THE UNIVERSITY OF CHICAGO

ESSAYS ON SCHOOL CHOICE AND CIVILIAN COLLABORATION IN CIVIL CONFLICTS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
MARIANA LAVERDE QUINTERO

CHICAGO, ILLINOIS
JUNE 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank my advisors Scott Ashworth, Kerwin Charles, Steven Durlauf and Seth Zimmerman. I am extremely grateful for their constant support and their generosity sharing with me their insights, their time and, the always needed encouragement. Many other Professors at the University of Chicago provided help that, in one way or another, was instrumental for the execution of this dissertation. I thank Michael Dinnerstein, Dan Black, Wioletta Dziuda, Ingvil Gaarder, Yana Gallen, Derek Neal, Luis Martinez, Maria Angelica Bautista, James Robinson, Austin Wright, Chris Blattman, Oeindrila Dube, Peter Ganong and Anthony Fowler. I am indebted to Mohammad Akbarpour, whose class and subsequent discussions I always found inspiring, and to Margaux Luflade for her generous help and for being a female image to look-up to.

The group of Ph.D. students at the Harris School are an exceptional group of people that contributed in many ways to my learning, as well as to my wellbeing. I am specially grateful to Wendy Wong, Yuvraj Pathak, Katherine Baird, Chenyu Qiu, Olga Namen, William Delgado, Mariella Gonzales, Val Michelman, Miguel Morales, Derek Wu, Justin Holz, Scott Lee, Laura Montenegro, Maria Adelaida Martinez and Victor Ruan. The Pearson Institute staff, fellows and scholars were always a joyful and energetic group. They provided meaningful support and a good share of adventures. A special word of appreciation goes to Cynthia Cook-Conley whose warmth and kindness makes the Harris School feel like home.

My time in Chicago left me unexpectedly with a group of one-of-a-kind friends. Many thanks to Esperanza Johnson, Nicolas Castro, Lucila Cardona, Pablo Robles, Daniela Podda, Simone Lenzu, Bettina Hilliger, Nick Tsivanidis, Silvia Acosta, Felipe Labbe, Mari Perez, Jose Tudon, Chien-Yu Lai, Celia Roussand, Molly Schwartz, Alejandro Hoyos, Sofia Correa and Ignacio Cuesta for making this time much more fun.

Finally, many thanks to Santiago who celebrated every one of my milestones, and brightened my days with laughter and music.

ABSTRACT

In this dissertation, I compile two papers. The first on assignments to public schools and the limits of schools choice; the second on the impact of commodity price shocks on civilian collaboration and violence in civil wars.

In chapter 1, I study the limits of school choice policies in the presence of residential sorting. Using data from the Boston Public Schools choice system, I show that [redacted]. [redacted], I use rich data on applicants' rank-order choices to estimate preferences over schools, and consider a series of counterfactual assignments. I find that between [redacted] of the gap in achievement at the schools assigned to black and Hispanic students relative to those assigned to white students is explained by travel costs to high-performing schools. Differences in preferences for schools explain about [redacted] of the gap, while algorithm rules [redacted]. [redacted].

Chapter 2 is joint work with Austin Wright. We propose a mechanism that rationalizes changes in violence in civil wars after economic shocks and test it with newly collected data. The rationale we propose relies on a theory of violence where political groups choose the type of violence -selective or indiscriminate- that maximize their expected control, and where information provided by civilian informants determine the relative effectiveness of these types of violence. Civilians, are producers of an agricultural good and choose the political group they will supply information to, if any, having considerations on expected revenue and chances of survival. We argue that increases in the price of commodities whose production relies on collaboration reduce the incentives of civilians to inform against fellow

community members. Since information is necessary to effectively carry out selective attacks, a reduction in the information available to political groups will cause a reduction in selective violence. We use a text analysis algorithms to classify violent attacks and test the mechanism.

CHAPTER 1

UNEQUAL ASSIGNMENTS TO PUBLIC SCHOOLS AND THE LIMITS OF SCHOOL CHOICE

1.1 Introduction[1]

Since the late 1980s, many cities across the United States have adopted centralized school choice systems.[2] These systems allow families a choice among public schools, as opposed to neighborhood assignments where school districts assign students to schools based on proximity to residences. Since typically lower-achievement schools are in low-income areas populated by racial and ethnic minorities, neighborhood assignments replicate residential segregation and sustain educational inequality across racial and income groups. By decoupling residences and schools, choice systems are believed to create opportunity for desegregation and equal access to educational quality. As Boston Public Schools' superintendent wrote in the proposal for the 1988 choice plan: "My overall goal is to create a student assignment plan that provides all Boston students with high-quality desegregated education" (Boston Desegregation Project 1988).

This paper asks how effectively choice systems reduce cross-racial gaps in access to quality education relative to a geographic assignment, and why. Using assignments data from Boston Public Schools (BPS), I begin by showing that under Boston's choice system white pre-kindergartners are assigned to schools with higher achievement than black and Hispanic students. Moreover, average achievement of the schools assigned to white, black, and Hispanic students is the same as that generated under a neighborhood system where students are

1. I thank Lisa Harvey and Apryl Clarkson from Boston Public Schools for providing data and guidance. Results are under review by data providers so parts of the text have been redacted accordingly. I gratefully acknowledge support by the Successful Pathway from School to Work initiative of the University of Chicago, funded by the Hymen Milgrom Supporting Organization. All errors are mine.

2. According to the non-profit *Education Commission of the States*, 47 states plus the District of Columbia have passed laws to allow or mandate a version of school choice. School districts that have implemented open enrollment include New York, Boston, Cambridge, Charlotte, and New Haven

assigned based on proximity to schools.[3] School choice assignments are identical to neighborhood assignments for about [redacted] of students. The remaining students are on average assigned to schools with marginally higher achievement, with the highest gains concentrated in the white population.[4] As a consequence, choice does not translate into more access to high-performing schools for Hispanic and black families, nor does it result in reductions in the achievement gap at the assigned schools compared to white students.

An effective policy response to the above depends on why the effects of choice are limited. I argue cross-race differences in choice-based assignments may stem from (i) differences in travel costs, (ii) differences in preference for schools, or from (iii) assignment rules that generate different probabilities of assignment conditional on parents' preferences. To distinguish between these channels, I combine detailed application data from BPS with a structural model of school demand to estimate racial-specific preferences for schools. I use the estimated parameters to generate counterfactual assignments that quantify the contribution of various mechanisms to the observed school achievement gap.

The mechanisms that explain the differential assignments of white, black and Hispanic students need to stem from either differences in the demand of high-achieving schools, or from assignment rules that generate different probabilities of assignment conditional on parents' preferences. In the next two paragraphs I'll discuss how parents' demand can explain the gap, as well as discuss which assignment rules have the potential to contribute to it.

In my model, parents' demand for schools is determined by two main components. First, the distance between the school and the families' residence, and second, the value parents place in all other school characteristics. The former determines the travel cost to a school. The latter is the location-independent value of a school, that is, the attractiveness of a school that is independent of distance to the students' residence.

3. I generate a neighborhood assignment matching students to schools in proximity order while taking into account school capacities. Specifically, I run a DA algorithm where preferences and priorities are fully determined by distance.

4. I use information of students assigned to schools in the first round of applications. Since there are more seats than students assigned in the first round, average improvements for all students are possible.

Assignment rules that depend on the residential location of students may contribute to explain the gap in two ways. First, school districts typically prioritize students for assignment based on proximity to schools. This means that students that live closer to high-achieving schools are more likely to get assigned to these schools.[5] Second, school districts can restrict the menu of schools parents can apply to based on closeness to a students' residence.[6] These rules may reduce the probability that black and Hispanic students get assigned to high-achieving schools if these schools are less often in their menus.

To disentangle the three mechanisms, I use detailed data on all first-round applicants to a seat in pre-kindergarten in BPS between 2010 and 2012. The data includes all rank-order lists of schools submitted by parents to BPS and the residential location and race of each applicant. I first estimate group-specific preference parameters from a random utility model using the rankings submitted by parents to BPS.[7] Under some identification assumptions, the structural demand model allows me to separately identify parents' assessment of travel costs, and the valuation of each school net of this cost. Using these parameters I simulate counterfactual rankings after changing the location-independent preferences for schools, the travel costs, or the assignment rules. I use these rankings and recreate the Deferred Acceptance (DA) algorithm Gale and Shapley 1962) used in Boston to generate counterfactual assignments and the resulting average achievement at the schools assigned to students of each race. Boston is a good setting to estimate school demand since the DA does not reward strategic play, and the district does not impose limits to the length of the rankings submitted. Mechanisms that do not meet these criteria may not generate truthful reports (Abdulkadiroglu et al. 2005, Haeringer and Klijn 2009, Calsamiglia et al. 2010).

In a first counterfactual, I quantify how much of the gap is explained by the residential

5. Dur et al. [2018] show that having a proximity priority under the precedence order used in Boston, does not importantly increase the fraction of walk-zone students admitted relative to an assignment where the proximity priority is abolished

6. Restricting school menus based on geography is not very common across school districts. BPS has had this type of restrictions since the early 1990s.

7. Similar model are estimated in Abdulkadiroglu et al. [2017] and Pathak and Shi [2013]

location of students. I simulate submitted rankings and assignments for black and Hispanic students after a change in residential location. Specifically, new locations for black and Hispanic students are randomly drawn from white students' residences. In each new location, I use the estimated parameters to generate rankings and subsequently assignments using the DA algorithm. I change the residential location of one student at a time, to make sure I can sustain the assumption that schools and hence preferences are unchanged. This counterfactual parallels the Moving to Opportunity (MTO) experiment that relocated families from high-poverty neighborhoods to low-poverty communities in the late 1990's.[8] Results from the counterfactual I propose show first-order implications of a change in residential location within a city.

Changing the residential location of a student is a bundled treatment. Students that are relocated face different travel costs to high-achieving schools, while assignment rules that are location-specific impact the probability of assignment to schools with higher achievement. To disentangle between the effect of assignment rules and travel costs, in a second counterfactual I independently vary assignment rules. Specifically, I first generate assignments assuming that there are no restrictions over choice menus, and later consider the case where proximity priorities are eliminated.

Finally, I simulate assignments under a change in the location-independent preferences for schools. In these counterfactuals, I generate assignments where black and Hispanic students take white students' preference parameters, while the original location of each student's residence is unchanged. Results from this counterfactual highlight how preferences for location-independent school characteristics impact the observed gap. Differences across races in these preferences may capture trade-offs made by parents between demographic and academic school characteristics, as well as any other dimensions of preference heterogeneity.

I find that after a change in residential location the gap in achievement at the schools

8. Papers that study the impacts of this experiment include Ludwig et al. [2013], Chetty et al. [2016], Katz et al. [2001], Kling et al. [2007], Clampet-Lundquist and Massey [2008]

assigned to treated students and white students reduced by about [redacted] relative to the original gap. A change in the location-independent preferences of schools explains [redacted] of the original gap. Finally, eliminating proximity priorities and choice menu restrictions [redacted]. This suggests that the impact of a residential location change is fully explained by changes in travel costs to high-achieving schools.

The salience of travel costs on the resulting school assignments has important policy implications. It suggests that school choice alone does not always mitigate the undesirable effects of residential sorting and that there may be gains from coordinating the efforts of school and housing authorities. Increasing investment in schools close to constrained students, while guaranteeing housing affordability can increase access to quality education and possibly reduce school segregation if less constrained students react to these investments. Alternatively, policies that incentivize residential desegregation can lead to more equity in schooling.

This paper contributes to the empirical literature that studies the impact of heterogeneity in ranking behavior on the results from school choice mechanisms (Hastings et al. 2009, Borghans et al. 2015, Glazerman and Dotter 2017, Oosterbeek et al. 2019, Burgess et al. 2015). Related to my findings, Hastings et al. [2009] find that black families in Charlotte trade-off high school performance with a low fraction of same-race peers. The authors show that this trade-off hinders the competitive pressures that are believed to deliver system-wide school improvements under choice. My work highlights the trade-off between distance and performance. I find that this trade-off undermines the equity goal of the policy, and I show it has sizable consequences relative to the effect of preference heterogeneity. My analysis complements evidence from Glazerman and Dotter [2017] by generating estimates of the contribution of several channels to the observed heterogeneity. This literature is under the umbrella of a broader set of papers that study theoretically and empirically the implications of school choice[9] on sorting and stratification (Epple and Romano 1998, Hsieh and Urquiola

9. In these set of papers, school choice is broadly defined to include vouchers and other forms of choice.

2006, Altonji et al. 2015, MacLeod and Urquiola 2015).

Moreover, this paper shows that choice systems alone may not be sufficient to create opportunity for residents of *low-opportunity neighborhoods*. Growing up in these areas has an important impact on adult earnings and education (Chetty and Hendren 2018), and some of these effects are related to access to school quality (Laliberte 2018). This paper shows that guaranteeing access to school quality for these populations requires not only including high-quality choices in their menus but having quality choices close to home. As a result, choice systems would benefit from parallel policies that aimed to reduce residential segregation. An example of such policies includes those proposed by Bergman et al. [2019].

Finally, my analysis adds to a recent series of studies leveraging preference data from centralized school assignments to study school demand (Hastings et al. 2009, Borghans et al. 2015, Abdulkadiroglu et al. 2017, Abdulkadiroğlu et al. 2017, Glazerman and Dotter 2017, Kapor et al. 2018, Agarwal and Somaini 2018, Luflade 2018). Some of these papers study parents' demand under mechanisms that provide incentives to misrepresent preferences, while others study the welfare and fairness associated with assignment rules that give parents incentives to strategize relative to strategy-proof mechanisms. I build on previous work by using data from a strategy-proof mechanism with no restriction on list length, to rationalize differences in assignments across racial and ethnic groups.

My analysis focuses on studying differences in average achievement at the schools assigned to white, black, and Hispanic students. Average achievement is a bundled measure of the academic ability of the students a school enrolls, and of the capacity of a school to generate improvements in student outcomes; that is the effectiveness of a school. In this paper, I am not able to speak of differences in effectiveness as opposed to peer composition, and how gaps in achievement map onto these. Nevertheless, schools that enroll high-achieving peers have been found to be more effective (Abdulkadiroglu et al. 2017). This suggests that higher achievement may be correlated with school effectiveness.

The rest of the paper is organized as follows. Section 1.2 discusses the institutional con-

text and the data restrictions,and it also summarizes the main observed differences in application behavior and assignments across races. Section 1.3 presents reduced form evidence on the mechanisms. Section 1.4 presents the model used to recover demand parameters, discusses the assumptions and analyzes the results. Section 1.5 describes the methodology and assumptions made to run counterfactual exercises and the results. I conclude in Section 1.6.

1.2 Elementary School Choice in Boston

1.2.1 The Assignment Mechanism

Parents of students entering pre-kindergarten in Boston who want to apply for a seat in a school within the district are required to submit a ranking of programs to the school district.[10] Students can rank any number of programs with the condition that they are housed at a school the student is eligible for. Schools typically have at least one general education program and some have programs for English language learners (ELL). Eligibility is determined according to a student's residential location. During the study period, Boston was divided into three zones as shown in Figure 1.2a. Students were eligible for any general education program in their residence zone, plus any within a mile from their home. There are also a handful of city-wide schools that can accept applications from all over the city. I refer to the set of schools a parent can rank as the parents' choice-menu.[11]

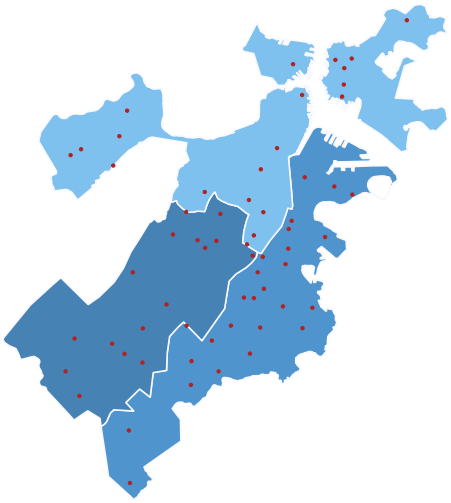
Students are prioritized for admission at each school using a priority structure determined by the school district that is common across schools. Under this priority, students who have a sibling at school j have a higher priority at j than students who do not have a sibling at

10. Students rank pairs of the form school-program

11. Eligibility criteria for applying to English language programs includes not being a native English speaker and score below a threshold in a BPS administered language test. There are also geographic restrictions similar to those of general education programs. Pathak and Shi [2013] discusses how geographic eligibility restrictions may not be binding for language programs. Given this I assume, as Pathak and Shi [2013] do, that English language learner students can apply to any program across the city

Figure 1.1: North, West and East Zones and Choice Menus

(a) Zones



(b) Students with the same Choice Menu

Note: Figure built using data from Boston Public Schools. Red points are schools with a pre-kindergarten program in 2010. Image not shown, results are under review by data providers.

the school. Also, students that live within a mile of a school -usually called the walk-zone- have priority at that school over students that live farther. Specifically, the first priority is given to students that both have a sibling and live in the walk-zone. The second priority is for students who have a sibling, and the third priority is for those that live in the walk-zone. In each group, ties are broken with a random number that is assigned to each applicant and guarantees that priorities generate a strict ordering of students.[12]

The assignment mechanism is a version of Gale and Shapley's (1962) student-proposing DA (Balinski and Sönmez 1999; Abdulkadiroğlu and Sönmez 2003). This algorithm uses the ranks submitted by parents and the described priorities to generate an assignment as follows:

- *Step 1:* Applicants are sorted in priority order in their first ranked schools and students in excess of capacity are rejected. Those who are not rejected are provisionally

12. A more complicated implementation of this priority structure is used in Boston since 1999. Seats at each school are split between those that are assigned using walk-zone considerations and those assigned without. The effects of the precedence of these priorities on the resulting assignment is studied in Dur et al. 2018. Details about the implementation of the priority structure are discussed in Appendix X

admitted.

- *Step k:* For students rejected in step $k-1$, their next preferred option is considered. Each school ranks by priority order the set of provisionally admitted students jointly with those new students who are being considered in k. The program provisionally admits those with the highest priority and rejects students in excess of capacity. The algorithm stops when every rank list has been exhausted or when there are no rejections.

Under the DA, parents do not have incentives to misrepresent their true preferences when submitting rankings (Dubins and Freedman 1981, Roth 1982). Incentives to strategize may generate rankings that are a result of true preferences for schools as well as beliefs about admission chances. Moreover, restrictions to the length of submitted rankings may not generate truthful reports (Haeringer and Klijn 2009, Calsamiglia et al. 2010). The assignment mechanism used in Boston is both strategy-proof and does not restrict students' rank length. This makes Boston a good setting to study parents' preferences for schools.

1.2.2 Data

I use data from BPS that covers the universe of first-round applicants to pre-kindergarten between the years 2010 and 2012. For each applicant I observe the rank-ordered list submitted, the school assigned or an indicator for whether the student was unassigned, and the priority that generated the assignment.[13] I also observe the residential location[14], race and language spoken at home for each applicant, as well as an indicator for whether the applicant

13. A student will be unassigned if he is rejected from every school on his submitted rank list. Students who are unassigned in the first round can reapply in the second round or search for options outside the school district

14. Residential locations are coded by the school district at the geocode level. The geocodes form a partition of the city in 868 polygons of average area of 0.1 sq. miles. The assignment algorithm is built using such geocodes and in consequence that level of aggregation does not represent any loss of information for purposes of the assignment algorithm

is an ELL. [15] First-round applicants represent over 80% of admitted students (Pathak and Shi 2017), the rest applied in the second round and face a more restrictive choice-set.

I use yearly data on school characteristics from the Massachusetts Department of Education (DOE). This includes data on the racial makeup of each school and the fraction of low-income students enrolled in Kindergarten[16]. I also observe the fraction of third grade students in each school scoring *advanced or proficient* in the math and English Massachusetts Comprehensive Assessment System (MCAS) test. Most of the schools that offer a pre-kindergarten program also offer a third grade program. Only a few schools offer up to first grade, for these I do not observe test score data[17].

Using the location of each school and the geocode of residence of each student, I measure the distance to each school as the linear distance between the geocode's centroid and the school. This is the main measure of distance between schools and students. I recreate the algorithm used by BPS to generate walk-zone priority status for each student-school pair: student i is in the walk-zone of school j if a one-mile radius from school j intersects the geocode of residence of i. Similarly, I define the choice-menu of each student using data on the zone in which each school and geocode lies.

Ideally, I would have the sibling priority status of every student at every school. Nevertheless, I only observe the sibling priority status of student i at school j, if i was assigned to j with this priority. Throughout the analysis, I assume that all students that are not assigned with a sibling priority do not have a sibling at any school, and that students assigned with a sibling priority at j do not have a sibling at other schools. A analysis of assignments data shows that this is a good assumption. For most schools I can show that no student with a sibling priority was rejected, and for only a handful I cannot rule out this happened. This means that in the set of schools that each student find acceptable the assumption is for the

15. I remove from my sample students with an invalid geocode that represent around 2% of the sample

16. Income status is measured by the DOE

17. 5 schools in each year offer up to first grade

Table 1.1: Descriptive Statistics: Applicants

[redacted]

Note: I do not observe the income of these families. I approximate income using the average income at the census tract of residence in 2010. I compute the distance between the school assigned and the students' residence as the linear distance between the centroid of the geocode of residence and the location of the assigned school. Table not shown, results are under review by data providers.

most part correct. If students have a sibling priority at multiple schools, I would only be able to account for the priority at the school ranked higher.[18]

The sample has [redacted] applicants to pre-kindergarten between 2010 and 2012. Close to [redacted] of the applicants to pre-kindergarten in Boston are Hispanic, while black and white students are around [redacted] of the sample each.[19] This composition is in contrast with Boston's resident makeup, where white residents account for about half of residents. Choice-menus have on average [redacted] schools, and students rank on average [redacted] options. Black students submit [redacted] lists while white students submit [redacted] lists. Students who are unassigned after running the assignment algorithm may apply in a subsequent round. Since pre-kindergarten attendance is not mandatory, there are applicants who are not assigned to any school and who need to search for options outside of the public school district. About [redacted] of the students that apply in the first round are unassigned, and out of all unassigned students about

18. If the following conditions are satisfied then a school did not reject a student with a sibling priority: First, if there are fewer assigned students than available seats then no student was rejected. Second, if a school accepted students with either the walk-zone or no priority then that school did not reject anyone with a sibling priority. If that were the case then the resulting match would fail to be stable. The number of schools that do not satisfy these in 2010 is 3, in 2011 is 2 and in 2012 is 6. Only for these schools I cannot rule out they rejected a student with a sibling priority.

19. Asian students are around [redacted] of the applicants in my sample. Due to the small sample size, I do not estimate preferences for this population, nor I analyze their assignments.

[redacted] do not enroll in any public school.

Table 1.2: Descriptive Statistics: Schools

	Mean	*St. Dev.*	*Min*	*Max*
Capacity	31.3	16.5	6.0	108.0
Achievement				
% Scoring Advanced-Proficient Math	44.0	19.2	2.0	86.0
% Scoring Advanced-Proficient English	38.4	15.9	10.0	86.0
Demographics				
% Black Students	32.4	19.3	2.1	79.7
% White Students	14.5	14.8	0.0	65.8
% Hispanic Students	43.8	19.0	14.3	90.8
% Low-Income Students in Kindergarten	69.9	18.5	8.3	96.3
Observations				189 (68 distinct schools)

Note: I do not observe achievement data for all schools in all years. There are a total of 17 missing observations (school-year pairs) of schools that do not offer third grade or for which data is restricted due to a small set of test takers.

Between 2010 and 2012, there were a total of 67 public schools that offered a pre-kindergarten program. Not all schools had pre-kindergarten seats in all years. The schools are far from being homogeneous in terms of demographics and achievement. While on average the share of 3rd grade students scoring *advanced or proficient* in math in each school is 44%, the school with the lowest achievement had 2% of students scoring *advanced or proficient* in math while for the highest-performing school the fraction was close to 90%. On average, schools have 32% of black students and 15% of white students. Since both white and black students represent about 20% of all applicants, this means that there are several schools with a high concentration of black students, while there are several schools with a low fraction of white students. Each school has on average 70% of low-income students, and the school with the lowest fraction of low-income students has 8%.

1.3 The Gap in School Achievement and the Possible Explanations

In this section I describe the two main facts that motivate the paper. Then, I discuss the mechanisms that can explain why in a choice setting we do not see a more equitable access to high-performing schools. Finally I give some reduced form evidence on the importance of each mechanism.

1.3.1 The Racial Gap in School Achievement

Between 2010 and 2012, white pre-kindergarteners in Boston were assigned to schools with higher average achievement and a smaller fraction of low-income students and minority students than their black and Hispanic peers. I measure achievement at school j for assignments in the school-year t, as the fraction of 3rd-grade students at j scoring *advanced or proficient* at the math MCAS test in $t-1$. Equivalently, the demographic characteristics of a school will be measured one year prior to the assignments. These are measures of the status of schools that was observable to parents when they apply for admission at pre-kindergarten programs. While white students were assigned to schools where more than [redacted] of students scored *advanced or proficient*, these measures were close to [redacted] for black and Hispanic students (Figure 1.3a). In terms of demographics, white students were assigned to schools with near [redacted] of low-income kindergarten students, and for Hispanic and black students the percentage is closer to [redacted].

Moreover, cross-race differences in school characteristics generated under the choice system are not lower than those generated under a neighborhood assignment. Comparing the distribution of school characteristics generated using parents stated preferences with those of a neighborhood assignment serves as a good benchmark. The latter shows how would these gaps look like if a neighborhood assignment was implemented under the current residential choices in Boston. I generate this alternative assignment running the DA algorithm with the

Figure 1.3: Distribution of School Achievement under School Choice and Neighborhood Assignments

(a) School Choice Assignments (b) Neighborhood Assignments

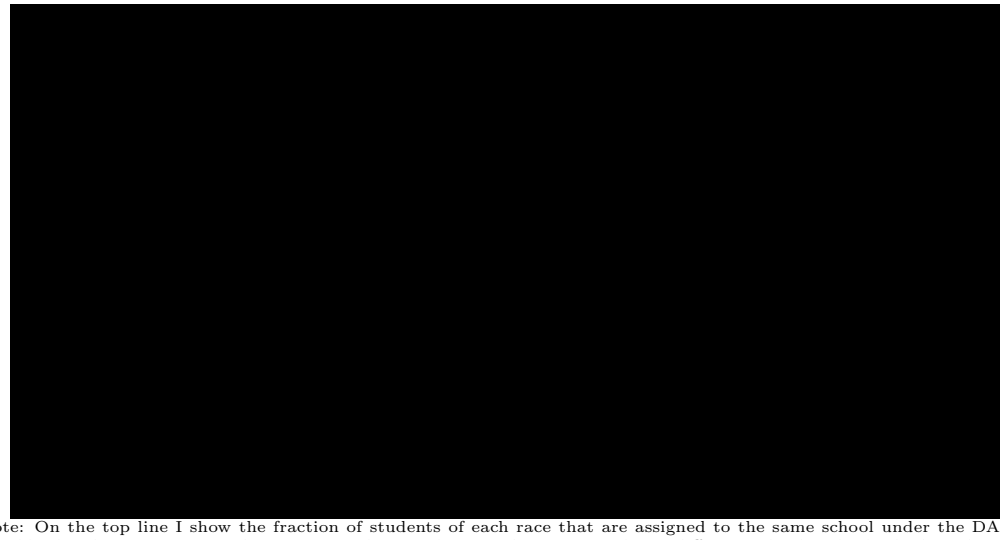
Note: Distribution of school achievement for students assigned to general education programs between 2010 and 2012, and counterfactual distributions built if these students were assigned to the school closest to their homes. Figures are not shown, results are under review by data providers.

set of all students assigned via the choice system, and redefining their preferences and priorities to be determined exclusively by proximity: students prefer schools closer to home, and schools prioritize students that live closer to schools. Under the proposed neighborhood assignment, the distribution of school achievement is similar to that obtained under the choice system (1.3b). [redacted].[20]

The choice and neighborhood assignments are different for around [redacted] of students, who under the choice system sort into schools that have marginally higher achievement but are about [redacted] than their neighborhood schools. These numbers are different across races. Under choice, white students have the [redacted] gain in achievement for every extra mile traveled. While the average black student travels [redacted] more and is assigned to a school with [redacted] higher achievement, the average white student travels [redacted] more and are assigned to school with [redacted] higher achievement (Table 1.3). Also, under choice

20. I cannot reject the null hypothesis that the [redacted]. Two tail p-values are [redacted], for white, black and Hispanic students, respectively.

Table 1.3: Descriptive Statistics: Neighborhood Assignments and DA Assignments



Note: On the top line I show the fraction of students of each race that are assigned to the same school under the DA and Neighborhood assignments. Below I restrict the sample to students assigned to a different school under both algorithms and compare the schools assigned to students under each. Table not shown, results are under review by data providers.

[redacted].

This raises the questions, why giving parents the option to choose does not translate into a more equitable access to high-achieving schools? Why black and Hispanic families are not assigned to schools with higher average achievement under choice-based assignments?

1.3.2 The Mechanisms

Under the choice system, two reasons may explain race differences in access to high-achieving schools. First, differences in parents' school demand that result in different stated preferences for schools or, second, by rules of the algorithm that generate uneven chances of assignment conditional on parents' demand.

Expanding on the former, parents may have a different school demand if their valuation for school attributes is different. For example, a different valuation of school achievement, or any other school characteristic that is correlated with achievement generates differences in demand that can explain the gap. Using similar data, Hastings et al. [2009] find evidence in support of such preference heterogeneity. They argue that black students in Charlotte demand high-achieving schools less often because these schools tend to have a low fraction

of black students. Then, heterogeneity in preferences for the demographic composition of schools results in a different demand for high-achieving schools. On the other hand, parents may have a different school demand if the costs of traveling to high-achieving schools is different across races. A common finding in the literature is that parents perceive the distance between the residence and the school as an important cost (Agarwal and Somaini 2019). If black and Hispanic students tend to live farther from high-achieving schools this may impact their demand for high-achieving schools even if their valuation for other school attributes was equal to that of white parents.

About the latter, choice-menu restrictions and walk-zone priorities are two assignment rules that have the potential to favor access of white students into high-achieving schools. These rules factor into the algorithm the students residential location and as such can generate an uneven probability of assignment into high-achieving schools given that school quality is not uniformly distributed in space. For instance, if choice-menus of black and Hispanic families have fewer seats in high-achieving schools, these students will mechanically face a lower likelihood of being assigned to these schools. Also, if white students live closer to high-achieving schools, a larger fraction of white students will have a walk-zone priority at these schools. In this case, white students would be more likely assigned to high-achieving schools.[21]

In this paper I estimate the contribution of the mechanisms discussed above to the cross-race gap in school achievement. Concretely, I quantify the contribution of three mechanisms. First, differences in demand that stem from differences in preferences for school attributes. Second, differences in demand that stem from differences in distance to high-achieving schools. Third, the contribution of the rules of the algorithm, namely choice-menu restrictions and walk-zone priorities.

21. Dur et al. [2018] show that having a proximity priority under the precedence order used in Boston, does not importantly increase the fraction of walk-zone students admitted relative to an assignment where the proximity priority is abolished

1.3.3 Reduced-form Evidence on the Contribution of the Mechanisms

Before going into the details of the structural model of demand and the counterfactual assignments that will be used to estimate the contribution of each mechanism, it is worth to analyze the raw data to get a sense of the contribution of each mechanism to the gap in school achievement.

An analysis of the submitted rankings and schools in the choice-menu of students reveals that choice-menu restrictions are [redacted] contributor to the gap. [redacted] students [redacted] (Figure 1.4). This means that [redacted].

On the contrary, Figure 1.4 shows that [redacted] to white families and that this may impact assignments via walk-zone priorities and parents' demand. Schools in the walk-zone of white families are [redacted] than those in the walk-zone of black and Hispanic families (Figure 1.4a).[22] This means that white students tend have a [redacted] at high-achieving schools. Moreover, while black and Hispanic families [redacted] (Figure 1.4b and Table 1.5). Having black and Hispanic families [redacted] may impact their demand for these. Figure 1.4a shows evidence consistent with this. The schools ranked first by white families are [redacted] families, and this [redacted].

Cross-race differences in school demand can also be stemming from heterogeneity in parents' valuation of school characteristics. A reduced form analysis of rankings is insufficient to disentangle the contribution of preference heterogeneity and travel costs. In the next section I discuss the structural model of school demand that will be the framework to disentangle between each of these.

22. Related to this Walters [2018] finds suggestive evidence that in Boston charter middle schools tends to locate in lower-achieving areas of the city.

Figure 1.4: School Achievement Geographic Distribution and School Demand

(a) Average School Achievement

(b) Distance to Schools by Race and Achievement

Note: Panel (a) shows the average school achievement weighted by capacity at the schools in the choice-menu and walk-zone of applicants by race. Also, the average school achievement for the schools ranked first and the schools assigned to students by race. In panel (b) I plot the average distance of schools to students of each race, by school achievement deciles. The positive slope for black and Hispanic students show that these families trade-off proximity and achievement. Also, white students live on average closer to schools in the top deciles of achievement. Figures are not shown, results are under review by data providers.

1.4 Estimating Parent Preferences

In this section, I present the model and assumptions used to recover parents' preferences for schools. At the end of the section I discuss the estimated parameters and the fit of the model.

1.4.1 Model and Identification

I model preferences using a random utility model where $i \in \mathcal{I}$ index students and $j \in \mathcal{J}$ index schools. Each student belongs to a group $r \in \{White, Black, Hispanic\}$, and $\mathcal{I}^r$ is the set of students in group r. These categories are exclusive as they are in the data. I refer to r as denoting race, following the convention used by BPS. The indirect utility of student i of race r from attending school j is:

$$U_{ij} = \delta^r_{jt} + \beta^r D_{ij} + X'_{ij}\gamma^r + \epsilon_{ij} \tag{1.1}$$

where δ^r_{jt} summarizes the racial-specific attractiveness of school j in year t that is independent of school location. D_{ij} is the linear distance from i's residence to j, measured in miles. The parameter β^r summarizes parents' preferences for proximity. Notice that each student i is observed only once in my sample, in consequence, each i is associated with a single t and a single r. I do not include these subscripts in all the variables and parameters in 1.1 for simplicity. The matrix X_{ij} includes three indicator variables, that capture within race individual heterogeneity. The first is an indicator whether student i has a sibling at school j, the second for whether i is a language learner and j offers a language program, and the third for whether j has a language program specializing in i's first language. ϵ_{ij} captures idiosyncratic tastes for schools. This is observed by the student but not by the econometrician. I assume ϵ_{ij} is iid. *T1EV*, with a scale parameter σ^r that I allow to vary across races.

Truth-telling. I assume that submitted rankings are truthful. This means that parents

rank all acceptable schools in true preference order. A school is acceptable if it is preferred to the outside option, which is the best substitute parents can access outside the school district. This assumption is motivated by the algorithm's incentive compatibility and the property that there are no restrictions on the number of schools parents can rank. Having restrictions over the length of submitted lists, even under a strategy-proof mechanism, can generate reports that are not truthful (Haeringer and Klijn 2009, Calsamiglia et al. 2010, Luflade 2018). Boston's choice system is one of some that satisfies both properties. This makes it a good setting to estimate and study parents' demand for schools.

Truth-telling can be violated if admission outcomes are largely predictable. Under this setting, parents may misrepresent preferences by not ranking schools that are desirable but are perceived to have low admission probabilities. This is more likely to be a worry in settings where priorities for admission and their distribution across applicants are known before submitting applications, and where historical cutoffs are observable to applicants (Fack et al. 2019).

Consistent with this assumption, if $R_i = (R_{i1}, \cdots, R_{il_i})$ is the rank-ordered list submitted by i and $\mathcal{J}_i$ is the choice-menu of i then,

$$R_{i1} = \underset{j \in \mathcal{J}_i}{\arg\max} \quad u_{ij} \tag{1.2}$$

$$R_{ik} = \underset{j \in \mathcal{J}_i \setminus \{R_{im}: m<k\}}{\arg\max} \quad u_{ij} \tag{1.3}$$

Moreover, if U_{i0} is the utility of the outside option then,

$$U_{ij} > U_{i0} \quad \forall \quad j \in R_i \tag{1.4}$$

$$U_{i0} > U_{ij} \quad \forall \quad j \in \mathcal{J}_i \setminus R_i \tag{1.5}$$

The utility U_{i0} represents the expected utility of the best accessible alternative outside the school district.

Identification. As is common in logit models, the parameters are identified modulo the scale parameter of the idiosyncratic shock, σ^r. Moreover, I normalize the utility of the outside option to zero, and in consequence the school mean-effects are estimated as deviations with respect to the outside option (Train 2009). Specifically, I estimate

$$\left(\frac{\delta^r_{jt} - \delta^r_{0t}}{\sigma^r}, \quad \frac{\beta^r}{\sigma^r}, \quad \frac{\gamma^r}{\sigma^r} \right) \qquad \text{for all } j, r, t \tag{1.6}$$

where δ^r_{0t} is the mean utility of the outside option.

Identification of the distance parameters relies on the assumption that ϵ_{ij} is independent of D_{ij} conditional on school j's fixed effect and X_{ij}. This means that families may sort into neighborhoods according to average tastes for observable and unobservable school characteristics and those in X_{ij}. The assumption will be violated if families sort according to idiosyncratic tastes ϵ_{ij}. In this case, the distance parameter may be biased downward showing that students care for distance more than they really do.

Figure 1.5: Proximity Priority and Ranking Behaviour

Note: Probability of ranking a school first as a function of the distance to the boundary of the proximity priority. Figure not shown, results are under review by data providers.

To asses the validity of this assumption, I use the geographic discontinuities generated by the assignment mechanism to study whether sorting is a concern, and I find little evidence in support of this. The walk-zone priority favors access to students into schools that are closer than a mile from their homes. Then, if parents are sorting near their preferred schools it is optimal to sort to the left of the one-mile boundary than to the right. In figure 1.12 I plot the probability that a student ranks a school first as the distance to the proximity boundary of that school changes. The zero in the x-axis represents the one-mile proximity threshold, that is, at zero a student lives at exactly one mile from the school in question. To the left of zero, students live closer than the mile. The downward trend responds to parents' valuation of proximity. The fact that there is no clear discontinuity at the proximity boundary is evidence in support of a lack of sorting in these boundaries, which suggests sorting is not common for this population across the city.

Two distinct sources of variation identify school mean utilities and preferences for proximity. Rankings of students who are equidistant to any two pairs of schools generate the variation used to identify school mean utilities. Students ranking schools farther over schools closer is the variation used to identify the preferences for proximity.

Estimation. I estimate utility parameters by maximum likelihood, using all the first-round rankings submitted to BPS between 2010 and 2012. Details about the likelihood function are shown in Appendix 1.B.

1.4.2 Parameter Estimates

Tables 1.4 and 1.6 show the estimated parameters for all races. Negative signs for the mean of the distance parameters show that parents value proximity, and the low standard deviations evidence low heterogeneity within race. The magnitude of the mean parameters is similar to the estimation done by Pathak and Shi 2013, who carry out a similar analysis for a sample

that overlaps mine.[23] [redacted]
and again the within race heterogeneity is low for these preference. Finally, parents value
[redacted] student parents.

Table 1.4: Preference Parameters: Distance, Sibling and Language Programs

[redacted]

Note: Standard errors computed with the inverse of the Hessian in parenthesis. Table not shown, results are under review by data providers.

School mean utilities summarize the overall attractiveness of a school after discounting the effect of distance. These parameters cannot be compared across races, but the order generated does provide a way to assess commonalities in the valuation of schools across races. Figure 1.6 shows a positive correlation between the coefficients of white parents and parents of other groups. This suggests, there are underlying school characteristics that all groups value. Hispanic families parameters have a higher correlation with white families parameters than black families do.

Parents of all races value schools that are [redacted], that have a higher fraction of [redacted] and a lower fraction of [redacted] (Table 1.7). Nevertheless, demographic characteristics contribute more to explain the variation in school mean utilities than achievement does (Table 1.8).

Preference parameters cannot be directly compared across races. Doing so requires assuming some relation between the scale parameters σ_r. A way to asses how preferences for

23. The authors do not carry-out a race-specific model estimation. A weighted average of the race-specific parameters is similar to the values they obtain.

Figure 1.6: Correlation of School Mean Utilities δ_j^r

Note: Scatter plot of school mean effects of black and Hispanic students with white students' school effects. The correlation between Hispanic and white students' parameters is 0.7, between black and white students parameters is 0.5 and between Hispanic and black is 0.7. Figure not shown, results are under review by data providers.

proximity compare across groups is to use the parameters of the model to simulate rankings, and evaluate how rankings change when the distance to a school is marginally increased. The number of positions a school looses after this change is a comparable metric across students of different races. Results from this simulation are determined not only by the estimated preferences for distance, but also by the distribution of school mean utilities and the geographical location of schools.

I generate a series of rankings using the demand model and compare the resulting rankings with other generated after increasing $D_{i\bar{j}}$ for a fix $\bar{j}$ and all i by 0.1 miles. Table 1.9 shows the average number of positions a school would gain after running the exercise for every school in the sample. I find [redacted] with respect to distance. On average, a school is ranked [redacted] after the proposed distance

change.

Fit. To evaluate the fit of the model I use the estimated parameters to generate rankings and subsequent assignments, and I compare these to the assignments generated running the DA using the rankings parents submitted to BPS. Using the demand model I generate a ranking for every student assuming that families rank only acceptable schools, that is, only rank schools that are preferred to the outside option. I use these simulated rankings to run the DA algorithm, and I compare the resulting assignment to the assignment obtained with the rankings submitted by families to BPS. I find that the parameters closely predict the distribution of school achievement for the assignments of white, black and Hispanic students in the year 2011. Also, the model fits well the distribution of distance to the assigned school for these groups (Figure 1.14). The average length of the simulated rankings coincides with the average length of the rankings submitted by parents to BPS.[24]

1.5 Counterfactual Assignments

In this section, I describe how and under what assumptions the counterfactual assignments are generated, and then discuss the results. The counterfactual assignments are produced by either generating alternative rankings using the parameters of the demand model, or by changing the rules of the algorithm while holding rankings fixed. The counterfactuals will be used to estimate the contribution of the mechanisms described in Section 1.3. As such, these are not intended to be thought of as policy counterfactuals, but instead as a tool to carry out an accounting exercise.

I consider three types of counterfactuals. In the first, I study how the assignment of minority students depends on their residential location. To study this, I generate assignments where new residential locations, randomly drawn from white students residences, are assigned to minority students. In the second, I hold minority students' residential locations but instead change their preference parameters to be those of white students. The former captures

24. Parents ranks on average 5.4 schools. Simulated rankings have an average length of 4.6.

the impact of the local school supply and parents' preference for proximity, plus the effect of location specific assignment rules. The latter captures the impact of heterogeneity in preferences for school characteristics. Finally, I run counterfactual assignments where I eliminate assignment rules that are location-specific. By doing this, I disentangle the effect of preferences for proximity from location-specific rules.

1.5.1 Changing the location of a student

Even in a choice-based system where the link between residential choices and school quality is weakened, the residential location of families may play a crucial role in their school assignment. If parents value proximity, the benefit of attending a high-achieving school may be upset by high travel costs. Moreover, assignment rules that constrain geographically the choices of families, or that prioritize students based on proximity to schools, could generate geographic inequities even in the absence of travel costs.

Studying how location effectively matters in choice-based settings is a first order concern to evaluate the equity effectiveness of choice policies. To estimate how much of the cross-race gap in school achievement can be attributed to the location of students in space, I evaluate how would the submitted rankings and subsequent assignments change as the residential location of students change. Concretely, I study how do assignments of minority students change if their residential location was randomly drawn from the set of white students' locations.

After drawing a new residential location for a single minority student, I use the demand model to generate the ranking that the student would have submitted at that new location. Demand parameters do not change, nevertheless the change in distance to all schools will shift travel costs. Also, choice-menu restrictions may limit and/or expand parents available choices.

To make sure that schools do not change as a result of the residential location change, I will consider the relocation of one student at a time. If I relocated several students simulta-

neously school characteristics may change, for instance, school demographics. Changing the residential location of a single student guarantees that schools are unchanged and then preference for schools should be the same. This means that under this counterfactual I estimate a partial equilibrium result, concretely, the average impact of relocating one single minority student.

I run the counterfactual taking all the students that applied for a seat in 2011, and all the schools open for admissions in that year. To build counterfactual locations, I randomly pair minority students and white students. In each counterfactual, the minority student will take the white students' residential location, choice-menu and, walk-zone and sibling priorities. After generating assignments for all minority students both at their original location and their counterfactual locations, I generate the distribution of school achievement for white and minority students at their original locations, and for minority students at their counterfactual locations.[25]

Figure 1.7: Change location of a student: Achievement at the Assigned School

(a) Change location of a black student (b) Change location of a Hispanic student

Note: Distribution of achievement in schools assigned to black and Hispanic students under a counterfactual assignment where they are randomly assigned to a new residence drawn from the distribution of whites' residences. This is compared to the distribution for black and white students in their original location. Figures are not shown, results are under review by data providers.

The proposed change in location [redacted] school achieve-

25. A more detailed description of the process used can be found on Section 1.A.5.

ment for [redacted]. Figure 1.7 shows the distribution of school achievement for the schools assigned to white, black and Hispanic students in their original residential locations, and for black and Hispanic students in their counterfactual locations. Mean school achievement increased [redacted]. Under the proposed change, the gap between black and white students [redacted] and the gap between Hispanic and white students [redacted].

1.5.2 Changing preference parameters

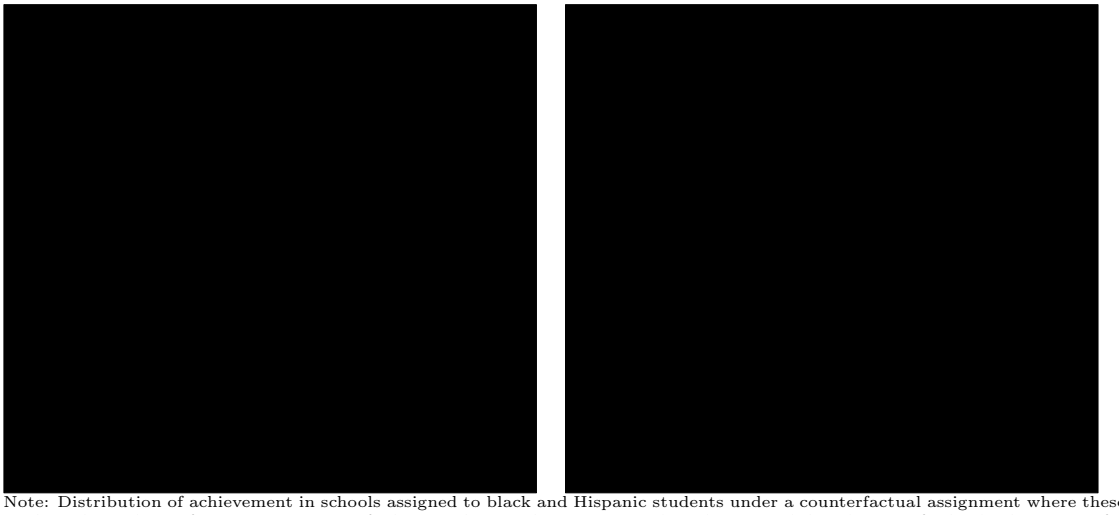
Differences in school assignments may be responding to differences in parents' preferences for school attributes. The gaps discussed can be generated if parents of different races place a different value in school achievement. Nevertheless, even if their valuation for achievement is the same, the gaps may be generated by a different valuation of any school characteristic that is correlated with achievement. For instance, Hastings et al. 2009 finds that black families in Charlotte trade-off schools with higher-achievement for schools with an ideal fraction of same-race peers.

To study the contribution of preference heterogeneity on the gap in school achievement, I evaluate how would the submitted rankings and subsequent assignments of minority students change if their preference parameters where those of white parents. In this counterfactual, the residential location of every student is unchanged and, for consistency with the previous counterfactual, I change the preference parameters of one student at a time. The counterfactual ranking describes how would a minority student rank schools in their original residential location if their preferences were those of white parents.

As before, I run the counterfactual taking all the students that applied for a seat in 2011, and all the schools open for admissions in that year. After changing the preference parameters of one minority student I generate new rankings and assignments for all students. Later I pool the assignment of all students who's preference parameters were changed and generate the distributions of school achievement for white and minority students using their original

Figure 1.8: Change preferences of a student: Achievement at the Assigned School

(a) Change preferences of a black student (b) Change preferences of a Hispanic student



Note: Distribution of achievement in schools assigned to black and Hispanic students under a counterfactual assignment where these students have the preference parameters of white students. This is compared to the original distribution of school achievement for black, Hispanic and white students. Figures are not shown, results are under review by data providers.

preference parameters, and for minority students with their counterfactual demand.[26]

Under the proposed change in preferences black and Hispanic students are assigned to [redacted]. Figure 1.8 shows the distributions of school achievement of minority students under the original and counterfactual preference parameters, and the distribution for white students. Mean school achievement [redacted] for black families and [redacted]. The gap [redacted].

1.5.3 Eliminate Choice Menus and Walk-zone Priorities

The effect of location on school assignments is sizeable. When a student changes locations not only his travel costs change, but also, that students' menu of schools will change, as well as the set of schools where the student has a walk-zone priority.

To disentangle the effect of the last two from that of travel costs, I run two additional counterfactual assignments. In the first, I eliminate choice-menu restrictions and allow parents to rank schools from across the city. Under these rules, a minority family can rank the

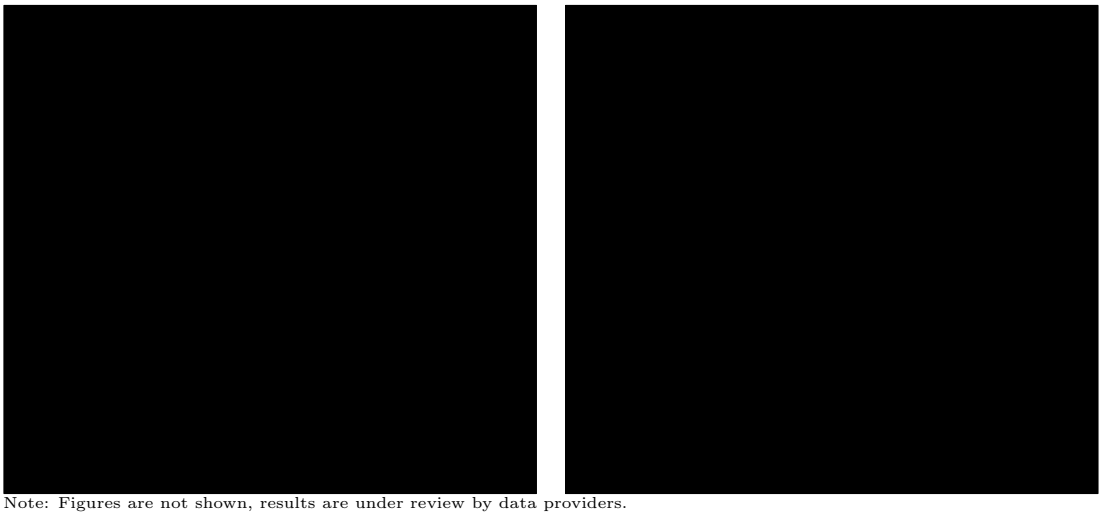
26. A more detailed description of the process used can be found on Section 1.A.5.

same schools that they would have ranked under the location change counterfactual. Then, the only reason why these rankings wouldn't coincide is explained by differences in travel costs. In the second, I eliminate walk-zone priorities and run the DA algorithm assuming no one has this priority. Eliminating priorities won't change parents' submitted rankings, but change assignments via priorities.[27]

When limits to choice-menus and walk-zone priorities are eliminated, expected school achievement does not change. Figure 1.9 shows the distribution and average school achievement for black students after eliminating choice-menu restrictions on the left, and the walk-zone priority on the right. After running the counterfactual [redacted].[28]

Figure 1.9: Eliminate location-specific rules: Achievement at the Assigned School

(a) Choice-Menu (b) Walk-Zones



Note: Figures are not shown, results are under review by data providers.

This results imply that neither the design of the matching algorithm, nor the implementation of it contribute to the cross-race gap in school achievement. In consequence, changing these rules won't have any impact on the gaps. Moreover, this result suggests that the sizable

27. A more detailed description of the process used can be found on Section 1.A.5.

28. I reject the null-hypothesis that [redacted].

effects of location are explained by differences in distance to high-achieving schools.

1.5.4 Summary

Figure summarize the reductions in the gap in achievement at the schools assigned to black, Hispanic, and white students. White students are assigned to schools that have [redacted].

Figure 1.10: Decomposition of the gap

Note: Figure not shown, results are under review by data providers.

1.5.5 Change in the School-Match After a Location Change

Using the model parameters I can assess whether black and Hispanic students are assigned to schools with higher value after a location change. To do this I compare, for each treated student, the location-independent value of the school assigned under the original setting and the counterfactual. Let $\mu(i) \in \mathcal{J}$ be the school assigned to i under the original setting and $\tilde{\mu}(i) \in \mathcal{J}$ be the school assigned to i under the counterfactual. If N^r is the number of students of race r, then the following is the average change in school value for students of race r expressed in miles

$$\frac{1}{N^r} \sum_{i \in \mathcal{I}^r} \frac{\delta^r_{\tilde{\mu}(i)} - \delta^r_{\mu(i)}}{\beta^r}.$$

I find that after a location change, black and Hispanic students are matched to schools [redacted]. The average change in school value for black and Hispanic students is equivalent to [redacted]. After the change in location, black students are assigned to schools that are on average [redacted]. Hispanic students, on the other hand, are assigned to schools that are [redacted] from home relative to the original assignments.

1.6 Conclusion

Choice-based assignments are designed to increase equity and foster diversity by offering students the option to sort into their preferred schools, and by weakening the link between residences and schools. [redacted]. I find that the main contributor to this gap are cross-race differences in distance to high-achieving schools. Higher travel costs to high-achieving schools in the form of longer commutes upset the perceived benefits of high-achieving schools for black and Hispanic families. Differences in parental valuation for school attributes accounts for a smaller share of the gap. Importantly, the de-

sign and implementation of the assignment algorithm does not explain any of the differences found.

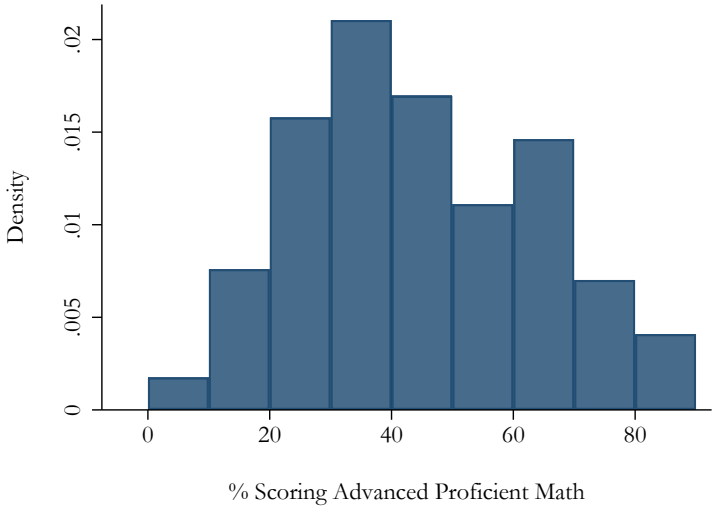
The salience of travel costs under choice-based assignments shows a first-order reason of why neighborhoods matter. It also implies that geography plays a crucial role in an equal provision of public goods, and more concretely on the effectiveness of educational policy. This suggests that we should not only evaluate educational systems by their overall measures of quality, but also by the distribution of quality in space.

APPENDICES

1.A Supplementary Tables and Figures

1.A.1 Descriptive Statistics

Figure 1.11: Histogram of School Achievement



Note: Histogram of school achievement measured as the fraction of 3rd grade students scoring advanced of proficient in the math MCAS tests.

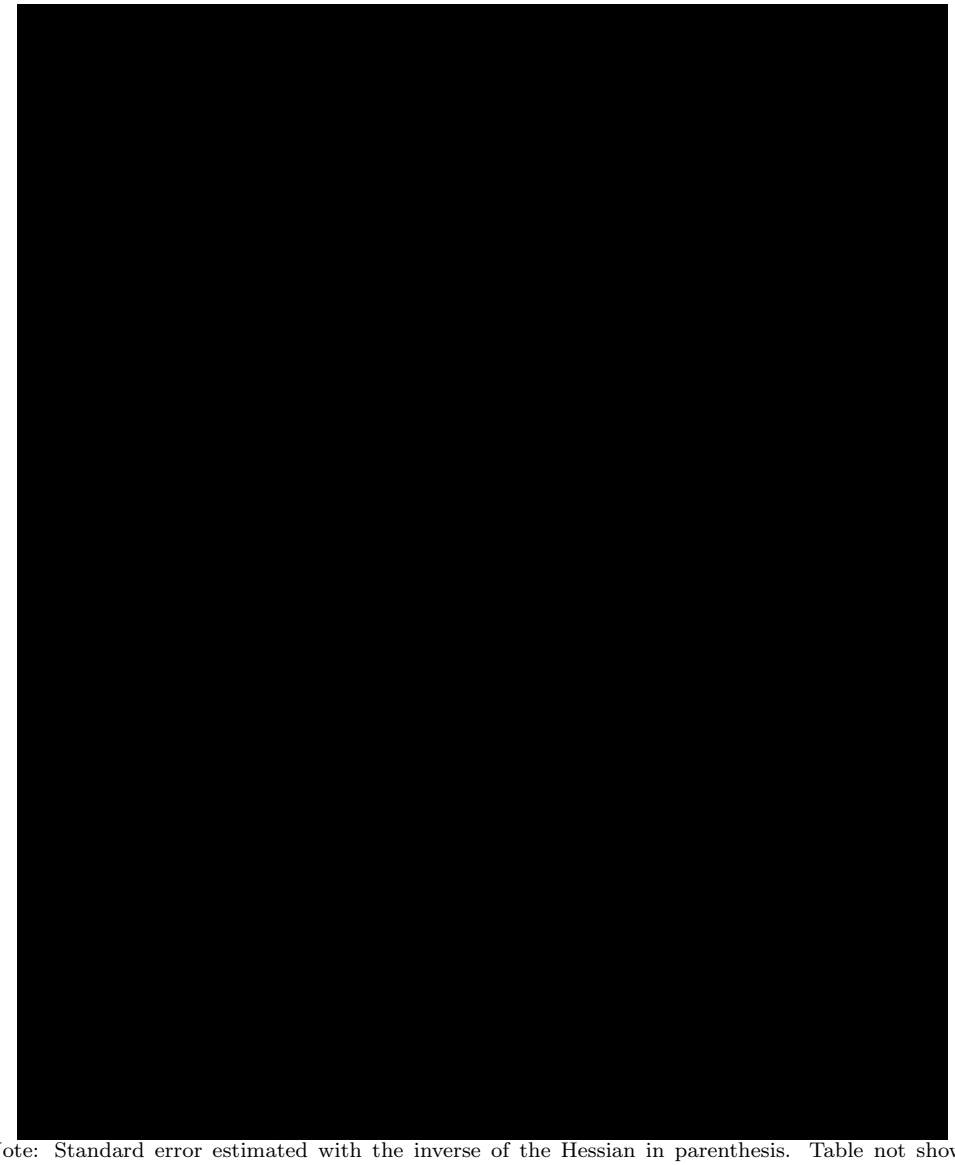
Table 1.5: Relation Between Distance to Schools and School Achievement

[redacted]

Note: Each column shows a regression between school achievement and distance. Each observation is a pair student-school for schools in the choice-menu of every student. Standard errors in parenthesis. Table not shown, results are under review by data providers.

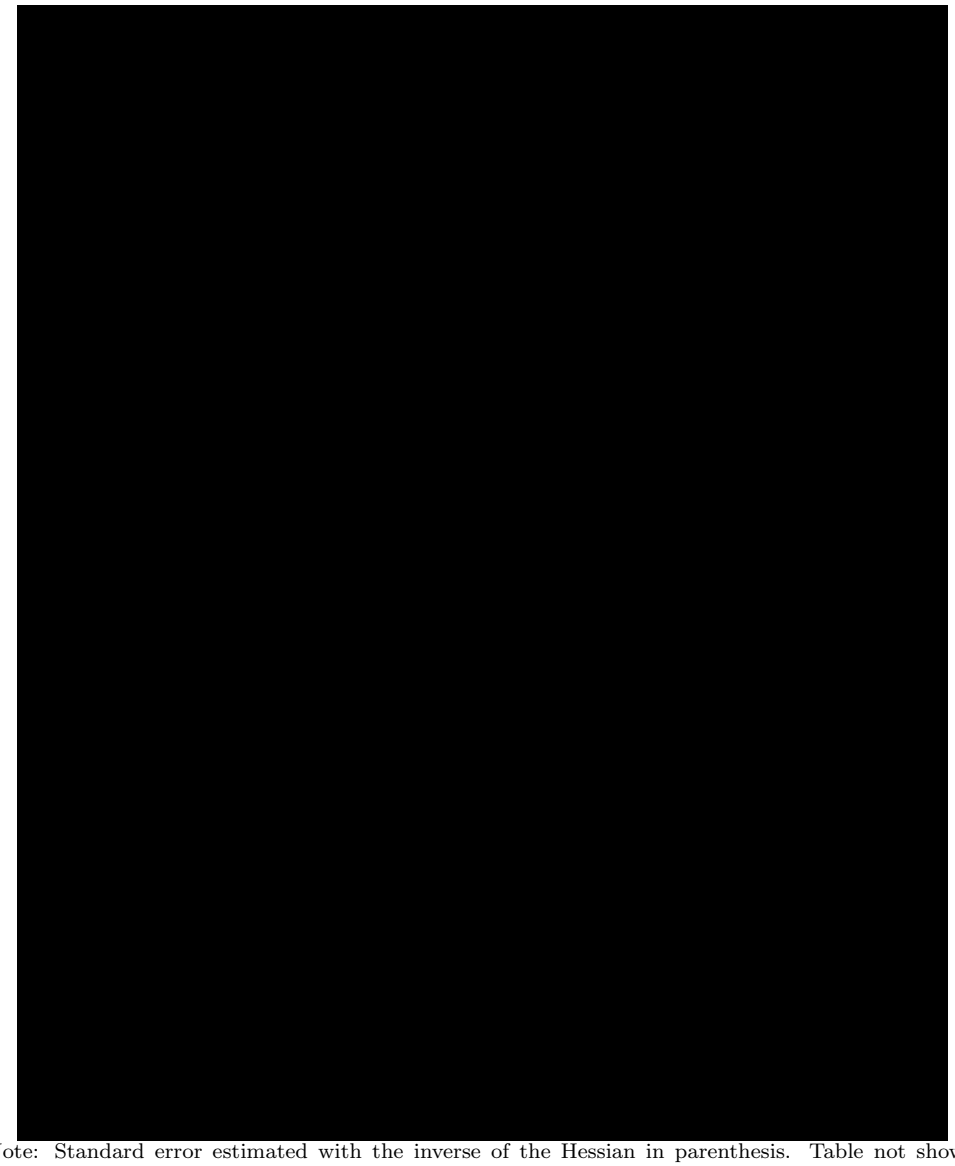
1.A.2 Preference Estimates

Table 1.6: School Mean Utilities



Note: Standard error estimated with the inverse of the Hessian in parenthesis. Table not shown, results are under review by data providers.

Table 1.6: School Mean Utilities Continued

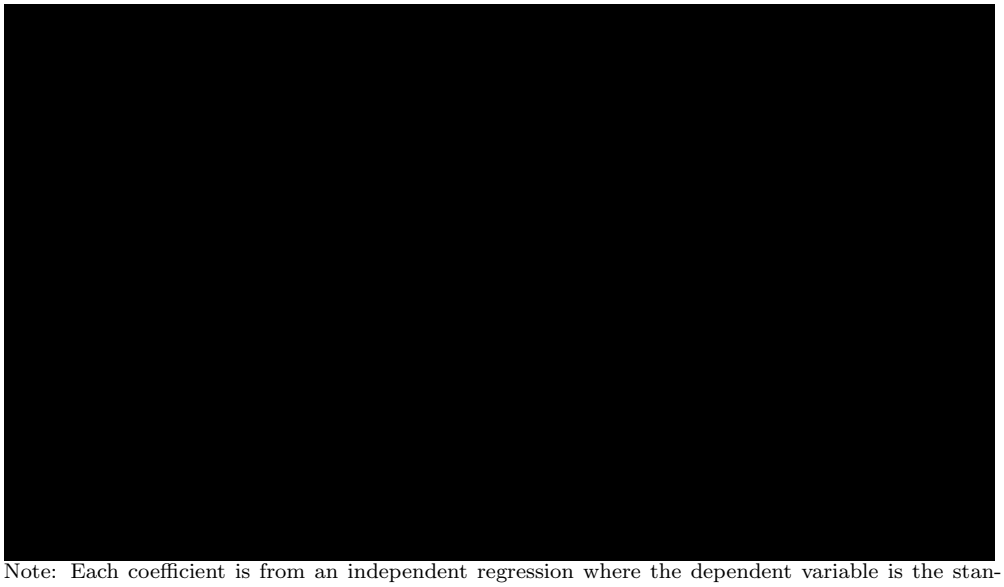


Note: Standard error estimated with the inverse of the Hessian in parenthesis. Table not shown, results are under review by data providers.

Table 1.6: School Mean Utilities Continued

Note: Standard error estimated with the inverse of the Hessian in parenthesis. Table not shown, results are under review by data providers.

Table 1.7: School Mean Utilities and School Characteristics - Individual Regressions



Note: Each coefficient is from an independent regression where the dependent variable is the standardized δ_j^r. Standard errors in parenthesis. Table not shown, results are under review by data providers.

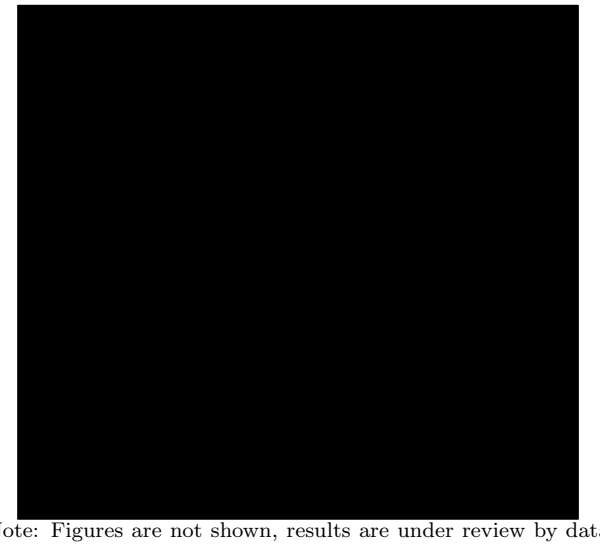
Table 1.8: School Mean Utilities and School Characteristics - Pooled Regressions

Note: Coefficient from regression between the standardized δ_j^r and school characteristics. Standard errors in parenthesis. Table not shown, results are under review by data providers.

Figure 1.12: Location of 2011 Schools by Deciles of Mean Utility

(a) Black Students (b) Hispanic Students

(c) White Students



Note: Figures are not shown, results are under review by data providers.

Figure 1.13: Distance and School Mean Utility

Note: Average distance between students of each race and schools by deciles of school mean utility δ^r_{jt}. Figure not shown, results are under review by data providers.

Table 1.9: Simulated change in positions after an extra 0.1 miles

Note: Average number of positions gained by a school after an increase in travelled distance of 0.1 miles. Simulations generated using the estimated preference parameters and random realizations of ϵ. Table not shown, results are under review by data providers.

1.A.3 Model Fit

Figure 1.14: Fit of Estimated Preference Parameters: Achievement and Distance to Assigned School

(a) White

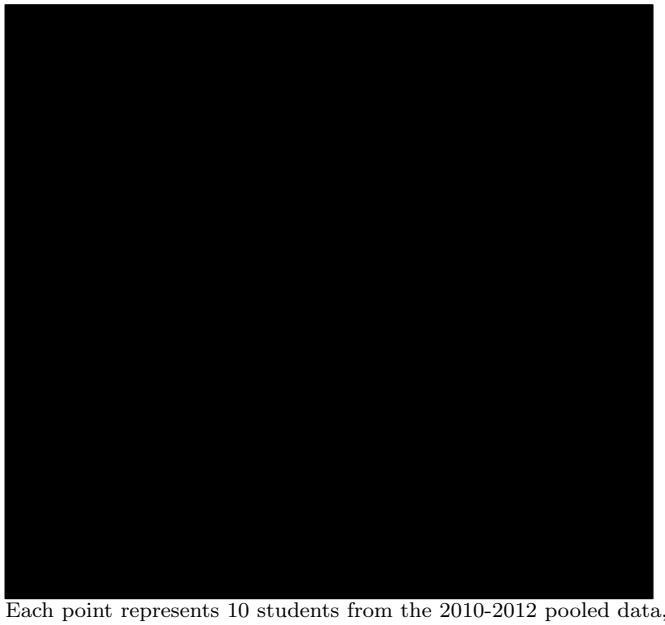
(b) Black

(c) Hispanic

Note: Submitted rankings distributions are obtained from running the DA using the rankings submitted by parents to BPS. Simulated rankings distributions are obtained from rankings generated using demand parameters and 100 random realizations of ϵ. I plot the piece-wise median density, and the 5% and 95% densities. Figures are not shown, results are under review by data providers.

1.A.4 Distribution of Students in Space

Figure 1.15: Spatial Distribution of Applicants by Race



Note: Each point represents 10 students from the 2010-2012 pooled data, randomly located at the census tract level. Figure not shown, results are under review by data providers.

Figure 1.16: Distribution of applicants by distance to schools

Note: Distribution of number of students from each race that applied to each school. Figure not shown, results are under review by data providers.

1.A.5 Counterfactual Assignments

Residential Location Change

1. Generate assignments under the original setting:

   (a) Take all applicants and schools from 2011. Using the parameters $(\tilde{\delta_{jt}^r}, \tilde{\beta^r}, \tilde{\boldsymbol{\gamma}^r})$, and a realization[29] of $\epsilon = (\epsilon_{01}, \cdots, \epsilon_{0\mathcal{J}}, \epsilon_{11}, \cdots, \epsilon_{\mathcal{I}\mathcal{J}})$ generate rank-order lists, R_i for all i. The length of the submitted list is determined by the position of the outside option in the ranking: only schools preferred to the outside option are ranked

   (b) Using the ranking profile $R = (R_1, \cdots, R_{\mathcal{I}})$, generate an assignment running the DA

   (c) Repeat for m realizations of ϵ

2. Generate counterfactual assignments:

   (a) Generate random locations by pairing each black or Hispanic student, i_b, in the 2011 sample, with k white students, i_w, in that year's sample[30]. Each pair (i_b, i_w) represents a location change for i_b. In the counterfactual, i_b will take i_w's choice-menu, walk-zone priorities, sibling priority, and distance to schools will be updated accordingly

   (b) Consider one pair (i_b, i_w). Under the counterfactual, all students will keep their location except for i_b

   (c) Consider only the same pair (i_b, i_w). For each realization of ϵ used in 1. generate rank-order lists assuming that each list has the same length of the ranking originally submitted to BPS. Notice that for each realization of ϵ, the lists of all untreated students under the original and counterfactual will be equal

29. Each ϵ_{ij} is drawn independently from a *T1EV* distribution with scale parameter 1.

30. This is done generating random draws of white students with replacement

(d) For each profile R generate an assignment running the DA

(e) Repeat for all pairs (i_b, i_w)

Preference Change

1. Generate assignments under the original setting:

   (a) Take all applicants and schools from 2011. Using the parameters $(\tilde{\delta_{jt}^r}, \tilde{\beta^r}, \tilde{\boldsymbol{\gamma}^r})$, and a realization[31] of $\epsilon = (\epsilon_{01}, \cdots, \epsilon_{0\mathcal{J}}, \epsilon_{11}, \cdots, \epsilon_{\mathcal{I}\mathcal{J}})$ generate rank-order lists, R_i for all i. The length of the submitted list is determined by the position of the outside option in the ranking: only schools preferred to the outside option are ranked

   (b) Using the ranking profile $R = (R_1, \cdots, R_{\mathcal{I}})$, generate an assignment running the DA

   (c) Repeat for m realizations of ϵ

2. Generate counterfactual assignments:

   (a) Take a student $i_b \in \mathcal{I}^b$. This will be the treated student

   (b) Replace the values of the parameters $(\tilde{\delta_{jt}^b}, \tilde{\beta^b}, \tilde{\boldsymbol{\gamma}^b})$ for $(\tilde{\delta_{jt}^w}, \tilde{\beta^w}, \tilde{\boldsymbol{\gamma}^w})$ only for i_b

   (c) For each realization of ϵ used in 1. generate rank-order lists for all students assuming that each list has the same length of the ranking originally submitted to BPS. Notice that each realization of ϵ, the lists of all untreated students under the original and counterfactual will be equal

   (d) For each profile R generate an assignment running the DA

   (e) Repeat for all $i_b \in \mathcal{I}^b$ and $i_h \in \mathcal{I}^h$

31. Each ϵ_{ij} is drawn independently from a *Gumbel* distribution with scale parameter 1.

Change in location-specific rules

1. Generate assignments under the original setting:

   (a) Take all applicants and schools from 2011. Using the parameters $(\tilde{\delta_{jt}^r}, \tilde{\beta}^r, \tilde{\boldsymbol{\gamma}}^r)$, and a realization[32] of $\epsilon = (\epsilon_{01}, \cdots, \epsilon_{0\mathcal{J}}, \epsilon_{11}, \cdots, \epsilon_{\mathcal{I}\mathcal{J}})$ generate rank-order lists, R_i for all i. The length of the submitted list is determined by the position of the outside option in the ranking: only schools preferred to the outside option are ranked

   (b) Using the ranking profile $R = (R_1, \cdots, R_{\mathcal{I}})$, generate an assignment running the DA

   (c) Repeat for m realizations of ϵ

2. Generate counterfactual assignments:

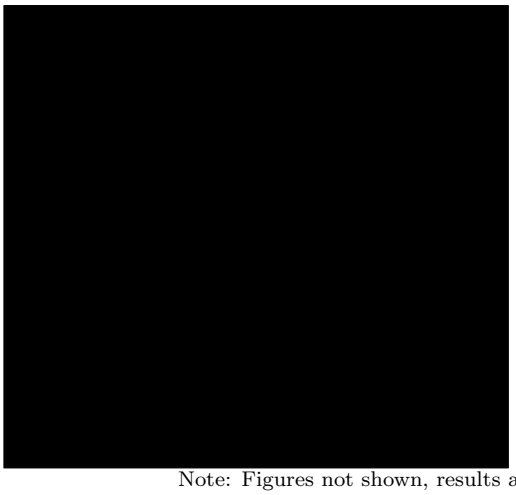
   (a) For each realization of ϵ in 1. generate rank-order lists for all students, assuming that there are no restrictions to choice-menus and that each list has the same length as the ranking originally submitted to BPS

   (b) For each profile R generate an assignment running the DA

32. Each ϵ_{ij} is drawn independently from a *Gumbel* distribution with scale parameter 1.

Figure 1.17: Change in Assignment Rules: School Achievement for Hispanic Students

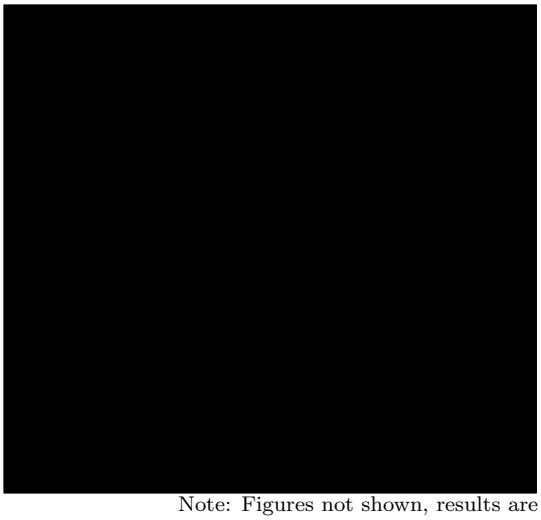
(a) Choice-Menu



(b) Walk-Zones

Note: Figures not shown, results are under review by data providers.

Figure 1.18: Location Change: Distance to Assigned School



Note: Figures not shown, results are under review by data providers.

Figure 1.19: Preference Change: Distance to Assigned School

Note: Figures not shown, results are under review by data providers.

1.B Maximum Likelihood Function

Let $R_i = (R_{i1}, \cdots, R_{il_i})$ be teh rank-order list submitted by i. The likelihood of R_i is

$$\mathcal{L}(R_i) = \left[\prod_{k=1}^{l_i} \frac{\exp(U_{iR_{ik}})}{1 + \sum_{j \in \mathcal{J}_i \setminus \{R_{im}: m<k\}} \exp(U_{iR_{ij}})} \right] \left[\frac{1}{\sum_{j \in \mathcal{J}_i \setminus \{R_{im}: m<l_i\}} \exp(U_{iR_{ij}})} \right] \quad (1.7)$$

I find the values of δ^r, β^r, γ^r that maximize

$$\Pr(R_1, \cdots, R_{\mathcal{I}}) = \prod_{i \in \mathcal{I}} \mathcal{L}(R_i) \quad (1.8)$$

CHAPTER 2

CIVILIAN COLLABORATION, PRICE SHOCKS AND VIOLENCE IN CIVIL WARS

with Austin Wright[1]

2.1 Introduction

The impact of price shocks on violence in civil conflicts has raised the interest of scholars in the social sciences for the past decade. As the quality of data improves, studies have moved from cross-country comparisons to intra-country studies that focus on smaller units of analysis: from departments in Colombia, which have an average size of 34.600 square km (Angrist and Kugler 2008) to 0.5×0.5 degrees latitude and longitude units in sub-Saharan Africa (Berman and Couttenier 2015). Despite the fact that data quality and, with this, identification strategies have improved, the literature is still vague in identifying the channels that drive changes in violence during civil wars as economic conditions change. In this paper we propose a mechanism that rationalizes the drivers of violence after economic shocks. This rationale, different from other mechanisms proposed by the literature, is consistent with a theory of violence where political groups contest for the control of a territory (Kalyvas 2006).

We argue that increases in the price of commodities that are produced with a technology that relies on collaboration across producers reduces the incentives of civilians to inform against fellow community members. This is the case since denouncing a fellow civilian will lower the gains from collaboration in production. Since informants feed armed groups with information necessary to effectively carry out selective attacks, a reduction in the information available to insurgent groups will cause a reduction in selective violence.

To test the mechanism we use data on the universe of violent attacks that occurred in Colombia between 2001 and 2016, and study changes in violence in municipalities that

1. We thank Elena Boadillo and Carla Solis for excellent research assistance

are producers of coffee as the prices of this commodity exogenously change. The proposed mechanism predicts that increases in the price of coffee should be associated with reductions in selective violence. To test the prediction, we use records of accounts by witnesses of each violent attack in our sample, and using these descriptions classify attacks as selective or indiscriminate using a text analysis algorithm. Later we test the causal impact of shocks to the price of coffee on the incidence of selective attacks.

On Section 2.2 we will talk about related literature and the conceptual framework of the theory of violence. On Section 2.3 we describe the mechanism that we propose to explain the connection between shocks to the price of coffee and the incidence of violence. On Section 2.4 we describe the dataset, the text classification algorithms used and the results and the empirical strategy. On Section 2.5 we will present the main results. Finally, on Section 2.6 we will conclude.

2.2 Related Literature

This paper contributes to the literature that studies the links between economic performance and civil conflict, specifically, the literature that tries to address the causal impact of shocks to the economy over violence in ongoing civil conflicts. It is not intended to shed light on the way economic shocks change the likelihood of a conflict starting or ending. It builds on the work of Dube and Vargas [2013] and Felter et al. [2016] who study the impact of commodity price shocks on violence in Colombia and Philippines, respectively.

Dube and Vargas [2013] argue that positive shocks to the price of commodities produced with a labor intensive technology will have a negative impact on violence since the shock will increase the opportunity cost of engaging in violent activities. On the other side, a positive shock to the price of commodities produced using a capital intensive technology will come with increases in violence due to an increase in the gains from appropriation. The authors test the mechanisms using exogenous changes in the price of coffee and oil in Colombia. Felter et al. [2016] propose a different rationale, they find that positive shocks to the price

of exporting bananas produced in the Philippines, usually by big international firms, come with reductions in violence, while on average there is no significant impact on violence after a shock to the price of bananas produced by local farmers for national consumption. Arguably, these are both produced in a labor intensive manner. The authors explain that in a more centralized business, as that of exporting bananas, the cost of extortion is lower and the return can be big, whereas, extortion to several small farmers might bring low returns at a high cost. Their story points at the rapacity effect being more profitable in large centralized markets, and that being the driver of increases in violence.

In addition, Felter et al. [2016] study heterogenous effects of the shock across different levels of control of insurgencies. This paper is to our knowledge the first that has a strong measure of control by political actors, which is known to be difficult to reliably measure. The authors find that for all kind of bananas, positive shocks to their price are associated with reductions in violence in those places where a group has high control. This is a finding that suggests that studying the way in which control shapes incentives of insurgents and civilians is important in understanding what drives changes in violence.

Both of the mechanisms studied by these papers are somewhat agnostic about the logic of violence in civil wars, neither picture insurgent groups as rational players or consider the possible benefits and costs of violence. In explaining how price shocks affect violence we are going to take a step back and take a stand on the rationale of violence, and propose a mechanism consistent with this general framework. We will take Kalyvas [2006] theory of violence and, in that sense, this paper contributes also to the empirical theory of violence in civil wars.

2.2.1 A Theory of Violence

Kalyvas [2006] states that violence is the main tool used by political actors to secure control of a territory. Control is defined as the exclusive collaboration of civilians. Civilians decide which group to support, if any, and decide whether to share information about fellow civilian

activities, specifically, information on collaboration with the adversary. This is what Kalyvas calls *denunciation*. Having more control will then be associated with having more information and authority over civilian activities.

In a race for control, political actors will use selective and indiscriminate violence to gain the exclusive collaboration of civilians. Selective violence targets defectors: those that support the adversary, but relies on gathering enough information about enemy support. Indiscriminate violence does not rely on having information on the victim's activities, and may randomly victimize civilians.

Three predictions from Kalyvas' theory are worth pointing out: first, the higher the level of an actor's control, the less likely he will use violence, selective or indiscriminate. Second, the lower the level of an actor's control, the less likely he will use selective violence; then, under dispute for control, selective violence will be primarily used by the agent with more control. Finally, under parity of control we are not likely to observe selective violence, since neither of the groups is strong enough to get information and secure protection for its informants.

We propose a theory of shocks to commodities and changes in violence that is consistent with Kalyvas' rationalization of violence. We argue that shocks to the price of a commodity that is produced with a technology where there are high gains from collaboration will reduce the incentives of civilians to denounce. This reduction in the information available to political actors will cause a reduction in selective violence. We test this mechanism using information on shocks to coffee price in Colombia. In the next section, we answer why the theory of violence presented is a good theory for the Colombian conflict from the mid-1990s onwards, and why coffee is a commodity produced with a technology with high gains from collaboration.

To the best of our knowledge, this is the first paper that tries to understand how commodity price shocks affect violence through a lens of denunciation. Nevertheless, this is not the first paper that identifies the role of civilians' supply of information as an important channel

to explain conflict dynamics. Berman et al. [2011] study the channels by which aid help rebuild social and economic order in conflict and post-conflict areas. They predict and test that, aid in the form of public goods and services will motivate civilians to share information about insurgents with the government who can then build more effective counterinsurgency strategies.

2.3 Mechanism: Collaboration and denunciation

2.3.1 The Colombian context

For most of the late 1970's and 1980's Colombia saw a big expansion of guerrilla groups. In this period, groups claimed control of rural territories without much opposition. In many of these places guerrilla groups were de facto rulers who punished thieves, solved disputes over the ownership of land and animals, and intervened in situations of domestic violence, among others (Centro Nacional De Memoria Histórica [2016]).

Memory reports from the *Centro Nacional de Memoria Histórica*, describe how in this period rural populations and guerrilla groups coexisted, in many cases in relative peace. Most of the violence perpetrated by the groups at the time was in urban centers, where they extorted businessmen and cattle farmers, robbed banks and hospitals, and attacked public infrastructure to sabotage the government and to signal strength.

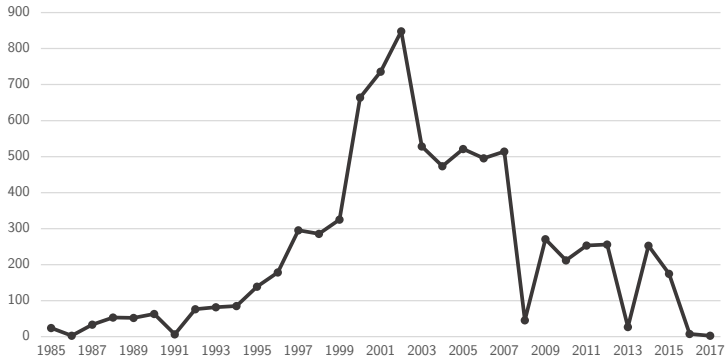
In the mid-1990s after the consolidation of the AUC as a unified paramilitary movement[2], the paramilitaries started a strong offensive against guerrilla groups. Paramilitary groups were financed by cattle farmers and owners of big extensions of land who opposed guerilla groups who constantly extorted them. From this period we see a sharp increase in the number of civilian victims of the conflict (Figure 2.1). We argue this trend is associated to the race between guerrilla and paramilitary groups over the control of the territories. During

2. AUC is the acronym for United Self-Defenders of Colombia.

this period the ELN and the FARC[3] made alliances to fight the paramilitaries who in many places worked with the connivance of the military.

Civilians were often victimized by paramilitaries under accusations of favoring guerrilla groups and viceversa. Lists with names of civilians who were suspected of being collaborators of one group or the other, and that were built with the help of civilian informants, were used to target and kill civilians (Centro Nacional De Memoria Histórica [2016]).

Figure 2.1: Number of Victims of the Armed Conflict



Source: National Victims Registry - Regístro Único de Víctimas (RUV), numbers in thousands of people

Collaboration didn't necessarily mean a premeditated action to aid one side. A civilian who was forced to sell goods to guerillas or who's son was forced to join a guerrilla group will be accused of being a supporter of the guerillas. Likewise, if he happened to "aid" paramilitaries. The following quote is found in the descriptions of the attacks in our dataset and exemplifies well the situation of many civilians in rural Colombia at the time.

> *"I worked to send money to my wife and kids but five months ago they arrived at the farm where I worked and recruited us. Since then, I worked without any pay, only violence and dead threats. As soon as I saw the chance, I escaped. I need to go far away otherwise they'll kill me. Some because I escaped and the others because they say I am a guerrilla member. I am neither"* (ONG Noche y Niebla)

3. ELN stands for National Liberation Army and FARC stands for Revolutionary Armed Forces of Colombia, these were the largest guerrilla group in Colombia

In summary, from the mid-1990s onwards, Kalyvas' theory of selective and indiscriminate violence is a good rationalization of the incentives political groups faced and the tactics they chose. Taking this framework as a benchmark, we answer the following question: How do this shocks impact civilian decisions to defect and denounce? And, how can shocks to commodity prices change the incentives of political groups to use one or other form of violence?

We argue that positive shocks to commodity prices that are produced with a technology that relies on collaboration across producers will make it more costly for civilians to denounce one another. This is the case since denunciation reduces the possible gains from collaboration, either because trust is broken or because the life of the denounced civilian is threatened. A reduction in denunciation will make political actors more uncertain about civilian support which will make selective violence less effective to the point it will be used less frequently.

In Colombia, coffee production relies heavily on collaboration across farmers. We will show that after a positive shock to the price of coffee there is a reduction in selective violence which is consistent with the mechanism we propose.

2.3.2 Coffee production in Colombia

Colombian coffee is produced by close to 550 thousand producers, most of whom are small family owned farms. Despite being a market of mainly small producers, the production chain is highly centralized and organized with an association -the National Federation of Coffee Growers (FNC)- that groups almost all producers and provides services that aim at improving the quality of the product and the revenue of producers.

The production of coffee demands a big influx of labor, especially when the time comes to pick ripe coffee grains from trees. The seasonal increase in labor is supplied mainly by informal workers who are paid by the day, and other forms of labor such as exchange labor. Exchange labor is an informal agreement where coffee growers provide labor to others in exchange of future labor. In addition, coffee growers not only collaborate with each other when it comes to labor exchange but also to reduce fix costs such as that of transportation.

All this points at the coffee production being one where there are big gains from collaboration across producers, and where collaboration takes the form, not of a contractual agreement but of an informal short term commitment. As such, it is ruled by informal processes: those who are hired for the day will likely be friends of friends. Exogenous increases in the price of coffee will make collaboration more profitable imposing a cost on denunciation.

2.4 Data and Empirical Design

2.4.1 Data

Geocoded violent actions in the context of Colombia's civil conflict are taken from the Conflict Analysis Resource Center (CERAC) dataset, introduced by Restrepo et al. [2003]. The dataset aggregates information on human rights violations and other violent attacks, from 1996 to 2014, collected from non-governmental organizations (NGO), media reports, as well as a network of Catholic priests in remote areas of the country. The dataset classifies the events as clashes or attacks (or both), and registers, when available, information on the groups participating in the events as well as the number of victims. We use violent attacks that occurred between 2004 and 2012 in our analysis.

The data set is complemented with descriptions of the events, collected by the NGO *Noche y Niebla* (CINEP) and the project *Rutas del Conflicto*, as well as journalistic reports. These sources collect detailed information on the apparent motives, the perpetrators and include a description of the events made by victims and witnesses.

Information on hectares used for coffee production is taken from Mora [2009]. He uses geo-coded information at the coffee field level (i.e. sub-farm level) from the Colombian National Information System (SICA) of the National Federation of Coffee Growers (FNC) on hectares allocated for coffee sowing and renewal of coffee plants to build a measure of the area used for production in each municipality and year.

Information on internal coffee prices is taken from the FNC. Coffee production in Brazil,

Indonesia, and Vietnam, that will be used to instrument for the municipal coffee revenue, is taken from the International Coffee Organization (ICO), yearly municipal population in Colombia is taken from the Center for Research in Economic Development (CEDE) municipal panel data set, and finally, information on municipal coca production in 1994 as well as average rainfall and temperature at the municipal levels are taken from Dube and Vargas [2013].

A novelty of this paper, is the classification of attacks as selective or indiscriminate. Interpreting Kalyvas' theory, a selective attack is one where perpetrators have information on the activities and/or affiliation of all the victims involved in an attack, as opposed to an indiscriminate attack where perpetrators lack any information on victims activities and affiliation.

2.4.2 Classifying attacks

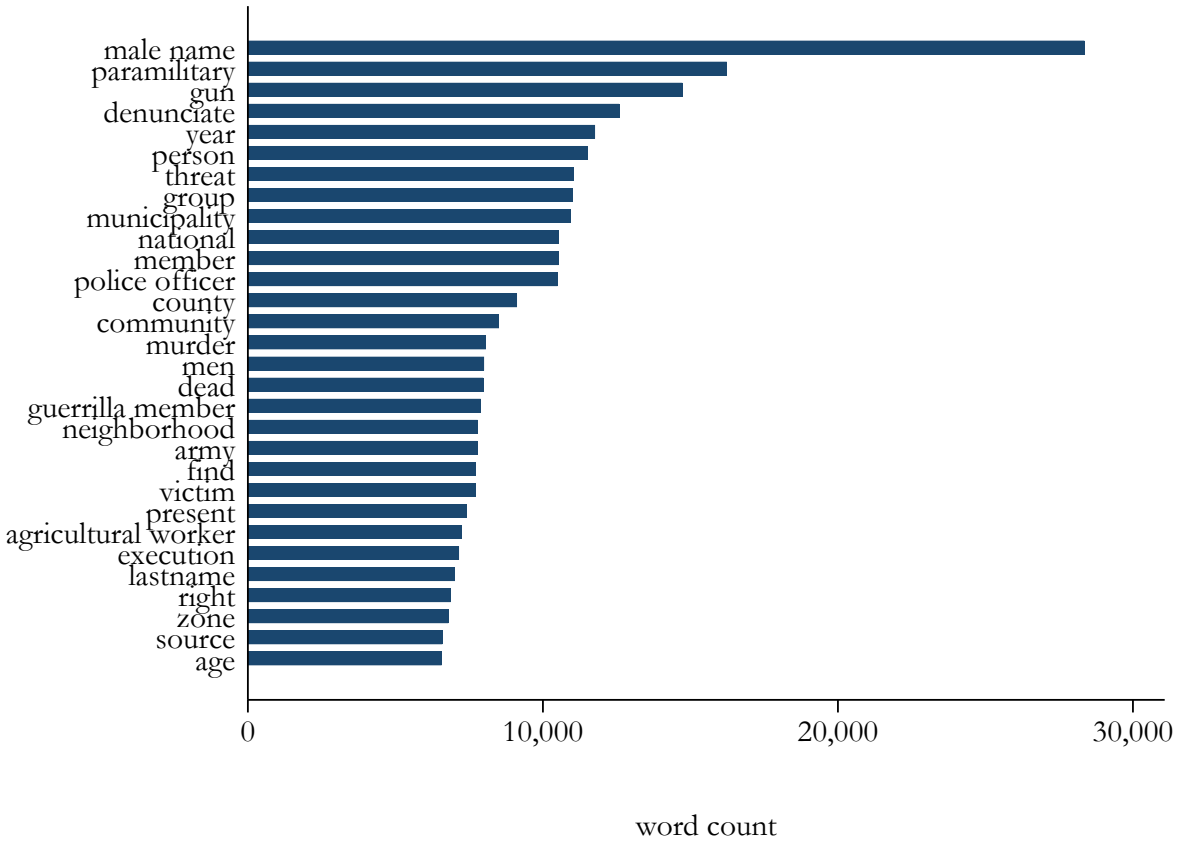
The data set has a total of 14,280 violent events between the years 2004 and 2012. This includes clashes between groups, attacks to infrastructure, land-mine explosions, kidnappings and other forms of victimization of civilians. We work with several classifications of the events between selective and indiscriminate. First, we leverage previous event classification done manually by CERAC and use these inputs to build the a classifier taking these categories as input, later we consider a couple of models of text classification to use all the potential of the descriptions in our dataset.

CERAC defines an *attack of political violence*, as an assassination, disappearance or other attacks directed toward civilians, that is perpetrated by a political actor. They exclude from this category clashes between armed groups, or any attack that is not directed toward civilians. As a starting point we take this classification and define a selective attack as an attack of political violence, also, we define an indiscriminate attack as one that is not of political violence.

Then we propose to use the descriptions of the attacks to carry out a text classification

analysis. Classifications will be done using three text classification methods, a Naive Bayes, Logit and Support Vector Machines (SVM) algorithms. To carry out the text classification analysis we clean the descriptions of every violent event. The text cleaning follows standard lematization, discarding prepositions as well as discarding words that appear only a handful of times in all the descriptions. Also, since the data set shows several male and female names, we replaced all male names with a unique identifier. The same is done for female names and last names.

Figure 2.2: Most frequent words and their word count



After cleaning the dataset of words each description has an average of 40 words that are used for the text analysis classification. Figure 2.2 shows the frequency of the 30th more common words in the cleaned set of descriptions. Male names are very common in all the descriptions, as well as words referring to paramilitary, or guerrilla group, and others that

refer to weapons, threats, killings and victims.

To classify the more than 14,000 violent events between indiscriminate and selective, a training set of 851 observations was built. These 851 violent events were manually classified by the same person in three categories: selective, indiscriminate or none. In the third category there are mainly clashes between insurgent groups and/or the army. A random 20% of the training set was used as a testing set. This means, we trained the algorithm with 80% of the training set and tested the predictions in the remaining 20% of the events.

For each text classification algorithm we run two models, the first classifies descriptions between *selective* and *not selective* attacks, the second classifies attacks between *indiscriminate* and *not indiscriminate*. For each description then we have two classifications.

Text classification methods model each description as a vector of words, or attributes, and model the probability that descriptions belong to each class. The class with the highest probability is selected. Under the Naive Bayes approach, Bayes' theorem is used to estimate the probabilities under the assumption that attributes in each description are drawn independently. The logistic method assumes that the probability follows a logistic functional form on a linear combination of the attributes in each description. Finally, the SVM finds the optimal hyperplane that separates the vectors of words in the training set. The planes are found by minimizing a loss function, in consequence, part of the data may be miss-classified.

Table 2.1: Performance of the classifiers

Model	Metric	Indiscriminate - Rest	Selective - Rest
NB	Accuracy	0.87	0.77
	Precision	0.61	0.71
	Recall	0.41	0.49
LOGIT	Accuracy	0.90	0.80
	Precision	1.00	0.92
	Recall	0.37	0.42
SVM	Accuracy	0.91	0.78
	Precision	1.00	0.91
	Recall	0.44	0.36

For each model we generate statistics on the performance of the classifiers on the testing

set. We compute three measures of performance: accuracy, precision and recall. Accuracy measures the fraction of observations in the testing set that are correctly classified, while precision and recall measure the incidence of false positives and false negatives, respectively. When these last two measures are one, false predictions go to zero. Table 2.1 shows these statistics for the setting that predicts indiscriminate attacks and for the setting that predicts selective attacks. While the model predicts accurately more than 80% of the classes in the testing set, there are differences in the proportion of false positives and false negatives. The precision of the model is higher than the recall, showing that false negatives are more frequent than false positives. Precision is particularly high for the logit and SVM models, but also the recall in these models is specially low. Miss-classification is mostly occurring because the model fails to classify as selective -or indiscriminate-, observations that are in this category.

Consistent with this, the classifiers -and specially the Logit and SVM- have a low count of selective and indiscriminate attacks. This implies that several attacks are left unclassified by the algorithm consistent with a high rate of false negatives. While using CERAC's classification, 68% of the attacks are selective and 32% are indiscriminate[4], these fractions are 51% and 35% for the NB classifier, and are even lower for the Logit and SVM classifiers with 26% and 8%.

Selective attacks classified with the classifying algorithms, as well as the manual classifier are all correlated (Table 2.2). This shows two things, first, that the results from the classifying algorithms are consistent with each other, and second, the manual classifier -although classified with different criteria- capture features common to the algorithms. Indiscriminate attacks classified with different methods are not as highly correlated. Selective and indiscriminate attacks are generally not correlated.

The model that dominates all others classifying indiscriminate events is the SVM. For classifying selective events, there is no one model that dominates all others in all measures

4. By definition these proportions add-up to one, that is not the case for the ML classifications where attacks can be neither selective nor indiscriminate, or attacks can be both.

Table 2.2: Correlation Across Classifications

		Selective				Indiscriminate			
		Manual	NB	Logit	SVM	Manual	NB	Logit	SVM
Selective	Manual	1.00							
	NB	0.79	1.00						
	Logit	0.70	0.78	1.00					
	SVM	0.66	0.75	0.97	1.00				
Indiscriminate	Manual	0.37	0.33	0.25	0.28	1.00			
	NB	0.35	0.60	0.32	0.32	0.52	1.00		
	Logit	0.19	0.15	0.10	0.11	0.50	0.35	1.00	
	SVM	0.24	0.19	0.14	0.16	0.55	0.38	0.92	1.00

although the model that arguably fit better the data is the logit model.

2.4.3 Coffee price shocks

In order to show some evidence on this channel, we show that after a positive shock to the price of coffee selective violence reduces importantly, more than indiscriminate does. The empirical strategy follows very closely that of Dube and Vargas [2013]. We estimate a *difference-in-difference* model where the treatment variable is the municipal revenue from the coffee activity. This measure of revenue is the product of the internal price of coffee that varies with time, and the average hectares used for coffee production for each municipality[5]. Since both the internal price of coffee and the intensity of production in each municipality may be endogenous to violence across the country, the municipal revenue from coffee is instrumented. Measures of average rainfall and temperature across time for each municipality will account for the potential of production that is exogenous of violence, and the total of production of the other three largest coffee producers: Brazil, Indonesia, and Vietnam (BIV), is used to capture variations in price that are uncorrelated with Colombia's supply.

More precisely, the revenue from coffee will be instrumented with three variables: the product of (i) temperature and production of BIV, (ii) rainfall and production of BIV, and

5. In this papers we use the average of production from 2000 to 2002. Results are robust when using the average for the period of analysis, 2003 to 2013.

(iii) rainfall, temperature and the production of BIV. Formally, the first stage of the model will look like:

$$q_{rj}p_t = \alpha_j + \beta_t + \delta_r t + \gamma coca_{jr} t + \phi ln(pop_{jt}) + \sum_{m=0}^{1}\sum_{n=0}^{1}(r_j^m t_j^n biv_t)\theta_{mn} + \eta_{jrt} \qquad \text{with } m+n > 0, \tag{2.1}$$

where r_j^m and t_j^n are rainfall and temperature in municipality j raised to the power of m and n in each case, and biv_t is the production of BIV in time t.

The first stage will be estimated using two stage least squares and standard errors will be clustered at the department level. The following second stage will be estimated:

$$y_{jrt} = \alpha_j + \beta_t + \delta_r t + \gamma coca_{jr} t + \rho(\hat{q_{jr}p_t}) + \phi ln(pop_{jt}) + \epsilon_{jrt} \ , \tag{2.2}$$

where y_{jrt} is a violence measure in region r municipality j at time t.

2.5 Results

Consistent with previous results, we find that a positive shock to the price of coffee impacts downward the violence perpetrated in coffee growing municipalities. Both the count of violent attacks and the number of casualties in these municipalities reduced as a consequence of an increase in the price of coffee. The results imply that if the median coffee producing municipality experiences and increase in coffee prices equivalent to the average yearly change in coffee prices for the study period, there is a reduction of about 58 violent events in a year and 13% reduction in casualties.

Now, once we classify attacks as selective or indiscriminate and run the regressions with these we find mixed results. Consistent with the model we propose, manually classified data shows that selective attacks in coffee growing municipalities significantly reduce as the price of coffee increases. Nevertheless, these results are not robust to every classification method

Table 2.3: Coffee revenues and violence

	Total Attacks	Casualties
Coffee Revenue	-0.045	-0.010
	(0.015)	(0.003)
Observations	8,964	8,964
KP F-statistic	11.69	11.69

Standard errors in parentheses

done using the machine learning classifiers. In these cases, we find a negative effect that is not statistically significant. As we know, some of these classifiers tend to have a high incidence of false negatives, which may be explaining the lack of power.

When we study the effect of coffee price shocks on indiscriminate attacks, we find no significant effect of these shock on manually classified attacks, or indiscriminate attacks classified by any other model. These result contrasts that of manually classified selective attacks. Although our theory gives us no prediction for indiscriminate attacks, tunning the model to reduce false negatives is key to have a more conclusive result.

Table 2.4: Coffee revenues and violence

	Selective				Indiscriminate			
	Manual	NB	LOGIT	SVM	Manual	NB	LOGIT	SVM
Coffee Revenue	-0.032	-0.006	-0.001	-0.001	-0.010	-0.001	-0.002	-0.002
	(0.01)	(0.007)	(0.002)	(0.002)	(0.006)	(0.006)	(0.001)	(0.001)
Observations	8,964	8,964	8,964	8,964	8,964	8,964	8,964	8,964
KP F-Statistic	11.69	11.69	11.69	11.69	11.69	11.69	11.69	11.69

Standard errors in parentheses

Using attacks classified with the manual classifier we find that reductions in selective violence are driven by reductions on attacks by the FARC and the paramilitaries. Indiscriminate attacks perpetrated by any group are not significantly reduced after positive shocks to the price of coffee.

Table 2.5: Coffee price shocks: Selective Attacks by each group

	FARC	ELN	Paramilitaries	Government
Coffee Revenue	-0.00369**	-0.00161	-0.00739**	-0.000911
	(0.00158)	(0.00108)	(0.00318)	(0.00131)
Observations	8964	8964	8964	8964
KP F-statistic	11.69	11.69	11.69	11.69

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Coffee price shocks: Indiscriminate Attacks by each group

	FARC	ELN	Paramilitaries	Government
Coffee Revenue	-0.00338	-0.000582	0.00295	-0.00372
	(0.00246)	(0.000821)	(0.00217)	(0.00263)
Observations	8964	8964	8964	8964
KP F-statistic	11.69	11.69	11.69	11.69

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6 Conclusion

This paper proposes a theory of the impact of price shocks on the use of selective violence by insurgents in ongoing civil conflicts. The mechanism is consistent with a theory where insurgents use violence strategically when they dispute the control of a territory, and civilians react strategically choosing to which group to support and whether to supply information on civilian activities. Anecdotal evidence confirms that in the Colombian conflict insurgencies collected information on civilian activities and used it to target individuals via selective violence. We find evidence in support of this mechanism, which highlights the crucial role of civilians in conflict dynamics.

References

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American economic review*, 93(3):729–747, 2003.

Atila Abdulkadiroglu, Parag A Pathak, Alvin E Roth, and Tayfun Sonmez. The boston public school match. *American Economic Review*, 95(2):368–371, 2005.

Atila Abdulkadiroğlu, Nikhil Agarwal, and Parag A Pathak. The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review*, 107(12):3635–89, 2017.

Atila Abdulkadiroglu, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. Do parents value school effectiveness? Technical report, National Bureau of Economic Research, 2017.

Nikhil Agarwal and Paulo Somaini. Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444, 2018.

Nikhil Agarwal and Paulo J Somaini. Revealed preference analysis of school choice models. Technical report, National Bureau of Economic Research, 2019.

Joseph G Altonji, Ching-I Huang, and Christopher R Taber. Estimating the cream skimming effect of school choice. *Journal of Political Economy*, 123(2):266–324, 2015.

Joshua D Angrist and Adriana D Kugler. Rural windfall or a new resource curse? coca, income, and civil conflict in colombia. *The Review of Economics and Statistics*, 90(2): 191–215, 2008.

Michel Balinski and Tayfun Sönmez. A tale of two mechanisms: student placement. *Journal of Economic theory*, 84(1):73–94, 1999.

Peter Bergman, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer. Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. Technical report, National Bureau of Economic Research, 2019.

Eli Berman, Jacob N Shapiro, and Joseph H Felter. Can hearts and minds be bought? the economics of counterinsurgency in iraq. *Journal of Political Economy*, 119(4):766–819, 2011.

Nicolas Berman and Mathieu Couttenier. External shocks, internal shots: the geography of civil conflicts. *The Review of Economics and Statistics*, 97(4):758–776, 2015.

Lex Borghans, Bart HH Golsteyn, and Ulf Zolitz. Parental preferences for primary school characteristics. *The BE Journal of Economic Analysis & Policy*, 15(1):85–117, 2015.

Boston Desegregation Project. (130). *Northeastern University Library, Archives and Special Collections*, 9(33), 1988.

Simon Burgess, Ellen Greaves, Anna Vignoles, and Deborah Wilson. What parents want: School preferences and school choice. *The Economic Journal*, 125(587):1262–1289, 2015.

Caterina Calsamiglia, Guillaume Haeringer, and Flip Klijn. Constrained school choice: An experimental study. *American Economic Review*, 100(4):1860–74, 2010.

Centro Nacional De Memoria Histórica. Granada: Memorias de guerra, resistencia y reconstrucción. *CNMH-Colciencias-Corporación Regional*, 2016.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3): 1107–1162, 2018.

Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

CINEP. Banco de datos de derechos humanos, dih y violencia poltica [en linea]. URL https://www.nocheyniebla.org/. Last downloaded on 2018-03-28.

Susan Clampet-Lundquist and Douglas S Massey. Neighborhood effects on economic self-sufficiency: A reconsideration of the moving to opportunity experiment. *American Journal of Sociology*, 114(1):107–143, 2008.

Oeindrila Dube and Juan F Vargas. Commodity price shocks and civil conflict: Evidence from colombia. *The Review of Economic Studies*, 80(4):1384–1421, 2013.

Lester E Dubins and David A Freedman. Machiavelli and the gale-shapley algorithm. *The American Mathematical Monthly*, 88(7):485–494, 1981.

Umut Dur, Scott Duke Kominers, Parag A Pathak, and Tayfun Sonmez. Reserve design: Unintended consequences and the demise of bostons walk zones. *Journal of Political Economy*, 126(6):2457–2479, 2018.

Dennis Epple and Richard E Romano. Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, pages 33–62, 1998.

Gabrielle Fack, Julien Grenet, and Yinghua He. Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109 (4):1486–1529, 2019.

Joseph H Felter, Benjamin Crost, et al. Export crops and civil conflict. Technical report, Empirical Studies of Conflict Project, 2016.

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

Steven Glazerman and Dallas Dotter. Market signals: Evidence on the determinants and consequences of school choice from a citywide lottery. *Educational Evaluation and Policy Analysis*, 39(4):593–619, 2017.

Guillaume Haeringer and Flip Klijn. Constrained school choice. *Journal of Economic theory*, 144(5):1921–1947, 2009.

Justine Hastings, Thomas J Kane, and Douglas O Staiger. Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper*, 2145:1–46, 2009.

Chang-Tai Hsieh and Miguel Urquiola. The effects of generalized school choice on achievement and stratification: Evidence from chile's voucher program. *Journal of public Economics*, 90(8-9):1477–1503, 2006.

Stathis N Kalyvas. *The logic of violence in civil war*. Cambridge University Press, 2006.

Adam Kapor, Christopher A Neilson, and Seth D Zimmerman. Heterogeneous beliefs and school choice mechanisms. Technical report, National Bureau of Economic Research, 2018.

Lawrence F Katz, Jeffrey R Kling, and Jeffrey B Liebman. Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics*, 116(2):607–654, 2001.

Jeffrey R Kling, Jeffrey B Liebman, and Lawrence F Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

Jean-William P Laliberte. Long-term contextual effects in education: Schools and neighborhoods. *V manuscript*, 2018.

Jens Ludwig, Greg J Duncan, Lisa A Gennetian, Lawrence F Katz, Ronald C Kessler, Jeffrey R Kling, and Lisa Sanbonmatsu. Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review*, 103(3):226–31, 2013.

Margaux Luflade. The value of information in centralized school choice systems. *Unpublished manuscript*, 2018.

W Bentley MacLeod and Miguel Urquiola. Reputation and school competition. *American Economic Review*, 105(11):3471–88, 2015.

Juan Carlos Muñoz Mora. *Los caminos del café: aproximación a los efectos del conflicto armado rural en la producción cafetera colombiana*. PhD thesis, Uniandes, 2009.

Hessel Oosterbeek, Sándor Sóvágó, and Bas Klaauw. Why are schools segregated? evidence from the secondary-school match in amsterdam. 2019.

Parag A Pathak and Peng Shi. Simulating alternative school choice options in boston-technical appendix, 2013.

Parag A Pathak and Peng Shi. How well do structural demand models work? counterfactual predictions in school choice. Technical report, National Bureau of Economic Research, 2017.

Jorge A Restrepo, Michael Spagat, and Juan F Vargas. The dynamics of the colombian civil conflict: A new data set. 2003.

Alvin E Roth. The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628, 1982.

Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

Christopher R Walters. The demand for effective charter schools. *Journal of Political Economy*, 126(6):2179–2223, 2018.