

Simulating Alternative School Choice Options in Boston - Technical Appendix*

Parag A. Pathak Peng Shi[†]



MIT School Effectiveness and Inequality Initiative[‡]

January 2013

Contents

1	Introduction	2
2	Data and Econometric Approach	2
3	Estimating the Demand Model	5
3.1	In-Sample Fit	7
3.2	Out-of-Sample Fit	9
4	Replicating BPS Assignment Algorithm and Computing Cutoffs	15
4.1	Describing the Algorithm	15
4.2	Estimating “Access” Using Bin Cutoffs	16
5	Length of Simulated Rank Order Lists	17
6	Validation of Simulation Approach	18
6.1	Access to Quality Ignoring Acceptability	19
6.2	Access to Top Menu Choice	19
6.3	Proximity	19
6.4	Diversity	19
6.5	Community	20

*Draft date: January 23, 2013. We thank Kamal Chavda, Dr. Carol Johnson, Carleton Jones, Timothy Nicolette, Peter Sloan, and Jack Yessayan for their expertise and for facilitating access to the data used in this study. Pathak acknowledges generous support from the National Science Foundation (grant SES-1056325).

[†]Pathak: Massachusetts Institute of Technology, Department of Economics, email: ppathak@mit.edu and Shi: Massachusetts Institute of Technology, Operations Research Center, email: pengshi@mit.edu.

[‡]<http://seii.mit.edu/>

1 Introduction

Boston Public Schools (BPS) is considering modifications to their school assignment plan for the 2013-2014 school year. We analyzed several proposals by simulating what would have happened in Round 1 of 2012-2013 if a new plan had been implemented that year, holding everything else fixed. The executive summary and main report is contained in Pathak and Shi (2013). The purpose of this technical appendix is to provide a detailed explanation of our methodology and assumptions.

There are three main steps to the analysis. First, we use data on choices submitted during the Round 1 in 2011-2012 to **estimate demand** for schools. These methods build on discrete-choice econometric models of demand applied to school choice, following among others Abulkadiroglu, Agarwal, and Pathak (2012) and Walters (2012). Next, we **validate** these preference estimates to ensure that they can account for aggregate patterns of choice within sample (using data from 2011-2012) and by comparing our forecasts of choices to those made out-of-sample (using data from 2012-2013). Finally, using the descriptions of the alternative choice plans provided by BPS together with our validated estimates of the preference distribution, we **simulate** how participants would rank schools under the alternative plans and how the assignment algorithm would take those choices together with school priorities and processing order to determine school offers. The approach only provides a snapshot of offers at the conclusion of Round 1 and does not factor in wait-lists or enrollment decisions.

2 Data and Econometric Approach

Boston Public Schools (BPS) assigns students using a mechanism based on the student-proposing deferred acceptance algorithm (see, e.g., Abdulkadiroğlu and Sönmez (2003), Abdulkadiroğlu, Pathak, Roth, and Sönmez (2005), and Pathak and Sönmez (2008)). Students interested in enrolling in or switching schools are asked to list schools each January. Students entering kindergarten can either apply for elementary school at Grade K1 or Grade K2 depending on whether they are four or five years old.

BPS provided us access to Round-1 choice data for 2011-2012 and 2012-2013 and a separate file containing demographic information such as race and free-lunch status of participants. It is our understanding that this demographic information is collected after students submit their choices and enroll in schools, and therefore coverage is not complete. If an applicant cannot be matched to the demographic file, we do not have information on race or free lunch status. Table 1 presents descriptive statistics on the 2012-2013 data set. There are a total of 6,696 applicants in our file, of which nearly 70% are new applicants who are not assigned to a present school and 52% also do not have a sibling at the school. There are more applicants at Grade K1 than Grade K2; however, many grade K2 applicants are continuing students who obtained a school assignment at Grade K1.

The student assignment mechanism prioritizes applicants in the following order: continuing students (students attending the same school at the earlier grade) have highest priority, followed by students who have a sibling at the school. Moreover, each school program is internally divided into

Type	Number of students	% of Sample
All	6696	100%
Present Schoolers	2271	34%
New Applicants (No Present School)	4625	69%
New Families (No Sibling, No Present School)	3468	52%
K1	2666	40%
K2	4030	60%
Black	1436	21%
White	1059	16%
Asian	445	7%
Hispanic	2892	43%
Other	196	3%
Missing	668	10%
Free Lunch	3251	49%
Reduced Lunch	236	4%
Non-Free Reduced Lunch	778	12%
Missing	2431	36%
LEP (Limited English Proficiency)	2675	40%

Table 1: Student dataset used for our analysis. (Non-substantially separate SPED K1, K2 applicants in 2012-2013 Round-1 choice.) Race information missing for 668 students and lunch information missing for 2431 students.

two halves, a “walk-half” and an “open-half.” For the walk-half, two students who either both have siblings or both don’t are prioritized based on who lives within 1 mile radius of the school. Students who live within 1-mile straight line distance of school are said to live in the school’s “walk-zone.” For the open-half, walk-zone priority plays no role. Each student receives a single, independently and identically distributed random number, and the same number is used for all programs. Students in the same priority group are ordered based on their random number.

The above defines for every program “half” a relative ranking over all students. After students submit rankings of programs, the assignment algorithm converts this to a ranking over program “halves” by having students in the walk-zone applying to the “walk-half” first, and students outside of the walk-zone applying to the “open-half” first. The assignment from students to program “halves” is then computed using the student-proposing deferred acceptance algorithm. Since the mechanism is based on student-proposing deferred acceptance and there is no restriction on the number of choices that can be listed, the assignment mechanism is strategy-proof. This means that truth-telling is a weakly dominant strategy.

The district also informs families of this property on the application form where BPS (School Guide 2012) advises families to

list your school choice in your true order of preference. If you list a popular school first, you won’t hurt your chances of getting your second choice school if you don’t get your first choice.

Since the mechanism is strategy-proof, we take the submitted preferences of families as representing their true preferences.

In the choice data, we observe a proxy for the student’s home address upon applying (called a “geocode”; Boston is partitioned into 868 geocodes). We observe the student’s race, socioeconomic status (whether the student receives free lunch, reduced lunch or neither) and English proficiency status. We also observe the student’s top 10 choices, random number, as well as where the student is assigned (if assigned at all). For each choice, we can infer from the priority whether the student has a sibling at that school or is a continuing student. For continuing students, if they do not choose their current program, the BPS mechanism appends their current program to their choice list, and assigns them a special priority known as “FinalPass.” For these students, we follow the BPS mechanism and append their current program to their choice list (capturing the apparent convention that if the students are not offered any other program they will choose to remain in their current program).

We begin by describing the tradeoffs expressed in student choices using a random utility model. Let i index students, j index school programs. j can be represented as ordered pair (s, p) , where s_j is the school and p_j is the program type. (*i.e.* for j =Quincy KED, s_j =Quincy and p_j =KED.) The following are data:

- Sibling $_{is}$: Indicator for i having sibling at s
- Walk $_{is}$: Indicator for s being in i ’s walkzone
- Present $_{ij}$: Indicator for j being i ’s current program
- d_{is} : Estimated Walk distance from i to s
- z_i^l : Student characteristic, indexed by l
- x_s^l : School characteristic, indexed by l .

The following parameters are estimated from choice data:

- δ_s : Fixed effect for school s
- δ_p : Fixed effect for program type p
- α, β, γ : Taste parameters.

We model student i ’s indirect utility for program j using the following specification:

$$\begin{aligned}
 u_{ij} &= \delta_j + \theta_{ij} + \kappa_{ij} + \varepsilon_{ij}, \text{ where} & (1) \\
 \delta_j &= \delta_{s_j} + \delta_{p_j} \\
 \theta_{ij} &= \gamma \cdot \vec{f}(d_{is_j}) + \beta_1 \text{Sibling}_{is_j} + \beta_2 \text{Walk}_{is_j} + \beta_3 \text{Present}_{ij} \\
 \kappa_{ij} &= \sum_l \alpha^l z_i^l x_{s_j}^l.
 \end{aligned}$$

The first set of variables represent school and program fixed effects. Each represents a single dimensional variable which captures what makes a school or program popular in general. The fixed effects represent the net overall effect of a school or program on the ranking decision and form a composite of attributes which do not to vary with student but only across schools (such as safety, facilities, and principal/teacher characteristics). We assume that tastes for schools and programs can be written in an additively separable form rather than school-program type interactions for statistical power. The second set of variables represent a student’s special affinities for a program, based on a function (or functions) of the walk-distance, whether the student has a sibling at the

school, whether the school is in the walk-zone, and whether the student is already enrolled in the same school program. The third set of variables are interaction terms between the student’s observed characteristics and the school’s observed characteristics. This formulation allows for differing tastes for particular school characteristics across demographic or socioeconomic groups.

Finally, ε_{ij} is an error term drawn from an extreme value type-I (Gumbel) distribution with variance normalized without loss of generality to $\frac{\pi^2}{6}$. Hence, our model is a multinomial logit discrete choice model or rank-ordered logit. We estimate this model by maximum likelihoods and report conventional standard errors.

The logit functional form allows us to write the likelihood in closed form, sidestepping the need for numerical integration. The logit assumption also comes with the Independence of Irrelevant Alternatives (IIA) property, which implies that if the choice set changes, the relative likelihood of ranking any two schools does not change.

It is worth noting that this model does not include an outside option in the choice set. The reason is that the majority of applicants rank 5 or fewer school programs. If we interpret the outside option as more preferred than any unlisted program, then most students prefer their outside option above most schools. This implication would be misleading since the majority of students who submit incomplete lists enroll in Boston Public Schools the following school year implying that they do not prefer their outside option. Moreover, families may only rank a few programs also because of limited information or confidence that they will get in to where they ranked, weighing against the assumption that unranked programs are less desirable than the outside options.

To forecast the outcomes under new plans with our preference estimates, it is important to have a way to simulate how many programs a typical applicant ranks. If an applicant submits a ranking over all possible schools in their menu, then the amount of competition will be vastly higher than in reality. Moreover, to estimate “access to quality” according to the definition in our main report (Pathak and Shi 2013), it is important to have some notion of individual rationality, or what programs are “acceptable” to each student. For this report, we make the behavioral assumption that families find their top 10 options in choice menu “acceptable”, while any programs outside of the top 10 are not listed. This behavioral assumption is convenient because it allows us to truncate every choice list exactly to 10 without worrying about strategic concerns. However, it also implies that what families find acceptable is directly influenced by what menus they are offered. We discuss this assumption in more detail in Section 5.

3 Estimating the Demand Model

In our model specified in equation (1), we have the following choices:

- What years of data, grades, and students to focus on?
- How do students evaluate distance to closest school, i.e., what distance transformation functions $\vec{f}(\cdot)$ to use?
- What student and school characteristics z_i^l and x_s^l influence the ranking decision?

Because we wish to incorporate models with substantial taste heterogeneity to predict the choices of all non-substantially separate special needs students in 2012-2013 Round-1 data set, we use the

largest possible sample of students. Moreover, because Grade K1 programs naturally transition into Grade K2 programs, and in conversations with parents it seems that many are making a conscious choice between waiting until Grade K2 and applying to a certain program or applying to Grade K1 program a year earlier, we assume that Grade K1 and K2 programs act as substitutes and therefore pool together choices at Grade K1 and K2 to increase precision of our estimates of school and program type fixed effects.

However, because school information and statistics (and even location) change from year to year, without a fuller analysis of the precise changes that occurred, we decided it conservative to fit the model one year at a time. Therefore, we estimate school and program type fixed effects for each years independently.

Our methodology for specification selection is as follows. We use 2011-2012 choice data as training set (in-sample), and 2012-2013 choice data as evaluation set (out-of-sample). For our actual simulations, we use the model fit using 2012-2013 choice data.

We report estimates from four specifications:

- 1) (SIMPLE): $\vec{f}(d) = (d)$; no cross interactions.
- 2) (SIMPLE2): $\vec{f}(d) = (d, \sqrt{d})$; no cross interactions.
- 3) (MEDIUM): $\vec{f}(d) = (d, \sqrt{d})$; interaction between student's race (indicator for each of black, white, Asian, Hispanic, other) and school's %Black, %White, %Asian and %Other.¹ Also interaction between the student's socioeconomic status (indicator for each of free lunch, reduced lunch, or none) with the school's % free lunch.
- 4) (MEDIUM2): Same as Medium except we also interact the distance vector with the student's socio-economic status (free lunch, reduced lunch, none). We also add an additional binary variable which is 1 if and only if both the school and the student are in East Boston, given that is geographically separate from the rest of the city.

For all models, distance is computed using the Google Map Walk Distance from the centroid of the family's geocode to the centroid of the school's geocode.

To fit the choice model, we also need to specify the choices offered to each student. Below are our assumptions on choice menus:

- Every English Proficient student is offered all same grade regular education programs within their walk-zone and the zone in which their geocode resides (home zone).
- Every Limited English Proficiency (LEP) student is offered all the programs English Proficient students are offered, plus the following:
 - Multi-lingual ELL programs: all within their walk-zone or home zone.
 - Spanish Language Specific ELL programs (only offered to Spanish speakers): all within their walk-zone or home zone.
 - Non-Spanish Language Specific ELL programs (only offered to students speaking that language): any in the city.

¹We do not include %Hispanic because it is linearly dependent of the others and serves as the omitted category.

We infer whether a student speaks a certain language by looking at the first language field in our choice data. Given that we are missing first languages for some students, we also augment this by assuming a student speaks a language if he/she chose or is assigned to a language-specific ELL program of that language. (We assume that Family Resource Center staff only allows students to choose a language specific ELL program if the language matches.) The reason we assume all non-Spanish Language Specific ELL programs are citywide is that we observed out-of-zone, out-of-walk-zone choices for these programs, and because the BPS choice website states:

If your child needs special education services or an English Language Learner program, as determined by testing or evaluation, the schools available to him/her could be different than those listed here. Please consult with staff at any FRC to learn more.²

Although ELL programs are only offered to LEP students whose ELD score is 1-3, in our choice data we have ELD scores only for 741 out of 2,676 LEP students, so we make the conservative assumption that every LEP student is potentially offered an ELL program.

3.1 In-Sample Fit

Figure 1 show the estimates and standard errors of different specifications. Table 2 tabulates the log likelihoods.

Specification	Log Likelihood (LL)	Gain in LL	# Additional Variables	LL Gain per variable
SIMPLE	-49486			
SIMPLE2	-49342	144	1	144
MEDIUM	-48766	576	24	24
MEDIUM2	-48716	50	7	7.1

Table 2: Log Likelihood (In-Sample Fit).

The total number of choices indicated by all 5,758 students in 2011-2012 choice file is 20,533. Since the specifications are nested, it is possible to evaluate the relative gain in likelihood by adding additional parameters. Under the Bayesian Information Criteria (BIC), we are justified in including an extra parameter if it produces a log likelihood gain of $\frac{\ln 20533}{2} = 4.96$.

Observe that going from SIMPLE to SIMPLE2, the addition of square root of distance produces a sizable improvement in fit. This corroborates the intuition that while 0.5 miles and 1.5 miles may represent a big difference for families sending children to kindergarten, 4 miles and 5 miles may be not so important because the student would already be on their commute. From SIMPLE2 to MEDIUM, we add interactions with demographics, and this yields a log likelihood gain of 24 per additional variable. Intuitively these variables are important because they allow families of different demographic groups to have differential preferences for demographics and for unobserved attributes correlated with demographics, such as school atmosphere or safety. From the BIC test the gain from these variables is justified.

²Taken from BPS “What Are My Schools?” web app. on January 20, 2013. <https://externalweb.mybps.org/ExternalWeb/home.aspx?tabindex=1&tabid=32>

Specification:	<u>2011 Round 1</u>				<u>2012 Round 1</u>
	Simple (1)	Simple2 (2)	Medium (3)	Medium2 (4)	Out-of-sample (5)
walkDistance	-0.481 (0.008)	-0.139 (0.021)	-0.152 (0.021)		-0.115 (0.017)
walkDistance * Student-Free Lunch				-0.114 (0.03)	
walkDistance * Student-Reduced Lunch				-0.010 (0.101)	
walkDistance * Student-Non-Free/Reduced Lunch				-0.200 (0.062)	
walkDistance * Student-No Lunch Info				-0.169 (0.037)	
sqrt walkDistance		-1.198 (0.069)	1.115 (0.069)		-1.082 (0.057)
sqrt walkDistance * Student-Free Lunch				-1.073 (0.093)	
sqrt walkDistance * Student-Reduced Lunch				-1.409 (0.31)	
sqrt walkDistance * Student-Non-Free/Reduced Lunch				-1.239 (0.186)	
sqrt walkDistance * Student-No Lunch Info				-1.140 (0.111)	
Sibling	2.984 (0.05)	2.958 (0.05)	2.888 (0.05)	2.891 (0.05)	2.860 (0.046)
Walk	0.350 (0.023)	0.170 (0.026)	0.154 (0.026)	0.161 (0.026)	0.262 (0.022)
Present Program	5.945 (0.085)	5.938 (0.085)	5.884 (0.085)	5.896 (0.085)	5.261 (0.068)
East Boston				0.293 (0.142)	
School fixed effects	Y	Y	Y	Y	Y
Program fixed effects	Y	Y	Y	Y	Y
Program demographics x Student demographics	N	N	Y	Y	Y
Grades	K1-K2	K1-K2	K1-K2	K1-K2	K1-K2
Log likelihood	-49486	-49342	-48766	-48716	-68798
Number of Parameters	88	89	112	119	112
Number of students	5758	5758	5758	5758	6696
Number of choices	20533	20533	20533	20533	27899

Notes. This table reports maximum likelihood estimates from the multinomial logit discrete choice demand model for five different specifications. Standard errors are in parenthesis. School fixed effects are 75 school-specific variables corresponding to each ranked school in the choice set. Program fixed effects are 12 program-specific variables corresponding to each program type in the choice set. Program types are KED, BES, BEM, KEM, TEE, BEV, IEE, REG, BEC, BEH, TES, BEK. Program demographics x student demographics is the interaction of school % black, %white, %asian, %other with student's own race indicators, as well as school % free lunch, interacted with the student's lunch status indicators. All models exclude an outside good.

Figure 1: Model Coefficients (both in-sample and out-of-sample)

Comparing MEDIUM to MEDIUM2, we add interaction effects between student’s socioeconomic status and distance, capturing the intuition that families of taste for distance may vary with socioeconomic background. The gain in likelihood per variable is 7.1, which passes the BIC criteria. However, it is not far from the threshold. Moreover, the estimates are noisier, with no statistically detectable differences between free lunch students and non-free-reduced lunch students. To err on the side of being conservative, we therefore focus on the MEDIUM specification rather than MEDIUM2.

The estimates for the MEDIUM specification fitted with 2012-2013 Round-1 data are in column (5) of Figure 1. Interestingly, many of the point estimates are similar to those using a different year of data from 2011-2012, suggesting that there are some common patterns of student preferences. These parameter estimates are used to evaluate all of the assignment plans.

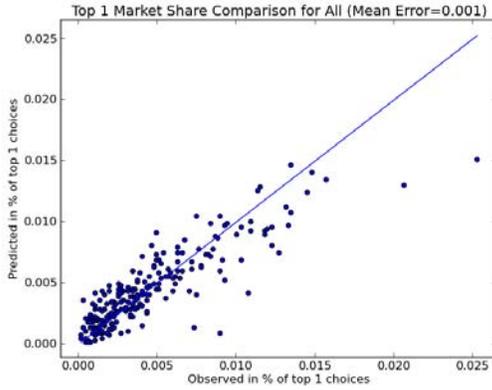
3.2 Out-of-Sample Fit

To test the predictive power of our demand model, we next examine how well the model fit using 2011-2012 data can predict the actual choices made by families in 2012-2013. When doing this, we use the specification MEDIUM but with the student set, student characteristics, and school characteristics from 2012-2013 data. This exercise constitutes an out-of-sample validation because without using their choice data, we estimate what families participating in 2012-2013 would have chosen and compare it to what they actually chose. If the predictions and actual choices are close, then we have confidence that our preference estimates capture the main patterns of behavior and provide a credible way to simulate alternatives.

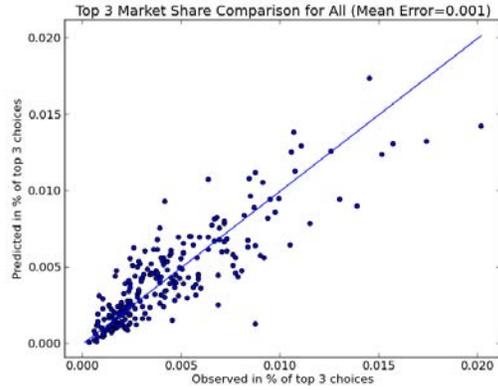
We report a series of scatter plots, each of which corresponds to a student set I and some parameter k . In the plots, each point corresponds to a school program j . The x-axis shows out of all students in I , the fraction of students for whom program j is among their top k *actual* choices. The y-axis shows for the same students, the fraction of students for whom program j is among their top k *predicted* choices. Predicted choices are computed from a single draw of the preference distribution, plus the idiosyncratic error for the empirical distribution of student and school characteristics. If we perfectly predict families’ relative choices, then points should lie on the 45 degree line. How closely the points hug the 45 degree line provides a measure of how well we can predict the top k choices of these families. Points that lie above this line means we over-estimated demand for this school program. Points below the line means we under-estimated demand.

Figure 2 shows how well we can predict the top 1, 3, 5 and 7 choices made by all families. The mean error reports the absolute difference between the prediction and the observed choices made averaged over all points and provides a convenient summary measure of how well the prediction matches actual choices. As can be seen although we only used 2011-2012 data in the model, we can predict choice shares relatively well in 2012-2013 in aggregate. While we successfully predict aggregate total choice share, it is possible that our demand estimates do not provide an adequate representation of the heterogeneity of preferences. To examine this possibility, we produce the top 3 choice share graph for students of different race (Figure 3), subsidized lunch status (Figure 4), grade (Figure 5), English proficiency (Figure 6) levels, and neighborhoods (Figure 7). If the fit is good for the majority of these subgroups, then we have confidence that our model captures the main aspects of preference heterogeneity. We focus on the top 3 to avoid too many extra exhibits, though examining top 1, 5 or 7 looks similar.

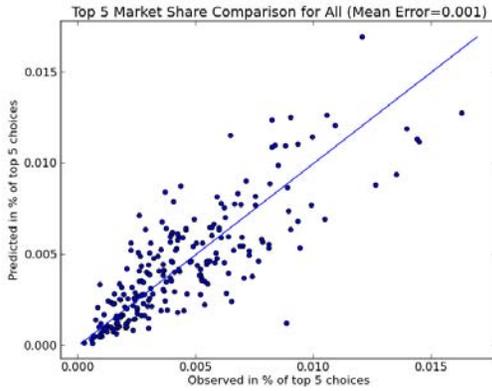
As shown by the out-of-sample choice share comparisons, our model predicts relative choice patterns between various programs reasonably well, both in the aggregate and for subgroups of students split by race, socio economic status, lep status, grade, or neighborhood. It is therefore possible to use these estimates to simulate the consequences of alternative assignment policies.



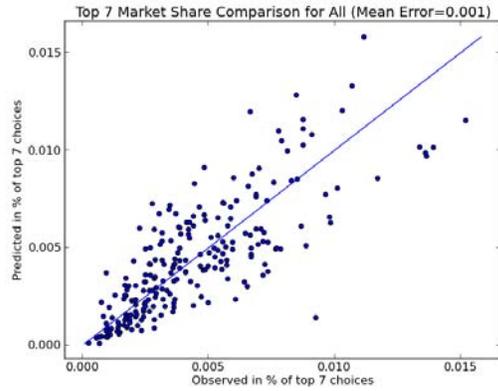
(a) Top 1 Choice



(b) Top 3 Choices

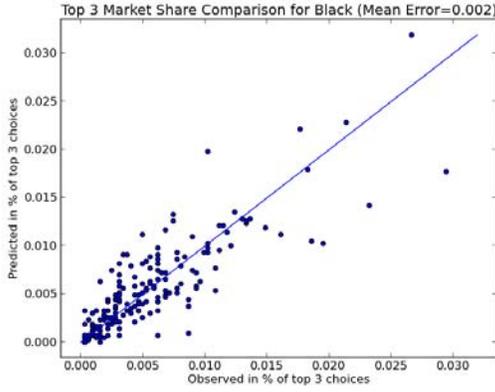


(c) Top 5 Choice

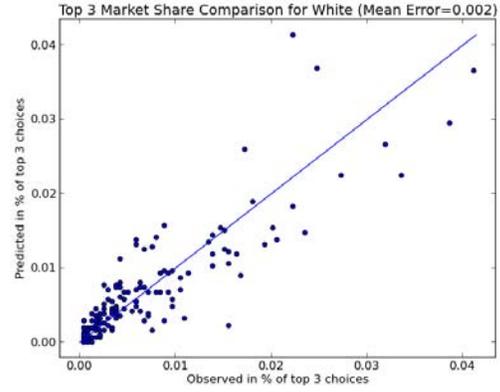


(d) Top 7 Choices

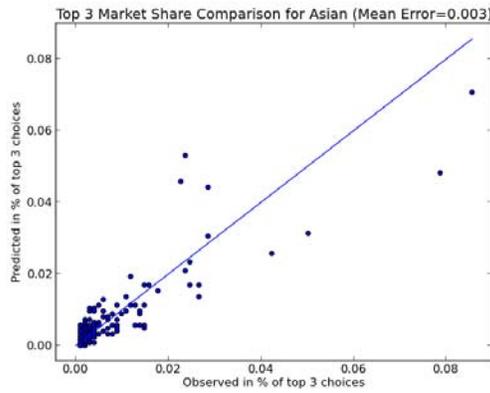
Figure 2: Out-of-Sample fit of MEDIUM specification, fitted using 2011-2012 data and evaluating on 2012-2013 data. Each point represents a school program. The x-axis shows the % of all families top k choices that is for this program. The y-axis shows what the model predicts.



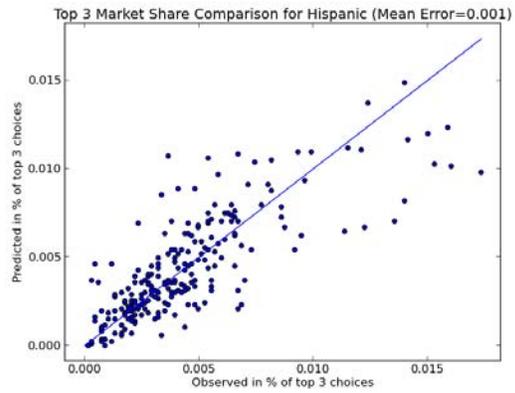
(a) Black



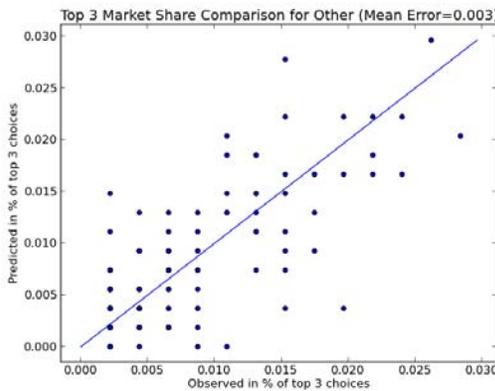
(b) White



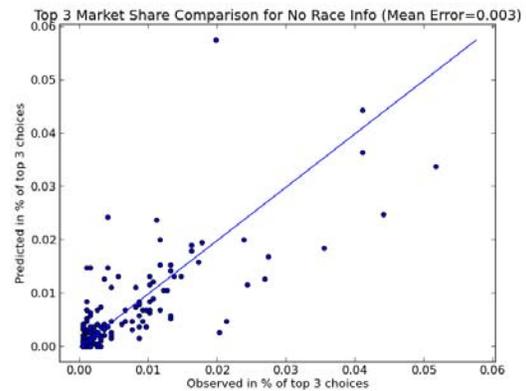
(c) Asian



(d) Hispanic



(e) Other

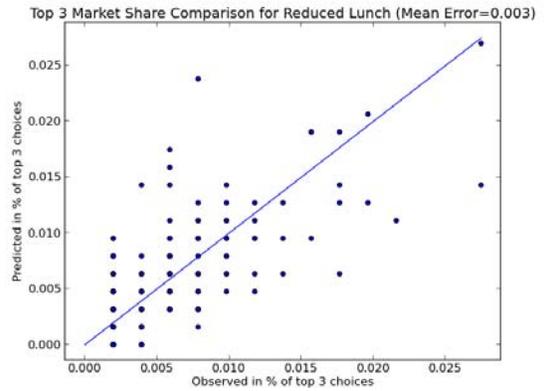


(f) No Race Information

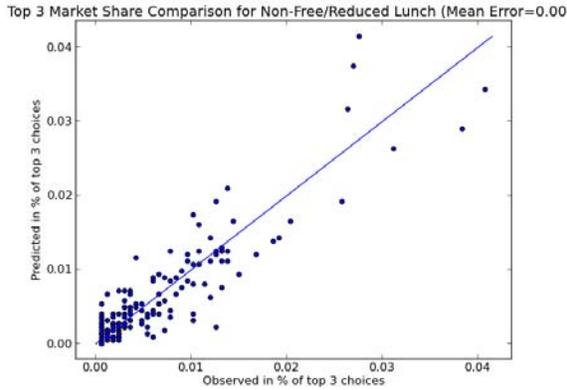
Figure 3: Out-of-Sample fit of MEDIUM specification by race. The x-axis shows the % of all families top 3 choices that is for this program. The y-axis shows what the model predicts. Note that we can predict well for all except “Other” which represents 3% of the data set and are spread across the city so their top programs are different from one another (so no program has a major share).



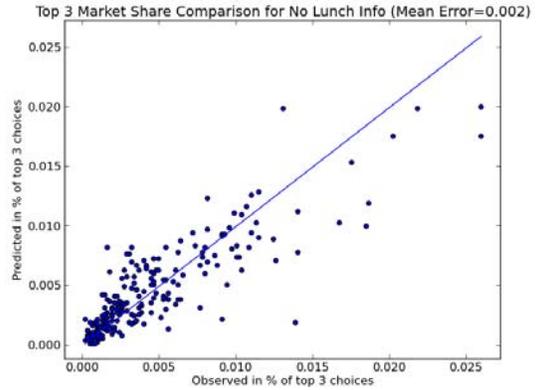
(a) Free Lunch



(b) Reduced Lunch

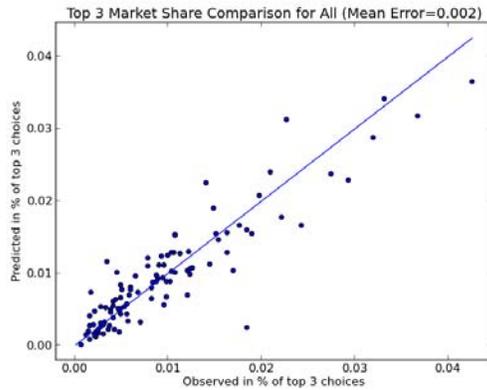


(c) Non-Free/Reduced Lunch

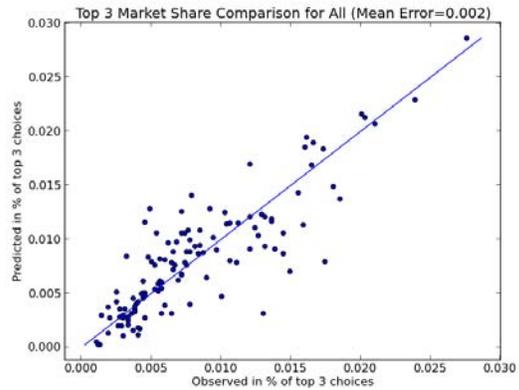


(d) No Lunch Information

Figure 4: Out-of-Sample fit of MEDIUM specification by subsidized lunch status. The x-axis shows the % of all families top 3 choices that is for this program. The y-axis shows what the model predicts. Note that we can predict well except for “Reduced Lunch” which represents 4% of our data and is spread across the city, so that their top programs are different from one another.

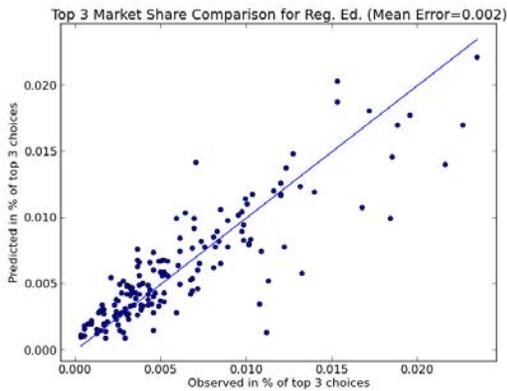


(a) Grade K1

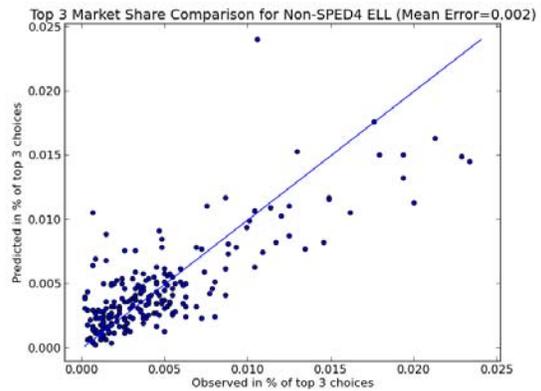


(b) Grade K2

Figure 5: Out-of-Sample fit of MEDIUM specification by grade. The x-axis shows the % of all families top 3 choices that is for this program. The y-axis shows what the model predicts.



(a) English Proficient



(b) Limited English Proficiency

Figure 6: Out-of-Sample fit of MEDIUM specification by LEP status. The x-axis shows the % of all families top 3 choices that is for this program. The y-axis shows what the model predicts. Note that we can predict choices better for English Proficient students than LEP students. This may be partly due to our lack of consistent data on which students have ELD 1-3 and are thus should be offered ELL-programs.

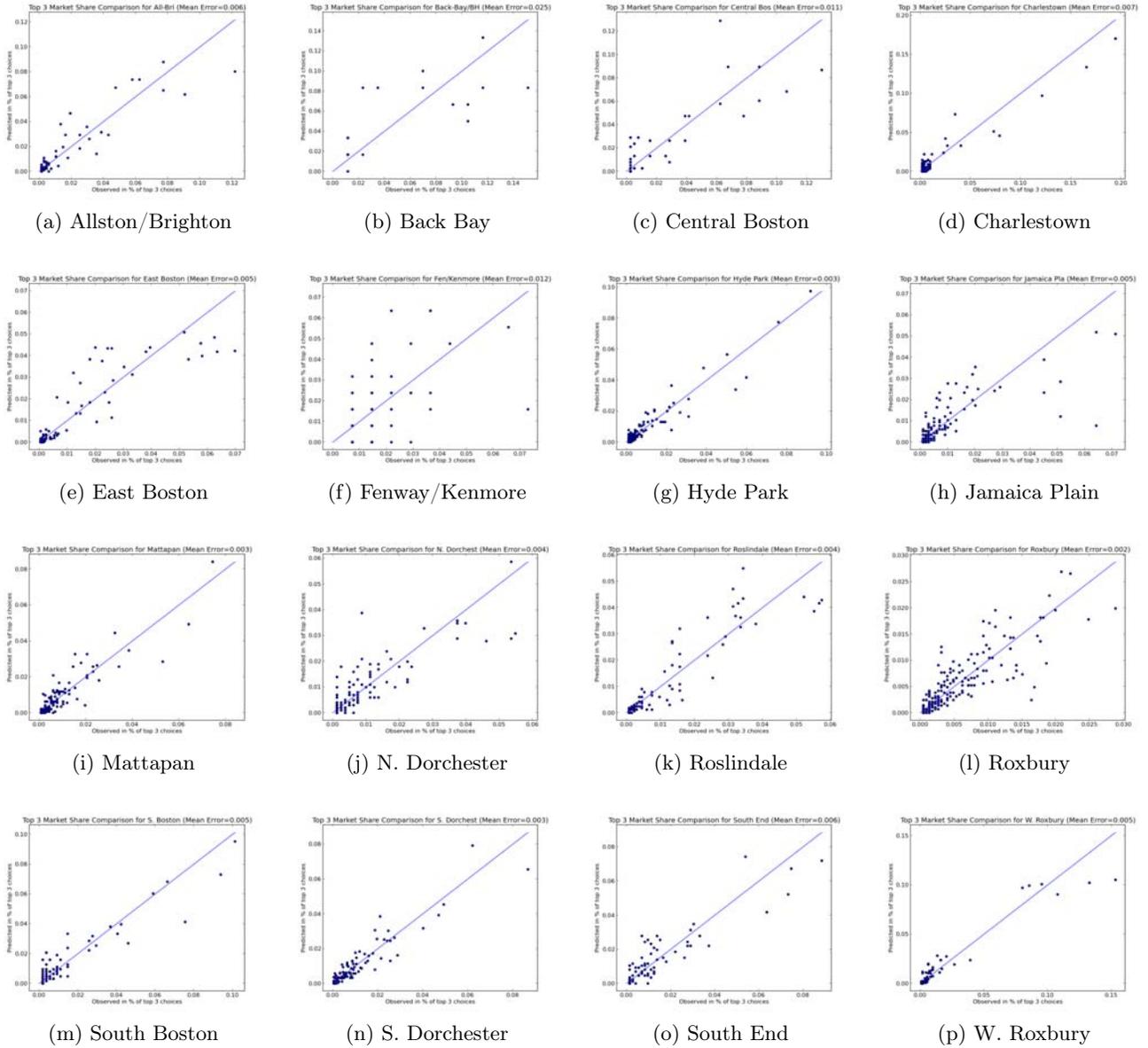


Figure 7: Out-of-Sample fit of Medium specification by neighborhood. The x-axis shows the % of all families top 3 choices that is for this program. The y-axis shows what the model predicts. For some neighborhoods, we have relatively few students in sample so the graphs for those neighborhood seems “pixelated.”

4 Replicating BPS Assignment Algorithm and Computing Cutoffs

4.1 Describing the Algorithm

In this section, we describe precisely how we implement the algorithm used by Boston Public Schools. There are a few reasons we do not expect to exactly replicate the BPS assignment. First, the BPS assignment algorithm uses all families choices, while our dataset only contains the top 10 ranked schools for each student. Next, the BPS mechanism has some minor exceptions, such as using the best random number for applicants who are twins, and also an administrative procedure for assigning unassigned students if capacity is available. Since we do not have all of the data that the algorithm uses, we aim to replicate the algorithm as closely as possible given the data we have. Table 3 shows that using 2012-2013 actual choice data, we can replicate 98.7% of BPS assignments (matching 6613/6698 students).

	# Students	% Replicated
Grade K1	2668	98.3%
Grade K2	4030	99.0%
All	6698	98.7%

Table 3: Replicating BPS assignment algorithm using our implementation using 2012-2013 Round 1 choice data. For every student, we look at whether he/she receives the same assignment (either to a school program or unassigned) in the BPS algorithm and in our algorithm, and we calculate the percentage of students we can match. (We treat administrative assigned as unassigned, so this is match with BPS algorithm before the administrative assignment stage.)

For each student i , there is a single random number r_i , which we normalize to be between 0 and 1, with 1 being best, 0 being worst. For each school program j (*i.e.* Quincy KED K2), we internally break the program into two portions—a walk-half and an open-half—according to the program’s “walk-percentage.” Hernandez has walk percentage of 0% since it is citywide, Orchard Gardens has walk percentage 75%, and all other programs have walk percentage 50%. When there are an odd number of seats, we allocate the extra seat to the walk half. For ease of exposition, we refer to each of these units (*i.e.* Quincy KED K2 Walk-half) as a “bin.”

For each bin b and student i , compute the student’s “score” π_{bi} as follows: Let j be the program that bin b represents.

1. Initialize $\pi_{bi} = r_i$
2. If the student has “Guarantee” or “FinalPass” priority to j (present program), then add 8 to π_{bi} and terminate at this point.
3. If the student has “PresentSchool” in his/her priority to j , then add 4 to π_{bi} .
4. If the student has Sibling in the school s_j , add 2 to π_{bi} .
5. If the program is a walk-half, and the student is in the program’s walkzone³, add 1 to π_{bi} .

³For precise definition of walk-zone, we are using a file from BPS which contains a geocode to school correspondence

The above scores π_{bi} induces a strict ranking of each bin b for all students.

For transforming students’ rankings of programs j to rankings of bins, there are two approaches described in Dur, Kominers, Pathak, and Sönmez (2012). In each approach, we replace where the program j is in the student’s ranked list by some ordered list of bins (specified below) representing that program, to form an ranking over bins.

- **Old Processing Order:** Students in the walk-zone rank first the walk-half, then the open-half. Students outside the walk-zone rank first the open-half, then the walk-half.
- **New Processing Order:** The open-half is divided into two: a first-open-quarter, and a second-open-quarter. The split is done evenly, rounding up for the second-open-quarter. All students, regardless of whether they are in the walk-zone, rank first the first-open-quarter, then the walk-half, then the second-open-quarter, in that order.

These two processing order defines a strict, partial ranking over bins for every student, as well as for every bin a strict, complete ranking over students. We then form a student to bin matching using the student-proposing deferred acceptance algorithm, and transform this into a student to program matching. For all of the validations and simulations involving the current 3-zone model, we use the Old Processing Order. For all of the new assignment plans, we use the New Processing Order, which in general is more advantageous to walk-zone students than the old processing order (Dur, Kominers, Pathak, and Sönmez 2012).

4.2 Estimating “Access” Using Bin Cutoffs

The scores provide a way to estimate “access.” In any assignment outcome, for each bin b , define its “bin cutoff” c_b as 0 if its capacity is not filled, and the minimum π_{bi} over the students assigned to it if it is filled. Intuitively, this is the minimum score needed to be offered an assignment in the bin, possibly displacing out the lowest-scoring applicant assigned to that bin.

For any student i and any program j , we define the student’s estimated access to program j , $a_j(i)$, as

$$a_j(i) = \max_{b \text{ representing } j} (\pi_{bi} - r_i) + 1 - c_b,$$

where $\pi_{bi} - r_i$ is the student’s “score boost” due to priority, as calculated in the way described before (*i.e.* 0 for NoPriority, 1 for Walk-Only, 2 for sibling, 3 for SiblingWalk, 4 for PresentSchool, etc). As a result, $1 - a_j(i)$ is the minimum random number r_i student i needs to get into one of the bins in j . The above calculation is exact if

1. when a student’s application in the algorithm causes another student to be rejected initiating a chain, it eventually returns to cause the first student to be rejected;
2. there is no difference between the minimum score that was assigned and the maximum score of someone who applied but was rejected out (distributed roughly as $\frac{1}{n}$ where n is the number of applicants of the cutoff priority level who apply to the bin).

of which schools are in each geocode’s walk-zone at each grade. This was computed by BPS by drawing a 1-mile circle from the school’s coordinate and adding any geocode that this circle touches. BPS has announced they will their definition of walk-zone to 1-mile straightline distance from school to home address starting 2013-2014.

In any finite problem, both of these situations are possible. However, when there are a large number of students compared to programs (as in our data where there are roughly 24 students per program), the approximation error from using $a_j(i)$ as a measure of student’s access to program j grows small.⁴

5 Length of Simulated Rank Order Lists

The model developed in Section 3 can tell us families’ *relative* preferences. However, since we do not model outside options (*i.e.* private schools, out-of-district charters, METCO, etc), this model does not tell us how many choices families will rank. We have not modeled what options families find “acceptable,” in the sense that they would actually attend in assigned. Because of the availability of outside options in reality presumably many families do have some “cutoff” for acceptability.

One possibility is to assume that families find every BPS option acceptable, in which case a utility maximizer would rank every single choice in their menu. However, the mode number of choices ranked (in 2012-2013 Round-1 data) is 5, and the mean is 4.2. If families were to rank every single choice in their menu, then we should see between 24 to 65 choices by everyone, which would imply a greater level of competition than what we observe in data.

Figure 8 plots how using the demand model from 2011-2012 data, we can predict the level of competition in 2012-2013 by various amounts of choice list truncation. We proxy competition using the program score cutoff (minimum bin cutoff of all bins b representing the program; see Section 4.2) changes as we truncate families choice lists to 1, 4, 7, 10, 13 and no-truncation respectively. The way to interpret the figure is as follows: the closer we are to the 45 degree line, the better we are predicting the level of competition for every individual program. For programs below the 45 degree line, we under-predict the level of competition. For programs above the 45 degree line, we over-predict competition.

As can be seen in Figure 8, without truncating choice lists, we over-predict demand (most points lie above the 45 degree line), and the mean error (0.298) is greater than if we had truncated at 10 (0.240). While 10 is not the truncation that yields the best fit (between 4 and 7 would be better), we truncate at 10 because of the following reasons:

- The choice data file we were provided truncated choice lists at 10.
- When using the demand model to simulate access to quality, we are conservative. Truncating at a lower number decreases the overall level of competition, which would increase the apparent access to quality.
- BPS recently proposed an option for families to have their rank lists automatically appended based on distance, which makes the effective choice lists of the families who ranked only a few programs longer.
- 10 performs reasonably well (error of 0.240 compared to 0.235 with 7).

When we simulate other options, the choice menu would change, but we have no systematic way to predict the right way to change our truncation point. To isolate the effects of changes in

⁴Kojima and Pathak (2009) formally demonstrate the first case is unlikely when the number of programs grows, holding fixed the length of applicant rank order lists.

choice menus, we truncate at 10 across all assignment plans. This implicitly is making the following **individual rationality** assumption:

Behavioral Assumption: Families consider only their top 10 choices in menu as “acceptable.” In other words, they prefer their outside option over their 11th choice.

This implies that the acceptability threshold is directly influenced by what menu they are presented. For example, if a family who has access to some schools considered “quality” in one assignment plan is offered more choices that are not considered “quality” but which they prefer (perhaps by virtue of being close to home and of good match for other reasons), then their “access to acceptable quality” would decrease. Therefore, offering participants more choices that are not considered “quality,” but which the applicant likes can decrease the person’s access to quality. Conversely, by taking away choices that are not considered “quality” but which the family prefers according to the demand model, one can increase their “access to quality” because suddenly the (perhaps far away) “quality” schools would fall within their top 10 list and they would find that acceptable. While it is possible that families’ acceptability threshold is influenced by framing and how the information on schools is presented, it is unclear about a more appropriate assumption. A restriction to ten choices seems more appealing to us than assuming that families find everything to be acceptable, which contradicts the fact that families do not rank all choices.

6 Validation of Simulation Approach

Equipped with a model for predicting families’ relative choices, a behavioral model of how many choices they will list, and an algorithm implementation that almost exactly matches the BPS algorithm, we can now simulate what might happen given a new assignment plan. As a validation of the simulation engine as a whole, we first predict what actually happened in 2012-2013 choice using 2011-2012 demand model.

For definitions of the metrics that we use to compare the different plans, see the main body of the report (Pathak and Shi 2013). We are unable to validate the following metrics because they all depend on the demand model (and for actual assignment we do not observe families complete preferences over all programs):

- Access to Acceptable Quality and Access to Acceptable Capacity (because we only count access if we estimate a program to be “acceptable” to student, and we do not observe what all that is “acceptable” in real choice data.)
- Element of Choice - Access to Top k Dream Choice.

We focus here on K2 New Families because these families do not have to weigh the influence of present school or sibling priority. As a result, we expect their choices to be more difficult to predict. The choice of K2 New Families is also consistent with the main report.

In the main report, all our figures are averages of 25 simulations (so we are showing expected values), where we draw from the estimated preference distribution 25 times. For this report, since we only observe one sample in our validation for 2012-2013, we draw preferences only once.

6.1 Access to Quality Ignoring Acceptability

The Access to Quality notion defined in the main report cannot be calculated using the actual choice data. This is because it is defined by access to high MCAS schools that the family finds “acceptable.” For simulated data, we can compute this metric because we can define what each family finds “acceptable” as their top 10 choices. For actual choices, we cannot identify what families find “acceptable” and assuming that they find unacceptable any choice they haven’t ranked, even for those who only rank a few choices, seems unrealistic.

However, we are able to evaluate an alternative measure of Access to Quality, that ignores their individual rationality conditions one at a time while maintaining the condition for other families. This corresponds to computing answering the following question: what is my access given the rankings of other participants and if I accept any BPS option if the MCAS is enough. While this notion is not self-consistent (we keep the acceptability threshold for everyone else, but ignore it for the family we are evaluating), it provides a simple way to evaluate the Access to Acceptable Quality in the main report, and it is an interesting to check whether we can predict this well.

For this, we simply simulate the assignment, then calculate the maximum estimated access $a_j(i)$ (See Section 4.2) for all “quality” programs j in student i ’s menu. Because $a_j(i)$ is based on random numbers of the bin-assignments, it is a well-defined quantity in both real and simulated assignments. To factor away the randomness due to random tie-breaking numbers, we endow student i with the same random number r_i as in the real assignment.

Figure 9 shows how we can predict this access to quality figure for Grade K2 new families in 2012-2013. To emphasize, this metric is different from that computed in the report because it ignores the individual rationality condition. The boxplots represent student data distributions. They are interpreted as follows: the red line shows the median. The box shows the 25% and 75% percentiles, respectively. The lower and upper line shows the min and the max, respectively.

6.2 Access to Top Menu Choice

Using score cutoffs, we compute for both actual and simulated assignment, the measures of access to top choice in the main report and the ranking of the choice received. Figure 10 shows that the proximity of our approach to what actually occurred.

6.3 Proximity

We aim to predict the distribution of distances to assignment in 2012-2013. Figure 11 shows that we are able to provide a close approximation to what happened that year.

6.4 Diversity

We aim to predict the distribution of % Assigned Class of a certain demographic group. Figure 12 shows that we can match both the distributions closely at all quartile levels.

6.5 Community

We calculate the actual # of same grade neighbors going to the same school (neighbors defined as within 0.5 miles of walking distance), and compare with what our simulation yields. Note that our utility model for students assume that families make decisions independently (when in reality several families may choose one school together). Moreover, the uniform truncation of choice list may also deviate from data (in which choice list length may be correlated based on geography). Both of these reasons tend to make our prediction lower than the actual, which is what we see in Figure 13. Nevertheless, on this dimension, the match is still reasonable (actual average is 4.44 compared to the predicted average of 3.85.)

7 Simulation Methodology for Evaluation

Having validated our simulation engine, we finally describe how we use it to evaluate various assignment plans in the main report (Pathak and Shi 2013).

In simulating assignment plans, we simulate each assignment plan 25 times and report the average of 25 draws. For each simulation, we hold the random tie-breakers and the realizations of the random variables fixed to ensure that the outcomes are comparable. More precisely, for each run (from 1 to 25), we generate the following for every student:

- A random tie-breaker for every student.
- A relative ranking of all BPS programs (including those that are not in the student’s choice menu). We call this the “dream ranking” (if the student could rank every BPS program, even those not in their menu, what would their relative ranking be.)

This sample of draws is held fixed regardless of the assignment plan.

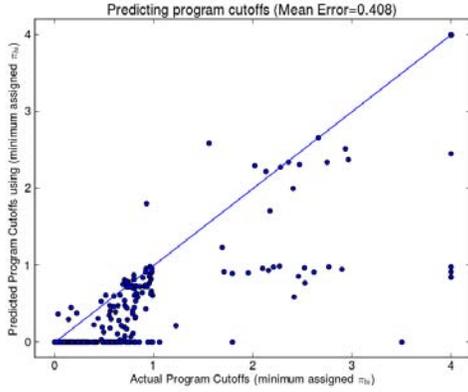
For each plan, we then compute the actual rankings of each student by taking the dream ranking and removing the programs not in his/her choice menu, and truncating this list to 10. We then feed this generated choice data and random numbers to the assignment algorithm, from which we obtain simulated assignment outcomes.

There are 25 simulated outcomes for every plan, and the outcomes are linked across plans in the underlying “randomness” so they are comparable. We then evaluate these outcomes using a variety of metrics. We average the 25 results and producing a consistent measure of relative performance across plans.

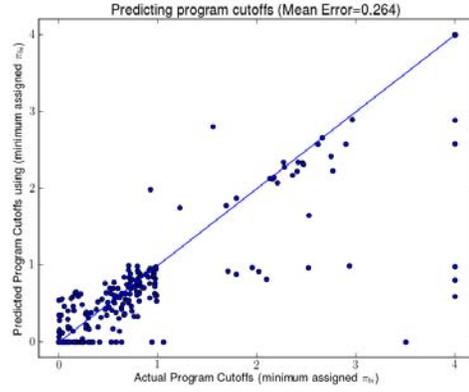
Note that this averaging makes all our estimates “expected values,” which is conditioned on the student’s observable characteristics, but taking expectation over the student’s ϵ_{ij} (random variable for “taste” in 1), random number, and everyone else’s random number and idiosyncratic taste shock. Suppose, for example, that we estimate a student’s distance to assignment is 2 miles. This number comes from a calculation where, given the student’s attributes, we draw preferences and the random-tie breaker 25 times for each student, and compute the average distance across 25 realizations of the assignment mechanism with these inputs.

References

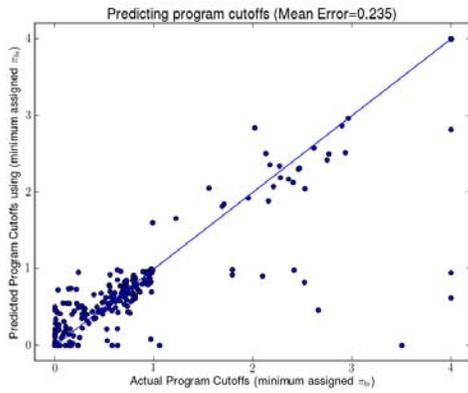
- ABDULKADIROĞLU, A., P. A. PATHAK, A. E. ROTH, AND T. SÖNMEZ (2005): “The Boston Public School Match,” *American Economic Review, Papers and Proceedings*, 95, 368–371.
- ABDULKADIROĞLU, A., AND T. SÖNMEZ (2003): “School Choice: A Mechanism Design Approach,” *American Economic Review*, 93, 729–747.
- ABULKADIROĞLU, A., N. AGARWAL, AND P. A. PATHAK (2012): “Sorting and Welfare Consequences of Coordinated School Admissions: Evidence from New York City,” Unpublished working paper, MIT.
- DUR, U., S. D. KOMINERS, P. A. PATHAK, AND T. SÖNMEZ (2012): “Priorities vs. Precedence in School Choice: Theory and Evidence from Boston,” Unpublished working paper, MIT.
- KOJIMA, F., AND P. A. PATHAK (2009): “Incentives and Stability in Large Two-Sided Matching Markets,” *American Economic Review*, 99, 608–627.
- PATHAK, P. A., AND P. SHI (2013): “Simulating Alternative School Choice Options in Boston,” Discussion paper, MIT School Effectiveness and Inequality Initiative.
- PATHAK, P. A., AND T. SÖNMEZ (2008): “Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism,” *American Economic Review*, 98(4), 1636–1652.
- WALTERS, C. R. (2012): “A Structural Model of Charter School Choice and Academic Achievement,” Unpublished working paper, MIT.



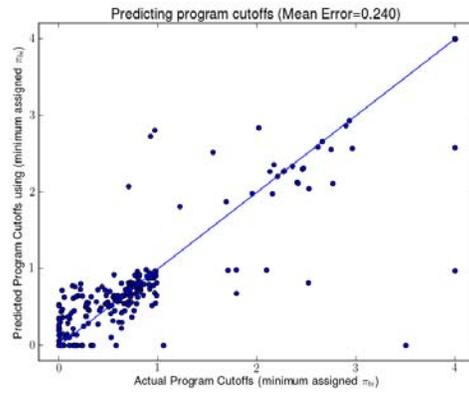
(a) Truncate to 1



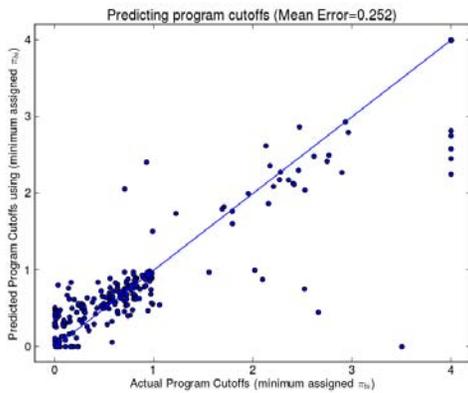
(b) Truncate to 4



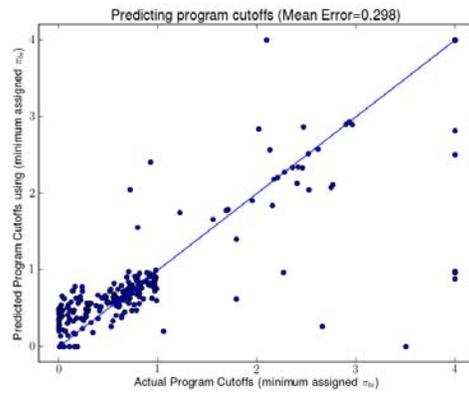
(c) Truncate to 7



(d) Truncate to 10

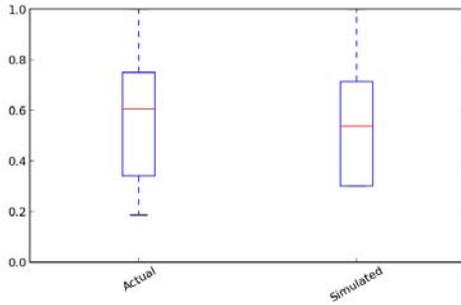


(e) Truncate to 13

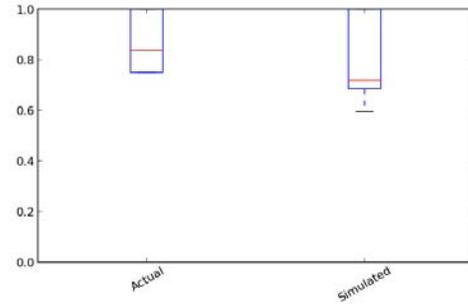


(f) No truncation

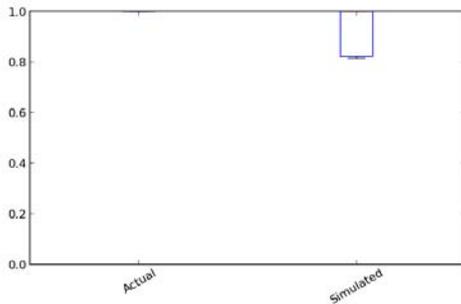
Figure 8: How well we predict competition in 2012-2013 (using Medium model fitted using 2011-2012 data). Each point is a program. The x-axis plot the observed minimum bin cutoff of the program (1 minus this cutoff is the exactly the access for student without any priority.) The y-axis plot what is predicted given we truncate choice list to what is indicated. Each plot shows an independent run of 1 sample. We use the actual random numbers in both cases. The more we hug the 45 degree line, the more we predict competition exactly. Programs below this line means we are under-predicting competition, and above this line means we are over-predicting.



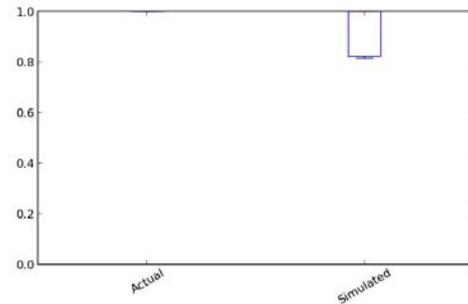
(a) Top 1/4 MCAS



(b) Top 1/3 MCAS

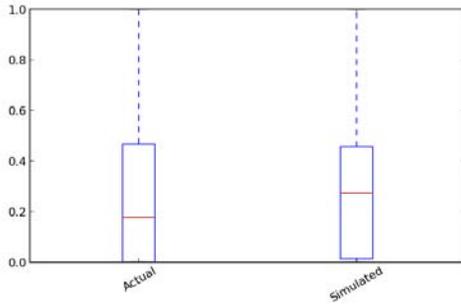


(c) Top 1/2 MCAS

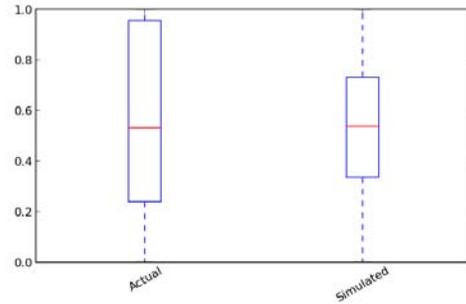


(d) Top 2/3 MCAS

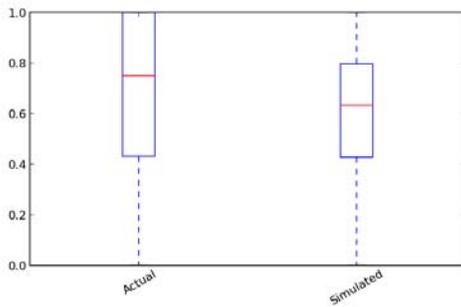
Figure 9: Box plots comparing simulated access and actual access, both estimated using score cutoffs for new families (non-present school, non-siblings) in Grade K2. (See Section 4.2). For each family, we estimate what would happen if everyone else ranks schools as they do, but they find all BPS programs to be acceptable and rank to maximize their access to each group of Top MCAS schools. We then aggregate this across students. (Note: this approach ignores the family’s acceptability threshold when calculating their access, but maintains it when calculating others’ access.) Simulated represents one simulation, using the MEDIUM 2011-2012 demand model, truncating rank order list length at 10. As can be seen in the box-plots, we reasonably match real access in various thresholds, but we slightly under estimate access because competition in our environment—where everyone ranks 10—is tougher than in actual assignment data, in which the average is less than 5.



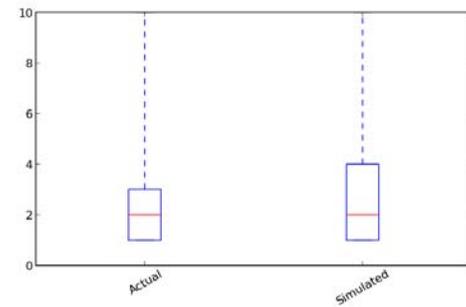
(a) Top 1 Choice



(b) Top 3 Choice



(c) Top 5 Choice



(d) Choice Number Got

Figure 10: Box plots comparing simulated access and actual measures of access to top choice measures, both estimated using score cutoffs.

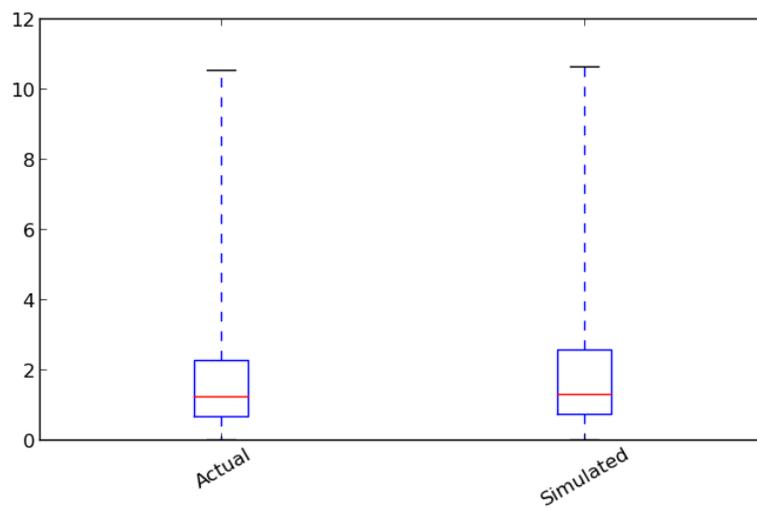
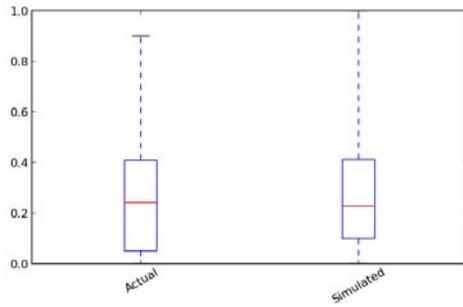
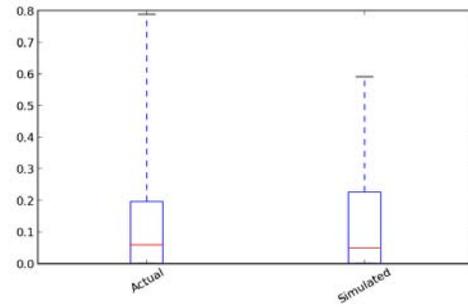


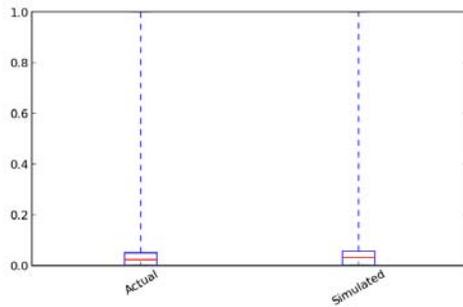
Figure 11: Box plot comparing actual distance to assignment versus predicted distance to assignment.



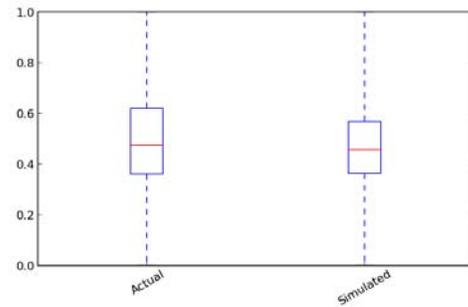
(a) Black



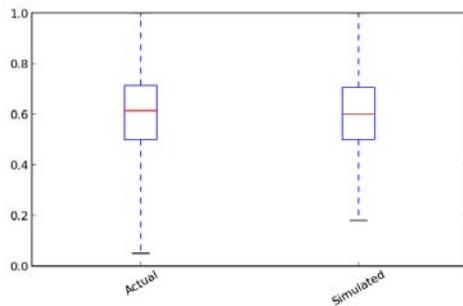
(b) White



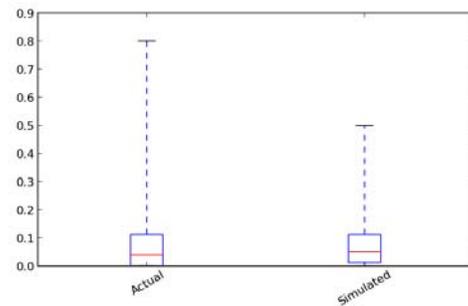
(c) Asian



(d) Hispanic



(e) Free Lunch



(f) Non-Free/Reduced Lunch

Figure 12: Box plot comparing simulated % of Assigned class (school program) of a certain demographic group and actual.

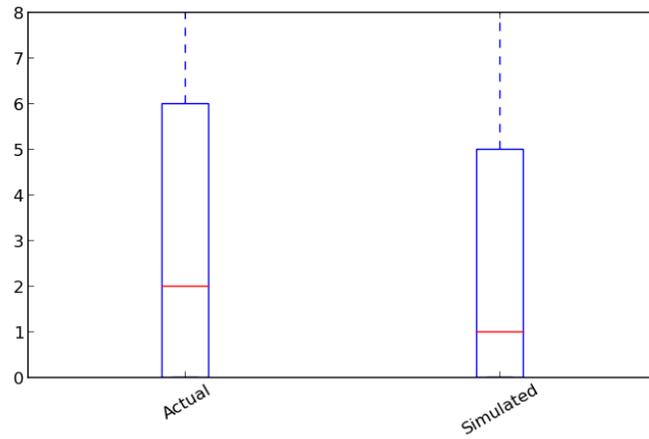


Figure 13: Comparison of actual same grade neighbor count (# others who live within 0.5 miles walking distance who go to the same school and same grade as me) with simulated result.