NBER WORKING PAPER SERIES

DEMAND ANALYSIS USING STRATEGIC REPORTS: AN APPLICATION TO A SCHOOL CHOICE MECHANISM

Nikhil Agarwal
Paulo Somaini

Working Paper 20775
http://www.nber.org/papers/w20775

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2014

We are grateful to Parag Pathak for sharing data on admissions in Cambridge Public Schools, and for several helpful discussions. We thank Dan Ackerberg, Steve Berry, Eric Budish, Phil Haile and Mike Whinston for several helpful discussions, and Susan Athey, Lanier Benkard, Victor Chernozhukov, William Diamond, Liran Einav, Glenn Ellison, Aviv Nevo, Whitney Newey, Ariel Pakes, Larry Samuelson, Joe Shapiro and Frank Wolak for their comments. Vivek Bhattacharya provided outstanding research assistance. Both authors thank the Cowles Foundation of Economic Research and the Department of Economics at Yale University for their generous hospitality. Support from the National Science Foundation (SES-1427231) is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Demand Analysis using Strategic Reports: An application to a school choice mechanism
Nikhil Agarwal and Paulo Somaini
NBER Working Paper No. 20775
December 2014, Revised July 2015, Revised November 2016
JEL No. C50,D47,I20

ABSTRACT

Several school districts use assignment systems in which students have a strategic incentive to misrepresent their preferences. Indeed, we find evidence suggesting that reported rank-order lists in Cambridge, MA respond to these incentives. Such strategizing can complicate the analysis of preferences. This paper develops a new method for estimating random utility models in such environments. Our approach views the report made by a student as a choice of a probability distribution over assignment to various schools. We introduce a large class of mechanisms for which consistent estimation is feasible and study identification of a latent utility model assuming equilibrium behavior. Preferences are non-parametrically identified under either sufficient variation in choice environments or sufficient variation in a special regressor. We then propose a tractable estimation procedure for a parametric model based on Gibbs' sampling. Estimates from Cambridge suggest that while 84% of students are assigned to their stated first choice, only 75% are assigned to their true first choice. The difference occurs because students avoid ranking competitive schools in favor of less competitive schools. Assuming that ranking behavior is described by a Bayesian Nash Equilibrium, the Cambridge mechanism produces an assignment that is preferred by the average student to that under the Deferred Acceptance mechanism by an equivalent of 0.07 miles. This difference is smaller if beliefs are biased, and for naïve students.

Nikhil Agarwal
MIT
Department of Economics, E17-224
77 Massachusetts Ave
Cambridge, MA 02139
and NBER
agarwaln@mit.edu

Paulo Somaini
Stanford Graduate School of Business
Knight Management Center
655 Knight Way E372D
Stanford, CA 94305
and NBER
soma@stanford.edu

A data appendix is available at:
http://www.nber.org/data-appendix/w20775

1 Introduction

Admissions to public schools throughout the world is commonly based on assignment mechanisms that use reported rankings of various school options (Abdulkadiroglu and Sonmez, 2003). The design of such school choice mechanisms has garnered significant attention in the theoretical literature (Abdulkadiroglu, 2013). Although mechanisms that incentivize truthful revelation of preferences have been strongly advocated for in the theoretical literature (see Pathak and Sonmez, 2008; Azevedo and Budish, 2013, for example), with rare exceptions, real world school choice systems are susceptible to gaming. Table 1 presents a partial list of mechanisms in use at school districts around the world. To our knowledge, only Boston currently employs a mechanism that is not manipulable.

The widespread use of manipulable school choice systems poses two important questions. First, what are the costs and benefits of manipulable mechanisms from the student assignment perspective? Theoretical results on this question do not yield an unambiguous answer. Abdulkadiroglu et al. (2011) use a stylized model to show that strategic choice in the erstwhile Boston mechanism can effectively elicit cardinal information on preferences and can improve average student welfare. However, this potential benefit comes at a cost of violating notions of fairness and stability of final assignments. Second, how does one interpret and analyze administrative data on reported rankings generated by these mechanisms if the reports cannot be taken as truthful? Information on preferences can be useful for academic research on the effects of school choice on student welfare (Abdulkadiroglu et al., 2015a), student achievement (Hastings et al., 2009), and school competition (Nielson, 2013). Additionally, student preference information can also be useful for directing school reforms by identifying which schools are more desirable than others.[1] However, ignoring strategic incentives for such purposes can lead to incorrect conclusions and mis-directed policies.[2]

This paper addresses these two questions by developing a general method for estimating the underlying distribution of student preferences for schools using data from manipulable mechanisms, and applying these techniques to data from elementary school admissions in Cambridge, MA. We make several methodological and empirical contributions. First, we

[1]School accountability and improvement programs, or district-wide reforms, often use stated rank order lists as direct indicators of school desirability or student preferences. Boston's Controlled Choice Plan used the number of applications to a school as a formal indicator of school performance in a school improvement program. Similarly, Glenn (1991) argues that school choice caused improvements in the Boston school system based on observing an increase in the number of students that were assigned to their top choice.

[2]Previous empirical work has typically assumed that observed rank order lists are a truthful representation of the students' preferences (Hastings et al., 2009; Abdulkadiroglu et al., 2015a; Ayaji, 2013), allowing a direct extension of discrete choice demand methods. The assumption is usually motivated by arguing that strategic behavior may be limited in the specific environment. He (2014), Calsamiglia et al. (2016) and Hwang (2016) are notable exceptions that allow for agents to be strategic and are discussed below.

document strategic behavior to show that student reports respond to the incentives present in the mechanism. Second, we propose a new revealed preference method for analyzing the reported rank-order lists of students. We use this technique to propose a new estimator for the distribution of student preferences and derive its limit properties. These technical results are applicable to a broad class of school choice mechanisms that includes the systems listed on Table 1 except for the Top Trading Cycles mechanism. Third, we derive conditions under which the distribution of preferences is non-parametrically identified. Finally, we apply these methods to estimate the distribution of preferences in Cambridge, which uses a variant of the Boston Mechanism. These preferences estimates are then be used to analyze how often students are assigned to their true first choice schools, and to compare the outcomes under the current mechanism to an alternative that uses the student-proposing Deferred Acceptance mechanism.

Interpreting observed rank-order lists requires a model of agent behavior. Anecdotal evidence from Boston (Pathak and Sonmez, 2008) and laboratory experiments (Chen and Sonmez, 2006; Calsamiglia et al., 2010) suggest that strategic behavior may be widespread in manipulable school choice systems. Indeed, our analysis of ranking behavior for admissions into public elementary schools in Cambridge indicates significant gaming. There are strong incentives for strategic behavior in Cambridge. Since it uses a variant of the Boston mechanism, some schools are rarely assigned to students that rank it second, while others are have spare capacity after all students have been considered. Students therefore risk losing their priority at a competitive school if they do not rank it first. We investigate whether students appear to respond to these incentives using a regression discontinuity design. The design leverages the fact that students receive proximity priority at the two closest schools. We find that student ranking behavior changes discontinuously with the change in priority. This finding is not consistent with a model in which students rank schools in order of their true preferences if the distribution of preferences is continuous with respect to distance.

Therefore, instead of interpreting stated rank order lists as true preferences, our empirical approach is based on interpreting a student's choice of a report as a choice of a probability distribution over assignments. Each rank-order list results in a probability of getting assigned to each of the schools on that list. This probability depends on the student's priority type and report, a randomly generated tie-breaker, as well as the reports and priorities of the other students. Given their beliefs of this assignment probability, the expected utility from the chosen report must be greater than other reports the agent could have chosen. Our baseline model assumes that students best respond to the distribution of reports submitted by other students applying for admission in that year. This rational expectations assumption is an important baseline model for accounting the strategic behavior observed. However, we also consider models in which students have less information or have biased beliefs. First, we consider a model in which agents are unaware of the fine distinctions in the mechanism between various priority and student types. Second, we consider a model with adaptive expectations in which beliefs are based on the previous year. In an extension, we also estimate a mixture model with both naïve and sophisticated players.

Once a model of strategic behavior has been assumed, it is natural to ask whether it can be used to learn about the distribution of preferences using a typical dataset. To address this question, we study identification of a flexible random utility model that allows for student and school unobservables (see Block and Marshak, 1960; McFadden, 1973; Manski, 1977). Under the models of agent beliefs discussed above, estimates of assignment probabilities obtained from the data can be substituted for the students' beliefs. As we discuss later, consistent estimates of the assignment probabilities (as a function of reports and priority types) can be obtained using the data and the knowledge of the mechanism. Our results show that two types of variation can be used to learn about the distribution of preferences. The first is variation in choice environments that may arise from two identical populations of students facing different mechanisms or a different number of school seats. Under such variation, we characterize the identified set of preference distributions. The second form of variation assumes the availability of a special regressor that is additively separable in the indirect utility function. Such a regressor can be used to "trace-out" the distribution of preferences (Lewbel, 2000). Similar assumptions are commonly made to identify preferences in discrete choice models (Berry and Haile, 2010, for example). In our application, we use distance to school as a shifter in preferences for schools. Our empirical specification therefore rules out within-district residential sorting based on unobserved determinants of school preferences.[3]

This analysis naturally suggests a two-step estimation procedure for estimating the distribution of preferences under these alternative assumptions on student behavior. In a first step, we estimate the probability of assignment associated with each report and priority type. In a second step, we estimate the parameters of the distribution of preferences using a likelihood based method. Specifically, we implement a Gibbs' sampler adapted from McCulloch and Rossi (1994).[4] This procedure is convenient in our setting because the set of utility vectors for which a given report is optimal can be expressed in terms of linear inequalities, and allows us to avoid computing or simulating the likelihood that a report is

[3] We investigated whether house prices or the fraction of residential units occupied by Cambridge elementary school students is higher on the side of a priority zone boundary where priority is accorded at better performing schools. Our results suggest that proximity priority does not influence the housing market, and housing location decisions. These findings are consistent with families not paying attention to the details of the admissions system at the time of choosing where to live. Details available upon request.

[4] We view our non-parametric identification results as justifying that parametric assumptions are not essential for learning about the primitives of interest but are made to assist estimation in finite samples.

optimal given a parameter vector. In an appendix, we prove that our estimator is consistent and asymptotically normal. The primary technical contribution is a limit theorem for the estimated assignment probabilities. This result requires a consideration of dependent data since assignments depend on the reports of all students in the market. That school choice mechanisms are usually described in terms of algorithms rather than functions with well-known properties further complicates the analysis. We solve this problem for new and large class of school choice mechanisms that includes all mechanisms in table 1, except the Top Trading Cycles mechanism.

We then apply our methods to estimate student preferences in Cambridge, and address a wide range of issues. First, we investigate the extent to which students avoid ranking competitive schools in order to increase their chances of assignment at less competitive options. Prevalence of such behavior can result in mis-estimating the attractiveness of certain schools if stated ranks are interpreted on face value. Ignoring strategic behavior may therefore result in inefficient allocation of public resources for improving school quality. Further, a large number of students assigned to their first choice may not be an indication of student satisfaction or heterogeneity in preferences. We therefore investigate whether strategic behavior results in fewer students being assigned to their true first choice as compared to their stated first choice.

Second, we study the welfare effects of a switch to the student proposing Deferred Acceptance mechanism. The theoretical literature supports strategy-proof mechanisms on the basis of their simplicity, robustness to information available to participants and fairness (see Azevedo and Budish, 2013, and references therein). However, it is possible that ordinal strategy-proof mechanisms compromise student welfare by not screening students based on the intensity of their preferences (Miralles, 2009; Abdulkadiroglu et al., 2011). We quantify student welfare from the assignment under these two mechanisms under alternative models of agent beliefs and behavior. This approach abstracts away from potential costs of strategizing and acquiring information, which are difficult to quantify given the available data. Nonetheless, allocative efficiency is a central consideration in mechanism choice in addition to other criteria such as differential costs of participating, fairness and strategy-proofness (Abdulkadiroglu et al., 2009).

Our baseline results assuming equilibrium behavior indicate that the average student prefers the assignments under the Cambridge mechanism to the Deferred Acceptance mechanism. Interestingly, this difference is driven by paid-lunch students who face stronger strategic incentives than free-lunch students due to quotas based on free-lunch eligibility. A cost of improved assignments in Cambridge is that a small fraction of students (4-13% depending on the specification and the student group) have justified envy.[5] We then evaluate the mechanisms assuming that agents have biased beliefs about assignment probabilities. These estimates suggest that biased beliefs may mitigate the screening benefits of the Cambridge mechanism because mistakes can be costly in some cases.

Finally, we evaluate a mixture model with naïve and sophisticated agents to assess the distributional consequences across agents that vary in their ability to game the mechanism. We estimate that about a third of paid-lunch and free-lunch students report their preferences sincerely even if it may not be optimal to do so. Although naïve agents behave suboptimally, we find that the average naïve paid-lunch student prefers the assignments under the Cambridge mechanism. This occurs because naïve students rank their most preferred school first and gain priority at this school at the cost of sophisticates who avoid ranking these schools. The cost of not receiving their true second or third choices turns out to be smaller than this benefit. Further, the Cambridge mechanism does not constrain the assignments to ensure that no student has justified envy, allowing for potential increases in student welfare.

Related Literature

These empirical contributions are closely related to a handful of recent papers that estimate preferences for schools using manipulable mechanisms (He, 2014; Calsamiglia et al., 2016; Hwang, 2016). He (2014) proposes an estimator based on theoretically deriving properties of undominated reports using specifics of the school choice implementation in Beijing.[6] Hwang (2016) proposes a subset of restrictions on agent behavior based on simple rules to derive a bounds-based estimation approach. Compared to our procedure, these approaches avoid using restrictions that are implied if agents have information about which schools are more competitive than others. Calsamiglia et al. (2016) estimates a mixture model in which strategic agents solve for the optimal report in the Boston mechanism using backwards-induction from lower to higher ranked choices. This backwards induction approach is specific to mechanism studied in the paper.

There are a few other general distinguishing features from the aforementioned papers worth noting. First, the papers mentioned above use approaches that are specifically tailored to the school choice mechanism analyzed, and it may be necessary for a researcher to modify the ideas before applying them elsewhere. In contrast, we allow analysis for a more general class of mechanisms that includes mechanisms with student priority groups. Second, results

[5] Student i has justified envy if another student i' is assigned to a school j that student i prefers to her assignment and student i has (strictly) higher priority at j than student i'.

[6] The estimators proposed in He (2014) that do not assume optimal play are based limited number of restrictions implied by rationality, the specific number of schools and ranks that can be submitted in Beijing based on the fact that the school district treats all agents symmetrically.

on identification and large market properties of the estimator are not considered in the papers mentioned above. Finally, our empirical exercise investigates the consequences of specific forms of subjective beliefs on the comparison between mechanisms.

Our technical results on the large sample properties of our estimator uses results from the work on large matching markets by Kojima and Pathak (2009), Azevedo and Leshno (2013) and Azevedo and Budish (2013). The results on identification build on work on discrete choice demand (Matzkin, 1992, 1993; Lewbel, 2000; Berry and Haile, 2010). While the primitives are similar, unlike discrete choice demand, each report is a risky prospect that determines the probability of assignment to the schools on the list. This feature is similar to estimation of preference models under risk and uncertainty (Cardon and Hendel, 2001; Cohen and Einav, 2007; Chiappori et al., 2012). Cardon and Hendel (2001) and Cohen and Einav (2007) estimate preferences over a discrete set of insurance contracts with uncertainty in outcomes within each contract rather than uncertainty over which option is ultimately allocated. Chiappori et al. (2012) focuses on risk attitudes rather than the value of underlying prizes.

Our paper provides an empirical complement to the large theoretical literature that has taken a mechanism design approach to the student assignment problem (Gale and Shapley, 1962; Shapley and Scarf, 1974; Abdulkadiroglu and Sonmez, 2003). A significant literature debates the trade-offs between manipulable and non-manipulable mechanisms Ergin and Sonmez (2006); Pathak and Sonmez (2008); Miralles (2009); Abdulkadiroglu et al. (2011); Featherstone and Niederle (2011); Troyan (2012); Pathak and Sonmez (2013). Theoretical results from this literature has been used to guide redesigns of matching markets (Roth and Peranson, 1999; Abdulkadiroglu et al., 2006, 2009).

A growing literature is interested in methods for analyzing preferences in matching markets, usually using pairwise stability (Choo and Siow, 2006; Fox, 2010b,a; Chiappori et al., 2015; Agarwal, 2015; Agarwal and Diamond, 2014). In some cases, estimates are based on the strategic decision to engage in a costly courting decisions in matching markets (Hitsch et al., 2010). Similar considerations are important when applying to colleges (Chade and Smith, 2006).

The proposed two-step estimator leverages insights from the industrial organization literature, specifically the estimation of empirical auctions (Guerre et al., 2000; Cassola et al., 2013), single agent dynamic models (Hotz and Miller, 1993; Hotz et al., 1994) and dynamic games (Bajari et al., 2007; Pakes et al., 2007; Aguirregabiria and Mira, 2007). As in the methods used in those contexts, we use a two-step estimation procedure where the distribution of actions from other agents is used in a first step estimator.

Overview

Section 2 describes the Cambridge Controlled Choice Plan and presents evidence that students are responding to strategic incentives provided by the mechanism. Section 3 and 4 present the model and the main insight on how to interpret submitted rank order lists. Section 5 and Section 6 discuss identification and estimation. A reader interested in the empirical application instead of the econometric techniques may skip these sections. Section 7 applies our techniques to the dataset from Cambridge, MA.

2 Evidence on Strategic Behavior

2.1 The Controlled Choice Plan in Cambridge, MA

We use data from the Cambridge Public Schools' (CPS) Controlled Choice Plan for the academic years 2004-2005 to 2008-2009. Elementary schools in the CPS system assigns about 41% of the seats through a partnerships with pre-schools (junior kindergarten or Montessori) or an appeals process for special needs students. The remaining seats are assigned through a school choice system that takes place in January for students entering kindergarten. We will focus on students and seats that are allocated through this process.

Table 2 summarizes the students and schools. The CPS system has 13 schools and about 400 students participating in it each year. One of the schools, Amigos, was divided into bilingual Spanish and regular programs in 2005. Bilingual Spanish speaking students are considered only for the bilingual program, and students that are not bilingual are considered only for the regular program.[7] King Open OLA is a Portuguese immersion school/program that is open to all students. Tobin, a Montessori school, divided admissions for four and five year olds starting 2007.

An explicit goal of the Controlled Choice Plan is to achieve socio-economic diversity by maintaining the proportion of students who qualify for the federal free/reduced lunch program in each school close to the district-wide average. Except Amigos and only for the purposes of the assignment mechanism, all schools are divided into paid-lunch and free/reduced lunch programs. Students eligible for federal free or reduced lunch are only considered for the corresponding program.[8] About 34% of the students are on free/reduced lunch. Each program has a maximum number of seats and the overall school capacity may be lower than the sum of the seats in the two programs. Our dataset contains both the total number of seats in the school as well as the seats available in each of the programs.

[7] A student voluntarily declares whether she is bilingual on the application form.

[8] Households with income below 130% (185%) of the Federal Poverty line are eligible for free (reduced) lunch programs. For a household size of 4, the annual income threshold was approximately \$27,500 (\$39,000) in 2008-2009.

The Cambridge Controlled Choice Mechanism

We now describe the process used to place students at schools. The process prioritizes students at a given school based on two criteria:

1. Students with siblings who are attending that school get the highest priority.

2. Students receive priority at the two schools closest to their residence.

Students can submit a ranking of up to three programs at which they are eligible. Cambridge uses an Immediate Acceptance mechanism (a variant of the Boston mechanism) and assigns students as follows:

Step 0: Draw a single tie-breaker for each student

Step k = 1,2,3: Each school considers all students that have not been previously assigned and have listed it in the k-th position and arranges them in order of priority, breaking ties using the randomly drawn tie-breaker. Starting from the first student in the list, students are considered sequentially. The student is considered for the paid-lunch program if she is not eligible for a federal lunch subsidy and for the free/subsidized lunch program otherwise. She is assigned to the corresponding program unless,

(a) There are no seats available in the program, or

(b) There are no seats available in the school.

If either of the conditions above are satisfied, the student is rejected.

There are a few notable features of this mechanism. First, the mechanism prioritizes students at higher ranked options. The effective priority therefore depends on the report of the student. Second, there is a cutoff effective priority for each program/school and all students below that cutoff are rejected. This cutoff is set so that the number of students assigned to the program/school does not exceed the capacity. Finally, students are assigned to their highest ranked option for which their effective priority is above the cutoff.

These features of the Cambridge Controlled Choice Plan are shared with a large class of mechanisms that we will formally introduce below. There are two clear reasons why such mechanisms can result in strategic incentives. First, the dependence of the effective priority on the report provides incentives to skew ranking towards options where priority is most valuable. Second, if the length of the list is limited, students should avoid ranking too many schools where their priority is likely to be below the cutoff. We now describe the ranking behavior and strategic decision-making in response to these incentives in Cambridge.

2.2 Descriptive Evidence on Ranking Behavior

Panels A and D in table 3 show that over 80% of the students rank the maximum allowed number of schools and over 80% of the students are assigned to their top ranked choice in a typical year. Researchers in education have interpreted similar statistics in school districts as indicators of student satisfaction and heterogeneity in student preferences. For instance, Glenn (1991) argues that school choice caused improvements in the Boston school system based on observing an increase in the number of students that were assigned to their top choice.[9] Similarly, Glazerman and Meyer (1994) interpret a high fraction of students getting assigned to their top choice in Minneapolis as indicative of heterogeneous student preferences.

Conclusions based on interpreting stated preferences as truthful are suspect when a mechanism provides strategic incentives for students. It is well understood that students risk "losing their priority" if a school is not ranked at the top of the list in mechanisms like the Boston mechanism (Ergin and Sonmez, 2006). Table 3, panel E shows that students tend to rank schools where they have priority closer to the top. For instance, schools where a student has sibling priority is ranked first 32% of the time as compared to 3% of the time anywhere else on the list. Likewise, schools where a student has proximity priority are also more likely to be ranked higher. These statistics do not necessarily indicate that this behavior is in response to strategic incentives because having priority may be correlated with preferences. However, given that strategic incentives may also result in similar patterns, it may be incorrect to estimate preferences by treating stated lists as true preference rankings. For example, Panels D and F of tables 2 and 3 show that the top-ranked school is closer than the average school, and closer than other ranked schools. One may incorrectly conclude that students have strong preferences for going to school close to home if proximity priority is influencing this choice.

2.3 Strategic Incentives in Cambridge

Table 4 takes a closer look at the strategic incentives for students in Cambridge. Panel A shows the frequency with which students rank the various school options, the capacity at the various schools as well as the the rank and priority type of the first rejected student in a school. Panels B and C present identical statistics, but split by free/reduced lunch status of students. The table indicates significant heterogeneity in the competitiveness of the schools. Baldwin, Haggerty, Amigos, Morse, Tobin, Graham & Parks, and Cambridgeport are competitive schools with many more students ranking them than there is capacity. Panel

[9] The argument is based on ranking and assignment data generated when Boston used a manipulable assignment system.

A indicates that a typical student would be rejected in these schools if she does not rank it as her top choice. Indeed, Graham & Parks rejected all non-priority students even if they had ranked it first in each of the five years. The other schools typically admit all students that were not assigned to higher ranked schools. Additionally, the competitiveness of schools differs by paid-lunch status. While Graham & Parks is very competitive for students that pay for lunch, it did not reject any free/reduced lunch students that applied to it in a typical year. More generally, a larger number of schools are competitive for paid-lunch students than for free-lunch students.

There are two other features that are worth highlighting. First, there are few schools that do not reject students that listed them first but do reject second or third choice students. Therefore, students must rank competitive school first in order to gain admission but may rank non-competitive schools at any position. This suggests that, in Cambridge at least, strategic incentives may be particularly important when considering which school to rank first. Second, several paid lunch students rank competitive schools as their second or third choice. This may appear hard to rationalize as optimal behavior. However, it should be noted that despite the large number of students ranking competitive schools second, these choices are often not pivotal, as evidenced by the extremely large number of students that are assigned to their top choice. Another possibility is that students are counting on back-up schools, either at the third ranked school, a private or a charter school in case they remain unassigned. Finally, students may simply believe that there is a small chance of assignment even at competitive schools. We further discuss these issues when we present our estimates.

2.4 Strategic Behavior: A Regression Discontinuity Approach

We now present evidence that students are responding strategically when choosing which schools to rank. Our empirical strategy is based on the assignment of proximity priority in Cambridge. A student receives priority at the two closest schools to her residence. We can therefore compare the ranking behavior of students that are on either side of the boundary where the proximity priority changes. If students are not behaving strategically and the distribution of preferences are continuous in distance to school, we would not expect the ranking behavior to change discretely at this boundary. On the other hand, the results in table 4 indicate that a student risks losing her proximity priority at competitive schools if she does not rank it first. Therefore, some students may find it optimal to manipulate their reports in order to avoid losing proximity priority. Strategic students may rank a competitive school at which they have priority as their first choice instead of their most preferred school. We now test whether students are responding to this strategic incentive.

Figure 1 and table 5 present the results. The figure plots the probability of ranking a school in a particular position against the distance from a proximity priority boundary. Specifically, let d_{i2} and d_{i3} be the second and third closest schools to student i. For any school j, the horizontal axis is the difference $\Delta d_{ij} = d_{ij} - \frac{1}{2}(d_{i2} + d_{i3})$. Since Cambridge assigns a student priority at the two closest schools, Δd_{ij} is negative if student i has priority at school j and positive otherwise. The vertical line represents this boundary of interest where we assess ranking behavior. The black dashed lines are generated from a local linear regression of the ranking outcome y_{ij} on the distance from this boundary, Δd_{ij}, estimated separately using data on either side of the boundary. The black points represent a bin-scatter plot of these data, with a 95% confidence interval depicted with the bars. The grey points control for school fixed effects. Table 5 presents the estimated size of the discontinuity using the procedure recommended by Imbens and Kalyanaraman (2011). We use their estimator to test whether the outcome studied changes discontinuity at the corresponding boundary discontinuity.

Panels (a) through (d) of figure 1 construct the boundary so that students have proximity priority at schools to the left of the vertical line. Panel (a) shows that the probability that a student ranks a school first decreases discontinuously at the proximity boundary. Further, the response to distance to school is also higher to the left of the boundary, probably reflecting the preference to attend a school closer to home. The jump at the boundary may be attenuated because a student can rank only one of the two schools she has priority as her top choice.[10] In contrast to panel (a), panels (b) and (c) do not show a large jump at the proximity boundary for the probability a school is ranked second or third. This should be expected because we saw earlier that one's priority is unlikely to be pivotal in the second or third choices. These panels also show that the probability of ranking in second or third position a school that is extremely close to a student's home is low. This is explained by the fact that students instead rank nearby schools first. Table 5 presents the estimated size of this discontinuity and the standard errors of these estimates. The first column shows that the probability that a school is ranked first drops by 5.75% at the boundary where the student loses proximity priority. This effect is statistically significant at the 1% level. Further, panels B and C of the table show that this change is larger for paid lunch students than for free lunch students. This is consistent with the theory that paid lunch students are responding to the stronger strategic incentives as compared to free lunch students. The next two columns present these estimates for the second and third ranked choices. As indicated by the figures, the estimated effects are smaller and often not statistically significant.

[10]Panel (a) of figure D.1 in the appendix focuses on the second and third closest schools and shows that the discontinuity is discernible.

Strategic pressures to rank a school first may be particularly important if the school is competitive. Panels (d) and (e) of figure 1 investigates the differential response to proximity priority by school competitiveness. Specifically, we split the schools based on whether they rejected some students in a typical year or not as delineated in table 4. Consistent with strategic behavior, panel (d) shows that the probability of ranking a competitive school first falls discontinuously at the boundary where proximity priority changes. In contrast, the discontinuity in panel (e), which focuses on non-competitive schools is smaller. Indeed, the fourth and fifth columns of table 5 confirm that the estimated drop in the probability of ranking a competitive schools first is 7.27%, which is larger than the overall estimate. Additionally, panels B and C of table 5 shows that the estimated response to proximity priority is larger for paid-lunch students at 11.07% as compared to 1.47% for free-lunch students.[11] Non-competitive schools, in stark contrast, have an estimated drop that is only 2.06% and not statistically significant. Consistent with strategic pressures being less stringent at non-competitive schools, the change in ranking probability at the boundary is statistically indistinguishable from zero for both paid-lunch and free-lunch students. However, we view the estimates for free-lunch students as inconclusive because the point estimates are fairly large and imprecise for both competitive and non-competitive schools. Our findings are consistent with paid-lunch students responding to significant strategic pressures in the Cambridge mechanism, and free-lunch students with an undetectable response given the lower strategic incentives they face.

Finally, we consider a placebo test in which we constructed the figures and estimates above assuming that proximity priority is only given at the closest school. Panel (f) in figure 1 shows no discernible difference in the ranking probability at this placebo boundary. The estimated size of the discontinuity, presented in the last column of table 5, is only 0.07% and statistically indistinguishable from zero. Figure D.1 (panel d) in the appendix presents a second placebo boundary, dropping the two closest schools and constructing priorities at the two closest remaining schools. As expected, we do not find a discontinuous response at this placebo boundary.

Together, these results strongly suggest that ranking behavior is discontinuous at the boundary where proximity priority changes. However, there are two important caveats that must be noted before concluding that agents in Cambridge are behaving strategically. First, the results do not show that all students are responding to strategic incentives in the mechanism, or that their expectations are correct. We therefore consider models with biases in beliefs and behavior in addition to a rational expectations model. Second, it is possible that the response is driven in part by residential choices with which parents picking a home so

[11]Figure D.1 (panels b and c) in the appendix shows the plots by free-lunch status.

that the student receives priority at a more preferred school. While the previous literature has found evidence of residential sorting across school districts (Black, 1999; Bayer et al., 2007), we are not aware of any research on the effects of priorities on the housing market within a unified school district. A boundary discontinuity design similar to Black (1999) does not suggest that house prices or residential location decisions respond to these priority boundaries.[12] A more thorough analysis of this issue or a full model that considers the joint decision residential and school choices is left for future research.

These results contrast with Hastings et al. (2009), who find that the average quality of schools ranked did not respond to a change in the neighborhood boundaries in the year the change took place. Assuming that students prefer higher quality schools, their finding suggests that students did not strategically respond to the change in incentives. As suggested by Hastings et al. (2009), strategic behavior may not be widespread if the details of the mechanism and the change in neighborhood priorities are not well advertised. Charlotte-Mecklenberg had adopted the school choice system just a year prior to their study and the district did not make the precise mechanism clear. In contrast, Cambridge's Controlled Choice Plan is published on the school district's website and has been in place for several years. These institutional features may account for the observed differences in the student behavior.

3 Model

We consider school choice mechanisms in which students are indexed by $i \in \{1, \ldots, n\}$ and programs indexed by $j \in \{0, 1, \ldots, J\} = \mathcal{J}$. Program 0 denotes being unmatched. Each program has $n \times q_j^n$ seats, with $q_j^n \in (0, 1)$ and $q_0 = 1$.[13] In addition, program j belongs to school $s_j \in \{0, 1, \ldots, S\} = \mathcal{S}$. The school capacities are such that $q_s^n \leq \sum_{\{j: s_j = s\}} q_j^n$. We now describe how students are assigned to these seats, their preferences over the assignments, and the equilibrium behavior.

3.1 Utilities and Preferences

We assume that student i's utility from assignment into program j is given by $V(z_{ij}, \xi_j, \epsilon_i)$, where z_{ij} is a vector of observable characteristics that may vary by program or student or

[12] Details available upon request. Note that the proximity priority system for kindergarten admission is not used for higher grades in Cambridge.

[13] This convention will be convenient since we will be considering limits as $n \to \infty$ in which $q_j^n \to q_j^n \in (0, 1)$.

both, and ξ_j and ϵ_i are (vector-valued) unobserved characteristics. Let

$$v_i = (v_{i1}, \ldots, v_{iJ})$$

be the random vector of indirect utilities for student i with conditional joint density $f_V(v_{i1}, \ldots, v_{iJ} | \xi, z_i)$, where $\xi = (\xi_1, \ldots, \xi_J)$ and $z_i = (z_{i1}, \ldots, z_{iJ})$. We normalize the utility of not being assigned through the assignment process, v_{i0}, to zero.[14] Therefore, v_{i0} is best interpreted as the inclusive value of remaining on the wait-list.

This formulation allows for heterogeneous and non-additive preferences conditional on observables. The primary restriction thus far is that a student's indirect utility depends only on their own assignment and not of other students. This rules out preferences for peer groups or for conveniences that carpool arrangements may afford.

While our identification results do not make parametric assumptions on utilities, we will make parametric assumptions to assist estimation in finite samples in our application. For the empirical application, we specify student i's indirect utility for school j as

$$\begin{aligned} v_{ij} &= \sum_{k=1}^{K} \beta_{kj} z_{ijk} - d_{ij} + \varepsilon_{ij} \qquad (1) \\ v_{i0} &= 0 \end{aligned}$$

where d_{ij} is the road distance between student i's home and school j; z_{ijk} are student-school specific covariates; β_{kj} are school specific parameters to be estimated; $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ}) \sim N(0, \Sigma)$ independently of z, d.[15] The normalization of $v_{i0} = 0$ is without loss of generality, and the scale normalization is embedded in the assumption that the coefficient on d_{ij} is -1. Our estimated specification constructs z_{ijk} by interacting indicators of student paid-lunch status, sibling priority, proximity priority, ethnicity, home-language and a constant with school-specific dummies. The assumption that $\varepsilon \perp z, d$ rules out residential sorting on unobservables, and is commonly made in the existing empirical work on school choice. As mentioned earlier, we find do not find evidence of residential decisions in Cambridge responding to incentives created by the mechanism.

[14] Our identification results will impose two different sets of scale normalizations that are both without loss of generality. These normalizations will be discussed in Section 5.

[15] Our specification allows for heteroskedastic errors ε_{ij} and arbitrary correlation between ε_{ij} and $\varepsilon_{ij'}$. This specification relaxes homoskedastic and independent preference shocks commonly used in logit specifications.

3.2 Assignment Mechanisms

School choice mechanisms typically use submitted rank-order lists and defined student priority types to determine final assignments. As is the convention in the school choice literature, let $R_i \in \mathcal{R}_i$ be a rank-order list, where jR_ij' indicates that j is ranked above j'.[16] Students, if they so choose, may not submit a report in which the most preferred schools are ranked in order of true preferences. Let student i's priority type be denoted $t_i \in T$. In Cambridge, t_i defines the free-lunch type, the set of schools where the student has proximity priority and whether or not the student has a sibling in the school.[17]

A mechanism is usually described as an outcome of an algorithm that takes these rank-order lists and priorities as inputs. To study properties of a mechanism and our methods, it will be convenient to define a mechanism as a function that depends on the number of students n.

Definition 1. *A* ***mechanism*** *is a function* $\Phi^n : \mathcal{R}^n \times T^n \to \left(\Delta^J\right)^n$ *such that for all* (R_i, t_i), *and* $R_{-i} = (R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n)$ *and* $T_{-i} = (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$, $\Phi^n_{ij}((R_i, t_i), (R_{-i}, T_{-i}))$ *denotes the probability that* i *is assigned to program* j.

In this notation, the $i-j$-th component of $\Phi^n((R_i, t_i), (R_{-i}, T_{-i}))$, is the probability that student i is assigned to program j. Hence, assignment probabilities for each student is in the J-simplex Δ^J. The final outcome depends on a random number used to break ties between students. Such tie-breakers are a common source of uncertainty faced by students.

In Cambridge, the function $\Phi^n((R_i, t_i), (R_{-i}, T_{-i}))$ is determined by a cutoff-rule, and the priority of the student. Since the mechanism prioritizes students based on the position where the program is listed, this priority depends on the student's report. Specifically, the priority of student i at school j is

$$e_{ij} = f_j(R_i, t_i, \nu_i) = \frac{3 - R_i(j) + \dfrac{t_{ij} + \nu_i}{4}}{3},$$

where $\nu_i \in [0, 1]$ is student i's draw of the random tie-breaker, $R_i(j)$ is the position of school j in the rank-order list R_i,[18] and t_{ij} is respectively 0, 1, 2 or 3 if the student has no priority, proximity priority only, sibling priority only, or both proximity and sibling priority. The

[16]The set $\mathcal{R}_i$ may depend on the student's priority type t_i and may be constrained. For example, students in Cambridge can rank up to three schools, and programs are distinguished by paid-lunch status of the student.

[17]We assume that students take their priority types as given. Cambridge verifies residence and free-lunch eligibility using documentary evidence. Since most schools are less competitive for free-lunch students and the classes of instruction are not split by free-lunch status, it's unlikely that not declaring free-lunch eligibility is beneficial for any student.

[18]$R_i(j)$ is set to 4 if school j is not ranked.

function f chosen for Cambridge ensures that students that rank a school higher than other students are given precedence, with ties broken first by proximity and/or sibling priority, and then by the random tie-breaker.

Given the priorities e_{ij} above, let p_j be the highest priority student that was rejected at program j. This algorithm places student i in program j if $e_{ij} > p_j$ and student i ranks program j above any other program j', with $e_{ij'} > p_{j'}$. Hence, the algorithm assigns student i to program j if $D_j^{(R_i,t_i,\nu_i)}(p) = 1$, where

$$D_j^{(R_i,t_i,\nu_i)}(p) = 1\{e_{ij} > p_j, jR_i 0\} \prod_{j' \neq j} 1\{jR_i j' \text{ or } e_{ij'} \leq p_{j'}\}.$$

The total fraction of students that would be assigned to program j, if the cutoffs were p is therefore given by $D_j(p) = \frac{1}{n} \sum_i D_j^{(R_i,t_i,\nu_i)}(p)$. Because the number of students assigned to program j (likewise school s) cannot exceed the number of available seats $n \times q_j^n$ (likewise $n \times q_s^n$), it is easy to verify that the cutoffs determined by the algorithm in Section 2.1 have the following property.

Definition 2. *The vector of cutoffs $p \in [0,1]^J$ is a **market clearing cutoffs** for $(D(p), q^n)$, if, for each program $j \in \mathcal{J}$,*

$$D_j(p) - q_j^n \leq 0,$$

with equality if $p_j > \min\{p_{j'} : j' \neq j, s_{j'} = s_j\}$,[19] and for each school $s \in \mathcal{S}$,

$$\sum_{\{j : s_j = s\}} D_j(p) - q_s^n \leq 0,$$

with equality if $\min\{p_j : s_j = s\} > 0$.

The first constraint ensures that program j has a cutoff larger than the other programs in the same school only if a student was rejected because the program has exhausted its capacity. The second constraint ensures that the lowest cutoff in the school is strictly positive only if the school has exhausted its total capacity.

In fact, this representation of the assignments in terms of the market clearing cutoffs and a priority that depends on a student's report is not unique to Cambridge. Many of school choice mechanisms prioritize students based on their rank-order list and implement a cutoff rule which rejects all students below a program/school specific threshold. We now formally define a large class of mechanisms for which such a representation is valid.

[19]We use the convention that $\min\{p_{j'} : j' \neq j, s_{j'} = s_j\} = 0$ if $\{j' \neq j, s_{j'} = s_j\} = \emptyset$.

Definition 3. *A mechanism* Φ^n *is a* ***Report-Specific Priority + Cutoff*** *mechanism if there exists a function* $f : \mathcal{R} \times T \times [0,1]^J \to [0,1]^J$ *and a measure* γ_ν *over* $[0,1]^J$ *such that*

(i) $\Phi_j^n((R_i, t_i), (R_{-i}, t_{-i}))$ *is given by*

$$\int \cdots \int D^{(R_i,t_i,\nu_i)}(p^n) \mathrm{d}\gamma_{\nu_1} \ldots \mathrm{d}\gamma_{\nu_n}$$

where $e_i = f(R_i, t_i, \nu_i)$ *is the eligibility score vector for student* i*,*

(ii) p^n *are market clearing cutoffs for* $(D(p), q^n)$ *where* $D(p) = \frac{1}{n} \sum_i D_j^{(R_i,t_i,\nu_i)}(p)$*.*

(iii) $f_j(R_i, t_i, \nu_i)$ *is strictly increasing in* ν_{ij}*, and does not depend on* $\nu_{ij'}$ *for* $j' \neq j$*.*[20]

The representation highlights two ways in which these mechanisms can be manipulable. First, the report of an agent can affect her eligibility score. Fixing a cutoff, agents may have the direct incentive to make reports that may not be truthful. Second, even if eligibility does not depend on the report, an agent may (correctly) believe that the cutoff for a school will be high, making it unlikely that she will be eligible. If the rank-order list is constrained in length, she may choose to omit certain competitive schools.[21]

Table 1 presents a partial list of school choice mechanisms currently used around the world. As we show in Appendix B.1, all mechanisms in table 1, except for the Top Trading Cycles Mechanism belong to this class of cutoff based mechanisms. A remarkable feature is that these mechanism differ essentially by the choice of f. The techniques we develop below are applicable to this entire class.[22]

3.3 Agent Beliefs and Best Responses

Evidence presented in Section 2 suggests that agents are responding to strategic incentives in the Cambridge mechanism. Further, anecdotal evidence suggests that parent groups and forums discussing ranking strategies are active (Pathak and Sonmez, 2008), and laboratory experiments suggests that strategic behavior is more common for manipulable mechanisms than strategy-proof mechanisms (Chen and Sonmez, 2006; Calsamiglia et al., 2010).

To model this strategic behavior, we assume that agents submit rank-order lists that are optimal given a set of beliefs about the probability of assignment at various schools. Let

[20] This allows for a single tie-breaker as in Cambridge, by assuming that ν_{ij} and $\nu_{ij'}$ are perfectly correlated.

[21] The representation extends the characterization of stable matchings by Azevedo and Leshno (2013) in terms of demand-supply and market clearing to discuss mechanisms. Particularly, we can use the framework to consider mechanisms that produce matchings that are not stable. The representation may therefore be of independent theoretical interest.

[22] Abdulkadiroglu et al. (2015b) use a related cutoff based approach for evaluating achievement gains from attending various types of schools for the case of the Deferred Acceptance Algorithm.

$L_{R_i,t_i} \in \Delta^J$ denote the belief that an agent with priority type t_i has about her assignment probabilities at the various schools. We will specify these beliefs as a function of the mechanism as well as the strategies of the other agents in the mechanism. This dependence is suppressed for notational simplicity.

Given a preference vector $v_i \in \mathbb{R}^J$ and priority type t_i, the agent's expected utility from reporting R_i is therefore $v_i \cdot L_{R_i,t_i}$. By choosing a different rank-order lists, this agent can achieve (believed) assignment probabilities in the set $\mathcal{L}_{t_i} = \{L_{R_i,t_i} : R_i \in \mathcal{R}_i\}$. Therefore, this agent reports R_i only if

$$v_i \cdot L_{R_i,t_i} \geq v_i \cdot L_{R_i',t_i}, \text{ for all } R_i' \in \mathcal{R}_i. \tag{2}$$

It is important to emphasize that optimality of R_i is with respect to the agent's belief about her assignment probabilities, which may or may not be the true assignment probabilities generated by the mechanism. In our application, we consider specifications with three alternative forms of beliefs below.

3.3.1 Rational Expectations

Our baseline model assumes that agents have correct beliefs about the probabilities of assignments given their own type (v, t), and the population distribution of types and ranking strategies used in the district.

Specifically, let $\sigma : \mathbb{R}^J \times T \to \Delta^{|\mathcal{R}|}$ be a (symmetric) mixed strategy used by the students in the district.[23] The first argument of σ is the vector of utilities over the various schools, and the second argument is the priority type of the student. Each student believes that her probability of assignment probability when she reports $R_i \in \mathcal{R}_i$ is given by the vector

$$L^{n,\sigma}_{R_i,t_i} \quad = \quad \mathbb{E}_\sigma[\Phi^n((R_i,t_i),(R_{-i},T_{-i}))|R_i,t_i] \tag{3}$$

where $\mathbb{E}_\sigma[\cdot]$ denotes expectations taken over the random variable $(R_{i'}, t_{i'})$ for $i' \neq i$ drawn iid with probability $f_\sigma(R,t) = f_T(t) \int \sigma_R(V,t) \mathrm{d}F_{V|T=t}$, and $F_{V,T}$ is the joint distribution of preference and priority types.

In this model of beliefs, the perceived probability of assignment depends on both the tie-breaker in the mechanism and the distribution of reports by the other students in the district. Therefore, a student perceives uncertainty due to both the reports submitted by other students as well the uncertainty in the lotteries within in the mechanism. This contrasts with models of complete information about the reports submitted by the other students,

[23] Note that, except for a measure zero set of types, a pure strategy is optimal when the distribution of utilities admits a density.

where the latter form of uncertainty is not present.[24] We believe that this model is more realistic than a complete information model since students are unlikely to be aware of all the reports that will be submitted by others. This assumption is commonly made in the analysis of other non-dominant strategy mechanisms, for example, in empirical analysis of auctions (Guerre et al., 2000; Cassola et al., 2013, among others).[25]

3.3.2 Adaptive Expectations

Assuming rational expectations implies a strong degree of knowledge and sophistication. One may reasonably argue that the primary source of information for parents may be based on prior year information. We address this possibility by alternatively specifying agent beliefs as

$$L^{n,-1}_{R_i,t_i} = \mathbb{E}_{\sigma_{-1}}[\Phi^{-1,n}((R_i, t_i), (R_{-i}, T_{-i}))|R_i, t_i] \tag{4}$$

where the notation -1 indicates the use of previous year quantities rather than the current year. Specifically, we assume that the agents have knowledge about the previous year strategy σ_{-1}, assume that the school/program capacities from the previous year, and the distribution of other student types are based on the previous year as well.

Agents with such beliefs form expectations about the competitiveness of various schools based on the experiences of parents that participated in the previous year. These agents may be systematically mis-informed, for example, about increases or decreases in capacity. Estimates from this model can be used to investigate whether the potential screening benefits of manipulable mechanisms hinge crucially on agents forming rational expectations.

3.3.3 Coarse Expectations

Another form of mis-information in the mechanism may be about the specific use of priorities and the program quotas based on free/reduced lunch status. If agents form expectations that are not sensitive to these details of the mechanism, their expected probability of assignment given report R_i will not depend on their priority type t_i. For example, beliefs that are based only on aggregate information about the number of applicants and capacities at each school

[24] Complete information Nash Equilibrium models are common in the literature on assignment mechanisms (see Ergin and Sonmez, 2006, for example). Results based on assuming that agents have knowledge of R_{-i} are both quantitatively and qualitatively similar to the ones presented here, and are available on request.

[25] Indeed, when all agents optimally respond to such beliefs, the behavior is consistent with a Bayesian Nash Equilibrium:

Definition 4. *The strategy σ^* is a* ***type-symmetric Bayesian Nash Equilibrium*** *if $v_i \cdot L^{n,\sigma^*}_{R_i,t_i} \geq v_i \cdot L^{n,\sigma^*}_{R'_i,t_i}$ for all $R'_i \in \mathcal{R}_i$ whenever $\sigma^*_{R_i}(v_i, t_i) > 0$.*

would have this property.

To address the possibility that beliefs are coarse, we consider a model in which an agent's expected assignment probability is given by

$$\bar{L}^{n,\sigma}_{R_i} = \int L^{n,\sigma}_{R_i,t_i} \mathrm{d}F_T. \tag{5}$$

Such beliefs may have distributional consequences, and may undo some of the goals of the Controlled Choice Plan of maintaining a diverse student mix within programs. It is possible that schools that are popular amongst paid-lunch students, for example Graham & Parks, may be under-subscribed by free/reduced lunch students because of such coarse beliefs.

4 A Revealed Preference Approach

This section illustrates the key insight that allows us to learn about the preferences of students from their (potentially manipulated) report, and present an overview of our method for estimating preferences.

Equation (2) reveals that a student's optimal choice depends on the expected assignment probabilities given her report and priority type. The choice of a report by a student can be interpreted as a choice over the set of lotteries,

$$\mathcal{L}_{t_i} = \{L_{R_i,t_i} : R_i \in \mathcal{R}_i\}.$$

These are the assignment probabilities that a student with priority type t_i believes she can achieve by making different reports to the mechanism. The various forms of beliefs described in Section 3.3 specify particular values for L_{R_i,t_i}. We will suppress the dependence on t_i in the notation for expositional simplicity, focusing on students with a given priority type.

Assume, for the moment, that a student's belief for the assignment probabilities is known to the analyst and consider her decision problem.[26] Figure 2 illustrates an example with two schools and an outside option. Each possible report corresponds to a probability of assignment into each of the schools and a probability of remaining unassigned. Figure 2(a) depicts three lotteries $L_R, L_{R'}, L_{R''}$ corresponding to the reports R, R' and R'' respectively on a unit simplex.[27] The dashed lines show the linear indifference curves over the lotteries for an agent with a utility vector parallel to the vector a. A student with a utility vector parallel to a_R will therefore find L_R optimal (figure 2(b)). A student that is indifferent between L_R

[26]Section 6.1 presents a consistent estimator for the student's believed assignment probabilities for each form of beliefs that we consider.

[27]The simplex is often referred to as the Marschak-Machina triangle.

and $L_{R'}$ must have indifference curves that are parallel to the line segment connecting the two points and, therefore, a utility vector that is parallel to $a_{R,R'}$ (figure 2(b)). Likewise, students with a utility vector parallel to $a_{R,R''}$ are indifferent between L_R and $L_{R''}$. In fact, L_R is optimal for all students with utility vectors that are linear combinations of $a_{R,R'}$ and $a_{R,R''}$ with positive coefficients. A similar reasoning can be applied to $L_{R'}$ and $L_{R''}$ resulting in the vector $a_{R',R''}$ depicted in figure 2(c). We now turn our attention to utility space in figure 2(d). The rays starting from the origin and parallel to each of these vectors partition this space. As argued above, L_R is optimal for students with utility vectors $v \in C_R$, for example if v is parallel to a_R. In symbols, for any J and set of lotteries $\mathcal{L}$, choosing L_R is optimal if and only if the utility vector belongs to the cone:

$$C_R = \{v \in \mathbb{R}^J : v \cdot (L_R - L_{R'}) \geq 0 \text{ for all } R' \in \mathcal{R}\}. \tag{6}$$

For all values of v in this cone, the expected utility from choosing R is at least as large as choosing any other report.[28] Similarly, reports R, R' and R'' are only optimal for students with utility vectors in the regions C_R, $C_{R'}$ and $C_{R''}$ respectively. Further, these regions may intersect only at their boundaries, and together cover the utility space. The choice of report therefore reveals the region in utility space to which a student's preference vector belongs.

Remarkably, a partition of this form is implied for all school choice mechanisms that use tie-breakers if a student's belief for assignment probabilities can be estimated. Further, these inequalities use all the restrictions on preferences given the beliefs of the agents and the mechanism. We can use this insight to construct the likelihood of observing a given choice as a function of the distribution of utilities:

$$\begin{aligned} \mathbb{P}(R|z,\xi) &= \mathbb{P}(R = \arg\max_{R' \in \mathcal{R}} v \cdot L_{R,t} | z, t, \xi; f) \\ &= \int 1\{v \in C_R\} f_{V|t,z,\xi}(v|z,t,\xi)\mathrm{d}v. \end{aligned} \tag{7}$$

This expression presents a link between the observed choices of the students in the market and the distribution of the underlying preferences, and will be the basis of our empirical approach. Note that the number of regions of the utility space that we can learn about from observed choices is equal to the number of reports that may be submitted to a mechanism, which grows rapidly with the number of schools or the number of ranks submitted.[29] In

[28] We allow for the possibility that $L_R = L_{R'}$ for two reports R and R'. This may occur if the student lists a school she is sure to be assigned to as her first choice. In such cases, our revealed preference method does not deduce any preference information from later ranked choices.

[29] Since students in Cambridge can rank up to three programs from 13, there are a total of 1,885 elements in $\mathcal{R}$.

the empirical application, we specify $f_V(v|z,t,\xi)$ according to equation (1) as a function of finite-dimensional parameters θ.

There are three remaining issues to consider which we address in the subsequent sections. First, we introduce a large class of mechanisms for which the equilibrium assignment probabilities can be consistently estimated. This is essential for determining the regions C_R needed to construct the likelihood. The objective is to estimate the assignment probabilities given the strategies used by students in the sample. Therefore, our procedure is robust to the possibility of multiple equilibria in case one wishes to assume that students play equilibrium strategies. Second, we provide conditions under which the distribution of utilities is non-parametrically identified. We can obtain point identification by "tracing out" the distribution of utilities with either variation in lottery sets faced by students or by using an additively separable student-school specific observable characteristic. Third, we propose a computationally tractable estimator based on Gibbs' sampling that can be used to estimate the parameters of the distribution of indirect utilities. Here, we use an estimate of the lotteries obtained from the first step.

5 Identification

In Section 4, we showed that the choice of report by a student allows us to determine the cone, $C_R \subseteq \mathbb{R}^J$ for $R \in \mathcal{R}$, that contains her utility vector v. This deduction required knowledge of student's beliefs for assignment probabilities $L_{R,t}$. Knowledge of the mechanism and the joint distribution of reports and types directly identifies beliefs over assignment probabilities as specified in equations (3)-(5).[30] This section presents our results on identification of the distribution of indirect utilities. Knowledge of this distribution is sufficient for positive analysis of various types. For example, it allows for counterfactual analysis of assignments under alternative mechanisms as well as analyzing the fraction of students that are assigned to their true first choice. Additionally, certain forms of normative analysis that involve comparing the proportion of students that prefer one mechanism over another can also be conducted.[31]

We now articulate how one can learn about the distribution of utilities using implications

[30] Section 6.1 presents a consistent estimator for $L_{R,t}$ as defined by each of the forms of beliefs in Section 3.3

[31] Solving the social planner's problem or comparing mechanisms using a Kaldor-Hicks criterion requires additional assumptions on the transferability of utility or a choice of Pareto weights.

of equation (7):

$$\mathbb{P}(R \in \mathcal{R}|z, t, \xi, b) = \int 1\{v \in C_{b,R,t}\} f_{V|z,t,\xi}(v|z, t, \xi)\mathrm{d}v,$$

where b is a market subscript and the dependence on t has been reintroduced for notational clarity. It allows us to consider different market conditions for the same set of schools or students with different priority types.

The expression above shows that two potential sources of variation are available to the analyst that can be used to "trace out" the densities $f_{V|T,z,\xi}(v|z, t, \xi)$. First, we can consider choice environments with different values of $C_{b,R,t}$. Second, we can consider variation in the observable characteristics z. We consider each of these in the subsequent sections.

As is standard in the literature on identification, our results in this section abstract away from sampling noise. Hence, we treat the assignment probabilities and the fraction of students that choose any report as observed. Non-parametric estimation of random utility models can be computationally prohibitive and imprecise in finite samples, particularly if the number of schools is large. We view these results as articulating the empirical content of the data and highlighting the sources of variation required to identify the model.

5.1 Identification Under Varying Choice Environments

In some cases, a researcher is willing to exclude certain elements of the priority structure t from preferences, or may observe data from multiple years in which the set of schools are the same, but the capacity at schools varies across years. For instance, some students are grandfathered into Kindergarten from pre-K before the January assignment in Cambridge. This affects the number of seats available at a school during this process. This variation assists in identification if it is excluded from the distribution of utilities. This section illustrates what can be learned from such variation without any further assumptions.

When t is excluded from the distribution of preferences, i.e. $v|z, \xi, t \sim v|z, \xi, \tilde{t}$ for $t, \tilde{t} \in T$, we effectively observe students with the same distribution of preferences facing two different choice sets for assignment probabilities. For example, assume that the choice sets faced by t and $\tilde{t}$ are $\mathcal{L} = \{L_R, L_{R'}, L_{R''}\}$ and $\tilde{\mathcal{L}} = \{L_R, \tilde{L}_{R'}, L_{R''}\}$ respectively. Figure 3(a) illustrates these choice sets. The change from $L_{R'}$ to $\tilde{L}_{R'}$ affects the set of utilities for which the various choices are optimal. Now, the set of types for which L_R is optimal also includes the dotted cone. These utilities in this cone can be written as linear combinations of $\tilde{a}_{R,R'}$ and $a_{R,R'}$ with positive coefficients. Observing the difference in likelihood of reporting R for students with the two types allows us to determine the weight on this region:

$$\mathbb{P}(R|z,\tilde{t}) - \mathbb{P}(R|z,t) = \int (1\{v \in \tilde{C}_R\} - 1\{v \in C_R\}) f_{V|z}(v|z) dv,$$

where explicit conditioning on z, ξ is dropped because it is held fixed. Since utilities may be determined only up to positive affine transformations, normalizing the scale as $\|v_i\| = 1$ for each student i is without loss of generality. Hence, it is sufficiently to consider the case when v_i has support only on the unit circle. Figure 3(b) illustrates that this variation allows us to determine the weight on the arc $\tilde{h}_R - h_R$. Appendix C.2 formalizes this argument and characterizes the identified set under such variation.

The discussion suggests that enough variation in the set of lotteries faced by individuals with the same distribution of utilities can be used to identify the preference distribution. If such variation is available, the arc above traces the density of utilities along the circle. However, typical school choice systems have only finitely many priority types and datasets typically cover a small number of years. Therefore, due to limited support, we will typically partially identify $f_{V|z}$. Our observations in this section articulate the sources of choice set variation that are implicitly used when utilities are not linked directly with priority types.

While this source of variation may not be rich enough for a basis for non-parametric identification, it makes minimal restrictions on the distribution of utilities. In particular, the result allows for the distribution to depend arbitrarily on residential locations. Although beyond the scope of this paper, this framework may be a useful building block for a model that incorporates both residential and schooling choices.

5.2 Identification With Preference Shifters

In this section we assume that the set of observables $z_{ij} \in \mathbb{R}^{K_z}$ can be partitioned $z_{ij}^2 \in \mathbb{R}^{K_z - 1}$ and $z_{ij}^1 \in \mathbb{R}$. The indirect utility function is additively separable in z_{ij}^1, and are therefore given by

$$V(z_{ij}, \xi_j, \epsilon_i) = U(z_{ij}^2, \xi_j, \epsilon_i) - z_{ij}^1. \tag{8}$$

We assume that $\epsilon_i \perp (z_{i1}^1, \ldots, z_{iJ}^1)$, which implies that any unobserved characteristics that affect the taste for schools is independent of z^1. The magnitude of the coefficient on z_{ij}^1 can be viewed as a scale normalization, and the model is observationally equivalent to one with student-specific tastes, α_i, for distance as long as it is negative for all i. This scale normalization replaces the normalization, $\|v_i\| = 1$, made in the previous section.

The term z_{ij}^1 is sometimes referred to as a special regressor (Lewbel, 2000; Berry and Haile, 2010). The combination of the additively separable form and independence of ϵ_i is the main restrictions in this formulation. In the school choice context, these assumptions needs to be made on a characteristic that varies by student and school. For instance, Abdulkadiroglu et al. (2015a) assume that distance to school is independent of student preferences. The assumption is violated if unobserved determinants of student preferences simultaneously determine residential choices. In our empirical application, we include an indicator for whether a student has priority at a school as a determinant of preference to (partially) control for residential preferences.

We now describe how variation in z^1 within a market can be used to learn about the distribution of indirect utilities. In this exercise, we fix the school unobservables ξ, and consider sets of students with identical values of z^2. The objective is to identify the joint distribution of $u_{ij} = U(z^2_{ij}, \xi_j, \epsilon_i)$ given ξ, z^2, where we drop this conditioning for simplicity of notation because it is held fixed. Since ϵ_i is independent of z^1_{ij}, we have that $f_{V|Z^1}(v|z^1) = g(v + z^1)$ where g is the density of $u = v + z^1$. Our objective is therefore to identify the density g since z^1_{ij} is observed.

Our next result shows that variation in z^1 can be used to "trace-out" the density of u. Consider the lottery set faced by a set of students in figure 2 and the corresponding region, C_R, of the utility space that rationalizes choice R. A student with $z^1_i = z$ chooses R if $v = u - z \in C_R$. The values of u that rationalize this choice is given by $z + \alpha_1 a_{R,R'} + \alpha_2 a_{R,R''}$ for any two positive coefficients α_1 and α_2. Figure 4 illustrates the values of u that make R optimal. As discussed in Section 4, observing the choices of individuals allows us to determine the fraction of students with utilities in this set. Similarly, by focusing on the set of students with $z^1_i \in \{z', z'', z'''\}$, we can determine the fraction of students with utilities in the corresponding regions. Figure 4 illustrates the sets that make R optimal for each of these values of z^1. By appropriately adding and subtracting the fractions, we can learn the fraction of students with utilities in the parallelogram defined by $z - z' - z''' - z''$. This allows us to learn the total weight placed by the distribution g on such parallelograms of arbitrarily small size. It turns out that we can learn the density of g around any point z in the interior of the support of z^1 by focusing on local variation around z. The next result formalizes this intuition.

Theorem 1. *Suppose C_R is spanned by J linearly independent vectors $\{w_1, \ldots, w_J\}$. If $h_{C_R}(z^1) = P(v \in C_R|z^1)$ is observed on an open set containing z^1, then $g(z^1)$ is identified. Hence, $f_V(v|z^1)$ is identified if $v + z_1 \in int(\zeta) \subseteq \mathbb{R}^J$, where ζ be the support of z^1.*

Proof. Let $W = (w'_1, \ldots, w'_J)'$ be the matrix containing linearly independent vectors such that $C_R = \{v : v = -Wa \text{ for some } a \geq 0\}$. Assume, wlog, $|\det W| = 1$. Evaluating h_{C_R} at

Wx, we have that

$$h_{C_R}(Wx) = \int_{\mathbb{R}^J} 1\{u - Wx \in C\} g(u)\, \mathrm{d}u.$$

After the change of variables $u = Wa$:

$$\begin{aligned} h_{C_R}(Wx) &= \int_{\mathbb{R}^J} 1\{W(a-x) \in C_R\} g(Wa) da \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_J} g(Wa)\, da \end{aligned}$$

where the second inequality follows because $1\{W(a-x) \in C_R\} = 1\{a - x\} \leq 0$. Then:

$$\frac{\partial^J h_{C_R}(Wx)}{\partial x_1 \ldots \partial x_J} = g(Wx)$$

and $g(z^1)$ is given by $\frac{\partial^J h_C(Wx)}{\partial x_1 \ldots \partial x_J}$ evaluated at $x = W^{-1} z^1$.

Note that $f_V(v|z_1) = g(v + z_1)$. Since $g(z)$ is identified on the interior of ζ, $f_V(v|z_1)$ is identified if $v + z_1$ is in the interior of ζ as well. □

Intuitively, z_i^1 shifts the distribution of indirect utilities of students. For example, when z_{ij}^1 denotes the distance to school j, then all else equal, students closer to a particular school should have stronger preferences for attending that school. These students should be more likely to rank it on their list, even if the probabilities of getting assigned are not particularly high. The extent to which students that are closer to a given school are more likely to rank it is indicative of the importance of distance relative to other factors that affect preferences that are captured by $U(\cdot)$. The result formalizes how such variation can be used to learn about the distribution of indirect utilities.[32]

The formal result above considers cones C_R that are spanned by linearly independent vectors. In cases where C_R is not spanned by linearly independent vectors, we can still identify g if z^1 has full support on $\mathbb{R}^J$ as long as the tails of g are exponentially decreasing. Theorem C.3 in Appendix C.3 states the results and conditions formally. The proof is based on Fourier-deconvolution techniques since the distribution of v is given by a location family parametrized by z^1. The conditions on g are quite weak, and are satisfied for commonly used distributions with additive errors such as normal distributions, generalized extreme value distributions or if u has bounded support.[33]

[32]The local nature of this identification result articulates precisely the fact that identification of the density at a point does not rely on observing extreme values of z^1. Of course, identification of the tails of the distribution of u will rely on support on extreme values of z^1. Also note that our identification result requires only one convex cone generated by a lottery, and therefore, observing additional lotteries with simplicial cones generates testable restrictions on the special regressor.

[33]We do not require that g has a non-vanishing characteristic function. When u has bounded support, the

When z^1_{ij} is assumed to be the road or walking distance from student i's residence to school j, then the support of z^1 will be limited by geographical constraints. In this case, such variation provides partial information on f_V. Our estimator, which is described in the next section, will use variation from this source in addition to variation in choice environments in a parametric specification.

6 Estimating Assignment Probabilities and Preference Parameters

We estimate our preference parameters, $\theta = (\beta, \Sigma)$, using a two-step estimator where we first estimate $L^{n,\sigma}_{R,t}$ for the assignment probabilities given report R and priority type t. For simplicity of notation, we omit the dependence of L on the application year. We use administrative data on all the submitted applications, data on school capacities in each year and knowledge of the assignment mechanism in this step.[34]

The second step is defined as an extremum estimator:

$$\hat{\theta} = \max_{\theta \in \Theta} \; Q_n(\theta, \hat{L}).$$

In our case, Q_n is a likelihood function. Specifically, equation (7) implies that the likelihood is given by

$$Q_n(\theta, \hat{L}) = \max_{\theta \in \Theta} \; \sum_{i=1}^{n} \log \mathbb{P}(R_i = \arg\max_{R \in \mathcal{R}_i} v_i \cdot \hat{L}_{R,t_i} | z_i, t_i; \theta), \tag{9}$$

where R_i is the report submitted by student i, z_i is the vector of observables that the distribution of v_i depends on, t_i is the priority type of agent i, and θ parametrizes the distribution of v as given in Section 3.1. We will implement our second-step as a Gibbs' sampler, and interpret the posterior mean of this sampler as asymptotically equivalent to the Maximum Likelihood Estimator (see van der Vaart, 2000, Theorem 10.1 (Bernstein-von-Mises)).

Our asymptotic framework assumes that the number of students n grows large, and the number of programs is held fixed. The capacity of the programs, $n \times q^n_j$, increases proportionally to the number of students, i.e. $q^n_j \to q_j \in (0, 1)$. These limits are meant to

support conditions on ζ can also be relaxed. In this case, we can allow for ζ to be a corresponding bounded set.

[34]Knowledge of the mechanism and school capacities is essential for our procedure. One may use a sample of reports from the population of students by weighting the observations so that the total number of applicants and the distribution of their types matches the true set of applicants.

capture an environment, such as the one in Cambridge, where the number of students is large relative to the number of schools.[35] It is sufficient for the researcher to observe data from a single year of the mechanism with many students for consistency and asymptotic normality of the estimator. For parsimony, the parameters governing the distribution of preferences given the observables, θ, is held constant across years. Therefore, in our applications, we use multiple years of data to improve the precision of our estimate for θ.

Theorem 2.1 and 6.1 of Newey and McFadden (1994) show conditions under which a two-step estimator is consistent and asymptotically normal. These results require that the first-step is consistent and asymptotically normal, and the second-step is reasonably well-behaved. As discussed above, our second-step estimator can be viewed as equivalent to a maximum likelihood estimator. We now study the limit behavior of our first step and outline the computational procedure used in the second step.[36]

6.1 First Step: Estimating Assignment Probabilities

The first step requires us to estimate the probabilities in equation (3) for each value of (R_i, t_i). The expression in this equation can be re-written as:

$$L^{n,\sigma}_{R_i,t_i} = \sum_{(R_{-i},T_{-i})} \Phi^n((R_i,t_i),(R_{-i},T_{-i})) \prod_{k\neq i} f_\sigma(R_k,t_k)$$

where $f_\sigma(R,t) = f_T(t)\int \sigma_R(v,t)\mathrm{d}F_{V|T=t}$ is the probability that an agent with priority type t reports R and $F_{V,T}$ is the distribution of utility and priority types.

The equation highlights two sources of uncertainty facing the student when forming this expectation. First, at the time of submitting the report, the student does not know the realization of (R_{-i}, T_{-i}). The agents form expectations for this realization based on population distribution of types and the forecast strategy σ. The second source of uncertainty lies in the lottery draw used to determine assignments is uncertain after the rank-order lists are submitted, as given by Φ^n.

The first source of uncertainty can be approximated using a resampling procedure since the data consists of a large sample of reports and priorities drawn from this distribution. This suggests that

$$L^{n,\sigma}_{R_i,t_i} \approx \frac{1}{B}\sum_{b=1}^{B} \Phi^n((R_i,t_i),(R_{-i},T_{-i})_b), \tag{10}$$

[35] Azevedo and Leshno (2013) use similar limits to analyze properties of stable matchings in a large market.

[36] Our results present bootstrap standard errors for our estimates using the procedure described in Appendix D.5.

where $(R_{-i}, T_{-i})_b$ is the b-th sample of $n-1$ reports and priority types.

The second source of uncertainty, given by Φ^n, arises from a known tie-breaking procedure as specified by the mechanism and can therefore be simulated. Here, we can use the fact that Cambridge uses a mechanism with an Report-Specific Priority + Cutoff (RSP+C) representation. For each draw of $(R_{-i}, T_{-i})_b$, and a draw of $\nu_{b,-i}$, we can compute a market clearing cutoff p_b^{n-1} by simulating the mechanism using only those agents. Our estimator approximates the cutoffs by ignoring the report of agent i since, in a large market, any single agent has a negligible impact on the cutoffs.[37] The assignment for a student with priority type t_i, report R_i, and a draw of the random tie-breaker ν_i is given by $D^{R_i,t_i,\nu_i}(p_b^{n-1})$. We therefore estimate $L^{n,\sigma}_{R_i,t_i}$ for each feasible pair of (R_i, t_i) by calculating

$$\hat{L}_{R_i,t_i} = \frac{1}{B}\sum_{b=1}^{B}\int D^{(R_i,t_i,\nu_i)}(p_b^{n-1})\mathrm{d}\gamma_{\nu_i}, \tag{11}$$

where γ_{ν_i} is the cdf of the tie-breaker. This estimator uses a single set of draws for p_b^{n-1} for the various values of (R_i, t_i), making the computation tractable even with a large number of possible rank-order lists.

Our resampling estimator therefore approximates the two sources of uncertainty faced by students when submitting their preference lists mentioned above. Even with rational expectations, these two sources of uncertainty imply that the realized empirical assignment probabilities differ from the agent's expectations.[38]

Equation (10) highlights why consistently estimating assignment probabilities for general mechanisms may be difficult. These probabilities are a result of mechanisms that are usually described in terms of algorithms that using a profile of reports and priority types of all the students in the district. There are few a priori restrictions on these algorithms, allowing for mechanisms that may be ill-behaved. For instance, a small changes in students' reports could potentially have large effects on the assignment probabilities.[39] Moreover, our objective is to estimate assignment probabilities simultaneously for all priority-types and each possible rank-order list that can be submitted by a student.

[37] This intuition is formalized in Appendix B, where we show that the approximation error in using only the other $n-1$ agents is of order $1/n$.

[38] Additionally, the sample assignment probabilities face the curse of dimensionality since the number of assignment probabilities to be estimated equals the number of possible rank-order lists times the number of priority types. Our resampling estimator uses the fact that the assignment probabilities depend only on the number of cutoffs, which is equal to the number of schools. In Cambridge, there are just under 2,000 feasible rank-order lists for each priority type. Our sample consists of 2,129 students.

[39] Two pathological examples allowed by Definition 1 are instructive. The first example is one in which the assignment of all students depends only student 1's report. The second is an algorithm that depends on whether an odd or even number of students apply to schools.

Equation (11) shows that the problem may be tractable for mechanisms that have an RSP+C representation. Specifically, it presents a route for obtaining results on the limiting distribution of $\hat{L}$ by examining the limit behavior of the cutoffs, p_b^{n-1}.

Theorem 2. *Suppose that Φ^n is an RSP+C mechanism with a lottery draw $\nu_{ij} = \nu_{ij'} = \nu_i$ that is drawn from the uniform distribution on $[0, 1]$.*

(i) If p^* is the unique market clearing cutoff for $(E[D(p)], q)$, where $q = \lim q_j^n$, then for each (R, t), $|\hat{L}_{R,t} - L_{R,t}^{n,\sigma}| \xrightarrow{p} 0$.

(ii) Further, if $p_j^* > 0$ for all j, $\|q^n - q\| = o_p(n^{-1/2})$ and $\nabla_p E[D(p^*)]$ is invertible, then for each each (R, t),

$$\sqrt{n}(\hat{L}_{R,t} - L_{R,t}^{n,\sigma}) \xrightarrow{d} N(0, \Omega^*)$$

where Ω^* is given in Theorem B.1.

Proof. The result is a special case of Theorem B.1 in the Appendix B, which relies on less restrictive assumptions, and does not assume that $p_j^* > 0$. □

This result is based on showing that the cutoffs determining the eligibility thresholds are close to p^* if the number of students is large, and then analyzing the limit distribution of $\sqrt{n}(p_b^{n-1} - p^*)$. As discussed in Appendix B, the uniqueness of the cutoff p^* is a generic property.[40] In the Cambridge mechanism uniqueness of p^* follows if for all j, $D_j(\cdot)$ is strictly decreasing in p_j (see Proposition B.5 in the Appendix). Further, the assignment probabilities $L_{R,t}$ are a smooth function of these cutoffs since ν has a non-degenerate distribution. This cutoff representation also highlights that agents need not have beliefs over a very high dimensional object in order to compute best responses.[41] Additionally, as formalized in Proposition B.6, our results also imply convergence of finite market equilibria to large-market limits.

6.2 Second Step: Preference Estimates

For this step, we adapt the Gibbs' sampler used by McCulloch and Rossi (1994) to estimate a discrete choice model. The Gibbs' sampler obtains draws of β and Σ from the posterior distribution by constructing a Markov chain of draws from any initial set of parameters $\theta^0 = (\beta^0, \Sigma^0)$. The invariant distribution of the Markov chain is the posterior given the prior

[40] Specifically, we use results from Azevedo and Leshno (2013) to show that market clearing cutoffs are generically unique. Using techniques from Berry et al. (2013) and Berry and Haile (2010), we can also derive stronger conditions for global uniqueness of the market clearing cutoffs.

[41] Work in progress by Kapor et al. (2016) uses a survey of students in New Haven to construct estimates of students' beliefs over these cutoffs. They then extend our procedure to estimate the preference distribution using these estimated beliefs.

and the data. It offers a computationally convenient likelihood-based method for estimating parameters in some cases when an analytic form for the likelihood function is not available.

As in the discrete choice case, we first use data augmentation to pick a utility vector for each agent consistent with their choice. Here, we initialize $v_i^0 \in C_{R_i}$ for each student i, where R_i is the report chosen by agent i. Since C_{R_i} is defined by linear inequalities given in equation (6), we set v_i^0 by using a linear programming solver to find a solution to the constraints $v_i^0 \cdot (\hat{L}_{R_i,t_i} - \hat{L}_{R',t_i}) \geq 0$ for all $R' \in \mathcal{R}$.[42]

The chain is then constructed by sampling from the conditional posteriors of the parameters and the utility vectors given the previous draws. The sampler iterates through the following sequence of conditional posteriors:

$$\begin{aligned} \beta^{s+1} &\mid v_i^s, \Sigma^s \\ \Sigma^{s+1} &\mid v_i^s, \beta^{s+1} \\ v_i^{s+1} &\mid v_i^s, C_{R_i}, \beta^{s+1}, \Sigma^{s+1} \end{aligned}$$

The first step updates the parameter β of equation (1). We use the standard procedure in Bayesian approaches to draw β^{s+1} from the posterior distribution of β given its prior, the augmented data (v^s, x) and the distribution of error terms $N(0, \Sigma^s)$. A new draw Σ^{s+1} is drawn from the posterior distribution of Σ given the prior and ε^{s+1}, which can be solved for using equation (1), v_i^s and β^{s+1}. The last step draws v_i^{s+1} for each student. This occurs by iterating through the various schools and sampling from the following conditional posteriors:

$$v_{ij}^{s+1} | v_{i1}^{s+1}, \ldots, v_{ij-1}^{s+1}, v_{ij+1}^{s}, \ldots, v_{iJ}^{s}, \beta^{s+1}, C_{R_i}, \Sigma^{s+1}.$$

This step requires us to draw from a (potentially two-sided) truncated normal distribution with mean, variance and truncation points determined by β^{s+1}, Σ^{s+1}, C_{R_i} and $v_{i,-j}$.[43] We can ensure that $v_i^{s+1} \in C_{R_i}$ for every student i in every step by calculating the bounds on v_{ij}^{s+1} conditional on $v_{i,-j}$ defined by the restriction $v_i^{s+1}(L_{R_i} - L_R) \geq 0$ for all $R \in \mathcal{R}$.

We specify independent and diffuse prior distributions for $\beta = \{\beta_{jk}\}_{j=1..J,k=1..K} \in \mathcal{R}^{JK}$ and Σ. It is convenient to use a normal prior on β, $\beta \sim N(\overline{\beta}, A^{-1})$ and an independent inverse Wishart prior on Σ,$\Sigma \sim IW(\nu_0, V_0)$. These priors are convenient because (conditional) conjugacy is maintained at each step of the algorithm. Additional details on the

[42] There are 1,885 constraints in Cambridge, one for each possible rank-order list. A linear programming solver can be used to eliminate linearly dependent constraints with positive coefficients in order to further simplify the later stages of the Gibbs' sampler.

[43] Our problem is therefore slightly is different from, although not more difficult than, a Gibbs' sampler approach to estimating standard discrete choice models in McCulloch and Rossi (1994). The standard discrete choice models only involve sampling from one-sided truncated normal distributions.

implementation of our Gibbs' sampler are in Appendix D.1.

7 Application to Cambridge

7.1 Estimates: Assignment Probabilities and Preference Parameters

This section presents our estimates for the assignment probabilities obtained using the approach described in Section 6.1 and estimates for the preference parameters. Table 6 presents the assignment probabilities for various schools, averaged over various student subgroups.[44] As in table 4, the estimates indicate considerable heterogeneity in school competitiveness. The typical student isn't guaranteed assignment at the more competitive schools even if she ranks it first. On the other hand, several schools are sure shots for students that rank them first. The probability of not getting assigned to a school also differs with paid-lunch status. A comparison of estimates in panel A with those in panels D and E indicates that having priority at a school significantly improves the chances of assignment. The differential is larger if the school is ranked first.

Panel A of table 7 presents the (normalized) mean utility for various schools net of distance, by student group for four specifications. The first specification treats the agent reports as truthful, while the second, third and fourth specifications assume that students best respond to beliefs given by rational expectations, adaptive expectations and coarse beliefs as defined in Section 3.3.[45] In each of these specifications, we find significant heterogeneity in willingness to travel for the various school options. Paid-lunch students, for instance, place a higher value on the competitive schools as compared to the non-competitive schools. Although not presented in the mean utilities, Spanish and Portuguese speaking students disproportionately value schools with bilingual and immersion programs in their home language. Students also place a large premium on going to school with their siblings.

A comparison between the first column and the others suggests that treating stated preferences as truthful may lead to underestimates of the value of competitive schools relative to non-competitive schools. This differential is best illustrated using Graham & Parks as an example. Treating stated preferences as truthful, we estimate that paid-lunch students have an estimated mean utility that is an equivalent of 1.29 miles higher than the average public school option. This is an underestimate relative to the models that assumes that students

[44] Table D.1 provides an estimate for standard errors of $\hat{L}$ constructed by bootstrapping the estimator.

[45] The underlying parameter estimates for the two baseline specifications, rational expectations and truthful reporting, are presented in tables D.3 and D.2.

correctly believe that Graham & Parks is a competitive school. In contrast, the value of Graham & Parks for free-lunch students is over-estimated by the truthful model relative to both the rational expectations and adaptive expectations model. The difference can be explained by observing that Graham & Parks is not competitive for free-lunch students, and therefore, the low number of applications it receives indicates particular dislike for the school from this group of students.

Overall, estimates based on modeling expectations as adaptive are strikingly similar to those from assuming rational expectations. In part, this occurs because the relative competitiveness of the various schooling options in Cambridge is fairly stable even though there is some annual variation in assignment probabilities across school. This result is comforting for the robustness of our estimates to small mis-specifications of agent beliefs. Estimates that endow agents with coarse beliefs continue to indicate that treating reports as truthful underestimates the relative preference for the most competitive schools such as Graham & Parks, Haggerty, Baldwin and Morse. The results are more mixed for the less desirable schools. As in the models that treat preferences as truthfully reported, free-lunch and paid-lunch students are in broad agreement on the relative ranking of the various schools.

Another significant difference between the estimates that treat agents as truthful and those that do not is in the number of schools students find preferable to the outside option. Panel B shows that estimates that treat stated preferences as truthful suggest that about half the students have five or more schools where assignment is preferable to the outside option. On the other hand, treating agents as best responding to one of the three forms of beliefs studied here suggests that about half the students find at most two schools in the system preferable to the outside option. Treating preferences as truthful extrapolates from the few students (about 13%) that do not have complete rank order lists. On the other hand, the model that treats students as being strategic interprets the decision to rank long-shots in the second and third choices as evidence of dislike for the remaining schools relative to the outside option.

These results should be viewed in light of Cambridge's thick after-market. About 92% of the students that are not assigned though the school choice process are assigned to one of the schools in the system. In fact, more than a quarter of unassigned students are placed at their top ranked school through the wait-list. There are also charter school and private school options that unassigned students may enroll in. The value of the outside option is therefore best interpreted in terms of the inclusive value of participating in this after-market.[46]

[46]Students that are assigned through the process can enroll in other schools with open seats, approximately 91% of the students register at their assigned school. Some differences between assignments and registrations can be caused by changed in student preferences or the arrival of new information Narita (2016). Nonetheless, since the wait-list process in Cambridge allows students choose a set of schools to apply for, we explored

The specifications estimated the preference parameters using the set of students that submitted a rank-order list consistent with optimal play (i.e. submitted a list corresponding to an extremal lottery). For the rational expectations model, a total of 2,071 students (97.3% of the sample) submitted a rationalizable list.[47] The large fraction of students with rationalizable lists may initially appear surprising. However, theorem C.1 in the appendix indicates that the lists that are not rationalized are likely the ones where assignment probabilities for one of the choices is zero. Our estimates in table 6 suggest that this is rare, except for a few schools. Most of the students with lists that cannot be rationalized listed Graham & Parks as their second choice. Indeed, the reports can be rationalized as optimal if agents believe that there is a small but non-zero chance of assignment at these competitive schools. One concern with dropping students with lists that cannot be rationalized is that we are liable to underestimate the desirability of competitive schools. Although not reported, estimates that add a small probability of assignment to each of the ranked options yield very similar results.

7.2 Ranking Behavior, Out-of-Equilibrium Truthtelling and Assignment to Top Choice

In this section we investigate the ranking strategy of agents, whether they would suffer large losses from out-of-equilibrium truth-telling, and how strategic manipulation may affect student welfare.

Table 8 presents the fraction of students that find truthful reporting optimal and losses from truthful behavior relative to optimal play as estimated using the two polar assumptions on student behavior and beliefs. The first three columns are based on the assumption that the observed reports are truthful and analyze the losses as a result of such naïvete. These estimates can be interpreted as analyzing the true loss to students from not behaving strategically if they are indeed out-of-equilibrium truth-tellers. The estimates suggest that the truthful report is optimal for 57% of the students. The average student suffers a loss equal to 0.18 miles by making a truthful report, or 0.42 miles conditional on regretting truthful behavior. We also estimate heterogeneous losses across student groups. Free-lunch students, for instance, suffer losses from truthful play less often and suffer lower losses conditional on any losses. This reflects the fact that the Cambridge school system is not competitive for

whether this feature results in significant bias. Specifically, we estimated the probability that a student is able to ultimately register at a school where she was rejected during the main application process. Beliefs based on these probabilities resulted in quantitatively similar results to our baseline specifications. We therefore avoid modeling the after-market.

[47] One student was dropped because the recorded home address data could not be matched with a valid Cambridge street address.

these students because of the seats specifically reserved for this group.

The last three columns use estimates based on rational expectations and tabulate losses from non-strategic behavior.[48] Again, these estimates suggest that about half the students, and disproportionately paid-lunch students have strategic incentives to manipulate their reports. Together, the observations suggest that markets where students face large competitive pressures are precisely the markets where treating preferences as truthful may lead to biased assessments of how desirable various schools are.

The estimated losses using both specifications may seem small on first glance, but can be explained by noting that whenever a student has a strong preference for a school, she will rank it as her first choice in her optimal report (and potentially manipulate lower ranked choices). The priority given to the first ranked choice results in a low chance that the student is not assigned to this highly desired school. This fact significantly lowers the potential of large losses from truthful reporting.

Our estimates that about half the students find it optimal to behave truthfully is likely to affect our assessment of how many students are assigned to their top choice. Table 9 presents this fraction by student paid-lunch status. The last column indicates that 85.2% of the students rank their top choice first. This occurs because many students avoid ranking competitive schools as their top rank in favor of increasing the odds of assignment to a less preferred option. As a result, fewer students rank Graham & Parks as their top choice, instead favoring Haggerty or Baldwin. We therefore see over-subscription to Haggerty and Baldwin by paid-lunch students relative to the true first choice. The last column indicates that while 83.4% were assigned to their stated first choice, only 72.3% were assigned to their true first choice. This pattern is particularly stark for paid-lunch students, who are assigned to their true first choice only 64.6% of the time. Table 6 indicated that assignment to competitive schools is less likely for paid-lunch students. Together, these results suggest that calculations of whether students are assigned to their preferred options based on stated preferences may be misleading, and differentially so by student demographics.

7.3 Evaluating Assignments under Alternative Mechanisms

A central question in the mechanism design literature is whether variants of the Boston Mechanism are worse for student welfare as compared to strategy-proof mechanisms such as the Deferred Acceptance Mechanism. This question has been debated in the theoretical literature with stylized assumptions on the preference distribution (see Miralles, 2009; Abdulkadiroglu et al., 2011; Featherstone and Niederle, 2011). The Boston mechanism exposes

[48]These estimates differ from the ones based of truthful reporting only because of differences in preference parameters.

students to the risk that they are not assigned to their top listed choices, which can harm welfare when they strategically choose not to report their most preferred schools. However, this risk has a countervailing force that only agents with particularly high valuations for their top choice will find it worthwhile listing competitive schools on top. Hence, the mechanism screens agents for cardinal preferences and can result in assignments with higher aggregate student welfare. Addtionally, assignments under the Boston mechanism may be preferable under a utilitarian criterion because they need not eliminate justified envy (equivalently, may not be stable). These are situations in which a student envies the assignment of another student even though the envied student has lower priority at that school.

Table 10 presents a quantitative comparison between the Cambridge mechanism and the Student Proposing Deferred Acceptance mechanism[49] using the preference estimates presented earlier. Because the Deferred Acceptance Mechanism is strategy-proof, evaluating the counterfactual market with this mechanism is relatively straightforward and does not require computing an equilibrium.[50]

An approach that treats agents' stated preferences in the Cambridge Mechanism as truthful finds little difference in the average welfare between the two mechanisms. Even though agents that behave truthfully risk losing out on their lower ranked choices, panel B shows that a large fraction of students are assigned to their top choice under the Cambridge mechanism due to the additional priority awarded to students at schools that are ranked first. This feature of the mechanism results in instances of justified envy. Treating preferences as truthfully reported, about 10% of students prefer the assignment of another student that has lower priority at that school. This may seem like a small number at first glance, but note that potential instances of justified envy are limited because a large majority of students are assigned to their top choice.

These estimates may be biased if strategic behavior is widespread. In contrast to estimates assuming truthful behavior, the results that treat agents as responding to strategic incentives indicate that the assignments produced by the Cambridge mechanism are preferable to those produced by the Deferred Acceptance mechanism. The fraction of students assigned to their true first choice choice remains higher under the Cambridge mechanism. Interestingly, the Cambridge mechanism also places students at their true second choices with high probability if agents are strategic. This is a consequence of strategic behavior be-

[49]We construct a Deferred Acceptance mechanism by adapting the Cambridge Controlled Choice Plan. Schools consider students according to their priority + tie-breaking number. A paid-lunch student's application is held if the total number of applications in the paid-lunch category is less than the number of available seats and if the total number of held applications is less than the total number of seats. Free-lunch student applications are held in a similar manner. We allow students to rank all available choices.

[50]It may be possible to simulate counterfactual equilibria for manipulable Report-Specific Priority + Cutoff mechanism since only equilibrium cutoffs need to be obtained.

cause some students report their true second choice as their top choice. Further, we estimate that there are fewer instances of justified envy if agents have rational expectations (5.6% instead of about 10%) because of the greater ability to obtain assignment at one of the top two choices.

Panel C shows that more students prefer the Cambridge mechanism's assignments to the Deferred Acceptance mechanism's assignments than the other way around. This observation suggests that the mechanism is effectively screening based on cardinal utilities. The average student prefers the assignments under the Cambridge mechanism by an equivalent of 0.07 miles. This magnitude is similar to the difference between Deferred Acceptance and Student Optimal Stable Matching in New York City, as measured by Abdulkadiroglu et al. (2015a). However, the Cambridge mechanism does not result in a Pareto improvement relative to the Deferred Acceptance mechanism. The table also illustrates differences across student groups. Paid-lunch students prefer the Cambridge assignments more than free/reduced lunch students.

However, specifications with biased beliefs indicate that the cardinal screening benefits of a Boston-like mechanism may be diminished and instances of justified envy may be larger if beliefs are not well aligned with true assignment probabilities. In the models with biased beliefs, free-lunch students tend to prefer the assignment produced by the Deferred Acceptance mechanism relative to the one produced by the Cambridge mechanism. Further, the benefits to paid-lunch students are lower than the model that treats agents as having rational expectations. The significant aggregate benefits to free-lunch students under the Deferred Acceptance mechanism is driven, in part, by the large fraction of students assigned to their top two choices. Paid-lunch students continue to prefer assignments in the Cambridge mechanism to the strategy-proof counterpart.

Our quantitative results contribute to the debate in the theoretical literature about the welfare properties of the Boston mechanism, which is similar to the Cambridge mechanism. The results are different in spirit from Ergin and Sonmez (2006), that suggests that full-information Nash equilibria of the Boston Mechanism are Pareto inferior to outcomes under Deferred Acceptance. This difference stems from our focus on beliefss that account for ex-ante uncertainty faced by the students. Abdulkadiroglu et al. (2011) theoretically show that the Boston mechanism can effectively screen for the intensity of preferences and can have better welfare properties than the Deferred Acceptance mechanism. Troyan (2012) shows that the theoretical results in this literature that are based on notions of interim efficiency are not robust to students having priorities, and advocates for an ex-ante comparison such as the one performed in this paper.

It is important to note that agents may face costs of strategizing since students may need to gather additional information about the competitiveness of various schools before formulating ranking strategies. These costs may weigh against using Boston-like mechanisms for school assignment. Additionally, there may be distributional consequences if agents vary in their ability to strategize (Pathak and Sonmez, 2008). While we cannot quantify the direct costs of strategizing and gathering information with out data, we extend our model to address distributional consequences of heterogeneous sophistication and biased beliefs in the next section.

7.4 Extension: Heterogeneous Agent Sophistication

Another possible violation of equilibrium behavior may arise from a population of agents that differ in their ability to strategize when reporting preferences. These differences may be driven by either heterogeneity in the information about the competitiveness of various schools or a mis-understanding of the mechanism. There are a large number of possible ways in which agents may differ in their ability to game the mechanism. The difficulty in empirically analyzing extremely flexible models of heterogeneous sophistication stems from the fact that a researcher has to disentangle heterogeneity in sophistication from preference heterogeneity by simply observing the actions of the agents. Theorem C.1 in the appendix shows it is typically possible to rationalize each submitted rank order list as optimal for some vector of utilities for the various schools. Simultaneously identifying preferences and heterogeneity in sophistication will therefore be based on restricting behavioral rules and parametric assumptions.

We estimate a stylized model with heterogeneous agent sophistication based on Pathak and Sonmez (2008).[51] They theoretically compare the Deferred Acceptance mechanism to the Boston mechanism using a model with two types of agents: naïve and sophisticated. Naïve agents report their preferences sincerely by ranking the schools in order of their true preferences. Sophisticated agents, on the other hand, recognize that truthful reporting is not optimal because schools differ in the extent to which they are competitive and the details of the mechanism. Reports made by sophisticated agents are optimal given the reports of the other agents.

In the model, the population consists of a mixture of sophisticated and naïve agents that have the same distribution of preferences but differ in their behavior. Naifs report their preferences truthfully while sophisticated agents report optimally given their (correct) beliefs about the probability of assignment at each option given their report. The distribution of

[51] See Calsamiglia et al. (2016) for another model of agents that are heterogeneous in their sophistication. They use a simulated maximum likelihood to estimate a parametric model with agents that follow one of two specific rules of thumb when making their reports.

preferences is parametrized as in equation (1). In addition to parametric assumptions, the model embeds several two strong restrictions. First, it is a mixture of two extreme forms of agent behavior: perfect sophistication and complete naïvete. Second, the distribution of preferences does not depend on whether the agent is sophisticated. These simplifications allow us to keep the estimation procedure tractable. Appendix D.2 details the Gibbs' sampler for this model, which needed to be modified. The model does not require us to re-estimate the first-step assignment probabilities.

Table 11 presents the estimated mean utilities and the fraction of agents that are naïve. The estimated mean utilities are similar to the estimates in the other specifications, and usually in between the specifications treating agents as either truthful or fully sophisticated (table 7). Panel B shows that about a third of paid-lunch and free-lunch students are estimated to be naïve. These results contrast from estimates obtained by Calsamiglia et al. (2016) in Barcelona, where they estimate that over 94% of households are strategic in their decisions. One potential reason driving this difference is that the 93% of students in Barcelona are assigned to their top ranked choice, while in Cambridge, this number is only 84%.

Table 12 describes the differences between outcomes in the Cambridge and the Deferred Acceptance mechanism. Since Deferred Acceptance is strategy-proof, both naïves and sophisticates report their preferences truthfully; therefore, their outcomes are identical. The fractions of students assigned to their first, second and third choices are similar to the results presented previously. We also see a similar overall increase in the fraction of students assigned to their top choice school in the Cambridge mechanism and a decrease in fractions assigned at lower ranked choices. Interestingly, the probability of a student assigned to their top choice under the Cambridge mechanism is larger for naïve agents than for sophisticated agents even though they have identical preferences (78.4% vs 76.2%). This relatively larger probability of assignment at the top choice is at the cost of a significantly lower probability of assignment at the second choice, which is 6.6% for naifs and 12.4% for sophisticates. These differences are particularly stark for the paid-lunch students who face a more complex strategic environment. Our estimates suggest that, relative to sophisticates, naive students effectively increase their chances of placement at their top choice school at the cost of loosing out at less preferred choices.

These results can be explained by the difference between the propensity of naifs and sophisticates for ranking popular schools. While naïve students disregard that a school is competitive, sophisticates are likely to avoid ranking competitive schools. Therefore, naifs effectively gain priority at their top choice school relative to sophisticated students with similar preferences that may not rank that school first. Although not reported, Graham

& Parks is estimated to be the top choice for 17.7% of students, but about a third of the sophisticated students for whom it is the top choice avoid ranking it first. Consequently, naive students are about 10% more likely to be assigned to Graham & Parks if it is their first choice. Qualitatively similar patterns hold for the other competitive schools such as Haggerty, Baldwin and Morse. This increase in assignment probability at the top choice comes at the cost of a reduction in the probability of assignment to the second choice. For example, while 14.7% of sophisticated paid-lunch students are assigned to their second choice school, only 6.6% of naïve paid-lunch students get placed at their second choice. As Pathak and Sonmez (2008) pointed out, naïve students effectively "lose priority" at their second and lower choice schools to sophisticated students that rank the school first. It is therefore not surprising that the instances of justified envy are largest amongst naïve students, and particularly paid-lunch naifs. About 17% of paid-lunch naifs remain unassigned while about 6% of paid-lunch sophisticates are unassigned. Further, of the 27.6% paid-lunch naifs that are not assigned to their top choice, about two-thirds have justified envy for another student's assignment.

The aggregate welfare effects for naïve students therefore depends on whether the benefits of increased likelihood of assignment at the top choice outweighs the lost priority at less preferred options. Although the naïve agents are making mistakes in the Cambridge mechansim, our comparison of assignments under the Deferred Acceptance mechanism to those under the Cambridge mechanism in panel B of table 10 shows that only 35.3% of the naive paid-lunch students prefer the Deferred Acceptance mechanism to the Cambridge mechanism. This compares with 24.9% for paid-lunch sophisticates and less than 50% for free-lunch naifs and free-lunch sophisticates. Overall, we find that the average naïve student prefers assignments under the Cambridge mechanism by an equivalent of 0.010 miles. Since sophisticates are optimally responding to incentives in their environment, their estimated value for the assignments in the Cambridge mechanism is larger, at an equivalent of 0.081 miles.

8 Conclusion

We develop a general method for analyzing preferences from reports made to a single unit assignment mechanism that may not be truthfully implementable. We view the choice of report as a choice from available assignment probabilities. These probabilities can be consistently estimated under a weak condition on the convergence of a sequence of mechanisms to a limit. The condition is verified for a broad class of school choice mechanisms including the Boston mechanism and the Deferred Acceptance mechanism. Using these probabilities, we characterize the identified set of preference distributions under the assumption that agents play optimally given their beliefs. The set of preference distributions are typically not point identified, but may be with sufficient variation in the lottery set. We then obtain point identification if a special regressor is available.

The baseline model in this paper assumes that sophisticated agents with correct beliefs are participating in the mechanism. Ranking behavior in Cambridge indicates that agents respond to the strategic incentives in the mechanism. Specifically, students that reside on either side of the boundary where proximity priority changes have observably different ranking behavior. We take this as evidence against the assumption that agents are ranking schools in order of true preferences. We then implement our method using the proposed estimator. Our results indicate that treating preferences as truthful is likely to result in biased estimates in markets where students face stiff competition for their preferred schools. The stated preferences therefore exaggerate the fraction of students assigned to their true top choice. We also illustrate how our method can be used to evaluate changes in the design of the market. Specifically, our baseline model finds that the typical student prefers the Cambridge mechanism's assignment to the Deferred Acceptance mechanism's assignment by an equivalent of 0.07 miles. These losses are concentrated for the paid-lunch students, who for whom the scarcity of seats at desirable programs results in the highest advantage from screening based on intensity of preferences. Free-lunch students, on the other hand, face a less complex strategic environment in the Cambridge mechanism and the average student is close to indifferent between the two mechanisms. Estimates from models in which agents have biased beliefs about assignment probabilities have a less optimistic view on the cardinal screening benefits of the Cambridge mechanism. A model with heterogeneously sophisticated agents finds that assignments under the Cambridge mechanism are preferable for paid-lunch naifs but not for free-lunch naifs. Across specifications, we find relatively few instances of justified envy in the Cambridge mechanism due to the significant majority of students that are assigned to their top choice in this school district. The most common instance of justified envy is estimated using the model with heterogeneous sophistication. We find about 18% of the roughly 30% naïve paid-lunch agents have justified envy. These differences across the two mechanisms should be weighed against potential costs of strategizing in a recommendation of mechanism choice. Quantifying these costs may be difficult without directly observing differences in information acquisition activites across mechanisms. More broadly, our results motivate further research on mechanisms that use the intensity of student preferences in allocation without some of the potential costs of strategic behavior (see Azevedo and Budish, 2013, for example).

Our methods can be extended in several directions. In the context studied here, schools are passive players who express their preferences with only coarse priorities and a random tie-breaker. Extending the techniques to allow for exam scores and finely defined priority groups will broaden the applicability of the results, but may require technical innovations for estimating the assignment probabilities. Another important extension is to consider a college admissions setting where students make application decisions while in consideration of chance of admission. A challenge in directly extending our approach is that we observe all priorities relevant for admissions in the data. In the college applications context, admission may depend on unobservables that also affect preferences, complicating the analysis. A closely related context is a multi-unit assignment mechanism such as course allocation mechanisms. The preferences in this context would need to be richer in order to allow for complementarities over the objects in a bundle that are assigned to an individual. These extensions are interesting avenues for expanding our ability to analyze agent behavior in assignment mechanisms.

References

Abdulkadiroglu, A. (2013). School Choice. In Vulkan, N., Roth, A. E., and Neeman, Z., editors, *The Handbook of Market Design*, chapter 5, pages 138–169. Oxford University Press, Cambridge, UK.

Abdulkadiroglu, A., Agarwal, N., and Pathak, P. A. (2015a). The Welfare Effects of Coordinated School Assignment: Evidence from the NYC High School Match. *NBER Working Paper*, No. 21046.

Abdulkadiroglu, A., Angrist, J., Narita, Y., and Pathak, P. A. (2015b). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation.

Abdulkadiroglu, A., Che, Y.-K., and Yasuda, Y. (2011). Resolving Conflicting Preferences in School Choice: The Boston Mechanism Reconsidered. *American Economic Review*, 101(1):399–410.

Abdulkadiroglu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-Proofness Versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match.

Abdulkadiroglu, A., Pathak, P. A., Roth, A. E., and Sonmez, T. (2006). Changing the Boston School Choice Mechanism. *NBER Working Papers 11963*.

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

Agarwal, N. (2015). An Empirical Model of the Medical Match. *American Economic Review*, 105(7):1939–78.

Agarwal, N. and Diamond, W. (2014). Identification and Estimation in Two-sided Matching Markets. *Cowles Foundation Discussion Paper No. 1905.*

Aguirregabiria, V. and Mira, P. (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica*, 75(1):1–53.

Ayaji, K. (2013). School Choice and Educational Mobility: Lessons from Secondary School Applications in Ghana.

Azevedo, E. M. and Budish, E. (2013). Strategyproofness in the large. *Chicago Booth, mimeo.*

Azevedo, E. M. and Leshno, J. (2013). A Supply and Demand Framework for Two-Sided Matching Markets.

Bajari, P., Benkard, C. L., and Levin, J. (2007). Estimating Dynamic Models of Imperfect Competition. *Econometrica*, 75(5):1331–1370.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, 115(4):588–638.

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

Berry, S. T. and Haile, P. A. (2010). Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, 114(2):577–599.

Block, H. and Marshak, J. (1960). Random Orderings and Stochastic Theories of Responses. In Olkin, I., Ghurye, S., Hoeffding, W., Mado, W., and Mann, H., editors, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis : The Kernel Approach with S-Plus Illustrations: The Kernel Approach with S-Plus Illustrations*. OUP Oxford.

Calsamiglia, C., Fu, C., and Guell, M. (2016). Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives.

Calsamiglia, C. and Guell, M. (2014). The Illusion of School Choice: Empirical Evidence from Barcelona.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained School Choice: An Experimental Study. *American Economic Review*, 100(4):1860–74.

Cardon, J. H. and Hendel, I. (2001). Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey. *2RAND Journal of Economics*, 32(3):408–427.

Cassola, N., Hortaçsu, A., and Kastl, J. (2013). The 2007 Subprime Market Crisis Through the Lens of European Central Bank Auctions for Short Term Funds. *Econometrica*, 81(4):1309–1345.

Chade, H. and Smith, L. (2006). Simultaneous Search. *Econometrica*, 74(5):1293–1307.

Chen, Y. and Kesten, O. (2013). From Boston to Chinese parallel to deferred acceptance: Theory and experiments on a family of school choice mechanisms. *Discussion Papers, Research Unit: Market Behavior*.

Chen, Y. and Sonmez, T. (2006). School choice: an experimental study. *Journal of Economic Theory*, 127(1):202–231.

Chiappori, P.-A., Salanie, B., Salanie, F., and Gandhi, A. (2012). From Aggregate Betting Data to Individual Risk Preferences.

Chiappori, P.-A., Salanié, B., and Weiss, Y. (2015). Partner Choice and the Marital College Premium. *CEPR Discussion Paper No. DP10403*.

Choo, E. and Siow, A. (2006). Who Marries Whom and Why. *Journal of Political Economy*, 114(1):175–201.

Cohen, A. and Einav, L. (2007). Estimating Risk Preferences from Deductible Choice. *American Economic Review*, 97(3):745–788.

Ergin, H. and Sonmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics2*, 90(1-2):215–237.

Featherstone, C. and Niederle, M. (2011). School choice mechanisms under Incomplete Information: an Experimental Investigation.

Fox, J. T. (2010a). Estimating Matching Games with Transfers. *University of Michigan, mimeo.*

Fox, J. T. (2010b). Identification in matching games. *Quantitative Economics*, 1(2):203–254.

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

Glazerman, S. and Meyer, R. H. (1994). Public school choice in Minneapolis, midwest approaches to school reform. *Proceedings of a Conference Held at the Federal Reserve Bank of Chicago.*

Glenn, C. L. (1991). Controlled choice in Massachusetts public schools. *Public Interest*, 103:88–105.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal Nonparametric Estimation of First-Price Auctions. *Econometrica*, 68(3):525–574.

Hastings, J. S., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous Preferences and the Efficacy of Public School Choice.

He, Y. (2014). Gaming the Boston School Choice Mechanism in Beijing. *Toulouse School of Economics, mimeo.*

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010). Matching and Sorting in Online Dating. *American Economic Review*, 100(1):130–163.

Hotz, V. J. and Miller, R. A. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *Review of Economic Studies*, 60(3):497–529.

Hotz, V. J., Miller, R. A., Sanders, S., and Smith, J. (1994). A Simulation Estimator for Dynamic Models of Discrete Choice. *Review of Economic Studies*, 61(2):265–89.

Hwang, S. I. M. (2016). A Robust Redesign of High School Match.

Imbens, G. and Kalyanaraman, K. (2011). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies*, 79(3):933–959.

Kapor, A., Neilson, C., and Zimmerman, S. (2016). Heterogeneous Beliefs and School Choice.

Kojima, F. and Pathak, P. A. (2009). Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review*, 99(3):608 – 27.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1):145–177.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, 8(3):229–254.

Matzkin, R. L. (1992). Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models. *Econometrica*, 60(2):239–70.

Matzkin, R. L. (1993). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics*, 58(1-2):137–168.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. pages 105 – 142.

Miralles, A. (2009). School Choice: The Case for the Boston Mechanism.

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

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Nielson, C. (2013). Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students.

Pakes, A., Ostrovsky, M., and Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics*, 38(2):373–399.

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

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

Pennell, H., West, A., and Hind, A. (2006). Secondary School Admissions in London.

Roth, A. E. and Peranson, E. (1999). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89(4):748–780.

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

Troyan, P. (2012). Comparing school choice mechanisms by interim and ex-ante welfare. *Games and Economic Behavior*, 75(2):936–947.

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

Table 1: School Choice Mechanisms

Mechanism	Manipulable	Examples
Boston Mechanism	Y	Barcelona[1], Beijing[2], Boston (pre 2005), Charlotte-Mecklenberg[3], Chicago (pre 2009), Denver, Miami-Dade, Minneapolis, Seattle (pre 1999 and post 2009), Tampa-St. Petersburg.
Deferred Acceptance		
w/ Truncated Lists	Y	New York City[4], Ghanian Schools, various districts in England (since mid '00s)
w/ Unrestricted Lists	N	Boston (post 2005), Seattle (1999-2008)
Serial Dictatorships		
w/ Truncated Lists	Y	Chicago (2009 onwards)
First Preferences First	Y	various districts in England (before mid '00s)
Chinese Parallel	Y	Shanghai and several other Chinese provinces[5]
Cambridge	Y	Cambridge[6]
Pan London Admissions	Y	London[7]
Top Trading Cycles		
w/ Truncated Lists	Y	New Orleans[8]

Notes: Source Table 1, Pathak and Sonmez (2008) unless otherwise stated. See several references therein for details. Other sources: [1] Calsamiglia and Guell (2014); [2]He (2014); [3]Hastings et al. (2009); [4]Abdulkadiroglu et al. (2009); [5]Chen and Kesten (2013); [6] "Controlled Choice Plan" CPS, December 18, 2001; [7]Pennell et al. (2006);
[8]`http://www.nola.com/education/index.ssf/2012/05/new_orleans_schools_say_new_pu.html` accessed May 20, 2014.

Table 2: Cambridge Elementary Schools and Students

Year	2004	2005	2006	2007	2008	Average
			Panel A: District Characteristics			
Schools	13	13	13	13	13	13
Programs	24	25	25	27	27	25.6
Seats	473	456	476	508	438	470
Students	412	432	397	457	431	426
Free/Reduced Lunch	32%	38%	37%	29%	32%	34%
Paid Lunch	68%	62%	63%	71%	68%	66%
			Panel B: Student's Ethnicity			
White	47%	47%	45%	49%	49%	47%
Black	27%	22%	24%	22%	23%	24%
Asian	17%	18%	15%	13%	18%	16%
Hispanic	9%	11%	10%	9%	9%	10%
			Panel C: Language spoken at home			
English	72%	73%	73%	78%	81%	76%
Spanish	3%	4%	4%	4%	3%	3%
Portuguese	0%	1%	1%	1%	1%	1%
			Panel D: Distances(miles)			
Closest School	0.43	0.67	0.43	0.47	0.45	0.49
Average School	1.91	1.93	1.93	1.93	1.89	1.92

Notes: Students participating in the January Kindergarten Lottery. Free/Reduced lunch based on student's application for Federal lunch subsidy.

Table 3: Cambridge Elementary Schools and Students

Year	2004	2005	2006	2007	2008	Average
	Panel A: Round of assignment					
First	81%	84%	85%	83%	75%	82%
Second	8%	3%	4%	7%	5%	5%
Third	5%	2%	2%	2%	4%	3%
Unassigned	6%	11%	9%	8%	16%	10%
	Panel B: Round of assignment: Paid Lunch Students					
First	80%	77%	78%	79%	68%	76%
Second	5%	4%	5%	8%	5%	5%
Third	6%	3%	4%	2%	3%	4%
Unassigned	9%	16%	14%	11%	24%	15%
	Panel C: Round of assignment: Free Lunch Students					
First	85%	95%	98%	94%	89%	92%
Second	14%	1%	2%	4%	6%	5%
Third	2%	1%	0%	1%	4%	1%
Unassigned	0%	4%	0%	2%	1%	1%
	Panel D: Number of Programs Ranked					
One	2%	6%	9%	5%	12%	7%
Two	5%	6%	9%	7%	7%	7%
Three	93%	89%	82%	88%	81%	87%
	Panel E: Students with Priority at Ranked Schools					
Sibling Priority at 1st Choice	38%	34%	32%	24%	34%	32%
Sibling Priority at 2nd Choice	4%	3%	1%	2%	2%	2%
Sibling Priority at 3rd Choice	0%	2%	1%	1%	0%	1%
Proximity at 1st Choice	53%	52%	50%	51%	52%	51%
Proximity at 2nd Choice	42%	34%	37%	33%	37%	36%
Proximity at 3rd Choice	22%	24%	24%	25%	21%	23%
	Panel F: Mean Distance (miles)					
Ranked first	1.19	1.18	1.24	1.29	1.19	1.22
All ranked schools	1.37	1.41	1.38	1.40	1.34	1.38
Assigned School	1.10	1.01	1.07	1.12	0.92	1.04

Notes: Sibling and proximity priority as reported in the Cambridge Public School assignment files. Students with older siblings enrolled in CPS receive priority at their sibling's school. Students also receive proximity priority at their two closest schools. Percentages, where reported, are based on the total number of applicants each year.

Table 4: School Popularity and Competitiveness

School	Graham Parks	Haggerty	Baldwin	Morse	Amigos	Cambridgeport	King Open	Peabody	Tobin	Flet Mayn	Kenn Long	MLK	King Open Ola
Panel A: All Students													
Ranked First	60	56	53	47	37	34	33	31	25	18	16	12	5
Ranked Second	72	37	66	25	18	44	39	38	17	10	18	20	0
Ranked Third	56	33	46	31	19	44	37	32	20	15	16	15	0
Ranked Anywhere	192	120	166	102	75	113	114	105	64	48	54	51	6
Capacity	41	41	41	42	41	27	51	48	35	38	41	37	15
First Rejected	1-P	1-R	1-R	1-R	1-R	1-R	NR	NR	1-R	NR	NR	NR	NR
Panel B: Paid Lunch Students													
Ranked First	49	45	40	29	25	24	25	17	13	4	7	4	2
Ranked Second	60	28	56	14	12	29	23	27	10	3	6	6	0
Ranked Third	47	29	33	19	15	34	24	18	11	4	8	10	0
Ranked Anywhere	152	95	128	60	51	87	70	65	33	9	21	20	3
Capacity	29	27	27	29	41	18	36	34	29	35	34	27	15
First Rejected	1-P	1-R	1-R	1-R	1-R	1-R	NR	NR	3-R	NR	NR	NR	NR
Panel C: Free Lunch Students													
Ranked First	9	12	12	17	12	11	13	10	12	16	10	9	2
Ranked Second	13	8	7	11	5	12	17	12	8	8	14	11	0
Ranked Third	10	4	9	10	4	12	13	13	9	10	11	4	0
Ranked Anywhere	29	24	25	40	20	36	44	38	31	36	34	25	2
Capacity	25	23	26	26	41	17	33	31	19	18	26	24	15
First Rejected	NR	NR	NR	1-R	1-R	2-P	NR	NR	1-R	NR	NR	NR	NR

Notes: Median number of applicants and seats over the years 2004-2008. First rejected is the round and priority of the first rejected student, e.g., 1-P indicates that a student with proximity priority was rejected in the first round. S: Sibling priority, PS: both proximity and sibling priority, R: regular/no prioirity, and NR: no student was rejected in any round. Free/Reduced lunch based on student's application for Federal lunch subsidy.

Table 5: Regression Discontinuity Estimates

	Baseline			Competitive School	Non-Competitive School	Placebo Boundary
	Rank First	Rank Second	Rank Third	Rank First	Rank First	Rank First
			Panel A: All Students			
Estimate	-5.75%	-2.38%	-0.86%	-7.27%	-2.06%	0.07%
	(0.013)	(0.012)	(0.011)	(0.018)	(0.019)	(0.024)
t-statistic	-4.54	-2.02	-0.80	-3.96	-1.10	0.03
			Panel B: Paid Lunch Students			
Estimate	-7.44%	-2.65%	-0.68%	-11.07%	-1.22%	1.88%
	(0.016)	(0.014)	(0.015)	(0.025)	(0.018)	(0.031)
t-statistic	-4.64	-1.90	-0.46	-4.45	-0.67	0.61
			Panel C: Free Lunch Students			
Estimate	-3.55%	-2.59%	-3.15%	-1.47%	-5.23%	-3.55%
	(0.022)	(0.021)	(0.022)	(0.031)	(0.031)	(0.033)
t-statistic	-1.60	-1.22	-1.43	-0.47	-1.67	-1.06

Notes: Regression discontinuity estimates based bandwidth selection rule proposed by Imbens and Kalyaraman (2011). All estimates use rankings by 2,128 students. Competitive schools are Graham & Parks, Haggerty, Baldwin, Morse, Amigos, Cambridgeport and Tobin. Placebo boundary at the mid-point of the two-closest schools. Standard errors clustered at the student level in parenthesis.

Table 6: Estimated Assignment Probabilities

Graham Parks	Haggerty	Baldwin	Morse	Amigos	Amigos Bilingual	Cambridgeport	King Open	Peabody	Tobin	Tobin K4	Tobin K5	Fletcher Maynard	Kenn Long
Panel A: All Students													
0.43	0.59	0.63	0.57	0.73	0.98	0.60	1.00	0.94	0.85	0.31	0.34	0.92	1.00
0.24	0.25	0.23	0.20	0.35	0.94	0.18	0.92	0.83	0.74	0.04	0.14	0.86	1.00
0.21	0.19	0.18	0.10	0.25	0.83	0.10	0.67	0.61	0.66	0.02	0.08	0.77	0.90
Panel B: Paid Lunch													
0.22	0.45	0.49	0.54	0.73	1.00	0.51	1.00	0.94	0.93	0.32	0.36	1.00	1.00
0.00	0.05	0.03	0.16	0.35	1.00	0.08	0.89	0.82	0.76	0.03	0.16	1.00	1.00
0.00	0.01	0.00	0.06	0.24	0.85	0.01	0.56	0.56	0.64	0.01	0.09	0.89	0.89
Panel C: Free/Reduced Lunch													
0.82	0.87	0.90	0.64	0.74	0.97	0.77	1.00	0.94	0.72	0.31	0.29	0.76	1.00
0.71	0.65	0.61	0.26	0.35	0.90	0.39	0.98	0.86	0.71	0.07	0.08	0.59	0.99
0.62	0.56	0.52	0.18	0.27	0.83	0.28	0.87	0.70	0.68	0.03	0.05	0.53	0.92
Panel D: Proximity Priority													
0.63	0.97	0.95	0.89	0.95	0.99	0.94	1.00	1.00	0.92	0.55	0.54	0.95	1.00
0.12	0.21	0.15	0.24	0.38	0.97	0.30	0.97	0.84	0.76	0.07	0.16	0.80	1.00
0.10	0.11	0.06	0.12	0.28	0.84	0.16	0.76	0.66	0.67	0.02	0.12	0.72	0.89
Panel E: No Priority													
0.37	0.55	0.60	0.52	0.71	0.97	0.55	1.00	0.94	0.85	0.28	0.33	0.92	1.00
0.28	0.25	0.23	0.19	0.34	0.96	0.17	0.92	0.83	0.75	0.03	0.14	0.88	1.00
0.25	0.20	0.18	0.10	0.24	0.85	0.10	0.66	0.60	0.66	0.02	0.08	0.78	0.90

verage estimates weighted by number of students of each type. Probabilities estimated using B=1,0
ority types of opposing students are drawn with replacement from the observed data. Second and t
nent probabilities are conditional on no assignment to higher ranked choices, averaged across feasil
sts.

Table 7: Estimated Mean Utilities

	Truthful		Rational Expectations		Adaptive Expectations		Coarse Beliefs	
	Paid Lunch	Free Lunch	Paid Lunch	Free Lunch	Paid Lunch	Free Lunch	Paid Lunch	Free Lunch
Panel A: Mean Utility								
Graham Parks	1.29	0.40	1.93	-0.20	1.89	-0.19	1.85	0.47
	[0.06]	[0.08]	[0.27]	[0.33]	[0.36]	[0.52]	[0.12]	[0.16]
Haggerty	1.39	0.72	1.45	0.69	1.41	0.89	1.56	0.68
	[0.07]	[0.11]	[0.16]	[0.24]	[0.20]	[0.31]	[0.13]	[0.18]
Baldwin	1.26	0.50	1.36	0.70	1.05	-0.14	1.57	0.72
	[0.05]	[0.09]	[0.14]	[0.16]	[0.17]	[0.37]	[0.10]	[0.15]
Morse	0.66	0.70	0.77	0.97	0.61	0.78	0.82	0.8
	[0.07]	[0.08]	[0.12]	[0.17]	[0.18]	[0.33]	[0.14]	[0.16]
Amigos	-0.01	-0.38	0.13	-0.29	0.05	-0.43	-0.11	-0.64
	[0.13]	[0.15]	[0.18]	[0.27]	[0.22]	[0.32]	[0.19]	[0.27]
Cambridgeport	0.77	0.18	0.60	0.29	0.41	0.25	0.91	0.21
	[0.06]	[0.08]	[0.14]	[0.18]	[0.21]	[0.26]	[0.12]	[0.16]
King Open	0.65	0.40	0.58	0.27	0.44	0.37	0.52	0.24
	[0.06]	[0.07]	[0.10]	[0.13]	[0.13]	[0.18]	[0.10]	[0.12]
Peabody	0.22	0.48	0.06	0.30	0.03	0.52	0.05	0.31
	[0.08]	[0.09]	[0.12]	[0.15]	[0.16]	[0.20]	[0.11]	[0.15]
Tobin	-0.49	0.64	-0.92	0.38	-0.83	0.39	-0.74	0.28
	[0.11]	[0.12]	[0.21]	[0.26]	[0.25]	[0.30]	[0.18]	[0.21]
Fletcher Maynard	-1.30	-0.05	-2.03	-0.14	-1.67	0.12	-2.21	-0.3
	[0.14]	[0.10]	[0.42]	[0.23]	[0.32]	[0.22]	[0.30]	[0.18]
Kenn Long	-0.19	0.47	-0.40	0.11	-0.54	0.08	-0.51	0.25
	[0.09]	[0.07]	[0.20]	[0.17]	[0.27]	[0.21]	[0.18]	[0.15]
MLK	-0.66	0.08	-1.14	-0.28	-0.82	0.16	-1.24	-0.27
	[0.10]	[0.09]	[0.23]	[0.19]	[0.22]	[0.22]	[0.18]	[0.17]
King Open Ola	-3.60	-4.13	-2.39	-2.79	-2.01	-2.79	-2.47	-2.75
	[0.35]	[0.39]	[0.53]	[0.67]	[0.62]	[0.93]	[0.47]	[0.63]
Outside Option	-2.08	-1.44	-0.55	-0.92	-0.64	-0.74	-0.49	-0.73
	[0.10]	[0.09]	[0.08]	[0.08]	[0.09]	[0.13]	[0.06]	[0.07]
Panel B: Percentage of Acceptable Schools								
up to 1	10%	20%	24%	14%	23%	19%	16%	21%
up to 2	16%	30%	61%	40%	61%	45%	58%	51%
up to 3	23%	40%	85%	62%	86%	69%	84%	74%
up to 4	34%	51%	95%	79%	95%	85%	95%	89%
up to 5	44%	61%	98%	91%	99%	95%	99%	96%

Notes: Average estimated utility for each school, normalizing the mean utility of the inside options to zero. Utilities calculated by averaging the predicted utility given the non-distance covariates. Bootstrap standard errors in brackets, except for Truthful reporting where we present the standard deviation of the posterior. Adaptive Expectations based on reported lists from 2005, 2006 and 2008 with assignment probabilites estimated using data from 2004, 2005 and 2007 respectively. This specification drops data from 2007 in preference estimates since Tobin split by entering age in that year. The fraction of students with rationalizable lists is 97.3%, 96.8%, and 99.3% for the Rational Expectations, Adaptive Expectations and Coarse Beliefs specifications respectively.

Table 8: Losses from Truthful Reports

	Truthful						Rational Expectations					
	No Loss		Mean Loss		Std Loss		No Loss		Mean Loss		Std Loss	
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
All	57%	0.01	0.18	0.02	0.53	0.05	46%	0.01	0.07	0.01	0.26	0.03
Free Lunch	68%	0.02	0.01	0.00	0.09	0.03	62%	0.02	0.01	0.00	0.07	0.03
Paid Lunch	51%	0.01	0.26	0.03	0.64	0.06	38%	0.01	0.10	0.01	0.31	0.04
Black	65%	0.02	0.06	0.02	0.30	0.07	56%	0.02	0.04	0.01	0.19	0.05
Asian	56%	0.03	0.19	0.04	0.56	0.09	46%	0.03	0.07	0.02	0.25	0.06
Hispanic	60%	0.04	0.10	0.03	0.36	0.09	51%	0.03	0.04	0.01	0.18	0.06
White	52%	0.01	0.24	0.03	0.62	0.06	40%	0.02	0.09	0.01	0.30	0.04
Other Race	47%	0.06	0.20	0.07	0.51	0.15	39%	0.05	0.08	0.04	0.24	0.09

Notes: Estimated loss from reporting preferences truthfully, relative to optimal report in distance units (miles).

Table 9: Ranking and Assignment of Top Choice

	Graham Parks	Haggerty	Baldwin	Morse	Amigos	Cambridgeport	King Open	Peabody	Tobin	Fletcher Maynard
Panel A: All Students										
ed School	24.8	11.6	9.4	10.0	7.9	6.1	7.2	6.1	5.4	4.0
#1 (simul)	16.2	12.6	11.6	11.2	8.4	8.0	8.6	7.0	4.9	3.9
#1 (data)	14.3	12.6	11.9	11.0	8.8	7.7	8.2	7.8	5.7	4.4
ed and Ranked #1	15.3	10.4	8.2	9.6	7.5	5.8	7.2	6.1	4.6	3.9
ed and Assigned	9.9	8.5	6.6	8.2	6.8	5.1	7.2	6.0	3.8	3.5
#1 and Assigned	10.4	10.1	9.0	9.5	7.6	6.8	8.6	6.9	4.2	3.6
Panel B: Free Lunch Students										
ed School	9.0	8.1	6.5	12.5	6.9	6.8	7.2	7.9	10.7	10.8
#1 (simul)	8.9	8.5	6.8	12.7	6.9	7.1	7.2	8.2	9.0	10.6
#1 (data)	6.7	8.3	8.0	12.2	7.8	6.4	7.7	9.0	8.5	10.8
ed and Ranked #1	8.6	8.0	6.4	12.2	6.7	6.6	7.2	7.8	8.9	10.3
ed and Assigned	7.9	7.5	6.0	10.7	6.0	6.1	7.2	7.8	7.1	9.3
#1 and Assigned	8.3	7.9	6.4	11.1	6.2	6.5	7.2	8.1	7.2	9.5
Panel C: Paid Lunch Students										
ed School	32.6	13.3	10.9	8.7	8.6	5.7	7.3	5.3	2.7	0.6
#1 (simul)	20.0	14.6	13.9	10.4	9.3	8.3	9.3	6.5	2.9	0.6
#1 (data)	18.8	14.6	13.4	10.6	9.5	7.9	8.7	7.2	3.7	1.2
ed and Ranked #1	18.9	11.5	9.2	8.2	8.1	5.3	7.3	5.3	2.4	0.6
ed and Assigned	11.0	8.9	7.0	6.9	7.3	4.6	7.3	5.2	2.2	0.6
#1 and Assigned	11.7	11.1	10.2	8.6	8.3	6.9	9.3	6.4	2.7	0.6

Unless otherwise noted, table presents averages over 1,000 simulations from the posterior mean of
:ed from the rational expectations model.

Table 10: Deferred Acceptance vs Cambridge

	Truthful			Rational Expectations			Coarse Beliefs			Adaptive Expectations		
	All Students	Paid Lunch	Free Lunch	All Students	Paid Lunch	Free Lunch	All Students	Paid Lunch	Free Lunch	All Students	Paid Lunch	Free Lunch
Panel A: Deferred Acceptance												
Assigned to First Choice	67.7	58.2	86.6	67.9	58.1	87.5	69.7	61.0	87.1	68.4	56.9	89.1
Assigned to Second Choice	12.1	14.2	8.1	15.5	18.6	9.4	11.9	13.7	8.5	13.6	17.3	7.1
Assigned to Third Choice	5.7	8.2	0.8	5.2	7.1	1.3	4.9	6.7	1.2	5.1	7.3	1.1
Assigned to Fourth Choice	3.5	5.3	0.1	1.4	2.0	0.2	1.9	2.7	0.2	1.6	2.3	0.1
Assigned to Fifth Choice	2.1	3.2	0.0	0.2	0.3	0.0	0.4	0.6	0.0	0.2	0.4	0.0
Panel B: Cambridge Mechanism												
Assigned to First Choice	79.0	74.5	87.8	72.3	63.9	88.8	73.9	67.3	86.9	72.3	63.0	88.9
Assigned to Second Choice	6.5	6.8	6.0	14.7	18.1	7.9	10.2	11.1	8.3	12.1	15.3	6.4
Assigned to Third Choice	3.1	4.0	1.4	3.9	5.1	1.3	3.5	4.6	1.5	3.7	4.9	1.4
Assigned to Fourth Choice	0.0	0.0	0.0	1.0	1.4	0.3	1.5	2.1	0.3	1.3	1.8	0.3
Assigned to Fifth Choice	0.0	0.0	0.0	0.2	0.2	0.0	0.4	0.5	0.0	0.3	0.4	0.1
Panel C: Deferred Acceptance vs Cambridge												
Mean Utility DA - Cambridge	-0.004	-0.010	0.008	-0.072	-0.109	0.003	-0.045	-0.074	0.013	-0.049	-0.097	0.037
	(0.017)	(0.025)	(0.006)	(0.011)	(0.015)	(0.013)	(0.011)	(0.013)	(0.016)	(0.028)	(0.035)	(0.040)
Std. Utility DA - Cambridge	0.230	0.280	0.047	0.171	0.142	0.197	0.174	0.146	0.207	0.213	0.142	0.282
Percent DA > Cambridge	26.8	26.0	28.3	16.5	14.2	21.1	22.6	21.3	25.1	19.1	16.5	23.9
Percent DA ≈ Cambridge	31.9	26.2	43.0	30.3	27.1	36.6	30.6	26.5	38.7	31.6	26.2	41.4
Percent DA < Cambridge	41.4	47.8	28.7	53.2	58.7	42.3	46.9	52.2	36.2	49.3	57.4	34.7
Percent with Justified Envy	9.93	12.69	4.46	5.6	5.1	6.4	7.1	7.8	5.6	6.7	8.0	4.4

Notes: Panels A and B present percentages of students assigned to true k-th choice. Panel C compares the expected utility difference between Deferred Acceptance and Cambridge Mechanism. Simulations of the Deferred Acceptance mechanism draw other student reports using the estimated utility distribution. We say DA ≈ Cambridge if the expected utility is within 10-5 miles. Point estimates use the estimated parameters and 1,000 simulations of the mechanisms. Bootstrap standard errors using 250 draws of the parameters in parentheses, except for Truthful reporting, where we use the posterior distribution.

Table 11: Estimated Mean Utilities using a Mixture Model

	Mixture Model		
	All Students	Paid Lunch	Free Lunch
	Panel A: Mean Utility		
Graham Parks	1.19	1.53	0.52
	[0.11]	[0.12]	[0.15]
Haggerty	1.27	1.53	0.76
	[0.14]	[0.13]	[0.22]
Baldwin	1.25	1.45	0.84
	[0.10]	[0.10]	[0.13]
Morse	0.74	0.68	0.86
	[0.11]	[0.11]	[0.13]
Amigos	-0.12	0.00	-0.38
	[0.21]	[0.19]	[0.30]
Cambridgeport	0.56	0.68	0.31
	[0.11]	[0.11]	[0.15]
King Open	0.48	0.57	0.32
	[0.09]	[0.10]	[0.12]
Peabody	0.08	0.03	0.16
	[0.13]	[0.13]	[0.16]
Tobin	-0.45	-0.81	0.26
	[0.16]	[0.18]	[0.23]
Flet Mayn	-1.02	-1.52	-0.04
	[0.24]	[0.30]	[0.18]
Kenn Long	-0.07	-0.23	0.24
	[0.14]	[0.16]	[0.15]
MLK	-0.68	-0.97	-0.10
	[0.14]	[0.17]	[0.15]
King Open Ola	-3.23	-2.96	-3.77
	[0.40]	[0.43]	[0.46]
Outside Option	-1.11	-1.03	-1.26
	[0.09]	[0.09]	[0.11]
	Panel B: Agent Behavior		
Fraction Naïve		0.378	0.316
		[0.0079]	[0.0079]

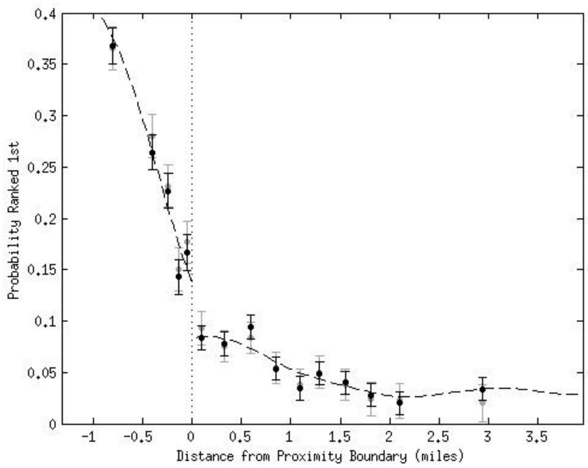
Notes: Panel A presents the average estimated utility for each school, normalizing the mean utility of the inside options to zero. Utilities calculated by averaging the predicted utility given the non-distance covariates. Panel B reports the estimated fraction of naive agents by free-lunch status. Bootstrap standard errors in brackets.

Table 12: Deferred Acceptance vs Cambridge using a Mixture Model

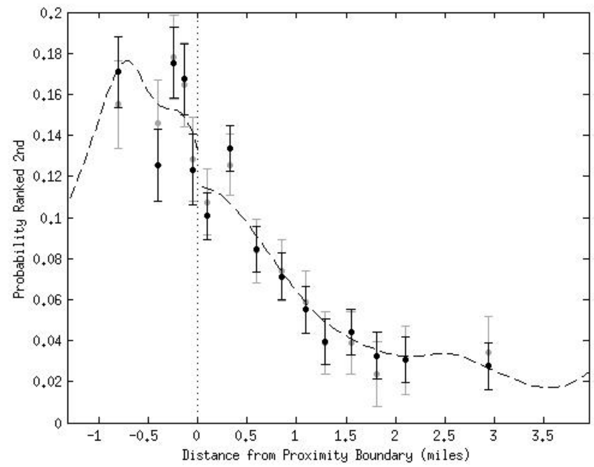
	All Students		Paid Lunch		Free Lunch	
	Panel A: Deferred Acceptance					
Assigned to First Choice	70.7		62.1		87.6	
Assigned to Second Choice	12.8		14.5		9.5	
Assigned to Third Choice	5.7		7.8		1.5	
Assigned to Fourth Choice	2.8		4.0		0.3	
Assigned to Fifth Choice	1.0		1.4		0.1	
	Panel B: Cambridge Mechanism					
	Naïve	Sophisticated	Naïve	Sophisticated	Naïve	Sophisticated
Percent of Students	35.7	64.3	37.8	62.2	31.6	68.4
Assigned to First Choice	78.4	76.2	72.4	69.5	90.2	89.6
Assigned to Second Choice	6.5	12.3	6.6	14.7	6.3	7.6
Assigned to Third Choice	3.3	4.3	4.1	5.9	1.7	1.3
Assigned to Fourth Choice	0.0	1.8	0.0	2.6	0.0	0.3
Assigned to Fifth Choice	0.0	0.5	0.0	0.8	0.0	0.1
	Panel C: Deferred Acceptance vs Cambridge					
	Naïve	Sophisticated	Naïve	Sophisticated	Naïve	Sophisticated
Mean Utility DA - Cambridge	-0.010	-0.081	-0.021	-0.122	0.010	0.000
	(0.012)	(0.013)	(0.016)	(0.016)	(0.013)	(0.013)
Std. Utility DA - Cambridge	0.162	0.172	0.144	0.149	0.192	0.184
Percent DA > Cambridge	23.2	14.1	23.8	10.7	22.1	20.9
Percent DA Cambridge	30.3	30.0	26.4	26.0	37.9	37.9
Percent DA < Cambridge	46.5	55.9	49.8	63.3	40.1	41.1
Percent with Justified Envy	15.4	2.7	18.1	2.6	9.2	3.0

Notes: Panels A and B present percentages of students assigned to true k-th choice. Panel C compares the expected utility difference between Deferred Acceptance and Cambridge Mechanism. Simulations of the Deferred Acceptance mechanism draw other student reports using the estimated utility distribution. We say DA ≈ Cambridge if the expected utility is within 10-5 miles. Bootstrap standard errors in brackets.

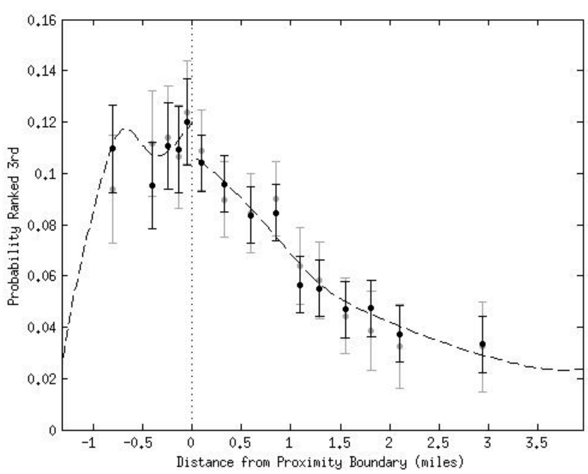
Figure 1: Effect of Proximity Priority on Ranking Behavior



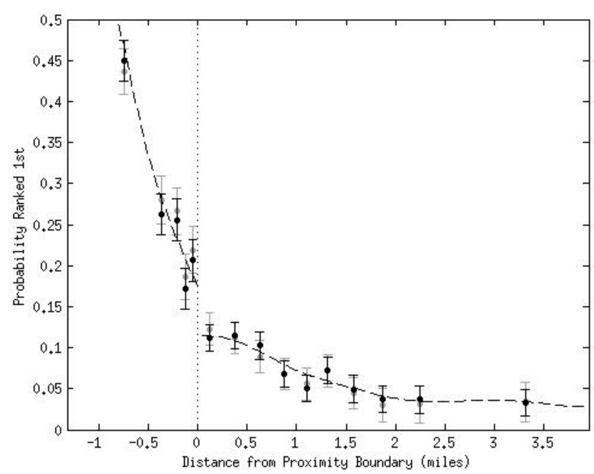
(a) First Rank



(b) Second Rank



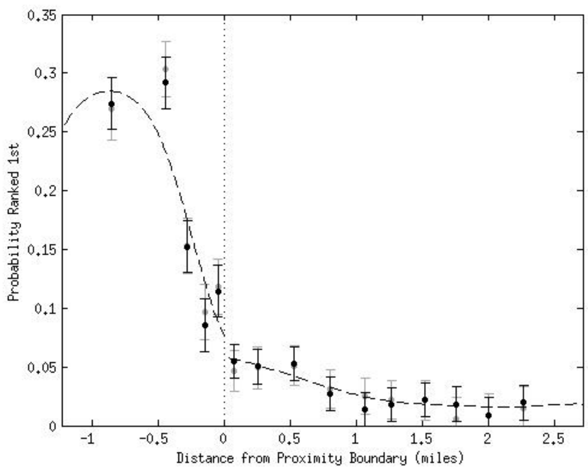
(c) Third Rank



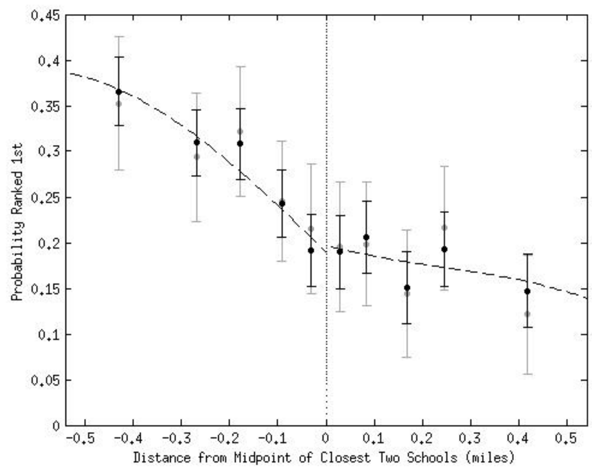
(d) Competitive schools

Cont'd...

Figure 1: Effect of Proximity Priority on Ranking Behavior (cont'd)



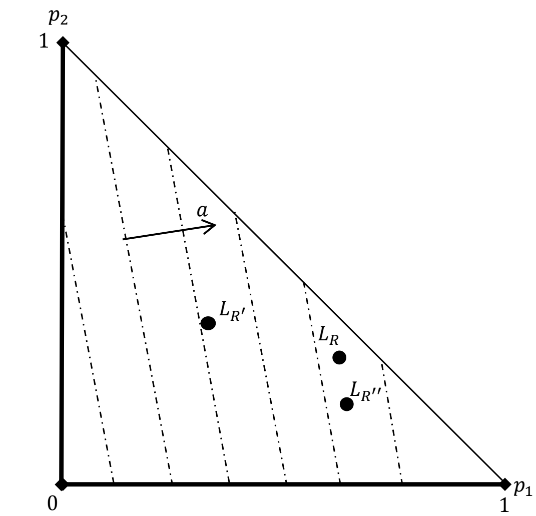
(e) Non-competitive schools



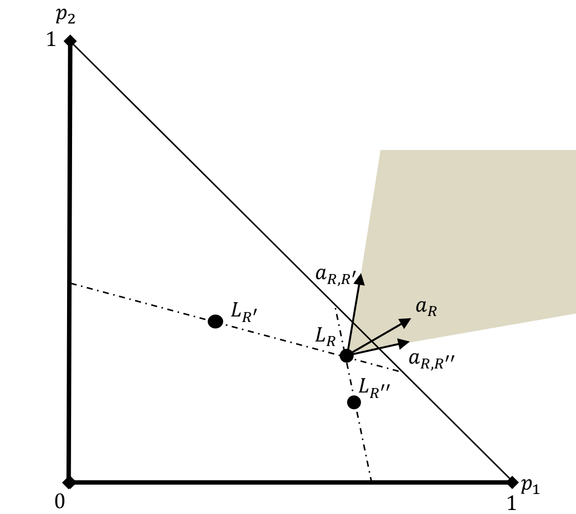
(f) Placebo at prioritized schools

Notes: The graphs are bin-scatter plots (based on distance) with equally sized bins on either side of the boundary. For each student, we construct a boundary distance, $\bar{d}_i$, based on her distance to the schooling options. For a given school-student pair, the horizontal axis represents $d_{ij} - \bar{d}_i$. The vertical axis is the probability that a student ranks the school in the relevant distance bin. Range plots depict 95% confidence intervals. Black plot points are based on the raw data, while the grey points control for school fixed effects. Dashed lines represent local linear fits estimated on either side of the boundary based on bandwidth selection rules recommended in Bowman and Azzalini (1997) (page 50). Panels (a) through (e) use the average distance between the second and third closest schools as the boundary. A student is given proximity priority at the schools to the left of the boundary and does not receive priority at schools to the right. Competitive schools considered in panel (d) are Graham & Parks, Haggerty, Baldwin, Morse, Amigos, Cambridgeport and Tobin. The remaining schools are considered non-competitive in panel (e). Panel (f) considers only the two closest schools and uses the average distance between the closest and second closest schools. Only schools where students have proximity priority are considered. Panels (a), (d), (e) and (f) plot the probability that a school is ranked first. Panels (b) and (c) plot the probability that a school is ranked second and third respectively. Distances as calculated using ArcGIS. Proximity priority recorded by Cambridge differs from these calculations in about 20% of the cases. Graphs are qualitatively similar when using only students with consistent calculated and recorded priorities. Details in data appendix.

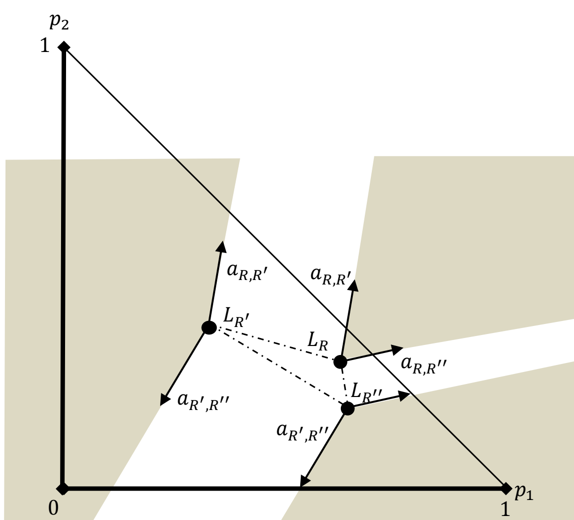
Figure 2: A Revealed Preference Argument



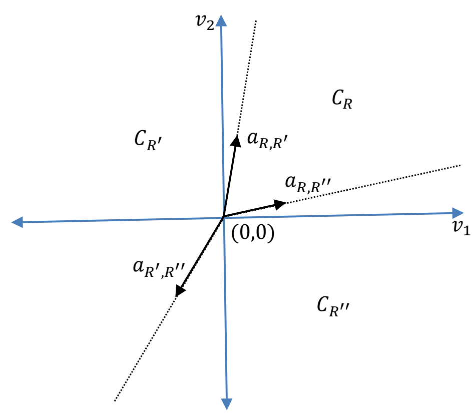
(a) Indifference curves for utility vectors parallel to a, and choices over three lotteries



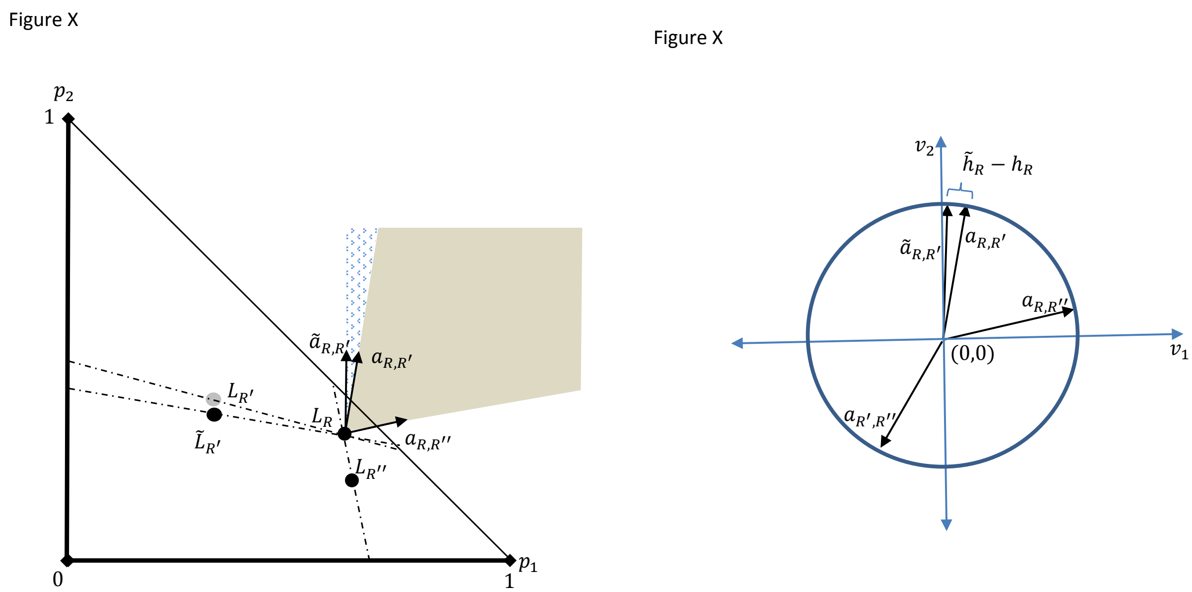
(b) L_R is optimal for a_R if $a_R = \alpha_1 a_{R,R'} + \alpha_2 a_{R',R''}$ for $\alpha_1, \alpha_2 > 0$



(c) Normal cones for each lottery



(d) Lottery choice reveals utility region



(a) Variation in the assignment probabilities induces variation in the normal cones

(b) Allows identifying the density of utility distribution on the unit circe

Figure 3: Variation in Lotteries

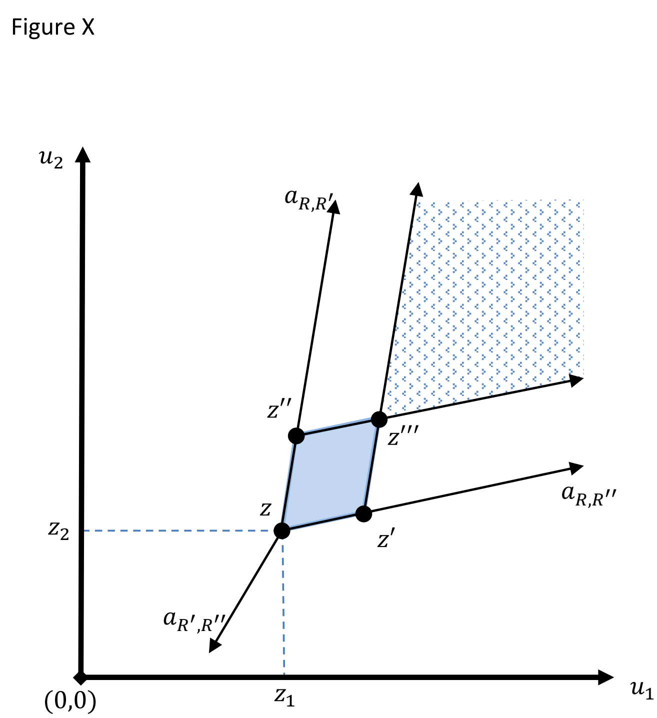


Figure 4: Local variation in z identifies the density of u