Prediction and Optimization in School Choice

by

Peng Shi


Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

© Massachusetts Institute of Technology 2016. All rights reserved.

Author ............................................................ Sloan School of Management
.................................................. April 14, 2016

Certified by .......................................................... Itai Ashlagi
............................................................................
Assistant Professor
Thesis Supervisor

Accepted by .......................................................... Dimitris Bertsimas
.......................................................... Boeing Leaders for Global Operations Professor
............................................................................
Co-Director, Operations Research Center
Prediction and Optimization in School Choice

by

Peng Shi

Submitted to the Sloan School of Management on April 14, 2016, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Operations Research

Abstract

In this thesis, I study how data-driven optimization can be used to improve school choice. In a typical school choice system, each student receives a set of school options, called the student’s menu. Based on his/her menu, each student submits a preference ranking of schools in the menu. Based on the submitted preferences, a centralized algorithm determines the assignment.

In Boston, New York City, Chicago, Denver, New Orleans, Washington DC, among other cities, the assignment algorithm is the student-proposing deferred acceptance (DA) algorithm, which can also incorporate a priority for each student at each school. These priorities may contain a deterministic as well as a random component. An advantage of this algorithm is incentive-compatibility, meaning that no student has incentives to misreport his/her preferences.

The first research question of this thesis is how to optimize the menus and priorities so that students have equitable chances to go to the schools they want, while the city’s school busing costs are controlled. The second question is how the assignment algorithm can be modified to keep the same assignment probability of every student to every school, while improving neighbors’ chances of going to the same school.

To answer these questions, I build a multinomial logit (MNL) model to predict how students will rank schools under new menus, and validate the predictive accuracy of this model out of sample. I also propose a simple plan for menus and priorities, called the Home-Based plan, and compare with other proposals using the MNL model. (As a result of this analysis, the Home-Based plan was adopted by Boston in 2013.) I then show how one can further optimize the menus and priorities under the MNL model, by developing a new theoretical connection between stable matching and assortment planning, as well as methodologies on solving a new type of assortment planning problem, in which the objective is social welfare rather than revenue. Finally, I show how to further optimize the correlations between students’ assignments to improve neighbors’ chances of going to the same school.

Thesis Supervisor: Itai Ashlagi
Title: Assistant Professor
Acknowledgments

First, I would like to thank my PhD adviser, Prof. Itai Ashlagi, for all his advice, mentorship, financial support and friendship during the past five years. I learned a lot from him during my PhD and really appreciate being his student. I would also like to thank my other thesis committee members: Prof. Parag Pathak, Prof. David Simchi-Levi, and Prof. Alvin Roth. They have all been instrumental in the development of my research direction and philosophy, and I’m grateful for their time and effort investing in me as a researcher.

I would also like to thank the faculty, colleagues, and staff of the Operations Research Center (ORC) for providing such a great academic program and environment for research. Thanks to Prof. Dimitris Bertsimas and Prof. Patrick Jaillet for being such resourceful and personable co-directors. Thanks to Laura Rose and Andrew Carvalho for the help in paperwork and logistics (and for the much appreciated reminders for administrative deadlines). Thanks also to all the fellow students for making the ORC such a fun and collegial place to work.

The practical and empirical components of this thesis would not have been possible without the kind support of staff at Boston Public Schools (BPS), who provided data and institutional insights about the student assignment system in Boston. Specific individuals I would like to thank by name include Timothy Nicolette, Carleton Jones, Peter Sloan, Kamal Chavda, Jack Yessayan, Jerry Burrell, Kim Rice, James Rcanelli, Barry Kaufman, and Superintendent Carol Johnson. Thanks also to city of Boston staff Rebecca Frisch for the help during the 2012-2013 Boston school assignment reform, as well as to all the members of the External Advisory Committee. I would also like to thank the many parents that have helped me during this time by sharing ideas and giving feedback to my work. They include Bruce Kiernan, Josh Weiss, John Mudd, Ian Whittle, Ann Walsh, Kim Janey, Frank Hart, Meira Levinson, Maura O’Tool, and Megan Wolf.

I would also like to thank my family and friends. Most importantly, thanks to my wife Annie for praying for me, giving me helpful and honest feedback on my
presentations and writing, and for all the various practical and emotional support. Thanks to my parents for supporting my PhD studies and for all the sacrifices made for my education and development. Thanks to David Kwabi, Gerald Pho, Sam Perli, Mark Ku, David Miculescu, Po-Ru Loh, and many other friends from the MIT Graduate Christian Fellowship for walking with me personally and spiritually during the past five years. I would also like to thank Lee and Miltinnie Yih for their personal mentorship and for inviting me to the Bible study in their home.

Finally, I would like to thank God, who is the ultimate source of all wisdom and intelligence, power and strength, opportunity and resource, and to whom belongs all praise and honor and thanks. May He establish this work and use it for His glory.
Contents

1 Introduction .......................... 23
   1.1 Motivating Problem: School Choice in Boston ............... 24
   1.2 Outline of Thesis and Research Contributions ......... 25
      1.2.1 Note to Readers .................................. 31

2 Building a Predictive Model of Assignment Outcomes .... 33
   2.1 Introduction ............................ 33
   2.2 Background .................................. 36
      2.2.1 Assignment Plan: Problems ....................... 39
      2.2.2 School Assignment Reform ....................... 41
   2.3 Problem Formulation ....................... 42
   2.4 Data and Analysis ......................... 45
   2.5 Demand Modeling .......................... 46
      2.5.1 Simulation .................................. 47
   2.6 Designing Simple Plans ...................... 49
      2.6.1 Other Zone-Based Plans ....................... 50
      2.6.2 The Home-Based Plan ....................... 50
   2.7 Results .................................. 52
   2.8 Discussion .................................. 55
   2.A The Home-Based plan Implemented in 2014 .......... 57
   2.B Estimating Multinomial Logit .................... 58
# Validating the Predictive Model Across a Policy Change

## 3.1 Introduction

## 3.2 Data

## 3.3 Methodology

### 3.3.1 Demand Models

### 3.3.2 Forecasting the Applicant Pool

### 3.3.3 Simulation

### 3.3.4 Evaluation

## 3.4 Results

### 3.4.1 Raw Prediction Errors

### 3.4.2 Decomposing the Prediction Error

### 3.4.3 Causes of Large Prediction Error in the Environment

## 3.5 Additional Analysis

### 3.5.1 Value of Random Coefficients

### 3.5.2 Tier Salience Effect

## 3.A Neighborhood Level Results

## 3.B Detailed Participation Forecasts

## 3.C Calibrating Mixed Logit using MCMC

### 3.C.1 Specifying the Likelihood Function

### 3.C.2 Overview of the MCMC

### 3.C.3 Our MCMC Sampler

# Optimizing Menus and Priorities to Improve Welfare Subject to Bus-

## 4.1 Introduction

### 4.1.1 Related Literature

## 4.2 Model

### 4.2.1 The Deferred Acceptance (DA) Framework

### 4.2.2 The Simulation Framework

### 4.2.3 The Optimization Problem

## 3 Validating the Predictive Model Across a Policy Change

## 3.1 Introduction

## 3.2 Data

## 3.3 Methodology

### 3.3.1 Demand Models

### 3.3.2 Forecasting the Applicant Pool

### 3.3.3 Simulation

### 3.3.4 Evaluation

## 3.4 Results

### 3.4.1 Raw Prediction Errors

### 3.4.2 Decomposing the Prediction Error

### 3.4.3 Causes of Large Prediction Error in the Environment

## 3.5 Additional Analysis

### 3.5.1 Value of Random Coefficients

### 3.5.2 Tier Salience Effect

## 3.A Neighborhood Level Results

## 3.B Detailed Participation Forecasts

## 3.C Calibrating Mixed Logit using MCMC

### 3.C.1 Specifying the Likelihood Function

### 3.C.2 Overview of the MCMC

### 3.C.3 Our MCMC Sampler

# Optimizing Menus and Priorities to Improve Welfare Subject to Bus-

## 4.1 Introduction

### 4.1.1 Related Literature

## 4.2 Model

### 4.2.1 The Deferred Acceptance (DA) Framework

### 4.2.2 The Simulation Framework

### 4.2.3 The Optimization Problem
List of Figures

2-1 This map shows the three zones in Boston’s three-zone plan. . . . . 37

2-2 These plots compare the relative market shares of programs as pre-
dicted by the demand model fitted using 2011 data with the actual
market shares in 2012. Although the predictions were not perfect,
they were reasonably good, because most of the dots are near the 45
degree line. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63

3-1 Number of new and continuing applicants to BPS (K0-2) . . . . . . . 68

3-2 Trend in total number of applicants to BPS. . . . . . . . . . . . . . . 81

3-3 Proportion of new applicants distributed into each grade, neighborhood
combination. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82

3-4 The proportion of new applicants of each grade. . . . . . . . . . . . 82

3-5 For each grade and neighborhood, the proportion of potentially contin-
uing students from the previous year who apply as continuing students
in the current year. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83

3-6 The percentage of top choices of each tier for K1. For example, to
see the actual percentage of K1 top 3 choices for tier 1, read the y-
coordinate for the Tier 1 plot with x-coordinate being 3. The predic-
tions from using the actual population and each of the demand models
are also listed. As can be seen, the actual outcome is leans toward
better tiers compared to logit or mixed logit, but the move is not as
drastic as predicted in the Naive model. . . . . . . . . . . . . . . . . 96
The percentage of top choices of each tier for K2. For example, to see the actual percentage of K1 top 3 choices for tier 1, read the y-coordinate for the Tier 1 plot with x-coordinate being 3. The predictions from using the actual population and each of the demand models are also listed. As can be seen, the actual outcome is leans toward better tiers compared to logit or mixed logit, but the move is not as drastic as predicted in the Naive model.

The diagram on the left shows the distribution of students and the capacities of schools. Each blue circle represents a neighborhood, with its area proportional to the expected number of students from that neighborhood. Each yellow circle represents a school, with its area proportional to the number of K2 seats available. The capacity schools are shaded. The distribution of students is based on the 2010-2013 average. The capacities are based on data from 2013. The right shows estimates of $Q_s$ (inferred quality) from the 2013 data. The size of the circle is proportional to the estimated $Q_s$, with higher quality schools having larger circles.

These plots show the parameters that define the optimal assortments from Section 4.4.3. Each neighborhood is giving an allowance of points, and each school costs a certain number of points. The diagram on the left shows the distribution of expected allowance $E[a_t]$ across neighborhoods, with each circle representing a neighborhood and the size of the circle proportional to the expected allowance. The diagram on the right plots the schools as circles, with the size of each circle proportional to the school costs. These are not monetary costs, but shadow costs from the optimization (OptAssortBoston) from Section 4.4.3.
4-3 These plots show the expected utilities of neighborhoods under the optimized plan. Each circle represents a neighborhood and the size of the circle is proportional to the expected utility. The left plots the values under the population distribution and utility distributions from Section 4.4.1. The right plots the values under the actual 2014 population and re-estimated utility distribution. The red oval in the plot on the right shows the biggest area of utility decrease. This corresponds to the Hyde Park region of Boston. 

4-4 These plots show the inferred qualities of schools in the utility distribution. Each circle represents a school and the size of the circle is proportional to the inferred quality. The left plots the values in the original utility distribution in Section 4.4.1, which is estimated from 2013 data. The right plots the updated utility distribution based on 2014 data. The red oval shows the decrease in school quality in the Hyde Park region. 

5-1 Random assignment in the 2-school case: students who prefer school 1 with higher than cutoff priority level (set $D$) get in for sure; students who prefer school 1 with exactly the cutoff priority level (set $E$) get in with probability $p = \frac{q}{|E|}$; students who prefer school 2 or whose priority level is lower than cutoff (set $F$) get in school 2 for sure.

5-2 Assignment within community $c$: horizontal dimension is students and vertical dimension is lottery numbers. $z_1^*, z_2^*, z_3^*$ and $z_4^*$ are lottery cutoffs for community $c$ of schools 1, 2, 3, 4 respectively (well-defined because we assume everyone in the same community are in the same priority class). $d_3(z'|c)$ is the demand function for school 3 at lottery $z'$. $p_3(c,y')$ is the probability student $y'$ is assigned to school 3. The area shaded with the same color represents the 1-dimensional measure of assignment probabilities of students from this community to a school ($\bar{p}_s(c)$).
Diagram illustrating structural identity (Proposition 5). The triangle represents 3 quantities that add up to a constant. There are 3 possible cases: A) high between-community variation in assignment probability, so baseline cohesion is already high. B) high within-community heterogeneity in assignment probability, so there is nothing we can do by Proposition 3. C) significant potential to improve cohesion by correlated lottery. Corollary 2 shows that potential to improve can be interpreted as baseline cohesion multiplied by a weighted average of the SCV (squared coefficient of variation) of the demand function (with respect to the lottery $z$). The SCV can be interpreted to correspond to a mixture of preference correlation and competition. Proposition 4 shows that the improvement is upper-bounded by a measure of total uncertainty of the lottery. This relationship holds not only overall but also school by school.

Partition of Boston into .5 mile \( \times \) .5 mile squares. We treat each square as a community. Each circle corresponds to a geocode, and the area of the circle is proportional to the number of students residing at the geocode. Defined in this way, the average number of students per community is 2666/205 = 13 for K1 and 4030/205 = 19 for K2.
5-5 Geographic distribution of baseline cohesion and cohesion improvement for K1 new families. Each circle corresponds to a geocode. Its color corresponds to the average value of students living at that location, and its size is proportional to the number of students. The bar on the right of each map shows a box-plot based on the same color scale of the student level distribution, with the ends of the box showing the 25 and 75 percentiles, and the red line shows the median (parts of box plot may be outside of range to have colors show reasonable contrast). From the left plot, we see that cohesion is highest in East Boston (upper-right island), and varies in different pockets in the city. Improvement in cohesion from correlated-lottery is unevenly distributed, being highest in north east, south west, and various pockets of the city.

5-6 Geographic distribution of baseline cohesion and cohesion improvement for K2 new families. From the left plot, we see that cohesion is highest in East Boston (upper-right island), and various pockets in the city. Improvement in cohesion from correlated-lottery is very uneven, concentrated in north east, extreme south west, and eastern part of the city, with very small gains in the center of the city. The median cohesion gain is only about 0.25.

5-7 Plot of choiceAgreement(i) for K1 and K2 new families. For each student i, this is the average # of programs she has in common in her top 3 choices with an uniformly random peer i' from the same community. As seen, the median is almost twice larger for K1 new families than for K2. Moreover, for K2, this is low in the middle portions of the city, which matches the pattern of low gain from correlated lottery from Figure 5-6.
Interaction of menu change and correlated lottery for K2 new families. The menu we change to is Home Based A (the one adopted by the city). As can be seen, while the median increase from correlated lottery alone is about 0.4, and the median increase from menu change alone is about 0.6, the median increase from both together is about 1.8. The red region in the middle of the city in the second and third graph effectively disappears in the last graph, showing that large cohesion gain is achieved throughout the city when we apply these reforms together.
List of Tables

2.1 The table shows simulated performance of the assignment plans that
the EAC considred for its final vote in February 2013. . . . . . . . . . . 53

3.1 Aggregate supply and demand in grades K0-2, in years 2010-2013. . . 68

3.2 Applicants’ demographics across neighborhoods. . . . . . . . . . . . 69

3.3 Estimated coefficients for logit models. Each model is estimated using
maximum likelihood, using the choice data for grades K1 and K2. The
standard errors are estimated using the inverse of the Hessian of the
log likelihood function at the maximum. . . . . . . . . . . . . . . . . . . 75

3.4 Coefficients for Mixed Logit (MLogit) models, compared to simple Logit. 79

3.5 RMSE across neighborhoods between forecasts and actual outcome. . 87

3.6 RMSE across neighborhoods for market shares between forecasts and
actual outcome. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88

3.7 Resimulating by using a combination of the updated demand model
(M), population (P), or rank-order list length (R). The original simu-
lation errors are given as references. . . . . . . . . . . . . . . . . . . . 90

3.8 Market shares when resimulating by using a combination of the up-
dated demand model (M), or population (P). The original simulation
errors are given as references. . . . . . . . . . . . . . . . . . . . . . . . . 91

3.9 Prediction error in number of applicants . . . . . . . . . . . . . . . . . 92
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.10 Resimulating by using a combination of the updated demographics (D), number of applicants (N), or school programs (S). The original simulation errors and when the exact population is fixed as actual are given as references.</td>
<td>94</td>
</tr>
<tr>
<td>3.11 The total variation between the estimated joint probability distribution of first and second choice, and the actual joint distribution. As can be seen, in most setups, mLogit outperforms Logit, with the exception of 2014 K1 under the setup “Updated-P”.</td>
<td>95</td>
</tr>
<tr>
<td>3.12 Comparing Access to Quality predictions with actual outcome for 2014 K1.</td>
<td>98</td>
</tr>
<tr>
<td>3.13 Comparing Distance predictions with actual outcome for 2014 K1.</td>
<td>98</td>
</tr>
<tr>
<td>3.14 Comparing Unassigned predictions with actual outcome for 2014 K1.</td>
<td>98</td>
</tr>
<tr>
<td>3.15 Comparing Access to Quality predictions with actual outcome for 2014 K2.</td>
<td>99</td>
</tr>
<tr>
<td>3.16 Comparing Distance predictions with actual outcome for 2014 K2.</td>
<td>99</td>
</tr>
<tr>
<td>3.17 Comparing Unassigned predictions with actual outcome for 2014 K2.</td>
<td>99</td>
</tr>
<tr>
<td>3.18 Forecast for total number of new applicants in 2014.</td>
<td>100</td>
</tr>
<tr>
<td>3.19 Forecast for proportion of the new applicants distributed to each grade and neighborhood in 2014.</td>
<td>100</td>
</tr>
<tr>
<td>3.20 Forecast for continuing ratios for each grade and neighborhood in 2014.</td>
<td>100</td>
</tr>
<tr>
<td>4.1 Parameters of the MNL model, estimated from preference data from 2013. The values can be interpreted in units of miles (how many miles a student is willing to travel for one unit of this variable).</td>
<td>141</td>
</tr>
<tr>
<td>4.2 Simulation Comparisons of 3-Zone, Home-Based, and Optimized plans.</td>
<td>149</td>
</tr>
<tr>
<td>4.3 Means and standard deviations of the proportion of K2 applicants from each neighborhood, estimated using 4 years of historical data.</td>
<td>152</td>
</tr>
<tr>
<td>4.4 Re-evaluation of 3-Zone, Home-Based, and Optimized plans using updated parameters from 2014 data. All the results are averages from 10,000 independent simulations.</td>
<td>155</td>
</tr>
</tbody>
</table>
4.5 Comparing the predictions from the large-market model of the Optimized plan and from simulations. The simulation results are different than those in Section 4.4.4 because we use a different set of quotas and a population distribution with less variation, in order to focus on market size aspect of the approximation. We also report a reduced set of metrics because certain metrics are difficult to evaluate in the large-market setting without simulations.

5.1 Assignment probabilities from example. For example, student A from community I prefers school 1 over 2, and in the randomized assignment she is assigned to school 1 with probability \( \frac{61}{140} \) and school 2 with probability \( \frac{9}{140} \). With remaining probability \( \frac{1}{2} \), she is not assigned (or assigned to an outside option).

5.2 Correlated lottery implementation of the assignment probabilities from Table 5.1. This randomizes over 4 assignments, each represented by a column. For example, assignment i is chosen with probability \( \frac{61}{140} \), and assigns A and B to school 1, E and F to 2, and leaves the rest unassigned. Note that no matter what the realized assignment is, communities stay together as much as space allows.

5.3 Order of priorities used in Boston elementary school assignment in 2012. The earlier level priorities are more important, with later levels only used to break ties. The 3rd level is applied only for walk-zone halves.

5.4 Summary statistics of the choice data.

5.5 Percentage of students of each type getting their top choice. For example, for K2, 47% of students are continuing from K1, and 95% of them get their top choice.
5.6 Impact of lottery correlation on cohesion for various groups of students.

For example, for K1 new families, conditional on being assigned, students on average can expect to find 1.36 same-community peers. If we use correlated-lottery, the number of same community peers improves by 79% to 2.44. Based on the given assignment probabilities, no correlation procedure can produce cohesion greater than 3.26. Using correlated lottery, the average K1 new family gains 1.08 additional same-community peers. No correlation procedure can produce a gain greater than 1.90.

5.7 Expected number of same-community peers co-assigned conditional on being assigned outside of walk-zone. This is the number of same-community neighbors a student who is traveling outside of his/her 1-mile walk-zone can expect to find at his/her assigned program. So for a K1 new family, conditioning on the student going outside of walk-zone for school, he/she on average has only 0.53 neighborhood peers in the current lottery. But with correlated lottery this more than doubles to 1.26. Without changing assignment probabilities, we cannot expect this to be larger than 1.73.

5.8 Baseline and Improved Cohesion with differing walk-zone percentages.

The first three columns correspond to students’ expected # of same-community peers co-assigned conditional on being assigned. The last three columns correspond to the same number conditional on being assigned outside of walk-zone.

5.9 Cohesion with differing choice menus. The first three columns correspond to students’ expected # of same-community peers co-assigned conditional on being assigned. The last three correspond to the expected # of same-community peers conditional on being assigned outside of walk-zone. The rows are sorted in increasing baseline cohesion.
5.10 Descriptive regression of improvement in cohesion for new families in various choice menus. This seeks to explain variation in the 18 rows of data in Table 5.9 in terms of grade, uncertainty of lottery, and preference correlation. “Grade” is dummy for K2. “Uncertainty” is the average individual assignment variance, $E_i[\sum_s p_{is}(1 - p_{is})]$. “Preference Correlation” is the average number of common programs in the top-3 lists of two randomly chosen students from the same neighborhood. Robust standard errors are shown in parenthesis. Consistent with Proposition 4, the magnitude of improvement is increasing in uncertainty and preference correlation.

5.11 Descriptive regression of individual improvement in expected # of same-community peers from correlated lottery. Uncertainty of lottery is measured by variance of individual lottery ($\sum_s p_{is}(1 - p_{is})$). Preference agreement is proxied by product of size of community and average # of top 3 choice agreements with community ($|c(i)|\text{choiceAgreement}(i)$).

Non-continuing sibling and continuing are dummies. The standard errors are from clustering on the 205 communities. As can be seen, students for whom the lottery is more uncertain, whose preferences are highly correlated with their communities, and who are not siblings nor continuing students can expect the greatest cohesion gains from correlated lottery.

5.12 Impact of correlated lottery on racial or socio-economic diversity. For example, when we use independent lottery implementation for K1, on average 36.7% of pairs of students assigned to the same school are of the same race. This percentage increases by only 0.2% to 36.9% when we use correlated lottery.
Chapter 1

Introduction

This thesis is about using predictive modeling and optimization to improve school choice, so that students regardless of home address have equitable chances to go to the schools they want, while local communities remain cohesive and the city's transportation burden is controlled. Public school choice is implemented in many US cities, including Boston, Cambridge, Charlotte-Mecklenburg, Chicago, Denver, Minneapolis, Miami-Dade, New York City, New Orleans, Newark, and San Francisco. In these cities, students apply to public schools by submitting their preference rankings for various schools to a centralized system, which computes the assignment based on various priorities and lotteries. There’s a large literature in Economics studying these systems, but few previous studies examine externalities of choice, such as busing costs, and most consider only Pareto efficiency. This thesis contributes to the literature by considering busing costs and global efficiencies, as measured by utilitarian and max-min welfare.

The motivating problem is the 2012-2013 Boston school assignment reform, in which I had the privilege of working with Boston Public Schools (BPS) to analyze various proposals. I also proposed my own proposal, the Home-Based Plan, which was implemented in 2014. This collaboration motivated several ensuing academic projects, which became different parts of this thesis.
1.1 Motivating Problem: School Choice in Boston

Since 1988, BPS has assigned students to elementary and middle schools by dividing the city into three geographic zones. About six months before school starts, new applicants submit a ranked-order list of any number of schools in their zone, plus a one-mile radius “walk-zone.” Given the submitted preferences, BPS assigns students using the Gale-Shapley Deferred Acceptance (DA) algorithm, which guarantees that students have no incentives to misreport their preferences.

Despite the good properties of this algorithm, there were several complaints to the school choice system:

1. High transportation burden: In 2012, Boston spent about $80 million dollars a year, about one tenth of the entire school board budget, on school busing. This is because the zones are so large, so each school has to pick up children from an area about one-third of the city.

2. Equity of access: families complained that the demand/supply ratio for quality schools is worse in the East zone compared to the West zone.

3. “Illusion of choice”: although families have about 30 choices to rank, many do not get their top choices. So the large variety of choice seems like an illusion. The high number of choices also increases the burden on families to research schools.

4. Low community cohesion: Twelve children on the same street might go to twelve different schools, so children do not know their neighbors, and local community is weakened.

Despite these issues, Boston did not want to move to a neighborhood schools system, because schools are heterogeneous, and choice helps families to find schools that best serve their children.

Our problem then is to find a choice plan that decreases transportation costs, improves community cohesion, while maintaining welfare and equity.
1.2 Outline of Thesis and Research Contributions

This thesis has four chapters. The first two chapters are about predicting what would happen under a different choice plan. The last two chapters are about how to optimize the choice plans to improve welfare, equity and community cohesion, while reducing transportation burdens.

Chapter 2: Building a Predictive Model of Assignment Outcomes

In this chapter, we develop a multinomial logit (MNL) discrete choice model of how students would rank schools given new sets of choice options, based on the student’s home location, historic demand for the school, and student and school characteristics. We validate the model by using 2011 data to fit the model and comparing predicted market shares with 2012 choices. Using this demand model, we forecast the performance of various proposed plans by simulation, evaluating in terms of equity of access to quality, proximity to home, variety of choice, predictability, bus coverage area, socio-economic diversity, and community cohesion.

We use the simulation engine to evaluate the 3-zone plan, a 10-zone plan proposed by BPS, and the following plan that I proposed, called the Home Based plan. Every student can rank the union of the following sets: a) every school within a one-mile radius; b) the closest 2 top 25% schools; c) the closest 4 top 50% schools; d) the closest 6 top 75% schools; e) the closest 3 capacity schools, which were to be chosen by BPS to accommodate excess demand.

We find that while no plan dominates another, the Home-Based Plan performs well throughout the portfolio of metrics. Comparing with the 3-zone plan, the Home-Based plan sacrifices variety of choice for large savings in busing distance and bus coverage area. It also achieves large gains in community cohesion, and maintains similar equity of access to quality, predictability, and socio-economic diversity.

The research contribution of this chapter is to mathematically model the key trade-offs in the 2012-2013 Boston school choice reform, and building an empirically
relevant predictive model to inform policy. The content of this chapter appeared before in Pathak and Shi (2013) and Shi (2015).

The policy impact of this analysis is that Boston adopted the Home-Based plan in 2014. The original Home-Based plan is described in Section 2.6.2 and the version of the plan implemented is described in Section 2.A.

Chapter 3: Validating the Predictive Model Across a Policy Change

One concern with the previous methodology is that it assumes how students choose schools remains stable across the reform. However, when the Home-Based plan was implemented, the city changed the way school options were presented. Previously, information about schools were buried in a thick booklet, and one has to do additional research to compare schools. After the reform, the city created a tiering of schools based on quality quartiles, and created a website that asked for the student’s home address and clearly presented in one page the quality quartile of each school and its distance to the student’s home. Hence, one might be concerned that if students’ choices were affected more by behavioral issues such as framing than by underlying preferences, then a predictive model based on past data might no longer be valid after the reform.

Another concern is that the MNL model restricts possible substitution patterns, so that the probability that a student chooses one school or another as his/her $k$th choice does not correlate with how he/she chose for the first $k-1$ choices. This rules out the scenario that a student who chooses a school with high test scores first would more likely choose other schools with high test scores for subsequent choices. In the literature, this is called the Independence of Irrelevant Alternatives (IIA) assumption. One way to bypass this restriction is to use a mixture of MNL model, in which students’ preferences for various characteristics are drawn from a distribution instead of being constants.

We compare the MNL model with a mixture of MNL model, as well as a hier-
archival model that follows how one would think families would choose under the new framing: choose schools based on quality quartiles first, then based on distance within each quartile. We show that conditional on the student population, the MNL model performs similarly to the mixture of MNL model, and both outperform the hierarchical model. Moreover, the choice model is stable before and after the reform, in that the demand model fitted from pre-reform data performs almost as well as the demand model fitted from post-reform data. This validates MNL as a reasonable way to model students’ preferences.

An caveat is that when we tried to predict outcomes using only pre-reform data, we were not able to adequately predict the population of students who apply. Hence, when we used the forecasted rather than actual applicants, the MNL model did worse than the hierarchical model in overall predictions for grade K2. The mis-predictions in population are largely due to 1) a simultaneous change in the BPS test for English Language Learners (ELL), which decreased the percentage of such students from about 40% to 15%, and 2) our lack of accurate data on continuing students, which presumably would not have been a problem if BPS were doing the analysis.

The research contribution of this chapter is being one of few works that evaluates the predictions of a structural model using the actual outcome. To prevent post-analysis bias, we published forecasts before the reform in a “pre-analysis report,” which is rare in the Economics or Operations literature, but common in the medical literature. The content of this chapter appears also in P. and Shi (2015).

Chapter 4: Optimizing Menus and Priorities to Improve Welfare Subject to Busing Constraint

Having arrived at a reasonable predictive model, we turn to optimization. Specifically, we consider how to alter the choice menus and priorities of students to maximize a weighted sum of utilitarian welfare and max-min welfare, while staying within a busing constraint. For each setting of menus and priorities, we use the Deferred Acceptance (DA) algorithm to compute the assignment, so students have no incentives
to misreport their preference rankings within their choice menus. To model the busing constraint, we restrict the average distance bused and the number of options outside of walk-zone to be within that of the Home-Based plan.

Direct optimization is difficult because of the high number of combinations of choice menus and priorities, and because Deferred Acceptance is difficult to analyze under random preferences and priorities. Instead, we optimize a fluid approximation with infinitesimal students, then evaluate by simulation in the discrete setting.

The research contributions of this chapter are three-fold. First, we connect stable matching with assortment planning. We show that any mechanism based on stable matching with given menus, priority distributions, and capacities, can be represented as random assortments: offering each student a random assortment of schools, and letting the student pick his/her favorite among the offered assortment. Here, the priority distributions can only depend on the student’s public information, such as continuing/sibling status or home location. Similarly, the probability of getting each assortment depends only on public information. Conversely, any random assortment mechanism can be represented as stable matching under certain menus, priority distributions and capacities.

This equivalence allows us to formulate the optimization as a Linear Program (LP), with the decision variables being the probability of offering each class of student each assortment of schools. Ignoring continuing students and siblings for now, the LP can be reduced to iteratively solving the following, socially-optimal assortment planning problem: for each neighborhood, find the optimal assortment of schools to offer to maximize expected utility, minus the opportunity cost to other neighborhoods. The opportunity costs arise from the shadow costs to the capacity and busing constraints. Since costs can be viewed as the negative of revenue, this can be equivalently viewed as assortment optimization to maximize the sum of revenue and a multiple of consumer welfare. This generalizes the revenue-maximizing assortment planning problem studied in the operations management literature.

The second research contribution is showing that many results in the revenue-maximizing assortment planning literature extend to the socially-optimal case. Specif-
ically, we give efficient algorithms for socially-optimal assortment planning under MNL utilities and matroid constraints, under multi-level Nested Logit utilities and a cardinality constraint at each nest, and under a Markov-chain-based model.

The third contribution is an optimized plan for Boston: we show that the optimal assortments from the LP have a simple structure: for each neighborhood, we give a quota of points, possibly randomized. Each school costs a number of points. The quotas represent how much we favor a neighborhood, and is adjusted to obtain equity of welfare across neighborhoods. The costs represent the opportunity cost of offering a school, and is higher for schools with higher popularity or lower capacity. The optimal assortment for a neighborhood includes

- any school within the walk-zone with cost less than the neighborhood’s quota;
- up to 8 schools outside of walk-zone with the highest score, defined as

\[
\text{score} = \text{average utility of neighborhood for school} + \log(\text{quota} - \text{cost of school} - \text{distance}).
\]

We translate these random assortments into menus and priorities, which we can implement with Deferred Acceptance in the discrete model. We show that the fluid approximation is fairly good in our dataset, being within one percent of the discrete simulation in utilitarian welfare, average distance, and probability of getting top choice. We also show that the optimized plan significantly improves over both Home-Based and the 3-zone plan in utilitarian welfare, max-min welfare, busing savings, and probability of getting top and top 3 choices. When taking into account possible errors in forecasts, we show that the gains are maintained in all metrics except for max-min welfare. This makes sense because max-min welfare delicately depends on the balance of supply and demand and quality of schools, and this is disturbed when school qualities and populations change.

The content of this chapter is based on my working paper, Shi (2016), which builds on Ashlagi and Shi (2015).
Chapter 5: Optimizing Lottery Implementation to Improve Community Cohesion

Adjusting menus and priorities affects assignment probabilities. An orthogonal direction to optimize how these probabilities map to actual assignments. In this chapter, we study how we can express the same assignment probabilities as different convex combinations of assignments, so as to maximize community cohesion, which we define as proportional to the number of pairs of students from the same community going to the same school.

The first research contribution is to show that while the optimization problem is NP-hard even with 2 schools, we have a heuristic that performs well in practice. We test our heuristic on real data from Boston, and show that we can significant increase community cohesion. In fact, the improvement under the 3-zone system from correlated-lottery alone is greater than the improvement from changing to any of the assignment plans considered by the city committee. However, the improvement is not equitably distributed across the city, with communities near the center of the city seeing little improvement. When we applied both the Home Based Plan and correlated lottery, we observe a multiplier-effect: The total gain of applying both is greater than the sum of applying each separately, being a factor of 3 improvement for grade K1 and a factor of 2 for K2. Moreover, the improvement is more equitably distributed, as all communities improved significantly in cohesion.

The second contribution is a large-market analysis that gives insights on how the benefit of lottery-correlation depends on the primitives of the school district. We study a fluid approximation with a continuum of communities, a continuum of students within each community, and finitely many schools. We prove that under this model, any assignment mechanism that is non-atomic, symmetric within each priority class, incentive compatible, and Pareto efficient within each priority class, can be represented as Deferred Acceptance with Single Tie-Breaker (DASTB): a student’s priority score for a school is the sum of a constant that depends on his/her priority class and the school, and the student’s lottery number, which can without loss of
generality be distributed uniformly between zero and one. Note that the same lottery number is used for the student at all schools. When each community falls under only one priority class, we show that the benefit from lottery-correlation can be written as a constant minus the Herfindahl index of school sizes, minus the between-community and the within-community variations in assignment probabilities. This suggests that lottery-correlation matters more when school sizes are similar, and when preference heterogeneity is low. The latter explains our empirical findings because the communities near the center of Boston have greater preference heterogeneity, so lottery-correlation initially does not help much; under the Home-Based Plan, preference heterogeneity in those communities is reduced, so the benefit from lottery-correlation increases.

The content of this chapter also appears in Ashlagi and Shi (2014).

1.2.1 Note to Readers

Each of the chapters originally come from its own paper, so can be read separately without need to reference other chapters. As a result, there may be differences in definitions and assumptions. To reduce confusion, I list the main differences here:

- The set of metrics to compare assignment plans is different between chapter 2 and 4. The difference is that in chapter 2, the metrics are based on what was chosen by the city committee. In chapter 4, the metrics are my interpretations of the best metrics to capture the underlying trade-offs of an assignment plan.

- The definition of community cohesion is different in chapters 2, 3, and 4. In all three chapters, community cohesion for a student is defined as the expected number of neighbors the student has at the assigned school. In chapters 2 and 3, a student’s neighbors are defined to be anyone who lives within 0.5 miles from the student. In chapter 4, for tractability, we define neighbors differently: we divide Boston into communities using a 0.5 mile by 0.5 mile grid, and define neighbors to be those living within the same cell of the grid.
• The data set used in the empirical evaluations are different between chapters. Chapter 2 and 5, which were written earliest, are based primarily on data from 2012. Chapter 3 and 4 are based on updated data from 2013 and 2014.

• The set of variables included in the multinomial logit choice model are different between chapters. The model used in chapter 4 is the most basic model, which includes a fixed effect for each school, a distance component, and an indicator for being within the one-mile walk-zone. The model in chapter 2 and 5 adds to this a term corresponding to the square root of distance, as well as interactions between socio-economic and racial data of the student as well as the school. The model in chapter 3 does not have the square root term but also include census tract data on the average household income in the student’s zip code.
Chapter 2

Building a Predictive Model of Assignment Outcomes

2.1 Introduction

Many public school districts in the United States have implemented school-choice systems to offer families more control in selecting the schools their children will attend, and to help school districts balance supply and demand across schools. Typically, families are offered a menu of potential school options, and apply by submitting a rank-order list of preferences for their children. A centralized algorithm assigns children to one of their top-choice schools if capacities allow; otherwise, it breaks ties using a system of priorities and lottery numbers. As of 2013, a partial list of public school districts that have implemented such a system includes Boston, Cambridge, Charlotte-Mecklenburg, Chicago, Denver, Minneapolis, Miami-Dade, New York City, New Orleans, and San Francisco.

Most of these districts include schools in which the demand is higher than the capacity, because families tend to select the best-performing schools based on standardized test scores and other quality metrics. Determining which children should be selected to attend which schools is a nontrivial task. Allowing all children to access all schools (i.e., giving each child the possibility of attending any school in the school system) and using a lottery to break ties is arguably the most equitable; how-
ever, this may result in unsustainably high transportation costs if a city must pay for school busing, because the school buses would need to pick up children from a large geographic area. Thus, a city has to navigate the delicate balance between offering sufficient access to quality schools and a variety of choices, while keeping choice (in the remainder of this chapter, choice refers to school choice) menus manageable. It must also consider families’ preferences, because access to a particular school depends on both the number of seats available and the number children competing for these seats.

The Boston public school system (BPS) was facing this problem. Since 1988, its choice system for elementary and middle schools had employed a three-zone plan, which partitioned the city into three geographic zones. A family could seek admission for its children to any school in its zone and any school within a walk zone (i.e., one mile) around its home by ranking that school. Thus, each family was given a large number of choices on its menu; partly because of this, the assignment system resulted in high transportation costs to the city, low predictability for families, and scattering children from the same neighborhood across the city. Furthermore, the quality of schools between the zones was a concern; the zone boundaries had been drawn in 1988 and demographics and the quality of some schools had changed since then. In January 2012, Boston launched an initiative to reform the three-zone plan, and a mayor-appointed external advisory committee (EAC) held public meetings to deliberate and to elicit feedback from the community. Important considerations for the new plan included equity of access to quality schools, predictability, variety of choice, proximity to home, transportation savings, transparency, community cohesion, and ability to adapt to future change. Despite the political push, there was no guarantee that the reform would result in a satisfactory solution, because similar reform attempts in 2004 and 2009 had failed as a result of concerns over the equity of access to quality schools in the new proposals.

I became aware of the reform by providence as I read online news articles about it, and was drawn to the issue because it aligned with my research interest of market design—creating allocation systems with the right incentives to induce the desired
outcomes. The internal EAC meetings were open to the public. So, I attended many of these meetings, and met concerned parents, BPS staff, city committee members, and activists. Initially, I mostly listened to what other attendees were saying. Through repeated interactions with them, I gained a clearer understanding of the reform, and formulated it as an optimization problem: find the choice menus and priorities that would induce the desired outcomes based on a variety of metrics set by the EAC. In October 2012, at a public meeting to elicit alternative proposals from the community, I gave a short presentation on an optimization-based proposal, which attracted positive interest. Based on feedback from city committee members and concerned parents, I simplified my proposal and presented it at an EAC meeting in November, at which it was well received by both BPS and the EAC. Subsequently, I began working more closely with BPS staff and was given access to student-level data on how families had previously ranked schools. I also began collaborating with Parag Pathak from the School Effectiveness and Inequality Initiative (SEII) at the Massachusetts Institute of Technology. In December, the EAC officially commissioned us to provide a rigorous analysis of a short list of proposed plans.

Using previous choice data, I fit a discrete choice model for how families would select schools under various choice menus, and built a simulation engine that predicted outcome metrics in any given plan. BPS used this engine to refine its proposals, and I used it to analyze new proposals. The quality of each school was ranked based on a metric specified by BPS and the EAC. In one of these proposals, home-based A, a family’s menu of options could include (1) any school within its one-mile walk zone, (2) the two closest schools ranked in the top 25 percent, (3) the four closest schools ranked in the top 50 percent, (4) the six closest schools ranked in the top 75 percent, and (5) the three closest capacity schools, which were designated to help BPS meet excess demand. On January 30, 2013, we published our commissioned report, which provided detailed analyses of the short list of proposals selected by BPS, showing differences by race, grade, student type, and neighborhood. The EAC deliberated on this list of proposals, elicited public feedback, and voted to adopt the home-based A-plan because the choices in the menus were close to students’ homes, while also
providing equity of access to quality schools based on a variety of thresholds; in addition, it could automatically adapt to future changes in school quality. On March 13, 2013, the Boston School Committee, the regulatory body for the school board, approved the plan for implementation in 2014.

This project illustrates the power of quantitative analysis and simulation in bypassing rhetorical gridlock in public policy and informing the debate. It shows the importance of listening to various stakeholders to formulate the correct quantitative problem, and of constraining the solution so that it is within the organization’s implementation capabilities. It also suggests the potential of further applications of simulation and optimization techniques in public service allocation.

2.2 Background

In 2012, about 9,500 students applied to BPS to attend one of the kindergarten grades, K0, K1, or K2. K2 is the main entry grade to elementary school, and corresponds to the age when mandatory-schooling laws take effect in Massachusetts. BPS had 77 elementary schools, and each school had at least a regular education (Reg. Ed.) program for K2; some also had also K1 or K0 programs. In addition to Reg. Ed. programs, some schools had specialized programs for English language learners (ELL), ELL students of a specific home language, or special education (SPED) students.

Since 1988, families in Boston had been applying to BPS via its three-zone choice plan, which partitioned Boston into three geographically contiguous zones, called the East, the West, and the North zones (see Figure 1). Each child’s family could apply on that child’s behalf to any school within the family’s geographic zone and any school within a one-mile radius of its home (i.e., walk zone). Each school could have multiple programs for the same grade, such as Montesorri, ELL, or SPED. When a child applied to a school, that child could apply to any program for which he (she) was eligible. These programs together constituted the child’s choice menu. For K2 students, the choice menu contained between 24 and 42 Reg. Ed. programs, depending on the student’s geographic zone and walk zone, with additional programs for ELL or SPED
Figure 2-1: This map shows the three zones in Boston’s three-zone plan.

In January of each year, a family planning to enroll a child for September of that year had to submit a rank-ordered list of programs within its choice menu, in the order of its preferences, and could rank and submit as many programs as it would like. A computer algorithm based on a system of priorities and lottery numbers assigned children to programs, taking as input the families’ choices and the programs’ available capacities.

Since 1999, the priorities for school selection had been the following:

1. Continuing students (i.e., students who had been enrolled at the school for the previous grade and were advancing a grade) had the highest priority at the schools at which they were continuing.

2. Among students who were tied based on continuing status, those with currently enrolled siblings at the school were given priority.

3. Among students who were tied based on continuing and sibling status, those who lived in the school’s walk zone (an approximate one-mile radius around the school) received priority; however, this priority was applicable to only 50 percent of the seats in each program. The remaining 50 percent of the seats were open; no consideration was given to whether students came from inside or
outside the walk zone.

4. A random lottery number for each student determined the remaining ties. Each student received only one lottery number, which was applicable to the child’s application to all schools.

Given students’ choices and schools’ priorities, BPS still needed a way to compute the final assignment in a fair and efficient way; this method of determining the assignment is called the assignment algorithm. Since 2006, Boston had been using the student-proposing deferred acceptance (DA) algorithm, as described below.

1. Find an unassigned student with a nonempty choice list, and let the student apply to his (her) first-choice program.

2. Let the program tentatively accept the student; if this acceptance causes the program to violate its capacity limit, then let the program find the tentatively accepted student with the least priority and reject that student. Remove the program from that student’s choice list and unassign that student.

3. Go back to Step 1 until each unassigned student has an empty choice list.

The output of this algorithm does not depend on the order in which students apply in Step 1 (Roth and Sotomayor (1990), Gale and Shapley (1962) originally proposed this algorithm, and Abdulkadiroglu and Sönmez (2003b) adapted it to school choice. Based on a theorem by Dubins and Freedman (1981) and Roth (1982), Abdulkadiroglu and Sönmez (2003b) showed that as long as the priorities are predetermined and do not depend on students’ choice submissions, the above algorithm cannot be influenced by any strategy (i.e., regardless of what other students submit, no student can benefit by manipulating his (her) own choice submissions). This property influenced Boston to adopt the algorithm in 2006 (Abdulkadiroglu et al. (2006)). Alvin Roth, one of the economists responsible for introducing this algorithm to school choice in Boston and other cities in the United States, as well using it for medical-residence matching, and Lloyd Shapley, who first proposed this algorithm, won the Nobel Prize in Economics in 2012.
After BPS had run the DA algorithm, some students could still be unassigned. For students applying to K0 or K1, BPS would place them on a wait list and accept them later if spaces opened up. Many of these students might not be accommodated that year because BPS did not guarantee placement in K0 or K1; however, it was required by law to accommodate every K2 student. Therefore, after the algorithm terminated, BPS would administratively assign unassigned K2 students to the remaining available seats based on distance; if necessary, it would open new classrooms before the school year began to make space for all K2 students. If all rooms in a school building were occupied, BPS might add outdoor modular classrooms.

Each year, about 80 percent of the K0, K1, and K2 applicants participated in the January-February round of the assignment algorithm, which BPS called Round 1. BPS also executed three smaller assignment rounds between February and June. At the completion of these rounds, any unassigned students were assigned manually based on idiosyncratic considerations.

Boston provided transportation to any elementary school student who lived further than a one-mile straight-line distance from his (her) school. During the 2011-2012 school year, it bused about 18,500 elementary school students, including students up to grade 5; the average straight-line distance between home and school for the bused students was 2.0 miles.

2.2.1 Assignment Plan: Problems

The major criticisms of the three-zone plan were inequitable access to quality schools, high transportation costs, low predictability of assignment, and community dispersion.

Inequitable access to quality schools: The original zone boundaries were drawn in 1988 to achieve racial and socioeconomic balance (Alves and Willie, 1990); however, by 2012, demographics had shifted; the East zone had a higher concentration of poverty than the other zones. The percentages of elementary school students eligible for a free lunch because they lived in a household with a gross income of not greater than 130 percent of Federal poverty guidelines were 58, 53, and 52 percent for the
East, North, and West zones, respectively. BPS statistics published in 2012 show that schools in which students exhibited lower performance and slower advancement in the Massachusetts comprehensive assessment system (MCAS) tests were also more concentrated in the East zone. Some argued that the inequity was exacerbated by the walk-zone priority, because students living outside the walk zones of higher-performing schools had lower priority for admittance to these schools. The debate about the correct proportion of seats at each school to which the walk-zone priority should be applied was ongoing because the proportion of seats fluctuated; for example, it changed from 50 percent in 1988 to 75 percent in 1990, then to 100 percent in 1996, and back to 50 percent in 1999. This illustrated the difficulty of achieving public consensus over questions of access and equity.

High transportation costs: In 2011-2012, Boston spent more than $80 million on school busing; this represented almost 10 percent of its total school board budget \cite{Russell and Ebbert (2011)}. One reason for this high cost was that the city was required by law to pay for transportation of elementary school children not only to public schools, but also to charter, private, and parochial schools, and for expensive door-to-door transportation for certain SPED students. Nevertheless, a significant share of this cost resulted from having to bus Reg. Ed. students to BPS schools, which represented about 35 percent of the routes \cite{BPS (2013b)}. The geographically large zones required BPS to send buses to cover an area potentially as large as one-third of the city.

Low predictability of assignment: In 2011, the Boston Globe published a series of articles about the experiences of 13 families who had participated in the choice process \cite{Russell et al. (2011)}. One salient theme in a majority of these articles is that families were frustrated with their inability to predict whether their children would be admitted to one of their top choices. If their children were not admitted to one of their top-choice schools, some families would consider moving to suburban districts so that their children could attend schools in these districts; however, moving is a major decision for a family to make. Therefore, these families wanted to know the school-selection outcome for their children as early as possible. For some families, however,
the uncertainty was prolonged because they hoped their children would be admitted
to a school (i.e., kindergarten) in a later round. Some families were discontent; they
had exerted much effort into researching many options; however, their children were
not selected to attend any of their top-choice schools. This prompted some parents
and activists to label the system as an illusion of choice, because being able to rank
a school did not imply being selected to attend that school. Unpredictability was a
central reason that the Seattle school board changed from a choice lottery system to
a neighborhood school system in 2009, with school assignments determined primarily
by home location ([Lilly} {2009}; [Woodward} {2011}).

Community dispersion: [Ebbert and Russell} {2011} document 19 school-aged chil-
dren on one street in Boston who traveled to 15 different schools. They argue that
this resulted in a loss of a sense of community within a neighborhood, because “fam-
ilies are less likely to know one another when their children don’t attend schools
together.” Having schools with strong ties to the community was also important to
Boston, because it allowed the city to provide health, unemployment, and parental
education services through the school. Boston’s mayor, Thomas Menino, elaborated
on community dispersion under the three-zone plan in his 2012 State of the City
Address:

"But something stands in the way of taking our system to the next lev el:
a student assignment process that ships our kids to schools across our
city. Pick any street. A dozen children probably attend a dozen different
schools. Parents might not know each other; children might not play
together. They can’t carpool, or study for the same tests. We won’t have
the schools our kids deserve until we build school communities that serve
them well" ([Menino} {2012a}).

2.2.2 School Assignment Reform

Since 2003, Mayor Menino had made several attempts to reform the three-zone plan.
In December 2003, he appointed a student-assignment task force, which developed
seven assignment plans, each with smaller zones. The proposed plans had 4, 6, or 12 zones for elementary schools, each of which covered a smaller geographic area. As a result, fewer buses would be needed to cover each school, thus reducing transportation costs and community dispersion; however, previous city committees had not approved any of the zone reforms because of concerns about equity of access to quality schools (Landsmark et al. (2004)).

In April 2009, the mayor launched another initiative to review student assignments. BPS presented a five-zone plan; however, concerns over equity of access to quality schools (some zones had fewer high-quality schools) and unpopular proposed changes to some schools caused heated public opposition, which halted this attempt at reform.

In 2012, Mayor Menino again tried to reform the assignment plan. He appointed the EAC discussed above to develop a new plan. The EAC held community meetings to elicit families’ input, and requested BPS to provide data on supply and demand, school quality, student demographics, and choice patterns. In September 2012, BPS proposed new zone-based plans with 6, 9, 11, and 23 zones, and a plan that assigned each child to the closest school (Vaznis (2012)). Shortly after the publication of these plans, an outside analysis highlighted serious inequities in them in terms of access to quality schools, especially for children in the poorest neighborhoods (Burge (2012); Levinson et al. (2012)). Doubts arose about whether Boston would be able to reform the three-zone system this time, or would have to abort this reform attempt, as it had abandoned its 2004 and 2009 tries.

2.3 Problem Formulation

From February to November 2012, I attended many EAC meetings and engaged parents, community organizations, activists, BPS staff, EAC members, and other academics. These interactions helped me to iteratively refine my understanding of the key issues, and eventually formulate a precise quantitative problem.

As I understood it, the primary flaw of the three-zone plan was that the choice
menus provided families with too many options. For example, offering distant schools as choices meant that some children might have to be bused very long distances, contributing to the high transportation costs. Furthermore, having menus with a high number of options contributed to unpredictability because a child might be assigned to any of more than 25 schools; this also imposed a burden on families to research the many potential options. Moreover, the large number of options increased the likelihood that students from the same neighborhood would be scattered across the city.

Reducing the number of school options, however, might result in children living in inner-city neighborhoods having less access to schools that were perceived as higher quality, because the inner city was associated with lower-performing schools. In addition, because such neighborhoods also had a greater proportion of children of color, any policy change that might reduce access to quality schools for such students would increase racial inequity, which would be politically unacceptable.

Thus, the problem was how to allocate choice among various neighborhoods so that the total number of choices is not so high that it increases transportation costs and community dispersion, but gives students living near lower-performing schools enough choices to provide them with sufficient access to quality schools. We can view this as an optimization problem; the decision variables are the choice menus and priorities of students from various neighborhoods; the objective is a combination of equitable access to quality schools, predictability, proximity to home, community cohesion, racial and socioeconomic diversity in schools, and variety of choice. The challenge was how to quantify these concepts and how to relate the decision variables of choice menus and priorities to these outcomes.

The plan needed to be simple, within the capabilities of BPS to implement, and understandable to families. I learned to appreciate these constraints only after some trial and error. At a community meeting in October 2012, I proposed a solution that used a linear program to optimize for proximity to home and small choice menus, while guaranteeing equitable access to quality in a quantitatively precise way. Despite its theoretical appeal, this plan was perceived as too complex and too much of
a black box. BPS viewed it as too technically complex to implement. This is understandable; its objective of educating children does not normally require a technically sophisticated operations research team. Some city staff members told me that for any plan to be implementable, I should be able explain it to a fifth grader. Simpler solutions also tend to be more robust. A technically sophisticated solution may be optimal for a given model; however, in a real-world scenario, which often involves multiple and hard-to-quantify goals, human behavior and perception, and uncertain long-term ramifications, any model is at best a crude approximation. It may be more important for a solution to be simple and intuitive, thus allowing policy makers to take ownership of it and adapt it as necessary when new issues emerge, than for the solution to be optimal for a quantitative model.

I decided on the following approach to solve the problem:

1. Use previous choice data to fit and validate a demand model for how families would choose schools under new choice menus. (Because the student-proposing DA could not be influenced by any strategy, I assumed that the previous choice data reflected the true preferences of families.)

2. Precisely define outcome metrics and build a simulation engine that can forecast any system of choice menus and priorities.

3. Work with BPS to propose a set of plans that reduces the choice menus, protecting equity of access, and that BPS can implement. Use the simulation engine to explore variations and narrow the number of variations to a short list.

4. Present a detailed analysis of the plans to BPS and the EAC, showing differences by neighborhood, grade, student type, race, and socioeconomic status.

The final decision would still result in debates among various stakeholders to determine suitable trade-offs among various goals. As a result, the decision makers would more accurately and precisely understand the trade-offs, and would be able to discuss them in a more informed way.
2.4 Data and Analysis

The project used two data sources. The first was publicly available data posted on the BPS *Improving School Choice* website (http://bostonschoolchoice.org/), which BPS created as part of its public engagement for the 2012-2013 reform. The second was student-level data from BPS (BPS made the data anonymous, such that no student was identifiable). I received access to these data following an October 2012 meeting at which I presented to the EAC a rough analysis using aggregate data, and petitioned for access to student-level data to permit more refined modeling.

The publicly available data contained a table of school characteristics for each school, including its location, enrollment numbers as of December 2011, proportion of students of each race, English proficiency status and SPED status of each student, aggregate MCAS results for the previous two years, and the school’s BPS rank, which was the key quality metric used during the reform. BPS computed this rank by ranking its schools using a weighted average of student performance levels and performance growth (i.e., a metric that measures the improvement in a student’s test score relative to average improvement) in the previous two years’ MCAS tests; it gave performance a weight of 67 percent and growth a weight of 33 percent.

BPS stored location data using an internal geocoding system that partitioned the city into 868 small regions, which it called geocodes. The average area of a geocode was about 0.1 square miles. The geocode of a student’s home determined the student’s zone and the schools in the student’s walk zone.

The student-level data BPS provided were the Round 1 choice data for grades K0, K1, and K2 in 2010, 2011, and 2012. In each year and for each applicant, the data contained the student’s application grade (i.e., K0, K1, or K2), geocode, race, lunch status (a proxy for socioeconomic status), and first-10 school choices. For each choice, the data included the school code, the program type (e.g., Reg. Ed., ELL, SPED), and the student’s priority status relative to that school (e.g., continuing, sibling status, walk zone). For children whose families submitted rankings, the median number of choices was three, five, and four, for grades K0, K1, and K2, respectively. For those
who ranked more than 10 school options, the data showed only the first 10; however, because 94 percent of the applications ranked fewer than 10 options, this was not a major issue.

2.5 Demand Modeling

The first component of the simulation model is a demand model—a model of how families would choose schools when faced with new choices. Demand modeling uses previous choice data to correlate families’ observable characteristics with the choice rankings they submitted; this predicts how families of various neighborhoods, races, and socioeconomic backgrounds would choose schools, given a new set of options. The predictions are statistical in nature; that is, they do not predict the choices of specific individuals, but only the demand patterns of subpopulations in a neighborhood. Below, the data available for building the demand model are listed.

- Distance: the walking distance from the student’s home to the school, estimated using the Google Maps application programming interface and the centroid of the student’s and the school’s geocodes as proxies for exact locations.

- Free-lunch, reduced-lunch, full-fare status: variables indicating the student’s lunch status. Students received free lunches if their family income was below 130 percent of the Federal poverty guidelines. Students received reduced-price lunches if their family income was between 130 and 180 percent of the Federal poverty guidelines.

- Student race: variables indicating race (e.g., black, white, Asian, Hispanic, other).

- Sibling, walk, continuing status: variables indicating whether the student had a sibling at the school, was in the school’s walk zone, or would be a continuing student at the school.
• East Boston: variable indicating whether the student and the school are both in East Boston, which is geographically isolated from the rest of Boston by bridges and tunnels that are inconvenient to cross.

• Racial proportion: percentage of students at the school who are black, white, Asian, or other. Hispanic is not included to avoid multicollinearity.

Parag Pathak and I worked on the demand modeling. We chose to include several combinations of the above variables in the model. The trade-off in choosing the right variables was that if we used more variables, the model might capture more nuanced effects; however, it might not generate as precise an estimate, because it must estimate more parameters using the same number of data points. We tried several combinations and chose one that provided a good trade-off. We validated the model by fitting a version of the model using data from the previous year to predict choice patterns in the next year. Called out-of-sample validation, this ensures that we are not over-fitting. The appendix shows details of the analysis.

2.5.1 Simulation

I used the demand model to build a simulation engine that took as input any assignment plan (i.e., any system of choice menus and priorities); it then used the actual student and capacity data from 2012 to forecast the 2012 outcome had the plan been implemented. In the simulation engine, for each student, I simulated his (her) utility in each program using the demand model, then computed the student’s choice menu, and sorted the programs in the choice menu using the utilities, truncating to the top-10 schools according to the simulated utilities. These represented the student’s choice submissions. I then independently drew a lottery number for each student and simulated the DA algorithm based on the priorities associated with the plan. In each simulation round, I estimated the following metrics for each student:

• Access to quality schools: the student’s chance of receiving a lottery number that would enable that student to gain admittance to a school with an above-average BPS rank, and that the student would rank as a top-10 choice. This
implicitly assumed that only the student’s first 10 choices were acceptable to that student.

- Distance: the student’s walking distance to the assigned school.

- School-choice rank: the rank of the school choice obtained. For example, being admitted to one’s first-choice school corresponded to a choice rank of 1, and being admitted to one’s second-choice school corresponded to a rank of 2. This was undefined if the student was unassigned.

- Number of neighbors coassigned: the number of other students who lived within .5 miles of the student and who were assigned to the same grade and school. Distances between students’ homes were approximated by the Google Maps walk distance between the centroids of the geocodes of the homes.

- Access to top-three dream schools (defined as top-three choices without restricting the student’s menu): the student’s chance of receiving a lottery number that would enable that student to attend one of his (her) dream schools—the schools that student would have chosen if he (she) could have ranked any school in the BPS. Note that if the dream choices were not in the student’s menu, access would be zero.

- Percentage of peers of a specific race or lunch status: the percentage of other students of a specific race or lunch status who were assigned to the same program.

I averaged the above estimates over many independent rounds of simulation to compute an expected value for each student. In each simulation, the utilities and the lottery numbers were redrawn. For metrics that were only defined for students who were assigned, I computed the expected value, conditional on the student being assigned.

Using these student-level expected values, I computed the following aggregate statistics for the entire population.
• Minimum access to quality schools: over all students, the minimum of their expected access to quality schools. This corresponded to the welfare of the student with the worst access to quality schools.

• Median expected number of neighboring students coassigned.

• Minimum access to capacity: over all students, the minimum of their chances of being assigned to any school.

• Median access to a dream school.

• Average walk distance from student’s home to school.

• Median expected choice rank.

• Standard deviation across students of expected percentage of peers of each race or lunch status.

These metrics measured (1) equity of access to quality schools, (2) community cohesion, (3) supply and demand shortage, (4) availability of desirable-school choice, (5) proximity to home, (6) predictability of assignment, and (7) racial or socioeconomic segregation. For the first four metrics, a larger value is better; for the last three metrics, a smaller value is more desirable.

As a proxy for transportation costs, for each school, I computed the area a school bus would need to cover to pick up students; this was the difference between the area for which the school may appear in a student’s menu and the area of the school’s walk zone. I averaged this across all schools and called it the average bus coverage area.

2.6 Designing Simple Plans

For the reform, BPS decided to vary only the proposals for Reg. Ed. programs. For ELL and SPED programs, BPS committed to use a separate six-zone system, which it called ELL clusters and which it would apply on top of any assignment plan.
2.6.1 Other Zone-Based Plans

For the allocation of Reg. Ed. programs, BPS used the simulation engine to tweak its zone-based plan. After three iterations, it presented a 10-zone plan to the EAC. This plan was designed to balance the proportion of schools in each zone that BPS ranked in the top 50 percent. The EAC had previously decided to use these schools (i.e., highest 50 percent) as a proxy for quality schools. Based on EAC feedback, BPS also proposed a modified 11-zone plan. In both plans, a student’s menu consisted of the schools in his (her) zone and possibly additional schools within the one-mile walk zone.

2.6.2 The Home-Based Plan

I proposed two plans that did not use zones. In the first plan, which BPS called the home-based-A plan, each family could rank any school within each of the following sets:

- Any school within the one-mile walk zone
- The two closest schools ranked within the highest 25 percent (by BPS rank or another quality metric)
- The four closest schools ranked within the highest 50 percent
- The six closest schools ranked within the highest 75 percent
- The three closest capacity schools—schools to be designated by BPS in which excess capacity is available or in which capacity can be expanded cost effectively

The second plan, which BPS called the home-based-B plan, is similar to the home-based-A plan; however, the parameters two, four, and six (i.e., two, four, and six closest schools) were replaced by three, six, and nine.

I designed these plans based on the following criteria.

- Families should be able to rank all sufficiently close schools.
• Students living near lower-performing schools should be compensated with additional, higher-performing alternatives. Each student should have a sufficient number of options of various quality thresholds.

• Choice menus should include schools with sufficient capacity.

• Choice menus should be determined based on simple rules, which should allow the menus to adjust automatically when a school opens or closes, or its quality changes; otherwise, BPS would require another costly reform every few years.

The parameters for the home-based-A plan were chosen so that if quality were evenly distributed geographically and if capacity were widely available, the choice menus would essentially become the closest eight schools, in expectation that the choice menu would include two schools ranked in the highest 25 percent, four schools in the highest 50 percent, six schools in the highest 75 percent, and at least three schools within the set of the closest eight schools, which have excess capacity. However, quality schools were not yet distributed evenly; therefore, the menus were designed so that students who lived near the top schools would have much of their sets intersect, because a school in the top 25 percent school would simultaneously be in the top 50 percent and in the top 75 percent, and a student’s options would essentially include the closest six schools. However, a student whose closest schools did not have a good BPS ranking might have to travel farther than the closest six schools to find a school ranked in the highest 25 percent or 50 percent of schools; therefore, that student would be given additional, although more distant, options as compensation. Hence, to maintain equitable access to quality schools in the long term, the plans provided some guarantee of the quality of the schools that students could rank, regardless of the current geographic distribution of the schools. Provided that the chosen quality metric averaged a few years’ of data and did not fluctuate erratically, the adjustment in choice menu from year to year would also be smooth, because only a small subset of any family’s menu would change. This was in contrast to zone-based plans, in which any small change in zone boundaries would drastically affect those living on the boundary. Furthermore, the home-based menus varied smoothly across geographies; however, in
a zone-based plan, menus might change abruptly across zone boundaries, such that
neighboring students might have very different choice menus. The capacity schools
helped BPS meet supply and demand by pooling excess demand and directing them
toward local centers of supply (i.e. larger schools with greater capacities or greater
potential to expand).

This framework can be generalized to include the closest set of other types of
schools, including regional magnets, early-learning centers, or advanced-work classes,
if BPS were to decide to allow each family to rank such options.

Note that this plan guaranteed only the ability to rank certain schools, but did
not guarantee placement. A student’s chances of gaining admittance to a school
would depend on his (her) priority and the level of competition for this school. I
had proposed an earlier plan that guaranteed equity of probabilities of placement;
however, that plan required solving a linear program and significantly changing the
assignment process, and did not appeal to the EAC or BPS because of its complexity
and black-box nature.

2.7 Results

Using the simulation engine, I analyzed the short-listed plans selected by the BPS:
the status quo (3-zone), 10-zone, modified 11-zone, home-based-A, and home-based-B
plans. I used 25 rounds of independent simulations for each. For the same round of
simulation, I chose to use the same idiosyncratic taste shocks (i.e., random numbers
drawn in the utility model to capture unexplained preferences) and lottery numbers
across the plans, to ensure that any differences found would result primarily from
the different menus, and not from different random numbers. I presented the results
to BPS, who simplified them and presented them to the EAC. The most important
student pool for analysis was K2 noncontinuing students with no siblings enrolled in
a BPS school, because K2 was the main entry grade to elementary school, and most
seats were assigned to this category. Moreover, the assignment of continuing students
and siblings was essentially predetermined by the school at which the student was
continuing or the school that the student’s siblings attended, both of which were artifacts of the previous assignment system. Applying these filters, the relevant pool contained 1,659 students, 34 percent of whom were ELL students. Table 1 shows the aggregate simulation results for these students in each of the short-listed plans. BPS created the 10-zone and modified 11-zone plans, and I proposed the home-based plans. Although no plan completely dominated the others, the home-based-A plan performed reasonably well in equity of access to quality and proximity to home, which were the most important criteria. The simulations considered all students; however, the statistics reported were for this pool.

Table 2.1: The table shows simulated performance of the assignment plans that the EAC considered for its final vote in February 2013.

<table>
<thead>
<tr>
<th>Minimum Access to Quality</th>
<th>Status Quo</th>
<th>10 zone</th>
<th>modified 11 zone</th>
<th>Home Based A</th>
<th>Home Based B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 33% Schools</td>
<td>14.6%</td>
<td>0.7%</td>
<td>0.9%</td>
<td>12.7%</td>
<td>13.2%</td>
</tr>
<tr>
<td>Top 50% Schools</td>
<td>19.5%</td>
<td>22.7%</td>
<td>22.7%</td>
<td>22.5%</td>
<td>27.8%</td>
</tr>
<tr>
<td>Top 67% Schools</td>
<td>31.4%</td>
<td>22.7%</td>
<td>22.7%</td>
<td>32.9%</td>
<td>32.9%</td>
</tr>
<tr>
<td>Median # of Neighbors Co-Assigned</td>
<td>3.12</td>
<td>3.84</td>
<td>4.20</td>
<td>3.92</td>
<td>3.80</td>
</tr>
<tr>
<td>Minimum Access to Capacity</td>
<td>47.2%</td>
<td>31.8%</td>
<td>31.9%</td>
<td>36.0%</td>
<td>35.6%</td>
</tr>
<tr>
<td>Median Access to Top 3 Dream</td>
<td>42.4%</td>
<td>32.0%</td>
<td>31.4%</td>
<td>31.0%</td>
<td>32.1%</td>
</tr>
<tr>
<td>Average Distance (miles)</td>
<td>2.03</td>
<td>1.24</td>
<td>1.19</td>
<td>1.25</td>
<td>1.30</td>
</tr>
<tr>
<td>Median Rank Obtained</td>
<td>2.83</td>
<td>2.65</td>
<td>2.62</td>
<td>2.78</td>
<td>2.91</td>
</tr>
<tr>
<td>St. Dev. Across Students in</td>
<td>% Peers Free Lunch</td>
<td>9.0</td>
<td>11.5</td>
<td>11.5</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>% Peers Black</td>
<td>14.1</td>
<td>15.4</td>
<td>15.5</td>
<td>15.5</td>
</tr>
<tr>
<td></td>
<td>% Peers White</td>
<td>9.4</td>
<td>12.2</td>
<td>12.2</td>
<td>11.2</td>
</tr>
<tr>
<td></td>
<td>% Peers Hispanic</td>
<td>17.7</td>
<td>19.2</td>
<td>19.3</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>% Peers Asian</td>
<td>9.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.3</td>
</tr>
</tbody>
</table>

The metrics that the EAC considered to be most important were minimum access to quality schools (i.e., to schools ranked in the 50th percentile), and average distance (i.e., walking distance from home to school). As Table 1 shows, in both of these metrics, all the new plans were better than the status quo. Furthermore, the new plans reduced the bus coverage area by a factor of three to four, and increased community cohesion by 20 to 34 percent, as measured by the number of neighboring children (i.e., children living within .5 mile) in the same grade and coassigned to the same
school. The new plans, however, because of the smaller menus, were projected to increase supply challenges (because restricting choice menus reduced BPS’ flexibility in assigning children), to reduce the variety of choice (because median access to dream-school choices was lower), and to decrease racial and socioeconomic diversity within classrooms. Nevertheless, BPS and the EAC considered the overall trade-offs to be positive: the supply challenges could be overcome by adding seats in constrained regions, and the reductions in choice and in diversity were reasonably mild. In terms of predictability, as measured by median rank obtained, the new zone-based plans and the home-based-A plan performed better than the status quo plan, and the home-based-B plan was worse; however, the differences were fairly minor.

The primary differences in the new plans were as follows: the home-based-A and home-based-B plans significantly outperformed the 10-zone and 11-zone plans based on minimum access to quality (i.e., top one-third and top two-thirds) schools. Although the zone-based plans were optimized for the one threshold (i.e., highest 50 percent), the home-based plans considered a range of thresholds. This illustrated the brittleness of the zone-based plans in maintaining equity: they only worked for the one metric for which they were optimized. If quality changed significantly in the future, these plans could not guarantee that equity would be maintained without modifying the zone boundaries, which might require another reform.

The initial report contained simulation results broken down by grade, race, neighborhood, lunch status, and English-learner status. It also included sensitivity analysis on the percentage of seats at each school for which to apply walk-zone priorities (Pathak and Shi [2013]). After the simulation results were made public, the EAC met several times to deliberate the various trade-offs, and held several community-engagement meetings to elicit feedback from the public. Parent groups, community organizations, activist groups, and local and state politicians participated in these discussions and voiced comments.

On February 25, 2013, the EAC met to take its the final vote on the five plans shown in Table 1. The EAC overwhelmingly favored the home-based-A plan: of 23 members present for the vote, 20 voted for this plan, 1 voted for the home-based-
B plan, and 2 did not take a position by voting present. According to the EAC’s recommendation letter to the superintendent of the BPS, it choose this plan because it “ensures every family has high-quality schools on its list of options,” and that “it adapts to changes in quality over time” (EAC [2013], p. 4). In a follow-up statement by the BPS superintendent, additional reasons behind the EAC decision included increased predictability, increased equity of access, significantly lower travel distances, more opportunities for families to understand their choices (because of the reduced menus), greater community cohesion, elimination of the zone system, and the association of choices with school quality (Johnson [2013]).

On March 13, 2013, the Boston school committee approved the home-based-A plan for implementation for the 2014 assignment cycle (Seelye [2013]; Vaznis [2013]). The implementation involved minor changes to the plan. The final plan implemented is in Section 2.A.

2.8 Discussion

Following the EAC’s approval of this plan, BPS made several modifications. First, it removed the walk-zone priority, because the schools on the menus were already close to the students’ homes. Second, it changed the capacity schools to be essentially the lowest 25 percent of the schools based on BPS rank, because BPS predicted these schools would have excess capacity. Third, it expanded the menus to include the closest early-learning center (i.e., full-day kindergartens) and the closest schools with advanced-work classes. In addition, it made available several more schools citywide or for specific neighborhoods. This illustrated the value of having a sufficiently simple and flexible plan that the host organization could own and adapt to its needs.

This project highlighted the power of quantitative analysis and simulation to move debates forward by making the objects of discussion precise and by quantifying trade-offs. For example, prior to this analysis, participants in the reform vigorously debated the pros and cons of having a greater percentage of seats with walk-zone priority to help students attend schools closer to home, or a smaller percentage to grant
greater equity of access. Both were important objectives; therefore, such debates often evolved to gridlock situations. Perhaps surprisingly, the simulation results showed that the walk-zone priority, as it was implemented in the three-zone plan, did not make a significant difference for the majority of schools, because a greater proportion of families applied from within the walk zone and filled up the 50 percent of seats set aside for walk-zone applicants. After seeing these results, the participants in these discussions were able to bypass this false dichotomy and perceive the actual trade-offs.

One important question this project unfortunately could not address was the long-term impacts of the reform. The demand model was based on choice data collected before the reform; this data might not accurately reflect how people would choose schools following the reform. For example, the reform itself or the new plan’s presentation might significantly alter the behavior of families. Parag and I analyzed the validity of using data collected prior to the reform to build a demand model in a follow-up project ([Pathak and Shi (2014)]), in which we make predictions using these data and the models fitted, and evaluate the prediction accuracy using data collected after the reform. Other issues not yet addressed include complex interactions between the assignment plan and qualities of schools in the future, as well as how the assignment plan may affect student learning and performance, which one might argue are the most important results. A conclusive study of the long-term impacts of the reform is possible only after several years of postreform data become available.

During this project, I constrained the solution to a specific level of simplicity; in the future, increased complexity might be possible if societal tolerance for complexity increases and computer technologies can shield end users from some complexities. For example, BPS chose to implement the home-based-A plan by directing families to a Web applet; the user entered an address and the applet showed a map that includes the choice menu. The families did not need to directly construct their menus using the home-based-A plan rules. Hence, provided that families trust this black box, we may hope to implement a plan that is more sophisticated and better optimized using the same method of presentation. Two follow-up papers explore these possibilities.
Ashlagi and Shi (2014) explore an optimization-based lottery-correlation procedure to improve community cohesion, and Ashlagi and Shi (2015) solve a convex optimization problem to obtain the optimal choice menus and priorities to maximize a weighted combination of efficiency and equity, while staying within an expected busing budget.

The modeling and simulation techniques in this chapter and the optimization procedures in these follow-up papers can be applied to school choice in other cities and to other areas of public service allocation; examples include allocating subsidized housing, high-demand college courses, and office space.

2.A The Home-Based plan Implemented in 2014

The Home-Based plan implemented in Boston in 2014 is precisely defined as follows. The menus are based on a partition of schools into four tiers based on standardized test-scores, with Tier 1 being the best. The menu of each regular education student is the union of:

- any school within 1 mile straight line distance;
- the closest 2 Tier 1 schools;
- the closest 4 Tier 1 or 2 schools;
- the closest 6 Tier 1, 2 or 3 schools;
- the closest school with Advanced Work Class (AWC);
- the closest Early Learning Center (ELC)[1]
- the 3 closest capacity schools[2]
- the 3 city-wide schools, which are available to everyone in the city.

---

1ELCs are extended-day kindergartens.
2Capacity schools are those which BPS has committed to expanding capacity as needed to accommodate all students. In the 2014 implementation of the Home Based Plan, for elementary schools, the capacity schools are exactly the Tier 4 schools.
Furthermore, for students living in parts of Roxbury, Dorchester, and Mission Hill, their menu includes the Jackson/Mann school in Allston/Brighton.

For students without older siblings in the system or special needs, the priorities were: 1) students living in East Boston had priority for East Boston schools; students outside of East Boston had priority for non-East Boston schools. 2) to break ties, each student $i$ is given an i.i.d. random number $\delta_i$. This version of the plan implemented no longer has any priority for living within the walk-zone (one-mile radius) of the school.

2.B Estimating Multinomial Logit

We modeled the family choice process using a multinomial logit discrete choice model with school and program-type fixed effects. Let $j$ be a program, with $j = (s, p)$, where $s$ denotes the school and $p$ the program type. We assumed that a student’s choice submission was driven by unobserved utilities of the form,

$$u_{ij} = q_j + \vec{\beta} \cdot \vec{z}_{ij} + \epsilon_{ij},$$

where $u_{ij}$ represents student $i$’s utility to program $j$, $q_j$ represents the quality or desirability of program $j$, $\vec{z}_{ij}$ is a vector of observed data for the student-program pair, $\vec{\beta}$ represents how these features impact the decision, and $\epsilon_{ij}$ represents idiosyncratic preference for the program and is assumed to be independently distributed as a standard Gumbel distribution; the reason for this distribution was to allow tractable estimation. We also assumed the quality term to be additively decomposable as a term for the school and a term for the program type, $q_j = a_s + b_p$. We called $\vec{a}$ the school fixed effects and $\vec{b}$ the program-type fixed effects. We estimated the parameters $\vec{a}$, $\vec{b}$, $\vec{\beta}$ using maximum likelihood. Beggs et al. (1981) and Hausman and Ruud (1987) introduced multinomial logit models for ranked-order data, and the estimation procedure is now standard. For more details on the estimation, see the appendix in Pathak and Shi (2013).
Choosing which features to include in the model represents a trade-off between model flexibility and estimation precision. A model with more parameters may be able to capture more nuanced effects; however, estimating more parameters using the same amount of data entails that each parameter may not be estimated with the same precision. We experimented with several models and used out-of-sample testing to determine the final model. We tried four sets of parameters, and we called the sets Simple, Simple2, Medium, and Medium2. The lists below show the features we included in each of the four sets. In the Demand Modeling section, we include an explanation of these variables.

- **Simple**: distance, sibling, walk, continuing.
- **Simple2**: all features in Simple plus the square root of distance.
- **Medium**: all features in Simple2, plus pairwise product of the student’s race and the school’s racial demographics, plus the pairwise product of the student’s lunch status and the school’s percentage of students who receive free lunches.
- **Medium2**: all features in Medium2, plus the East Boston indicator (i.e., whether both the student and the school are in East Boston), plus the product of distance and the student’s lunch status, plus the product of the square root of distance and the student’s lunch status.

Note that the models are nested. Moving from Simple to Simple2, we allow for nonlinear preferences for distance. The square root is to capture the intuition that although a school near a student’s home may be more desirable than a school one mile away, a school that is four miles away may not seem much better than a school that is five miles away, because the child will be on a bus for a long time in either case. Moving from Simple2 to Medium, we allow students of different races and socioeconomic status to have different preferences for aspects of the school correlated with the school’s racial or socioeconomic composition. Such aspects may include school climate, safety, and culture. Moving from Medium to Medium2, we allow differential preferences for distances for students of different socioeconomic status,
as well as potential preferences for East Boston students to not have to cross the frequently congested bridges or tunnels to attend school.

Because the school characteristics were for December 2011, we used only the choice data in the years 2011 and 2012. We designated 2011 as the in-sample data set and 2012 as the out-of-sample data set. We considered only grades K1 and K2, because most school programs started in K1 and only a few seats were allocated in K0. To select the model, we fit the models using 2011 data, and considered the Bayesian information criteria (BIC), which takes into account the goodness of fit and penalizes having too many parameters. In our case, this criteria persuaded us to select the model with more parameters if the incremental gain in log likelihood divided by the number of additional variables was more than \( \frac{\ln \# \text{ of observations}}{2} = \frac{\ln 20533}{2} = 4.96 \). Moving from Simple to Simple2 and from Simple2 to Medium, the gains per variable were 144 and 24, respectively, which significantly exceeded the threshold of 4.96; however, moving from Medium to Medium2, the gain was 7.1, which was greater than the threshold, but not by much. Furthermore, many of the additional variables in Medium2 turned out to be statistically insignificant. We ultimately chose to be conservative in the number of parameters to include; therefore, we chose the Medium model. We also fit this model using the out-of-sample data, to test for stability of coefficients. Table 2 shows the coefficient estimates of all four models using in-sample data, as well as the chosen model using out-of-sample data. Each model uses a different set of variables. The numbers without parentheses represent the impact of each variable and the numbers within parentheses are the standard errors (or uncertainty). The absolute value of the impacts are hard to interpret in isolation; however, greater magnitude implies greater relative importance. We found the Medium model to be superior and used it for out-of-sample testing.

To validate the Medium model, we compared the forecasted relative market shares of programs as predicted by the demand model fitted using the 2011 data with the actual market shares in 2012. We did this for the highest one, three, five, and seven choices; having several cutoffs gave us a picture of the goodness of our prediction for the entire choice list. The relative top \( k \) market share of a program is defined
Table 2.2: The table shows estimated coefficients of demand models used to predict how families living in different areas would choose between different schools. Each model uses a different set of variables. The numbers without parenthesis represent the impact of each variable and the numbers in parenthesis are the standard errors (or uncertainty). The absolute value of the impacts are hard to interpret in isolation; however, greater magnitude implies greater relative importance. We found the Medium model to be superior and used it for out-of-sample testing.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Simple</th>
<th>Simple2</th>
<th>Medium</th>
<th>Medium2</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>-0.481</td>
<td>-0.139</td>
<td>-0.152</td>
<td>-0.169</td>
<td>-0.115</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.037)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Distance x free-lunch</td>
<td>0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance x reduced-lunch</td>
<td>0.159</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.107)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance x full-fare</td>
<td>-0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sqrt(distance)</td>
<td>-1.198</td>
<td>-1.115</td>
<td>-1.140</td>
<td>-1.082</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.069)</td>
<td>(0.111)</td>
<td>(0.057)</td>
<td></td>
</tr>
<tr>
<td>Sqrt(distance) x free-lunch</td>
<td>0.066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.145)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sqrt(distance) x reduced-lunch</td>
<td>-0.270</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sqrt(distance) x full-fare</td>
<td>-0.099</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.217)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sibling</td>
<td>2.984</td>
<td>2.958</td>
<td>2.888</td>
<td>2.891</td>
<td>2.860</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>Walk</td>
<td>0.350</td>
<td>0.170</td>
<td>0.154</td>
<td>0.161</td>
<td>0.262</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Continuing</td>
<td>5.945</td>
<td>5.938</td>
<td>5.884</td>
<td>5.896</td>
<td>5.261</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.085)</td>
<td>(0.085)</td>
<td>(0.085)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>East-Boston</td>
<td></td>
<td>0.293</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.142)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School % race x student race</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>School % free lunch x student lunch status</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>
as the proportion of the students’ top $k$ choices for this program. Figure 2-2 shows the results. In the plots, each point corresponds to a program in K1 or K2. Its $x$-value corresponds to the predicted market share, and its $y$-value to the actual market share. If predictions were perfect, then all points would lie on the 45 degree line. As the plot shows, although the predictions were not perfect, the majority of points lie close to the 45 degree line, indicating that the predictions were reasonable. Moreover, the predictions were best for the first choices, and less accurate for other choices. An explanation of this is that a multinomial logit model implies a certain type of substitution patterns as we go down the choice list; however, the actual data might not satisfy these patterns. In the literature, this substitution pattern is called independence of irrelevant alternatives (IIA); see McFadden et al. (1977).

Our demand model does not contain outside options, because we had little data on families’ preferences for nonBPS alternatives, such as private, charter, parochial, or other schools. We decided to simply assume that students all ranked up to 10 options, which was the maximum number contained in our input data. We did sensitivity analyses on the number of ranked options, and found that 10 was a reasonable choice. The details are in the appendix of Pathak and Shi (2013).
Figure 2-2: These plots compare the relative market shares of programs as predicted by the demand model fitted using 2011 data with the actual market shares in 2012. Although the predictions were not perfect, they were reasonably good, because most of the dots are near the 45 degree line.

(a) First choice
(b) Highest three choices
(c) Highest five choices
(d) Highest seven choices
Chapter 3

Validating the Predictive Model Across a Policy Change

3.1 Introduction

As shown in Chapter 2, the multinomial logit (MNL) model for choice modelling seems to reasonably describe choice patterns of families. In using it to compare various proposed assignment plans, one assumption is that the policy change of assignment plan would not affect the underlying choice behavior. However, this may not hold, because along with the adoption of the Home-Based plan (see Section 2.A), the presentation of school options also changed. For example, the differentiation of schools into four tiers based on test scores is now more salient. If changing the assignment plan were to drastically affect how families rank schools, then no predictive model using based on past data can predict the preferences of families under a new assignment plan.

In this chapter, we evaluate the predictive accuracy of the multinomial logit (MNL) model across the policy change from the 3-Zone plan to the Home-Based Plan. We do this by using the MNL model estimated using data before the policy change to predict the outcomes of the reform, and comparing with the actual outcome based on the actual choices of families after the reform.

To quantify the error, we consider a variety of metrics. One set of metrics is the prediction error in various moments of interests, such as the number of unassigned
students in each neighborhood, the access to Tier 1 or 2 schools in each neighborhood, and the average distance students travel in each neighborhood. We also consider prediction errors in the choice rankings themselves, by comparing the distribution of Top 1, Top 2 and Top 3 choices in various neighborhoods.

We also compare the logit model with two alternative choice models. The first is a hierarchical model which assumes that every student follows a fixed rule in ranking schools. For example, for regular education students who are not continuing from a previous grade and do not have siblings in any school, this model assumes that they will rank schools first by tier, then by distance. The second is a mixed logit model, which allows different students to have different coefficients in the logit model. This is also called a random coefficient model in the literature.

We find that the raw predictions of logit and mixed logit models do not clearly outperform the hierarchical model, especially for grade K2, in which the hierarchical model actually performs the best. However, we show that this is largely due to the prediction errors in what we call the environment, which is the number of students who apply from each neighborhood and the demographics of students. We show that when one controls for the environment, the logit and mixed logit perform much better than the hierarchical model. Moreover, the logit and mixed logit models perform similarly well. This validates the MNL model as a reasonable way of predicting preferences in the Boston school choice context.

### 3.2 Data

The main data sources for this project are BPS Round 1 choice data and enrollment data for 2010-2013. For each year, the Round 1 choice data was collected in January or February of that calendar year, for application to the school year that began in September of the same calendar year. The enrollment data is a snapshot taken in December of the same year, 11 months after the choice data and three months after the school year began.

The choice data contains for each student who participated in round 1 his/her
student identification number; grade; English proficiency status and first language; special education or disability status; geocode (a geographic partitioning of the city used by BPS); school program to which the student has guaranteed priority (designation for continuing students); student identification numbers of the student’s siblings currently enrolled in BPS; lottery number; first 10 choices and priorities to each; school program to which the student was assigned and the priority under which he/she was assigned. Using the assigned school and program codes, we infer the capacity available for Round 1 assignment for each school program. We assume that this reflects true Round 1 capacities.

The enrollment data, which covers the vast majority of the students in the Round 1 choice data contains student identification numbers; enrolled school and program; grade; geocode; address; gender; race; languages spoken at home; dates of entrance and, if applicable, withdrawal from BPS; food service code (whether the student’s socio-economic situation qualifies for free or reduced lunch). (Since BPS began offering free lunch to all students after September 2013, the food service code will not be available in the future.)

In addition, we have access to a data set of school characteristics for each of the four years. The school dataset includes for each year and each school the school code, address, school type, % of students of each race, % of students who are English Language Learners (ELL), % of students who have Special Education (SPED) requirements, and % of students who scored Advanced or Proficient in grades 3, 4 and 5 for MCAS math and English in the previous year.

Since the assignment reform is mainly for elementary school assignment, we focus on the entry grades kindergarten 0 to 2 (K0, K1, K2). K2 is the main entry grade to elementary schools in Boston. Table 3.1 shows the total number of Round 1 applicants in each of the kindergarten grades, as well as the total Round 1 capacity. As can be seen, only a small fraction of seats are available in grade K0, while more than half the seats are available in K1.

Students who are assigned to K0 or K1 in the previous grade enter the assignment system the next year as continuing students. These have priority to their current
Table 3.1: Aggregate supply and demand in grades K0-2, in years 2010-2013.

<table>
<thead>
<tr>
<th>Year</th>
<th>K0</th>
<th>K1</th>
<th>K2</th>
<th>K0</th>
<th>K1</th>
<th>K2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>803</td>
<td>2134</td>
<td>3473</td>
<td>148</td>
<td>1676</td>
<td>3139</td>
</tr>
<tr>
<td>2011</td>
<td>704</td>
<td>2202</td>
<td>3556</td>
<td>170</td>
<td>1689</td>
<td>3328</td>
</tr>
<tr>
<td>2012</td>
<td>1001</td>
<td>2660</td>
<td>3985</td>
<td>181</td>
<td>1921</td>
<td>3689</td>
</tr>
<tr>
<td>2013</td>
<td>913</td>
<td>2599</td>
<td>4038</td>
<td>155</td>
<td>1890</td>
<td>3979</td>
</tr>
</tbody>
</table>

seat over new students. We define every non-continuing student as “new.” Figure 3-1 plots the total number of new and continuing applicants to BPS for four years.

![Figure 3-1: Number of new and continuing applicants to BPS (K0-2)](image)

To make use of geocode data, we take the latitude and longitude centroids of each geocode (provided by BPS) and compute a mapping from geocodes to “neighborhoods,” which gives a geographic partition of Boston into 14 regions. Table 3.2 shows the neighborhood breakdown of the Round 1 applicants and demographic profiles for each neighborhood, including an estimate of household income using the median household income from the 2010 Census of the census block group where the centroid of the student’s geocode lies, the percentage of applicants who are English Language Learners (ELL), and the racial composition of the applicant pool from this

---

1 Originally BPS has 16 neighborhoods, but we combined three small contiguous neighborhoods Central Boston, Back Bay, and Fenway/Kenmore into one neighborhood that we call “Downtown.” This is because these neighborhoods approximately make up the Boston downtown, and they each have very few number of applicants. Combining them still yields one of the smallest neighborhoods by number of applicants.

2 We used the 2012 ESRI demographics data set, which is available at [http://www.esri.com/data/esri_data/demographic-overview/demographic](http://www.esri.com/data/esri_data/demographic-overview/demographic).
neighborhood. This table aggregates all four years of Round 1 choice data for all 3 kindergarten grades.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>% Total</th>
<th>Income Est. [K]</th>
<th>ELL %</th>
<th>Black %</th>
<th>Hispanic %</th>
<th>White %</th>
<th>Asian %</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>5%</td>
<td>53.9</td>
<td>51%</td>
<td>8%</td>
<td>30%</td>
<td>28%</td>
<td>22%</td>
<td>5%</td>
</tr>
<tr>
<td>Charlestown</td>
<td>2%</td>
<td>60.7</td>
<td>24%</td>
<td>12%</td>
<td>27%</td>
<td>49%</td>
<td>10%</td>
<td>2%</td>
</tr>
<tr>
<td>Downtown</td>
<td>3%</td>
<td>58.6</td>
<td>41%</td>
<td>10%</td>
<td>18%</td>
<td>36%</td>
<td>30%</td>
<td>5%</td>
</tr>
<tr>
<td>East Boston</td>
<td>14%</td>
<td>33.8</td>
<td>76%</td>
<td>2%</td>
<td>77%</td>
<td>16%</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>6%</td>
<td>53.4</td>
<td>28%</td>
<td>40%</td>
<td>42%</td>
<td>13%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>6%</td>
<td>47.6</td>
<td>31%</td>
<td>13%</td>
<td>44%</td>
<td>30%</td>
<td>5%</td>
<td>8%</td>
</tr>
<tr>
<td>Mattapan</td>
<td>7%</td>
<td>36.0</td>
<td>30%</td>
<td>56%</td>
<td>40%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>6%</td>
<td>40.5</td>
<td>49%</td>
<td>28%</td>
<td>40%</td>
<td>11%</td>
<td>17%</td>
<td>4%</td>
</tr>
<tr>
<td>Roslindale</td>
<td>8%</td>
<td>54.9</td>
<td>28%</td>
<td>17%</td>
<td>42%</td>
<td>34%</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>Roxbury</td>
<td>14%</td>
<td>29.6</td>
<td>36%</td>
<td>45%</td>
<td>49%</td>
<td>2%</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>South Boston</td>
<td>3%</td>
<td>37.8</td>
<td>36%</td>
<td>16%</td>
<td>38%</td>
<td>36%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>14%</td>
<td>44.2</td>
<td>35%</td>
<td>33%</td>
<td>35%</td>
<td>13%</td>
<td>16%</td>
<td>3%</td>
</tr>
<tr>
<td>South End</td>
<td>4%</td>
<td>47.0</td>
<td>38%</td>
<td>29%</td>
<td>40%</td>
<td>15%</td>
<td>11%</td>
<td>4%</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>7%</td>
<td>65.6</td>
<td>21%</td>
<td>16%</td>
<td>27%</td>
<td>47%</td>
<td>6%</td>
<td>4%</td>
</tr>
</tbody>
</table>

| All Neighborhoods | 100%    | 44.3            | 40%   | 20%     | 44%        | 19%     | 8%      | 3%    |

Our analysis also uses distance estimates between each student’s home and each school. To account for geographic barriers and road availabilities, we use walking distances provided by Google Maps API. For students for whom we cannot find a valid address by matching to the enrollment data, we used the centroid of the student’s geocode as a proxy for home location.

### 3.3 Methodology

The aim of our analysis is to explore alternative approaches of forecasting outcomes and to evaluate the accuracy of each approach. We target outcomes that are important to BPS operations and decision-making.

We forecast three assignment outcomes. The first is the number of unassigned students per neighborhood. This outcome is important for BPS because they have publicly committed to assign each K2 student to a set within his/her Home Based choice menu. Two other assignment outcomes are average access to quality, as defined by students’ chances of getting into a Tier 1 or 2 school, and average distance to assigned school for each neighborhood. These were the two most important metrics

---

3 This occurs when the student is either not found in the enrollment data or the spelling of the address does not yield a valid result on Google maps, or the result ended up being more than 0.5 mile from the centroid of the applicant’s geocode, indicating a data error.
by which the city committee during the 2012-2013 school assignment reform made their decisions, and the numbers they examined were based on forecasts arising from demand modeling. This analysis examines the accuracy of such an approach.

We also forecast market shares, as more direct measures of the choice patterns themselves, apart from interactions with the assignment system. We examine school market shares for each neighborhood, for top 1, top 2, and top 3 choices. This represents the demand and substitution patterns of families’ choices across different neighborhoods. Because most of the available data is for K1 and K2, and because these grades are more important for BPS strategic and operational policies, we focus on these two grades in the analysis.

Given the actual choice data and lottery numbers, and given a table of school program capacities, the above moments can be computed deterministically. Program capacities are control variables that BPS often varies over the assignment cycle. Our simulation engine, which can be seen as a function mapping program capacities to outcome forecasts. To commit to specific forecasts, it is necessary to specify program capacities. For simplicity, we use Round 1 inferred capacities from the previous year.

Forecasting the above moments involves forecasting the application pool in 2014, how applicants choose schools, and simulating the BPS assignment algorithm to yield final outcomes. We describe these steps in detail in the following subsections.

3.3.1 Demand Models

The focus of our study is alternative approaches to predict families’ demand. While a full demand model should include families’ decisions to apply to BPS, we do not have sufficient data about each family’s outside options to precisely estimate such a model, so in this project we choose to focus on choices among BPS alternatives, and forecast the application pool using an ad-hoc approach described in Section 3.3.2.

We consider three types of demand models, which are ordered in increasing complexity.

- **Hierarchical Model:** Assume that all families choose based on a simple rule that
agrees with intuition and naturally arises from how the new assignment system has been presented.

- **Multinomial Logit Model**: This model is one of the simplest and most widely used approaches in demand modeling, especially for industrial organization applications. The 2012-2013 Boston school assignment reform heavily leaned on an analysis based on such a model, described in Pathak and Shi (2013).

- **Mixed-Logit Model**: This model is a popular alternative to the multinomial logit model due to its greater flexibility in capturing complex substitution patterns that violate the Independence of Irrelevant Alternatives (IIA) property of pure logit models. One theoretical justification of such a model is that any Random Utility Maximization (RUM) consistent demand model can be approximated to arbitrary accuracy by a mixed-logit model McFadden et al. (2000).

We now describe the details and specifications of each model.

**Hierarchical Model**

There are many possibilities for specifying statistical models of demand, which are not based on the random utility framework. We chose one particular model so that our investigation of more sophisticated models can be compared to an alternative. For instance, Nevo and Whinston (2010) point out that when evaluating the performance of simulation based merger analysis, it is important to compare these methods to other possibilities. Even though it necessarily requires some ad hoc choices, our Hierarchical benchmark represents such an alternative.

From interactions with parents and BPS staff, we learned that many people expect families to simply choose schools with the best Tier schools first, breaking ties with distance. For ELL students, BPS staff stated that if given sufficient information, families would place a premium on ELL programs since they offer targeted programming, especially language-specific ELL programs in their home language. Other patterns are suggested by the choice data. For example, the vast majority of continuing students (91%) select the next grade level of their current program first, an expected
pattern since families may not like having to move schools. Furthermore, for students who have a sibling currently attending BPS, 66% of them rank a school first that a sibling goes to, an expected pattern if families value having siblings attend the same school for transportation or other reasons.

If families’ choices are not based on coherent and stable preferences, but are strongly influenced by BPS’ publicity or framing efforts, the following simple model may adequately approximate families’ choices. Each family ranks the school programs in their personalized menu based on the following hierarchy:

<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(most important) present school program</td>
</tr>
<tr>
<td>2</td>
<td>another program in current school</td>
</tr>
<tr>
<td>3</td>
<td>school where sibling attends</td>
</tr>
<tr>
<td>4</td>
<td>(for ELL students) ELL program</td>
</tr>
<tr>
<td>5</td>
<td>(for ELL students) ELL program in home language</td>
</tr>
<tr>
<td>6</td>
<td>better Tier school</td>
</tr>
<tr>
<td>7</td>
<td>closer walking distance</td>
</tr>
</tbody>
</table>

Students only consider the hierarchy that pertain to them. For example, new applicants do not consider hierarchies 1 or 2, and non-ELL students do not consider hierarchies 4 and 5.

Outside of a random utility framework, a model of this type is a natural choice: for example, such a hierarchical model was used by the independent consulting group WXY, when commissioned by the BPS to analyze various counterfactuals.

**Multinomial Logit**

This model assumes that rankings of each student $i$ are induced by underlying utilities for each school program $j$, and that the utilities can be approximated by the following model:

$$u_{ij} = \beta \cdot F_{ij} + \epsilon_{ij},$$
where $F_{ij}$ is a vector of observable characteristics pertaining to student $i$ and choice option $j$, $\epsilon_{ij}$’s are iid random variables following a standard Gumbel distribution, and $\beta$ is a vector of parameters to be fitted. As usual, we standardize the error term without loss of generality since the model is invariant under multiplication or addition by constants. By the same reason we normalize one of the components of $\beta$ to zero.

A key implication of this model is that choices follow the Independence of Irrelevant Alternatives (IIA) property: the relative market shares of two programs does not depend on whether a third option is available. This means that substitution between programs follows the same proportional pattern across all choices. Although this property may be unrealistic for school choice, as two choices made by the same family may be correlated due to common, unobservable characteristics, it is plausible that the model may nevertheless provide a reasonable forecast of the moments that matter for decision making.

We fit this model by Maximum Likelihood Estimation (MLE) and obtain the covariance matrix of estimated coefficients by taking the inverse of the Hessian of the log likelihood function at the maximum. Table 3.3 shows the estimated coefficients for various specifications, using each of the 2012 and 2013 Round 1 choice data for grades K1 and K2. There are three specifications Simple (which does not use students’ demographics), Full (which fully interacts students’ race and income estimates with several key school characteristics), and Reduced (which removes the insignificant terms in Full and combines terms for efficiency). The features used are the following:

- distance: walk distance from home to school.

- continuing: indicator for whether the student has guaranteed status for the school program.

- sibling: indicator for whether student has sibling at the school.

- ell match: indicator for the student being English Language Learner (ELL) and the program being specialized for ELL.
• ell language match: indicator for the student being ELL and the program having a language-specific ELL program in the student’s first language.

• walk zone: indicator for whether student lives in the school’s walk-zone, which is approximately a 1-mile circle around the school.\footnote{The one mile circle is only approximate because the walk zones in our data are defined by drawing a one-mile circle from the school and including all geocodes that intersect the circle. A family from a geocode on the circle’s boundary may be a little further than one-mile from the school, but still in the walk zone.}

• black/hispanic: indicator for whether the student is black or hispanic.

• mcas: the proportion of students at the school who scored Advanced or Proficient in the previous year’s MCAS standardized test for math, averaging the proportions for grades 3, 4 and 5. (The MCAS test begins at grade 3. Grade 5 is the highest grade in many elementary schools. We only choose math because it is highly correlated with English.\footnote{This correlation is about .84 in years 2012 and 2013.})

• % white/asian: the proportion of students at the school who are white or asian.

In each model, we include a fixed effect for each school, which captures any common propensities to choose a school, due to perceived school quality, facilities, and other unobserved characteristics. In the Full specification, we interact the student’s income estimate, along with indicators for race (black, asian, hispanic, other, or unknown), with the school’s mcas, % white/asian and distance. (This represents $6 \times 3 = 18$ terms).

All of the specifications in Table 3.3 yield a significant and negative coefficient on distance. The magnitudes of the coefficients can be interpreted as follows: for the Simple specification fitted using 2012 data, the distance coefficient is $-0.395$, which means that for two school programs that are otherwise identical but one is one mile further from home, the student is more likely to choose the closer one $e^{-0.395} \approx 60\%$ of times. All estimates yield highly significant and positive coefficients for continuing, sibling, ell match, and ell language match. To gain further intuition, one can examine the ratios of these estimates with the distance coefficient. For example, all else being
Table 3.3: Estimated coefficients for logit models. Each model is estimated using maximum likelihood, using the choice data for grades K1 and K2. The standard errors are estimated using the inverse of the Hessian of the log likelihood function at the maximum.

<table>
<thead>
<tr>
<th></th>
<th>2012 Data</th>
<th>2013 Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance</td>
<td>-0.395***</td>
<td>-0.438***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.018)</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>sibling</td>
<td>2.143***</td>
<td>2.101***</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>ell match</td>
<td>1.548***</td>
<td>1.550***</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>ell language match</td>
<td>0.719***</td>
<td>0.610***</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>walk zone</td>
<td>0.570***</td>
<td>0.497***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>distance * black/hispanic</td>
<td>0.115***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>distance * income est.</td>
<td>-0.233***</td>
<td>-0.262***</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>mcas * black</td>
<td>-0.599***</td>
<td>-0.874***</td>
</tr>
<tr>
<td></td>
<td>(0.166)</td>
<td>(0.105)</td>
</tr>
<tr>
<td>mcas * income est.</td>
<td>0.506**</td>
<td>0.424*</td>
</tr>
<tr>
<td></td>
<td>(0.232)</td>
<td>(0.221)</td>
</tr>
<tr>
<td>% white/Asian * black/hispanic</td>
<td>-2.581***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td></td>
</tr>
<tr>
<td>% white/Asian * income est.</td>
<td>1.908***</td>
<td>1.982***</td>
</tr>
<tr>
<td></td>
<td>(0.218)</td>
<td>(0.211)</td>
</tr>
<tr>
<td>school fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>full interaction</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Log Likelihood | -70,969 | -70,013 | -70,090 | -65,944 | -64,763 | -64,829 |
# of Parameters | 81      | 99      | 87      | 83      | 101     | 89      |
# Students      | 6,644   | 6,644   | 6,644   | 6,627   | 6,627   | 6,627   |
# Choices       | 27,905  | 27,905  | 27,905  | 26,901  | 26,901  | 26,991  |

Note: *p<0.1; **p<0.05; ***p<0.01
equal, students in the Simple specification in 2012 are on average willing to travel 
\[
\frac{4.070}{0.395} \approx 10.3 \text{ extra miles to go to a continuing program, } 5.4 \text{ extra miles to go to school with a sibling, } 3.9 \text{ extra miles to go to an ELL program (if the student is ELL), and } 1.8 \text{ extra miles to go to an ELL program specialized to his/her home language.}
\]

Being in the walk zone is relevant because only students who live outside the walk zone are provided busing. Moreover, it is correlated with extreme proximity. Because of these potentially conflicting influences, the positive coefficient for walk zone is difficult to interpret. Another complication is that before 2014, students in the walk zone get walk zone priority to go to the school. Although this should not theoretically affect choice rankings because the mechanism is strategyproof, families may not fully appreciate this property and rank walk-zone schools higher because they think they have better chances to get into them. Alternative, the significance of this variable may indicate that student’s perceive a significant fixed costs to having to attend a school that requires a bus ride.

In creating the Reduced specification, we first fit the Full model. However, we found that the estimates for black and hispanic students are statistically indistinguishable, except for interaction with mcas, in which the coefficient for black students are significantly negative while the coefficient for hispanic students is insignificant. Moreover, the results for the other races are unstable between the years or insignificant, most likely due to lack of data because as seen in Table 3.2, few students are asian (8%) or other (3%). Thus, in the Reduced specification, we group black and hispanics together, except for the interaction with mcas, and we remove the other race dummies, implicitly grouping them with whites. We included all the terms involving income estimates since they tend to be statistically significant. The coefficients suggest that black/hispanic students are willing to travel further than other students, and tend to choose schools that have lower % white/asian. This may be due to demographic preferences, or preferences for unobserved characteristics that are correlated with demographics, such as school culture or environment. Black students seem to disproportionately choose schools with lower math scores even with our controls for distance and neighborhood income.
Because the Reduced specification captures all of the significant and stable interaction terms in the Full specification, but has more precise estimates, so we opt for this specification and simply refer to it by “Logit” in the rest of this chapter.

**Mixed Logit (MLogit)**

This model adds random coefficients to multinomial logit. Specifically, we use the following formulation,

\[ u_{ij} = \beta \cdot F_{ij} + \gamma_i \cdot G_{ij} + \epsilon_{ij}, \]

where \( F_{ij}, \epsilon_{ij} \) and \( \beta \) are observed features, iid taste shocks, and fixed coefficients as before. However, we allow the addition of a subset of features \( G_{ij} \) to interact with random coefficients \( \gamma_i \), which we assume to be zero mean jointly Gaussian distributed random variables, with possible covariance restrictions. The assumption of zero mean is without loss of generality since the means are captured by the fixed coefficients.

The assumption of Gaussian distribution is for convenience.

In terms of features we include, we use the fixed coefficients as in the Reduced specification of the logit model, but add random coefficients to the following features, which we organize into “blocks,” assuming independence across blocks, but allowing arbitrary covariance within each block.

<table>
<thead>
<tr>
<th>Block</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ell match</td>
</tr>
<tr>
<td>B</td>
<td>walk zone</td>
</tr>
<tr>
<td>C</td>
<td>distance, mcas, % white/asian</td>
</tr>
</tbody>
</table>

This formulation allows students to have heterogeneous preferences of going to an ELL program (if applicable), of choosing schools in the one-mile walk zone\(^6\), and of trading off distance, academics, and school demographics. We also include school fixed effects in order to capture the many unobserved school characteristics, such as safety, reputation, facilities, environment, and teaching quality.

\(^6\)The walk zones are only approximately one mile disc, because they were originally defined using geocodes.
Because the model no longer has closed form log-likelihood functions, and the log-likelihood functions are no longer guaranteed to be globally concave, we fit the model by Markov Chain Monte Carlo (MCMC), which is a commonly employed method for fitting such models in practice \cite{Train2003}. One technical difficulty in our situation is that we have many school fixed effects. As far as we are aware, the state of the art MCMC techniques for including fixed effects in mixed-logit models is described in \cite{Train2003}, and it involves adding a layer of Gibbs sampling and simulating the conditional distribution of the fixed effects using the Random Walk Metropolis-Hasting algorithm. However, in our case, there are 75 schools, so this step requires simulating a 75-dimensional distribution, which is prohibitively slow using Random Walk Metropolis (RWM).\footnote{See Katafygiotis and Zue\v{s} \cite{Katafygiotis2008} for geometric insight to why RWM breaks down in high dimensions when the dimensions are correlated.} Hence, we speed up computation by using Hamiltonian Monte Carlo (HMC)\footnote{See Neal \cite{Neal2011}.}, which incorporates the gradient of the log likelihood function, so can more quickly update the 75-dimensional estimate for fixed effects. We fit the above model by using 1,000,000 iterations of MCMC sampling, throwing out the first half as burn-in. To check for the convergence of the estimates, we repeated this 6 times with independent draws, sometimes with random starting values, and found the results to be near identical. Details of how we fit the mixed logit model are in Appendix 3.C.

The estimates are in Table 3.4. Note that beside the fixed coefficients in the Reduced specification of the simple logit model, we also estimate the standard deviations of the random coefficients, denoted by $\sigma(\text{ell match})$, $\sigma(\text{walk zone})$, $\sigma(\text{distance})$, $\sigma(\text{mcas})$, and $\sigma(\% \text{ white/asian})$. These are the square roots of the respective variances. We also estimate the correlation coefficients $\rho(\text{distance, mcas})$, $\rho(\text{distance, } \% \text{ white/asian})$, and $\rho(\text{mcas, } \% \text{ white/asian})$, which are computed by dividing the respective covariance terms by the product of the standard deviations. In the rest of the chapter, we refer to this model as “MixedLogit” or “MLogit” for short.
Table 3.4: Coefficients for Mixed Logit (MLogit) models, compared to simple Logit.

<table>
<thead>
<tr>
<th></th>
<th>2012 Data</th>
<th>2013 Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance</td>
<td>-0.365***</td>
<td>-0.638***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>continuing</td>
<td>4.027***</td>
<td>4.777***</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>sibling</td>
<td>2.104***</td>
<td>2.478***</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>ell match</td>
<td>1.548***</td>
<td>1.892***</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>ell language match</td>
<td>0.606***</td>
<td>0.610***</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>walk zone</td>
<td>0.500***</td>
<td>0.339***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>distance * black/hispanic</td>
<td>0.115***</td>
<td>0.188***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>distance * income est.</td>
<td>-0.262***</td>
<td>-0.295***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>mcas * black</td>
<td>-0.874***</td>
<td>-1.100***</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.153)</td>
</tr>
<tr>
<td>mcas * income est.</td>
<td>0.424*</td>
<td>1.065***</td>
</tr>
<tr>
<td></td>
<td>(0.221)</td>
<td>(0.299)</td>
</tr>
<tr>
<td>% white/Asian * black/hispanic</td>
<td>-2.581***</td>
<td>-3.732***</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.162)</td>
</tr>
<tr>
<td>% white/Asian * income est.</td>
<td>1.982***</td>
<td>2.633***</td>
</tr>
<tr>
<td></td>
<td>(0.211)</td>
<td>(0.322)</td>
</tr>
<tr>
<td>σ(ell match)</td>
<td>1.638***</td>
<td>1.358***</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>σ(walk zone)</td>
<td>0.981***</td>
<td>0.878***</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>σ(distance)</td>
<td>0.392***</td>
<td>0.409***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>σ(mcas)</td>
<td>2.275***</td>
<td>2.121***</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>σ(% white/Asian)</td>
<td>2.672***</td>
<td>2.512***</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>ρ(distance, mcas)</td>
<td>-0.232***</td>
<td>-0.285***</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>ρ(distance, % white/Asian)</td>
<td>-0.089**</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>ρ(mcas, % white/Asian)</td>
<td>0.035</td>
<td>-0.110*</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.061)</td>
</tr>
</tbody>
</table>

*Note:* *p<0.1; **p<0.05; ***p<0.01
3.3.2 Forecasting the Applicant Pool

An important driver of the number of unassigned students and the average access to Tier 1 or 2 schools is the number of applicants from each neighborhood. Regardless of how applicants choose schools, a large influx of new applicants from a neighborhood would drive up the number of unassigned from that neighborhood and drive down the average access to top Tier schools from that neighborhood. If we had data on all potential applicants and their non-BPS options, we might include this aspect as part of the structural model. In the absence of such data, we still need to reflect this uncertainty and to capture any first-order trends in the neighborhood participation patterns.

In forecasting the applicant pool, we consider new and continuing students separately. This is because continuing students are already in the enrollment data of the previous year, while for new students we need to use previous year’s applicants’ demographics as proxies. Figure 3-2 plots the number of new applicants to BPS in Round 1 for grades K0-2 for years 2010-2013, as well as the regression line with respect to the year. As seen, the number of applicants is on average increasing each year, at a rate of 6% a year on average, although it is not steady. For example, in 2012 there was an above-expected number of applicants. We model the next year’s total number of new applicants by a normal distributed random variable, having mean and standard deviation being the predicted mean and standard error of the regression line.

Figure 3-3 shows what proportion of this total is distributed into each grade and neighborhood combination. Since we study two grades and there are 14 neighborhoods, there are 28 time series in these plots. Most of the time series do not exhibit obvious trends. We model each of the 28 proportions next year as a normally distributed random variable. To estimate the mean and standard deviation, we run a regression with respect to year for each of these 28 time series, and discard all regressions for which the slope has less than 95% significance level. For the neighborhood and grade combinations in which we discard the regression, we forecast next year’s

---

9The influx of applicants in 2012 Round 1 raised operational pressures for BPS, as it had to add about 10 new classrooms to accommodate.
Figure 3-2: Trend in total number of applicants to BPS.

proportion using the previous 4 years’ sample mean and sample standard deviation. For the neighborhood and grade combinations in which the regression slope has 95% significance, we use the predicted mean and standard error of the regression. The regression lines we kept are for K1 Charlestown and K2 Downtown, for which we detected a steady upward trend in the number of applicants.

We therefore model the total number of new applicants from each neighborhood in the next year as a product of two independent normals, one representing a BPS wide shock and one a neighborhood shock. The common shock captures the uncertain effect that BPS publicity or policy initiatives have on the propensity for families to apply to BPS round 1. The neighborhood shock captures local population surges or unobserved reasons that affect participation. By using one common shock for all grades, we implicitly assume that different grades exhibit the same reactions to BPS policies, and are trending in the same directions. To check this, we plot in Figure 3-4 the proportion of new applicants of each grade through the four years. As seen, the relatively horizontal lines suggest that modeling the aggregate participation of both K1 and K2 using the same random variable may be a reasonable approximation.

For continuing students, we define the Round 1 continuing ratio for a grade and neighborhood as the proportion of relevant students from the previous year’s enrollment data who decide to continue in this year’s Round 1. Figure 3-5 plots these for

---

10 Note that the standard error has one fewer degree of freedom than the sample standard deviation.
11 The means and standard deviations of these estimates are tabulated in Appendix 3.B. After multiplying the two normals, we truncate at zero if the product is negative and round to the nearest integer.
Figure 3-3: Proportion of new applicants distributed into each grade, neighborhood combination.

Figure 3-4: The proportion of new applicants of each grade.
grades K1 and K2. As seen, due to the lower number of continuing students in K1 (recall that comparatively few students enroll in K0), the estimates for K1 are highly variable, while the K2 continuing ratios are around 70% to 90%.

Figure 3-5: For each grade and neighborhood, the proportion of potentially continuing students from the previous year who apply as continuing students in the current year.

We use the same approach to detect linear trends. However, in this data we failed to find any significant trends in continuing ratios, so we model them as normally distributed random variables, with mean and standard deviation according to the sample mean and sample standard deviation for years 2010-2013. The estimates are in Appendix 3.B

To simulate the pool applicant pool for next year, we independently draw the total number of new applicants, the proportion of this to allocate to each grade and
neighborhood\textsuperscript{12} and the continuation ratio for each grade and neighborhood. Fixing these realizations for one sample of choice data, we draw new applicants by sampling with replacement the previous year’s new applicants, but treating them as if they applied in the new year, and we sample the given number of new applicants from each neighborhood and grade. For continuing students, we go through all potentially continuing students and decide whether to include them independently, with probability according to the generated continuation ratio for that grade and neighborhood. We use this method to generate the applicant pool for all our simulations.

3.3.3 Simulation

We use 4 layers of random draws for our simulations.

1. **Population Draw**: Draw a pool of applicants according to the steps in Section 3.3.2. This represents uncertainty in participation rates.

2. **Coefficient Draw**: For the logit and mixed-logit demand models, draw the coefficients as jointly normal random variables, using the estimated means and covariance matrix. This represents uncertainty in the demand model.

3. **Preference Draw**: Having fixed a specific demand model and parameters, simulate for each student a complete ranking over his/her personalized menu of options, according to the randomness inherent in the demand model. Truncate this to the first ten choices.

4. **Lottery Draw**: Generate iid lottery numbers for each student. The lottery numbers are distributed uniformly between zero and one, with lower lottery numbers being better.

After doing these steps once, we have one set of simulated choice data, just as what we might have received from BPS. From this we can deterministically compute all of our outcome metrics by imitating the BPS assignment algorithm.

\textsuperscript{12}Since the total includes K0 but we do not estimate proportions for K0, we do not require these draws to add up to 1.
The reason we truncate to first 10 choices is as follows: currently our choice data from BPS truncates to first 10 choices, although students can rank arbitrarily many. The previous report, Pathak and Shi (2013), provides evidence that assuming everyone ranks 10 choices yields reasonably accurate forecasts. Moreover, our earlier report on which the city committee based the 2012-2013 reform assumed that families ranked 10 options, and we keep the same assumptions to validate or invalidate the methodology of the earlier report.

An alternative approach is to model outside options and assume that unranked programs are inferior to the outside option. However, we observe in the data that often students end up enrolling to options they did not rank but could have ranked, suggesting that this assumption is invalid. In our interactions with parents and BPS staff, it seems that many families are ranking few options not because they have better outside options, but because they feel confident they would get into the ones they picked and did not bother, or because they do not understand that ranking more options do not harm their chances to top choices. Future work is needed to better model this situation.

3.3.4 Evaluation

Having computed the assignment using the simulated choice data and lottery numbers, we compute the outcome measures as follows. In all of the analysis, we compute measures for grades K1 and K2 separately.

- **Unassigned**: Tally the number of unassigned students from each neighborhood after each round.

- **Access to Quality**: For each student, define his/her access to quality as the highest (worst) lottery number he/she can have and still be able to be assigned to a Tier 1 or 2 school. (Recall that lottery numbers are uniformly distributed between zero and one, so this can be interpreted as a probability.) We estimate this by finding for each Tier 1 or 2 school program the highest (worst) lottery number he/she needs to obtain an offer. More precisely, if the program is not
filled to capacity, the student can get in even with the worst lottery number; if
it is filled to capacity, we look at the worst lottery number the student needs to
be able to displace one student to obtain an offer.\footnote{For computing this metric, we ignore the possibility of that student displacing someone else at his/her next choice and starting a chain reaction that cycles back to the first student, since this is unlikely to occur in markets with a large number of participants \cite{Kojima and Pathak (2009)}.} Then we take the maximum
over all of his/her Tier 1 or 2 options and term this his/her “access to quality.”
Finally, we average the access to quality of all students of a neighborhood to
compute the neighborhood average access.

- **Distance**: For each neighborhood, take all the assigned students from this neigh-
  borhood and average their walk distances to assigned school.

- **Top k Market Share**: Take all the top \( k \) choices of students from a neighborhood.
  For each school, find the proportion of these choices that are to this school.\footnote{If a student ranks multiple programs of the same school as part of his/her top \( k \) choice, he/she contributes multiple “votes” for that school, since we treat each top \( k \) choice as one “vote.”}

We compute these metrics for each generated choice data and average across many
simulations to find the mean prediction. For a given choice data, we compute the
above metrics and define the prediction error for each neighborhood as follows: for
unassigned, access to quality, and distance, since we have a scalar value for each
neighborhood, we simply take the absolute value of the difference between predicted
mean and actual realization. For top \( k \) market share, since for each neighborhood we
have a vector of school market shares that add up to one, we define prediction error
as the total variation distance between these probability vectors. Given vectors of
market shares \( \mathbf{p} \) and \( \mathbf{q} \), define the total variation distance between them as

\[
\text{total variation distance} = \frac{1}{2} \sum_j |p_j - q_j|.
\]

This metric is standard for probability distributions, and can be interpreted as the
least total movement needed to redistribute market shares by moving from one school
to another, to turn the predicted shares into the actual shares.
Finally we judge the overall prediction accuracy by taking the Root Mean Squared Error (RMSE) over the neighborhoods. Specifically, for each metric, we take the prediction error for each neighborhood, square these numbers, take the mean of the squared errors, and take the square room of the mean. We choose this over Mean Absolute Deviation (MAD) because we want to penalize being far off for any specific neighborhood over being slightly off for many neighborhoods. This is because for capacity planning and for policy evaluation, the penalties are large for not foreseeing a large shortfall of seats in a neighborhood or not foreseeing vast inequities between neighborhoods.

3.4 Results

3.4.1 Raw Prediction Errors

For each neighborhood and each moment of interest, we compare the predictions with the actual outcome. For the moments “access to quality,” “distance,” and “unassigned,” this amounts to comparing two scalars. We compute the root mean square error (RMSE) across neighborhoods and tabulate the results in Table 3.5. The details for each neighborhood are in Appendix 3.A.

Table 3.5: RMSE across neighborhoods between forecasts and actual outcome.

<table>
<thead>
<tr>
<th>2014 Moment</th>
<th>Naive</th>
<th>K1</th>
<th>mLogit</th>
<th>Naive</th>
<th>K2</th>
<th>mLogit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to Quality</td>
<td>0.26</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>expected error</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>P-value</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Distance</td>
<td>0.34</td>
<td>0.19</td>
<td>0.19</td>
<td>0.14</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td>expected error</td>
<td>0.14</td>
<td>0.12</td>
<td>0.12</td>
<td>0.09</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>P-value</td>
<td>0</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Unassigned</td>
<td>30.22</td>
<td>21.65</td>
<td>21.21</td>
<td>34.02</td>
<td>40.92</td>
<td>40.44</td>
</tr>
<tr>
<td>expected error</td>
<td>15.95</td>
<td>16.26</td>
<td>16.54</td>
<td>8.89</td>
<td>7.49</td>
<td>7.85</td>
</tr>
<tr>
<td>P-value</td>
<td>0.03</td>
<td>0.18</td>
<td>0.20</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

As can be seen, while the Logit and Mixed Logit models continue to outperform Naive for K1 in prediction accuracy, they for the most part performed worse than
Naive for K2. However, recall that the Naive model did not require fitting choice data at all. In terms of absolute error, while the Logit and Mixed Logit prediction errors yielded a p-value of 1% to 20% for various moments for K1, all of the models yielded a near zero p-value for K2. Especially for the number of unassigned students, all of the models yielded prediction errors many times what would have been expected from the simulations. This at first glance is a very negative result for predicting counterfactuals using demand modeling. Using a complex choice model may yield worse results than having no models at all.

We also examine the prediction errors in market shares. For these, we calculate for each neighborhood the total variation distance between the predicted distribution across schools with the observed distribution. We also calculate the RMSE across neighborhoods and tabulate the results in Table 3.6. The detail market shares for each neighborhood and each school are in ancilliary tables attached to this chapter.

Table 3.6: RMSE across neighborhoods for market shares between forecasts and actual outcome.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Naive</th>
<th>K1 Logit</th>
<th>mLLogit</th>
<th>Naive</th>
<th>K2 Logit</th>
<th>mLLogit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1 Mkt. Share</td>
<td>0.49</td>
<td>0.24</td>
<td>0.24</td>
<td>0.30</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>expected error</td>
<td>0.10</td>
<td>0.12</td>
<td>0.12</td>
<td>0.08</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>P-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Top 2 Mkt. Share</td>
<td>0.46</td>
<td>0.22</td>
<td>0.21</td>
<td>0.38</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>expected error</td>
<td>0.08</td>
<td>0.09</td>
<td>0.09</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>P-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Top 3 Mkt. Share</td>
<td>0.45</td>
<td>0.21</td>
<td>0.20</td>
<td>0.39</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td>expected error</td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>P-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In contrast to for the previous moments, the market share predictions using Logit and Mixed Logit consistently outperform Naive. This means that the more sophisticated choice models are better describing the actual rank order choices of students.

### 3.4.2 Decomposing the Prediction Error

Recall that in forming predictions for the moments of interest, we had to estimate two components:
1. A choice model \( Y = f(X, S) \) that for given matrix of characteristics of students \( X \), and matrix of characteristics of school programs \( S \), predicted the rank order list for each student, denoted by \( Y \).

2. Forecasts for the student population \( X \) and school characteristics \( S \).

Let \( X^a \), \( S^a \) and \( Y^a \) be the actual student population, actual school program characteristics, and

Let \( M(Y, S) \) be the outcome in the moment of interest given student choices \( Y \) and school characteristics \( S \) (which includes school program capacities). For access to quality, distance, and unassigned, \( M \) is a 14-dimensional vector with each dimension denoting a neighborhood. Let \( Y^a \) be the actual rank-order choices, the total prediction error can be expressed as

\[
\|M(Y^a) - E[M(f(X, S))]\|
\]

where the norm is \( L_2 \). There are many potential causes for error in prediction:

1. The choice model \( f \) can be off. Since our models did not estimate the value of outside options, this in turn has two components: a) the within school model can be off b) our assumption that each student choose 10 schools could be causing overall prediction errors.

2. The population forecasts \( X \) and school forecasts \( S \) can be off.

For ease of reference, we call the first error in model and the second error in environment.

To disentangle between these, we redo the simulations with the following changes in setups:

- Updated-M: use the choice model fitted using 2014 choice data. This partials out the possibility that the reform fundamentally changed the choice model from 2013 to 2014. Let \( f^u \) be the choice model using the updated data. In symbols, this is \( \|M(Y^a) - E[M(f^u(X, S))]\| \)
- Updated-P: use the actual student population in 2014, along with the actual school program characteristics. This partials out any error in the environment. In symbols, this is $||M(Y^a) - E[M(f(X^a, S^a))]|$

- Updated-PM: using both the 2014 choice model and the 2014 students and schools. This uses both the updated choice model and the right environment. In symbols, this is $||M(Y^a) - E[M(f^u(X^a, S^a))]|$

- Updated-PMR: doing the above and also using the actual number of choices ranked for each student.

The result of the resimulation are in Tables 3.7 and 3.8.

Table 3.7: Resimulating by using a combination of the updated demand model (M), population (P), or rank-order list length (R). The original simulation errors are given as references.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Setup</th>
<th>K1</th>
<th>K2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Naive</td>
<td>Logit</td>
</tr>
<tr>
<td>Access to Quality</td>
<td>Raw Error</td>
<td>0.26</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Updated-M</td>
<td>0.26</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.22</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>0.22</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Updated-PMR</td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>Distance</td>
<td>Raw Error</td>
<td>0.34</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Updated-M</td>
<td>0.33</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.21</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>0.20</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Updated-PMR</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Unassigned</td>
<td>Raw Error</td>
<td>30.22</td>
<td>21.65</td>
</tr>
<tr>
<td></td>
<td>Updated-M</td>
<td>28.81</td>
<td>21.67</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>7.80</td>
<td>6.68</td>
</tr>
<tr>
<td></td>
<td>Updated-PMR</td>
<td>26.82</td>
<td>6.03</td>
</tr>
</tbody>
</table>

As can be seen, the prediction error changes little going from raw error to using the 2014 choice model (Updated-M). Similarly, when we use the 2014 population, having the updated choice model also makes little difference (going from Updated-P to Updated-PM). Moreover, although using the right number of choices ranked does improve estimates for Logit and Mixed Logit, the improvements are not so drastic.
Table 3.8: Market shares when resimulating by using a combination of the updated demand model (M), or population (P). The original simulation errors are given as references.

<table>
<thead>
<tr>
<th>Market Share</th>
<th>Setup</th>
<th>K1 Naive</th>
<th>K1 Logit</th>
<th>K1 Mlogit</th>
<th>K2 Naive</th>
<th>K2 Logit</th>
<th>K2 Mlogit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1</td>
<td>Raw Error</td>
<td>0.49</td>
<td>0.24</td>
<td>0.24</td>
<td>0.30</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Updated-M</td>
<td>0.49</td>
<td>0.24</td>
<td>0.23</td>
<td>0.30</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.41</td>
<td>0.22</td>
<td>0.23</td>
<td>0.23</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>0.41</td>
<td>0.20</td>
<td>0.18</td>
<td>0.23</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>Top 2</td>
<td>Raw Error</td>
<td>0.46</td>
<td>0.22</td>
<td>0.21</td>
<td>0.38</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Updated-M</td>
<td>0.46</td>
<td>0.23</td>
<td>0.22</td>
<td>0.38</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.44</td>
<td>0.17</td>
<td>0.19</td>
<td>0.40</td>
<td>0.15</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>0.44</td>
<td>0.16</td>
<td>0.15</td>
<td>0.40</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>Top 3</td>
<td>Raw Error</td>
<td>0.45</td>
<td>0.21</td>
<td>0.20</td>
<td>0.39</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Updated-M</td>
<td>0.45</td>
<td>0.21</td>
<td>0.20</td>
<td>0.39</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.41</td>
<td>0.15</td>
<td>0.17</td>
<td>0.40</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>0.41</td>
<td>0.14</td>
<td>0.13</td>
<td>0.40</td>
<td>0.12</td>
<td>0.12</td>
</tr>
</tbody>
</table>

However, there is large improvements in prediction errors when we use the right population (going from raw error to Updated-P). This means that the majority of prediction error is not caused by the choice model, but by errors in our forecasted environment. Moreover, although the result seem to have the potential to improve if we had a more sophisticated model for outside option, this is dominated by the errors in the environment.

### 3.4.3 Causes of Large Prediction Error in the Environment

Examining the data qualitatively, we find the following errors in our predictions of the environment.

- Errors in school program characteristics: there were some changes in programming that BPS made that were not aware to us. The most significant change was that the special inclusive programs IEE were merged with regular programs KED, so that these special programs no longer existed in schools, but every KED program now had to be able to have special education students. This decreases the number of potential programs in many schools, which impacts the Naive model the most because of its lexicographic ranking rule, which would put both
IEE and KED together. Other changes include changes of language specialization of English Language Learner (ELL) programs at 4 schools, the movement of kindergarten programs to of Higginson/Lewis school to a new school called Higginson K-2 at a different address, and the movement of Montessori programs from East Boston EEC to a new Montessori-only school at a new location.

- Error in student demographics: Unknown to us (and also to the team in charge of student assignment at BPS), the English Language Learners (ELL) division altered their tests for determining whether a student qualifies for ELL in grade K2. From assignment years 2010 to 2013, the percentage of K2 students who are ELL has been between 33% and 45%, but in 2014 with the new test, only 16% of students qualified as ELL. The result is many empty seats in ELL programs and overdemand for KED programs.

- Error in number of students applied: table 3.9 tabulate the predicted and actual students in K1 and K2 for both continuing and new students. As can be seen, we over-predicted the number of new students and under-predicted the number of continuing students. For K2, the discrepancy in continuing students is especially striking. This is surprising because we used the December 2013 BPS enrollment file, which was only 2 months prior to the round 1 choice submission, in order to forecast continuing students. It turned out that there were students who applied who were not in those enrollment files. Moreover, our previous simulations did not account for the change in program code as students went from K1 to K2, and for students whose program code did not exist in K1, we assumed that they would be assigned outside of the system. But in reality, these program codes were changed and the students had continuing status at K2 for new programs.

Table 3.9: Prediction error in number of applicants

<table>
<thead>
<tr>
<th>Grade</th>
<th>Type</th>
<th>Predicted</th>
<th>(Std. Error)</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>Continuing</td>
<td>91.4</td>
<td>6.9</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>2067.0</td>
<td>185.0</td>
<td>2313</td>
</tr>
<tr>
<td>K2</td>
<td>Continuing</td>
<td>1482.7</td>
<td>30.7</td>
<td>2051</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>2189.5</td>
<td>150.6</td>
<td>1875</td>
</tr>
</tbody>
</table>

92
To quantify the effect of each of these above components, we repeat the simulation with the following setups:

- **Updated-D (Demographics):** Fixing the number of continuing and new students from each neighborhood in each sample run of the original simulations, redraw these students' characteristics (race, income, continuing school, etc) from the actual population pool that showed up in the 2014 data. Then resimulate the outcome.

- **Updated-N (Numbers):** For each sample run of the original simulations, resample students in each neighborhood to match the actual number of continuing and new students in each neighborhood.

- **Updated-DN:** Using both the actual number of continuing and new students from each neighborhood, and also drawing student characteristics from the actual student pool. This is like a bootstrap version of the original simulations using the right demographics and population counts, but using old school characteristics.

- **Updated-DNS:** Doing the above but also using the updated school characteristics.

We compare these setups with the original simulations and “Updated-P” from before. (Updated-P is similar to Updated-DNS except it use the exact student population, without randomness.) The results are tabulated in Table 3.10.

As can be seen, the most improvements in prediction error happens when we are using the right number of students. For the Naive model, having the right set of school programs has a big effect, but this is not so important for the Logit and Mixed Logit models (which are not affected as much by the merging of IEE programs into KED.)
Table 3.10: Resimulating by using a combination of the updated demographics (D), number of applicants (N), or school programs (S). The original simulation errors and when the exact population is fixed as actual are given as references.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Setup</th>
<th>Raw Error</th>
<th>Naive</th>
<th>Logit</th>
<th>Mlogit</th>
<th>Naive</th>
<th>Logit</th>
<th>Mlogit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to Quality</td>
<td>Raw Error</td>
<td>0.26</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.15</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Updated-D</td>
<td>0.27</td>
<td>0.12</td>
<td>0.13</td>
<td>0.20</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Updated-N</td>
<td>0.22</td>
<td>0.07</td>
<td>0.07</td>
<td>0.14</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Updated-DN</td>
<td>0.23</td>
<td>0.06</td>
<td>0.07</td>
<td>0.23</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Updated-DNS</td>
<td>0.22</td>
<td>0.05</td>
<td>0.05</td>
<td>0.21</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.22</td>
<td>0.05</td>
<td>0.06</td>
<td>0.22</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

| Distance   | Raw Error | 0.34      | 0.19  | 0.19  | 0.14   | 0.13  | 0.14  |
|            | Updated-D | 0.26      | 0.15  | 0.15  | 0.12   | 0.12  | 0.13  |
|            | Updated-N | 0.31      | 0.19  | 0.19  | 0.20   | 0.11  | 0.11  |
|            | Updated-DN| 0.25      | 0.14  | 0.15  | 0.16   | 0.09  | 0.10  |
|            | Updated-DNS| 0.21   | 0.15  | 0.15  | 0.15   | 0.08  | 0.09  |
|            | Updated-P | 0.21      | 0.15  | 0.15  | 0.15   | 0.08  | 0.09  |

| Unassigned | Raw Error | 30.22     | 21.65 | 21.21 | 34.02  | 40.92 | 40.44 |
|            | Updated-D | 31.53     | 23.13 | 23.51 | 21.97  | 29.09 | 28.91 |
|            | Updated-N | 15.86     | 7.61  | 7.91  | 26.00  | 33.77 | 33.38 |
|            | Updated-DN| 18.18     | 5.99  | 6.36  | 16.01  | 15.16 | 15.90 |
|            | Updated-DNS| 8.27  | 7.02  | 7.35  | 15.84  | 15.73 | 16.35 |

3.5 Additional Analysis

3.5.1 Value of Random Coefficients

One question discussed in the literature is the importance of random coefficients, which allows the model to better capture substitution pattern between choices. As seen in Table 3.8, mLogit indeed outperforms Logit across majority of market share predictions. To directly measure substitution, we compute the joint distribution of first and second choice for each model and each grade, and compute the total variation between this predicted joint distribution and the actual joint distribution. We also explore varying the simulation setup, using the original simulations, as well as updating the population and demand model to the current year as before. The results are in Table 3.11.

As before, we see that having random coefficients allows better fit to the actual joint probability distribution of first and second choices. However, the improvement is
Table 3.11: The total variation between the estimated joint probability distribution of first and second choice, and the actual joint distribution. As can be seen, in most setups, mLogit outperforms Logit, with the exception of 2014 K1 under the setup “Updated-P”.

<table>
<thead>
<tr>
<th>Year</th>
<th>Setup</th>
<th>Naive</th>
<th>Logit</th>
<th>mLogit</th>
<th>Naive</th>
<th>Logit</th>
<th>mLogit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>Raw error</td>
<td>0.74</td>
<td>0.44</td>
<td>0.40</td>
<td>0.77</td>
<td>0.51</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.75</td>
<td>0.44</td>
<td>0.40</td>
<td>0.76</td>
<td>0.51</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>0.75</td>
<td>0.43</td>
<td>0.39</td>
<td>0.76</td>
<td>0.50</td>
<td>0.47</td>
</tr>
<tr>
<td>2014</td>
<td>Raw error</td>
<td>0.81</td>
<td>0.46</td>
<td>0.46</td>
<td>0.76</td>
<td>0.53</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Updated-P</td>
<td>0.70</td>
<td>0.42</td>
<td>0.44</td>
<td>0.70</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Updated-PM</td>
<td>0.70</td>
<td>0.41</td>
<td>0.38</td>
<td>0.70</td>
<td>0.50</td>
<td>0.48</td>
</tr>
</tbody>
</table>

small in magnitude. There’s also an exception: for 2014 K1, using the right population but the previous year’s choice model, Logit actually outperforms mLogit, which may occur because allowing random coefficients requires fitting extra coefficients, so there’s a higher danger of overfitting and the estimates are not as precise.

So in our setting, although having random coefficients indeed improves prediction of substitution patterns in majority of cases, one may argue that the improvement is too small to be worth the trouble. Examining the underlying joint market shares in detail, we find qualitative intuition for why this is so: the actual choices have high correlations between certain schools in specific neighborhoods, which even mLogit cannot capture. Indeed, we could not have expected it to, because the data we use to apply the random coefficients—mcas, distance, and % white/asian—do not complete capture the similarities of these schools to families (such as whether they are in these specific neighborhoods). In other words, random coefficients can help only as much as the characteristics we apply these random coefficients capture the underlying similarity of options, so without better data, random coefficients alone are of limited value.

3.5.2 Tier Salience Effect

Although the majority of our prediction errors were because of errors in forecasts of population and schools, there are some effect of Tier salience changing how students ranked schools. In figures 5-6 and 5-7 we plot for each tier and each model the
proportion of top $k$ choices of that tier. As can be seen, the actual choices ranked Tier 1 schools in top 1 and 2 choices more than predicted by Logit and Mixed Logit, but not as much as predicted by Naive (which after continuing, sibling, and ELL status ranked schools entirely by Tiers.)

Figure 3-6: The percentage of top choices of each tier for K1. For example, to see the actual percentage of K1 top 3 choices for tier 1, read the y-coordinate for the Tier 1 plot with x-coordinate being 3. The predictions from using the actual population and each of the demand models are also listed. As can be seen, the actual outcome leans toward better tiers compared to logit or mixed logit, but the move is not as drastic as predicted in the Naive model.

(a) Tier 1

(b) Tier 2

(c) Tier 3

(d) Tier 4 or unranked
Figure 3-7: The percentage of top choices of each tier for K2. For example, to see the actual percentage of K1 top 3 choices for tier 1, read the y-coordinate for the Tier 1 plot with x-coordinate being 3. The predictions from using the actual population and each of the demand models are also listed. As can be seen, the actual outcome is lean toward better tiers compared to logit or mixed logit, but the move is not as drastic as predicted in the Naive model.
### 3.A Neighborhood Level Results

Table 3.12: Comparing Access to Quality predictions with actual outcome for 2014 K1.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Naive</th>
<th>Logit</th>
<th>mLLogit</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>0.40</td>
<td>0.82</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>Charlestown</td>
<td>0.37</td>
<td>0.41</td>
<td>0.41</td>
<td>0.80</td>
</tr>
<tr>
<td>Downtown</td>
<td>0.43</td>
<td>0.58</td>
<td>0.59</td>
<td>0.65</td>
</tr>
<tr>
<td>East Boston</td>
<td>0.53</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>0.36</td>
<td>0.39</td>
<td>0.40</td>
<td>0.46</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>0.41</td>
<td>0.62</td>
<td>0.63</td>
<td>0.66</td>
</tr>
<tr>
<td>Mattapan</td>
<td>0.27</td>
<td>0.46</td>
<td>0.46</td>
<td>0.57</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>0.33</td>
<td>0.58</td>
<td>0.57</td>
<td>0.72</td>
</tr>
<tr>
<td>Roslindale</td>
<td>0.38</td>
<td>0.44</td>
<td>0.44</td>
<td>0.53</td>
</tr>
<tr>
<td>Roxbury</td>
<td>0.43</td>
<td>0.64</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>South Boston</td>
<td>0.39</td>
<td>0.51</td>
<td>0.50</td>
<td>0.67</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>0.30</td>
<td>0.58</td>
<td>0.58</td>
<td>0.67</td>
</tr>
<tr>
<td>South End</td>
<td>0.50</td>
<td>0.67</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>0.53</td>
<td>0.55</td>
<td>0.56</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 3.13: Comparing Distance predictions with actual outcome for 2014 K1.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Naive</th>
<th>Logit</th>
<th>mLLogit</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>1.23</td>
<td>1.22</td>
<td>1.23</td>
<td>1.16</td>
</tr>
<tr>
<td>Charlestown</td>
<td>1.82</td>
<td>1.41</td>
<td>1.38</td>
<td>1.40</td>
</tr>
<tr>
<td>Downtown</td>
<td>1.57</td>
<td>1.43</td>
<td>1.42</td>
<td>1.72</td>
</tr>
<tr>
<td>East Boston</td>
<td>1.97</td>
<td>1.44</td>
<td>1.38</td>
<td>1.51</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>2.05</td>
<td>1.90</td>
<td>1.87</td>
<td>2.10</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>1.31</td>
<td>1.27</td>
<td>1.27</td>
<td>1.14</td>
</tr>
<tr>
<td>Mattapan</td>
<td>2.30</td>
<td>1.77</td>
<td>1.76</td>
<td>1.84</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>1.47</td>
<td>1.26</td>
<td>1.20</td>
<td>1.00</td>
</tr>
<tr>
<td>Roslindale</td>
<td>2.81</td>
<td>1.62</td>
<td>1.61</td>
<td>1.18</td>
</tr>
<tr>
<td>Roxbury</td>
<td>1.43</td>
<td>1.29</td>
<td>1.26</td>
<td>1.35</td>
</tr>
<tr>
<td>South Boston</td>
<td>1.24</td>
<td>1.12</td>
<td>1.12</td>
<td>1.05</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>1.58</td>
<td>1.48</td>
<td>1.40</td>
<td>1.42</td>
</tr>
<tr>
<td>South End</td>
<td>1.50</td>
<td>1.35</td>
<td>1.33</td>
<td>0.93</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>2.08</td>
<td>1.76</td>
<td>1.76</td>
<td>1.52</td>
</tr>
</tbody>
</table>

Table 3.14: Comparing Unassigned predictions with actual outcome for 2014 K1.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Naive</th>
<th>Logit</th>
<th>mLLogit</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>21.36</td>
<td>14.75</td>
<td>11.96</td>
<td>14</td>
</tr>
<tr>
<td>Charlestown</td>
<td>36.05</td>
<td>40.67</td>
<td>40.18</td>
<td>15</td>
</tr>
<tr>
<td>Downtown</td>
<td>30.25</td>
<td>31.91</td>
<td>31.91</td>
<td>34</td>
</tr>
<tr>
<td>East Boston</td>
<td>123.49</td>
<td>132.59</td>
<td>132.58</td>
<td>91</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>80.39</td>
<td>82.77</td>
<td>81.96</td>
<td>56</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>81.56</td>
<td>60.26</td>
<td>56.88</td>
<td>55</td>
</tr>
<tr>
<td>Mattapan</td>
<td>84.71</td>
<td>70.77</td>
<td>69.33</td>
<td>47</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>55.55</td>
<td>28.06</td>
<td>27.34</td>
<td>22</td>
</tr>
<tr>
<td>Roslindale</td>
<td>108.77</td>
<td>114.96</td>
<td>114.25</td>
<td>77</td>
</tr>
<tr>
<td>Roxbury</td>
<td>81.77</td>
<td>55.48</td>
<td>54.89</td>
<td>36</td>
</tr>
<tr>
<td>South Boston</td>
<td>17.26</td>
<td>15.72</td>
<td>15.85</td>
<td>5</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>142.72</td>
<td>109.65</td>
<td>109.08</td>
<td>80</td>
</tr>
<tr>
<td>South End</td>
<td>34.00</td>
<td>23.71</td>
<td>23.00</td>
<td>27</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>81.46</td>
<td>93.77</td>
<td>92.67</td>
<td>93</td>
</tr>
</tbody>
</table>
### Table 3.15: Comparing Access to Quality predictions with actual outcome for 2014 K2.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Naive (95 % C.I.)</th>
<th>Logit (95 % C.I.)</th>
<th>mLogit (95 % C.I.)</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>0.80 (0.70,0.94)</td>
<td>0.98 (0.87,1.00)</td>
<td>0.98 (0.88,1.00)</td>
<td>0.75</td>
</tr>
<tr>
<td>Charlestown</td>
<td>0.94 (0.83,1.00)</td>
<td>0.98 (0.91,1.00)</td>
<td>0.99 (0.90,1.00)</td>
<td>1</td>
</tr>
<tr>
<td>Downtown</td>
<td>0.85 (0.76,0.93)</td>
<td>0.92 (0.85,0.97)</td>
<td>0.91 (0.84,0.96)</td>
<td>0.88</td>
</tr>
<tr>
<td>East Boston</td>
<td>0.80 (0.72,0.90)</td>
<td>0.87 (0.77,0.98)</td>
<td>0.88 (0.77,0.98)</td>
<td>0.89</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>0.69 (0.59,0.81)</td>
<td>0.86 (0.76,0.94)</td>
<td>0.86 (0.76,0.94)</td>
<td>0.73</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>0.68 (0.60,0.78)</td>
<td>0.89 (0.78,1.00)</td>
<td>0.90 (0.77,1.00)</td>
<td>0.66</td>
</tr>
<tr>
<td>Mattapan</td>
<td>0.50 (0.40,0.60)</td>
<td>0.97 (0.85,1.00)</td>
<td>0.96 (0.83,1.00)</td>
<td>1.00</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>0.50 (0.42,0.58)</td>
<td>0.77 (0.63,0.94)</td>
<td>0.77 (0.63,0.93)</td>
<td>0.58</td>
</tr>
<tr>
<td>Roxbury</td>
<td>0.74 (0.66,0.84)</td>
<td>0.98 (0.89,1.00)</td>
<td>0.98 (0.88,1.00)</td>
<td>0.82</td>
</tr>
<tr>
<td>South Boston</td>
<td>0.48 (0.40,0.57)</td>
<td>0.72 (0.61,0.86)</td>
<td>0.70 (0.58,0.82)</td>
<td>0.54</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>0.60 (0.52,0.68)</td>
<td>0.87 (0.76,0.97)</td>
<td>0.86 (0.75,0.96)</td>
<td>0.77</td>
</tr>
<tr>
<td>South End</td>
<td>0.67 (0.59,0.74)</td>
<td>0.82 (0.73,0.92)</td>
<td>0.80 (0.72,0.89)</td>
<td>0.71</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>0.78 (0.70,0.88)</td>
<td>0.89 (0.81,0.95)</td>
<td>0.89 (0.81,0.95)</td>
<td>0.78</td>
</tr>
</tbody>
</table>

### Table 3.16: Comparing Distance predictions with actual outcome for 2014 K2.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Naive (95 % C.I.)</th>
<th>Logit (95 % C.I.)</th>
<th>mLogit (95 % C.I.)</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>1.44 (1.23,1.66)</td>
<td>1.27 (1.10,1.47)</td>
<td>1.28 (1.09,1.50)</td>
<td>1.22</td>
</tr>
<tr>
<td>Charlestown</td>
<td>0.97 (0.70,1.26)</td>
<td>0.94 (0.76,1.12)</td>
<td>0.95 (0.78,1.14)</td>
<td>0.91</td>
</tr>
<tr>
<td>Downtown</td>
<td>1.30 (1.06,1.64)</td>
<td>1.23 (1.08,1.40)</td>
<td>1.23 (1.08,1.39)</td>
<td>1.28</td>
</tr>
<tr>
<td>East Boston</td>
<td>1.75 (1.60,1.93)</td>
<td>1.21 (1.05,1.44)</td>
<td>1.27 (1.06,1.51)</td>
<td>1.38</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>2.04 (1.89,2.20)</td>
<td>1.80 (1.67,1.94)</td>
<td>1.78 (1.64,1.92)</td>
<td>2.06</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>1.29 (1.19,1.39)</td>
<td>1.17 (1.07,1.26)</td>
<td>1.16 (1.07,1.25)</td>
<td>1.20</td>
</tr>
<tr>
<td>Mattapan</td>
<td>1.80 (1.69,1.93)</td>
<td>1.71 (1.61,1.83)</td>
<td>1.71 (1.60,1.84)</td>
<td>1.78</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>1.29 (1.17,1.42)</td>
<td>1.17 (1.08,1.27)</td>
<td>1.17 (1.07,1.27)</td>
<td>1.19</td>
</tr>
<tr>
<td>Roxbury</td>
<td>1.73 (1.62,1.82)</td>
<td>1.53 (1.44,1.63)</td>
<td>1.52 (1.43,1.62)</td>
<td>1.60</td>
</tr>
<tr>
<td>South Boston</td>
<td>1.37 (1.28,1.46)</td>
<td>1.21 (1.15,1.28)</td>
<td>1.21 (1.15,1.28)</td>
<td>1.32</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>1.46 (1.36,1.60)</td>
<td>1.21 (1.06,1.37)</td>
<td>1.20 (1.04,1.35)</td>
<td>1.43</td>
</tr>
<tr>
<td>South End</td>
<td>1.52 (1.43,1.61)</td>
<td>1.43 (1.35,1.51)</td>
<td>1.43 (1.35,1.51)</td>
<td>1.46</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>1.89 (1.76,2.02)</td>
<td>1.70 (1.57,1.83)</td>
<td>1.69 (1.56,1.82)</td>
<td>1.95</td>
</tr>
</tbody>
</table>

### Table 3.17: Comparing Unassigned predictions with actual outcome for 2014 K2.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Naive (95 % C.I.)</th>
<th>Logit (95 % C.I.)</th>
<th>mLogit (95 % C.I.)</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>4.07 (0.00,14.00)</td>
<td>1.28 (0.00,9.00)</td>
<td>1.18 (0.00,10.00)</td>
<td>28</td>
</tr>
<tr>
<td>Charlestown</td>
<td>0.68 (0.00,8.00)</td>
<td>1.25 (0.00,10.00)</td>
<td>1.38 (0.00,11.00)</td>
<td>6</td>
</tr>
<tr>
<td>Downtown</td>
<td>1.24 (0.00,5.00)</td>
<td>1.09 (0.00,7.00)</td>
<td>1.36 (0.00,9.00)</td>
<td>17</td>
</tr>
<tr>
<td>East Boston</td>
<td>47.22 (2.00,96.00)</td>
<td>49.22 (5.00,107.00)</td>
<td>49.00 (0.00,111.03)</td>
<td>96</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>2.90 (0.00,16.00)</td>
<td>5.09 (0.00,18.00)</td>
<td>5.12 (0.00,18.00)</td>
<td>27</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>18.15 (6.00,34.00)</td>
<td>1.68 (0.00,7.00)</td>
<td>2.07 (0.00,8.00)</td>
<td>34</td>
</tr>
<tr>
<td>Mattapan</td>
<td>17.73 (3.00,37.02)</td>
<td>3.96 (0.00,18.00)</td>
<td>5.17 (0.00,20.00)</td>
<td>47</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>18.79 (7.00,34.00)</td>
<td>2.07 (0.00,8.00)</td>
<td>3.02 (0.00,11.00)</td>
<td>29</td>
</tr>
<tr>
<td>Roxbury</td>
<td>4.66 (0.00,22.00)</td>
<td>3.69 (0.00,20.00)</td>
<td>4.27 (0.00,22.00)</td>
<td>35</td>
</tr>
<tr>
<td>South Boston</td>
<td>42.07 (20.98,67.03)</td>
<td>4.24 (0.00,15.00)</td>
<td>5.97 (0.00,18.00)</td>
<td>66</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>8.93 (1.00,20.00)</td>
<td>0.35 (0.00,4.00)</td>
<td>0.63 (0.00,5.00)</td>
<td>9</td>
</tr>
<tr>
<td>South End</td>
<td>19.80 (3.00,49.00)</td>
<td>10.52 (0.00,36.00)</td>
<td>11.18 (0.00,39.00)</td>
<td>107</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>9.30 (0.00,22.00)</td>
<td>4.03 (0.00,14.00)</td>
<td>4.70 (0.00,15.00)</td>
<td>23</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>5.43 (0.00,25.00)</td>
<td>8.78 (0.00,30.00)</td>
<td>8.42 (0.00,26.02)</td>
<td>49</td>
</tr>
</tbody>
</table>
3.B Detailed Participation Forecasts

Table 3.18: Forecast for total number of new applicants in 2014.

<table>
<thead>
<tr>
<th>Total New Applicants</th>
<th>Predicted Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5798.5</td>
<td>366.4</td>
</tr>
</tbody>
</table>

Table 3.19: Forecast for proportion of the new applicants distributed to each grade and neighborhood in 2014.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Predicted Mean</th>
<th>Standard Deviation</th>
<th>Predicted Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K1</td>
<td></td>
<td>K2</td>
<td></td>
</tr>
<tr>
<td>Allston-Brighton</td>
<td>0.0177</td>
<td>0.0016</td>
<td>0.0150</td>
<td>0.0022</td>
</tr>
<tr>
<td>Charlestown</td>
<td>0.0136</td>
<td>0.0006</td>
<td>0.0190</td>
<td>0.0013</td>
</tr>
<tr>
<td>Downtown</td>
<td>0.0155</td>
<td>0.0027</td>
<td>0.0198</td>
<td>0.0006</td>
</tr>
<tr>
<td>East Boston</td>
<td>0.0533</td>
<td>0.0047</td>
<td>0.0577</td>
<td>0.0067</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>0.0315</td>
<td>0.0021</td>
<td>0.0206</td>
<td>0.0008</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>0.0327</td>
<td>0.0038</td>
<td>0.0210</td>
<td>0.0007</td>
</tr>
<tr>
<td>Mattapan</td>
<td>0.0308</td>
<td>0.0024</td>
<td>0.0262</td>
<td>0.0021</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>0.0245</td>
<td>0.0031</td>
<td>0.0179</td>
<td>0.0035</td>
</tr>
<tr>
<td>Roslindale</td>
<td>0.0451</td>
<td>0.0027</td>
<td>0.0267</td>
<td>0.0030</td>
</tr>
<tr>
<td>Roxbury</td>
<td>0.0566</td>
<td>0.0040</td>
<td>0.0522</td>
<td>0.0027</td>
</tr>
<tr>
<td>South Boston</td>
<td>0.0136</td>
<td>0.0010</td>
<td>0.0138</td>
<td>0.0013</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>0.0630</td>
<td>0.0041</td>
<td>0.0460</td>
<td>0.0036</td>
</tr>
<tr>
<td>South End</td>
<td>0.0190</td>
<td>0.0020</td>
<td>0.0207</td>
<td>0.0007</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>0.0408</td>
<td>0.0036</td>
<td>0.0223</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Table 3.20: Forecast for continuing ratios for each grade and neighborhood in 2014.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Predicted Mean</th>
<th>Standard Deviation</th>
<th>Predicted Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K1</td>
<td></td>
<td>K2</td>
<td></td>
</tr>
<tr>
<td>Allston-Brighton</td>
<td>0.81</td>
<td>0.04</td>
<td>0.78</td>
<td>0.07</td>
</tr>
<tr>
<td>Charlestown</td>
<td>0.72</td>
<td>0.30</td>
<td>0.76</td>
<td>0.10</td>
</tr>
<tr>
<td>Downtown</td>
<td>0.52</td>
<td>0.35</td>
<td>0.72</td>
<td>0.02</td>
</tr>
<tr>
<td>East Boston</td>
<td>0.85</td>
<td>0.09</td>
<td>0.85</td>
<td>0.03</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>0.72</td>
<td>0.19</td>
<td>0.79</td>
<td>0.04</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>0.77</td>
<td>0.40</td>
<td>0.78</td>
<td>0.06</td>
</tr>
<tr>
<td>Mattapan</td>
<td>0.83</td>
<td>0.17</td>
<td>0.80</td>
<td>0.02</td>
</tr>
<tr>
<td>North Dorchester</td>
<td>0.71</td>
<td>0.20</td>
<td>0.77</td>
<td>0.11</td>
</tr>
<tr>
<td>Roslindale</td>
<td>0.63</td>
<td>0.32</td>
<td>0.81</td>
<td>0.04</td>
</tr>
<tr>
<td>Roxbury</td>
<td>0.84</td>
<td>0.31</td>
<td>0.79</td>
<td>0.03</td>
</tr>
<tr>
<td>South Boston</td>
<td>0.53</td>
<td>0.46</td>
<td>0.78</td>
<td>0.08</td>
</tr>
<tr>
<td>South Dorchester</td>
<td>0.68</td>
<td>0.06</td>
<td>0.79</td>
<td>0.03</td>
</tr>
<tr>
<td>South End</td>
<td>0.69</td>
<td>0.27</td>
<td>0.80</td>
<td>0.06</td>
</tr>
<tr>
<td>West Roxbury</td>
<td>0.81</td>
<td>0.39</td>
<td>0.80</td>
<td>0.05</td>
</tr>
</tbody>
</table>

100
3.C Calibrating Mixed Logit using MCMC

Unlike in the simple Logit model, the log likelihood function associated with the mixed logit model is difficult to evaluate directly, involving many multi-dimensional integrals. Hence, we calibrate it using Markov Chain Monte Carlo (MCMC) instead of maximum likelihood.

The basic framework to calibrate mixed logit models using MCMC have been established in previous works such as Train (2003), and is based on Gibbs sampling and the Metropolis-Hasting algorithm. However, the examples there do not have so many fixed coefficients as we have. (We have a fixed effect for every school and over 80 schools, and since only the relative difference between fixed effects matter for preferences, but not their absolute values, these fixed effects are likely to be highly correlated.) It is known that the simple Metropolis Hasting with random walk proposals does not perform well when estimating many dimensions (see Katafygiotis and Zuev (2008)), especially if the dimensions are correlated. So we modify the framework to use Metropolis-Within-Gibbs (MWG), which samples blocks of coordinates iteratively and not everything at once, and Hamiltonian Monte Carlo (HMC), which incorporates gradient information to suggest good directions to sample. We will describe these methods in Section 3.C.2.

3.C.1 Specifying the Likelihood Function

The first step of applying MCMC techniques is specifying the full likelihood function of observing the data given the model parameters. To do this, we will restate the mixed logit model in a more precise way and clearly define the parameters.

The model of interest is as follows. For student $i$ and program $j$, let $F_{ij}$ be a vector in which the corresponding components correspond to the following features for this student-program pair: “continuing,” “sibling,” “ell language match,” “distance*black/hispanic,” “distance*income est.,” “mcas*black,” “mcas*income est.,” “% white/asian*black/hispanic,” and “% white/asian*income est.” Let $G_{ij}$ be a vector with components corresponding to the following features: “ell match,” “walk zone,”
“distance,” “mcas” and “% white/asian.” The explanations for these features are in Section 3.3.1. The features in $F_{ij}$ are assumed to have the same coefficients for all students, while the features in $G_{ij}$ correspond to features that have random coefficients. These random coefficients are in turn partitioned into 3 blocks, with the first block being “ell match,” the second “walk zone,” and the third “distance,” “mcas,” and “% white/asian.” The coefficients within each block are arbitrarily correlated, while the coefficients across blocks are independent. There is also a fixed effect for every school, capturing its general, common attractiveness. (Note that this departs from the notation in Section 3.3.1 as $F_{ij}$ there corresponds to not only the above but also all of the features in $G_{ij}$ and also the school-specific fixed-effects. This current notation makes the following exposition easier.)

More precisely speaking, the utility of student $i$ for program $j$ in school $s(j)$ is

$$u_{ij} = \tilde{\alpha}_{s(j)} + \beta \cdot F_{ij} + \gamma_i \cdot G_{ij} + \epsilon_{ij},$$

where vector $\tilde{\alpha} = \begin{pmatrix} \alpha^T & 0 \end{pmatrix}$ corresponds to the school fixed-effects. (The last component is normalized to zero because only relative differences between fixed effects drive utilities.) $\beta$ corresponds to the other fixed coefficients. $\gamma_i$ is the instantiation of the random coefficients for student $i$, and $\epsilon_{ij}$ is an unobservable idiosyncratic taste shock. The terms $\gamma_i$ are i.i.d. and distributed $\text{Normal}(b, W)$, where $b$ is the mean vector and $W$ is the covariance matrix. The idiosyncratic shocks $\epsilon_{ij}$ are i.i.d. standard Gumbel distributed. Because of our partition of the random coefficients into 3 blocks, with the blocks having 1, 1 and 3 variables respectively, the covariance matrix can be written in the block diagonal form

$$W = \begin{pmatrix} W_1 & & \\ & W_2 & \\ & & W_3 \end{pmatrix},$$

where $W_1$, $W_2$ and $W_3$ are $1 \times 1$, $1 \times 1$ and $3 \times 3$ symmetric positive definite matrices. In summary, the parameters to estimate are $\alpha$, $\beta$, $b$, $W_1$, $W_2$ and $W_3$. 

102
The data to fit these parameters are the observed choices of every student along with the observed characteristics $F_{ij}$ and $G_{ij}$. Suppose that student $i$ makes $m_i$ choices, and let the chosen programs from best to worst be $y_{i1}, y_{i2}, \ldots, y_{im_i}$. Given the instantiation of $\gamma_i$, the likelihood function on $\alpha$ and $\beta$ is

$$\phi_i(\alpha, \beta | \gamma_i) = \prod_{c=1}^{m_i} \frac{\exp(\tilde{\alpha} s(y_{ic}) + \beta \cdot F_{icyc} + \gamma_i \cdot G_{icyc})}{\sum_{d=1}^{m_i} \exp(\tilde{\alpha} s(y_{id}) + \beta \cdot F_{icyd} + \gamma_i \cdot G_{icyd})}. \quad (3.1)$$

This is the logit likelihood function. (Recall that $\tilde{\alpha} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ is the fixed effects including the last school that is normalized to zero, and $s(j)$ denotes the school where program $j$ is in.)

The full likelihood function incorporating all the data is

$$\Phi(\alpha, \beta, b, W) = \prod_{i=1}^{n} \int_{\mathbb{R}^5} \phi_i(\alpha, \beta | \gamma_i) \exp\left(-\frac{1}{2}W^{-1}\|\gamma_i - b\|^2\right) d\gamma_i. \quad (3.2)$$

Here, $n$ is the number of students; recall that the random coefficients $\gamma_i$ each has five dimensions.)

Our estimates will be based on sampling the parameters based on this likelihood function $\Phi$. Because $\Phi$ is complex, we do this by MCMC. As a detour, we will give an overview of MCMC and the specific techniques we use. Readers who are familiar with these techniques can jump to Section 3.C.3.

3.C.2 Overview of the MCMC

The idea behind Markov Chain Monte Carlo (MCMC) is to samples from a distribution by constructing a Markov chain whose unique stationary distribution is the desired distribution of interest. So if the chain is easy to simulate and if it is fast-mixing, meaning that it converges quickly to the stationary distribution, then we can sample by simply simulating the chain. After throwing out a so-called “burn-in” period at the beginning, we would have arrived at samples from the desired distribution.

The work horse of MCMC are Gibbs sampling and Metropolis-Hasting. Gibbs sampling is used when the desired distribution can be factored into several marginal
distributions that are easier to sample. For example, to sample from a joint distribution on $x$, $y$ and $z$, one might iteratively sample one variable at a time conditional on the other ones. Precisely speaking, we initialize $x^0$, $y^0$ and $z^0$ arbitrarily. For each $t \geq 1$, sample iteratively from the following conditional distributions:

$$
\begin{align*}
x^t &| y^{t-1}, z^{t-1} \\
y^t &| x^t, z^{t-1} \\
z^t &| x^t, y^t
\end{align*}
$$

After a sufficient number $S$ of samples, and after throwing out the initial burn-in of $B$ samples, $\{(x^t, y^t, z^t) : B < t \leq S\}$ would approximate samples from the original distribution, although successive samples are not independent. One can remove the serial correlation by either sampling independently from this set, or by keeping only samples in which $t$ is a multiple of $\Delta$, where $\Delta$ is a chosen positive integer.

Metropolis-Hasting is a technique to sample from an arbitrary distribution with given likelihood function $L(x)$. There are many variants, but the common idea is to use a proposal distribution that is easy to sample from and reject certain samples to get the likelihood ratios to be correct. The proposal distribution may depend on the current iterate $x$. Let transition probability density be $T(y|x)$; this is the probability density of proposing $y$ given that the current sample is $x$. In order to obtain the correct likelihoods, we can only accept a fraction of the samples proposed, and reject the others. The probability that we accept proposal $y$ given the previous iterate being $x$ is

$$
A(y|x) = \min(1, \frac{L(y)T(x|y)}{L(x)T(y|x)}).
$$

Note that if $T(y|x)$ is proportional to $L(y)$, then the acceptance probability is always 1 as the proposal distribution already matches the target. Otherwise, the above formula is tuned so that the following identity, called “detailed balance” in the literature, holds:

$$
L(x)T(y|x)A(y|x) = L(y)T(x|y)A(x|y).
$$

This equation guarantees that the desired density $p(x)$ is a stationary distribution
of the Markov chain induced by the proposal and acceptance process. Furthermore, if the chain is ergodic, which is true for example if the proposal distribution has full support, then $p(x)$ is the only stationary distribution.

The sampling procedure is then to initialize $x^0$ arbitrarily, and for each $t \geq 1$

1. Draw $y$ according to $T(y|x^{t-1})$.

2. Set $x^t = \begin{cases} y & \text{with prob. } A(y|x^{t-1}), \\ x^{t-1} & \text{otherwise.} \end{cases}$

By iterating this many times and discarding sufficiently many burn-in samples, we would have arrived at the desired distribution.

Because of the flexibility in the proposal distributions, there are many variants of the above techniques. The goal is to find a proposal distribution that strikes a good balance of being easy to sample from and approximating the target distribution locally. If it is not easy to sample from, then each step would take too long; if it is too far from the target distribution, then the acceptance probabilities would be very low and the chain may get stuck at a certain iterate for a very long time. In the following sections we present the three variants we use: Random Walk Metropolis (RWM), Metropolis-Within-Gibbs (MWG), and Hamiltonian Monte Carlo (HMC).

**Random Walk Metropolis (RWM)**

This method is the easiest to sample from, as it uses a simple random walk to propose the next value: if the current iterate is $x$, it proposes $y = x + \epsilon$, where $\epsilon$ is multivariate normal distributed, $\epsilon \sim \text{Normal}(0, \rho I)$, where $I$ is the identity matrix and $\rho$ is a scale parameter. Other covariance matrices can also be used instead of the identity but it must be the same for every $x$. The scale parameter is tuned to match the overall variance of the desired distribution. Too small a $\rho$ and successive samples and there will be too much serial correlation; too large a $\rho$ and acceptance probability might be near zero so the chain may get stuck. We tune $\rho$ by multiplying it up or down so that the average acceptance ratio since last tuning is between 0.4 and 0.6, which
is the ball park value suggested by the literature.\textsuperscript{15} The number of steps we wait before tuning increases exponentially, so that after our burn-in sample until our last iteration there is no tuning.

This method performs well when the target distribution has not too many dimensions, and has approximately the same scale in each dimension. However, when there are many dimensions, it becomes exponentially harder to guess the right direction, and the method may take very long to converge; when there are dimensions that are at very different scales, then there may exist no \( \rho \) that is good for all dimensions.

**Metropolis Within Gibbs (MWG)**

This is a simple extension of RWM that allows various sub-blocks of coordinates to have different scales. It is simply to sample each sub-block iteratively, conditional on the others, much like running several RWM within a Gibbs sampling framework. This also reduces the number of dimensions sampled at each step. The drawback is that more samples are needed.

Precisely speaking, instead of sampling all dimensions of vector \( x \) simultaneously, write it in terms of sub-vectors \( x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \). Each sub-vector may represent several coordinates. Initialize \( x^0 \) arbitrarily and for \( t \geq 1 \), sample

\[
\begin{align*}
    x^t_1 & \mid x^{t-1}_{2}, \ldots x^{t-1}_{k} \\
    x^t_2 & \mid x^t_1, x^{t-1}_{3}, \ldots x^{t-1}_{k} \\
    \vdots \\
    x^t_k & \mid x^t_1, \ldots x^t_{k-1}
\end{align*}
\]

Each of the above is sampled using RWM, perhaps with different scale parameters for different sub-vectors. In each Gibbs iteration, for each of the variables, we only take one step of Metropolis-Hasting, which involves one proposal and possible acceptance. Because of detailed balance, embedding Metropolis-Hasting into Gibbs

\footnote{See \cite{Roberts1997}.}
sampling in this way also works.

**Hamiltonian Monte Carlo (HMC)**

This method uses the gradient of the log likelihood function to inform the proposals, which can significantly improve the acceptance probabilities in high dimensions. The drawback is that each iteration is slower as several gradient calls is needed. The method is motivated by Hamiltonian dynamics in physics. It models the current iterate $x$ as a location vector, and treats the negative log likelihood function as an energy potential. In each step, it samples a random momentum vector and simulates the trajectory of the object by discretizing time and alternatively updating the momentum using the potential function and updating the position using the momentum. To make detailed balance work out, the first and last steps of simulation are half-steps. Precisely speaking, let the gradient of the log likelihood function be $G(x) = \nabla (\log(L(x)))$. Let $\epsilon$ and $\Delta$ be tuning parameters, representing the discretization in time and the number of steps to simulate respectively. The proposal is based on the pseudocode in this algorithm (this is taken from Neal (2011)):

**Algorithm 1** Pseudocode for one step of HMC

Function HMC_STEP($x$):
Draw momentum $p_0 \sim \text{Normal}(0, I)$.
Initialize $y = x, p = p_0$.
Update $p = p - \epsilon G(y)/2$.
for $\Delta - 1$ iterations do
  Update $y = y + \epsilon p$
  Update $p = p - \epsilon G(y)$.
end for
Update $y = y + \epsilon p$.
Update $p = p - \epsilon G(y)/2$.
return $\begin{cases} y \text{ with prob. } A(y|x) = \min(1, \frac{L(y)}{L(x)} \exp(\frac{||p_0||^2 - ||p||^2}{2})) \\ x \text{ otherwise} \end{cases}$

Note that the chance of proposing $y$ given $x$ is simply the chance of drawing momentum $p_0$. Moreover, by the reversibility of the intermediate steps of discrete simulation, if we started at $y$ and drew a momentum of $-p$ (where $p$ is the final
momentum vector in HMC_STEP), then the proposal would be $x$. This implies that

$$
\frac{T(y|x)}{T(x|y)} = \frac{\exp(-\frac{1}{2}\|p_0\|^2)}{\exp(-\frac{1}{2}\|p\|^2)},
$$

which implies that

$$\frac{T(y|x)A(y|x)}{T(x|y)A(x|y)} = \frac{\exp(-\frac{1}{2}\|p_0\|^2) L(y)}{\exp(-\frac{1}{2}\|p\|^2) L(x)} \exp(\frac{\|p_0\|^2 - \|p\|^2}{2}) = \frac{L(y)}{L(x)}.$$

So detailed balance holds and the following is a valid Metropolis-Hasting sampler: Initialize $x^0$ arbitrarily. For $t \geq 1$, set $x^t = \text{HMC\_STEP}(x^{t-1})$.

One can show that as the time discretization $\epsilon \to 0$, for any fixed total simulation time $\epsilon \Delta$, the acceptance probability goes to 1. Hence, we would like $\epsilon$ to be small enough so the chain does not get stuck and $\epsilon \Delta$ large enough so that successive samples are not too serially correlated. In practice, we fix $\Delta = 20$ and tune $\rho$ so that the empirical acceptance rate since last tuning is between 0.5 and 0.8. As before, we increase the interval between tuning times exponentially so that no tuning happens in the sample we keep (after burn-in and before the last iteration). Another detail is that to prevent cases in which $\epsilon \Delta$ is exactly what makes the proposal $y$ go back to original point $x$, instead of using the same $\epsilon$, we draw $\tilde{\epsilon} \sim \text{Uniform}(0.85\epsilon, 1.15\epsilon)$ before each call to HMC_STEP, and use $\tilde{\epsilon}$ as the step size throughout that call. Because this distribution is a-priori fixed, detailed balance is also preserved. All these are according to the best practices for applying HMC as outlined in [Neal, 2011].

### 3.C.3 Our MCMC Sampler

Our MCMC procedure is based on the one in [Train, 2003] but breaking up the estimation of the fixed coefficients into two steps, one step using Hamiltonian Monte Carlo (HMC) and the other Metropolis Within Gibbs (MWG). We use HMC to estimate the school fixed effects and MWG to estimate the other fixed coefficients. These techniques allow us to accommodate the large number of school fixed effects and the unequal scales across the other fixed coefficients.
To sample from the full likelihood function \( \Phi(\alpha, \beta, b, W) \) (Equation 3.2), we initialize \( \alpha^0, \beta^0, b^0, W_1^0, W_2^0, W_3^0 \) arbitrarily. For each \( t \geq 1 \), we do a few layers of Gibbs sampling. In some of the layers we embed a form of Metropolis-Hasting; but in each Gibbs iteration we only take one step of Metropolis-Hasting, much as it is in MWG. Furthermore, let \( T \) be a parameter indicating how long we wait before tuning. We initialize \( T \) to be 1 and increase this parameter steadily, so that tuning becomes exponentially less frequent. For \( t \geq 1 \), each MCMC step is as follows:

1. Draw \( \gamma_i^t|\alpha^{t-1}, \beta^{t-1}, b^{t-1}, W^{t-1} \). This is done using one iteration of RWM with likelihood function

\[
L(x) = \phi_i(\alpha^{t-1}, \beta^{t-1}, x) \exp\left( -\frac{1}{2} (W^{t-1})^{-1} |x - b^{t-1}|^2 \right)
\]

and starting value \( \gamma_i^{t-1} \). (See Equation 3.1 for definition of \( \phi_i \).) We initialize \( \rho = 0.05 \) and initially to tune for each \( i \) every Uniform(1000\( T \), 1500\( T \)) steps.

2. Draw \( b^t|\gamma_i^t, W^{t-1} \). This is sampling from Normal\( (\frac{1}{n} \sum_{i=1}^{n} \gamma_i^t, \frac{1}{m} W^{t-1}) \).

3. Draw \( W^t|\gamma_i^t, b^t \). This can be done as follows: For \( l \in \{1, 2, 3\} \), let \( C_l^t \) be the covariance matrix of the \( l \)th block of \( \gamma_i^t \) assuming mean as in the \( l \)th block of \( b^t \). (Recall that the random coefficients are organized into 3 blocks, with ell match being the first block, walk zone being the second, and distance, mcas, and % white/asian being the third.) Let \( k_l \) be the number of variables in the \( l \)th block. Draw \( W_l^t \) according to the Inverse Wishart Distribution with degree of freedom \( \nu = k_l + n \) and scale matrix \( \Psi = k_l I_{l \times l} + n C_l^t \).

4. Draw \( \alpha^t|\gamma_i^t, \beta^{t-1} \). This is done using one step of HMC with likelihood function

\[
L(x) = \prod_{i=1}^{n} \phi_i(x, \beta^{t-1}|\gamma_i^t).
\]

We initialize \( \epsilon = 0.015 \), and \( \Delta = 20 \). We tune every 1000\( T \) steps.

5. Draw \( \beta^t|\gamma_i^t, \alpha^t \). This is done using one iteration of MWG with likelihood
function

\[ L(x) = \prod_{i=1}^{n} \phi_i(\alpha_i^t, x|\gamma_i^t). \]

We break the fixed coefficients $\beta$ into 6 subvectors: 1) “continuing;” 2) “sibling;” 3) “ell language match;” 4) “distance*black/hispanic” and “distance*income est.”; 5) “mcas*black” and “mcas*income est.”; 6) “% white/asian*black/hispanic” and “% white/asian*income est.” We initialize the scales $\rho$ for each subvector to be .5, .5, .1, .1, .5, and .5 respectively. We tune every Uniform$(100T, 150T)$ steps.

We run these steps 1,000,000 times, increasing the tuning interval parameter $T$ by a factor of 1.2 every 5000 iterations. We throw out the first 500,000 iterations as burn-in. Note that in the interval we keep, no tuning happens. This ensures the correctness of the Markov chain in this period.

For robustness check, we re-ran this procedure 6 times, sometimes with different initial values, and we found near identical results each time.
Chapter 4

Optimizing Menus and Priorities to Improve Welfare Subject to Busing Constraint

4.1 Introduction

In many public school systems across the US, school choice has become the preferred alternative to the traditional method of assigning each student to a designated school based on home location. In a typical school choice system, each student is given a set of school options, which we refer to as the student’s menu, and submits a preference ranking of the schools in this menu. These preference rankings are collected by the school board many months before the school year starts, and the school board computes the school assignment using a centralized algorithm, which takes into account possible priorities between various types of students and admission quotas at schools.

In Boston, New York City, Chicago, Denver, New Orleans, Washington DC, among other cities, the assignment algorithm is the student-proposing deferred acceptance algorithm, originally proposed by Gale and Shapley (1962) and applied to school choice by Abdulkadiroğlu and Sönmez (2003b). The algorithm computes a stable matching, which means that no student is rejected by a school that has left over seats,
and that no student is rejected by a school that accepted another with a lower priority. Moreover, the algorithm is strategyproof, meaning that students have no incentives to misreport their preference rankings. This incentive property was especially important in the adoption of this algorithm in New York and Boston (see Abdulkadiroğlu et al. (2006) and Abdulkadiroğlu et al. (2009)).

In the student-proposing deferred acceptance algorithm (which we simply refer to as the DA algorithm), there are three important policy levers. The first is the menu of each student, which is the set of schools each student can rank in the preference submission. We assume that the menus are only determined by the student’s observable characteristics, such as home location and special needs status, and we refer to the observable characteristics as the student’s type. The second lever is how schools prioritize between different students. We assume that every student is given a number at every school, called the student’s priority at the school, and in situations of conflict the student with the higher priority is admitted. We assume that the vector of priorities for each student is drawn from a priority distribution which depends on the student’s type. The third lever is a upperbound at each school for the number of students admitted, which we call the school’s quota.

Almost all existing literature treat the menus, priority distributions, and quotas as exogenously given, over which the social planner has little control. However, in practice, these policy levers are set by the school board to induce desirable outcomes. This is because in most public school districts, the school board owns all the schools and has centralized control over menus, priority distributions, and quotas.

For example, in 2004, 2009, and 2012, Boston Public Schools tried three times to reform students’ menus in order to achieve a better balance of variety of choice, equity, and school busing costs. Previously, the city used a 3-zone plan, which divided the city into 3 geographic zones and allowed students to rank any number of schools in their zone. This gave students about 30 schools to rank, but resulted in a staggering busing cost of about $80 million per year, since each school had to pick up children from one-third of the city (See Russell and Ebbert (2011)). Furthermore, there were perceived inequities in the distribution of schools across zones. In both 2004 and 2009,
the school board proposed alternative zoning plans that divided the city into more zones and so reduced the menus and the bussing burden. However, none of these plans were approved because of equity concerns. In 2013, after conducting an extensive simulation study comparing various proposed plans on a range of metrics (see Pathak and Shi (2013), Shi (2015)), the city finally approved a new proposal, called the Home-Based Plan, which gave students about 9-15 choices around their home based on proximity and standardized test scores. Furthermore, the city revised the priority distributions to no longer depend on whether students live within a one-mile radius of the school, called the school’s walk-zone. Hence, the menus and priority distributions adopted in Boston in 2013 are not exogenously given, but carefully determined policy levers to induce a desirable outcome.

This chapter studies how to systematically optimize the menus, priority distributions, and quotas in the DA algorithm to maximize a given objective function, which may take into account the expected utilities of students, equity measures, socio-economic diversity, and bussing costs. The input of the optimization is the school board’s objective function, a distribution for the number of students of each type, and a distribution of utilities for students of each type. (The population distribution and utility distributions can be estimated from past data.) The output is the optimized menu and priority distribution for each type of student, as well as the optimized quota for each school.

The optimization methodology is based on a new connection between stable matching and assortment planning. We first define an alternative, simpler framework of assigning students, which we call a random assortment mechanism. In this mechanism, each student is given a set of schools, which we call an assortment, and the student is assigned to his/her favorite school in the assortment. The assortment of a student may be random, but the distribution over assortments is dependent only on the type of the student. This mechanism is characterized by the probability of giving each type of student each assortment of schools, which we call assortment probabilities. We show that the DA algorithm, given menus, priority distributions and quotas, can be equivalently expressed as a random assortment mechanism with certain assortment
probabilities. This implies that when we define a new optimization problem of finding a random assortment mechanism that maximizes the school board’s objective, this problem is a relaxation of the original optimization over menus, priority distributions and quotas.

In the other direction, given the utility distribution for each type of student, we show that any random assortment mechanism can be expressed as the asymptotic behavior of the DA algorithm with certain fixed menus, priority distributions and quotas, in the limit when both the number of students of each type and the quota for each school is scaled proportionally to infinity. We call this asymptotic behavior a large-market limit, and it is based on the large-market model of stable matching of Azevedo and Leshno (2015). We augment their theory by showing an equivalent characterization of large-market stable matching as a random assortment mechanism. The proof of this yields a mapping between assortment probabilities and menus, priority distributions and quotas that are equivalent in the large-market limit.

The optimization methodology is as follows: we first optimize the school board’s objective over the space of random assortment mechanisms, and then use the optimal assortment probabilities and the above mapping to construct menus, priority distributions and quotas, which we take as the output. The outputted policy levers may no longer be exactly optimal under a finite number of students, but we expect them to be approximately optimal if the number of students is large, which is the case in practice.

The optimization over random assortment mechanisms can be formulated as a convex optimization problem, which is similar in structure to the choice-based linear program (CBLP) studied in the network revenue management literature (Liu and van Ryzin (2008)). The convex optimization has exponentially many variables, but it can still be solved efficiently by repeatedly solving a sub-problem. This sub-problem is to find an assortment of schools for a specific type of students, such that if each student picks his/her favorite school among the assortment, and if assigning each seat of a school incurs a certain externalities, then the weighted sum of the expected utility of students of this type plus the expected externalities is maximized. We call this the
socially-optimal assortment planning problem, and show that it is a strict generalization of the revenue-maximizing assortment planning problem studied in the revenue management literature. We show that many techniques from the revenue-maximizing case can be extended to the socially-optimal case, including efficient algorithms for multinomial logit (MNL), nested logit, and Markov chain based choice models.

To demonstrate the effectiveness of this methodology, we apply it on real data from Boston Public Schools (BPS). Specifically, we take the menus and priority distributions from the Home-Based plan adopted in the 2012-2013 school assignment reform, and produce optimized menus and priority distributions that induce higher expected utilities for students while using less school busing, in terms of both the average distances traveled by students, and the average area schools have to cover to pick up students from. All of the evaluations are done by discrete simulations, and do not rely on large-market assumptions. We show that the optimized plan also performs well in aspects that do not explicitly appear in the optimization, including high chances for students to get their top choices in menu, and high chances for neighbors to go to the same school. We show that these results are robust to moderate changes in the population and utility distributions.

4.1.1 Related Literature

Optimization models of school assignment trace back to Clarke and Surkis (1968), Belford and Ratliff (1972), and Franklin and Koenigsberg (1973). However, these papers do not allow students to choose, but assume that the assignment is completely determined by the student’s home location. In their formulation, students are partitioned into local neighborhoods, and the goal is to find an assignment of neighborhood to schools in order to minimize the total distances traveled by students, subject to capacity, diversity, and other constraints. Sutcliffe et al. (1984) survey this literature and propose a formulation that treats the constraints as not necessarily binding but desirable goals that incur a penalty for being not fulfilled. Lemberg and Church (2000) study the dynamic problem of how to update zone boundaries smoothly across years to reflect changes in demographics and capacity. Caro et al. (2004) combine
the optimization with a Geographic Information System (GIS) for easy visualization and better treatment of spatial information.

Previous literature on school choice systems have focused on finding assignment mechanisms that achieve certain axiomatic properties, and much of it supports using a variant of the DA algorithm. Abdulkadiroğlu and Sönmez (2003b) show that while both the DA algorithm and another algorithm called top trading cycles (TTC) are strategyproof, they differ in their trade-off between efficiency and fairness: DA achieves Pareto-optimality among assignments that respect the priorities, while TTC achieves Pareto-optimality among all assignments. Abdulkadiroğlu et al. (2005a) apply this theory to the school assignment system in New York city (NYC), and their work helped NYC adopt the DA algorithm in 2004. In a follow-up paper, Abdulkadiroğlu et al. (2009) review the NYC data and argue that having a strategyproof algorithm is important to allow students to participate straightforwardly. In Boston, the algorithm before 2005 was not strategyproof. The algorithm, called the Boston Mechanism in the literature, had schools prioritize students who rank them earlier in their preferences, and gave students incentives to strategize in their preference submissions. Abdulkadiroğlu et al. (2005b) argue for the adoption of a strategyproof algorithm in Boston, and their work led to Boston adopting the DA algorithm in 2005. However, Miralles (2012) and Abdulkadiroğlu et al. (2011) argue that the Boston Mechanism may sometimes yield a more efficient assignment because it can better elicit the preference intensities of students. Erdil and Ergin (2008) and Abdulkadiroğlu et al. (2015) suggest ways to modify the DA algorithm to improve in efficiency.

However, the vast majority of previous literature treats the menus, priority distributions, and quotas as exogenous inputs, rather than potential policy levers. There are two strands of literature that are exceptions: the first strand considers how to implement priorities or quotas in the DA algorithm to induce diversity in the assignment, such as socio-economic or racial diversity. However, this literature focuses on how to implement a certain type of priority or quota, rather than what priorities or quotas to set. Abdulkadiroğlu and Sönmez (2003b) extend both DA and TTC
to allow quotas for various types of students at each school. Kojima (2012) show that there are cases in which these quotas may hurt the minority students they are intended to help. Hafalir et al. (2013) propose a way to remedy this by having soft bounds that may be violated depending on the preference of students, rather than hard bounds that always have to be satisfied. Ehlers et al. (2014) generalize these soft bounds to multiple types of students. Kominers and Sönmez (2015) study how to implement more complex diversity constraints in the DA algorithm by sub-dividing each school into slots and having possibly different priorities for students in each slot. Dur et al. (2013) demonstrates the impact of altering the order these slots are processed within the DA algorithm, and argue that this understanding led to the removal of the walk-zone priority in Boston in 2013.

The second strand of literature compares two specific priority distributions and discuss their impact on students. This current chapter is different in that we optimize over all possible priority distributions, rather than compare only two. The two priority distributions compared in the literature both involve a combination of a strict priority between categories of students, and a tie-breaking rule within each category. The differences are in the tie-breaking rule. In single tie-breaking (STB), a single random number is given to each student, which is the same at every school. In multiple tie-breaking (MTB), a different random number is given to each student at each school, and the random numbers are independently drawn. Abdulkadrioglu et al. (2009) compare STB and MTB by simulation using data from New York, and find that while neither of these dominate the other, STB gives more students their top choice, while MTB leaves fewer students unassigned. Ashlagi et al. (2015) and Ashlagi and Nikzad (2015) prove that similar results hold in general in a theoretical model with large number of students and seats. Arnosti (2015) show similar results in a theoretical model with a different asymptotic setup.

The most similar work to this current work is Ashlagi and Shi (2015), which is also motivated by optimizing the assignment system in Boston. However, this chapter is different in three important aspects. Firstly, Ashlagi and Shi (2015) seek to optimize over all possible mechanisms, and for tractability introduce additional assumptions,
which they show imply using the DA algorithm with single tie-breaking; this current chapter assumes that the DA algorithm will be used, and optimize over all possible priority distributions, which include multiple tie-breaking as well as more complicated priority structures. Secondly, we develop the connection with assortment planning in greater detail, and this allows us to leverage recent algorithms from the revenue management literature to handle more complex utility distributions and constraints. Thirdly, in the empirical exercise, we consider not only the average distance students travel, but also the area buses have to cover to pick up students, and this provides a more realistic model of transportation burden. This ability to bound the bus coverage area is only made possible through the new connection with assortment planning, and through the algorithm we develop for the socially-optimal assortment planning problem with cardinality constraints.

In terms of methodology, our work draws upon the large-market theory of stable matching developed by Azevedo and Leshno (2015) and Abdulkadiroğlu et al. (2015). The techniques used to solve the socially-optimal assortment planning are based on algorithms developed for the revenue-maximizing case, including the algorithms for the MNL model by van Ryzin and Mahajan (1999) and Rusmevichientong et al. (2010), for the nested logit model by Gallego and Topaloglu (2014) and Li et al. (2015), and for the Markov chain model by Blanchet et al. (2013) and Feldman and Topaloglu (2014b).

4.2 Model

There is a finite number of students to be assigned to a finite number of schools. The student population is partitioned into \( l \) types, according to observable characteristics such as residential neighborhood, special needs, test scores, etc. For each \( t \in [l] = \{1, 2, \ldots, l\} \), let \( n_t \) be the number of students of type \( t \). We assume that the vector \( \vec{n} \) is drawn from a known population distribution, denoted \( H \). Let \( m \) be the number of schools, and let the set of schools be \([m] = \{1, 2, \ldots, m\}\). We assume that there is also an outside option denoted by \( 0 \). Let \( \Omega = [m] \cup \{0\} \) be the set of all possible
options. Each student is to be assigned to exactly one of these options.

For each type $t \in [l]$, let $I_t$ be the set of students of type $t$. The preference of student $i \in I_t$ for schools is described by a utility vector $\vec{u}_i \in \mathbb{R}^{m+1}$, where $u_{ij}$ is the student’s utility for option $j \in \Omega$. We assume that $\vec{u}_i, i \in I_t$, are independent and identically distributed according to a known utility distribution $F_t$. Moreover, we assume that $F_t$ is continuous, which implies that the probability that a student is completely indifferent between two schools is zero.

As an example of utility distributions, consider the case where

$$u_{ij} = \bar{u}_{tj} + \epsilon_{ij} \quad \forall i \in I_t,$$

where $\bar{u}_{tj}$ is a constant representing the average utility of type $t$ students for option $j$, and $\epsilon_{ij}$ is a random perturbation drawn i.i.d. from a Gumbel distribution with location parameter 0 and scale parameter 1. This utility distribution is called multinomial logit (MNL), and it is one of the most common ways to model discrete choice because the parameters $\bar{u}_{tj}$ are easy to estimate from data. This is the form of the utility distribution we use in our empirical application in Section 4.

4.2.1 The Deferred Acceptance (DA) Framework

Students are assigned to schools according to the student-proposing deferred acceptance algorithm, which we call the DA algorithm. The algorithm has three sets of policy levers controlled by the social planner, which we call menus, priority distributions, and quotas. We define the deferred acceptance framework as the set of assignment systems that use the DA algorithm with certain settings of policy levers. These policy levers are defined as follows.

- The menu $M_t \subset [m]$ for students of type $t$ is a subset of schools offered to every student of this type. The menu $M_t$ and the utility vector $\vec{u}_i$ of a student $i \in I_t$ uniquely determine his/her preference ranking of schools, which is a permutation of options in $M_t \cup \{0\}$ sorted according to decreasing utilities. We denote this preference ranking of student $i$ by $R_i$. 

119
The priority distribution $\Pi_t$ of type $t$ is a continuous measure on $[0, 1)^{m+1}$. Each student $i \in I_t$ is given a priority vector $\vec{\pi}_i \sim \Pi_t$, where $\pi_{ij}$ is a number, which we call the priority of student $i$ at school $j$. A higher number $\pi_{ij}$ corresponds to a higher priority.

The quota $q_j$ of school $j \in [m]$ is the maximum number of students allowed to be assigned to that school.

As described above, the menus and the priority distributions induce for each student $i$ a preference ranking $R_i$ and a priority vector $\vec{\pi}_i$. Moreover, there is a quota for every school. The DA algorithm is defined in terms of these preference rankings, priority vectors, and quotas:

1. Find a student $i$ who is not yet assigned anywhere, and have him/her apply to his/her first ranked option in $R_i$.

2. If this choice is the outside option, assign the student there. Otherwise, if this choice is a school, say school $j$, then tentatively accept the student at this school.

3. If the quota of school $j$ is not exceeded as a result of this new acceptance, go back to Step 1. Otherwise, find the tentatively accepted student $i'$ at school $j$ with the lowest priority $\pi_{i'j}$, reject the student, and go to Step 2 with student $i'$ applying to his/her next choice in $R_{i'}$.

The algorithm terminates when every student is assigned to an option, which happens eventually as the outside option never rejects any student. The output of the algorithm is an assignment of every student to a school or to the outside option.

The algorithm was first proposed by [Gale and Shapley (1962)] and was first adapted for school choice by [Abdulkadiroğlu and Sönmez (2003b)]. One can show that the output does not depend on the order that students are chosen in Step 1, and that it is a dominant strategy for every student to report his/her true preference ranking (See [Roth (1982)] and [Dubins and Freedman (1981)]). This means that for each student $i$, regardless of the preference rankings of other students $i' \neq i$ and regardless
of the priority vectors, the student’s utility is maximized when he/she reports his/her truthful ranking $R_i$.

### 4.2.2 The Simulation Framework

The assignment outcome depends on all the parameters described so far according to the following process, which we call the simulation framework.

1. Draw $\vec{n} \sim H$.

2. For each student $i \in I_t$, draw the student’s utility vector $\vec{u}_i \sim F_t$. Use this utility vector and the menu $M_t$ to define the student’s preference ranking $R_i$.

3. For each student $i$ of type $t$, draw the student’s priority vector $\vec{\pi}_i \sim \Pi_t$.

4. Compute the assignment using the DA algorithm, using the students’ preference rankings $\{R_i\}$, the priorities $\{\pi_{ij}\}$, and the schools’ quotas $\{q_j\}$.

We call this the simulation framework because by repeatedly simulating the above process and averaging, one can estimate to any degree of accuracy various outcome measures. The two measures we focus on are as follows: the average expected utility of students from type $t$, which we denote by $\bar{v}_t$, and the probability that a student of type $t$ is assigned to option $j \in \Omega$, which we denote by $p_{tj}$. We denote the vector of all expected utilities by $\vec{v}$ and the matrix of all assignment probabilities by $\mathbf{p}$.

### 4.2.3 The Optimization Problem

Given the population distribution $H$ and the utility distributions $\{F_t\}$, the optimization problem of the school board is to set menus $\{M_t\}$, priority distributions $\{\Pi_t\}$ and quotas $\{q_j\}$ in a way to optimize a given objective function $W$, which we assume to be a function of the expected utilities $\vec{v}$ and assignment probabilities $\mathbf{p}$. The expected utilities and assignment probabilities are related to the decision variables, $\{M_t\}$, $\{\Pi_t\}$ and $\{q_j\}$ according to the following mapping:

$$(\vec{v}, \mathbf{p}) = DA(\{M_t\}, \{\Pi_t\}, \{q_j\}),$$  \hspace{1cm} (4.1)
which is well defined under the simulation framework for given population distribution
and utility distributions.

Under this notation, the optimization problem, which we call (OptDA), can be
expressed as follows.

$$(\text{OptDA}) \quad \text{Maximize: } W(\vec{v}, p)$$

subject to: 
$$(\vec{v}, p) = DA(\{M_t\}, \{\Pi_t\}, \{q_j\})$$

For the objective function $W$, we assume that it is non-decreasing in each com-
ponent of the expected utility vector $\vec{v}$ and that it is globally concave. This level of
generality allows it to incorporate a combination of diverse considerations, such as
average utility of students, minimum utility of any type, capacity constraints, and
convex penalty function for the average distance students travel to school (as a proxy
for school busing costs). At the same time, the objective is only dependent on the
expectations of utilities and assignments, which makes it tractable (as will be seen in
Section 4.3.2). We consider a specific form of the objective function in the empirical
exercise in Section 4.4.

4.2.4 Sufficiency of the Deferred Acceptance Framework

Before proceeding to the analysis, we first comment on why we confine the optimiza-
tion to be within the deferred acceptance framework, instead of a more general class
of assignment systems.

For example, one can imagine an alternative formulation in which the decision
variables are directly the assignment outcome, which can be represented by binary
variables $x_{ij}$ that indicate whether student $i$ is assigned to option $j$. The inputs
to this optimization would be the students’ elicited preference over schools, and the
school board’s objectives.
The issue with such an approach is that students might not have the incentive to report their true preferences, so the optimization might be based on erroneous inputs. This was a real issue in Boston before 2005, during which Boston used a non-incentive compatible assignment algorithm instead of DA. This previous algorithm, called the Boston Mechanism in the literature, prioritized satisfying students’ first choices, then only considered second choices after all first choices were filled, and third choices after second choices, and so on. This made one’s first choice very important, and parents strategized in not wasting their first choice on very popular schools, but using it on a school that they think they have a fair chance at. This resulted in prevalent gaming of preferences and a large proportion of students simply ended up unassigned. (See Abdulkadiroğlu et al. (2005b).) Because of these incentive issues, the Boston school committee adopted the DA algorithm in 2005.

To resolve the incentive issue, one would need to add incentive-compatibility constraints. Assuming that the total number of students is fixed and is equal to \( n \), this requires reformulating the optimization as maximizing over allocation functions \( x : \mathbb{R}^{n(m+1)} \rightarrow \mathbb{R}^{n(m+1)} \), in which the input is everyone’s utility reports and the output is each student \( i \)'s assignment probabilities to schools. Assuming common priors \( F_t \) on the utility distribution of students of type \( t \), the incentive-compatibility constraints are

\[
E[\vec{u}_i \cdot \vec{x}_i(\cdots, \vec{u}_i, \cdots)] \geq E[\vec{u}_i' \cdot \vec{x}_i(\cdots, \vec{u}_i', \cdots)] \quad \forall \vec{u}_i
\]

In words, this says that the expected utility for student \( i \) from reporting his true utility vector \( \vec{u}_i \), whatever that may be, must be greater than his/her expected utility from reporting any other vector \( \vec{u}_i' \). This optimization problem, being over the space of functions and having uncountably many constraints, is very complex, and the state-of-art techniques can only tractably solve this for 2 schools. (See Miralles (2012) for a paper studying a simplified version of the 2-schools case.)

For tractability, one has to assume more structure, and a natural structure is the framework already implemented in many US cities, which is choosing menus, priority distributions and quotas within the DA algorithm. This is the class of policies
considered by the city committee in the 2012-2013 Boston school assignment reform (see Shi (2015) and Pathak and Shi (2013)). An advantage of using DA is that it naturally solves the incentive problem. Furthermore, Ashlagi and Shi (2015) show that under the large-market limit in which each student becomes infinitesimal, then any method of assigning students that satisfy certain natural assumptions can be represented as DA with certain menus and priority distributions. The assumptions are 1) only using preference rankings instead of preference intensities; 2) treating students symmetrically within each type; 3) incentive compatibility; and 4) no subset of students can trade assignments within their type and all improve. This shows that under certain assumptions, restricting within the deferred acceptance framework is without loss of generality.

4.3 Analysis

While one can evaluate by simulation the school board’s objective given any combination of menus, priority distributions and quotas, it is unclear how to optimize over these policy levers. Without simulating the entire deferred acceptance (DA) algorithm, it is difficult to express students’ utilities and assignment probabilities in terms of the policy levers, as whether a student can get into a school depends on how many others with higher priority rank it highly, which in turn depends on whether they can get into their more preferred choices.

Our approach is to approximate the original problem, (OptDA), with a simpler problem, which we call the optimal random assortment problem. In this problem, the assignment is no longer by the DA algorithm, but by a random assortment mechanism, which is to offer each type of students a randomized set of schools, and assigning each student to his/her favorite school among the offered set. Such a random assortment mechanism is characterized by the probability of offering each set of schools to each type of students, and we call these the assortment probabilities. These assortment probabilities induce a realization of the the expected utilities of each type and the assignment probabilities, which in turn induce a realization of the school board’s ob-
jective function. We call the problem of finding assortment probabilities to maximize the school board’s objective the random assortment approximation to (OptDA).

Solving the random assortment approximation is useful for two reasons. Firstly, we show that the random assortment approximation is a relaxation to (OptDA), so its optimal objective upper-bounds the optimal objective in (OptDA). Secondly, we show that random assortment mechanisms are equivalent to the large-market approximation of stable matching due to Azevedo and Leshno (2015) and Abdulkadiroğlu et al. (2015). This implies that there is a mapping from assortment probabilities to menus, priority distributions, and quotas. Under this mapping, the corresponding random assortment mechanism can be interpreted as the asymptotic behavior of the corresponding DA algorithm, in the limit in which the number of student types and the number of schools are fixed, but the number of students of each type and the quotas of the schools are scaled up to infinity. This mapping allows us to use the optimal assortment probabilities to construct optimized menus, priority distributions, and quotas, that are approximately optimal solutions to (OptDA) when the number of students is large.

To solve the random assortment approximation, we formulate the optimization as a convex program, which can be solved by repeatedly solving a sub-problem of finding a deterministic assortment of schools for a single student type, such that if each student picks his/her favorite school among the assortment, and if assigning each seat of a school incurs a certain externalities, then the weighted sum of the expected utility of students of this type and the expected externalities is maximized. We call this the socially-optimal assortment planning problem, and show that it generalizes the revenue-maximizing assortment planning problem studied in the revenue management literature. We show that many algorithms for the revenue-maximizing case can be extended to the socially optimal case and use this to derive efficient algorithms under MNL utility distributions.
4.3.1 The Random Assortment Approximation

Suppose that instead of running the DA algorithm with certain menus, priority distributions and quotas, we simply offered each student of type $t$ an assortment $S \subseteq [m]$ of schools with probability $x_{tS}$, and assign each student to his/her favorite option in $S \cup \{0\}$. (Recall that 0 represents the outside option.) We call this a random assortment mechanism. Since the assortment probabilities $x_{tS}$ are exogenously given, any random assortment mechanism is strategyproof.

Define

$$V_t(S) = E[\max_{j \in S \cup \{0\}} u_{ij}],$$

(4.2)

where $\bar{u}_i \sim F_t$, and

$$P_t(j, S) = \begin{cases} F_t(\{u_{ij} \geq u_{i{j'}} \text{ for every } j' \in S \cup \{0\}\}) & \text{if } j \in S \cup \{0\}, \\ 0 & \text{otherwise.} \end{cases}$$

(4.3)

$V_t(S)$ is the expected utility of students of type $t$ when offered assortment $S$ and $P_t(j, S)$ is the probability that a type $t$ student chooses school $j$ when offered assortment $S$. Note that it is possible that the student chooses the outside option, so $j$ can be 0.

The benefit of considering random assortment mechanisms is that optimizing over them can be done by solving the following convex optimization problem, which we call the random assortment convex program. (The decision variables are $\{x_{tS}\}$. We assume also that the assortment $S$ is restricted to be within $\Psi_t$, which for now we define as $\Psi_t = [m]$. Having the more abstract form of $\Psi_t$ simplifies the exposition in Section 4.3.2 and is used to approximate a non-convex objective function in our empirical application in Section 4.4.)
(OptAssortments) Maximize: \( W(\vec{v}, \vec{p}) \)

subject to:

\[
v_t = \sum_{S \in \Psi_t} V_t(S) x_{tS}, \quad t \in [l] \quad (4.4)
\]

\[
p_{tj} = \sum_{S \in \Psi_t} P_t(j, S) x_{tS}, \quad t \in [l], j \in \Omega \quad (4.5)
\]

\[
\sum_{S \in \Psi_t} x_{tS} = 1, \quad t \in [l] \quad (4.6)
\]

\[
x_{tS} \geq 0, \quad t \in [l], S \in \Psi_t \quad (4.7)
\]

While there are exponentially many decision variables, we can use standard techniques in convex optimization to reduce this to repeatedly solving simpler, tractable problems, which we do in Section 4.3.2. Before solving this optimization, we will elucidate the connection between this problem, (OptAssortments), with the original optimization, (OptDA). Section 4.3.1 shows that (OptAssortments) is a relaxation of (OptDA). Section 4.3.1 interprets (OptAssortments) as a large-market approximation of (OptDA), and Section 4.3.1 use this relationship to map the solution of (OptAssortments) back into a feasible solution of (OptDA), which is the final output of our optimization.

Random Assortment as Relaxation

The first interpretation of (OptAssortments) is that it is a relaxation of (OptDA). This comes from the following theorem, which says that in the DA algorithm, each student is being assigned his/her favorite school from a set which we call his/her accessible assortment. The accessible assortment is independent of the student’s own preferences, but possibly dependent on the number of students of each type, the preferences of all other students and the priorities of every student.

**Definition 1.** For a student \( i \), define his/her accessible assortment as the set of options he/she can be assigned to in the student-proposing deferred acceptance (DA) algorithm if he/she chooses that option first, and ranks the outside option as more
preferred than any other option.

**Theorem 1.** In the student-proposing deferred acceptance algorithm, every student is assigned to his/her favorite school among his/her accessible assortment.

*Proof of Theorem 1.* The desired result is implied by the strategyproofness of student-proposing DA for the students (Roth (1982) and Dubins and Freedman (1981)). To see this, suppose that in the student-proposing DA, student \( i \) is rejected by schools \( j_1, \cdots, j_{k-1} \) and finally assigned to option \( j_k \).

By the strategyproofness of student-proposing DA, if \( i \) had ranked \( j_k \) first, \( i \) would still be matched to \( j_k \). This is because if \( i \) cannot get \( j_k \) by ranking it first, but can get it by ranking other schools first, then the student has incentives to misreport preferences if \( j_k \) happened to be his/her true first choice.

It suffices to show that none of the previous options, \( j_1, \cdots, j_{k-1} \) are in the accessible assortment. This again follows from strategyproofness, because if \( i \) can get any of these schools by ranking it first, then \( i \) would have an incentive to deviate because that improves upon \( i \)'s current assignment of \( j_k \). \( \square \)

In particular, the independence between a student’s accessible assortment and own preferences imply that if \( x_{tS} \) is the probability that the accessible assortment of students of type \( t \) is \( S \), then the expected utility of students of type \( t \) in the DA algorithm is exactly as in Equations 4.4 of (OptAssortments), and the assignment probabilities are as in Equation 4.5. Furthermore, the probabilities \( x_{tS} \), must sum to one for each type, so satisfy Equation 4.6. This implies that (OptAssortments) with \( \Psi_t = [m] \) is a relaxation of (OptDA), so its optimal objective must upper-bound the optimal objective of (OptDA).

**Random Assortment as Large-Market Approximation**

An alternative interpretation of (OptAssortments) is that it is solving for the optimal menus, priority distributions and quotas in the large-market approximation of DA from Azevedo and Leshno (2015). This relationship allows us to map the optimal assortment probabilities back into menus, priority distributions and quotas, which can
be interpreted as approximately optimal solutions to the original problem, (OptDA).

In the following definition, we adapt the machinery developed by Azevedo and Leshno (2015) into our setting.

**Definition 2.** Given the number of students of each type \( \{n_t\} \) and the utility distributions \( \{F_t\} \), a large-market stable matching is a tuple \((\{M_t\}, \{\Pi_t\}, \{q_j\}, \{h_j\})\), such that \( M_t \) and \( \Pi_t \) are the menu and priority distribution of type \( t \), and \( q_j \) is the quota of school \( j \in [m] \). The additional parameters are priority cutoffs \( h_j \geq 0 \) for every option \( j \in \Omega \), which is defined to be zero for the outside option \( j = 0 \). This stable matching represents a scenario in which every student \( i \) of type \( t \) receives a priority vector \( \bar{\pi}_i \sim \Pi_t \), and is assigned to his favorite option \( j \) within \( M_t \cup \{0\} \) for which his priority meets the cutoff: \( \pi_{ij} \geq h_j \). The priority cutoffs are related to the quotas in that for schools whose quota is not filled, the priority cutoff is restricted to be zero.

The intuition behind this large-market approximation is as follows: consider multiplicatively scaling up both the number of students of each type \( \{n_t\} \) and the quotas \( \{q_j\} \) by the same multiplicative parameter \( \gamma \), then the stochasticity in the preference and priority draws is averaged away, and the priority needed to get into each school \( j \) converges to a constant, represented by the cutoff \( h_j \). This is zero for schools with unfilled quota, as the DA algorithm would accept anyone who applies to such schools. For schools in which the quota is filled, this is the lowest priority among the accepted students, which is the priority one needs to beat to be assigned in the DA algorithm. A full derivation of this large market limit and its properties can be found in Azevedo and Leshno (2015).

Under this large-market stable matching model, one can analogously define the accessible assortment as the set of options that are in menu and for which one’s priority meets the cutoff. For a student of type \( t \), the probability that the accessible assortment is \( S \) is exactly

\[
x_{ts} = \begin{cases} 
\Pi_t(\{\pi_{ij} \geq h_j \forall j \in S \text{ and } \pi_{ij} < h_j \forall j \in M_t \setminus S\}) & \text{if } S \subseteq M_t, \\
0 & \text{otherwise.}
\end{cases}
\] (4.8)
Using the proof of part (1) of Theorem 1 of Azevedo and Leshno (2015) (page 5 in their appendix) and the proof of part (1) of their proposition 3 (page 16 in their appendix), one can show the following.

**Proposition 1.** Given the number of students of each type \(\{n_t\}\) and utility distributions \(\{F_t\}\). Assume that each \(F_t\) satisfies full-support, which means that for every permutation of the options \(\Omega = [m] \cup \{0\}\), utilities drawn according to \(F_t\) would induce this preference ranking with positive probability. Then for every combination of menus \(\{M_t\}\), priority distributions \(\{\Pi_t\}\), quotas \(q_j\), there exists a unique set of cutoffs \(\{h_j\}\) for which \((\{M_t\}, \{\Pi_t\}, \{q_j\}, \{h_j\})\) is a large-market stable matching.

Furthermore, consider any sequence of finite markets indexed by \(k\), with \(n_t^k\) students of type \(t\), and a quota of \(q_j^k\). Then as \(k \to \infty\), if \(n_t^k \to n_t\) and \(q_j^k \to q_j\), then the distribution of the accessible assortment under the DA algorithm (see Definition 4) converges to the distribution of accessible assortment under the above large-market model, with assortment probabilities given by Equation 4.8.

This formally establishes that the large-market approximation is well defined and approximates the outcome of the DA algorithm when the number of students of each type and the quota of each school is scaled up to infinity. Having connected the original DA framework to large-market stable matchings, the following theorem connects large-market stable matchings with random assortment mechanisms.

**Theorem 2.** Fix the number of students of each type \(\{n_t\}\) and utility distributions \(\{F_t\}\). Any large-market stable matching \((\{M_t\}, \{\Pi_t\}, \{q_j\}, \{h_j\})\) can be equivalently represented as a random assortment mechanism with certain assortment probabilities \(\{x_{tS}\}\). Conversely, any random assortment mechanism with assortment probabilities \(\{x_{tS}\}\) can be represented as a large-market stable matching with certain parameters \((\{M_t\}, \{\Pi_t\}, \{q_j\}, \{h_j\})\).

**Proof of Theorem 2.** The first direction is immediate. Given \((\{M_t\}, \{\Pi_t\}, \{q_j\}, \{h_j\})\), define \(x_{tS}\) as in equation 4.8. The random assortment mechanism with these assortment probabilities assigns every student to their favorite school in menu for which their priority meets the cutoff.
The second direction takes a little more work. Given the assortment probabilities $x_{ts}$, let

$$M_t = \bigcup_{S: x_{ts} > 0} S \quad \text{for all } t \in [l], \quad (4.9)$$

$$h_j = \frac{1}{2} \quad \text{for every school } j \in [m], \quad (4.10)$$

$$q_j = \sum_{t=1}^{l} n_t \sum_{S \ni j} P_t(j, S) x_{ts} \quad \text{for every school } j \in [m], \quad (4.11)$$

$$\quad (4.12)$$

For the priority distributions $\{\Pi_t\}$, define for each student two random variables, $S_i$ and $\delta_i$, where $S_i \subseteq [m]$ is distributed according to $x_{ts}$ and $\epsilon_i \sim \text{Uniform}[0, 1]$. Let each student’s priority vector $\pi_i$ be such that

$$\pi_{ij} = \frac{1}{2}(1(j \in S_i) + \delta_i). \quad (4.13)$$

This priority distribution is constructed so that for each student $i$ of type $t$, the set of schools for which the student’s cutoff is above the the cutoff of $h_j \frac{1}{2}$ is exactly $S_i$, which is distributed according to the assortment probability $x_{ts}$. The definition of quotas make it so that every school is exactly filled, which removes any restrictions on the cutoffs.

The Optimized Menus, Priority Distributions, and Quotas

The proof of Theorem 2 yields a possible mapping of assortment probabilities $\{x_{ts}\}$ into menus, priority distributions, and quotas. While this mapping can be applied to general assortment probabilities, the following simpler mapping works in the special case when the assortments for each type $C_t = \{ S : x_{ts} > 0 \}$ are nested. The advantage of the following mapping is that it is simpler to implement, being the sum of a constant term $h_{ij}$ and an random number $\epsilon_i$. Moreover, the final assignment is also guaranteed to be Pareto optimal within each type, meaning that students cannot trade assignments within their type and all improve. This form of priorities is used
also in Ashlagi and Shi (2015). The mapping has the same menus, cutoffs, and quotas as in Equations 4.9, 4.10, and 4.11 but the priorities become.

\[ \pi_{ij} = \frac{1}{2} (h_{tj} + \epsilon_i) \]  

(4.14)

where \( \epsilon_i \sim \text{Uniform}[0,1] \) as before and

\[ h_{tj} = \sum_{S \ni j} x_{tS}. \]

It is this mapping that we will apply in Section 4.4.

### 4.3.2 Solution Method to the Random Assortment Convex Program

The random assortment convex program as formulated in Section 4.3.1 (OptAssortments), has exponentially many decision variables \( x_{tS} \). A standard approach for such problems is simplicial decomposition, which is a generalization of the column generation technique from linear programming. This is an iterative algorithm that maintains for each type a subset \( C_t \subseteq \Psi_t \). From any feasible solution of (OptAssortments), we initialize each \( C_t \) to be the set of assortments \( S \) for which \( x_{tS} > 0 \) in the feasible solution.

The algorithm is iterative, and the sets \( C_t \) expand after each iteration. In each iteration, we first solve a master problem, which is the same as the original formulation of (OptAssortments) except that we replaces \( \Psi_t \) by the smaller set \( C_t \), which greatly simplifies the optimization if the cardinalities of the \( C_t \)'s are small.

Given the optimal solution of the master problem, we compute a super-gradient \((\vec{\alpha}, \vec{r})\) of the concave objective \( W(\vec{v}, \vec{p}) \), such that \( \alpha_t \) is the component of the super-gradient for \( v_t \) and \( r_{tj} \) is the component for \( p_{tj} \). Note that \( \alpha_t \geq 0 \) by our assumption that the objective is non-decreasing in every component of \( \vec{v} \).

The sub-problem for each type \( t \) is to find an assortment \( S \in \Psi_t \) that optimizes
the following linear objective:

$$\max_{S \in \Psi_t} \alpha_t V_t(S) + \sum_{j \in \Omega} r_{tj} P_t(j, S), \quad (4.15)$$

where $V_t(S)$ is the value of assortment $S$ for type $t$ students as in Equation 4.2 and $P_t(j, S)$ is the probability of preferring option $j$ the most among $S \cup \{0\}$ as in Equation 4.3.

For each type $t$, let the optimal solution to the above sub-problem be $S_t^*$. We append $S_t^*$ to $C_t$ for every type $t$, and iterate again to resolve the master problem. The algorithm terminates if the optimal objective of the master problem does not improve between two successive iterations. See [Von Hohenbalken 1977] for the development of the theory of simplicial decomposition and proof of correctness. For a more recent exposition, see Chapter 4 of [Bertsekas 2015].

The Key Sub-Problem: Socially-Optimal Assortment Planning

The sub-problem (Equation 4.15) has the following interpretation: find for each type $t$ a deterministic assortment $S$ within a constraint set $\Psi_t$ to maximize the weighted sum of two terms: the expected utility of type $t$ and the externalities on others. The expected utility is $V_t(S)$ and is multiplied by parameter $\alpha_t \geq 0$, which represents how much to weigh the utility of this type. There is an externality of $r_{tj}$ of assigning each student of type $t$ to each option $j \in \Omega$, so the average externalities from assortment $S$ is $\sum_{j \in \Omega} r_{tj} P_t(j, S)$. For the MNL model,

$$V_t(S) = \log\left(\sum_{j \in S \cup \{0\}} \exp(\bar{u}_{tj})\right) + \gamma_{\text{Euler}},$$

where $\gamma_{\text{Euler}} = .577\ldots$ is the Euler constant, and

$$P_t(j, S) = \begin{cases} \frac{\exp(\bar{u}_{tj})}{\sum_{j' \in S \cup \{0\}} \exp(\bar{u}_{tj'})} & \text{if } j \in S \cup \{0\}, \\ 0 & \text{otherwise.} \end{cases}$$
If $\alpha_t = 0$, then this is exactly the revenue-maximizing assortment planning problem studied in the revenue management literature. In that problem, $r_{tj}$ is interpreted as the unit revenue of assigning each person of type $t$ to option $j$. With positive $\alpha_t$, we call this the socially-optimal assortment planning problem, since the objective can be interpreted as the weighted sum of consumer welfare and revenue.

In the following, we extend classical results for the MNL model (defined in Section 4.2) from the revenue-maximizing case to the socially-optimal case. This gives us efficient algorithms to solve the sub-problem (Equation 4.15) under MNL utilities, which allows us to efficiently solve the assortment planning convex program, (OptAssortments). For ease of exposition, for the remainder of this section, we drop the suffix $t$ in $\alpha_t$, $\Psi_t$, $V_t$, $P_t$, $r_{tj}$, and $\bar{u}_{tj}$, since everything in this section pertains to a fixed type $t$. For each option $j \in \Omega$, define $w_j = \exp(\bar{u}_j)$. Plugging in the formula for $V(S)$ and $P(j, S)$ into Equation 4.15, the problem we need to solve is

$$\max_{S \in \Psi} \alpha \log(\sum_{j \in S \cup \{0\}} w_j) + \frac{\sum_{j \in S \cup \{0\}} r_j w_j}{\sum_{j \in S \cup \{0\}} w_j}. \quad (4.16)$$

In the revenue management literature in which $\alpha = 0$, a classical result due to van Ryzin and Mahajan (1999) is that when $\Psi = [m]$, the optimal assortment is revenue-ordered, which means in our setting that whenever the assortment contains any option $j$, it must contain any option $j'$ with better unit externalities, $r_{j'} > r_j$. Rusmevichientong et al. (2010) show that with cardinality constraints, in which $\Psi = \{S \subseteq [m] : |S| \leq k\}$ for some parameter $k > 0$, the revenue-ordered property no longer holds, but one can still compute the optimal assortment efficiently as follows. Define function

$$f(\lambda) = \max_{S \in \Psi} \{ \sum_{j \in S \cup \{0\}} w_j (r_j - \lambda) \}. \quad (4.17)$$

This is the maximum of finitely many decreasing linear functions, so is convex and decreasing. There exists a unique $\lambda^*$ such that $f(\lambda^*) = 0$. Rusmevichientong et al. (2010) show that the optimal objective to the assortment planning problem (Equa-
tion 4.16 with $\alpha = 0$) is exactly $\lambda^*$ and an assortment $S^* \subseteq \Psi$ is an optimal solution if and only if

$$S^* \in \arg \max_{S \in \Psi} \sum_{j \in S \cup \{0\}} w_j(r_j - \lambda^*).$$

This generalizes the first result because with $k = \infty$, $\Psi = [m]$, the above implies that an assortment is revenue-optimal if and only if it contains every option with $r_j > \lambda^*$ and no option with $r_j < \lambda^*$.

This result yields a polynomial-time algorithm to compute the optimal assortment because $f(\lambda)$ can be computed efficiently under cardinality constraints. This is because at any $\lambda$, $f$ is simply the sum of the largest $k$ terms $z(\lambda) = w_j(r_j - \lambda)$, as long as all of them are positive. (If some of the largest $k$ terms are negative at $\lambda$, then we do not include them in the sum.) Treating $z(\lambda)$ as a linear function with $\lambda$ as the x-axis, we see that the identity of the included terms is fixed within any region in which the lines do not cross one another or cross the x-axis. Since there are at most $O(m^2)$ such crossing points, one can compute a piece-wise linear representation of $f$ by first sorting the intersection points, and computing the linear representation within each region between successive intersection points. This piece-wise linear representation has at most $O(m^2)$ pieces and can be computed in $O(m^2 \log(m))$ time. (See Rusmevichientong et al. (2010).)

We generalize these results to the case with positive $\alpha$ using the following theorem. The intuition is to use quasi-convexity and convex duality to reformulate the original optimization as a simpler, one-dimensional optimization.

**Theorem 3.** With $\alpha > 0$, define

$$\Lambda^* = \arg \max_{\lambda \in \mathbb{R}} \{ f(\lambda) \exp\left(\frac{\lambda}{\alpha}\right) \},$$

then an assortment $S^*$ is an optimal solution to the socially-optimal assortment plan-
ning problem with MNL utilities (Equation 4.16) if and only if

\[ S^* \in \bigcup_{\lambda^* \in \Lambda^*} \arg \max_{S \in \Psi} \sum_{j \in S \cup \{0\}} w_j (r_j - \lambda^*). \]  

(4.19)

Proof of Theorem 3. Define \( x(S) = \sum_{j \in S \cup \{0\}} w_j \), \( y(S) = \sum_{j \in S \cup \{0\}} r_j w_j \), \( D = \{(x(S), y(S)) : S \in \Psi\} \), and \( g(x, y) = \alpha \log(x) + \frac{y}{2} \). Since there is a correspondence between points in \( D \) and assortments, the original problem (Equation 4.16) can be formulated as

\[ \max_{(x, y) \in D} g(x, y) \]  

(4.20)

Note that in the domain \( R = (0, \infty) \times (-\infty, \infty) \), the function \( g(x, y) \) is quasi-convex, continuous, and strictly increasing in \( y \). This allows us to use the following lemma, which exploits these properties of \( g \) to reformulate the problem as a simpler, one-dimensional optimization.

**Lemma 1.** Let \( R \) be an open convex subset of \( \mathbb{R}^2 \) and let \( g(x, y) : R \to \mathbb{R} \) be a quasi-convex, continuous function that is strictly increasing in \( y \). Let \( D \) be a finite set of points from \( R \). Define

\[ f(\lambda) = \max_{(x, y) \in D} \{y - \lambda x\}, \]  

(4.21)

\[ A(\lambda) = \arg \max_{(x, y) \in D} \{y - \lambda x\}, \]  

(4.22)

\[ h(\lambda, f) = \inf_{(x, y) \in R} \{g(x, y) : y = f + \lambda x\}. \]  

(4.23)

Then

\[ \max_{(x, y) \in D} g(x, y) = \sup_{\lambda \in \mathbb{R}} h(\lambda, f(\lambda)). \]  

(4.24)

Moreover, the supremum on the right is attainable, and \( (x^*, y^*) \) is an optimal solution of the left hand side if and only if \( (x^*, y^*) \in A(\lambda^*) \) for some optimal solution \( \lambda^* \) of
the right hand side.

The proof of this Lemma is in Appendix 4.E.

Plugging in the specific form of \( g \), we find that the \( h \) function in our case is
\[
h(\lambda, f) = \alpha \log(\frac{f}{\alpha}) + \lambda + \alpha.
\]
Optimizing this with respect to \( \lambda \) is equivalent to optimizing \( \exp(\frac{\lambda}{\alpha}) \). This implies the desired result.

Theorem 3 is generalized in Appendix 4.F for any Generalized Extreme Value (GEV) utility distribution, which includes nested logit, and multi-level nested logit utility distributions.

This theorem gives us a polynomial-time algorithm to solve problem 4.16 if we can find in polynomial-time a piece-wise linear representation of \( f(\lambda) \) with polynomially-many pieces. This is because for each linear segment, \( a - b\lambda \), the optimal solution to Equation 4.18 can be found analytically, by checking the end points of the line segment and by checking \( \lambda = \frac{a}{b(1+\alpha)} \).

This yields a polynomial-time algorithm for problem 4.16 if \( \Psi \) is a matroid, which is defined as follows and generalizes cardinality constraints.

**Definition 3.** A matroid over ground set \([m]\) is a set \( \Psi \) of subsets of \([m]\) such that
1) \( \emptyset \in \Psi \); 2) if \( S' \subset S \subseteq [m] \) and \( S \in \Psi \), then \( S' \in \Psi \); and 3) if \( S, S' \in \Psi \) and \( |S'| > |S| \), then there exists \( j \in S'\backslash S \) such that \( S \cup \{j\} \in \Psi \).

A well-known result in combinatorial optimization is that when optimizing a sum over a matroid as in Equation 4.17, the greedy algorithm is optimal. This implies that in computing \( f(\lambda) \) of Equation 4.17 for each \( \lambda \), we can go through options \( j \) in decreasing order of \( w_j(r_j - \lambda) \) and greedily include \( j \) as long as this does not violate the matroid constraint. (Precisely speaking, we compute a set \( S \) by initializing it as the empty set, and iteratively going through options \( j \) sorted in this way, and adding elements to \( S \) as long as this does not violate \( S \in \Psi \). After this terminates, we define \( f(\lambda) = \sum_{j \in S \cup \{0\}} w_j(r_j - \lambda). \) Since the construction of \( S \) only depends on the order of the lines \( z(\lambda) = w_j(r_j - \lambda) \), there are at most \( O(m^2) \) segments and \( f(\lambda) \) can be calculated in \( O(m^2 \log(m)) \) time (assuming that checking whether a set \( S \) satisfies \( S \in \Psi \) takes constant time).
4.4 Empirical Application

We apply the optimization framework on data from Boston Public Schools (BPS) to yield improved menus and priority distributions for elementary school assignment. The motivation is that Boston launched a reform to alter their menus and priority distributions in 2012-2013. (The quotas were based on capacity limits and were considered fixed and not a part of the reform.) The menus and priority distributions before the reform constituted what is called the 3-Zone plan, which divided the city into three geographic zones. The menu of a student was the union of any school within the zone containing the student’s home, and any school within a one-mile radius of the student’s home. This one-mile radius is called the student’s walk-zone, and if the student travels outside the walk-zone, then the school board is obligated to send a school bus to pick up the student.

Because of these large menus, each school had to pick up children from about one-third of the city, and this resulted in Boston paying over 80-million dollars a year\(^1\) in 2012, which is about 10% of its budget, on busing. (See Russell and Ebbert (2011).) The menus and priority distributions after the reform was called the Home-Based plan, which was designed to reduce the busing burden of the city by reducing the menus, while giving students the schools they want the most by constructing a menu centered around the student’s home, based on test-scores, capacities, and special programs (See Appendix 2.A for a precise definition of the Home-Based plan). The plan also compensated families living in worse neighborhoods with more choices.

The city decided to adopt the Home-Based plan after reviewing detailed simulation results. The original analysis is found in Pathak and Shi (2013) and Shi (2015), and uses a similar setup as Section 4.2. In this section, we use our optimization framework to evaluate how much room for improvement there is for the Home-Based plan. Specifically, we use the same busing allowance as in Home-Based, but optimize the sum of the average utilities of students and the minimum utility of any neighbor-

\(^1\)As a comparison, the per capita spending on busing in 2012 for Boston is over 3 times the national average and 30% higher than in New York City. (Source: US Census Public-Secondary Education Finance Data 2012.)
hood. The output of the optimization are menus and priority distributions, which can be incorporated within the DA algorithm similar to the 3-Zone and Home-Based plan. We find that substantial improvement is possible, as the optimized plan not only achieves better expected utilities, but also improves students’ chances of getting top choices in menu and improves students’ chances of going to school with their neighbors.

The distributional assumptions are in Section 4.4.1. The outcomes of interest are described in Section 4.4.2. We apply the optimization framework in Section 4.4.3 and discuss the results in Section 4.4.4.

4.4.1 Distributional Assumptions

In this section, we outline the distributional assumptions behind both our simulation results and our optimization. These assumptions are the same as in the empirical exercise in Ashlagi and Shi (2015). As in Section 4.2, we specify a partitioning of students into types, a population distribution, and a utility distribution for each type. Since the quotas was not part of the reform, we will also specify the quotas for each school. These parameters are based on BPS data from 2010-2013 for grade Kindergarten-2, which is the main entry grade to the elementary school system.

Student Types

We partition students into types based on geographic location alone. The BPS dataset partitions Boston into 868 geographic tracts, which we will call neighborhoods. We define each neighborhood to be a type.

The data also groups the 868 neighborhoods into 14 larger regions, which are based on natural divisions of the city. (For example, downtown is a region by itself.)

---

2It is also possible to consider other differences across students, such as race, older siblings, special education needs, and language learning needs. However, for clarity of analysis, we focus on the geographic aspects in this exercise.
Population Distribution

Based on projections from previous years, we model \( n_t \), the number of students from neighborhood \( t \), as follows. Define a normal random variable with mean 4294 and standard deviation 115. This represents the total number of applicants and is estimated from historic data from 2010-2013. To accommodate medium-scale regional variations, we generate an independent normal random variable for each of the 14 regions, which represents the proportion of students who come from this region. The means and standard deviations as in Table 4.3 of Appendix 4.A. The total number of students of each region is the product of the overall normal variable with the region-specific term, rounded to the nearest integer. Having computed this regional total, we sample the neighborhood \( t \) of each student based on the historic density in 2010-2013. This process induces random variables \( n_t \) for each neighborhood \( t \), which are positively correlated both across the city and within each region.

Utility Distributions

As in Ashlagi and Shi (2015), we estimate a MNL model of students’ preferences. Let \( i \) be a student of neighborhood \( t \) and let \( j \) be a school. Assume that utilities take the form

\[
  u_{ij} = \bar{u}_{tj} + \beta \epsilon_{ij},
\]

\[
  \bar{u}_{tj} = Q_j - \text{Distance}_{tj} + \gamma \cdot \text{Walk}_{tj}.
\]

The data in the above equations are Distance\(_{tj}\), and Walk\(_{tj}\), and the parameters are \( Q_j \), \( \gamma \) and \( \beta \). Distance\(_{tj}\) is the walking distance from the centroid of neighborhood \( t \) to school \( j \) according to Google Maps. Walk\(_{tj}\) is an indicator that school \( j \) is within the walk-zone of neighborhood \( t \). \( Q_j \) is a school-specific fixed effect capturing overall school popularity, and we call this the inferred quality of school \( j \). \( \beta \) is the size of the perturbation term in the utility (recall that \( \epsilon_{ij} \) is Gumbel distributed with location parameter 0 and scale parameter 1). \( \gamma \) is a coefficient for living within one-mile.

Note that the normalization in the above is different from that in Section 4.2.
Table 4.1: Parameters of the MNL model, estimated from preference data from 2013. The values can be interpreted in units of miles (how many miles a student is willing to travel for one unit of this variable).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_j$</td>
<td>0-6.29</td>
<td>Quality of schools. For a school of $\Delta Q$ additional quality, holding fixed other components, a student would be willing to travel $\Delta Q$ miles further. The value for each school is graphically displayed in Figure 4-1b of Appendix 4.A.</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.86</td>
<td>Additional utility for going to a school within the walk-zone.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.88</td>
<td>Scale parameter of the Gumbel term.</td>
</tr>
</tbody>
</table>

Instead of normalizing the scale of the Gumbel term to one, we allow a scale of $\beta$, while normalizing the distance coefficient to one. Either normalizations induce the same choice behavior, and the difference is entirely in interpretation. We estimate the parameters $Q_j$, $\gamma$ and $\beta$ from rank-order preference rankings from the past, using the maximum likelihood technique of Hausman and Ruud (1987). The estimates are shown in Table 4.1. For the inferred qualities of schools, we plot them on a map in Figure 4-1b in Appendix 4.A. We normalize the lowest $Q_j$ to zero.

Note that this model seeks to estimate from students’ preference rankings also their preference intensities. The logic behind this is as follows: assuming that the differences in how students from different neighborhoods rank schools can entirely be explained by geography, then we can infer students’ preference strengths by observing how quickly they trade these preferences for distance. For example, suppose that students generally prefer school A over school B. In neighborhoods equidistant from the two schools, then we would expect more students rank A before than B. However, as we move through neighborhoods going closer to B, we may see students preferring B more. By observing the speed at which their preference distribution change, we can have a rough estimate of the strength of their preferences.

It is possible also to add non-linear terms of distance as well as interactions between students’ race and income and the school’s demographics and test-scores, as in Shi (2015) and P. and Shi (2015). P. and Shi (2015) also compare the above model to a mixed MNL model, which allows random coefficients, and they show that the

3In the estimation, we use the normalization in which the scale of the idiosyncratic term is one. After estimating the parameters, we re-normalize to make the distance coefficient one.
models perform similarly in prediction accuracy in the Boston data.

Schools and Quotas

There are $m = 77$ schools in our dataset, each of which has a capacity constraint, which we model as a hard quota $\bar{q}_j$. In contrast to our model in Section 4.2, the quotas are given inputs rather than decision variables in the optimization. Figure 4-1a plots the school quotas and locations.

One complication with hard quotas is that the school board is obligated by law to assign every applicant, and in practice, the school board may open up new classrooms or slightly increase class sizes in order to accommodate unassigned students. In Boston, this is often done manually after the main assignment has taken place. In the Home-Based plan, there is a set of 19 schools which are called “capacity schools,” which in the original proposal were intended to be schools that BPS can expand capacities at to accommodate excess demand. So many of the unassigned students may be later assigned to one of these capacity schools.

For concreteness, we assign each neighborhood a default school, which is the closest capacity school. Moreover, we treat the default school for a type not as a regular school option, but as the outside option of that type, so assignment to one’s outside option is synonymous with assignment to one’s default school. This implies that students assigned to the default school are not counted against the quota of the default school, which guarantees that every student can at least be assigned to his/her default school. For each school $j$, define $T_j$ as the types for which the school $j$ is not the type’s default school. These are the types on which the quota $\bar{q}_j$ applies.

4.4.2 Outcomes of Interest

Given any setting of menus and priority distributions, we use the quotas from Section 4.4.1 and compute the following set of metrics. The first and second metrics measure the transportation burden for the city. The third measures efficiency and equity in terms of students’ welfare. The fourth measures predictability, and the fifth
measure community cohesion. The first three outcomes are the most important, and they will directly enter the optimization in Section 4.4.3. The last two are secondary, and do not enter the optimization.

1. **Average busing distance:** Define $b_{tj} = \text{Distance}_{tj}(1 - \text{Walk}_{tj})$. This is the distance between school $j$ and neighborhood $t$ if bus transportation is needed, and zero otherwise. The average busing distance is the average of $b_{tj}$ among all students.

2. **Average bus coverage area:** This is the area the average school has to cover to pick up children from. Let $\text{Area}_t$ be the size of neighborhood $t$, and let $\text{Walk}_t$ be the set of schools within the one-mile walk zone of neighborhood $t$. The average bus coverage area is

$$\frac{1}{m} \sum_{t \in [l]} |M_t\setminus \text{Walk}_t|\text{Area}_t. \quad (4.25)$$

3. **Expected utilities of neighborhoods:** For each neighborhood $t$, this is the expected utility of a student from neighborhood, without conditioning on the student’s own utility draw. This is exactly the $v_t$ as defined in Section 4.2.2. To aggregate the results from all 868 neighborhoods, we compute:

   a) Weighted average: $\frac{\sum_{t \in [l]} E[n_t]v_t}{\sum_{t \in [l]} E[n_t]}$.
   b) 10th percentile from the bottom.
   c) Minimum $v_t$.

4. **Probability of getting top choice in menu:** This is estimated from the empirical frequency of getting one’s top choice within menu from the simulation framework in Section 4.2.2. Similarly, we compute the chance of getting one of the top three choices.

5. **Number of neighbors co-assigned:** For each neighborhood $t$, we compute the empirical average across many simulations of the number of peers students
from this neighborhood have at their assigned school who live less than half mile away according to Google Maps walking distance. We report the median across neighborhoods. A similar metric for community cohesion is studied in Ashlagi and Shi (2014).

Note that we use distance and bus coverage as proxies for busing costs. This is because these were the metrics used to measure busing burden by the city committee during the 2012-2013 reform (See Pathak and Shi (2013)). A more direct estimate is also very difficult to compute as bus routes depend on previous year’s assignments and institutional constraints. Furthermore, the actual bus routes used by BPS is computed by an independent company using proprietary software, which we have no access to.

4.4.3 Details in Applying the Random Assortment Approximation to Boston

We formulate the optimization as finding menus and priority distributions that, when evaluated using the simulation framework of Section 4.2.2 with the quotas and distributional assumptions in Section 4.4.1 maximize the sum of the weighted average and minimum expected utilities of neighborhoods, subject to staying within the Home-Based plan in the average busing distance and bus coverage area.

One difficulty in applying the optimization framework in Section 4.3 is that the bus coverage area in Equation 4.25 cannot be directly incorporated into a concave objective $W$ in terms of utilities and assignment probabilities. As a heuristic to overcome this, we take advantage of the flexibility of having arbitrary set constraints $\Psi_t$ in the formulation of (OptAssortments) in Section 4.3.1. For each neighborhood $t$, we limit the assortment to have at most $k$ schools outside of the one-mile walk-zone, with $k$ being a parameter to be determined later. Precisely speaking, let

$$\Psi_t = \{S \subseteq [m] : |S \setminus \text{Walk}_t| \leq k + \text{Walk}_{t0} - 1\}, \quad (4.26)$$
where \( \text{Walk}_t \) is the set of schools within one-mile, and \( \text{Walk}_{t0} \) is an indicator for whether the default school of neighborhood \( t \) is within the walk-zone. (Recall from Section 4.4.1 that we model the default school as the outside option, so in order to have \( k \) be the limit including the default school, we add the term \( \text{Walk}_{t0} - 1 \).) This cardinality constraint in the assortment is intended to limit the cardinality of the menus, and so induce a low bus coverage area.

After applying this heuristic, the optimization can now be expressed in a form as in (OptAssortments). Let \( \bar{q}_j \) be the fixed quota from section 4.4.1 which only applies to neighborhoods in the subset \( T_j \). Let \( B \) be a limit on the average busing distance. For the MNL model in Section 4.4.1, the assortment valuations and assignment probabilities are

\[
V_t(S) = \beta \log \left( \sum_{j \in S \cup \{0\}} \exp(\bar{u}_{tj}/\beta) \right)
\]  
(4.27)

\[
P_t(j, S) = \begin{cases} 
\frac{\exp(\bar{u}_{tj}/\beta)}{\sum_{j' \in S \cup \{0\}} \exp(\bar{u}_{tj'}/\beta)} & \text{if } j \in S \cup \{0\}, \\
0 & \text{otherwise.} 
\end{cases}
\]  
(4.28)

Using these parameters, the random assortment convex program for Boston becomes as follows:
Maximize: $A$ Utility + MinUtility

subject to:

\[
\sum_{t \in [l]} \sum_{S \in \Psi_t} E[n_t]v_t = A \text{ Utility} \quad (4.29)
\]

\[
v_t \geq \text{MinUtility}, \quad t \in [l] \quad (4.30)
\]

\[
\sum_{t \in T_j} \sum_{S \in \Psi_t} E[n_t]p_{tj} \leq \bar{q}_j, \quad j \in [m] \quad (4.31)
\]

\[
\sum_{S \in \Psi_t} V_t(S)x_{tS} = v_t, \quad t \in [l]
\]

\[
\sum_{S \in \Psi_t} P_t(j, S)x_{tS} = p_{tj}, \quad t \in [l], j \in [m] \cup \{0\}
\]

\[
\sum_{S \in \Psi_t} x_{tS} = 1, \quad t \in [l]
\]

\[
x_{tS} \geq 0, \quad t \in [l], S \in \Psi_t
\]

This is efficiently solvable using the technique in Section 4.3.2 because $\Psi_t$ in Equation 4.26 describes a matroid (see Definition 3). The solution depends on two parameters, an allowance $B$ for average busing distance, and an allowance $k$ for the cardinality of assortments outside of one’s walk-zone, and we tweak these parameters by trial and error so the corresponding menus and priority distributions from Section 4.4.3 use less busing distance and coverage area than the Home-Based Plan. The final choice of these parameters are $B = 0.55$ and $k = 8$.

**The Optimal Random Assortment Mechanism**

From Theorem 3 the optimal assortments have the following structure: each neighborhood $t$ is given an allowance $a_t$ of points, and each school $j$ costs $c_j$ points. Given allowance $a_t$, the assortment for the neighborhood is the union of

\[\text{4Note that the default school, being modeled as the outside option, is not technically in the assortment, but these schools are always offered to students and students can always attend them if they wish.}\]
• any school $j$ within the one-mile walk-zone which costs less than the allowance, $c_j < a_t$.

• The $k_t = k + \text{Walk}_{i0} - 1$ schools outside of the walk-zone with the highest score $\sigma_{tj}$, defined as

$$\sigma_{tj} = \bar{u}_{tj} + \log(a_t - \text{Distance}_{tj} - c_j) \quad (4.32)$$

If fewer than $k_t$ schools have a positive sum within the logarithm, then include only the ones that do.

The costs $c_j$ are fixed, but the allowance $a_t$ may be random, and so the associated assortment can also be random. However, the assortment is deterministic for 63% of neighborhoods, and involves at most two values for 35% of neighborhoods.

These assortments can be interpreted intuitively as follows: The cost $c_j$ represent the opportunity cost of removing one unit of capacity for other students. This is higher for schools that are more popular but have lower capacities. The allowance $a_t$ represents how much “affirmative action” neighborhood $t$ gets in the allocation, with neighborhoods with lower expected utility given a higher value of $a_t$, since the minimum utility of neighborhoods is in the objective. Figure 4-2a in Appendix 4.A plots the distribution of expected allowance across the city and Figure 4-2b plots the school costs. For schools within the one-mile walk-zone, there is no need of limiting options for the bus coverage constraint, so such schools are included as long as they are not too highly demanded, $c_j < a_t$. For schools outside of the walk-zone, we limit to $k_t = k + \text{Walk}_{i0} - 1$ schools so that including the default school (which is modeled as the outside option as in Section 4.4.1), the total number of schools outside of one’s walk-zone is at most $k$. The formula for deciding which schools to include (Equation 4.32) favors schools that the neighborhood likes on average (high $\bar{u}_{tj}$), and penalizes schools that are far away or have high cost. This formula optimally balances the utility of neighborhood $t$ with externalities of others.
The Optimized Menus and Priority Distributions

From these optimal assortment probabilities \( \{x_{tS}\} \), we obtain corresponding menus, priority distributions, and quotas. For the menus, let the assortments with positive probability for each type be \( C_t = \{S : x_{tS} > 0\} \), we define the menu to be \( M_t = \bigcup_S C_t S \) as in Equation 4.9.

As mentioned before, because of institutional constraints from 2012-2013 reform, we use the quotas \( \bar{q}_j \) from Section 4.4.1 instead of the optimized quotas from Equation 4.11. By Equation 4.11, the optimized quota cannot be more than the capacity limit, so this departure from theory only increases quotas at schools. By well-known properties of the DA algorithm, this yields a Pareto improvement for the students in their utilities, but may increase the busing distance.

Moreover, for the priority distributions, instead of using Equation 4.13, which involves the sum of two random components, we use Equation 4.14, which involves the sum of a neighborhood-specific deterministic term and an i.i.d. random number for each student. This is because this second form of priorities is already used in the Home-Based plan, so it is easier to adopt and explain to families. However, in order to apply Equation 4.14, the assortments should be nested for every neighborhood, which does not hold in our case. As a heuristic, we modify the assortments to make them nested. Let the original assortments be:

\[
C_t = \{S_1, S_2, \ldots\}, \tag{4.33}
\]

with \(|S_1| \leq |S_2| \leq \cdots \) \hspace{1cm} (4.34)

Then define new assortments \( S'_1 = S_1, S'_2 = S'_1 \cup S_2, S'_3 = S'_2 \cup S_3 \) and so on, with assortment probabilities \( x'_{tS'} = x_{tS} \). These new assortments \( C'_t = \{S'_1, S'_2, \ldots\} \) are nested by construction, so we can apply Equation 4.14. Let \( h_{tj} = \sum_{S' \in C'_t, S' \subseteq j} x'_{tS'} \), and let \( \delta_i \sim \text{Uniform}[0,1] \), then we define the priority of student \( i \) for school \( j \) to be:

\[
\pi_{ij} = h_{tj} + \delta_i \tag{4.35}
\]

\footnote{Note that we removed the factor of \( \frac{1}{2} \), but this is without loss of generality.}
We refer to the menus and priority distributions described above as the Optimized plan.

### 4.4.4 Simulation Results

We evaluate the outcomes of interest from Section 4.4.2 for the 3-Zone plan, the Home-Based plan, and the Optimized plan in Table 4.2. All the results are averages from 10,000 independent simulations according to the process described in Section 4.2.2, which involves sampling population, utilities, priority distributions and applying the DA algorithm. In each of the plans, the corresponding menus and priority distributions are used. The capacity limit of Section 4.4.1 is used as quotas for all three plans.

Table 4.2: Simulation Comparisons of 3-Zone, Home-Based, and Optimized plans.

<table>
<thead>
<tr>
<th></th>
<th>3-Zone</th>
<th>Home-Based</th>
<th>Optimized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. # of choices</td>
<td>27.5</td>
<td>14.8</td>
<td>15.0</td>
</tr>
<tr>
<td>Av. busing distance (miles)</td>
<td>1.3</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Av. bus coverage (sq. miles)</td>
<td>20.9</td>
<td>8.5</td>
<td>8.3</td>
</tr>
<tr>
<td>Expected utilities of neighborhoods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>7.4</td>
<td>7.0</td>
<td>7.5</td>
</tr>
<tr>
<td>10th perc.</td>
<td>6.6</td>
<td>6.1</td>
<td>6.7</td>
</tr>
<tr>
<td>Lowest</td>
<td>5.1</td>
<td>4.5</td>
<td>6.5</td>
</tr>
<tr>
<td>% getting top choices in menu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1</td>
<td>63%</td>
<td>64%</td>
<td>85%</td>
</tr>
<tr>
<td>Top 3</td>
<td>84%</td>
<td>85%</td>
<td>96%</td>
</tr>
<tr>
<td>Median # of neighbors co-assigned</td>
<td>3.1</td>
<td>5.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Comparing first the 3-Zone plan and the Home-Based plan, we see that the Home-Based plan gives students a smaller sized menu on average (15 as opposed to 28), and uses much less busing. Both the average distance and the bus coverage area are less than half of that in the 3-Zone plan. However, in obtaining these savings, the Home-Based plan sacrifices the expected utilities of students, both on average and for the neighborhoods with lowest utilities, and this is driven by the large decrease in the variety of choice. However, this decrease is arguably small. As a ballpark comparison, we translate this utility decrease in terms of distance, as the distance coefficient in the
utilities is one. This welfare decrease is equivalent to having students travel 0.5 miles longer to every school. In terms of predictability, the Home-Based plan performs similarly to the 3-Zone plan, as measured by giving students high chances to the schools they rank first. In terms of community cohesion, students from the median neighborhood have 2 more neighbors assigned to their school under the Home-Based Plan. This is a sizable increase because the baseline number of neighbors in the 3-Zone plan is about 3.

Comparing the Optimized plan with the other two, we see that it attains similar savings in busing distance and bus coverage as the Home-Based plan, which is by construction. However, it does not sacrifice expected utility of students compared to the 3-Zone plan, even though the average number of schools in menu is decreased as in the Home-Based plan. It achieves this win-win situation by better optimizing the set of schools each neighborhood has access to and better accounting for the externalities on others of giving one neighborhood access to a school. (See Section 4.4.3 for intuition on how it accomplishes this balancing.)

For predictability, the Optimized plan outperforms both of the other plans. In other words, if a school appears in the menu, then the Optimized plan gives students access to the school with high probability. This is a side product of the fact that the assortment probabilities are found by solving a linear program, which produces basic feasible solutions, which naturally include a few number of non-zeros. For community cohesion, the Optimized plan achieves the gains of the Home-Based plan.

Hence, by optimization, one can produce a plan that dominates both the 3-Zone and Home-Based plans in all the outcomes of interest. The plan uses the smaller busing allowance of the Home-Based plan but gives students utilities higher than the 3-Zone plan. It also achieves better results for predictability and community cohesion, although it did not explicitly optimize for these moments.

Since the optimization involves heuristics, it may not be completely optimal. In Appendix 4.C we upper-bound the possible performance of any plan that uses at most .6 miles of busing per student as in the Home-Based plan, and we show that the average utility is at most 7.8. Note that the Optimized plan achieves 7.5 and also
attains a small bus coverage area.

In Appendix 4.B we evaluate the robustness of these findings under changes in the population and utility distributions. We show that without re-optimizing, most of these gains of the Optimized plan are preserved. This shows that optimization yields gains even if the inputs are not completely correct.

4.5 Conclusion

We show that the menus and priority distributions in school choice systems can be systematically optimized to induce a desirable outcome, which for the Boston case study entails giving students better chances to go to the schools they want while decreasing busing burdens. The plan also improves the system’s predictability and help local communities stay cohesive.

The methodology is based on a new connection between matching and assortment planning (Theorems 1 and 2). This allows us to approximate the optimal school choice problem as a convex program, which can be efficiently solved by iteratively solving a deterministic assortment planning problem for each type of students. This subproblem is analogous to the revenue-maximizing assortment planning problem from the revenue management literature, except that the objective is social welfare rather than revenue, and we adapt algorithms from the literature to efficiently find socially-optimal assortments (Theorem 3).

To apply the optimization methodology in practice, one would estimate the population and utility distributions from past years’ data, and compute optimal menus and priority distributions for the next few years. (Since only the preferences of past years’ students enter into the optimization, the system remains strategyproof.) One would not re-optimize the menus and priority distributions every year, but only do so every 5-10 years.\(^6\)

\(^6\)Having ever changing menus is costly for busing, because each neighborhood needs a bus not only from the schools in the current menu, but in the schools in the past menus, since those assigned in previous years and have not graduated yet also need transportation.
4.A  Additional Tables and Figures

Table 4.3 shows the forecasted proportion of students applying from each neighborhood. Figures 4-1a and 4-1b give a big picture view of the distribution of supply and demand for schools and of inferred school quality in Boston.

Table 4.3: Means and standard deviations of the proportion of K2 applicants from each neighborhood, estimated using 4 years of historical data.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Neighborhood</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allston-Brighton</td>
<td>0.0477</td>
<td>0.0018</td>
<td>North Dorchester</td>
<td>0.0522</td>
<td>0.0047</td>
</tr>
<tr>
<td>Charlestown</td>
<td>0.0324</td>
<td>0.0024</td>
<td>Roslindale</td>
<td>0.0771</td>
<td>0.0048</td>
</tr>
<tr>
<td>Downtown</td>
<td>0.0318</td>
<td>0.0039</td>
<td>Roxbury</td>
<td>0.1493</td>
<td>0.0096</td>
</tr>
<tr>
<td>East Boston</td>
<td>0.1335</td>
<td>0.0076</td>
<td>South Boston</td>
<td>0.0351</td>
<td>0.0014</td>
</tr>
<tr>
<td>Hyde Park</td>
<td>0.0388</td>
<td>0.0022</td>
<td>South Dorchester</td>
<td>0.1379</td>
<td>0.0065</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>0.0570</td>
<td>0.0023</td>
<td>South End</td>
<td>0.0475</td>
<td>0.0022</td>
</tr>
<tr>
<td>Mattapan</td>
<td>0.0759</td>
<td>0.0025</td>
<td>West Roxbury</td>
<td>0.0638</td>
<td>0.0040</td>
</tr>
</tbody>
</table>

4.B  Robustness of the Optimization to Errors in Parameters

The optimization in Section 4.4 depends on distributional assumptions on the student population and preferences. In this section, we evaluate the robustness of the Optimized plan to errors in these assumptions.

The population distribution in Section 4.4.1 is based on data from 2010-2013, and the utility distribution is estimated from students’ submitted preferences from 2013. In this section, we re-evaluate the various plans by using the real population data from 2014 and by using a utility distribution estimated from 2014 preferences. The amount of perturbation in parameters from this computational experiment represent the typical perturbation one may observe from year to year.

The changes in the distributional assumptions are significant. For the population, instead of a forecasted total of 4294 students, only 3964 students applied in 2014. This difference is about 3 times the standard deviation of 115 students in the original population distribution. For the utility distribution, the inferred qualities of schools shifted, with the average shift being 0.69. (Recall that in the utility distribution,
Figure 4-1: The diagram on the left shows the distribution of students and the capacities of schools. Each blue circle represents a neighborhood, with its area proportional to the expected number of students from that neighborhood. Each yellow circle represents a school, with its area proportional to the number of K2 seats available. The capacity schools are shaded. The distribution of students is based on the 2010-2013 average. The capacities are based on data from 2013. The right shows estimates of $Q_s$ (inferred quality) from the 2013 data. The size of the circle is proportional to the estimated $Q_s$, with higher quality schools having larger circles.

Magnitudes are normalized to distance, so this is equivalent to changing students’ travel distances to a school by $\pm 0.69$ miles.) The estimated scale of the Gumbel distribution $\beta$ changed from 1.88 to 1.64, and the estimated effect of coefficient for the walk-zone term $\gamma$ changed from 0.86 to 0.37.

The simulation results using these updated parameters are in Table 4.4. We do not display the average number of choices and the bus coverage area since these are the same as before. Note that although the parameters for evaluation changed, the Optimized plan is based on parameters from before, and has not been re-optimized.

We find that the optimized plan still dominates the 3-Zone and Home-Based plans in busing savings, average expected utility of students, predictability and community
Figure 4-2: These plots show the parameters that define the optimal assortments from Section 4.4.3. Each neighborhood is giving an allowance of points, and each school costs a certain number of points. The diagram on the left shows the distribution of expected allowance $E[a_i]$ across neighborhoods, with each circle representing a neighborhood and the size of the circle proportional to the expected allowance. The diagram on the right plots the schools as circles, with the size of each circle proportional to the school costs. These are not monetary costs, but shadow costs from the optimization (OptAssortBoston) from Section 4.4.3.

cohesion, despite not having the right distributional assumptions as inputs. Moreover, the magnitude of its improvement over Home-Based in these moments are similar to before.

However, the Optimized plan no longer dominates the other two in measure of equity: the 10th percentile of expected utilities of neighborhoods and the minimum expected utility. Figures 4-3a and 4-3b compare the expected utility for each neighborhood from the original simulations and the simulates with the updated parameters, and we find that much of the decrease is in the Hyde Park region of Boston, which is shown using a red oval. To understand what happened, we compare the school qualities from the original and updated utility distributions in Figures 4-4a and 4-
Table 4.4: Re-evaluation of 3-Zone, Home-Based, and Optimized plans using updated parameters from 2014 data. All the results are averages from 10,000 independent simulations.

<table>
<thead>
<tr>
<th></th>
<th>3-Zone</th>
<th>Home-Based</th>
<th>Optimized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. busing distance (miles)</td>
<td>1.1</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Expected utilities of neighborhoods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>6.7</td>
<td>6.3</td>
<td>6.7</td>
</tr>
<tr>
<td>10th perc.</td>
<td>5.8</td>
<td>5.5</td>
<td>5.7</td>
</tr>
<tr>
<td>Lowest</td>
<td>4.4</td>
<td>4.8</td>
<td>4.5</td>
</tr>
<tr>
<td>% getting top choices in menu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 1</td>
<td>62%</td>
<td>61%</td>
<td>82%</td>
</tr>
<tr>
<td>Top 3</td>
<td>85%</td>
<td>83%</td>
<td>96%</td>
</tr>
<tr>
<td>Median # of neighbors co-assigned</td>
<td>3.1</td>
<td>4.5</td>
<td>4.6</td>
</tr>
</tbody>
</table>

and we find that the inferred qualities in 2014 of schools in Hyde Park is much lower than the estimates from 2013. This shows that the equity performance of the Optimized Plan is delicate, and can be easily disturbed if school qualities in a region change in a systematic way. This makes sense because these equity measures, of 10th percentile and minimum, are susceptible to outliers. We leave the problem of finding an assignment plans that preserve equity in a way that is more robust to errors in parameters to future work.

Overall, the results are positive, as the Optimized plan preserves its gains in performance for the majority of the outcomes of interest, even under sizable errors in its distributional assumptions for population and preferences.

4.C Upper-bounds on the Optimality Gap

The Optimized Plan in Section 4.4.3 involve several heuristics, so may not be completely optimal. In this section, we take advantage of the fact that (OptAssortments) is a relaxation of (OptDA) (see Section 4.3.1), and use its optimal objective to upper-bound the optimal objective of (OptDA). We compute the average expected utility achievable by any plan that uses no more miles of busing per student than the Home-Based plan. To do this, we solve (OptAssortBoston) with a few modifications:

- Set a budget $B = 0.6$ for average busing. This is the average busing using in
Figure 4-3: These plots show the expected utilities of neighborhoods under the optimized plan. Each circle represents a neighborhood and the size of the circle is proportional to the expected utility. The left plots the values under the population distribution and utility distributions from Section 4.4.1. The right plots the values under the actual 2014 population and re-estimated utility distribution. The red oval in the plot on the right shows the biggest area of utility decrease. This corresponds to the Hyde Park region of Boston.

the Home-Based plan.

- Remove the MinUtility term from the objective, to only optimize the weighted average utility.

- Relax the assortment constraints to $\Psi_t = [m]$.

By the arguments in Section 4.3.1, the assortment probabilities (defined in terms of accessible assortment for each neighborhood $t$) of any plan that uses less than 0.6 miles of busing per student must be a feasible solution to this revised convex program. This is true in spite of possible randomness in the population vector, and also in spite of the fact that we are not using a large-market approximation. This
Figure 4-4: These plots show the inferred qualities of schools in the utility distribution. Each circle represents a school and the size of the circle is proportional to the inferred quality. The left plots the values in the original utility distribution in Section 4.4.1 which is estimated from 2013 data. The right plots the updated utility distribution based on 2014 data. The red oval shows the decrease in school quality in the Hyde Park region.

A convex program can be solved using the techniques in Section 4.3.2 and the optimal objective is 7.8. The Optimized plan from Section 4.4.4 achieves 7.5, despite its also optimizing minimum utility and limiting bus coverage areas. This shows that the optimality gap, at least in terms of utilitarian welfare of students, is small.
4.D Validation of the Large-Market Approximation for Boston

The assortment planning approximation in Section 4.3.1 is based on a theory requiring the number of students of each type to go to infinity. However, in the empirical exercise, the total number of students is 4294, and there are 868 types, so the average number of students per type is only 5. Nevertheless, there are reasons to expect the large-market approximation to be reasonable. Firstly, neighborhoods that are close to one another tend to have similar menus, utility distributions, and priority distributions, so there are regional pooling effects. Secondly, the independence in preferences make it so that the number of students who prefer a school from a certain area converge quickly to its expectation.

In this section, we empirically test the goodness of the large-market approximation in this data set. We do this by comparing the outcomes of interests in Section 4.4.2 as predicted by the large market model with the outcomes from discrete simulations.

Before showing the results, we first comment on the possible sources of discrepancies between these two types of estimates. The first possible source of discrepancy is that the market size in the Boston data is not large enough for the large-market approximation of Proposition 1 to set in. The second source is that the simulations involve randomness in the student population, while the large-market approximation assumes the number of students of each type is fixed. The third is that the quotas in the large-market approximation should come from Equation 4.11 while the simulation in Section 4.4.4 use the capacity limits instead as a heuristic.

In this section, we focus on the first issue of market size. In order to do this, we modify the population distribution and quotas in order to remove the latter two sources of discrepancy. For the quotas, we use the quotas from Equation 4.11 rounded down to an integer. For the number of students of each type $n_t$, we set it to be as close to the original mean $E[n_t]$. Precisely speaking, let $\mu_t = E[n_t]$, we redefine a
modified population distribution so that

\[ n'_t = \begin{cases} 
\mu_t + 1 & \text{with probability } \mu_t - \lfloor \mu_t \rfloor, \\
\mu_t & \text{otherwise.}
\end{cases} \]

Table 4.5 tabulates the simulation results of the Optimized plan from Section 4.4.3 using the modified population distribution and quotas, and compare with the predictions from the large-market model. As can be seen, the estimates are all very similar, with the largest discrepancy coming from the minimum expected utility of neighborhoods. This makes sense because the minimum is equivalent to the \( \frac{100}{868} \)th percentile, which has a larger variance and requires a larger sample to converge to the mean. For all of the other moments, the simulation results are within 3% of the large market estimate. This shows that the market size in Boston is large enough for the large-market approximation to be adequate.

<table>
<thead>
<tr>
<th>Large-market</th>
<th>Simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. busing distance (miles)</td>
<td>0.56</td>
</tr>
<tr>
<td>Expected utilities of neighborhoods</td>
<td></td>
</tr>
<tr>
<td>Weighted average</td>
<td>7.62</td>
</tr>
<tr>
<td>10th perc.</td>
<td>6.79</td>
</tr>
<tr>
<td>Lowest</td>
<td>6.79</td>
</tr>
<tr>
<td>% getting top choices in menu</td>
<td></td>
</tr>
<tr>
<td>Top 1</td>
<td>89%</td>
</tr>
</tbody>
</table>

4.4 Omitted Proofs

Proof of Lemma 4.4.3 For convenience, denote the supremum on the right hand side of Equation 4.24 as RHS, and the optimal objective on the left as LHS.
First, we show that \( \text{RHS} \leq \text{LHS} \). This is because for any \( \lambda_0 \in \mathbb{R} \), let \((x_0, y_0) \in A(\lambda_0)\), then \((x_0, y_0) \in \{(x, y) \in R : y = f(\lambda_0) + \lambda_0 x\}\). This implies that

\[
h(\lambda_0, f(\lambda_0)) \leq g(x_0, y_0) \leq \text{LHS}.
\]

Conversely, we show that \( \text{LHS} \leq \text{RHS} \). If \((x^*, y^*)\) is an optimal solution for the LHS, with optimal objective \( c = \text{LHS} \). Consider the lower contour set

\[B = \{(x, y) \in R : g(x, y) \leq c\}\]

By the optimality of \( c \), we have \( D \subseteq B \). Since \( g \) is quasi-convex and \( R \) is convex, \( B \) is a convex subset of \( \mathbb{R}^2 \). Since \( g \) is increasing in \( y \), \((x^*, y^*)\) cannot be in the interior of \( B \), but must lie on its boundary. By duality of convex sets, there exists an outward pointing normal of \( B \) at \((x^*, y^*)\) with direction \((-\lambda_0, 1)\). (The \( y \)-coordinate is 1 without loss of generality because \( g \) is strictly increasing in \( y \).) Let \( f_0 = y^* - \lambda_0 x^* \), then we have that \( B \), and also \( D \) is contained in the half-plane:

\[
\{(x, y) : y - \lambda_0 x \leq f_0\}.
\]

We now show that \( h(\lambda_0, f(\lambda_0)) = g(x^*, y^*) \), from which it would follow that \( \text{LHS} \leq \text{RHS} \). First, note that the above implies that \( f(\lambda_0) = f_0 \), so \((x^*, y^*) \in A(\lambda_0)\). Second, since \((x^*, y^*) \in \{(x, y) \in R : y = f(\lambda_0) + \lambda_0 x\}\), we have by the definition of \( h \) that \( h(\lambda_0, f(\lambda_0)) \leq g(x^*, y^*) \). Now, suppose on the contrary that \( h(\lambda_0, f(\lambda_0)) < g(x^*, y^*) \), then there must exist \((x_0, y_0) \in R\) such that \( g(x_0, y_0) < c \) and \( y_0 - \lambda_0 x_0 = f(\lambda_0) \). Since \( R \) is open and \( g \) is continuous and increasing in \( y \), there exists a sufficiently small \( \epsilon_0 \), such that if \( y_1 = y_0 + \epsilon \), then \((x_0, y_1) \in R\), \( g(x_0, y_1) < c \) and \( y_1 - \lambda_0 x_0 > f_0 \). Therefore,

\[
(x_0, y_1) \in B \text{ but } y_1 - \lambda_0 x > f_0,
\]

which is a contradiction because \( B \) is contained in the half-plane specified by Equation 4.36. Therefore, \( h(\lambda_0, f(\lambda_0)) = g(x^*, y^*) \), as desired.
This shows that LHS = RHS. By the above construction for \( \lambda_0 \) from optimal \((x^*, y^*)\), we get that \( \lambda_0 \) is an optimal solution to the RHS, with \((x^*, y^*) \in A(\lambda_0)\). Finally, for any optimal solution \( \lambda^* \) for the RHS, for any \((x_0, y_0) \in A(\lambda^*)\), the argument in the first paragraph shows that \((x_0, y_0)\) is also an optimal solution to the LHS.

4.F Socially-Optimal Assortment Planning for Other Utility Distributions

Given a set \([m] = \{1, \cdots, m\}\) of options and an outside option 0, a utility distribution \(F\), a constraint set \(\Psi \subseteq 2^{[m]}\) of allowable assortments, and an externalities \(r_j\) for every option \(j \in [m] \cup \{0\}\). The socially-optimal assortment planning problem is to find an allowable assortment that maximizes social welfare, which is defined as a weighted sum of the expected utilities of agents and the expected externalities:

\[
\max_{S \in \Psi} \alpha V(S) + \sum_{j \in [m] \cup \{0\}} P(j, S) r_j, \tag{4.37}
\]

where \(\alpha > 0\) is the weighted of the expected utilities term in the objective. Moreover, \(V(S)\) is the value of assortment \(S\) for agents and \(P(j, S)\) is the probability of choosing option \(j\) given assortment \(S\). These are induced by the utility distribution \(F\) as follows: if \(\bar{u} \sim F\), then

\[
V(S) = E[\max_{j \in S \cup \{0\}} u_j], \tag{4.38}
\]

\[
P(j, S) = \begin{cases} P(u_j = \max_{j' \in S \cup \{0\}} u_{j'}) & \text{if } j \in S \cup \{0\}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.39}
\]

This generalizes the revenue-maximizing assortment planning problem, which has \(\alpha = 0\).

In section 4.3.2, we propose an algorithm for the above problem with a MNL utility model. The algorithm is polynomial-time solvable under matroid constraints.
In this section, we study the following utility distributions:

1. The Generalized Extreme Value (GEV) model of Mcfadden (1978), which generalizes MNL and nested logit models.


The extensions to the latter two models are necessary because the original model only defines choice probabilities, but not the assortment valuations.

We give polynomial-time algorithms for the model nested logit under cardinality constraints. For the unconstrained case, we give efficient algorithms for the the multi-level nested logit model and the Markov chain based model.

4.F.1 Generalized Extreme Value (GEV) Model

Assortment Valuation and Choice Probabilities

In the MNL model, the utility of agent $i$ for option $j \in [m] \cup \{0\}$ is distributed as

$$u_{ij} = \bar{u}_j + \epsilon_{ij},$$

where $\bar{u}_j$ is a constant and $\epsilon_{ij}$ is i.i.d. Gumbel distributed, with location parameter zero and scale parameter $\beta$. The independence in the $\epsilon_{ij}$’s gives this model the Independence of Irrelevant Alternatives (IIA) restriction, which says that regardless of the assortment $S$, as long as options $j$ and $k$ are both in the assortment, the ratio $\frac{P(j,S)}{P(k,S)}$ is constant. This severely limits the type of substitution patterns possible, as adding a new option to the assortment would always attract market shares away from options $j$ and $k$ in a proportional way.

A model that bypasses this restriction is the Generalized Extreme Value (GEV) model, which has the same form as above, except that the $\epsilon_{ij}$’s are no longer independent. In the GEV model, the errors $\epsilon_{i0}, \epsilon_{i1}, \ldots, \epsilon_{im}$ are jointly distributed according
to the following CDF,

\[ F(\epsilon_{ij} \leq \delta_j \text{ for all } j \in [m]) = \exp(-G(e^{-\delta_0}, e^{-\delta_1}, \ldots, e^{-\delta_m})), \]

where \( G : \mathbb{R}^{m+1} \rightarrow \mathbb{R} \) satisfies the following properties:

1. **Non-negativity on the positive orthant:** \( G(\bar{w}) \geq 0 \) for \( \bar{w} \geq 0 \). (\( \bar{w} \) is a \((m+1)\)-dimensional vector and \( \bar{w} \geq 0 \) means that every component is non-negative.)

2. **Homogeneous of degree \( 1/\beta \):** For all \( \alpha \geq 0 \), \( G(\alpha \bar{w}) = a^{\frac{1}{\beta}} G(\bar{w}) \).

3. **Differentiable**, with mixed partial derivatives satisfying the sign restriction:
   \((-1)^k \frac{\partial^k G}{\partial \bar{w}^k} \leq 0.\)

We call such a function \( G \) a **GEV generating function with scale** \( \beta \). The MNL model is a special case of GEV model with generating function

\[ G(\bar{w}) = \sum_{j \in [m] \cup \{0\}} w_j^{\frac{1}{\beta}}. \quad (4.40) \]

For every option \( j \in [m] \cup \{0\} \), define attraction weight \( w_j = e^{\bar{w}_j} \), and let

\[ w_j(S) = \begin{cases} 
  w_j & \text{if } j \in S \cup \{0\}, \\
  0 & \text{otherwise.} 
\end{cases} \]

Define \( \bar{w}(S) \) to be the vector \((w_0(S), w_1(S), \ldots)\). This is restricting the components in \([m]\)\(\setminus S\) to be zero.

A well-known result for the GEV model is that the assortment valuations and

\[7\text{ The GEV model is originally proposed in McFadden (1978) with scale } \beta = 1, \text{ and the form with arbitrary scale appears in Bierlaire et al. (2003).} \]
choice probabilities are (see Mcfadden (1978) and Bierlaire et al. (2003)):

\[ V(S) = \beta (G(\bar{w}(S)) + \gamma_{\text{Euler}}) \],

\[ P(j, S) = \frac{w_j(S)\partial_j G(\bar{w}(S))}{G(\bar{w}(S))}, \]

where \( \gamma_{\text{Euler}} = .577 \ldots \) is the Euler constant and \( \partial_j G \) is the partial derivative of \( G \) in component \( j \in [m] \cup \{0\} \).

**Solution Method**

Using the above formula for assignment probabilities, the expected externalities of assortment \( S \) is

\[ R(S) = \frac{\sum_{j \in S \cup \{0\}} r_j w_j(S)\partial_j G(\bar{w}(S))}{G(\bar{w}(S))}, \]

and the socially-optimal assortment problem becomes

\[ \max_{S \in \Psi} \alpha \beta G(\bar{w}(S)) + R(S) \] (4.44)

The following theorem generalizes Theorem 3 for the MNL model to all GEV models.

**Theorem 4.** For the GEV utility model with generating function \( G \) with scale \( \beta \), let

\[ f(\lambda) = \max_{S \in \Psi} \{ G(\bar{w}(S))(R(S) - \lambda) \}, \]

\[ A(\lambda) = \arg \max_{S \in \Psi} \{ G(\bar{w}(S))(R(S) - \lambda) \}, \]

\[ \Lambda^* = \arg \max_{\lambda \in \mathbb{R}} \{ f(\lambda)e^{\frac{\lambda}{\alpha \beta}} \}. \]

then an assortment \( S^* \) is an optimal solution to Equation 4.44 if and only if \( S^* \in A(\lambda^*) \) for some \( \lambda^* \in \Lambda^* \).

**Proof of Theorem 4.** The proof is almost identical to that of Theorem 3. Let \( x(S) = G(\bar{w}(S)) \), \( y(S) = R(S)G(\bar{w}(S)) \), \( g(x, y) = \alpha \beta \log(x) + \frac{y}{x} \). We plug in this formula for
$g$ into Lemma 1 as in the proof of Theorem 3 and the desired result follows.

Theorem 4 suggests this solution method for the socially-optimal assortment problem.

1. Find a set $\mathcal{A}$, which we call a set of candidate solutions, which is a subset of $\Psi$ such that for any $\lambda \in \mathbb{R}$, $A(\lambda) \cup \mathcal{A} \neq \emptyset$, where $A(\lambda)$ is from Equation 4.46.

2. Find the optimal assortment among the reduced set $\mathcal{A}$ by enumeration.

This is an efficient algorithm if the set $\mathcal{A}$ can be found efficiently and if $|\mathcal{A}|$ is guaranteed to be small. It turns out that many algorithms for the revenue-maximizing assortment planning problem follow the exact recipe, and since the definition of $\mathcal{A}$ does not depend on $\alpha$, those same algorithms can be adapted to solve the socially-optimal case. The only difference is that the enumeration in step 2 uses a different objective.

For example, Gallego and Topaloglu (2014) and Feldman and Topaloglu (2014a) study the nested logit model, which is a GEV model with a generating function constructed as follows. Let the set of options $[m]$ be partitioned into disjoint sets indexed by $k$. $[m] = \Omega_1 \cup \Omega_2 \cdots$. Each $\Omega_k$ is called a nest.

$$G(\vec{w}) = w_0 + \sum_k \left( \sum_{j \in \Omega_k} w_j \right)^{\nu_k},$$

where $\nu_k \in (0, 1]$ is called the dissimilarity parameter for the $k$th nest. Section 4 of Gallego and Topaloglu (2014) describe a polynomial time algorithm to find a set of candidate solutions $\mathcal{A}$ for the nested logit problem, with a cardinality constraint at each nest. Section 4.2 of Feldman and Topaloglu (2014a) do the same for the version of the problem in which the cardinality constraint can also be across nests. The set $\mathcal{A}$ they find satisfies the conditions we need and is guaranteed to have no more than polynomially many elements (polynomial in $m$ and the number of cardinality constraints). By Theorem 4, this immediately yields polynomial-time algorithms for socially-optimal assortment planning for nested logit utilities with cardinality constraints.
Similarly, Li et al. (2015) define a generalization of the nested logit model with multiple level of nests. In their model, the nest structure is represented by a rooted tree. Each node of the tree corresponds to a subset of options. For each node $k$, let $\Omega_k$ denote the subset of options it represent. Each node $k$ can either be an internal node or a leaf node. If it is an internal node, then it has a set of children nodes, denoted by $\text{children}(k)$. The subset of options $\Omega_{k'}$ of the children $k' \in \text{children}(k)$ are disjoint from one another and partition $\Omega_k$, so

$$
\Omega_k = \bigcup_{k' \in \text{children}(k)} \Omega_{k'}.
$$

Each internal node $k$ also has an associated dissimilarity parameter $\nu_k$. The root node is an internal node corresponding to all the options, $\Omega_{\text{root}} = [m]$.

Under this model, the GEV generating function is

$$
G(\bar{w}) = w_0 + \sum_{k \in \text{children}(\text{root})} G_k(\bar{w}(\Omega_k)),
$$

where the functions $G_k$ are recursively defined for internal nodes:

$$
G_k(\bar{w}) = \left( \sum_{k' \in \text{children}(k)} G_{k'}(\bar{w}(\Omega_{k'})) \right)^{\nu_k}.
$$

For leaf nodes $k$, the function $G_k$ is a MNL generating function limited to components $\Omega_k$ and has scale parameter equal to the product of $\beta$ as well as the dissimilarity parameter of all the internal nodes between the root node and that leaf node. (See Appendix G of Li et al. (2015) for more details on the definition of $G$.)

Section 5 of Li et al. (2015) shows how to efficiently find a set $\mathcal{A}$ of candidate solutions for this problem, and the set they find has cardinality at most $O(m^2)$. This yields a polynomial-time algorithm for socially optimal assortment planning for multi-level nested logit models.
4.F.2 Markov Chain Based Utility Model

Assortment Valuation and Choice Probabilities

The Markov chain model of preferences is proposed by Blanchet et al. (2013) as a tractable approximation to the mixed MNL model, which McFadden et al. (2000) show can approximate any random utility model to any degree accuracy. We extend this preference model in a natural way to add a measure of preference intensity.

The Markov chain based utility model we study is as follows: each agent has an initial utility $v_0$, which is an arbitrary constant. Let $\Omega = [m] \cup \{0\}$. Define a Markov chain with $m + 1$ nodes, in which each node corresponds to an option $j \in \Omega$. There are two sets of parameters in this utility model, an arrival rate $a_j \geq 0$ for each state $j \in \Omega$, which sum to one, and a transition probability $\rho_{kj}$ from each state $k \in [m]$ to each state $j \in \Omega$. Note that there are no transitions out of the outside option.

Given any assortment $S \subseteq [m]$, consider the following stochastic process: customers arrive at each state according to the arrival rates. Whenever they arrive at one of the nodes $S \cup \{0\}$, they leave the system. Otherwise, at each time step, they follow the transition probabilities to their next state, and continue in the system until they arrive at one of the states $S \cup \{0\}$. The assortment valuations and choice probabilities are defined as follows:

$$V(S) = v_0 - E[\# \text{ of time steps before leaving}]$$

$$P(j, S) = \mathbb{P}(\text{The state when they leave the system is } j)$$

Solution Method

We adapt the LP-based approach of Feldman and Topaloglu (2014b) to solve the socially-optimal assortment planning problem with Markov chain utilities and no constraints ($\Psi = 2^m$).

**Theorem 5.** Consider a Markov chain based utility model with arrival probabilities $\bar{a}$ and transition probabilities $\rho$. Let $(x^*, z^*)$ be an optimal basic solution to the linear
program:

\[
\begin{align*}
\text{Maximize:} & \quad v_0 + \sum_{j \in \Omega} r_j x_j - \alpha \sum_{j \in [m]} z_j \\
\text{subject to:} & \quad x_k + z_k = a_k + \sum_{j \in [m]} \rho_{jk} z_j \quad \forall k \in \Omega
\end{align*}
\]

(4.50)

(4.51)

Let \( S^* = \{ j \in [m] : x_k^* > 0 \} \). \( S^* \) is a solution to the socially-optimal assortment planning problem for this utility model.

**Proof of Theorem 5.** As in Section 1 of [Feldman and Topaloglu (2014b)](#), for any assortment \( S \), let \( R(j, S) \) be the steady state rate of people leaving state \( j \). We have that

\[
P(k, S) = \begin{cases} 
  a_k + \sum_{j \in [m]} \rho_{jk} R(j, S) & \text{if } k \in S \cup \{0\} \\
  0 & \text{otherwise.}
\end{cases}
\]

\[
R(k, S) = \begin{cases} 
  a_k + \sum_{j \in [m]} \rho_{jk} R(j, S) & \text{if } k \in [m]\backslash S \\
  0 & \text{otherwise.}
\end{cases}
\]

Therefore, for any assortment \( S \), setting \( x_k = P(k, S) \) and \( z_k = R(k, S) \) yields a feasible solution to the LP. The social welfare of this assortment is \( \sum_{j \in \Omega} r_j x_j + \alpha(v_0 - \sum_{j \in [k]} z_j) \), which is exactly the objective function of the LP. So the optimal social welfare is upper-bounded by the optimal solution of the LP.

Moreover, by Lemma 1 of [Feldman and Topaloglu (2014b)](#), the polyhedron described by inequality (4.51) is such that for any vertex and any \( k \in \Omega \), either \( x_k = 0 \) or \( z_k = 0 \). This implies that if \( S^* \) are as defined in the theorem, then \( P(k, S^*) \) and \( R(k, S^*) \) are exactly given by \( x_k \) and \( z_k \), so we can attain the optimal social welfare with assortment \( S^* \).
4.F.3 Non-Parametric Utility Model

Assortment Valuation and Choice Probabilities

The non-parametric model of preferences in Farias et al. (2013) is that agents are randomly drawn from finitely many customer segments, each of which has a deterministic preference ordering. They show that having such a flexible model may yield large gains in prediction accuracy compared to parametric models. For our setting, we endow each customer segment with a deterministic utility vector, which is necessary as we also need a measure of preference intensity.

Specifically, the model is that there are $K$ customer segments, and for each $k \in \{1, \cdots K\}$, there is a deterministic utility vector $\vec{u}_k$ and a probability $p_k$. The probabilities sum up to one over all $K$ segments. For simplicity, we assume that none of the utility vectors $\vec{u}_k$ has any two components being equal, so there are no ties in preferences.

The assortment valuations and choice probabilities are

\[
V(S) = \sum_{k=1}^{K} p_k \max_{j \in S \cup \{0\}} \{u_{kj}\} 
\]

(4.52)

\[
P(j, S) = \begin{cases} 
\sum_{k=1}^{K} p_k 1 \left( u_{kj} = \max_{j' \in S \cup \{0\}} \{u_{kj'}\} \right) & \text{if } j \in S \cup \{0\}, \\
0 & \text{otherwise.} 
\end{cases} 
\]

(4.53)

Solution Method

Since this utility model generalizes the choice model of Farias et al. (2013) and since the socially-optimal assortment problem is a generalization of the revenue-maximizing problem, the hardness results in Aouad et al. (2015) carry through to our case, so the optimal solution is NP-hard to approximate within any factor $O(\min(m, K)^{(1-\epsilon)})$ for any $\epsilon > 0$.

For small number of options or number of segments, we can adapt the mixed integer program (MIP) formulation of Bertsimas and Misic (2015). There is a binary variable $x_j \in \{0, 1\}$ for whether option $j \in [m]$ should be included in the assortment,
and continuous variable \( y_{kJ} \in [0, 1] \) for whether the \( k \)th segment chooses option \( j \in [m] \cup \{0\} \).

Maximize: \[ \sum_{k=1}^{K} p_k \sum_{j=0}^{m} y_{kJ} (\alpha u_{kJ} + r_j) \]

Subject to:

\[ y_{kJ} \leq x_j \quad k \in [K], j \in [m] \]

\[ \sum_{j' \in [m] \cup \{0\}: u_{kJ} > u_{kJ'}} y_{kJ'} \leq 1 - x_j \quad k \in [K], j \in [m] \cup \{0\} \]

\[ \sum_{j=0}^{m} y_{kJ} \leq 1 \quad k \in [K] \]

\[ x_j \in \{0, 1\} \quad j \in [m] \]

\[ y_{kJ} \geq 0 \quad k \in [K], j \in [m] \cup \{0\} \]

Bertsimas and Misic (2015) report positive computational results for this model in their computational experiments. Since the above formulation is the same as theirs except for the objective function (which adds the term \( \alpha u_{kJ} \)), similar results should hold in this case.
Chapter 5

Optimizing Lottery Implementation to Improve Community Cohesion

5.1 Introduction

In various school choice mechanisms, students submit a ranked list of schools they would like to attend, and schools have priorities over students; a centralized assignment algorithm, which may randomize to break ties, determines the assignment. Such mechanisms are used to assign children to public schools in many metropolitan areas in the US, including Boston, New York, New Orleans, San Francisco, and Chicago.

One drawback of existing school choice systems is that children from the same community end up going to many different schools, thus weakening community ties. Ebbert and Ulmanu (2011) document 19 children on one street in Boston going to 15 different schools, as an example of the community dispersion due to a choice lottery. Community dispersion also raises transportation burdens: Boston Public Schools spent $80 million in 2012 on busing students, which represents almost 10% of its total budget (Sutherland (2012)). In his January 2012 State of the City address, Boston’s mayor Menino said,

“Pick any street. A dozen children probably attend a dozen different schools. Parents might not know each other; children might not play
together. They can't carpool, or study for the same tests. We won't have
the schools our kids deserve until we build school communities that serve
them well.” [Menino (2012b)]

One important facet of building communities is to have children go to school with
others from their community, so the families get to know one another through com-
mon activities. Partly motivated by this, some have advocated abandoning school
choice altogether and switching to a neighborhood-based system, in which kids pre-
dominately go to the closest school. But choice adds value by allowing families to find
an option that fits their individual tastes, and many parents vocally defend their right
to choose. Is it possible to improve community cohesion without sacrificing choice?

The main insight of this chapter is that much of the community dispersion is ar-
tificial, caused purely by the allocation algorithm using independently drawn lottery
numbers. If we had used a “correlated-lottery,” then we could implement the same as-
signment probabilities but improve cohesion. Consider the following example: There
are 2 schools with 2 seats each, and 8 students from 2 communities. Students A, B, C,
D are from community I, and E, F, G, H are from community II. Students A, B, E, F
prefer school 1, and C, D, G, H prefer school 2. Suppose for simplicity that the choice
mechanism is Random Serial Dictatorship– students are ordered uniformly randomly
and they pick schools sequentially according to this order. It is straightforward to
work out the assignment probabilities (see Table 5.1).

Conditional on being assigned, what is a student’s chance of being assigned with
someone else from the same community? Because there are 2 communities and ev-
erything is symmetric, we expect this to be roughly $\frac{1}{2}$. Working out the details, we
get that the answer is in fact $\frac{13}{35} \approx 37\%$.

However, the random assignment in Table 5.2 generates the same assignment
probabilities for everyone, while always keeping communities together.

From each individual’s perspective, the second random assignment is “equivalent”
to the first, in the sense that the individual has the same probabilities of being assigned
to each school as before. As a result, every individual’s expected distance to school,
expected “academic quality” of assigned school, probabilities of getting into some set
Table 5.1: Assignment probabilities from example. For example, student A from community I prefers school 1 over 2, and in the randomized assignment she is assigned to school 1 with probability $\frac{61}{140}$ and school 2 with probability $\frac{9}{140}$. With remaining probability $\frac{1}{2}$, she is not assigned (or assigned to an outside option).

<table>
<thead>
<tr>
<th>Community</th>
<th>Student</th>
<th>Preference</th>
<th>Assignment Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A</td>
<td>1 ≻ 2</td>
<td>School 1: $\frac{61}{140}$, School 2: $\frac{9}{140}$</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1 ≻ 2</td>
<td>School 1: $\frac{61}{140}$, School 2: $\frac{9}{140}$</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2 ≻ 1</td>
<td>School 1: $\frac{9}{140}$, School 2: $\frac{61}{140}$</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>2 ≻ 1</td>
<td>School 1: $\frac{9}{140}$, School 2: $\frac{61}{140}$</td>
</tr>
<tr>
<td>II</td>
<td>E</td>
<td>1 ≻ 2</td>
<td>School 1: $\frac{61}{140}$, School 2: $\frac{9}{140}$</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1 ≻ 2</td>
<td>School 1: $\frac{61}{140}$, School 2: $\frac{9}{140}$</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>2 ≻ 1</td>
<td>School 1: $\frac{9}{140}$, School 2: $\frac{61}{140}$</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>2 ≻ 1</td>
<td>School 1: $\frac{9}{140}$, School 2: $\frac{61}{140}$</td>
</tr>
</tbody>
</table>

Table 5.2: Correlated lottery implementation of the assignment probabilities from Table 5.1. This randomizes over 4 assignments, each represented by a column. For example, assignment i is chosen with probability $\frac{61}{140}$, and assigns A and B to school 1, E and F to 2, and leaves the rest unassigned. Note that no matter what the realized assignment is, communities stay together as much as space allows.

<table>
<thead>
<tr>
<th>Community</th>
<th>Student</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A</td>
<td>i: 1</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>ii: 2</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>iii: 1</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>iv: 2</td>
</tr>
<tr>
<td>II</td>
<td>E</td>
<td>i: 2</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>ii: 1</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>iii: 2</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>iv: 1</td>
</tr>
</tbody>
</table>

Formally, we define the community cohesion of a lottery as the expected number of same-community school peers a student can expect to see. We propose the following approach to increase community cohesion in any randomized allocation mechanism: estimate the assignment probabilities for every student to every school in the current mechanism, and implement the same assignment probabilities in a “community-
correlated” lottery. In other words, we seek a convex combination of deterministic assignments that matches the original assignment probabilities but that maximizes cohesion. We term this optimization problem correlated-lottery implementation.

In this chapter, we address the following questions:

1. For the most prevalent mechanisms used in practice, by how much can we hope to improve community cohesion using a correlated lottery? In what settings can we expect the most improvement?

2. How to solve the correlated-lottery optimization problem in practice?

3. For Boston (where community cohesion in school choice has been the focus of much debate), how much can correlated lottery improve community cohesion? How would such a method interact with other possible reforms?

To help us address the first question, we prove a large market characterization of all mechanisms that satisfy the following four properties: (1) non-atomicity (a single individual has negligible effect on assignment probabilities of others), (2) asymptotic Bayesian incentive compatibility (given distribution of students’ preferences, taking the limit as the market size goes to infinity, students reporting truthfully forms a Bayes-Nash equilibrium), (3) symmetry within each priority class (students with same priorities to every school and same submitted preferences receive the same assignment probabilities), and (4) asymptotic efficiency within each priority class (no trading cycles within each priority class). These properties are generally satisfied by commonly implemented mechanisms, such as Deferred Acceptance (DA) or Top Trading Cycles (TTC) with randomized tie-breakers, when we take a suitable limit as market size is scaled up.\(^1\) We show that any mechanism that satisfies these properties in the large

\(^1\)Following Azevedo and Leshno (2014), a suitable limit would be any sequence of finite economies with market size going to infinity, but with the proportion of students of each preference and priority, as well as the relative capacities of programs, converging to the large market limit. Both DA-STB and TTC-STB satisfy incentive compatibility and symmetry regardless of market size. Azevedo and Leshno (2014) Theorem 1 and 2 imply that except when capacities fall into some measure-zero set, DA-STB satisfies asymptotic non-atomicity always, because another student changing her preferences only affects me if her change is affecting the allocation of the “last seat” at a school, which happens with negligible probability. Both mechanisms satisfy efficiency within each priority class because in the large market limit, the
market can be interpreted as a “lottery-plus-cutoff” mechanism, which means that students are divided into priority classes and are each given an identically distributed lottery number; given the distribution of preferences, schools set a lottery cutoff for each priority class; students are assigned to their most preferred school for which they meet the lottery cutoff. This generalizes a result by Liu and Pycia (2012) to allow for priorities.

Using this characterization, we derive clean expressions for cohesion with independent lotteries and optimally correlated lotteries. In a large market framework, we show that baseline cohesion (using independent lotteries) is equal to the sum of a measure of variation in school size and between-community variation in assignment probabilities. Under an additional assumption that each community has the same priority to all schools, we show that maximum cohesion (using optimally correlated lotteries) is the sum of baseline cohesion and the average variance of a certain “demand function” of communities for schools. This improvement term can be interpreted as a measure of preference correlation, with greater improvement under higher uncertainty of lottery or under higher within-community preference correlation. Under a random utility model, we show that if there are no priorities or between-community variation, cohesion gain from lottery correlation increases when preferences are more correlated.

We address the second question by showing that the problem is NP-hard to solve optimally and introduce a heuristic that performs reasonably well in practice. The underlying optimization problem is related to the Quadratic Assignment Problem, which is in general notoriously intractable (see Burkard et al. (1998)), but there is more structure in our case, which is exploited in our heuristic. We also derive an upper-bound to test the optimality gap of our heuristic.

We address the third question by applying our heuristic to real data from Boston elementary school choice, simulating what would have happened if we implemented lottery correlation in 2012 Round 1 assignment. The main grades under consideration are kindergarten 1 (K1) and kindergarten 2 (K2). Defining each community to be a distribution of preferences of students of the same priority class is fixed regardless of lottery number, and this precludes trading cycles within each priority class.
.5 mile by .5 mile square, we show that our method improves community cohesion by 79% for grade K1 and 37% for K2. Conditional on the student traveling outside their walk-zone (1-mile radius), cohesion improves by 140% for K1 and 64% for K2.

We also compare our approach to reforms discussed by a mayor-appointed city committee during the 2012-2013 Boston school choice reform. The main reforms discussed were to increase the walk-zone percentage and to reduce the choice menu. As of 2012, school programs were split into two halves, one half that prioritizes students living within 1 mile (walk-zone), and one half that does not have this priority. For most programs, the walk-zone half represented 50% of seats. By increasing this percentage, policy makers can induce a closer-to-home assignments, thus increasing cohesion. However, we showed that even if we had made the percentage 100%, the gain in community cohesion would not be as much as if we had kept the walk-zone percentage unchanged but used correlated lottery. Furthermore, while increasing the walk-zone percentage would not increase the number of same-community-peers for students traveling out of their walk-zone, correlated lottery would.

The other reform discussed by the city committee was to reduce the choice menus of students. During this process many plans for choice menus were proposed, some involving dividing the city into more assignment zones, and others involving a customized menu that depended on students' addresses. Using simulated choices from a discrete choice model fitted with real data, we show that the cohesion gain from lottery correlation is comparable to sizable reductions in choice menus. More interestingly, the two strategies amplify one another. For example, consider the choice menu called “Home Based A.” (This was the choice menu eventually chosen by the city committee.) If we were to apply this choice menu reform alone, we would improve cohesion by 46% for K1 and 30% for K2. If we were to apply correlated-lottery alone, the cohesion gains are 88% for K1 and 39% for K2. So correlated-lottery achieves more gain. However, if we were to apply both reform at the same time, cohesion would more than triple for K1 and more than double for K2. So the number of neighbors students can expect to see at their assigned school would dramatically increase.

We also analyze the geographic distribution of cohesion gains due to correlated
lottery, and show that while performing correlated lottery alone yields uneven gains for K2, this would be largely mitigated if we simultaneously reduced the choice menu. Furthermore, we show that lottery correlation has minimal impact on racial or social-economic diversity. These analyses are in the Electronic Companion.

5.1.1 Related work

While there is much existing literature on school choice (see Abdulkadiroğlu and Sönmez (2003b); Abdulkadiroğlu et al. (2009, 2006); Pathak (2011)), most of the literature focuses on individual students’ assignments and ignore the correlations between different students’ assignments. One reason for this is that complementarities in matching is difficult to analyze theoretically.

The idea that different random assignments can represent the same assignment probabilities has appeared before in the literature. Abdulkadiroğlu and Sönmez (2003a), note that such random assignments may differ in their ex-post efficiency. However, in their setup there is no guideline to decide between such random assignments. In our setup, there is the added performance measure of community cohesion. Piantadosi et al. (2007) and Asadpour and Saberi (2010) seek for an “optimal” convex combination of assignments that yields the highest entropy while maintaining the same assignment probabilities. In their setting, they maximizes a concave function, which is computationally tractable. However, in our setting, we seek to maximize a convex function, and the optimization is NP-hard. This difficulty cannot be avoided by an alternative definition of cohesion because cohesion is inherently convex, as it corresponds to having greater variation between assignments (either have children from a neighborhood mostly go to one school or mostly go to another).

Our approach of defining a mechanism by implementing the marginal assignment probabilities is similar to Budish et al. (2013). They study a more general framework and address the issues of group-specific quotas, ex-ante efficiency, ex-post fairness, and implementability of lotteries under general constraints. However, their techniques do not handle issues involving complementarities, such as community cohesion, and our work expands on the applicability of their framework in this domain. Another
work that studies the decomposition of assignment probabilities into deterministic assignments to achieve certain properties is [Pycia and Ünver (2012)].

Randomization has been much studied in school choice mechanisms. The currently most adopted mechanisms break ties in school’s preferences for students using independently generated lottery numbers, and apply the deferred acceptance (DA) algorithm or the top trading cycles (TTC) algorithm. (See [Abdulkadiroğlu and Sönmez (2003b)].) [Abdulkadiroğlu et al. (2009)] study whether to use a single tie-breaker for all schools or different tie-breakers for different schools, and their simulations using New York City data shows that single tie-breaking is better. [Pathak and Sethuraman (2011)] show that in the absence of school priorities both tie-breaking methods are equivalent. [Erdil and Ergin (2008)] illustrate the potential ex-ante inefficiencies from running an ex-post efficient mechanism after random tie-breaking, and propose a method to deal with such inefficiencies, but the resultant mechanism is no longer incentive compatible. [Azevedo and Leshno (2010)] show that this proposed improvement may yield Nash equilibria in which the outcome is Pareto-dominated by the original mechanism. [Che and Kojima (2011)] show that in the large market, without school priorities, the deferred acceptance algorithm with a randomized tie-breaker is equivalent to the ordinal efficient probabilistic serial mechanism, which [Liu and Pycia (2012)] show is equivalent in the large market to any asymptotically efficient, symmetric, and asymptotically strategyproof ordinal allocation mechanisms. These works suggest that there may be little room for improvement over the status quo in terms of individual students’ welfare, given requirements of strategyproofness and fairness. Our work is different from these because we focus on community cohesion, which can be seen as “orthogonal” to students’ individual assignment probabilities to schools. We show that there is in fact much room for improvement in this direction, while also maintaining most of the good properties of the current mechanisms.

In terms of implementing other social objectives in school choice, there has been previous work in “controlled school choice,” most of which focus on achieving diversity (see e.g., [Ehlers (2010)], [Ehlers et al. (2011)], [Echenique and Yenmez (2012)] and [Kominers and Sönmez (2015)].)
A recent paper that studies neighborhood interactions in school choice is Weiwei (2013), which uses secondary school choice data from New York City to show that students tend to choose similar schools as their immediate neighbors. This provides empirical support that students value going to school with their neighbors.

Outside of school choice, there has been other studies of assignment externalities. On the theoretical side, general settings of matching with externalities usually yields negative results. (For example, in many-to-one matching of workers to firms, if preferences can be over colleagues then the game theoretic core can be empty. See Echenique and Yenmez (2007) for an example. Klaus and Klijn (2005) show a similar impossibility result even when joint preferences are between only two agents.) One empirical study that has similar flavor to ours is Mariagiovann et al. (2012), in which they consider the assignment of faculty members to offices in a US professional school, and study how institutional and social ties between faculty affect their choices and the final assignment. They quantify the effects of these network externalities and assess the matching protocol from a welfare perspective.

5.2 Model

There are $n$ students to be assigned to $m$ schools. The set of individual students is $I$ and the set of schools is $S$. Each school $s \in S$ has capacity $q_s$. Without loss of generality, we assume that all students must be assigned. This is because we can model unassignment if needed by including a dummy school $s_0$ with infinite capacity to denote “unassigned.”

The students are partitioned into $k$ disjoint communities:

$$I = I_1 \cup I_2 \cdots \cup I_k.$$ 

A community-membership function $c : I \rightarrow \{1, \cdots k\}$ maps each student to the index of the community she belongs to. (For clarity of exposition, we refer to students using the feminine gender and the social planner using the male gender.) As a slight abuse
of notation, we may also index communities using $c$.

An assignment $a$ is a mapping that takes students to schools, and we require that no school capacity is violated. Formally, $a : I \rightarrow S$, $|a^{-1}(s)| \leq q_s$, where $a^{-1}(s) = \{i : a(i) = s\}$. A random assignment $x$ is a random variable whose realizations are assignments. Denote the set of all random assignments $X$.

Slightly abusing notation, we sometimes represent $a$ as an indicator matrix, in which $a_{is}$ is 1 if and only if student $i$ is assigned to school $s$. In this notation $x_{is}$ becomes a binary random variable for whether $i$ is assigned to $s$. And $p = E[x]$ becomes the matrix of assignment probabilities. ($p_{is}$ is probability student $i$ is assigned to school $s$.)

For each student $i$ and school $s$, define the school-specific utility of $i$ for $s$ to be $u_{is}$. Abuse notation slightly and define $u_i(a) := u_{ia(i)}$, which is student $i$’s school-specific utility under assignment $a$. Define $v_i(a)$ to be $i$’s number of same-community peers under assignment $a$. This is the number of other students assigned to the same school and from the same community, and can be written as $v_i(a) = |(a^{-1}(a(i)) \cap I_c(i)) \setminus \{i\}|$.

Assume that student $i$’s preference over random assignments is induced lexicographically by the ordered pair

$$(E[u_i(x)], E[v_i(x)]).$$

So students in our model care foremost what school they are assigned to, and within a given school they prefer to be assigned with more peers from their community. We call this preference structure weakly community-preferring. We assume that for each student $i$, $u_{is}$ is different for different $s$. Hence, the $u_{is}$’s induce for student $i$ a complete strict ordering $\succ^*_i$ over schools, in which her more preferred schools are ranked better. We call $\succ^*_i$ the true preference ranking of student $i$.

One intuitive measure of community cohesion under assignment $a$ is simply the average number of same-community peers assigned to the same school. We define
\( f(a) \), the cohesion of assignment \( a \), as follows:

\[
f(a) = \frac{1}{n} \sum_i v_i(a) = -1 + \frac{1}{n} \sum_i \sum_{j \in I_i} 1\{a(i) = a(j)\},
\]

where \( 1\{a(i) = a(j)\} \) is the indicator for students \( i \) and \( j \) being assigned together. Note that \( nf(a)/2 \) is simply the total number of pairs of students from the same community who are assigned to the same school. The cohesion of a random assignment \( x \) is defined as \( E[f(x)] \). (In the example in the introduction, the expected cohesion of the first lottery is .37 and of the second is 1.)

An assignment mechanism \( M \) is a function that takes as input a strict ranking of schools \( \succ_i \) from every student, and outputs a random assignment. Note that the function \( M \) may implicitly incorporate various rules for prioritizing one student over another, but these are given a-priori and are therefore not treated as inputs.

We define the max-cohesion correlated lottery implementation of mechanism \( M \) as a mechanism induced by the following maximization program:

\[
M_{\text{max-cohesion}}(\text{input}) = \arg \max_x E[f(x)] \tag{5.1}
\]

\[
\text{s.t. } E[x] = E[M(\text{input})] \tag{5.2}
\]

\( x \in X. \)

This optimization maximizes cohesion subject to maintaining the same assignment probabilities as the original mechanism. This is equivalent to maximizing the second component of expected social welfare in our utility model, which implies that if the original mechanism is Pareto-efficient with respect to school-specific utilities, the correlated lottery implementation becomes Pareto-efficient with respect to the full lexicographic utility. By inheriting the assignment probabilities from the original mechanism, we preserve any statistic of the original mechanism that can be expressed in assignment probabilities, including expected distance to assignment, probability of attending a certain set of schools, and expected value of the first component of the
preference tuple. This is attractive because the decisions of what choices students
have and who gets what priority is often the result of an intensely-debated political
process, so being able to maintain the exact same assignment probabilities for every-
one, while improving community cohesion, makes this approach less controversial.

5.2.1 NP-Hardness even with 2 schools

Consider the case of 2 schools. Let the schools be labeled 1 and 2, with capacities $q_1$
and $q_2$ respectively. Suppose that both schools are acceptable to every student, and
that the total number of seats matches the number of students, $q_1 + q_2 = n$.

Without loss of generality, let school 1 be the over-demanded school. We assume
that students’ priorities for school 1 are based on their “priority class,” and a higher
priority student always takes precedence over a lower priority student. For students
from the same priority class, we require that if they both prefer school 1, then they
get in with the same probability.

Despite the generality of the priority structure, we can characterize the structure
of the assignment probabilities. There is a “cutoff priority level” for school 1 at which
any student who prefers school 1 with higher than cutoff priority will get in school
1; any student with lower than cutoff priority level or prefers school 2 will get into
school 2; students who who prefer school 1 with exactly the cutoff priority level will
be allocated based on a fair lottery. Respectively denote these sets of students $D$ (get
in school 1 for sure), $F$ (get in school 2 for sure), and $E$ (allocated based on lottery).
Define the number of seats to be assigned by lottery to be $q^* = q_1 - |D|$. Students in
$E$ are assigned to school 1 with probability

$$p = \frac{q^*}{|E|},$$

and school 2 otherwise. The random assignment is illustrated in figure 5-1.

An upper-bound on cohesion is if communities in $E$ are always assigned together.
In order to achieve this, we need to be able to partition $E$ into subsets of size $q^*$,
which is a hard “packing” problem. So it is computationally hard to decide whether
Figure 5-1: Random assignment in the 2-school case: students who prefer school 1 with higher than cutoff priority level (set $D$) get in for sure; students who prefer school 1 with exactly the cutoff priority level (set $E$) get in with probability $p = \frac{q}{|E|}$; students who prefer school 2 or whose priority level is lower than cutoff (set $F$) get in school 2 for sure.

this upper-bound can be achieved.

**Proposition 2.** Unless $P=NP$, even with 2 schools, there exists no polynomial time algorithm (or FPTAS) to compute the maximum achievable cohesion by correlated lottery implementation.

However, if there are many communities and no single community dominates in size, one can show that we can achieve this upper-bound approximately. By analyzing the “large market limit” as the number of communities go to infinity, one can obtain an exact formula for max-cohesion. We defer this to the Online Appendix because the insights are similar to the insights from the large market analysis in Section 5.4 which uses a model that can encompass many schools.

This NP-hardness result precludes us from having nice expressions for cohesion from correlated lottery in the finite market model. To circumvent this difficulty, we “smooth away” the NP-hardness by considering the “large market” environment, in which there is a continuum of communities. Although in practice the number of communities is finite, the continuum model allows for simple analytic expressions for maximum cohesion and improvement by correlated lottery, thus illustrating the underlying insights in a clean way.

In Section 5.3, we setup a large market model and prove a useful characterization result, which may be of independent interest. In Section 5.4, we use this characterization to gain insights about how much we can gain in cohesion from lottery correlation.
and in what environments we can expect the most improvement.

5.3 Large market characterization of reasonable one-sided matching mechanisms with priorities

In this section, we define a large market model for one-sided matching markets with priorities. We show that in this setup, any mechanism that satisfies certain regularity conditions (non-atomicity, Bayesian incentive compatibility, symmetry and efficiency within each priority class) can be interpreted as “lottery-plus-cutoff”: each student is given an identically distributed lottery number and each school sets a lottery-cutoff for each priority class (the cutoffs may depend on the distribution of preferences for each priority class); a student is assigned her most-preferred school for which she meets the cutoff. This sets up the basis for our analysis of cohesion in Section 5.4. However, this characterization itself does not have to do with communities or cohesion.

5.3.1 Large market model

Let the set of students, $I$, be represented by a subset of Euclidean space of Lebesgue measure 1. Let $S$ be a finite set of schools, with $|S| = m$. For $s \in S$, let $q_s$ be the school’s capacity.

As before, each student $i$ submits preferences $\succ_i$, which is a ranking of schools. We assume that every school is acceptable to every student, and that every student ranks all schools. There are $m!$ possible rankings, and we assume that for each possible ranking, the set of students that submit this ranking is measurable.

Our model is one-sided matching with priorities, which means the following. Students are partitioned into priority classes $\Pi$. We assume that for each priority class, the set of students of that priority class is measurable and of positive measure. Furthermore, the distribution of priority classes and the distribution of preferences within each priority class is common knowledge.

Given students’ preferences and priorities, an assignment mechanism is repre-
sented by a random indicator function $1_s(i)$, which equals 1 if student $i$ is assigned to school $s$, and 0 otherwise. We assume that the assignment mechanism satisfies the following regularity conditions in the large market setting:

- Non-atomicity: Any single student changing her preferences has no effect on the assignment probabilities of others.

- Bayesian incentive compatibility in school-specific utilities: Given the distribution of students’ preferences within each priority class, students reporting truthfully is a Bayes-Nash equilibrium. In other words, given the measure of students of each priority class and of each possible preference ranking, assuming that everyone else reports truthfully, a student cannot improve her expected school-specific utility $E[u_i]$ by submitting a false preference. Henceforth we denote this simply by “incentive compatibility.”

- Symmetry: Students in the same priority class with the same preferences receive the same assignment probabilities.

- Efficiency within each priority class: For students in the priority class, there does not exist a Pareto improvement “trading cycle” where by $s_0 \prec_i s_1$, $s_1 \prec_i s_2 \cdots s_l \prec_i s_0$ and $i_0$ has positive probability of being assigned $s_0$, $i_1$ has positive probability for $s_1$, etc.

Note that since these conditions only need to hold in the large market, one can interpret them as only needing to be “asymptotically” true. This means that for example, the “incentive compatibility” condition in the above is actually the less-restrictive incentive compatibility “in the large” condition described in Azevedo and Budish (2012). Taking the suitable limit and ignoring “knife-edge” cases, two of the most widely used mechanisms—Deferred Acceptance with Single Tie-breaking (DA-STB) and Top Trading Cycle with Single Tie-breaking (TTC-STB), assuming that students are allowed to rank as many choices as they would like, both satisfy the above conditions in the large market, so both fall into our framework.² DA-STB is used in

²See endnote 1
Boston, New York, Denver and San Francisco. TTC-STB is used in New Orleans. For descriptions of these mechanisms, see Abdulkadiroğlu and Sönmez (2003b) and Abdulkadiroğlu et al. (2009).

This model is similar to the continuum two-sided matching model in Azevedo and Leshno (2014). The main difference is that preference structure is one-sided in our model and two sided in their model: only students have preferences over schools in our model, while schools also have strict preferences over students in their model. In their paper, they show that a continuum two-sided matching model can be interpreted as a limit of discrete matching models. This provides a theoretical foundation for such continuum models. Other works that use continuum matching models include Abdulkadiroğlu et al. (2008), Miralles (2008), and Budish and Cantillon (2012).

5.3.2 A characterization theorem

We show that in the large market model, any mechanism that satisfies our four regularity conditions can be described as a “lottery-plus-cutoff” mechanism: each student gets an identically distributed lottery number and each school sets a lottery-cutoff for every priority class; a student is assigned her most preferred school for which her lottery meets the cutoff.

**Definition 4.** A mechanism is lottery-plus-cutoff if it can be described as follows: given preference submissions, each student $i$ receives an identically distributed lottery number $z_i$ (may be jointly correlated). WLOG, $z_i \sim \text{Uniform}[0,1)$. Given the measure of students in each priority level submitting each ranking, schools have a priority-dependent lottery cutoff $z^*_{\pi,s}$. Student $i$ is assigned her most preferred school for which she meets the cutoff ($z_i \geq z^*_{\pi(i),s}$).

If in addition the lottery numbers $z_i$ are independently generated for different students, then we call the lottery independently implemented.

**Theorem 6.** In the continuum model, an assignment mechanism is non-atomic, Bayesian incentive compatible, symmetric and efficient within each priority class if and only if it is a lottery-plus-cutoff mechanism.
The core ideas behind the proof appeared previously in Liu and Pycia (2012), which shows that without priorities and in the large market, any mechanism that is non-atomic, strategyproof, symmetric and efficient is equivalent to the so-called probabilistic serial mechanism, or equivalently lottery-plus-cutoff with one priority class. The main difference with our result is that while their setup does not have priorities, we allow arbitrary priorities. Moreover, while their analysis studies the limit as a discrete model is scaled up, our analysis is directly in the continuum setting. Our proof is given in the Online Appendix 5.C.

One way to interpret the above result is that in a large one-sided many-to-one matching market, asymptotic incentive compatibility and asymptotic efficiency within each priority class constrain the mechanism significantly, leaving policy makers with only two control levers: (1) cutoffs for each priority class and (2) lottery correlation. Assuming in addition that the mechanism does not waste desirable resources, then the first lever, determining cutoffs, is a purely distributional question: how much claim does each priority class have on each resource and whether such claims can be traded. The second lever, lottery correlation, is what we study in this chapter.

5.4 Cohesion in the large market model

The characterization in Section 5.3.2 allows us to study the impact of lottery correlation in a wide class of mechanisms. Once we know that a mechanism is “lottery-plus-cutoff,” we can isolate the effects of lottery-correlation by treating the cutoffs as given. By doing this, we insulate the analysis from the complexities of how the mechanism treats different priority classes, as all such subtleties are endogenized into the cutoffs.

We first define communities and cohesion in the large market model. Let the unit interval $C = [0, 1]$ represent a continuum of communities. (We need infinitely many communities in order to show analytic results, because the hardness result in Section 5.2.1 still hold even if we had infinitely many students per community but a

\[3\] In DA-STB, the claims cannot be traded, while in DA-TTC, the claims can.
finite number of communities.) Let the set of students within each community be also represented by the unit interval \([0, 1]\), so the set of all students can be represented by the Cartesian product of two unit intervals, which is the unit box, \(I = [0, 1]^2\). Each student \(i \in I\) can be represented by its coordinates \((c, y)\), in which \(c\) is the student’s community and \(y\) is the student’s index in this community. (This implicitly assumes that communities have equal size, but this can be relaxed.) The community membership function \(c(i)\) is simply the first coordinate of \(i\). For simplicity, assume that the total supply of seats equals total demand, so \(\sum_s q_s = 1\).

Define the cohesion for school \(s\), \(f^s = E[\mathbb{1}_s(i)\mathbb{1}_s(i')|c(i) = c(i')]\). This is the expected measure of same community pairs assigned to \(s\). Define \(f = \sum_s f^s\) as the total cohesion, which is the measure of pairs of students from the same community assigned to the same school.

Our analytical results on max-cohesion correlated lottery require an additional assumption, which we call homogeneity of cutoffs within communities.

**Assumption 1.** For each school, everyone from the same community sees the same cutoffs to this school.

This would hold if everyone in the same community belongs to the same priority class, which would be the case if communities and priorities were purely geographically based and communities stay together in any geographic division. Under this assumption, maximum community cohesion can be attained by giving everyone in the community the same lottery number. Without this assumption, all of the expressions for max-cohesion in this section would still hold as upper-bounds, but they might not be attainable.

Given students submitted preferences and the mechanism, define the following quantities:
\[ p_s(i) := \text{probability of student } i \text{ of being assigned to school } s. \text{ For } i = (c, y), \text{ we also write } p_s(c, y) := p_s(i). \]

\[ 1_s(z|i) := \text{Indicator for whether student } i \text{ at lottery } z \text{ would be assigned to school } s. \text{ (It is 1 iff at lottery } z, \text{ student } i \text{ finds school } s \text{ to be her most desirable option for which she meets the lottery cutoff.) We call this student } i\text{'s demand function for school } s. \]

\[ \bar{p}_s(c) := \int_y p_s(c, y)dy. \text{ The total assignment probabilities to } s \text{ of students in community } c. \]

\[ d_s(z|c) := \int_y 1_s(z|(c, y))dy. \text{ The total demand for } s \text{ from community } c \text{ at lottery } z. \text{ We call this community } c\text{'s demand function for } s. \]

These quantities, as well as the cutoff structure, are illustrated in Figure 5-2. This graphs the students in a community on the horizontal axis and the lotteries on the vertical axis. It is a “slice” of a cube which would correspond to \( I = [0, 1]^2 \) in two dimensions and lottery \( z \in [0, 1] \) in the third dimension.

Note from Figure 5-2 that \( \bar{p}_s(c) = E_z[d_s(z|c)] \), because both are equal to the “area” assigned to school \( s \). The first is integrating horizontally, and the second is integrating vertically.

The following theorem provides a simple analytical formula for cohesion to any given school in the independent and max-cohesion lottery implementations. The formula for the max-cohesion case uses the additional assumption that everyone in the same community has the same cutoff for the school.

**Theorem 7.** In the continuum model, for any school \( s \), the cohesion from independent lottery implementation is:

\[
\mathcal{F}_{s,\text{indep}} = E_c[\bar{p}_s^2] = \bar{q}_s^2 + \text{Var}_c(\bar{p}_s).
\]
Figure 5-2: Assignment within community \( c \): horizontal dimension is students and vertical dimension is lottery numbers. \( z^*_1, z^*_2, z^*_3 \) and \( z^*_4 \) are lottery cutoffs for community \( c \) of schools 1, 2, 3, 4 respectively (well-defined because we assume everyone in the same community are in the same priority class). \( d_3(z'|c) \) is the demand function for school 3 at lottery \( z' \). \( p_3(c, y') \) is the probability student \( y' \) is assigned to school 3. The area shaded with the same color represents the 1-dimensional measure of assignment probabilities of students from this community to a school (\( \bar{p}_s(c) \)).

Under Assumption \[, \]

\[
f^s_{max} = E_{c,z}[d_s^2] = f^s_{independent} + E_c[Var_z(d_s|c)].
\]

Thus, the total cohesion,

\[
f_{independent} = \sum_s q_s^2 + \sum_s Var_c(\bar{p}_s),
\]

\[
f_{max} = f_{independent} + \sum_s E_c[Var_z(d_s|c)].
\]

We call \( f_{independent} \) the Baseline cohesion, and \( (f_{max} - f_{independent}) \) the potential gain in cohesion from correlated lottery.

The proof (see Online Appendix 5.C) follows from the cutoff structure, re-arranging orders of integration, and the law of total variance. We interpret the terms as follows.
• $\sum_s q_s^2$: Herfindahl index of school size. This is a measure of variation in school sizes. The more varied school sizes are, the higher the baseline cohesion.

• $\sum_s \text{Var}_c(\bar{p}_s)$: Between community variation in assignment probabilities. The more varied assignment probabilities are between communities, the higher the baseline cohesion.

• $\sum_s E_c[\text{Var}_s(d_s|c)]$: Potential gain in cohesion from using correlated lottery. For each school, the summand is the average across communities of the variation in demand function. In other words, it is how much, on average, the lottery number affects the mass of students from a community assigned to a school. Intuitively, this has to do with how competitive schools are (high cutoffs), and how correlated preferences within a community are. This intuition is made more precise in Corollary 2.

By manipulating the expressions in Theorem 7 we immediately yield 2 corollaries. The first shows that the maximum possible improvement ratio is upper-bounded by the number of schools, which can be achieved if all schools have size $\frac{1}{m}$ and all students share the same preferences. The second corollary interprets the gain in cohesion as a weighted average of an aggregate statistic representing competition and preference correlation.

**Corollary 1.** (Upper-bound on improvement ratio)

$$\frac{f_{\text{max}}}{f_{\text{independent}}} \leq \frac{1}{\sum_s q_s^2} \leq m.$$

**Corollary 2.** Under Assumption 7 the proportional improvement in cohesion for lottery correlation can be interpreted as a weighted measure of the Squared Coefficient of Variation (SCV) of the demand function. Define the Squared Coefficient of Variation $\text{SCV}_{d_s}(c) = \frac{\text{Var}_s(d_s|c)}{E_z[d_s|c]^2}$. Intuitively, the SCV of the demand function is a combination of competition (high cutoffs) and high within-community preference correlation.
Define weights \( w_c = E_z[d_s|c]^2 = \bar{p}_s(c)^2 \). For any school \( s \), the proportional improvement in cohesion is a weighted average of the SCV:

\[
\frac{f^s_{\text{max}} - f^s_{\text{independent}}}{f^s_{\text{independent}}} = \frac{E_c[w_cSCV_{ds}(c)]}{E_c[w_c]}.
\]

We now show that maximum cohesion in our large market setup can also be expressed as a constant minus a measure of within-community heterogeneity. More precisely, define \( \Delta_s(i, i') = |p_s(i) - p_s(i')| \). This is the absolute difference in assignment probabilities to school \( s \) for individuals \( i \) and \( i' \). Define the mean absolute difference in assignment probability for school \( s \) and community \( c \) to be

\[
\bar{\Delta}_s(c) = E_{i, i' \in c}[\Delta_s(i, i')].
\]

This is the expected value of \( \Delta_s(i, i') \) for two randomly drawn individuals \( i, i' \) from community \( c \). Averaging across communities, \( E_c[\bar{\Delta}_s(c)] \) is then an aggregate measure of within-community heterogeneity in assignment probabilities to school \( s \). The following proposition shows that maximum cohesion to a school \( s \) is equal to the size of the school minus one half of this measure of within-community heterogeneity.

**Proposition 3. (Cohesion is limited by within-community heterogeneity)** Under Assumption 7

\[
f^s_{\text{max}} = q_s - \frac{1}{2}E_c[\bar{\Delta}_s(c)].
\]

So that summing across all schools, \( f_{\text{max}} = 1 - \frac{1}{2} \sum_s E_c[\bar{\Delta}_s(c)] \), where the term being subtracted is an aggregate measure of within-community heterogeneity in assignment probabilities.

The expression for gain in cohesion in Corollary 2 is exact but difficult to think about. The SCV of demand function is hard to estimate. The following proposition bounds potential gain from lottery correlation by an easier-to-estimate measure of lottery uncertainty.
Proposition 4. (Decomposition of potential to improve) Under Assumption 1, the gain from cohesion from correlated lottery is equal to the average individual assignment variance minus a function of within-community differences in assignment probabilities.

\[ f_{\text{max}}^* - f_{\text{independent}}^* = E_s[p_s(1 - p_s)] - \frac{1}{2}E_{c(i) = c(i')}[\Delta_s(i, i')(1 - \Delta_s(i, i'))] \leq E_s[p_s(1 - p_s)]. \]

In the first line, the first term is the average across individuals of assignment variance to school \( s \); this is a measure of how uncertain the lottery is for school \( s \). The second term is minimized when \( \Delta_s(i, i') \) is either close to 1 or close to 0. In other words, for correlated lottery to be most effective, we desire that for two students within the same community, they either get assigned to a school with very similar probabilities, or one gets assigned with very high probability and the other with very low probability. Because the second term is always non-negative, gain from cohesion is upper-bounded by the uncertainty of the lottery.

The following proposition shows exactly when lottery-correlation is useful. It turns out that for every school, the following three quantities add up to a constant:

1. Between-community heterogeneity in assignment probability.
2. Within-community heterogeneity in assignment probability.
3. Potential to improve cohesion by lottery-correlation.

So that we are generally in one of the following 3 cases:

1. High between-community heterogeneity: cohesion with independent lottery is already high.
2. High within-community heterogeneity: there is nothing we can do.Maximum cohesion is severely limited.
3. High potential to improve.
Proposition 5. *(Structural identity)* For every school $s$, the following three terms add to a constant:

$$q_s(1-q_s) = \text{Var}(\bar{p}_s) + \frac{1}{2} E_c[\Delta_s(c)] + (f^*_s - f^\text{independent})$$

between-community within-community potential to improve heterogeneity heterogeneity

This relationship is illustrated in Figure 5-3.

Figure 5-3: Diagram illustrating structural identity (Proposition 5). The triangle represents 3 quantities that add up to a constant. There are 3 possible cases: A) high between-community variation in assignment probability, so baseline cohesion is already high. B) high within-community heterogeneity in assignment probability, so there is nothing we can do by Proposition 3. C) significant potential to improve cohesion by correlated lottery. Corollary 2 shows that potential to improve can be interpreted as baseline cohesion multiplied by a weighted average of the SCV (squared coefficient of variation) of the demand function (with respect to the lottery $z$). The SCV can be interpreted to correspond to a mixture of preference correlation and competition. Proposition 4 shows that the improvement is upper-bounded by a measure of total uncertainty of the lottery. This relationship holds not only overall but also school by school.

5.4.1 Embedding a model with preferences

To more directly illustrate the relationship between preference structure and cohesion, we consider an explicit model of students’ preferences. Suppose that student $i$’s preferences for schools is driven by random utility model

$$u_{is} = \alpha \nu_s - \beta \omega_{c(i)s} + \epsilon_{is},$$

194
where $\nu_s$ corresponds to “quality” of school $s$, $\omega_{cs}$ is distance from community $c$ to school $s$ (more generally can capture any community-specific propensities for specific schools), and $\epsilon_{is}$ is an idiosyncratic shock drawn from a standard Gumbel distribution. This is a realistic structure of preferences that has been used to fit data in empirical studies. (See Pathak and Shi (2013).) For simplicity, we assume that for different schools $s, s'$, $\nu_s \neq \nu_{s'}$ and $\omega_{cs} \neq \omega_{cs'}$. We assume that there are no priorities, so Assumption 1 trivially hold. We study the behavior of $f_{\text{independent}}$ and $f_{\text{max}}$ as functions of $\alpha$ and $\beta$.

The following proposition shows the limit behavior of $f_{\text{independent}}$ and $f_{\text{max}}$ as $\alpha$ (how much quality matters) or $\beta$ (how much distance matters) goes to infinity. It shows that if there is no between-community heterogeneity ($\beta = 0$), cohesion with independent lotteries is fixed and is the lowest possible, regardless of how correlated students preferences are. On the other hand, with correlated lottery, perfect cohesion can be achieved without between-community heterogeneity, if preferences for quality are very correlated. With high between-community heterogeneity, perfect cohesion can be approached in both independent and correlated cases. This corroborates the triangle diagram in Figure 5-3.

**Proposition 6.** Assume that capacities are such that when everyone goes to their closest school, the closest school can accommodate. For any $\alpha_0$,

$$f_{\text{independent}}(\alpha_0, 0) = \sum_s q_s^2,$$  

(5.3)

$$\lim_{\beta \to \infty} f_{\text{independent}}(\alpha_0, \beta) = 1.$$  

(5.4)

For the correlated case, regardless of capacities, for any $\alpha_0$ and $\beta_0$,

$$\lim_{\alpha \to \infty} f_{\text{max}}(\alpha, \beta_0) = 1,$$  

(5.5)

$$\lim_{\beta \to \infty} f_{\text{max}}(\alpha_0, \beta) = 1.$$  

(5.6)
While the above result is only for the limit, the following proposition shows comparative statics for the finite case, in the special case that $\beta = 0$. It shows that if capacities are not too different so that more desirable schools are also more over-demanded (more applicants per seat), then as preferences become more correlated, cohesion from independent lottery stays fixed, while cohesion from correlated lottery increases.

**Theorem 8.** (Pure vertical differentiation) Suppose that $\beta = 0$ (no between-community heterogeneity). Let $r_s = e^{v_s}$. Suppose that the schools are ordered so that the more desirable schools are first, so $r_1 > r_2 > r_3 \cdots$. Assume that capacities are not “too different,” so that dividing by capacities do not change this relative order, so $\frac{r_1}{q_1} \geq \frac{r_2}{q_2} \geq \frac{r_3}{q_3} \cdots$. Then for every school $s$, while $f_{\text{independent}}^s(\alpha, 0) = q_s^2$ is constant in $\alpha$, $f_{\text{max}}^s(\alpha, 0)$ is strictly increasing in $\alpha$.

### 5.5 Computing the correlated lottery implementation in practice

Fixing students’ submitted rankings, let $p$ be the assignment probabilities of the original mechanism. ($p_{is}$ is the probability student $i$ is assigned to school $s$ under the mechanism.) The max-cohesion correlated-lottery implementation problem can be written as

$$
\begin{align*}
\text{Max} & \quad E[f(x)] \\
\text{s.t.} & \quad E[x] = p \\
& \quad x \in X.
\end{align*}
$$

(5.7)

Define the assignment polytope

$$
\mathcal{P} = \left\{ a \in \mathbb{R}^{n \times m} : \sum_s a_{is} = 1, \sum_i a_{is} \leq q_s, 0 \leq a_{is} \leq 1 \right\}.
$$

Represent random assignment $x$ explicitly as $\{(\lambda^l, a^l) : a^l \in \text{Vertices}(\mathcal{P}), \lambda^l \in [0, 1], \sum_l \lambda^l = 1\}$. 

196
1}, so that \( x = a^t \) with probability \( \lambda^t \), and let \( A \) be the \( nm \) by \( |\text{Vertices}(\mathcal{P})| \) matrix in which the columns encode the vertices of assignment polytope \( \mathcal{P} \). Define \( f(A) \) naturally as the vector in which the \( l \)th component is the cohesion of the \( l \)th column of \( A \). We can rewrite the above in the more explicit form

\[
\begin{align*}
\text{Max} & \quad f(A) \cdot \lambda \\
\text{s.t.} & \quad A\lambda = p \\
& \quad \mathbf{1} \cdot \lambda = 1 \\
& \quad \lambda \geq 0.
\end{align*}
\]

(5.8)

which is a standard form linear program (albeit with exponentially many variables). The theory of LP implies that there is an optimal solution with only \( nm \) positive components of \( \lambda \) (because \( \text{rank} \left( \begin{bmatrix} A \\ \mathbf{1} \end{bmatrix} \right) \leq nm \)). In other words, for any assignment probabilities \( \{p_{is}\} \), there exists a random assignment \( x \) with \( E[x] = p \) which randomizes over at most \( nm \) deterministic assignments and achieves maximum cohesion.

This suggests the following mechanism.

1. Estimate the individual assignment probabilities \( p_{is} \) of the original mechanism by independently simulating \( T \) times. Note that the computed estimates are unbiased and have component wise standard deviation \( \sqrt{\frac{(1-p_{is})p_{is}}{T}} \leq \frac{1}{2\sqrt{T}} \).

2. Use the estimated \( p \) as inputs to program (5.8) and compute a convex combination of deterministic assignments \( \{(\lambda^t, a^t) : \sum_i a^t = 1\} \). Output assignment \( a^t \) with probability \( \lambda^t \).

For any \( T \), the resultant randomized mechanism induces the same individual assignment probabilities as the original mechanism (using crucially the unbiasedness of the simulation in first step).

The only difficulty is what algorithm to use to solve the intractably large program (5.8). As shown in Proposition 2, solving the cohesion optimization is NP-hard even with 2 schools. In Online Appendix 5.A, we show that the case with many schools
is related to the notoriously hard Quadratic Assignment Problem (QAP), which is NP-hard to approximate to any constant factor (Burkard et al. 1998; Sahni and Gonzalez 1976).

We propose a simple heuristic to solve this in practice. This heuristic is related to the Birkhoff-von Neumann theorem, as it seeks to express the original assignment probabilities as a convex combination of deterministic assignments by iteratively breaking off one deterministic assignment at a time. To find a deterministic assignment at each iteration, it solves a max-weighted bipartite perfect matching problem, with the students on one side of the graph to be matched to the schools on the other side. As input to the max-weighted matching procedure at each iteration, we define an edge from a student to a school if and only if the assignment probability of that student to the school is positive, after having subtracted off the deterministic assignments from previous iterations. The weight of this edge is randomly generated, but we constrain the weights to be the same for everyone from the same community to the same school. An intuition of why this might work is that conditional on student $i$ being assigned to school $s$ in the max-weighted perfect matching, the edge weight of that student to the school is probably high, and so the edge weight of other students from $i$’s community to $s$ is probably high, so we expect many of them to be co-assigned with $i$ to $s$. Another intuition is that giving the same edge weights to students of the same community reduces “local minima” in which a trading cycle can increase cohesion. Details of the heuristic and elaborated intuition is in Appendix 5.A. We implemented our correlated-lottery implementor in Java (code is available upon request). Our heuristic does not require Assumption 1 to hold. As shown in Section 5.6, this heuristic achieves good results even when students of the same community have different priorities. Assumption 1 was only needed to prove exact analytical results in the large market case.

5.5.1 An upper-bound on maximum cohesion

To evaluate the optimality gap of our heuristic, we derive a simple upper-bound to the correlated lottery implementation mathematical program (5.7). Consider student
conditional on the student being assigned to $s$, the expected number of same community peers that can be co-assigned to school $s$, $E[v_i(x)|x_{is}]$ is upper-bounded by

$$E[v_i(x)|x_{is}] \leq \min \left( (q_s - 1), \sum_{c(i')=c(i),i' \neq i} \min(1, \frac{p_{i's}}{p_{is}}) \right).$$

Where the first term follows from the capacity constraint of $s$, and the second term follows from $E[x_{i's}|x_{is}] = \frac{E[x_{i's}x_{i's}]}{p_{is}} \leq \frac{\min(p_{i's},p_{i's})}{p_{is}}$.

Summing over all students and taking the expectation over school $s$, we get that cohesion is upper-bounded as follows.

$$E[f(x)] = \frac{1}{n} \sum_{i,s} p_{is} E[v_i(x)|x_{is}] \quad (5.9)$$

$$\leq \frac{1}{n} \sum_{i,s} \min \left( p_{is}(q_s - 1), \sum_{c(i')=c(i),i' \neq i} \min(p_{is},p_{i's}) \right). \quad (5.10)$$

## 5.6 Application to Boston elementary school choice

### 5.6.1 Description of school choice in Boston

School Choice in Boston Public Schools (BPS) began with the adoption of the Controlled Choice Student Assignment Plan in 1988. The plan organized public elementary and middle schools into three zones–East, North, and West–and students were given the option to apply to any set of schools within their zone. To apply, students submitted ranked lists of their preferred schools, and a centralized lottery produced the assignment. Since then, policies regarding the assignment process have been revised numerous times, including the lottery algorithm itself, but the overall framework of the process remained the same.

Our empirical study focuses on BPS elementary school assignment in 2012, which is when Mayor Menino made his call to improve community cohesion in school assignment. We focus on elementary schools because this is arguably the time when
going to school with neighbors is most important, and because this was the focus of the mayor’s call for reform. The goal of this study is to analyze what would have happened if we had adopted a correlated lottery procedure to improve community cohesion in 2012, and to compare with alternative approaches to improve cohesion considered by the mayor-appointed city committee.

The vast majority of students enter BPS elementary schools via entry grades K1 and K2 (K for kindergarten). To enroll, students participate in one of 4 application rounds by submitting rankings over specific programs in schools (a school may offer several programs: regular, English Language Learner, Montessori, etc). They can rank as many schools as they would like. The first round occurs in January, and this is when the majority of families participate (about 80% of families who eventually apply first apply in Round 1.) For families that come later, there are 3 smaller subsequent rounds that take place from March to June. There is also a wait-list process in which families may get a seat at a subsequent round if more capacity becomes available or if others drop out. The wait-list favors applicants from earlier rounds so it is always the best to apply in Round 1. For simplicity, since the majority of seats are allocated in Round 1, we focus on Round 1 in this chapter.

After families submit choices, the assignment is computed by Deferred Acceptance with Single Tie-Breaker (DA-STB). This was adopted in 2005 to eliminate the need for strategic manipulation. See [Abdulkadiroğlu et al. (2006)]. More precisely, each program is internally divided into 2 halves, a walk-zone half and an open half. Students’ preferences on programs are augmented into preferences on halves, such that students living in the walk-zone (within one mile of school) prefers the walk-half and students from outside the walk-zone prefer the open half. (Preferences between different programs are maintained in the augmentation.)

Each of the program halves also rank students in the following way: each student is given an i.i.d. random lottery number. The ranking over students is induced by several levels of priorities (1st level is most important, 2nd is to break 1st level ties, 3rd level is to break 2nd level ties, etc). The priorities are given in Table 5.3.

Given students’ rankings on program halves and program halves’ rankings on stu-
Table 5.3: Order of priorities used in Boston elementary school assignment in 2012. The earlier level priorities are more important, with later levels only used to break ties. The 3rd level is applied only for walk-zone halves.

1. Find an unassigned student, have her apply to her top remaining choice.

2. If the program half is not full, accept her; otherwise, bump out the least preferred student from that program half (which may be her), and remove this program half from that student’s ranking.

3. Iterate until all unassigned students have gone through all their choices.

It is well-known that the above algorithm induces a unique assignment regardless of the order of application. (See Roth and Sotomayor (1990).) This induces an assignment of students to school programs, and this assignment is mailed to families. It is well-known that if families can submit as many choices as they would like and if their submissions do not influence the priorities, then this assignment process is strategyproof for all students. (See Abdulkadiroğlu and Sönmez (2003b).)

5.6.2 Data

We use 2012 Round 1 choice data for grades K1-2, which have been anonymized but still contain information on students’ demographics, geocode (division of Boston into 868 smaller regions), top 10 choices, and final assignment. Although we were not given capacities of programs, we were able to infer them from final assignments. As a check, we were able to replicate 98.2% of K1 assignment and 99.0% of K2 assignment (excluding students who were administratively assigned by BPS after the
Table 5.4: Summary statistics of the choice data.

<table>
<thead>
<tr>
<th></th>
<th>K1</th>
<th>K2</th>
<th>K1-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schools</td>
<td>66</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Programs</td>
<td>106</td>
<td>123</td>
<td>229</td>
</tr>
<tr>
<td>Seats</td>
<td>1921</td>
<td>3689</td>
<td>5610</td>
</tr>
<tr>
<td>Continuing students</td>
<td>167</td>
<td>1904</td>
<td>2071</td>
</tr>
<tr>
<td>(6%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-continuing siblings</td>
<td>690</td>
<td>467</td>
<td>1158</td>
</tr>
<tr>
<td>(26%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New families</td>
<td>1809</td>
<td>1659</td>
<td>3469</td>
</tr>
<tr>
<td>(68%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Students</td>
<td>2666</td>
<td>4030</td>
<td>6696</td>
</tr>
</tbody>
</table>

Table 5.5: Percentage of students of each type getting their top choice. For example, for K2, 47% of students are continuing from K1, and 95% of them get their top choice.

assignment algorithm described before has finished). Table 5.4 summarizes the supply and demand data.

In the data it turns out that the lottery mostly matters only for new families (non-continuing, non-siblings). Table 5.5 tabulates the fraction of students of each type getting their top choice.

Since continuing students and siblings are very likely to be assigned their first choice (so there is essentially no lottery for most of them), we can only hope to significantly impact via lottery correlation those who are new families. So in reporting outcomes, we focus on the new families.

Our approach also takes as input delineations of community. The city may want to do this based on natural dividing lines or other considerations. For the purpose of this study, we simply use a square grid of .5 miles in length, with each .5 mile × .5 mile square defining a community. Figure 5-4 plots the geographic distribution of all 6696 K1-2 students, partitioned into 205 non-empty communities.
Figure 5-4: Partition of Boston into .5 mile × .5 mile squares. We treat each square as a community. Each circle corresponds to a geocode, and the area of the circle is proportional to the number of students residing at the geocode. Defined in this way, the average number of students per community is $2666/205 = 13$ for K1 and $4030/205 = 19$ for K2.
<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Improved</th>
<th>Upperbound</th>
<th>Improvement</th>
<th>Bound on Impr.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All students</td>
<td>1.35</td>
<td>2.11</td>
<td>2.70</td>
<td>0.75</td>
<td>1.34</td>
</tr>
<tr>
<td>Continuing students</td>
<td>1.30</td>
<td>1.32</td>
<td>1.38</td>
<td>0.02</td>
<td>0.08</td>
</tr>
<tr>
<td>Non-continuing siblings</td>
<td>1.35</td>
<td>1.43</td>
<td>1.56</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>New families</td>
<td>1.36</td>
<td>2.44</td>
<td>3.26</td>
<td>1.08</td>
<td>1.90</td>
</tr>
<tr>
<td><strong>K2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All students</td>
<td>2.48</td>
<td>2.89</td>
<td>3.39</td>
<td>0.42</td>
<td>0.91</td>
</tr>
<tr>
<td>Continuing students</td>
<td>2.26</td>
<td>2.27</td>
<td>2.30</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Non-continuing siblings</td>
<td>2.91</td>
<td>3.01</td>
<td>3.23</td>
<td>0.10</td>
<td>0.32</td>
</tr>
<tr>
<td>New families</td>
<td>2.61</td>
<td>3.58</td>
<td>4.69</td>
<td>0.97</td>
<td>2.08</td>
</tr>
</tbody>
</table>

Table 5.6: Impact of lottery correlation on cohesion for various groups of students. For example, for K1 new families, conditional on being assigned, students on average can expect to find 1.36 same-community peers. If we use correlated-lottery, the number of same community peers improves by 79% to 2.44. Based on the given assignment probabilities, no correlation procedure can produce cohesion greater than 3.26. Using correlated lottery, the average K1 new family gains 1.08 additional same-community peers. No correlation procedure can produce a gain greater than 1.90.

### 5.6.3 Impact of correlated lottery implementation

We take the actual choices, simulate the current assignment algorithm 1000 times with independently drawn lottery numbers to estimate the assignment probabilities, and run the correlated lottery heuristic described in the Online Appendix 5.A.I to produce another 1000 assignments with the same assignment counts for each student-program pair, but correlated so within the same assignment students from the same community are more likely to be co-assigned to the same school. Drawing one of the 1000 correlated assignments is then a correlated lottery that replicates the estimated assignment probabilities for all students, but has improved cohesion. In Table 5.6 we tabulate for various groups of students their average baseline cohesion (without correlation), improved cohesion (with correlated lottery), upper-bound to cohesion (from Section 5.5.1), amount of improvement (improved minus baseline), and upper-bound on improvement (upper-bound minus baseline).

As can be seen, while lottery correlation does little for continuing students and siblings (who mostly get their 1st choice regardless of the lottery), it increases average cohesion for new families by about 1 in both K1 or K2. This is one additional “neighbor” these students can find at their school assignment, which is a significant
Table 5.7: Expected number of same-community peers co-assigned conditional on being assigned outside of walk-zone. This is the number of same-community neighbors a student who is traveling outside of his/her 1-mile walk-zone can expect to find at his/her assigned program. So for a K1 new family, conditioning on the student going outside of walk-zone for school, he/she on average has only 0.53 neighborhood peers in the current lottery. But with correlated lottery this more than doubles to 1.26. Without changing assignment probabilities, we cannot expect this to be larger than 1.73.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Improved</th>
<th>Upperbound</th>
<th>Improvement</th>
<th>Bound on Improv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.53</td>
<td>1.11</td>
<td>1.48</td>
<td>0.58</td>
<td>0.95</td>
</tr>
<tr>
<td>Continuing students</td>
<td>0.55</td>
<td>0.58</td>
<td>0.60</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Non-continuing siblings</td>
<td>0.54</td>
<td>0.64</td>
<td>0.73</td>
<td>0.10</td>
<td>0.19</td>
</tr>
<tr>
<td>New families</td>
<td>0.53</td>
<td>1.26</td>
<td>1.73</td>
<td>0.74</td>
<td>1.20</td>
</tr>
<tr>
<td>K2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.98</td>
<td>1.34</td>
<td>1.71</td>
<td>0.36</td>
<td>0.73</td>
</tr>
<tr>
<td>Continuing students</td>
<td>0.93</td>
<td>0.95</td>
<td>0.99</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Non-continuing siblings</td>
<td>1.01</td>
<td>1.12</td>
<td>1.31</td>
<td>0.12</td>
<td>0.31</td>
</tr>
<tr>
<td>New families</td>
<td>1.00</td>
<td>1.64</td>
<td>2.27</td>
<td>0.64</td>
<td>1.27</td>
</tr>
</tbody>
</table>

increase since the baselines are 1.36 and 2.61 respectively. Moreover, even if we had solved the max-cohesion problem to optimality, we cannot expect the improvement in cohesion to be more than 1.9 for K1 and 2.08 for K2, so to achieve greater gains we would need to alter the assignment probabilities.

One motivation for increasing cohesion is so that families can car-pool and students can have neighborhood friends, but to some extent this matters only when the student is going to school not in his/her neighborhood (otherwise he/she would not need to go by car and would have neighborhood friends regardless). In Table 5.7, we examine the expected # of same-community peers conditional on being assigned outside of own walk-zone. The upper-bound uses a similar formula as the one in Section 5.5.1 except that it conditions differently. As seen, the proportional impact of correlated lottery is more pronounced in this case, more than doubling baseline cohesion for K1 new families and achieving a 64% gain over baseline cohesion for K2 new families.
5.6.4 Comparison and interaction with other reforms

During the 2012-2013 school choice reform, two types of reforms proposed to the city committee were increasing the walk-zone percentage and reducing the choice menu (the set of schools students from various neighborhoods could rank from). Both strategies were intended to affect assignment probabilities to result in closer to home assignment. By theorem 7, this would increase cohesion as it would increase between-community heterogeneity. We empirically estimate the increase in cohesion due to these potential reforms and compare to correlated lottery (which does not affect anyone’s assignment probabilities.) We also evaluate the interaction of these reforms with lottery correlation, to see how much we can improve by applying both at the same time.

Increasing the walk-zone percentage

As described in Section 5.6.1, the BPS assignment algorithm in 2012 had 50% of seats of a program allocated to the walk-half (the side that respected walk-zone priority) and the rest to the open half. One approach to increase community cohesion is to induce closer-to-home assignment by increasing the percentage of seats allocated to the walk-half. For K1 and K2 new families, we show in Table 5.8 the result of this on their baseline cohesion, on their improved cohesion, as well as on their cohesion conditional on traveling outside of walk-zone (1 mile radius).

As can be seen, while increasing the walk-zone percentage increases cohesion, the maximum magnitude of increase (to 1.91 for K1 and 3.32 for K2) is less than if we kept the same walk-zone percentage but switched to correlated lottery (to 2.44 for K1 and 3.58 for K2). For the students who are traveling outside of their walk-zone, Table 5.8 shows that altering walk-zone percentage has almost no effect on their expected # of same-community peers, while lottery correlation significantly increases it. From the perspective of increasing community cohesion, both in terms of overall increase and in terms of helping those who need it the most, lottery correlation is more effective than increasing the walk-zone percentage.
Walk-zone %s  Baseline  Improved  Improvement  Baseline  Improved  Improvement

K1
50  1.36  2.44  1.08  0.53  1.26  0.74
60  1.44  2.56  1.12  0.48  1.13  0.65
70  1.56  2.72  1.16  0.46  1.13  0.67
80  1.69  2.86  1.18  0.44  1.07  0.64
90  1.82  3.01  1.19  0.46  0.95  0.50
100 1.91  3.11  1.20  0.54  0.86  0.33

K2
50  2.61  3.58  0.97  1.00  1.64  0.64
60  2.74  3.73  0.99  0.95  1.59  0.65
70  2.90  3.87  0.97  0.91  1.53  0.62
80  3.09  4.05  0.96  1.00  1.62  0.61
90  3.21  4.14  0.93  1.05  1.72  0.66
100 3.32  4.21  0.89  1.08  1.70  0.62

Table 5.8: Baseline and Improved Cohesion with differing walk-zone percentages. The first three columns correspond to students’ expected # of same-community peers co-assigned conditional on being assigned. The last three columns correspond to the same number conditional on being assigned outside of walk-zone.

Reducing the choice menus

Another approach to increase cohesion is to decrease the choice menu, so as to focus choices from the same community to similar schools. To evaluate such a reform, we need a model for how students would choose given a new menu. We use the same demand model as in the study commissioned by the city committee during the 2012-2013 school choice reform to evaluate a range of potential outcomes. This is a multinomial logit model fitted using the same data as in this study, and includes a fixed effect for each school, and linear controls for distance to choices, racial/socio-economic interactions and school-specific affinities (whether a student is continuing student, has sibling at a school, or lives in the walk-zone of a school). The demand model is documented in detail in [Pathak and Shi (2013)](#).

During the 2012 school choice reform, many different plans for choice menus were proposed. Some involved subdividing the city into smaller zones (such as the 6-zone, 9-zone, 11-zone, or 23-zone plans) and some involved giving the closest schools of certain types to each family (such as the Home Based A plan). For the purpose of this study we do not go into the details of each plan. These choice menus are
Table 5.9: Cohesion with differing choice menus. The first three columns correspond to students’ expected # of same-community peers co-assigned conditional on being assigned. The last three correspond to the expected # of same-community peers conditional on being assigned outside of walk-zone. The rows are sorted in increasing baseline cohesion.

documented in Pathak and Shi (2013), BPS (2013a), and Shi (2013). Table 5.9 shows the impact of various menus on baseline and correlated cohesion, as well as the impact on cohesion conditional on traveling outside of walk-zone. The statistics are averages of 25 draws of simulated choice data. (For each draw, the different plans share the same underlying random utility shocks so as to minimize sampling variance in the comparison. This is the simulation approach used in Pathak and Shi (2013).)

As can be seen in Table 5.9, for K1, correlated lottery beats the cohesion improvement from any menu change. For K2, correlated lottery accomplishes the same effect as tripling the number of zones to 9. For either K1 or K2, for students who are traveling outside of their walk-zone, correlated lottery alone delivers greater increase in cohesion than any choice menu change.

What is most interesting is that the impact of menu change and correlated lottery magnify one another. As can be seen, as the first column (baseline cohesion) increases, the third column (improvement due to correlated lottery) also tends to increase. For example, for K2, while changing to Home Based A (the plan adopted by the city

\[\text{Table 5.9: Cohesion with differing choice menus. The first three columns correspond to students’ expected # of same-community peers co-assigned conditional on being assigned. The last three correspond to the expected # of same-community peers conditional on being assigned outside of walk-zone. The rows are sorted in increasing baseline cohesion.} \]

<table>
<thead>
<tr>
<th>Choice Menu</th>
<th>Baseline</th>
<th>Improved</th>
<th>Improvement</th>
<th>Baseline</th>
<th>Improved</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1 3-Zone</td>
<td>1.11</td>
<td>2.09</td>
<td>0.98</td>
<td>0.42</td>
<td>1.16</td>
<td>0.74</td>
</tr>
<tr>
<td>K1 6-Zone</td>
<td>1.32</td>
<td>2.64</td>
<td>1.32</td>
<td>0.53</td>
<td>1.81</td>
<td>1.28</td>
</tr>
<tr>
<td>K1 9-Zone</td>
<td>1.52</td>
<td>3.16</td>
<td>1.64</td>
<td>0.69</td>
<td>2.42</td>
<td>1.73</td>
</tr>
<tr>
<td>Home Based A</td>
<td>1.62</td>
<td>3.34</td>
<td>1.73</td>
<td>0.66</td>
<td>2.48</td>
<td>1.82</td>
</tr>
<tr>
<td>11-Zone</td>
<td>1.66</td>
<td>3.37</td>
<td>1.72</td>
<td>0.79</td>
<td>2.74</td>
<td>1.95</td>
</tr>
<tr>
<td>23-Zone</td>
<td>1.93</td>
<td>3.85</td>
<td>1.92</td>
<td>0.59</td>
<td>1.76</td>
<td>1.17</td>
</tr>
</tbody>
</table>

\[\text{Table 5.9: Cohesion with differing choice menus. The first three columns correspond to students’ expected # of same-community peers co-assigned conditional on being assigned. The last three correspond to the expected # of same-community peers conditional on being assigned outside of walk-zone. The rows are sorted in increasing baseline cohesion.} \]

<table>
<thead>
<tr>
<th>Choice Menu</th>
<th>Baseline</th>
<th>Improved</th>
<th>Improvement</th>
<th>Baseline</th>
<th>Improved</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2 3-Zone</td>
<td>2.10</td>
<td>2.91</td>
<td>0.81</td>
<td>0.74</td>
<td>1.38</td>
<td>0.63</td>
</tr>
<tr>
<td>K2 6-Zone</td>
<td>2.62</td>
<td>3.97</td>
<td>1.35</td>
<td>0.92</td>
<td>2.19</td>
<td>1.27</td>
</tr>
<tr>
<td>Home Based A</td>
<td>2.84</td>
<td>4.49</td>
<td>1.65</td>
<td>1.10</td>
<td>2.93</td>
<td>1.83</td>
</tr>
<tr>
<td>9-Zone</td>
<td>2.91</td>
<td>4.47</td>
<td>1.56</td>
<td>1.01</td>
<td>2.56</td>
<td>1.55</td>
</tr>
<tr>
<td>11-Zone</td>
<td>3.18</td>
<td>4.84</td>
<td>1.66</td>
<td>1.08</td>
<td>2.63</td>
<td>1.55</td>
</tr>
<tr>
<td>23-Zone</td>
<td>3.55</td>
<td>5.37</td>
<td>1.82</td>
<td>0.91</td>
<td>2.27</td>
<td>1.36</td>
</tr>
</tbody>
</table>

\[\textbf{Table 5.9: Cohesion with differing choice menus. The first three columns correspond to students’ expected # of same-community peers co-assigned conditional on being assigned. The last three correspond to the expected # of same-community peers conditional on being assigned outside of walk-zone. The rows are sorted in increasing baseline cohesion.} \]

4In Pathak and Shi (2013) and Shi (2013), Home Based A is called “ClosestTypes2463.”
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Improvement in cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>-0.741***</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>6.183***</td>
</tr>
<tr>
<td></td>
<td>(1.609)</td>
</tr>
<tr>
<td>Pref. Correlation</td>
<td>2.503***</td>
</tr>
<tr>
<td></td>
<td>(0.354)</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.307***</td>
</tr>
<tr>
<td></td>
<td>(1.055)</td>
</tr>
<tr>
<td>Observations</td>
<td>18</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.931</td>
</tr>
</tbody>
</table>

Table 5.10: Descriptive regression of improvement in cohesion for new families in various choice menus. This seeks to explain variation in the 18 rows of data in Table 5.9 in terms of grade, uncertainty of lottery, and preference correlation. “Grade” is dummy for K2. “Uncertainty” is the average individual assignment variance, $E_i(\sum_s p_{is}(1 - p_{is}))$. “Preference Correlation” is the average number of common programs in the top-3 lists of two randomly chosen students from the same neighborhood. Robust standard errors are shown in parenthesis. Consistent with Proposition 4, the magnitude of improvement is increasing in uncertainty and preference correlation.

committee) increases cohesion by $2.84 - 2.10 = 0.74$, and correlated lottery alone increases cohesion by $2.91 - 2.10 = 0.81$, doing both at the same time more than doubles cohesion, increasing it by $4.49 - 2.10 = 2.39 > 0.74 + 0.81$.

This phenomenon can be understood in terms of proposition 2. In the large market, the gain in cohesion is the product of the baseline cohesion and a weighted average of SCV (Squared Coefficient of Variation) of community demand function. The SCV is positively related with within-community preference correlation. When we focus the preferences by reducing the choice menus, both the baseline cohesion and the SCV increase, thus yielding magnified gains. By Proposition 4, this should be additively decomposable into a term for individual lottery variance and a term related to within-community heterogeneity. To check this intuition, we run a regression of the improvement in cohesion from Table 5.9 on a dummy for grade, a measure of lottery uncertainty, and a measure of within-community preference correlation. The results are in Table 5.10. As seen, all of the terms are individually significant, and the R-squared is very high (93.1%), supporting the intuition from the theory.
5.7 Discussion

In this chapter, we examine the potential of correlating the school choice lottery to increase community cohesion without affecting individual assignment probabilities. This is desirable as altering individual assignment probabilities inevitably raises questions of equity and access, which are hard to settle without a difficult political process. In contrast, a method that provides cohesion gains without changing assignment probabilities may face less opposition.

In our analysis, we show that while maximizing cohesion is an NP-hard, messy optimization problem in a discrete setting, studying the large-market limit “smooths away” the hardness and yields very clean analysis. Under such a setting, all “reasonable” one-sided many-to-one matching mechanisms can be described as lottery-plus-cutoff. This allows us to decouple the allocation from the lottery correlation, and obtain fairly general results for the potential to improve cohesion by lottery correlation. We can also build a random utility model on top and prove comparative statics, which would be unthinkably difficult if we had remained in the discrete setting.

An empirical finding is that while lottery correlation and reducing choice menus both improve community cohesion, they work best together. For Boston, in the main entry grade K2, correlated lottery alone can improve cohesion for new families by 39%, and choice menu reduction alone can improve it by 30%. But doing both at the same time more than doubles cohesion.

A question for future work is whether lottery correlation can be used to achieve other desirable social outcome. A first attempt might be to increase racial or socio-economic diversity at schools. Intuitively, one may try to “negatively correlate” race or socio-economic status in the lottery. Unfortunately, we show in the Electronic Companion that in the large market, using a correlated lottery produces no gain in diversity, because independent lotteries already achieves the optimum. Significant additional gain in diversity will require altering assignment probabilities.

From an implementation perspective, a potential criticism of our approach is that the lottery becomes less intuitive and transparent. Drawing random lottery numbers
and using it to determine priorities is very intuitive, and in the current Boston system, a family may ask the school board for their lottery number. One may imagine a family dissatisfied about their assigned choice may theoretically “audit” the system by checking everyone’s lottery number. However, in our approach, there is no longer a simple mapping between lottery numbers and assignment outcomes. Nevertheless, some of the intuitiveness and transparency may be restored if the school board explains the mechanism as a lottery on joint assignments instead of on individual priorities: the school board may print out a large table in which each row denotes a student, and each column denotes a joint assignment. The entry in each cell is the student’s assigned school in an assignment. The row counts of each student being assigned to each school would be consistent with the simulated assignment probabilities under the original mechanism. The lottery is then to draw a random column from this table, and a family dissatisfied about the outcome may theoretically “audit” the system by checking this table.

While in this chapter we assumed that increasing cohesion is desirable from a policy perspective, this claim warrants further investigation. Although families may in general prefer having more neighboring kids going to the same school so that the children can travel together, share homework help, car-pool on certain occasions, or play together after school, how much they value this and how much this contributes to the students’ development are yet to be determined. Moreover, it may be that not everyone wants to be with their neighbors, and for families living in high crime, high-poverty neighborhoods, it is conceivable that some may want the opposite. To accommodate this, one may want to use an opt-in or opt-out system in practice.

Another potential criticism of our approach is that it is unclear how one ought to define “community.” In the empirical study, we used a .5 mile by .5 mile grid, and a natural extension would be to optimize the grid size. But doing this optimization requires more in-depth understanding of what community means: too find a grid and the community definition may be too restrictive, too course and we may no longer be capturing “neighborhoods.” More broadly, it is unclear if any geographic delineation can adequately capture the notion of community, since this may be more determined...
by family background, work proximity, and other factors than home proximity. One way to micro-found our setting is to assume students preferences are lexicographic in the sense that each student first cares to which school she is assigned to, and for each assignment, she prefers to be co-assigned with more students from her own community. To relax these issues one potential is to allow families to define communities for themselves, or even let students report with which peers they would prefer to be assigned with. But this makes preferences for peers private knowledge and raises difficult incentive problems due to the complementarities such preferences introduce. For example, in the closely related matching markets, when couples exist in the market (who have joint preferences) a “stable” assignment may not exist [Klaus and Klijn (2005)]. One can similarly show that an “envy-free” allocation (with respect to the priorities) in the school choice problem need not exist even when just two students wish to be co-assigned. One direction is to limit the complementarities in the market. Ashlagi et al. (2010) show that as long as the number of couples in the market does not grow too fast, a stable matching exist. Students however may have asymmetric preferences over who they wish to be co-assigned with. Dur and Wiseman (2014) adopts a mechanism design approach that allows students to submit preferences over peers but relaxing the stability notion.\footnote{See also Lavy (2012) for empirical findings from a natural experiment in Tel Aviv which allowed students to report also preferences over peers and not just students.}

5.A Solving the Correlated-Lottery Implementation Problem

The natural approach to tackling (5.8) in Section 5.5 is column generation. Using the language of the simplex algorithm for LP, suppose we are currently at basis \( B \), let \( \beta = A_B^{-1} f(A_B) \), the column generation sub-problem of finding the deterministic assignment with the highest rate of improvement when we pivot to it (in the sense of
the Simplex algorithm) is

$$\text{Max } f(a) - \beta \cdot a \quad (5.11)$$

s.t. $a \in \mathcal{P}$.

Note that in the above we’ve relaxed constraint $a \in \text{Vertices}(\mathcal{P})$ to $a \in \mathcal{P}$. We can do this because the objective is convex. In fact it is quadratic as cohesion can be written as

$$f(a) = \frac{1}{n} \left( -n + \sum_{s,c} \left( \sum_{i \in I_c} a_{is} \right)^2 \right).$$

Either the optimum objective of (5.11) is positive in which case we pivot to the corresponding column $a^l$ representing the optimal solution found, or the optimum is non-positive and we yield a certificate of optimality for (5.8).

However, this subproblem (5.11) seeks to maximize a quadratic over the assignment polytope, which in its general case is exactly the notoriously difficult quadratic assignment problem (QAP) (see Burkard et al. (1998)), which is NP-hard to approximate to any constant factor (Sahni and Gonzalez (1976)). One hope is that since our quadratic is of a simpler form, we may still find a polynomial time algorithm. Unfortunately, one can show that the decision problem for both the original (5.8) and subproblem (5.11) are NP-complete in the strong sense, even when community sizes are constrained to be 2 or 3. The proof can be done by reduction from Not-All-Equal-3-SAT.

5.A.1 A practical heuristic

We present an efficient heuristic for correlated-lottery implementation.

The idea of this heuristic is similar to in the Birkhoff-von Neumann theorem. We express the original assignment probabilities as a convex combination of deterministic assignments, by iteratively breaking off deterministic assignments with high cohesion. In each iteration, we find a deterministic assignment with high cohesion that limits to
assigning students to schools for which they have positive assignment probability. This ensures that we can break off a positive multiple of this deterministic assignment. To find such a deterministic assignment, we solve a max-weight perfect matching problem on a bipartite graph, in which one side of the graph are students and the other is schools. We define an edge in this bipartite graph between a student and a school if and only if the student still has positive assignment probability to that school, after subtracting off deterministic assignments found in earlier iterations. The weight of each edge is random, but we constrain the weights to be equal for everyone from the same community to the same school. The intuition of why this may work is that conditional on student \( i \) being assigned to school \( s \) in the max-weighted matching, the weight between \( i \) and \( s \) is probably high, so the weight of everyone else from same community as \( i \) to \( s \) is probably high, so many of them are probably assigned to school \( s \).

To solve the max-weight bipartite perfect matching problem in each iteration, we use a specialized primal-dual implementation with worst case running time \( O(n^2m) \), but which is closer to \( O(nm^2) \) in practice. (This helps because in our case \( n \gg m \).) The total number of iterations is at most \( \min(mn, T) \), because each iteration reduces the number of non-zero assignment probabilities by at least 1. The number of iterations is also upper-bounded by \( T \) because if all assignment probabilities are multiples of \( \frac{1}{T} \), then the amount subtracted each time is at least \( \frac{1}{T} \). The total running time guarantee is \( O(n^2m \min(nm, T)) \).

A precise pseudo-code of the algorithm is given below.

**Explanation for the inner step**

Another intuition behind our method of finding a deterministic assignment in each iteration with high cohesion is as follows. Let \( e_{is} \) be the indicator for whether the assignment probability of \( i \) to \( s \) is positive. We want to find deterministic assignment \( a \leq e \) each time that maximizes cohesion \( f(a) \), but since this is NP-hard, we settle for “not-too-bad” solutions. It turns out that our method of generating community-specific random weights and solving max-weight assignment always avoids a kind of
Algorithm 2 Heuristic for program 5.8

Require: \( p \in \mathcal{P} \) (assignment probabilities to implement)

\[
x \leftarrow \emptyset
\]

while \( p \neq \vec{0} \) do

\[
\begin{align*}
  u_{cs} & \leftarrow \text{random weight, } \forall c \in C, s \in S \\
  w_{is} & \leftarrow u_{c(i)s}, \forall i \in I, s \in S \\
  e_{is} & = \begin{cases} 
    1 & \text{if } p_{is} > 0 \\
    0 & \text{otherwise}
  \end{cases}, \forall i \in I, s \in S \\
  a & \leftarrow \arg \max_{a'} \{w \cdot a' : a' \in \mathcal{P}, a' \leq e\} \\
  \lambda & \leftarrow \max \{\lambda' : \frac{p - \lambda a}{1 - \lambda} \in \mathcal{P}\} \\
  x & \leftarrow x \cup (\lambda, a) \\
  p & \leftarrow p - \lambda a
\end{align*}
\]

end while

return random assignment \( x = \{(\lambda_j, a_j)\} \)

“locally-suboptimal” solutions, in which cohesion can be improved by trading cycles.

To describe our notion of local sub-optimality, we first define trading cycles: Given an assignment \( a \), we say that there is a trading cycle (between communities and schools) \( c_0 \rightarrow s_0 \rightarrow c_1 \rightarrow s_1 \cdots \rightarrow s_{l-1} \rightarrow c_0 \) if for each community \( c_j \) there is some student \( i_j \in c_j \) such that \( a_{i_j s_{j-1}} = 1 \) and \( e_{i_j s_j} = 1 \) (where arithmetic on \( j \) is modulo \( l \)). In other words, by undoing the assignments \( i_j \rightarrow s_{j-1} \) and re-assigning \( i_j \rightarrow s_j \), we arrive at another feasible assignment \( a' \). We say that this trading cycle leads \( a \) to \( a' \).

Our notion of local sub-optimality occurs in assignment \( a \) when both trading cycles \( c_0 \rightarrow s_0 \rightarrow c_1 \rightarrow s_1 \cdots \rightarrow s_{l-1} \rightarrow c_0 \) and the reverse \( c_0 \leftarrow s_0 \leftarrow c_1 \leftarrow s_1 \cdots \leftarrow s_{l-1} \leftarrow c_0 \) exist in \( a \). Suppose the first leads to \( a' \) and the second to \( a'' \), then we have that the demographic counts, encoding for each community \( c \) and school \( s \) the number of assignments from \( c \) to \( s \), satisfies \( d^a = \frac{1}{2}(d^{a'} + d^{a''}) \). However, cohesion \( f(\cdot) \) can be written as a strictly convex function on \( d^a \), so by convexity \( f(a) < \frac{1}{2}(f(a') + f(a'')) \) where both \( a' \) and \( a'' \) are feasible assignments \( \leq e \). So \( a \) is suboptimal.

However, this situation cannot occur by our method because for any cycle \( c_0 \rightarrow s_0 \rightarrow c_1 \rightarrow s_1 \cdots \rightarrow s_{l-1} \rightarrow c_0 \) and any \( \{i_j \in c_j\} \), either \( \sum_j w_{i_j s_j} - \sum_j w_{i_j s_{j-1}} > 0 \) for all such \( \{i_j \in c_j\} \) or the reverse for all (since weights from the same community to the same school is defined to be the same, and the difference is non-zero with probability...
1 as weights are randomly generated); since a max-weight matching pushes as much as possible along positive weight trading cycles and reverses as much as possible any negative weight trading cycle, the above kind of local non-optimality can never occur in our inner step.

5.B  Additional empirical results

5.B.1  Distribution of gains

We examine how the gains in cohesion from correlated lottery are distributed among various neighborhoods. Figure 5-5 shows baseline cohesion and improvement (correlated baseline) for various geocodes for K1. Figure 5-6 does the same for K2. The plots show that the gains from correlated lottery for K1 is reasonably evenly distributed geographically, with some amounts of “green” throughout the city. On the other hand, the gains for K2 is very uneven, being concentrated in north east corner, south west corner, and a small wedge on the eastern part of the city. This is reflected also in the box plots, as the median improvement for K1 is about 0.6, while the median improvement for K2 is only 0.25. (The mean improvement is roughly 1 in both cases.) This is a disappointing finding as K2 is where the majority of students enter BPS.

How do we make sense of the uneven distribution of improvement in cohesion? The theory in Section 5.3 suggests that cohesion improvement should benefit those whose lottery is most uncertain (high variance) and whose communities have lowest within-community heterogeneity. Moreover, we expect that within a community, lottery correlation should help the most those whose preferences are most similar to the rest of the community. For ease of interpretation, we define the following proxy: for student $i$, comparing her with a randomly drawn peer from her community, what is the expected # of programs they have in common among their top 3 choices? We label this proxy $\text{choiceAgreement}(i)$.

We plot this for K1 and K2 new families in Figure 5-7. As shown, this metric of
Figure 5-5: Geographic distribution of baseline cohesion and cohesion improvement for K1 new families. Each circle corresponds to a geocode. Its color corresponds to the average value of students living at that location, and its size is proportional to the number of students. The bar on the right of each map shows a box-plot based on the same color scale of the student level distribution, with the ends of the box showing the 25 and 75 percentiles, and the red line shows the median (parts of box plot maybe outside of range to have colors show reasonable contrast). From the left plot, we see that cohesion is highest in East Boston (upper-right island), and varies in different pockets in the city. Improvement in cohesion from correlated-lottery is unevenly distributed, being highest in north east, south west, and various pockets of the city.

within-community preference correlation is about twice higher for K1 than for K2, and the geographic distribution seems to match the distribution of cohesion gains in Figures 5-5 and 5-6. To illustrate this relationship more precisely, we regress individual-level improvement in cohesion on 4 variables: individual lottery variance ($\sum_{s} p_{is}(1 - p_{is})$), scaled choice agreement ($|c(i)|choiceAgreement(i)$), and dummies for whether student $i$ is continuing student or non-continuing sibling. The scaling in the choice agreement is to control for community size. The regression results are tabulated in Table 5.11. As can be seen, much of the variation in gains from cohesion can be explained by uncertainty of lottery and preference correlation and incoming status, with highest gains for new families who face uncertain lottery and whose
preferences are highly correlated with their community.

5.B.2 Equitable distribution when applying both menu change and correlated lottery

While the cohesion gain from correlated lottery alone is not equitably distributed across the city, it turns out that doing correlated lottery along with a reduced choice menu remedies this inequity. Recall that the choice menu reduction chosen by the city in the 2012-2013 school assignment reform was Home Based A. By applying correlated lottery to Home Based A, we can yield cohesion improvements across all neighborhoods, with 75% of students experiencing a cohesion gain of about 1 or more. This is shown in Figure 5-8 which plots the geographic distribution of baseline cohesion, improvement from correlation alone, improvement from choice menu reduction alone, and improvement from applying both at the same time.
<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty</td>
<td>0.396***</td>
<td>0.381**</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.158)</td>
</tr>
<tr>
<td>Pref. Agreement</td>
<td>0.0386***</td>
<td>0.0434***</td>
</tr>
<tr>
<td></td>
<td>(0.00390)</td>
<td>(0.00493)</td>
</tr>
<tr>
<td>Non-continuing sibling</td>
<td>-0.823***</td>
<td>-0.726***</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.188)</td>
</tr>
<tr>
<td>Continuing</td>
<td>-0.394***</td>
<td>-0.161</td>
</tr>
<tr>
<td></td>
<td>(0.0958)</td>
<td>(0.111)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.089</td>
<td>-0.153*</td>
</tr>
<tr>
<td></td>
<td>(0.0877)</td>
<td>(0.0863)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,654</td>
<td>4,020</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.434</td>
<td>0.533</td>
</tr>
</tbody>
</table>

Table 5.11: Descriptive regression of individual improvement in expected # of same-community peers from correlated lottery. Uncertainty of lottery is measured by variance of individual lottery ($\sum s p_{is}(1-p_{is})$). Preference agreement is proxied by product of size of community and average # of top 3 choice agreements with community ($|c(i)|\text{choiceAgreement}(i)$). Non-continuing sibling and continuing are dummies. The standard errors are from clustering on the 205 communities. As can be seen, students for whom the lottery is more uncertain, whose preferences are highly correlated with their communities, and who are not siblings nor continuing students can expect the greatest cohesion gains from correlated lottery.
Figure 5-7: Plot of choiceAgreement($i$) for K1 and K2 new families. For each student $i$, this is the average # of programs she has in common in her top 3 choices with an uniformly random peer $i'$ from the same community. As seen, the median is almost twice larger for K1 new families than for K2. Moreover, for K2, this is low in the middle portions of the city, which matches the pattern of low gain from correlated lottery from Figure 5-6.

5.B.3 Impacts on racial and socio-economic diversity

One concern of using a correlated lottery is that it may cause racial or socio-economic segregation, because race and socio-economic status are correlated with geography, so higher chances to go to school with neighbor may entail higher chances to go to school with others of same race or socio-economic status. We check whether this is a cause for concern in Boston.

One practical concern with our approach is that it might harm racial or socio-economic diversity, which historically is a key reason why many school choice systems were created in the first place. However, in the data, we find that lottery correlation has minimal impact on diversity. As a measure of diversity, we compute the probability that two random students assigned to the same school are of the same race or socio-economic status.

To do this, we compute the probability for a random pair of students who are
Figure 5-8: Interaction of menu change and correlated lottery for K2 new families. The menu we change to is Home Based A (the one adopted by the city). As can be seen, while the median increase from correlated lottery alone is about 0.4, and the median increase from menu change alone is about 0.6, the median increase from both together is about 1.8. The red region in the middle of the city in the second and third graph effectively disappears in the last graph, showing that large cohesion gain is achieved throughout the city when we apply these reforms together.
assigned to the same school to be of the same race or socio-economic status. The races in our data are black (21%), white (16%), asian (7%), hispanic (43%), other (3%), or missing (10%). The percentages in the parenthesis denote the percentage of applicants of that race counting both K1 and K2. For socio-economic status, we use eligibility to receive free or reduced lunch as proxy. The options are free lunch (49%), reduced lunch (4%), non-free/reduced lunch (12%), and missing (36%). To computing the diversity measure, we go through the 1000 deterministic assignments that form the lottery, cycle through all pairs of distinct students who are assigned to the same program, and count the percentage of times they are of the same race or lunch status. We call these the “peer same race %” and “peer same lunch status %.” Table 5.12 shows the results. This is based on actual Round 1 choice data.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Improved</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Same Race %</td>
<td>36.7%</td>
<td>36.9%</td>
</tr>
<tr>
<td>Peer Same Lunch Status %</td>
<td>41.6%</td>
<td>41.8%</td>
</tr>
<tr>
<td>K2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Same Race %</td>
<td>42.6%</td>
<td>42.7%</td>
</tr>
<tr>
<td>Peer Same Lunch Status %</td>
<td>45.2%</td>
<td>45.2%</td>
</tr>
</tbody>
</table>

Table 5.12: Impact of correlated lottery on racial or socio-economic diversity. For example, when we use independent lottery implementation for K1, on average 36.7% of pairs of students assigned to the same school are of the same race. This percentage increases by only 0.2% to 36.9% when we use correlated lottery.

As can be seen, the impact of correlated lottery on racial or socio-economic diversity is minimal. For example, for K1, the same-race probability increases from 36.7% with independent lottery to 36.9% with correlated lottery; for K2, this increases from 42.6% to 42.7%. The impact is similarly small for same-lunch-status probability. One reason why the impact is so small is that geographic patterns of racial and socio-economic concentration in Boston are at a larger scale than the neighborhood sizes we define (.5 mile by .5 mile squares). Moreover, because choices are heterogeneous to begin with, the impact of lottery correlation is on the order of magnitude of about 1 additional neighbor, which does not significantly impact diversity because class sizes are usually 22 or more.
5.C Proofs

Proof of Proposition 2 (NP-hardness of 2 school case): Since there are only 2 schools and since only students in $E$ have non-deterministic assignments, the max-cohesion problem becomes selecting a random $q^*$-subset of $E$ so that each element of $E$ is selected with probability $p = \frac{q^*}{|E|}$, and the $E$’s stay in their communities as much as possible. Partition $E$ into communities, defining $E_c = E \cap I_c$. (And similarly $D_c = D \cap I_c$, $F_c = F \cap I_c$.) Let $x_i$ be the binary random variable that equals 1 iff $i$ is assigned to school 1. To simplify notation, we restrict attention to components of $x$ that corresponds to $E$ (because $D$ is assigned to school 1 with probability 1, and $E$ to school 2 with probability 1). Define $A$ to be all elements of $\{0,1\}^{|E|}$ that sum to $q^*$. The max-cohesion problem becomes:

$$f_{\text{max}} = \text{Max } \Gamma + \left[ \sum_{c=1}^{k} \sum_{i \neq j \in E_c} (x_i x_j + (1-x_i)(1-x_j)) \right]$$

s.t. $E[x_i] = p \ \forall i \in E$

$$x \in \Delta(A),$$

where the constant $\Gamma = 2 \sum_{c=1}^{k} \left[ (\frac{|D_c|}{2}) + (\frac{|F_c|}{2}) + p|D_c||E_c| + (1-p)|E_c||F_c| \right].$

An upper-bound on the maximum cohesion is $\Gamma + 2 \sum_{c} \left( \frac{|E_c|}{2} \right)$, which corresponds to $|E|$’s always being assigned with their community members.

It suffices to prove that it is strongly NP-hard to decide whether this upper-bound is achievable. To do this, we reduce from the following version of 3-PARTITION: Given a multi-set of $k = 3h$ integers $\{a_c\}$, for which $\sum_{c=1}^{k} a_c = hB$, and $\frac{B}{4} < a_c < \frac{B}{2}$. Can the $a_c$’s be partitioned so the sum of each partition is $B$? This is well-known to be strongly NP-hard. A proof is given in Garey and Johnson (1979).

Reduction: given an instance of 3-PARTITION, construct a case of the 2-school max-cohesion problem for which $|E_c| = a_k$, $q^* = B$, then we can achieve perfect cohesion of $U$ iff the $|E_c|$’s can be partitioned so the sum of each partition is exactly $q^*$, which is what we needed. □
Proof of Theorem 6 (Large Market Characterization): We show that a mechanism is non-atomic, incentive compatible, symmetric and efficient within each priority class, if and only if it is lottery-plus-cutoff.

If a mechanism is lottery-plus-cutoff, then it is non-atomic because any single agent changing preferences does not affect the measure of students in each priority class submitting each ranking, which is what cutoffs depend on. The mechanism is symmetric because within the same priority class, the cutoffs are the same for everyone, so since the lottery numbers are identically distributed, the assignment probabilities are the same for those submitting the same preferences. The mechanism is incentive compatible in school-specific utilities because no individual student can change her cutoff, so for each lottery realization her set of feasible schools is fixed. Submitting a false preference, if it changes her assignment at some lottery number, can only place her in a suboptimal school for those lottery numbers. (This is because with a truthful submission, she would already be assigned her most preferred schools among the feasible set by definition.) The mechanism is efficient within each priority class because if student \( i_0 \) prefers school \( s_1 \) over \( s_0 \) (\( s_0 \prec_{i_0} s_1 \)) and \( i_0 \) is assigned to the less-preferred school \( s_0 \) with positive probability, then the cutoff for \( s_1 \) must be higher: \( z^*_{\pi(i_0),s_0} < z^*_{\pi(i_0),s_1} \), because otherwise there would be no lottery number at which \( i_0 \) would select \( s_0 \) as her most preferred school. So trading cycles are prohibited by the arithmetic impossibility:

\[
z^*_{\pi,s_0} < z^*_{\pi,s_1} < z^*_{\pi,s_2} < \cdots < z^*_{\pi,s_l} < z^*_{\pi,s_0}.
\]

Conversely, suppose a mechanism is non-atomic, incentive compatible, and symmetric and efficient within each priority class, we show that it is lottery-plus-cutoff. To do this, we first use non-atomicity to produce a finite set of “representative” students for each priority class, such that every one of the \( m! \) possible preference orderings are submitted by someone from the representative set. By symmetry within each priority class, this pins down the assignment probabilities of everyone in that priority class. We then use incentive compatibility and efficiency within each priority class to
produce a set of cutoffs, and show that the assignment probabilities of everyone in the representative set is consistent with these cutoffs. This implies that the assignment probabilities from the original mechanism is consistent with a lottery-plus-cutoff description. To show that the mechanism itself is lottery-cutoff, we use the flexibility of arbitrary correlation of lottery numbers to define the lottery numbers so that they produce the exact random assignment of the original mechanism. The fact that the assignment probabilities are consistent with the cutoffs will mean that our lottery numbers can be distributed as Uniform[0, 1).

As outlined above, the first step is to produce a finite set “representative” students for each priority class using non-atomicity. For any priority class $\pi \in \Pi$, let the set of students in this class be $I_\pi$. Its measure is $\mu(I_\pi) > 0$. Because there are finitely many rankings, there exists some ranking $\succ_0$ that is picked by a positive measure of students in $I_\pi$. We choose $m!$ of these students who submitted $\succ_0$ and have them alter their preference submission to each of the $m!$ different possible rankings. By iteratively applying non-atomicity, this does not alter the assignment probabilities of anyone else. By symmetry, the assignment probabilities of each of these $m!$ students is representative of the assignment probabilities of anyone with their preference within the whole priority class. Denote the set of representative students $I_{\text{rep}}$. For any ranking $\succ$, define $p(\succ, s)$ as the assignment probability to school $s$ of the student in $I_{\text{rep}}$ corresponding to ranking $\succ$.

The second step is to use incentive compatibility and efficiency within each priority class to produce a set of valid cutoffs. First, note that incentive compatibility in school-specific utilities implies that if any student submits preference list $s_1 \succ s_2 \succ \cdots s_l \succ s_{l+1} \succ \cdots$, reshuffling the relative order of the top $l$ choices among themselves, or reshuffling the later choices among themselves, cannot affect the total assignment probability to $\{s_1, \ldots, s_l\}$ (any of the top $l$ choices). To see this, suppose her utilities for top $l$ schools are all $M + \epsilon_s$ and for later schools $\epsilon_s$ with $M >> \epsilon_s$ for all $s$, then the above condition is necessary for her not to have incentive to deviate from truthful preferences at some preference ranking (because for sufficiently large $M$, the potential to gain a greater probability of getting $M$ trumps any other considerations).
Fix a priority class, focusing on the representative student of this class, we define a complete ordering $>$ over schools, such that $s > s'$ if there exists a student in $I_{\text{rep}}$ who prefers the first school $s$ but receives a positive probability of being assigned to the second school $s'$. The lack of trading cycles among $I_{\text{rep}}$ makes this a well-defined ordering. In the case two schools are not comparable directly, we use transitivity to define their relative order. If after taking into account all ordering relations two schools are still not comparable, we can order them arbitrarily. Re-label the schools so that $s_1 > s_2 > s_3 \cdots > s_m$. Call this the over-demand ordering for this priority class. Intuitively, the earlier schools in this ordering are more highly demanded relative to the supply allocated to this priority class. For each $j$, define “run-out time” $t^*_{\pi,s_j}$ as the assignment probability to $s_j$ of students in $I_{\text{rep}}$ who rank $s_j$ first. (The nomenclature is motivated from the probabilistic serial mechanism.) This is well-defined because of the previous argument from incentive compatibility.

Having defined the “run-out times,” the key step is to note that if school $s$ is more over-demanded than school $s'$ according to the cutoff-ordering, then the run-out time of school $s$ must be weakly smaller, that is $t^*_{\pi,s} \leq t^*_{\pi,s'}$. To see this, note that $t^*_{\pi,s}$ is a student’s chances of getting into $s$ if she ranks it first, and other schools in arbitrary order. Similarly, $t^*_{\pi,s'}$ is a student’s chances of getting into $s'$ if she ranks $s'$ first and $s$ second, and other schools in arbitrary order. Now, if this student reorders her top two rankings to rank $s$ first and $s'$ second, then her total probability of getting into one of these schools must still be $t^*_{\pi,s'}$ by incentive compatibility, but her chance of getting into school $s$ is $t^*_{\pi,s}$, which has to be smaller than the total probability. So $t^*_{\pi,s} \leq t^*_{\pi,s'}$. This means that the run-out times follow a reverse ordering to that of the over-demand ordering of schools, $t^*_{\pi,s_1} \leq t^*_{\pi,s_2} \leq \cdots t^*_{\pi,s_m}$.

We now can define valid cutoffs using these run-out times. Define cutoffs $z^*_{\pi,s} = 1 - t^*_{\pi,s}$. We need to show that everyone’s assignment probabilities are consistent with applying lottery-plus-cutoff on these cutoffs. For any subset of schools $U \subseteq S$, let $\min U$ denote the least over-demanded school in $U$, which by the above paragraph is also the school with the largest run-out time. We observe that if a student in priority class $\pi$ ranks schools $U$ first in arbitrary order, followed by other schools
in arbitrary order, her total assignment probability to some school in $U$ must be equal to $t_{\pi, \text{min } U}^* = \max_{s \in U} t_{\pi, s}^*$. This is because this probability must not depend on her relative ranking within $U$ by incentive compatibility, and so must equal to the probability when she ranks school min $U$ first, which equals to $t_{\pi, \text{min } U}^*$ because she gets into school min $U$ with this probability and other schools in $U$ with zero probability since those are more over-demanded.

The above observation allows us to pin down everyone’s assignment probabilities to every school, since if a student ranks schools $U$ first, followed by $s$, followed by other schools, her assignment probability to $s$ must be

$$\max_{s' \in U \cup \{s\}} t_{\pi, s'}^* - \max_{s'' \in U} t_{\pi, s''}^*.$$

Therefore, if a student in priority class $\pi$ submits ranking $\succ$, her assignment probability to school $s$ is equal to

$$p(\succ, s) = \max(0, t_{\pi, s}^* - \max_{s' \succ s} t_{\pi, s'}^*).$$

This can be re-written as

$$p(\succ, s) = \max(0, \min_{s' \succ s} z_{\pi, s'}^* - z_{\pi, s}^*),$$

which is exactly the assignment probability induced by cutoffs $z_{\pi, s}^*$ and lottery numbers $\propto \text{Uniform}[0, 1]$.

Having shown that the assignment probabilities are consistent with some set of cutoffs, we now show that by correlating the lottery numbers suitably, we can recover the original random assignment exactly. We do this by reverse construction. For student $i$ in priority class $\pi$ who submits ranking $\succ_i$, suppose an instantiation of the random assignment $x$ of the original mechanism assigns $i$ to $s$. Among the schools which $i$ prefers over $s$, let $s'$ be the one with the lowest cutoff. It must be that the cutoff of $s'$ is higher than $s$ because otherwise the student would have never been
assigned to \( s \). We generate

\[
(z_i \text{ conditional on } x_i = s) \sim \text{Uniform}\left[z_{\pi,s}^*, z_{\pi,s'}^*ight).
\]

Since \( i \) is assigned to \( s \) with probability exactly \( z_{\pi,s'}^* - z_{\pi,s}^* \), generated in this way, the unconditional \( z_i \) would be distributed \( \text{Uniform}[0, 1] \). The lottery-plus-cutoff mechanism using these lottery numbers \( z_i \), along with the cutoffs \( z_{\pi,s}^* \), generates the exact same random assignment as the original mechanism, which is what we needed to show.

**Proof of Theorem 7:** Using the cutoff structure, we can express cohesion in both the independent and max-cohesion cases in terms of simple integrals. The theorem follows from re-arranging orders of integration and the law of total variance.

\[
f_s^{\text{independent}} = \int_c \int_y p_s(c, y) p_s(c, y') dy' dc
= \int_c \left( \int_y p_s(c, y) dy \right)^2 dc
= E_c[\bar{p}_s(c)^2]
= E_c^2[\bar{p}_s] + \text{Var}_c(\bar{p}_s)
= q_s^2 + \text{Var}_c(\bar{p}_s).
\]

Where the last equality follows from our assumption that total supply exactly equals demand.

For max-cohesion, we first note that by giving everyone in the same community the same lottery number, we can achieve cohesion of \( E_{c,z}[d_s^2(z|c)] \). This is because at lottery number \( z \), the measure of students from community \( c \) that would be assigned to school \( s \) is \( d_s(z|c) \), so the measure of pairs of students is \( d_s^2(z|c) \). Because there is a continuum of communities, this lottery-correlation scheme is feasible, since the total measure of students that would be assigned to school \( s \) remains equal to supply.
Hence, maximum cohesion satisfies

\[ f_{\text{max}}^s \geq E_{c,z}[d_s^2]. \]

On the other hand, the probability of any two students being co-assigned to school \( s \) is upper-bounded by \( \min(p_s(i), p_s(i')) \). But if \( i \) and \( i' \) are in the same community, they see the same school cutoffs, because we assumed that everyone in the same community are in the same priority class, so

\[
\min(p_s(i), p_s(i')) = \int_z 1_s(z|i)1_s(z|i')dz.
\]

Therefore,

\[
f_{\text{max}}^s \leq \int_c \int_y \int_y' \min(p_s(c, y), p_s(c, y'))dy'dy'dc
= \int_c \int_y \int_y' \int_z 1_s(z|(c, y))1_s(z|(c, y'))dzdy'dy'dc
= \int_c \int_z \int_y \int_y' 1_s(z|(c, y))1_s(z|(c, y'))dy'dydzdc
= \int_c \int_z \left( \int_y 1_s(z|(c, y))dy \right)^2 dzdc
= \int_c \int_z d_s^2(z|c)dzdc
= E_{c,z}[d_s^2].
\]

Combining the two, we get that

\[
f_{\text{max}}^s = E_{c,z}[d_s^2].
\]

Note that the above correlation scheme would not be feasible in a finite market, but our having a continuum of communities allows the variation in demand from this level of extreme correlation to be averaged away so that we do not violate any capacity constraints. In a certain sense, the continuum of communities smooths away
the NP-hardness of the max-cohesion problem in the finite case.

**Proof of Proposition 3**

\[ f_{\text{max}} = \int_c \int_y \int_{y'} \min(p_s(c, y), p_s(c, y')) dy'dydc \]
\[ = \int_c \int_y \int_{y'} \frac{\max(p_s(c, y), p_s(c, y')) + \min(p_s(c, y), p_s(c, y'))}{2} dy'dydc \]
\[ - \int_c \int_y \int_{y'} \frac{\max(p_s(c, y), p_s(c, y')) - \min(p_s(c, y), p_s(c, y'))}{2} dy'dydc \]
\[ = \int_c \int_y p_s(c, y) dydc - \frac{1}{2} \int_c \int_y \int_{y'} |p_s(c, y) - p_s(c, y')| dy'dydc \]
\[ = q_s - \frac{1}{2} E_c[\Delta_s(c)], \]

which is what we needed to show.

**Proof of Proposition 4**

\[ f_{\text{max}} - f_{\text{independent}} = \int_c \int_y \int_{y'} \min(p_s(c, y), p_s(c, y')) - p_s(c, y)p_s(c, y') dy'dydc \]
\[ = \int_c \int_y \int_{y'} p_s(c, y) + p_s(c, y') - \frac{p_s^2(c, y) + p_s^2(c, y')}{2} dy'dydc \]
\[ - \int_c \int_y \int_{y'} |p_s(c, y) - p_s(c, y')| - (p_s(c, y) - p_s(c, y'))^2 dy'dydc \]
\[ = E_s[p_s(i)(1 - p_s(i))] - \frac{1}{2} E_{c(i)=c(i')}[\Delta_s(i, i')(1 - \Delta_s(i, i'))], \]

which is what we needed.

**Proof of Proposition 5** The structural identity follows directly from Proposition 3 and Theorem 7.

\[ q_s(1 - q_s) = q_s - f_{\text{independent}} + \text{Var}_c(\bar{p}_s) \]
\[ = \frac{1}{2} E_c[\Delta_s(c)] + f_{\text{max}} - f_{\text{independent}} + \text{Var}_c(\bar{p}_s). \]
Proof of Proposition 6: The first line follows directly from Theorem 7, and observing that there is no between community variation in assignment probabilities (since communities have identical preferences and sizes).

The second line follow from when $\beta \to \infty$, we get the case in which everyone goes to their closest school.

The third and fourth line follows from almost everyone’s preferences in the same community become the same as $\alpha \to \infty$ or $\beta \to \infty$, so the demand function $d_s(z|c)$ in any community for any school at any lottery number become close to 1. The result follows from the identity in Theorem 7.

Note that in the above, for independent lottery to achieve perfect community cohesion, we needed school capacities to be just right. No such assumption is needed in the proof for the correlated lottery case.

Proof of Theorem 8: If $\beta = 0$, then there cannot be between community variation in assignment probabilities, so by Theorem 7 $f_{\text{independent}}^s = q_s^2$.

Label the schools so $r_1 > r_2 > \cdots$. (Recall that $r_s = e^{\nu_s}$.) Because of the logit utility structure, in any community, the probability that school $s$ is chosen first is

$$\frac{r_s}{\sum_{j=1}^{m} r_j}.$$ 

Because we assumed no priority, the mechanism is equivalent to probabilistic serial (or lottery-plus-cutoff with a single priority class). Define run-out time $t_s^* = 1 - z_s^*$. The cutoff for the most highly demanded school, school 1, is such that

$$t_1^* = q_1\left(\frac{\sum_{j=1}^{m} r_j}{r_1}\right).$$

By the independence of irrelevant alternative properties of logit, these students, given a lower lottery number, would choose school $s$ with probability $\frac{r_s}{\sum_{j=2}^{m} r_j}$. The cutoff
for the 2nd most demanded school is such that

\[ t_2^* - t_1^* = \left( q_2 - t_1^* \sum_{j=2}^{m} r_j \right) \frac{\sum_{j=2}^{m} r_j}{r_2} \]

\[ = \left( q_2 - q_1 \right) \sum_{j=2}^{m} r_j. \]

Define tail sum \( R_s = \sum_{j=s}^{m} r_j \). Continuing in this way by induction we get

\[ t_s^* = \sum_{j=1}^{s-1} q_j + \frac{R_s}{r_s} q_s, \]

and

\[ t_s^* - t_{s-1}^* = \left( \frac{q_s}{r_s} - \frac{q_{s-1}}{r_{s-1}} \right) R_s, \]

which implies the formula for \( f_{max}^s \),

\[ f_{max}^s = \sum_{k=1}^{s} \left( t_k^* - t_{k-1}^* \right) \left( \frac{R_s}{R_k} \right)^2 = r_s^2 \sum_{k=1}^{s} \left( \frac{q_k}{r_k} - \frac{q_{k-1}}{r_{k-1}} \right) \frac{1}{R_k}. \]

Since increasing \( \alpha \) increases the ratios \( \frac{r_1}{r_2}, \frac{r_2}{r_3}, \ldots, \frac{r_{m-1}}{r_m} \), it suffices to show that if one of these ratios increase, while the others stay the same, maximum cohesion increases. In other words, it suffices to show that for any \( l \), if we perform the transformation

\[ r_1, r_2, \ldots, r_l, r_{l+1}, \ldots, r_m \]

\[ \downarrow \]

\[ \gamma r_1, \gamma r_2, \ldots, \gamma r_l, r_{l+1}, \ldots, r_m \]

with \( \gamma > 1 \), we increase maximum cohesion \( f_{max}^s \).

Now, for \( l \leq s - 1 \), we have

\[ f_{max}^s = r_s^2 \sum_{k=1}^{s} \left( \frac{q_k}{r_k} - \frac{q_{k-1}}{r_{k-1}} \right) \frac{1}{R_k} \]

\[ = \frac{q_s r_s}{R_s} - r_s^2 \sum_{k=1}^{s-1} \frac{q_k}{R_k R_{k+1}}. \]

so \( f_{max}^s \) increase because the transformation fixes \( r_s, R_s \) and the \( q \)'s, while only
increasing some of the $R_k$’s in the denominator of the negative term. (Recall that $R_k = \sum_{j=k}^{m} r_j$.)

For $l \geq s$, we write $f_{\text{max}}^s$ in a different way

$$f_{\text{max}}^s = r_2^s \sum_{k=1}^{s} \left( \frac{q_k}{r_k} - \frac{q_{k-1}}{r_{k-1}} \right) \frac{1}{R_k}$$

$$= \sum_{k=1}^{s} \left[ \frac{q_k r_s}{r_k} - \frac{q_{k-1} r_s}{r_{k-1}} \right] \frac{r_s}{R_k},$$

so $f_{\text{max}}^s$ also increases by the transformation, since each of $\frac{q_k}{r_k}$ and $\frac{r_s}{r_{k-1}}$ stays fixed, while $\frac{r_s}{R_k}$ increase because the numerator increase by a factor of $\gamma$ while the denominator $R_k = \sum_{j=k}^{m} r_j$ increase by a factor less than or equal to $\gamma$ (some of the summands increase by factor $\gamma$ while others stay the same). This completes the proof.

5.D Detailed analysis with two schools

Continuing from the notation in Section 5.2.1 and proof of Proposition 2 in the 2 school case a mechanism using independent lotteries simply chooses a $q^*$-subset of $E$ uniformly randomly, and assign them to school 1. We call the cohesion obtained by independent implementation of the lottery $f_{\text{independent}}$. It is straightforward to work out

$$f_{\text{independent}} = \Gamma + \left( 1 - \frac{2q^* (|E| - q^*)}{|E|(|E| - 1)} \right) \left( 2 \sum_c \left( \frac{|E|}{2} \right) \right) \quad (5.13)$$

We want to bound the ratio $\frac{f_{\text{max}}}{f_{\text{independent}}}$, which is how much correlated lottery improves cohesion in this setting. However, $f_{\text{max}}$ is computationally hard to solve.

However, we can come up with fairly good bounds on $f_{\text{max}}$ when the number of communities is large. Imagine that we arrange students in $E$ in a circle, with members of the same community being adjacent to one another. We can randomly split this circle into a group of $q^*$ and a group of $|E| - q^*$ by uniformly randomly selecting a student and counting $q^*$ students clockwise from her. This keeps communities together except for at most 2 communities. The expected number of same-community pairs

233
split is at most \[2 \sum_c |E_c| \left( \frac{1}{\max_c (|E_c|^2)} \sum_a |E_c| a(|E_c|-a) \right) = 2 \sum_c |E_c| \frac{(|E_c|-1)|E_c|+1}{\max_c (|E_c|^2)} \leq 2 \sum_c |E_c| \frac{|E_c|}{\max_c (|E_c|^2)} \leq \max_c 2 (|E_c|).
\]

So this randomization, while not necessarily keeping all communities together, gives up at most the largest community. Thus,

\[\Gamma + 2 \sum_c \left( \frac{|E_c|}{2} \right) - 2 \max_c \left( \frac{|E_c|}{2} \right) \leq f(\text{max}) \leq \Gamma + 2 \sum_c \left( \frac{|E_c|}{2} \right) \tag{5.14}\]

Combining this with previous expression for \(f_{\text{independent}}\),

\[1 \leq \frac{f_{\text{max}}}{f_{\text{independent}}} \leq \frac{\Gamma + 2 \sum_c \left( \frac{|E_c|}{2} \right)}{\Gamma + (1 - \frac{2q^* (|E|-q^*)}{|E|(|E|-1)})(2 \sum_c \left( \frac{|E|}{2} \right))} \tag{5.15}\]

When the number of communities \(k\) is large and \(|E_c|\)'s are approximately equal, \(\frac{\max_c (\frac{|E_c|}{2})}{\sum (\frac{|E|}{2})} \to 0\), so the upper-bound is asymptotically achievable using the lottery correlation scheme defined above.

The maximum improvement occurs if \(\Gamma\) is small, and \(\frac{2q^* (|E|-q^*)}{|E|(|E|-1)} \approx p(1-p)\) is large, in which case the improvement ratio \(\frac{f_{\text{max}}}{f_{\text{independent}}} \approx \frac{1 - \frac{2q^* (|E|-q^*)}{|E|(|E|-1)}}{1 - \frac{2q^* (|E|-q^*)}{|E|(|E|-1)}} \leq 2 + \frac{2}{|E|-2}\). So the maximum possible improvement from correlated lottery is about a factor of 2. (This is achieved if for example there are no priorities and the lottery is most random, so \(\Gamma = 0\), \(p = \frac{1}{2}\), in which case an independent lottery splits any given same-community pair about \(\frac{1}{2}\) the time, and a correlated lottery, using the scheme described above, keeps communities together almost all the time.)

5.D.1 Large market approximation with two schools

The NP-hardness in the finite case precludes precisely pinning down max-cohesion \(f_{\text{max}}\), but as seen in the above this difficulty disappears when the number of communities is large. We examine this “large-market” case more closely to yield more intuition on when exactly we can expect correlated lottery to improve cohesion the most.

We approximate the expressions for \(f_{\text{independent}}\) and \(f_{\text{max}}\) in Equations 5.13 and
5.14 In this large market case. For each community \( c \), define \( n_c = |I_c| \) and define \( u_c, v_c, w_c \) to be the fraction of this community in sets \( D, E, F \) respectively. So \( u_c = \frac{|D_c|}{|n_c|} \), \( v_c = \frac{|E_c|}{|n_c|} \), \( w_c = \frac{|F_c|}{|n_c|} \), and we can approximate \( 2\binom{|D_c|}{2} \approx n_c^2 u_c^2 \), \( 2\binom{|E_c|}{2} \approx n_c^2 v_c^2 \) and \( 2\binom{|F_c|}{2} \approx n_c^2 w_c^2 \). Moreover, we approximate the term in Equation 5.13, \( \frac{q^*(|E|-q^*)}{|E|(|E|-1)} \approx \frac{q^*}{|E|^2} = p(1-p) \). To simplify notation, treat \( c \) as a random variable that takes a particular value \( c_0 \) with probability \( \frac{n_{c_0}}{\sum_{c'} n_c'} \). Rescale cohesion to be between 0 and 1 by defining \( \tilde{f} = \frac{f}{\sum_{c'} n_{c'}} \). We have

\[
\tilde{f}_{\text{independent}} \approx \mathbb{E}[u_c^2 + v_c^2 + 2pu_c v_c + 2(1-p)v_c w_c + (1-2p(1-p))v_c^2] \\
= \mathbb{E}[(u_c+pv_c)^2] + \mathbb{E}[(w_c+(1-p)v_c)^2] \\
= \mathbb{E}[u_c + pv_c] + \mathbb{E}[w_c + (1-p)v_c] + \text{Var}(u_c + pv_c) + \text{Var}(w_c + (1-p)v_c) \\
= \tilde{q}_1^2 + \tilde{q}_2^2 + 2\text{Var}(u_c + pv_c).
\]

Where \( \tilde{q}_1 = \frac{q_1}{n} \), \( \tilde{q}_2 = \frac{q_2}{n} \) are the proportion of seats that are in school 1 and 2 respectively. Moreover, when the number of communities is large and when the relative size of the largest community is small, we have that the upper-bound in Equation 5.14 becomes equality, so

\[
\tilde{f}_{\text{max}} \approx \mathbb{E}[u_c^2 + v_c^2 + 2pu_c v_c + 2(1-p)v_c w_c + v_c^2] \\
\approx \tilde{f}_{\text{independent}} + 2\mathbb{E}[p(1-p)v_c^2].
\]

We can interpret the terms as follows:

- \( \tilde{q}_1^2 + \tilde{q}_2^2 \): Herfindahl index of school size. This is a measure of school size concentration. It is higher when school sizes are more unequal.

- \( 2\text{Var}(u_c + pv_c) \): Between community variation in assignment probabilities. \( u_c + pv_c \) is exactly the expected proportion of people from community \( c \) who go to school 1. This variation is 0 if assignment probabilities in communities are identical, and increases if preferences and/or priorities result in more varied proportion of people assigned to each school in different communities.
• $E[2p(1 - p)v_c^2]$: How much independent lottery hurts cohesion. This is the expected number of same community pairs split up by independent lottery. It is higher when communities have more people affected by the lottery ($|E|$ is higher) and when the lottery is more “uncertain” (closer to $p = \frac{1}{2}$). For fixed $p$, it is higher when the variation across communities of the number of people affected by the lottery ($Var(v_c)$) is higher (or equivalently when within-community heterogeneity is low).

An interesting special case is when there are no priorities and school sizes equal, in which the above expressions immediately yield.

**Proposition 7.** When there are no priorities ($u_c = 0$) and $q_1 = q_2 = \frac{1}{2}$, we have that the proportional improvement from correlated lottery is

$$\frac{\tilde{f}_{\text{max}} - \tilde{f}_{\text{independent}}}{\tilde{f}_{\text{independent}}} = \frac{1 - p}{p} = 2E[v_c] - 1,$$

which can be interpreted as a measure of overall preference correlation or competition.

**Proof of Proposition 7.** When $u_c = 0$ and $q_1 = q_2 = \frac{1}{2}$, we have $p = \frac{1}{2E[v_c]}$, so

$$\tilde{f}_{\text{independent}} = E[(u_c + pv_c)^2] + E[(w_c + (1 - p)v_c)^2]$$

$$= p^2E[v_c^2] + E[(1 - pv_c)^2]$$

$$= 2p^2E[v_c^2] + 1 - 2pE[v_c]$$

$$= 2p^2E[v_c^2].$$

So

$$\frac{\tilde{f}_{\text{max}} - \tilde{f}_{\text{independent}}}{\tilde{f}_{\text{independent}}} = \frac{2p(1 - p)E[v_c^2]}{2p^2E[v_c^2]}$$

$$= \frac{1 - p}{p}.$$

$\square$
Based on this, we gather the following qualitative insights: to have higher cohesion when the lottery is implemented independently, one needs to have more varied school sizes, more between-community variations in preferences, or have the over-demanded school give more priority to communities that most demand it. Furthermore, the potential to improve cohesion via correlated lottery is greater when more people are affected by the lottery, when within-community preference correlation is high, or when the lottery is more uncertain. When school sizes are equal, the cohesion improvement is greater when more people desire the over-demanded school.

5.E Inability of lottery correlation to increase diversity

Besides community cohesion, another desirable outcome might be racial or socio-economic diversity. Unfortunately, we show in this section that running an independent lottery already achieves near optimal diversity, so any significant gain requires changing the assignment probabilities.

To see why this is true, consider the assignment of students of different races to a school of capacity $q$. Suppose there are $r$ different races. In the random assignment, let $n_1, n_2, \ldots, n_r$ be random variables denoting the number of students of each race assigned to the school. Suppose that the school is always assigned at capacity, so $\sum_{j=1}^{r} n_j = q$. A metric for diversity is the number of pairs of students of different races assigned to the same school, so maximizing diversity at this school is to maximize

$$ E[\sum_{j_1 \neq j_2} n_{j_1} n_{j_2}] .$$

This is the same as minimizing

$$ q^2 - 2E[\sum_{j_1 \neq j_2} n_{j_1} n_{j_2}] = E[\sum_{j=1}^{n} n_j^2] .$$
However, by Jensen’s inequality,

\[ E\left[ \sum_{j=1}^{n} n_j^2 \right] \geq \sum_{j=1}^{n} E[n_j]^2. \]

This lower bound is approximately achieved by running an independent lottery, because in that case \( n_j \) is almost always close to \( E[n_j] \) by Chernoff bound. In fact, in the large market model of Section 5.3, \( n_j \) would be equal to \( E[n_j] \) always, and this lower bound is exactly achieved by independent lottery implementation.

In a finite market, one can achieve small diversity gains by constraining \( n_j \) to be always between \( \lfloor E[n_j] \rfloor \) and \( \lceil E[n_j] \rceil \). More precisely, one would express the original assignment probabilities as a convex combination of such constrained deterministic assignments. This idea has been previously explored in Budish et al. (2013). However, by the above argument, one would expect gains in diversity to be small.
Chapter 6

Concluding Remarks

In this concluding chapter, I summarise the main insights of this thesis that might be generalizable to broader contexts, and comment on the limitations of the thesis and the remaining work needed before one can implement these techniques in practice.

6.1 Generalizable Insights

The main generalizable insights from this thesis are the following:

- Quantifying outcomes and simulating counterfactuals can help inform policy debate and bypass rhetorical gridlocks.

- Accurately predicting counterfactuals remains a challenge even with state-of-the-art tools. Changes in aspects of the system that are not modelled can completely skew the results; however, if one controls for these “changes in the environment,” multinomial logit and mixed logit models of discrete choice seem to work well for predicting students’ rankings over schools.

- Assortment planning and choice set design in matching markets are related. Techniques for assortment planning from the revenue management literature can be adapted to design optimal menus, priorities and quotas for many-to-one matching markets that use the deferred acceptance algorithm.
In many-to-one matching markets that use lotteries, correlated implementation of the lotteries can be a powerful tool to improve upon second-order moments of the assignment (such as community cohesion), without altering anyone’s assignment probabilities.

Large market (continuum) models of matching markets yield tractable analysis and generate useful insights that are approximately true even in discrete settings.

### 6.2 Remaining Work Needed Before Implementation

This thesis provides the proof of concept to the effective application of several data-driven techniques to school choice, using data from Boston for numerical tests. However, certain techniques need to be further refined before they can be directly implemented in practice. The remaining work before implementation can be classified into two categories: the first represents additional engineering effort to conduct numerical tests and robustness checks and to integrate other institutional details important in practice; the second represent resolving fundamental gaps that still remain between theory and practice, which warrant additional research.

Within the first category, one project that remains to be done is to simulate the performance of the optimized menus and priorities from Chapter 4 under a more detailed setup that incorporates English Language Learners, continuing students from earlier grades, and students who have older siblings at other schools. This requires estimating forecasts of the number of such students from each neighborhood and running a simulation that spans several years (which is needed to incorporate continuing students). Building a full simulation model of everything that goes on in Boston Public Schools (BPS) is very ambitious. Although it may be very useful to BPS for planning purposes, the engineering effort required makes it outside of the scope of this thesis.

Another remaining project within the first category is the re-evaluation of various techniques in this thesis under a unified numerical setup. Currently, the main criteria
for the welfare of students is different in Chapter 2 and in Chapter 4. In Chapter 2, the criteria is access to quality, which defines desirable schools using a test-score-based metric; in Chapter 4, the criteria is expected utility, which is based on students’ choices. While the second criteria is academically more sound, the first criteria is more easily accepted by the public and currently more important in policy debates. So it would be interesting to also evaluate the performance of the optimized plan in Chapter 4 in terms of access to quality as defined in Chapter 2. Analogously, it would be interesting to evaluate the correlated lottery implementation of Chapter 5 under the distance-based student-centric definition of community cohesion in Chapter 4, which does not require a-priori partitioning the city into communities. The student-centric definition does not fit the current framework in Chapter 5 but is easier to interpret and more convincing in policy debates.

While the first category of remaining work are relatively straightforward to conduct, the second category of remaining work represent fundamental gaps between theory and practice. One example of such a gap is that the perspective of school choice presented in this thesis is static and one-shot, with all students submitting choices at the same time and the capacity of schools a-priori known and fixed. In practice, there are several rounds in which students may participate in the choice system; furthermore, a family can send the child to school for the first time in grade K2, or send the child earlier in grades K0 or K1. Furthermore, the capacity of schools can be adjusted by the school board, possibly based on the number of applicants. Because of these dynamic aspects of student application and these flexibilities on the part of the school board, it might not be the best idea to exactly adopt the complex menus and priorities from Chapter 4 based on exact balancing of supply and demand. If one relaxes the constraints of exact balance and if one make use of the supply flexibility of the school board, perhaps one can come up with a much simpler plan that works better.

A related gap between theory and practice is that the current theory evaluates the performance for a given year, while in reality aspects such as predictability and transportation savings are dependent on the assignment of multiple years. With
the current framework, this across-year optimization of menus and priorities seems intractably complex. Further techniques are needed to better study this across-year dependence.

Another unresolved research question the feedback loop between changes in the assignment system and changes in students’ preferences or application rates. For example, the current framework does not consider families who move to another neighborhood or move to a private school in response to the changes in menus and priorities, nor does it consider changes in the assignment system affecting the qualities or funding levels of schools and in turn changing families’ preferences. Currently, there is little empirical evidence in Boston that the reform to the Home-Based system resulted any substantial changes in residential choice or school qualities, but these effects are theoretically plausible and further research is needed to understand how important they really are.
Bibliography


Ashlagi, I., A. Nikzad, A. Romm. 2015. Assigning more students to their top choices: A tiebreaking rule comparison. Available at SSRN 2585367.


245


Menino, T. 2012a. The honorable mayor thomas m. menino state of the city address. Available at http://www.cityofboston.gov/mayor/speech.asp.


