that is, unique combinations of applicant preferences over school programs and school priorities in those programs). Robust standard errors are reported in parentheses. P-values for robust joint significance tests are estimated by stacking outcomes and clustering at the student level.

*significant at 10%; **significant at 5%; ***significant at 1%
### Table 5b: Statistical tests for balance in student characteristics

<table>
<thead>
<tr>
<th>Student characteristics</th>
<th>No controls</th>
<th>Linear control</th>
<th>Simulated score</th>
<th>Propensity score controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3) (hundredths)</td>
<td>(4) (thousandths)</td>
</tr>
<tr>
<td>DA score (frequency)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin school is charter</td>
<td>0.108***</td>
<td>-0.051**</td>
<td>-0.037**</td>
<td>-0.029*</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.024)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.005</td>
<td>0.024</td>
<td>0.021</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.034)</td>
<td>(0.034)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.095***</td>
<td>-0.022</td>
<td>-0.013</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.031)</td>
<td>(0.028)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.033***</td>
<td>-0.002</td>
<td>-0.005</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.026)</td>
<td>(0.025)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Gifted</td>
<td>-0.028**</td>
<td>-0.026</td>
<td>-0.028</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Bilingual</td>
<td>0.023***</td>
<td>0.016</td>
<td>0.014</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Subsidized lunch</td>
<td>0.073***</td>
<td>-0.003</td>
<td>-0.004</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.027)</td>
<td>(0.025)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Limited English proficient</td>
<td>0.086***</td>
<td>-0.002</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Special education</td>
<td>0.004</td>
<td>0.034**</td>
<td>0.032*</td>
<td>0.032*</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Baseline scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>-0.002</td>
<td>-0.087</td>
<td>-0.083</td>
<td>-0.082</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.061)</td>
<td>(0.060)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>Reading</td>
<td>-0.085***</td>
<td>-0.096*</td>
<td>-0.100*</td>
<td>-0.108*</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.057)</td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Writing</td>
<td>-0.072***</td>
<td>-0.097*</td>
<td>-0.096*</td>
<td>-0.101*</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.056)</td>
<td>(0.054)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>N</td>
<td>4,964</td>
<td>1,436</td>
<td>1,289</td>
<td>1,247</td>
</tr>
<tr>
<td>Baseline scores: p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>0.000</td>
<td>0.278</td>
<td>0.329</td>
<td>0.454</td>
</tr>
<tr>
<td>Reading</td>
<td>19.1</td>
<td>1.20</td>
<td>1.13</td>
<td>0.99</td>
</tr>
<tr>
<td>Writing</td>
<td>1.04</td>
<td>1.35</td>
<td>0.71</td>
<td></td>
</tr>
</tbody>
</table>

See notes to Table 5a.
*significant at 10%; **significant at 5%; ***significant at 1%
Table 6: Comparison of 2SLS and semiparametric estimates of charter effects

<table>
<thead>
<tr>
<th></th>
<th>Frequency (saturated)</th>
<th>Formula (saturated)</th>
<th>Simulation (hundredths)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2SLS</td>
<td>Semiparametric</td>
<td>2SLS</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>0.496***</td>
<td>0.443***</td>
<td>0.524***</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.105)</td>
<td>(0.071)</td>
</tr>
<tr>
<td>Reading</td>
<td>0.127*</td>
<td>0.106</td>
<td>0.120*</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.107)</td>
<td>(0.073)</td>
</tr>
<tr>
<td></td>
<td>{0.065}</td>
<td>{0.069}</td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>0.325***</td>
<td>0.326***</td>
<td>0.356***</td>
</tr>
<tr>
<td></td>
<td>(0.079)</td>
<td>(0.102)</td>
<td>(0.082)</td>
</tr>
<tr>
<td></td>
<td>{0.077}</td>
<td>{0.080}</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1,102</td>
<td>1,093</td>
<td>1,083</td>
</tr>
</tbody>
</table>

Notes: This table compares 2SLS and semiparametric estimates of charter attendance effects on the 2012-13 TCAP scores of Denver 4th-10th graders. The instrument is an any-charter offer dummy. The semiparametric estimator is described in Section 3.5. In addition to score variables, 2SLS estimates include controls for grade tested, gender, origin school charter status, race, gifted status, bilingual status, subsidized price lunch eligibility, special education, limited English proficient status, and baseline test scores. Semiparametric models use these same variables as controls when computing the score weighting function. Standard errors in parentheses are from a Bayesian bootstrap. Conventional robust standard errors for 2SLS estimates are reported in braces.

*significant at 10%; **significant at 5%; ***significant at 1%
Table 7: Comparison of 2SLS and OLS estimates of charter attendance effects

<table>
<thead>
<tr>
<th>DA score</th>
<th>2SLS estimates</th>
<th>OLS with score controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (saturated)</td>
<td>Formula (saturated)</td>
</tr>
<tr>
<td>First stage</td>
<td>0.410*** (0.031)</td>
<td>0.389*** (0.032)</td>
</tr>
<tr>
<td>Math</td>
<td>0.496*** (0.071)</td>
<td>0.524*** (0.076)</td>
</tr>
<tr>
<td>Reading</td>
<td>0.127** (0.065)</td>
<td>0.120* (0.069)</td>
</tr>
<tr>
<td>Writing</td>
<td>0.325*** (0.077)</td>
<td>0.356*** (0.080)</td>
</tr>
<tr>
<td>N</td>
<td>1,102</td>
<td>1,083</td>
</tr>
</tbody>
</table>

Notes: This table compares 2SLS and OLS estimates of charter attendance effects using the same sample and instruments as for Table 6. The OLS estimates in column 6 are from a model that includes saturated control for frequency estimates of the DA score. In addition to score variables, all models include controls for grade tested, gender, origin school charter status, race, gifted status, bilingual status, subsidized price lunch eligibility, special education, limited English proficient status, and baseline test scores. Robust standard errors are reported in parentheses.

*significant at 10%; **significant at 5%; ***significant at 1%
<table>
<thead>
<tr>
<th>Offer instrument with DA score (frequency) controls (saturated)</th>
<th>First choice charter offer with risk set controls</th>
<th>Qualification instrument with risk set controls</th>
<th>Sample size increase for equivalent gain (col 2 vs col 1)</th>
<th>Sample size increase for equivalent gain (col 3 vs col 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First stage for charter offers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. First stage estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First stage for charter offers</td>
<td>1.000</td>
<td>0.774***</td>
<td>0.476***</td>
<td></td>
</tr>
<tr>
<td>First stage for charter enrollment</td>
<td>--</td>
<td>(0.026)</td>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>First stage for charter enrollment</td>
<td>0.410***</td>
<td>0.323***</td>
<td>0.178***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.035)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>B. 2SLS estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>0.496***</td>
<td>0.596***</td>
<td>0.409***</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>(0.102)</td>
<td>(0.149)</td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.127**</td>
<td>0.227**</td>
<td>0.229</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.102)</td>
<td>(0.144)</td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>0.325***</td>
<td>0.333***</td>
<td>0.505***</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.119)</td>
<td>(0.162)</td>
<td></td>
</tr>
<tr>
<td>N (students)</td>
<td>1,102</td>
<td>1,125</td>
<td>1,969</td>
<td></td>
</tr>
<tr>
<td>N (schools)</td>
<td>30</td>
<td>15</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table compares alternative 2SLS estimates of charter attendance effects using the same sample and control variables used to construct the estimates in Tables 6-7. Column 1 repeats the estimates using a DA offer instrument from column 1 in Table 7. The row labeled "First stage for charter offers" reports the coefficient from a regression of any-charter offer dummy (the instrument used in column 1) on other instruments, conditioning on the same controls used in the corresponding first stage estimates for charter enrollment. Column 2 reports 2SLS estimates computed using a first-choice charter offer instrument. Column 3 reports charter attendance effects computed using an any-charter qualification instrument. These alternative IV models control for risk sets making the first-choice and qualification instruments conditionally random; see Section 4.5 for details. Columns 4 and 5 report the multiples of the first-choice offer sample size and qualification sample size needed to achieve a precision gain equivalent to the gain from using the any-charter offer instrument. The last row counts the number of schools for which we observe in-sample variation in offer rates conditional on the score controls included in the model.

*significant at 10%; **significant at 5%; ***significant at 1%
Table 9: Enrollment destinies for charter applicants

<table>
<thead>
<tr>
<th></th>
<th>All charter applicants</th>
<th>Charter applicants with DA score (frequency) in (0,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No charter offer</td>
<td>Charter offer</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Enrolled in a study charter</td>
<td>0.147</td>
<td>0.865</td>
</tr>
<tr>
<td>... in a traditional public</td>
<td>0.405</td>
<td>0.081</td>
</tr>
<tr>
<td>... in an innovation school</td>
<td>0.234</td>
<td>0.023</td>
</tr>
<tr>
<td>... in a magnet school</td>
<td>0.192</td>
<td>0.021</td>
</tr>
<tr>
<td>... in an alternative school</td>
<td>0.009</td>
<td>0.005</td>
</tr>
<tr>
<td>... in a contract school</td>
<td>0.012</td>
<td>0.004</td>
</tr>
<tr>
<td>... in a non-study charter</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>N</td>
<td>2,555</td>
<td>1,833</td>
</tr>
</tbody>
</table>

Notes: This table describes school enrollment outcomes for charter applicants in the sample used to construct the estimates reported in Table 7. Columns 1-2 show enrollment by sector for all applicants without and with a charter offer. The remaining columns look only at those with a DA (frequency) score strictly between zero and one. Column 4 adds the non-offered mean in column 3 to the first stage estimate of the effect of charter offers on charter enrollment. School sectors are classified by grade. Innovation schools design and implement innovative practices to improve student outcomes. Magnet schools serve students with particular styles of learning. Alternative schools serve students struggling with academics, behavior, attendance, or other factors that may prevent them from succeeding in a traditional school environment; the latter offer faster pathways toward high school graduation, such as GED preparation and technical education. There is a single contract school, Escuela Tlatelolco, a private school contracted to serve DPS students, and a single non-study charter that closed in May 2013. Complier means in columns 5 and 6 were estimated using the 2SLS procedures described by Abadie(2002), with the same propensity score and covariate controls as were used to construct the estimates in Table 7.
Table 10: DPS charter and innovation school attendance effects

<table>
<thead>
<tr>
<th></th>
<th>DA score (frequency) controls (saturated)</th>
<th>Simulated score controls rounded (hundredths)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter and innovation</td>
<td>Charter and innovation</td>
</tr>
<tr>
<td></td>
<td>Charter only</td>
<td>Innovation only</td>
</tr>
<tr>
<td></td>
<td>Additive score controls</td>
<td>Additive score controls</td>
</tr>
<tr>
<td></td>
<td>Joint score controls</td>
<td>Joint score controls</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>(7)</td>
<td>(8)</td>
</tr>
<tr>
<td>Charter First Stage</td>
<td>0.410***</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>0.405***</td>
<td>0.398***</td>
</tr>
<tr>
<td>Innovation First Stage</td>
<td>0.348***</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>0.347***</td>
<td>0.348***</td>
</tr>
<tr>
<td>A. Math</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charter</td>
<td>0.496***</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.071)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>0.524***</td>
<td>0.517***</td>
</tr>
<tr>
<td>Innovation</td>
<td>--</td>
<td>-0.035</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td>(0.134)</td>
</tr>
<tr>
<td></td>
<td>0.177</td>
<td>0.286*</td>
</tr>
<tr>
<td>B. Reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charter</td>
<td>0.127**</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>0.076</td>
<td>0.072</td>
</tr>
<tr>
<td>Innovation</td>
<td>--</td>
<td>-0.285**</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.153)</td>
</tr>
<tr>
<td></td>
<td>-0.231</td>
<td>-0.190</td>
</tr>
<tr>
<td>C. Writing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charter</td>
<td>0.325***</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>0.357***</td>
<td>0.334***</td>
</tr>
<tr>
<td>Innovation</td>
<td>--</td>
<td>-0.119</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td>(0.148)</td>
</tr>
<tr>
<td></td>
<td>0.115</td>
<td>0.052</td>
</tr>
</tbody>
</table>

N 1,102 546 1,418 1,274 1,137 613 1,583 1,274

Notes: This table reports 2SLS estimates of charter and innovation attendance effects for applicants to schools in one or both sectors. The estimates for charter applicants in columns 1 and 5 are the same as reported in column 1 of Table 7. Columns 2 and 6 report innovation attendance effects for innovation applicants, estimated in models using an innovation offer instrument and innovation-specific saturated score controls constructed like those used for charter applicants. Columns 3 and 7 report coefficients from a two-endogenous-variable/two-instrument 2SLS model for the attendance effects of charters and innovations, conditioning additively on charter-specific and innovation-specific saturated score controls. Columns 4 and 8 show results from joint-effect models that add interactions between the two scores to the specification that generated column 7.

*significant at 10%; **significant at 5%; ***significant at 1%
3.6 Theoretical Appendix

3.6.1 Defining DA: Details

Our general formulation defines the DA match as determined by cutoffs found in the limit of a sequence. Recall that these cutoffs evolve according to

\[
\begin{align*}
    c_{s}^{t+1} &= \begin{cases} 
        K + 1 & \text{if } F(Q_{s}(c^t)) < q_s, \\
        \max \{ x \in [0, K + 1] \mid F(\{ i \in Q_{s}(c^t) \text{ such that } \pi_{is} \leq x \}) \leq q_s \} & \text{otherwise},
    \end{cases}
\end{align*}
\]

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

\[
Q_{s}(c^t) = \{ i \in I \mid \pi_{is} \leq c_{s}^t \text{ and } s \succ_i \tilde{s} \text{ for all } \tilde{s} \in S \text{ such that } \pi_{is} \leq c_{\tilde{s}}^t \}. \tag{3.16}
\]

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

**Proposition 3.** Consider an economy described by a distribution of students \( F \) and school capacities as defined in Section 3.3.1. Construct a sequence of cutoffs, \( c^t_s \), for this economy as described above. Then, \( \lim_{t \to \infty} c^t_s \) exists.

**Proof.** \( c^t_s \) is well-defined for all \( t \geq 1 \) and all \( s \in S \) since it is either \( K + 1 \) or the maximizer of a continuous function over a compact set. We will show by induction that \( \{ c^t_s \} \) is a decreasing sequence for all \( s \).

For the base case, \( c^2_s \leq c^1_s \) for all \( s \) since \( c^1_s = K + 1 \) and \( c^2_s \leq K + 1 \) by construction.

For the inductive step, suppose that \( c^t_s \leq c^{t-1}_s \) for all \( s \) and all \( t = 1, \ldots, T \). For each \( s \), if \( c^T_s = K + 1 \), then \( c^{T+1}_s \leq c^T_s \) since \( c^t_s \leq K + 1 \) for all \( t \) by construction. Otherwise, suppose to the contrary that \( c^{T+1}_s > c^T_s \). Since \( c^T_s < K + 1 \),

\[
F(\{ i \in Q_{s}(c^{T-1}) \text{ such that } \pi_{is} \leq c^T_s \}) = q_s. \tag{3.17}
\]

\[
F(\{ i \in Q_{s}(c^{T}) \text{ such that } \pi_{is} \leq c^T_s \}) = F(\{ i \in Q_{s}(c^{T}) \text{ such that } \pi_{is} \leq c^{T-1}_s \}) + F(\{ i \in Q_{s}(c^{T}) \text{ such that } c^T_s < \pi_{is} \leq c^{T+1}_s \}) \tag{3.18}
\]

\[
\geq q_s + F(\{ i \in Q_{s}(c^{T}) \text{ such that } c^T_s < \pi_{is} \leq c^{T+1}_s \}) \tag{3.19}
\]

\[
> q_s.
\]
Expression (3.17) follows because

\[ \{i \in Q_s(c^T) \text{ such that } \pi_{is} \leq c_{s}^{T} \} \]
\[ = \{i \in I \mid \pi_{is} \leq c_{s}^{T} \text{ and } s \succ_i \hat{s} \text{ for all } \hat{s} \in S \text{ such that } \pi_{i\hat{s}} \leq c_{\hat{s}}^{T} \} \]
\[ \supseteq \{i \in I \mid \pi_{is} \leq c_{s}^{T} \text{ and } s \succ_i \tilde{s} \text{ for all } \tilde{s} \in S \text{ such that } \pi_{i\tilde{s}} \leq c_{\tilde{s}}^{T-1} \} \quad \text{(by } c_{\tilde{s}}^{T} \leq c_{\tilde{s}}^{T-1}) \]
\[ = \{i \in Q_s(c^{T-1}) \text{ such that } \pi_{is} \leq c_{s}^{T} \}. \]

Expression (3.18) follows by the inductive assumption and since \( c_{s}^{T} < K + 1 \).

Expression (3.19) follows since if \( F(\{i \in Q_s(c^T) \text{ such that } c_{s}^{T} \leq \pi_{is} \leq c_{s}^{T+1} \}) = 0 \), then

\[ F(\{i \in Q_s(c^{T-1}) \text{ such that } \pi_{is} \leq c_{s}^{T+1} \}) = F(\{i \in Q_s(c^{T-1}) \text{ such that } \pi_{is} \leq c_{s}^{T} \}) \leq q_s, \]

while \( c_{s}^{T+1} > c_{s}^{T} \), contradicting the definition of \( c_{s}^{T} \).

Expression (3.19) contradicts the definition of \( c_{s}^{T+1} \) since the cutoff at step \( T + 1 \) results in an allocation that exceeds the capacity of school \( s \). This therefore establishes the inductive step that \( c_{s}^{T+1} \leq c_{s}^{T} \).

To complete the proof of the proposition, observe that since \( \{c'_{s}\} \) is a decreasing sequence in the compact interval \([0, K + 1]\), \( c'_{s} \) converges by the monotone convergence theorem.

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

\[ \max\{x \in [0, K + 1] \mid F(\{i \in Q_s(c') \text{ such that } \pi_{is} \leq x \}) \leq q_s \} \]
\[ = \max\{x \in [0, K + 1] \mid |\{j \in Q_s(c') : \pi_{js} \leq x \}| \leq k_s \}, \]

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

### 3.6.2 Conditional Independence of DA-generated Offers

**Proposition 4.** Let \( W_i \) be any variable that is independent of lottery numbers and write \( P[D_i(s) = 1|W_i, \theta] \) for \( P[D_i(s) = 1|W_i, \theta_i = \theta] \). Then

\[ P[D_i(s) = 1|W_i, \theta] = P[D_i(s) = 1|\theta]. \]
Proof. Suppose that DA converges at iteration $T$. Demand given cutoffs, $Q(c^T)$, a function of random numbers, preferences, and priorities, determines the distribution of offers. Equation (3.16) therefore implies that

$$P[D_i(s) = 1|W_i, c^T, \theta] = P[D_i(s) = 1|c^T, \theta].$$  \hfill (3.20)

Equation (3.20) does not contradict the fact that $c^T$ is determined in part by interactions between types and realized lottery numbers, interactions that may distort the distribution of lottery numbers conditional on cutoffs. In particular, (3.20) holds even if conditioning on $c^T$ makes the lottery number distribution depend on $\theta$. This follows from the fact that, as a consequence of the definition of demand for DA, offers and cutoffs are jointly independent of $W_i$ given $\theta$.\footnote{Using the shorthand notation $P[D_i(s), W_i, c^T, \theta]$ to denote joint probability statements and the associated conditionals without specifying realized values, we have:

$$P[D_i(s)|W_i, c^T, \theta] = \frac{P[D_i(s), c^T|W_i, \theta]}{P[c^T|W_i, \theta]} = \frac{P[D_i(s), c^T|\theta]}{P[c^T|\theta]} = P[D_i(s)|c^T, \theta].$$

Joint independence is used for the second equality.}

Note also that cutoffs satisfy

$$c^T \perp W_i|\theta,$$ \hfill (3.21)

a consequence of the fact that $c^T$ is the same for all $i$ in every lottery draw.

Finally, using (3.20), we have

$$P[D_i(s) = 1|W_i, \theta] = E\{E[D_i|W_i, \theta, c^T]|W_i, \theta\} = E\{E[D_i|c^T, \theta]|W_i, \theta\},$$

and the further implication by (3.21) that

$$E\{E[D_i|c^T, \theta]|W_i, \theta\} = E\{E[D_i|c^T, \theta]|\theta\} = P[D_i(s) = 1|\theta],$$

completing the proof for finite markets.

Extension to the continuum follows from the definition of $Q(c)$ for the limiting allocation in the continuum and the fact that the limiting cutoff in the continuum, $c$, is non-stochastic.

3.6.3 Proof of Theorem 2

Admissions cutoffs $c$ in a continuum economy are invariant to lottery outcomes $(r_i)$: DA in the continuum depends on $(r_i)$ only through $F(I_0)$ for sets $I_0 = \{i \in I | \theta_i \in \Theta_0\}$ with various choices of $\Theta_0$. In particular, $F(I_0)$ doesn’t depend on lottery realizations. Likewise, marginal priority $\rho_s$ is uniquely determined for every school $s$.\hfill $\square$
Consider the propensity score for school $s$. Students who don’t rank $s$ have $\varphi_s(\theta) = 0$. Among those who do rank $s$, those of type $\theta \in \Theta_s^a$ have $\rho_{\theta s} > \rho_s$. Therefore $\varphi_s(\theta) = 0$ for every $\theta \in \Theta_s^a \cup \Theta_s^c$.

Students of type $\theta \in \Theta_s^a \cup \Theta_s^c$ may be assigned $\tilde{s} \in B_{\theta s}$, where $\rho_{\theta s} = \rho_\tilde{s}$. Since lottery numbers are uniform, the proportion of type $\theta$ students assigned some $\tilde{s} \in B_{\theta s}$ where $\rho_{\theta s} = \rho_\tilde{s}$ is $MID_{\theta s}$. In other words, the probability of not being assigned any $\tilde{s} \in B_{\theta s}$ where $\rho_{\theta s} = \rho_\tilde{s}$ for a type $\theta$ student is $1 - MID_{\theta s}$. Every student of type $\theta \in \Theta_s^a$ who is not assigned a higher choice is assigned $s$ because $\rho_{\theta s} < \rho_s$, and so

$$\varphi_s(\theta) = (1 - MID_{\theta s}) \text{ for all } \theta \in \Theta_s^a.$$

Finally, consider students of type $\theta \in \Theta_s^c$ who are not assigned a higher choice. The fraction of students $\theta \in \Theta_s^c$ who are not assigned a higher choice is $1 - MID_{\theta s}$. Also, the random numbers of these students is larger than $MID_{\theta s}$. If $\tau_s < MID_{\theta s}$, then no such student is assigned $s$. If $\tau_s \geq MID_{\theta s}$, then the ratio of students that are assigned $s$ within this set is given by $\frac{\tau_s - MID_{\theta s}}{1 - MID_{\theta s}}$. Hence, conditional on $\theta \in \Theta_s^c$ and not being assigned a choice higher than $s$, the probability of being assigned $s$ is given by $\max\{0, \frac{\tau_s - MID_{\theta s}}{1 - MID_{\theta s}}\}$. Therefore,

$$\varphi_s(\theta) = (1 - MID_{\theta s}) \times \max\left\{0, \frac{\tau_s - MID_{\theta s}}{1 - MID_{\theta s}}\right\} \text{ for all } \theta \in \Theta_s^c.$$

### 3.6.4 Proof of Theorem 3

We complete the proof of Theorem 3 in Section 3.3.3 by proving the following two intermediate results.

**Lemma 13.** (Cutoff almost sure convergence) $\hat{c}_n \overset{a.s.}{\rightarrow} c$.

**Lemma 14.** (Propensity score almost sure convergence) For all $\theta \in \Theta$ and $s \in S$, $p_n(\theta) \overset{a.s.}{\rightarrow} \varphi_s(\theta)$.

**Proof of Lemma 13**

We use the Extended Continuous Mapping Theorem (Theorem 19.1 in van der Vaart (2000)) to prove the lemma. We first show deterministic convergence of cutoffs in order to verify the assumptions of the theorem.

Modify the definition of $F$ to describe the distribution of lottery numbers as well types: For any set of student types $\Theta_0 \subset \Theta$ and for any numbers $r_0, r_1 \in [0, 1]$ with $r_0 < r_1$, define the set of students of types in $\Theta_0$ with random numbers worse than $r_0$ and better than $r_1$ as

$$I(\Theta_0, r_0, r_1) = \{ i \in I \mid \theta_i \in \Theta_0, r_0 < \tau_i \leq r_1 \}.$$
In a continuum economy,

\[ F(I(\Theta_0, r_0, r_1)) = E[1\{\theta_i \in \Theta_0\}] \times (r_1 - r_0), \]

where the expectation is assumed to exist. In a finite economy with \( n \) students,

\[ F(I(\Theta_0, r_0, r_1)) = \frac{|I(\Theta_0, r_0, r_1)|}{n}. \]

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

\[ d(F, F') = \sup_{\Theta_0 \in \Theta_0, r_0, r_1 \in [0, 1]} |F(I(\Theta_0, r_0, r_1)) - F'(I(\Theta_0, r_0, r_1))|. \]

The notation is otherwise as in the text.

**Proof.** Consider a deterministic sequence of economies described by a sequence of distributions \( \{f_n\} \) over students, together with associated school capacities, so that for all \( n \), \( f_n \in \mathcal{F} \) is a potential realization produced by randomly drawing \( n \) students and their lottery numbers from \( F \). Assume that \( f_n \to F \) in metric space \( (\mathcal{F}, d) \). Let \( c_n \) denote the admissions cutoffs in \( f_n \). Note the \( c_n \) is constant because this is the cutoff for a particular realized economy \( f_n \).

The proof first shows deterministic convergence of cutoffs for any convergent subsequence of \( f_n \). Let \( \{\tilde{f}_n\} \) be a subsequence of realized economies \( \{f_n\} \). The corresponding cutoffs are denoted \( \{\tilde{c}_n\} \). Let \( \tilde{c} \equiv (\tilde{c}_n) \) be the limit of \( \tilde{c}_n \). The following two claims establish that \( \tilde{c}_n \to c \), the cutoff associated with \( F \).

**Claim 2.** \( \tilde{c}_s \geq c_s \) for every \( s \in S \).

**Proof of Claim 2.** This is proved by contradiction in 3 steps. Suppose to the contrary that \( \tilde{c}_s < c_s \) for some \( s \). Let \( S' \subset S \) be the set of schools the cutoffs of which are strictly lower under \( \tilde{c} \). For any \( s \in S' \), define \( I^s_n = \{i \in I|c_n \leq i \leq c_s \text{ and } i \text{ ranks } s \text{ first}\} \) where \( I \) is the set of students in \( F \), which contains the set of students in \( f_n \) for all \( n \). In other words, \( I^s_n \) are the set of students ranking school \( s \) first who have a student rank in between \( c_{ns} \) and \( c_s \).

**Step (a):** We first show that for our subsequence, when the market is large enough, there must be some students who are in \( I^s_n \). That is, there exists \( N \) such that for any \( n > N \), we have \( \tilde{f}_n(I^s_n) > 0 \) for all \( s \in S' \).
To see this, we begin by showing that for all $s \in S'$, there exists $N$ such that for any $n > N$, we have $F(I_n^s) > 0$. Suppose, to the contrary, that there exists $s \in S'$ such that for all $N$, there exists $n > N$ such that $F(I_n^s) = 0$. When we consider the subsequence of realized economies $\{f_n\}$, we find that

$$
\tilde{f}_n(\{i \in Q_s(c_n) \text{ such that } \pi_is \leq c_s\})
= \tilde{f}_n(\{i \in Q_s(c_n) \text{ such that } \pi_is \leq \tilde{c}_ns\}) + \tilde{f}_n(\{i \in Q_s(c_n) \text{ such that } \tilde{c}_ns < \pi_is \leq c_s\})
\leq q_s.
$$

Expression (3.22) follows from Assumption 1 by the following reason. (3.22) does not hold, i.e., $\tilde{f}_n(\{i \in Q_s(c_n) \text{ such that } \pi_is < \pi_is \leq c_s\}) > 0$ only if $F(\{i \in I|\tilde{c}_ns < \pi_is \leq c_s\}) > 0$. This and Assumption 1 imply $F(\{i \in I|\tilde{c}_ns < \pi_is \leq c_s \text{ and } i \text{ ranks } s \text{ first}\}) \equiv F(I_n^s) > 0$, a contradiction to $F(I_n^s) = 0$. Since $\tilde{f}_n$ is realized as $n$ iid samples from $F$, $\tilde{f}_n(\{i \in I|\tilde{c}_ns < \pi_is \leq c_s\}) = 0$. Expression (3.23) follows by our definition of DA, which can never assign more students to a school than its capacity for each of the $n$ samples. We obtain our contradiction since $\tilde{c}_ns$ is not maximal at $s$ in $\tilde{f}_n$ since expression (3.23) means it is possible to increase the cutoff $\tilde{c}_ns$ to $c_s$ without violating the capacity constraint.

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

Step (a) allows us to find some $N$ such that for any $n > N$, $\tilde{f}_n(I_n^s) > 0$ for all $s' \in S'$. Let $\tilde{s}_n \in S$ and $t$ be such that $\tilde{c}_ns_n^{-1} \geq c_s$ for all $s \in S$ and $\tilde{c}_n\tilde{s}_n < c_s$. That is, $\tilde{s}_n$ is one of the first schools the cutoff of which falls strictly below $c_{\tilde{s}_n}$ under the DA algorithm in $\tilde{f}_n$, which happens in round $t$ of the DA algorithm. Such $\tilde{s}_n$ and $t$ exist since the choice of $n$ guarantees $\tilde{f}_n(I_n^s) > 0$ and so $\tilde{c}_ns < c_s$ for all $s \in S'$.

**Step (b):** We next show that there exist infinitely many values of $n$ such that the associated $\tilde{s}_n$ is in $S'$ and $\tilde{f}_n(I_n^s) > 0$ for all $s \in S'$. It is because otherwise, by Step (a), there exists $N$ such that for all $n > N$, we have $\tilde{s}_n \not\in S'$. Since there are only finitely many schools, $\{\tilde{s}_n\}$ has a subsequence $\{\tilde{s}_m\}$ such that $\tilde{s}_m$ is the same school outside $S'$ for all $m$. By definition of $\tilde{s}_n$, $\tilde{c}_m\tilde{s}_m \leq c_m\tilde{s}_m < c_{\tilde{s}_m}$ for all $m$ and so $\tilde{c}_{\tilde{s}_m} < c_{\tilde{s}_m}$, a contradiction to $\tilde{s}_m \not\in S'$. Therefore, we have our desired conclusion of Step (b).
Fix some \( n \) such that the associated \( \tilde{s}_n \) is in \( S' \) and \( \tilde{f}_n(l_n^s) > 0 \) for all \( s \in S' \). Step (b) guarantees that such \( n \) exists. Let \( \tilde{A}_{n\tilde{s}_n} \) and \( A_{\tilde{s}_n} \) be the sets of students assigned \( \tilde{s}_n \) under \( \tilde{f}_n \) and \( F \), respectively. All students in \( I_n^{\tilde{s}_n} \) are assigned \( \tilde{s}_n \) in \( F \) and rejected by \( \tilde{s}_n \) in \( \tilde{f}_n \). Since these students rank \( \tilde{s}_n \) first, there must exist a positive measure (with respect to \( \tilde{f}_n \)) of students outside \( A_{\tilde{s}_n} \) who are assigned \( \tilde{s}_n \) in \( \tilde{f}_n \) and some other school in \( F \); denote the set of them by \( A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n} \). If \( \tilde{f}_n(A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n}) > 0 \) since otherwise, for any \( n \) such that Step (b) applies, \( f_n(A_{n\tilde{s}_n}) - f_n(A_{\tilde{s}_n}) = f_n(I_n^{\tilde{s}_n}) \),

which by Step (a) converges to something strictly smaller than \( F(A_{\tilde{s}_n}) \) since \( f_n(A_{\tilde{s}_n}) \to F(A_{\tilde{s}_n}) \) and \( f_n(I_n^{\tilde{s}_n}) > 0 \) for all large enough \( n \) by Step (a). Note that \( F(A_{\tilde{s}_n}) \) is weakly smaller than \( q_{\tilde{s}_n} \). This implies that for large enough \( n \), \( \tilde{f}_n(A_{n\tilde{s}_n}) < q_{\tilde{s}_n} \), a contradiction to \( A_{n\tilde{s}_n} \)’s being the set of students assigned \( \tilde{s}_n \) at a cutoff strictly smaller than the largest possible value \( K + 1 \). For each \( i \in A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n} \), let \( s_i \) be the school to which \( i \) is assigned under \( F \).

**Step (c):** To complete the argument for Claim 2, we show that some \( i \in A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n} \) must have been rejected by \( s_i \) in some step \( t \leq t - 1 \) of the DA algorithm in \( \tilde{f}_n \). That is, there exists \( i \in A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n} \) and \( t \leq t - 1 \) such that \( \pi_{i,s_i} > \tilde{c}_n^{s_i} \). Suppose to the contrary that for all \( i \in A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n} \) and \( t \leq t - 1 \), we have \( \pi_{i,s_i} \leq \tilde{c}_n^{s_i} \). Each such student \( i \) must prefer \( s_i \) to \( \tilde{s}_n \) because \( i \) is assigned \( s_i \neq \tilde{s}_n \) under \( F \) though \( \pi_{i,s_i} \leq \tilde{c}_n^{s_i} < c_{\tilde{s}_n} \), where the first inequality holds because \( i \) is assigned \( \tilde{s}_n \) in \( F \), while the second inequality does because \( \tilde{s}_n \in S' \). This implies none of \( A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n} \) is rejected by \( s_i \), applies for \( \tilde{s}_n \), and contributes to decreasing \( \tilde{c}_n^{s_i} \) at least until step \( t \) and so \( \tilde{c}_n^{s_i} < c_{\tilde{s}_n} \) cannot be the case, a contradiction. Therefore, we have our desired conclusion of Step (c).

Claim 2 can now be established by showing that Step (c) implies there are \( i \in A_{n\tilde{s}_n} \setminus A_{\tilde{s}_n} \) and \( t \leq t - 1 \) such that \( \pi_{i,s_i} > \tilde{c}_n^{s_i} \geq \tilde{c}_n^{s_i} \), where the last inequality is implied by the fact that in every economy, for all \( s \in S \) and \( t \geq 0 \), we have \( c_{s}^{t+1} \leq c_{s}^{t} \). Also, they are assigned \( s_i \) in \( F \) so that \( \pi_{i,s_i} \leq c_{s_i} \). These imply \( c_{s_i} > \tilde{c}_n^{s_i} \). That is, the cutoff of \( s_i \) falls below \( c_{s_i} \) in step \( t \leq t - 1 < t \) of the DA algorithm in \( \tilde{f}_n \). This contradicts the definition of \( \tilde{s}_n \) and \( t \). Therefore \( \tilde{c}_n \geq c_{s} \) for all \( s \in S \), as desired.

**Claim 3.** By a similar argument, \( \tilde{c}_s \leq c_{s} \) for every \( s \in S \).

Since \( \tilde{c}_s \geq c_{s} \) and \( \tilde{c}_s \leq c_{s} \) for all \( s \), it must be the case that \( \tilde{c}_n \to c \). The following claim uses this to show that \( c_n \to c \).

**Claim 4.** If \( \tilde{c}_n \to c \) for every convergent subsequence \( \{\tilde{c}_n\} \) of \( \{c_n\} \), then \( c_n \to c \).
Proof of Claim 4. Since \( \{c_n\} \) is bounded in \([0, K + 1]\), it has a convergent subsequence by the Bolzano-Weierstrass theorem. Suppose to the contrary that for every convergent subsequence \( \{\tilde{c}_n\} \), we have \( \tilde{c}_n \to c \), but \( c_n \not\to c \). Then there exists \( \varepsilon > 0 \) such that for all \( k > 0 \), there exists \( n_k > k \) such that \( \|c_{n_k} - c\| \geq \varepsilon \). Then the subsequence \( \{c_{n_k}\}_{k} \subseteq \{c_n\} \) has a convergent subsequence that does not converge to \( c \) (since \( \|c_{n_k} - c\| \geq \varepsilon \) for all \( k \)), which contradicts the supposition that every convergent subsequence of \( \{c_n\} \) converges to \( c \). \( \square \)

The last step in the proof of Lemma 13 relates this fact to stochastic convergence.

Claim 5. \( c_n \to c \) implies \( \hat{c}_n \overset{a.s.}{\to} c \)

Proof of Claim 5. This proof is based on two off-the-shelf asymptotic results from mathematical statics. First, let \( F_n \) be the distribution over \( I(\Theta_0, r_0, r_1)'s \) generated by randomly drawing \( n \) students from \( F \). Note that \( F_n \) is random since it involves randomly drawing \( n \) students. \( F_n \overset{a.s.}{\to} F \) by the Glivenko-Cantelli theorem (Theorem 19.1 in van der Vaart (2000)). Next, since \( F_n \overset{a.s.}{\to} F \) and \( c_n \to c \), the Extended Continuous Mapping Theorem (Theorem 18.11 in van der Vaart (2000)) implies that \( \hat{c}_n \overset{a.s.}{\to} c \), completing the proof of Lemma 13. \( \square \)

Proof of Lemma 14

Proof. Consider any deterministic sequence of economies \( \{f_n\} \) such that \( f_n \in F \) for all \( n \) and \( f_n \to F \) in the \((F, d)\) metric space. Let \( p_n(\theta) \) be the (finite-market, deterministic) propensity score for a particular \( f_n \). Note that this subtly modifies the definition of \( p_n(\theta) \) from that in the text. The change here is that the propensity score for \( f_n \) is not a random quantity, because economy \( f_n \) is viewed as fixed.

For Lemma 14, it is enough to show deterministic convergence of this finite-market score, that is, \( p_n(\theta) \to \varphi_s(\theta) \) as \( f_n \to F \). To see this, let \( F_n \) be the distribution over \( I(\Theta_0, r_0, r_1)'s \) induced by randomly drawing \( n \) students from \( F \). Note that \( F_n \) is random and that \( F_n \overset{a.s.}{\to} F \) by the Glivenko-Cantelli theorem (Theorem 19.1 in van der Vaart (2000)). \( F_n \overset{a.s.}{\to} F \) and \( p_n(\theta) \to \varphi_s(\theta) \) allow us to apply the Extended Continuous Mapping Theorem (Theorem 18.11 in van der Vaart (2000)) to obtain \( \tilde{p}_n(\theta) \overset{a.s.}{\to} \varphi_s(\theta) \).

We prove convergence of \( p_n(\theta) \to \varphi_s(\theta) \) as follows. Let \( \tilde{c}_{ns} \) and \( \tilde{c}_{ns'} \) be the random cutoffs at \( s \) and \( s' \), respectively, in \( f_n \), and

\[
\begin{align*}
\tau_{\theta s} &\equiv c_s - \rho_{\theta s}, \\
\tau_{\theta s'} &\equiv \max_{s' > \theta s} \{c_{s'} - \rho_{\theta s'}\}, \\
\tilde{c}_{\theta s} &\equiv \tilde{c}_{ns} - \rho_{\theta s}, \text{ and} \\
\tilde{c}_{\theta s'} &\equiv \max_{s' > \theta s} \{\tilde{c}_{ns'} - \rho_{\theta s'}\}.
\end{align*}
\]
We can express $\varphi_s(\theta)$ and $p_{ns}(\theta)$ as follows.

$$\varphi_s(\theta) = \max\{0, \tau_{\theta s} - \tau_{\theta s_-}\}$$

$$p_{ns}(\theta) = P_n(\tau_{\theta s} \geq R > \tau_{\theta s_-})$$

where $P_n$ is the probability induced by randomly drawing lottery numbers given $f_n$, and $R$ is any type $\theta$ student’s random lottery number distributed according to $U[0,1]$. By Lemma 13, with probability 1, for all $\epsilon > 0$, there exists $N_1$ such that for all $n > N_1$,

$$|\tilde{c}_{ns'} - c_{s'}| < \epsilon_1$$

for all $s'$,

which implies that with probability 1,

$$|\tilde{\tau}_{n\theta s_-} - \tau_{\theta s_-}|$$

$$= |\{\tilde{c}_{ns_1} - \rho_{\theta s_1}\} - \{c_{s_2} - \rho_{\theta s_2}\}|$$

$$< \left\{ \begin{array}{ll}
|\{\tilde{c}_{ns_1} - \rho_{\theta s_1}\} - \{\tilde{c}_{ns_1} - \rho_{\theta s_2}\} + \epsilon_1| & \text{if } c_{s_2} - \rho_{\theta s_2} \geq \tilde{c}_{ns_1} - \rho_{\theta s_1} \\
|\{\tilde{c}_{ns_1} - \rho_{\theta s_1}\} - \{\tilde{c}_{ns_1} - \rho_{\theta s_2}\} - \epsilon_1| & \text{if } c_{s_2} - \rho_{\theta s_2} < \tilde{c}_{ns_1} - \rho_{\theta s_1}
\end{array} \right.$$  

$$= \epsilon_1$$

where in the first equality, $s_1 \equiv \arg\max_{s' > \theta s'}\{\tilde{c}_{ns'} - \rho_{\theta s'}\}$ and $s_2 \equiv \arg\max\{c_{s'} - \rho_{\theta s'}\}$. The inequality is by $|\tilde{c}_{ns'} - c_{s'}| < \epsilon_1$ for all $s'$. For all $\epsilon > 0$, the above argument with setting $\epsilon_1 < \epsilon/2$ implies that there exists $N$ such that for all $n > N$,

$$p_{ns}(\theta)$$

$$= P_n(\tilde{\tau}_{n\theta s} \geq R > \tilde{\tau}_{n\theta s_-})$$

$$\in (\max\{0, \tau_{\theta s} - \tau_{\theta s_-} - \epsilon, \max\{0, \tau_{\theta s} - \tau_{\theta s_-} + \epsilon\})$$

$$\in (\varphi_s(\theta) - \epsilon, \varphi_s(\theta) + \epsilon),$$

where the second-to-last inclusion is because with probability 1, there exists $N$ such that for all $n > N$ such that $|\tilde{\tau}_{n\theta s} - \tau_{\theta s}|, |\tilde{\tau}_{n\theta s_-} - \tau_{\theta s_-}| < \epsilon_1$ and $R \sim U[0,1]$. This means $p_{ns}(\theta) \rightarrow \varphi_s(\theta)$, completing the proof of Lemma 14.

### 3.6.5 First Choice and Qualification Instruments: Details

Let $D_i^f$ be the first choice instrument defined in section 3.4.5 and let $\tilde{s}_i$ be $i$’s first choice school. The first choice risk set is $Q(\tilde{s}_i) \equiv (\tilde{s}_i, \rho_{\tilde{s}_i})$.

**Proposition 3.** In any continuum economy, $D_i^f$ is independent of $\theta_i$ conditional on $R(\theta_i)$.

**Proof.** In general,

$$\Pr(D_i^f = 1|\theta_i = \theta)$$
\[
\begin{align*}
&= \Pr(\pi_{i,s} \leq c_{s_i} | \theta_i = \theta) \\
&= \Pr(\rho_{i,s} + r_i \leq c_{s_i} | \theta_i = \theta) \\
&= \Pr(r_i \leq c_{s_i} - \rho_{i,s} | \theta_i = \theta) \\
&= c_{s_i} - \rho_{i,s},
\end{align*}
\]

which depends on \( \theta_i \) only through \( R(\theta_i) \) because cutoffs are fixed in the continuum. \( \square \)

Let \( D_i^q \) and \( R(\theta_i) \) be the qualification instrument and the associated risk set defined in section 3.4.5. The latter is given by the list of schools \( i \) ranks and his priority status at each, that is, \( R(\theta_i) \equiv (S_i, (\rho_{is})_{s \in S_i}) \) where \( S_i \) is the set of charter schools \( i \) ranks.

**Proposition 4.** In any continuum economy, \( D_i^q \) is independent of \( \theta_i \) conditional on \( R(\theta_i) \).

**Proof.** In general, we have

\[
\begin{align*}
&\Pr(D_i^q = 1 | \theta_i = \theta) \\
&= \Pr(\pi_{i,s} \leq c_s \text{ for some } s \in S_i | \theta_i = \theta) \\
&= \Pr(\rho_{i,s} + r_i \leq c_s \text{ for some } s \in S_i | \theta_i = \theta) \\
&= \Pr(r_i \leq c_s - \rho_{i,s} \text{ for some } s \in S_i | \theta_i = \theta) \\
&= \Pr(r_i \leq \max_{s \in S_i} (c_s - \rho_{i,s}) | \theta_i = \theta) \\
&= \max_{s \in S_i} (c_s - \rho_{i,s}),
\end{align*}
\]

which depends on \( \theta_i \) only through \( R(\theta_i) \) because cutoffs are fixed in the continuum. \( \square \)

### 3.6.6 Extension to a General Lottery Structure

Washington DC, New Orleans, and Amsterdam use DA with multiple lottery numbers, one for each school (see, for example, de Haan et al. (2015)). Washington, DC uses a version of DA that uses a mixture of shared and individual school lotteries. This section derives the DA propensity score for a mechanism with any sort of multiple tie-breaking.

Let a random variable \( R_{is} \) denote student \( i \)'s lottery number at school \( s \). Assume that each \( R_{is} \) is drawn from \( U[0, 1] \), independently with schools. We consider a general lottery structure where \( R_{is} \neq R_{is'} \) for some (not necessarily all) \( s, s' \in S \) and \( i \in I \).

Recall \( B_{\theta s} \) is defined as \( \{s' \in S | s' \succ_\theta s\} \). Partition \( B_{\theta s} \) into \( m \) disjoint sets \( B_{\theta s}^1, ..., B_{\theta s}^m \), so that \( s' \) and \( s'' \) use the same lottery if and only if \( s', s'' \in B_{\theta s}^m \) for some \( m \). Note that this partition is specific to type \( \theta \). With single-school lotteries, \( m \) simplifies to \( |B_{\theta s}| \), the number of schools type \( \theta \) ranks ahead of \( s \).

The **most informative disqualification**, \( MID_{\theta s}^m \), is defined for each \( m \) as

\[
MID_{\theta s}^m = \begin{cases} 
0 & \text{if } \rho_{\theta s} > \rho_s \text{ for all } \tilde{s} \in B_{\theta s}^m, \\
1 & \text{if } \rho_{\theta s} < \rho_s \text{ for some } \tilde{s} \in B_{\theta s}^m, \\
\max\{\tau_{\tilde{s}} | \tilde{s} \in B_{\theta s}^m \text{ and } \rho_{\theta \tilde{s}} = \rho_s \} & \text{if } \rho_{\theta s} = \rho_s \text{ for some } \tilde{s} \in B_{\theta s}^m \text{ and } \rho_{\theta \tilde{s}} > \rho_s \text{ otherwise.}
\end{cases}
\]

Let \( m^* \) be the value of \( m \) for schools in the partition that use the same lottery as \( s \). Denote the associated \( MID \) by \( MID_{\theta s}^{m^*} \). We define \( MID_{\theta s}^0 = 0 \) when the lottery at \( s \) is
unique and there is no $m^*$. The following result extends Theorem 2 to a general lottery structure. The proof is omitted.

**Theorem 2 (Generalization).** For all $s$ and $\theta$ in any continuum economy, we have:

$$
\Pr[D_i(s) = 1 | \theta_i = \theta] = \varphi_s(\theta) = \begin{cases} 
0 & \text{if } \theta \in \Theta_s^n, \\
\Pi_{m=1}^{\overline{m}}(1 - MID_{\theta_s}^m) & \text{if } \theta \in \Theta_s^a, \\
\Pi_{m=1}^{\overline{m}}(1 - MID_{\theta_s}^m) \times \max \left\{ 0, \frac{\tau_s - MID_{\theta_s}^s}{1 - MID_{\theta_s}^s} \right\} & \text{if } \theta \in \Theta_s^c.
\end{cases}
$$

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

Note that in the single tie breaker case, the expression for $\varphi_s(\theta)$ reduces to that in Theorem 2 since $\overline{m} = 1$ in that case.

### 3.6.7 The Boston (Immediate Acceptance) Mechanism

Studies by Hastings-Kane-Staiger (2009), Hastings-Neilson-Zimmerman (2012), and Deming-Hastings-Kane-Staiger (2013), among others, use data generated from versions of the Boston mechanism. Given strict preferences of students and schools, the Boston mechanism is defined as follows:

- **Step 1:** Each student applies to her most preferred acceptable school (if any). Each school accepts its most-preferred students up to its capacity and rejects every other student.

In general, for any step $t \geq 2$,

- **Step $t$:** Each student who has not been accepted by any school applies to her most preferred acceptable school that has not rejected her (if any). Each school accepts its most-preferred students up to its remaining capacity and rejects every other student.

This algorithm terminates at the first step in which no student applies to a school. Boston assignments differ DA in that any offer at any step is fixed; students receiving offers cannot be displaced later.

This important difference notwithstanding, the Boston mechanism can be represented as a special case of DA by redefining priorities as follows:

**Proposition 5.** (Ergin and Sönmez (2006)) The Boston mechanism applied to $(\succ_i)_i$ and $(\succ_s)_s$ produces the same assignment as DA applied to $(\succ_i)_i$ and $(\succ_s^*)_s$ where $\succ_s^*$ is defined as follows:
(1) For \( k = 1, 2, \ldots \), \( \{ \text{students who rank } s \text{ } k\text{-th} \} \succ_s^* \{ \text{students who rank } s \text{ } k + 1\text{-th} \} \\

(2) Within each category, \( \succ_s^* \) ranks the students in the same order as original \( \succ_s \).

This equivalence allows us to construct a Boston propensity score by redefining priorities so that priority groups at a given school consists of applicants who (i) share the same original priority status at the school and (ii) give the same rank to the school.

3.7 Empirical Appendix

3.7.1 Data

The Denver Public Schools (DPS) analysis file is constructed using application, school assignment, enrollment, demographic, and outcome data provided by DPS for school years 2011-2012 and 2012-2013. All files are de-identified, but students can be matched across years and files. Applicant data are from the 2012-2013 SchoolChoice assignment file and test score data are from the CSAP (Colorado Student Assessment Program) and the TCAP (Transitional Colorado Assessment Program) files. The CSAP was discontinued in 2011, and was replaced by the TCAP beginning with the 2012-2013 school year. Enrollment, demographic, and outcome data are available for students enrolled in DPS only; enrollment data are for October.

Applications and assignment: The SchoolChoice file

The 2012-2013 SchoolChoice assignment file contains information on applicants’ preferences over schools (school rankings), school priorities over applicants, applicants’ school assignments (offers) and lottery numbers, a flag for whether the applicant is subject to the family link policy described in the main text and, if so, to which sibling the applicant is linked. Each observation in the assignment file corresponds to an applicant applying for a seat in programs within schools known as a bucket.\(^{29}\) Each applicant receives at most one offer across all buckets at a school. Information on applicant preferences, school priorities, lottery numbers, and offers are used to compute the DA propensity score and the simulated propensity score.

Appendix Table B1 describes the construction of the analysis sample starting from all applicants in the 2012-2013 SchoolChoice assignment file. Out of a total of 25,687

\(^{29}\)Since applicants’ rankings are at the school-level but seats are assigned at the bucket level, the SchoolChoice assignment mechanism translates school-level rankings into bucket-level rankings. For example, if an applicant ranked school A first and school B second, and if all seats at both A and B are split into two categories, one for faculty children (“Faculty”) and one for any type of applicant (“Any”), then the applicant’s ranking of the programs at A and B would be listed as 10 for Faculty at A, 11 for Any at A, 20 for Faculty at B, 21 for Any at B where numbers code preferences (smaller is more preferred).
applicants seeking a seat in DPS in the academic year 2012-2013, 5,669 applied to any charter school seats in grades 4 through 10. We focus the analysis on applicants to grades 4 through 10 because baseline grade test scores are available for these grades only. We further limit the sample to 4,964 applicants who were enrolled in DPS in the baseline grade (the grade prior to the application grade) in the baseline year (2011-2012), for whom baseline enrollment demographic characteristics are available.

**Enrollment and demographic characteristics**

Each observation in the enrollment files describes a student enrolled in a school in a year, and includes information on grade attended, student sex, race, gifted status, bilingual status, special education status, limited English proficiency status, and subsidized lunch eligibility. 30 Demographic and enrollment information are from the first calendar year a student spent in each grade.

**Applicant outcomes: CSAP/TCAP**

Test scores and proficiency levels for the CSAP/TCAP math, reading, and writing exams are available for grades 3 through 10. Each observation in the CSAP/TCAP data file corresponds to a student’s test results in a particular subject, grade, and year. For each grade, we use scores from the first attempt at a given subject test, and exclude the lowest obtainable scores as outliers. As a result, 41 observed math scores, 19 observed reading scores, and 1 observed writing score are excluded from the sample of charter applicants that are in DPS in baseline year. After outlier exclusion, score variables are standardized to have mean zero and unit standard deviation in a subject-grade-year in the DPS district.

**School classification: Parent Guide**

We classify schools as charters, traditional public schools, magnet schools, innovation schools, contract schools, or alternative schools (i.e. intensive pathways and multiple pathways schools) according to the 2012-2013 Denver SchoolChoice Parent Guides for Elementary and Middle Schools and High Schools. School classification is by grade, since some schools run magnet programs for a few grades only. Schools not included in the Parent Guide (i.e. SIMS Fayola International Academy Denver) were classified according to information from the school’s website.

---

30 Race is coded as black, white, asian, hispanic, and other. In DPS these are mutually-exclusive categories.
<table>
<thead>
<tr>
<th>Applicants to</th>
<th>Applicants</th>
<th>Types</th>
<th>Applicants</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>All applicants</td>
<td>25,687</td>
<td>16,087</td>
<td>15,487</td>
<td>9,564</td>
</tr>
<tr>
<td>Applicants to grades 4 through 10</td>
<td>12,507</td>
<td>7,480</td>
<td>10,898</td>
<td>6,642</td>
</tr>
<tr>
<td>Applicants to any charters (grades 4 through 10)</td>
<td>5,669</td>
<td>4,833</td>
<td>4,964</td>
<td>4,282</td>
</tr>
</tbody>
</table>

Notes: All applications are for the 2012-2013 academic year. Columns 1 and 2 include all applicants in the SchoolChoice assignment file (see Data Appendix for details). Columns 3 and 4 exclude applicants who were not in DPS at the baseline grade (the grade prior to application grade) in baseline year (2011-2012). Applicants to grade "EC" (early childhood, or pre-kindergarten) are excluded from columns 3 and 4 because there is no baseline grade for those applicants. Columns 2 and 4 count unique combinations of applicant preferences over school programs and school priorities in those programs.
<table>
<thead>
<tr>
<th>Non-offered mean</th>
<th>No controls</th>
<th>Linear control</th>
<th>Rounded (hundredths)</th>
<th>Saturated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>A. DA score (frequency)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolled in DPS in follow-up year</td>
<td>0.905</td>
<td>0.029***</td>
<td>0.041**</td>
<td>0.040**</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Has scores in follow-up year</td>
<td>0.881</td>
<td>0.032***</td>
<td>0.050**</td>
<td>0.049**</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>N</td>
<td>2,939</td>
<td>4,964</td>
<td>1,436</td>
<td>1,289</td>
</tr>
<tr>
<td>B. DA score (formula)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolled in DPS in follow-up year</td>
<td>0.905</td>
<td>0.029***</td>
<td>0.036**</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.017)</td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Has scores in follow-up year</td>
<td>0.881</td>
<td>0.032***</td>
<td>0.032*</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>N</td>
<td>2,939</td>
<td>4,964</td>
<td>1,508</td>
<td>1,472</td>
</tr>
<tr>
<td>C. Simulated score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolled in DPS in follow-up year</td>
<td>0.905</td>
<td>0.029***</td>
<td>0.037**</td>
<td>0.040**</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td></td>
</tr>
<tr>
<td>Has scores in follow-up year</td>
<td>0.881</td>
<td>0.032***</td>
<td>0.040**</td>
<td>0.043**</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.020)</td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2,939</td>
<td>4,964</td>
<td>1,523</td>
<td>1,290</td>
</tr>
</tbody>
</table>

Notes: This table reports coefficients from regressions of DPS enrollment and test-score availability indicators on charter offers, for the sample of charter applicants potentially available to construct the 2SLS estimates reported in Table 7. Column 1 reports follow-up rates for charter applicants who did not receive a charter offer. The propensity score control schemes used to construct the estimates in columns 3-5 parallel those used for Table 7. All models control for the covariates used for that table as well. Robust standard errors are reported in parentheses.

*significant at 10%; **significant at 5%; ***significant at 1%
<table>
<thead>
<tr>
<th>Application variable</th>
<th>Non-offered mean (1)</th>
<th>No controls (2)</th>
<th>Linear control (3)</th>
<th>Rounded (hundredths) (4)</th>
<th>Saturated (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of schools ranked</td>
<td>4.375</td>
<td>-0.341***</td>
<td>-0.317***</td>
<td>-0.056</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.093)</td>
<td>(0.086)</td>
<td>(0.094)</td>
<td></td>
</tr>
<tr>
<td>Number of charter schools ranked</td>
<td>1.425</td>
<td>0.476***</td>
<td>0.062</td>
<td>0.016</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.043)</td>
<td>(0.041)</td>
<td>(0.044)</td>
<td></td>
</tr>
<tr>
<td>First school ranked is charter</td>
<td>0.291</td>
<td>0.612***</td>
<td>0.003</td>
<td>-0.005</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.023)</td>
<td>(0.020)</td>
<td>(0.019)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2,939</td>
<td>4,964</td>
<td>1,508</td>
<td>1,472</td>
<td>1,224</td>
</tr>
<tr>
<td>Risk set points of support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>156</td>
<td>43</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robust F-test for joint significance</td>
<td>1190</td>
<td>8.06</td>
<td>0.47</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.701</td>
<td>0.986</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table reports balance coefficients and standard errors like those shown in Table 5a, with the modification that score control uses the formula version of the DA score. Robust standard errors are reported in parentheses. P-values for robust joint significance tests are estimated by stacking outcomes and clustering standard errors at the student level.

*significant at 10%; **significant at 5%; ***significant at 1%
Table B3b: Statistical tests for balance in student characteristics

<table>
<thead>
<tr>
<th>Student characteristics</th>
<th>Non-offered mean</th>
<th>No controls</th>
<th>Linear control (hundredths)</th>
<th>Rounded (hundredths)</th>
<th>Saturated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Origin school is charter</td>
<td>0.086</td>
<td>0.108***</td>
<td>0.085***</td>
<td>-0.012</td>
<td>-0.037**</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.022)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.520</td>
<td>-0.005</td>
<td>0.014</td>
<td>0.041</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.030)</td>
<td>(0.032)</td>
<td>(0.035)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.595</td>
<td>0.095***</td>
<td>-0.004</td>
<td>-0.031</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.029)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.183</td>
<td>-0.033***</td>
<td>-0.008</td>
<td>0.008</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.024)</td>
<td>(0.025)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Gifted</td>
<td>0.203</td>
<td>-0.028**</td>
<td>-0.047**</td>
<td>-0.040*</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.023)</td>
<td>(0.024)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Bilingual</td>
<td>0.289</td>
<td>0.086***</td>
<td>0.021</td>
<td>-0.002</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.029)</td>
<td>(0.030)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>Subsidized lunch</td>
<td>0.767</td>
<td>0.073***</td>
<td>-0.007</td>
<td>0.011</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Limited English proficient</td>
<td>0.289</td>
<td>0.086***</td>
<td>0.021</td>
<td>-0.002</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.029)</td>
<td>(0.030)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>Special education</td>
<td>0.084</td>
<td>0.004</td>
<td>0.027*</td>
<td>0.036**</td>
<td>0.033*</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.016)</td>
<td>(0.017)</td>
<td>(0.018)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2,939</td>
<td>4,964</td>
<td>1,508</td>
<td>1,472</td>
<td>1,224</td>
</tr>
</tbody>
</table>

Baseline scores
Math                      0.022            | -0.002         | 0.018                      | -0.049               | -0.080    |
                                 | (0.027)        | (0.056)       | (0.057)                  | (0.063)             |           |
Reading                  0.040            | -0.085***      | -0.023                    | -0.067               | -0.100*   |
                                 | (0.026)        | (0.053)       | (0.053)                  | (0.057)             |           |
Writing                 0.035            | -0.072***      | -0.039                    | -0.068               | -0.108*   |
                                 | (0.026)        | (0.051)       | (0.051)                  | (0.055)             |           |
N                              | 2,891           | 4,889        | 1,491                      | 1,455                | 1,213     |

Robust F-test for joint significance
p-value                    19.1            | 2.38           | 1.16                       | 1.18                 | 0.290     |

Notes: This table reports balance coefficients and standard errors like those shown in Table 5b, with the modification that score control uses the formula version of the DA score. Robust standard errors are reported in parentheses. P-values for robust joint significance tests are estimated by stacking outcomes and clustering standard errors at the student level.
*significant at 10%; **significant at 5%; ***significant at 1%
Table B4: Expected covariate balance by market size

<table>
<thead>
<tr>
<th></th>
<th>No controls</th>
<th>Actual size</th>
<th>Double size</th>
<th>Four times larger</th>
<th>Eight times larger</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Number of schools ranked</td>
<td>-0.341</td>
<td>0.052</td>
<td>0.023</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>Number of charter schools ranked</td>
<td>0.474</td>
<td>0.055</td>
<td>0.019</td>
<td>0.004</td>
<td>-0.001</td>
</tr>
<tr>
<td>First school ranked is charter</td>
<td>0.616</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Origin school is charter</td>
<td>0.115</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Female</td>
<td>-0.007</td>
<td>0.003</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.094</td>
<td>-0.005</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>Black</td>
<td>-0.031</td>
<td>0.001</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td>Gifted</td>
<td>-0.022</td>
<td>-0.005</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td>Bilingual</td>
<td>0.020</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Subsidized lunch</td>
<td>0.073</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Limited English proficient</td>
<td>0.084</td>
<td>-0.005</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td>Special education</td>
<td>-0.004</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td>Baseline scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>0.010</td>
<td>-0.020</td>
<td>-0.013</td>
<td>-0.009</td>
<td>-0.009</td>
</tr>
<tr>
<td>Reading</td>
<td>-0.070</td>
<td>-0.014</td>
<td>-0.009</td>
<td>-0.005</td>
<td>-0.005</td>
</tr>
<tr>
<td>Writing</td>
<td>-0.056</td>
<td>-0.016</td>
<td>-0.008</td>
<td>-0.005</td>
<td>-0.006</td>
</tr>
</tbody>
</table>

Average sample size 4,964 1,419 2,636 5,436 11058

Notes: This table repeats the expected balance calculations reported in Table 4 with markets of increasing size. Columns 1 and 2 are the same as columns 2 and 5 in Table 4. Columns 3-5 show balance after scaling market size by factors of 2, 4, and 8; this is accomplished by drawing additional lottery numbers and multiplying the number of seats accordingly. Except for column 1, the sample size reported at the bottom of the table shows the average number of participants in the appropriately scaled market with variation in the any-charter offer dummy conditional on the propensity score estimate that is relevant for that column.
Table B5: Comparison of 2SLS and OLS estimates of charter attendance effects without covariate controls

<table>
<thead>
<tr>
<th>DA score</th>
<th>2SLS estimates</th>
<th>OLS with score controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency (saturated)</td>
<td>Formula (saturated)</td>
</tr>
<tr>
<td>First stage</td>
<td>0.399*** (0.032)</td>
<td>0.376*** (0.032)</td>
</tr>
<tr>
<td>Math</td>
<td>0.339** (0.148)</td>
<td>0.363** (0.158)</td>
</tr>
<tr>
<td>Reading</td>
<td>-0.102 (0.136)</td>
<td>-0.108 (0.144)</td>
</tr>
<tr>
<td>Writing</td>
<td>0.116 (0.137)</td>
<td>0.134 (0.144)</td>
</tr>
<tr>
<td>N</td>
<td>1,102</td>
<td>1,083</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates analogous to those reported in Table 7, computed in models without covariate controls. Robust standard errors are reported in parentheses.

*significant at 10%; **significant at 5%; ***significant at 1%
Table B6: DPS innovation schools

<table>
<thead>
<tr>
<th>School</th>
<th>Total applicants</th>
<th>Applicants offered seats</th>
<th>DA score (frequency)</th>
<th>DA score (formula)</th>
<th>Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elementary and middle schools</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cole Arts and Science Academy</td>
<td>31</td>
<td>15</td>
<td>11</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>DCIS at Ford</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DCIS at Montbello</td>
<td>412</td>
<td>125</td>
<td>163</td>
<td>156</td>
<td>170</td>
</tr>
<tr>
<td>Denver Green School</td>
<td>153</td>
<td>62</td>
<td>29</td>
<td>46</td>
<td>52</td>
</tr>
<tr>
<td>Godsman Elementary</td>
<td>10</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Green Valley Elementary</td>
<td>53</td>
<td>15</td>
<td>3</td>
<td>23</td>
<td>35</td>
</tr>
<tr>
<td>Martin Luther King Jr. Early College</td>
<td>427</td>
<td>177</td>
<td>117</td>
<td>120</td>
<td>121</td>
</tr>
<tr>
<td>McAuliffe International School</td>
<td>406</td>
<td>165</td>
<td>91</td>
<td>115</td>
<td>112</td>
</tr>
<tr>
<td>McGlone</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Montclair Elementary</td>
<td>15</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Noel Community Arts School</td>
<td>288</td>
<td>108</td>
<td>92</td>
<td>97</td>
<td>105</td>
</tr>
<tr>
<td>Valdez Elementary</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Whittier K-8 School</td>
<td>47</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>High schools</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collegiate Preparatory Academy</td>
<td>433</td>
<td>125</td>
<td>173</td>
<td>158</td>
<td>153</td>
</tr>
<tr>
<td>DCIS at Montbello</td>
<td>506</td>
<td>125</td>
<td>208</td>
<td>169</td>
<td>174</td>
</tr>
<tr>
<td>High-Tech Early College</td>
<td>481</td>
<td>125</td>
<td>209</td>
<td>193</td>
<td>214</td>
</tr>
<tr>
<td>Manual High School</td>
<td>390</td>
<td>130</td>
<td>152</td>
<td>159</td>
<td>187</td>
</tr>
<tr>
<td>Martin Luther King Jr. Early College</td>
<td>515</td>
<td>144</td>
<td>179</td>
<td>151</td>
<td>162</td>
</tr>
<tr>
<td>Noel Community Arts School</td>
<td>334</td>
<td>78</td>
<td>112</td>
<td>112</td>
<td>107</td>
</tr>
</tbody>
</table>

Notes: This table describes DPS innovation applications in a format like that used for charters in Table 1 (excluding column 6).
Table B7: Covariate balance and differential attrition for DPS innovation schools

<table>
<thead>
<tr>
<th></th>
<th>Propensity score controls</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DA score (frequency)</td>
<td>Simulated score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonparametric</td>
<td>Linear saturation</td>
<td>Rounded (hundreds)</td>
<td>Saturated (thousands)</td>
<td>Linear control</td>
<td>Rounded (hundreds)</td>
<td>Saturated (thousands)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-offered</td>
<td>Linear control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>No controls</td>
<td>No controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of schools ranked</td>
<td>4.657^-0.142***</td>
<td>0.164</td>
<td>0.012</td>
<td>0.034</td>
<td>0.135</td>
<td>0.132</td>
<td>0.190</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of innovation schools ranked</td>
<td>1.279</td>
<td>0.710***</td>
<td>0.192***</td>
<td>0.086</td>
<td>0.035</td>
<td>0.034</td>
<td>0.121</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First school ranked is innovation</td>
<td>0.052**</td>
<td>0.611***</td>
<td>-0.003</td>
<td>-0.007</td>
<td>-0.005</td>
<td>-0.030</td>
<td>-0.030</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin school is innovation</td>
<td>0.116</td>
<td>0.125***</td>
<td>0.032</td>
<td>0.045</td>
<td>0.044</td>
<td>0.010</td>
<td>0.040</td>
<td>0.100*</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.526</td>
<td>-0.011</td>
<td>0.030</td>
<td>0.028</td>
<td>0.028</td>
<td>0.063</td>
<td>0.060</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race Hispanic</td>
<td>0.491</td>
<td>0.136***</td>
<td>0.028</td>
<td>0.015</td>
<td>-0.001</td>
<td>0.037</td>
<td>0.043</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.262</td>
<td>-0.064***</td>
<td>0.018</td>
<td>0.018</td>
<td>0.030</td>
<td>0.003</td>
<td>0.009</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gifted</td>
<td>0.198</td>
<td>-0.056***</td>
<td>-0.019</td>
<td>-0.028</td>
<td>-0.041</td>
<td>0.017</td>
<td>0.020</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilingual</td>
<td>0.018</td>
<td>0.007</td>
<td>-0.025</td>
<td>-0.027*</td>
<td>-0.029*</td>
<td>-0.020</td>
<td>-0.014</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsidized lunch</td>
<td>0.763</td>
<td>0.047***</td>
<td>0.029</td>
<td>0.034</td>
<td>0.016</td>
<td>0.011</td>
<td>0.013</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited English proficient</td>
<td>0.253</td>
<td>0.047***</td>
<td>0.016</td>
<td>0.032</td>
<td>0.031</td>
<td>0.007</td>
<td>-0.001</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special education</td>
<td>0.092</td>
<td>0.004</td>
<td>-0.021</td>
<td>-0.031</td>
<td>-0.036</td>
<td>-0.026</td>
<td>-0.037</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1,176</td>
<td>2,483</td>
<td>769</td>
<td>717</td>
<td>623</td>
<td>888</td>
<td>705</td>
<td>279</td>
</tr>
<tr>
<td>Baseline scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>-0.017</td>
<td>-0.186***</td>
<td>-0.032</td>
<td>-0.018</td>
<td>-0.057</td>
<td>0.023</td>
<td>0.042</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.036</td>
<td>-0.220***</td>
<td>-0.066</td>
<td>-0.047</td>
<td>-0.047</td>
<td>-0.013</td>
<td>0.002</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>0.000</td>
<td>-0.163***</td>
<td>0.025</td>
<td>0.041</td>
<td>0.030</td>
<td>0.079</td>
<td>0.081</td>
<td>0.119</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1,158</td>
<td>2,434</td>
<td>752</td>
<td>704</td>
<td>614</td>
<td>869</td>
<td>689</td>
<td>273</td>
</tr>
</tbody>
</table>

Robust F-test for joint significance

p-value

143 | 1.10 | 0.99 | 0.91 | 0.80 | 0.92 | 1.52

N | 1,176 | 2,483 | 769 | 717 | 623 | 888 | 705 | 279

C. Differential attrition

Enrolls in Denver in follow-up year | 0.920 | -0.001 | -0.017 | -0.012 | -0.011 | -0.015 | -0.020 | -0.008 |
|            | (0.011)                  |            |            |            |            |            |            |            |
| Has scores in follow-up year | 0.897 | -0.011 | -0.019 | -0.014 | -0.018 | -0.008 | -0.017 | 0.018 |
|            | (0.012)                  |            |            |            |            |            |            |            |

Notes: Panels A and B report covariate balance tests for innovation offers in a manner analogous to that used for charter offer balance in Tables 5a and 5b. Panel C reports attrition differentials for innovation offers in a manner analogous to that used for charter offer in Appendix Table B2. Robust standard errors are reported in parentheses. P-values for robust joint significance tests are estimated by stacking outcomes and clustering standard errors at the student level.

*significant at 10%; **significant at 5%; ***significant at 1%
Bibliography


_, Saskia-Levy Thompson, and Rebecca Unterman, “Transforming the High School Experience: How New York City’s New Small Schools are Boosting Student Achievement and Graduation Rates,” 2010. MDRC.


202


205


206


