On the Spatial Determinants of Educational Access*

Francesco Agostinelli† Margaux Luflade‡ Paolo Martellini§

January 2023

Abstract

We study the role of local institutions and regulations—school boundaries, school transportation provision, and zoning restrictions—in determining inequalities of educational opportunities for children. Motivated by our empirical findings on how the demand for both neighborhoods and schools responds to quasi-experimental variation in school quality and transportation, we build and estimate a spatial equilibrium model of residential sorting and school choice. We validate the model with our empirical quasi-experimental findings as well as with experimental estimates from an influential voucher program. We find that the evaluation of both people-based and place-based policies heavily hinges on spatial equilibrium effects. Abstracting from those would lead to either overestimating (voucher) or overturning (school choice expansion) the impact of these policies on the inequality of educational access.

JEL Classification: I24, J13, R23, R31.

*We thank seminar participants at the Saieh Family Fellows Workshop (Becker Friedman Institute), Arizona State University, Cornell University, the University of Arizona, the University of Minnesota, the University of Oslo, the University of Pennsylvania, and the Urban Economics Association Conference. We also thank the Office of Student Assignment of the Wake County Public School District, in particular the Student Information, Demographic and Spatial Analytics office, for sharing school boundary maps and helping us understand the WCPSS institutional setting. Ornella Darova provided outstanding research assistance. All errors are our own.

†Department of Economics, University of Pennsylvania. Email: fagostin@sas.upenn.edu

‡Department of Economics, University of Pennsylvania. Email: mluflade@sas.upenn.edu

§Department of Economics, University of Wisconsin-Madison. Email: martellini@wisc.edu

1 Introduction

Most cities in the United States are characterized by a significant degree of neighborhood income segregation. Such spatial heterogeneity in neighborhood composition translates into inequality in access to local amenities, such as high-quality public schools. This topic has attracted the attention of policymakers, given the importance of early childhood education and peer effects on children's outcomes. To mitigate the connection between residential location and educational access, school choice policies have been implemented in multiple cities in recent years. However, the effectiveness of these policies crucially depends on the specific institutional design (e.g., the provision of transportation), on the geography of cities (e.g., the geographical distance between homes and schools), and on the equilibrium response of households' neighborhood and school choice.

In this paper we study how local institutions and regulations—neighborhood school portfolios, school transportation services, and local housing restrictions—interact with residential and school choices of families in the formation of the observed inequality in educational access and outcomes. Toward this aim, we develop the first empirical equilibrium model of neighborhood and school choice that allows us to jointly account for three key features of public school demand: (i) spatial heterogeneity in schooling opportunities due to differences in the set of available schools across neighborhoods; (ii) demand for schools that depends on geographical distance from home, making school transportation a key determinant of school choice; and (iii) local zoning regulations that affect the supply of housing in certain neighborhoods, with consequences on the equilibrium residential sorting in the city and school composition.

In the first part of the paper we provide direct empirical evidence on the extent to which demand for neighborhoods and schools responds to changes in the school quality and the availability of school transportation, and how this translates into children's educational outcomes. Our empirical analysis focuses on Wake County (North Carolina), which is a natural setting to study our questions of interest for several reasons. First, it is covered by a large county-wide school district (the Wake County Public School System, WCPSS) spanning roughly 850 square miles, which makes geographical access and transportation relevant questions. Second, a number of institutional changes regarding the boundaries of catchment areas and the public school choice network have been made over the last two decades.

We combine several administrative data sources for a comprehensive new dataset. We use student-level administrative data from the North Carolina Education Research Data Center (NCERDC) to access information about the universe of elementary school children, the school they attend, their test scores, and residential information. We merge this information with (i) information about school geographical enrollment boundaries, the set of available schools for each neighborhood, and admission probabilities to these schools; (ii) the map of the school transportation system; and (iii) data on house prices and residential zoning regulation for the entire county.

Our first set of results documents how neighborhood demand responds to school quality. We exploit longitudinal variation in school boundaries at the neighborhood level—that is, the set of schools available to children in each neighborhood—to construct quasi-experimental changes in peer composition, which is our measure of school quality. Our preferred intention-to-treat analysis shows that a 10 percent change in elementary school quality induces a 0.3 percent increase in house prices within the treated neighborhoods. We interpret this finding as evidence that neighborhood demand depends on the local school quality and that this willingness to pay from families is capitalized into house prices.

The same variation in school boundaries also affects the accumulation of children's skills. Our intention-to-treat analysis reveals that a 10 percent improvement in neighborhood base school quality leads to an increase of approximately 1 percent in children's test scores. The results are mostly driven by disadvantaged children who display the highest benefits from improved school quality. We see these results as evidence that local institutions directly impact the inequality of educational opportunities and outcomes for children.

Finally, we show that the demand for schools depends on the availability of school transportation services. By exploiting longitudinal variation in transportation availability between neighborhoods within the same school boundaries, our results highlight that once schools start providing transportation to and from a specific neighborhood, families increase their demand for those schools and substitute away from their assigned neighborhood (base) schools. This pattern of substitution between schools depends on the geographical location of schools: as home–school distance increases, the effect of the transportation on the demand for schools vanishes.

In the second part of the paper we develop and estimate a heterogeneous-agent equilibrium model of residential sorting and school choice. Conditional on their residential location, families have access to a neighborhood-specific portfolio of schools they can apply to. Families' school choices are determined by school quality and disutility from commuting, which depends on transportation provision. Seats in oversubscribed schools are allocated through a lottery system that grants applicant families admission with a certain probability. Given the value of the school portfolios associated with each neighborhood, families decide where to live. Such choice is further affected by the cost of housing, the quality of neighborhood amenities, and the existence of neighborhood-specific zoning restrictions that impose a lower bound on housing demand. Modeling the joint decision of neighborhoods and schools allows us to capture the key determinants of educational access: portfolios of schools that vary by residential location, home–school distance and transportation, and zoning restrictions to housing demand. Crucially for our policy counterfactuals, house prices, admission probabilities, and school quality are equilibrium objects, the latter being shaped by the composition of enrolled children.

We estimate the model via the method of simulated moments. Our identification strategy exploits both cross-sectional and longitudinal variation in neighborhood composition and school quality, distance between home and schools attended, and admission probabilities to oversubscribed schools. Our model replicates neighborhood income segregation as well as the heterogeneity in school composition in terms of children's skills.

We validate the estimated model in terms of the elasticities of the demand for neighborhood and school, which are key for our counterfactual results. We find that the model replicates—without being targeted in estimation— several valuable pieces of evidence in the data. First, we show that the model replicates the quasi-experimental reduced-form estimates on how the demand for schools responds to changes in school transportation. Second, we simulate the Moving to Opportunity (MTO) experiment in our estimated model by providing rent subsidies to low-income families so they can live in higher-income neighborhoods. We show that simulations from the models replicate the empirical findings of the MTO evaluation in Galiani et al. (2015) in terms of neighborhood income of the chosen destinations for families who take the voucher.

We use the estimated and validated model to consider three different policies that are aimed at expanding educational access for economically disadvantaged children. In the first counterfactual exercise, we study the consequences for residential sorting and school enrollment of expanding school choice for children in poor neighborhoods. We perform this by including the highest-quality schools as additional school options for children living in some low-income neighborhoods that are otherwise associated with the lowest-quality schools of the district. The outcomes of the policy heavily hinge on whether transportation to the added schools is provided: while distance to school is a clear barrier to educational access—with fewer than 1 percent of children applying to the newly available schools when transportation is not available—the policy becomes more successful once targeted families are provided with transportation service (a take-up rate up to 20 percent for a school seven miles away from home).

We find that this policy also has a positive effect on the local base school quality in the poor neighborhoods. By making the school portfolios of those neighborhoods more appealing, the policy generates an inflow of some higher-income families. Their children, when they are not admitted to the (oversubscribed) high-quality schools, enroll in the local base schools of the originally disadvantaged neighborhoods. This outcome is a distinct feature of our model in which families choose both schools and neighborhoods. We make this point explicit by performing the same policy experiment under the assumption that residential location was policy invariant, as it is commonly assumed in the school choice literature. Strikingly, ignoring changes in residential composition would lead to the opposite conclusion in terms of changes in school quality of the targeted disadvantaged neighborhoods.

Our second counterfactual policy focuses on expanding educational opportunities for children via housing vouchers, in the spirit of the Moving to Opportunity (MTO) experiment. We first consider only the individual treatment effect of this policy, as in our validation exercise, and find that children of beneficiary families increase their skills by 22.1 percent of a standard deviation. In contrast, our general equilibrium analysis of the policy reveals and average effect on children of families who take the voucher that is 40 percent smaller (+13.2 percent of a standard deviation). We also document a slightly negative impact on children who lived in the receiving neighborhoods at baseline (-6.6 percent of a standard deviation), even after accounting for endogenous residential resorting, and a small positive impact on all other neighborhoods (1.7 percent of a standard deviation). This result highlights the importance of accounting for general equilibrium effects when evaluating large-scale MTO-like policies.

Finally, we analyze the role of zoning regulations on educational opportunities for children. We focus on the same high-income neighborhoods that were studied in the voucher analysis, as those neighborhoods turn out to be highly regulated in terms of minimum lot size. The results of our upzoning policy highlight how house-size regulations effectively reduce competition in the local housing market by creating barriers to entry for low-income families and consequently lower the cost paid by high-income families to access high-quality schools.

Related Literature. This paper connects several strands of literature. First, this paper contributes to the literature on residential choice and school valuation, motivated by the well-documented fact that school quality capitalizes into house prices (Black, 1999). Epple and Sieg (1999) first proposed an empirical equilibrium framework to test Tiebout sorting across different neighborhoods and local jurisdictions. In their framework, agents sort into neighborhoods based on their preferences for housing and local public goods. The provision of neighborhood's public goods is financed via local property taxes. [1] Bayer

[1] Sieg et al. (2004) further developed an empirical method to estimate willingness to pay for large changes

et al. (2007) develop an empirical framework of neighborhood choice, where families have preferences over school quality and neighborhood amenities. The authors exploit boundary discontinuity design to identify household preferences for schools and neighborhood amenities, accounting for the endogenous socioeconomic composition of the neighborhood. Our paper departs from these previous studies in three important ways. First, we consider an environment with public school choice rather than a neighborhood-assignment setting in which families' residential location fully determines the school their children attend. This allows us to evaluate the extent to which, and the conditions under which, public school choice decouples school and residential choices. In addition, our model treats the composition of schools and neighborhoods as equilibrium objects, which respond to policy changes. Finally, on the identification side, we use time—instead of purely cross-sectional—variation in assignments of schools to neighborhoods to isolate the elasticity of price to school quality from the value of other neighborhood amenities.

Our paper also contributes to the empirical literature interested in the determinants of school choice. While recognizing the importance of the distance from home in families' choice of school for their children, previous studies (e.g., Hastings et al., 2009; Abdulkadiroğlu et al., 2017; Agarwal and Somaini, 2018; Kapor et al., 2020; Laverde, 2020) have treated home location as exogenously fixed. Our counterfactual analysis of school choice policies highlights the importance of accounting for endogenous residential choices by contrasting the policy outcomes in our setting with one in which families were not allowed to choose their residential location.

Our equilibrium model is also in the spirit of Nechyba (2000)'s, who studies the effect of private school vouchers on household sorting and educational opportunities by accounting for state and local taxes. A similar public finance focus within a model of school choice is also present in Epple and Romano (2003). In our work, we explore the relationship between neighborhood choice and educational access within a single school district, Wake County, where local financing is homogeneous across schools.[2] Avery and Pathak (2021) build a model in which school quality and housing prices are determined in equilibrium and study the extent to which public school choice increases access to high-quality schools for low-income children. In sharing their interest in the equilibrium response to school choice expansion, we investigate the geographic dimension of such policy in terms of home–school distance, school transportation, and zoning regulations.

in local public goods, in the presence of household spatial sorting, unobserved preference heterogeneity, and general equilibrium effects in the local housing market.

[2]In addition, our empirical framework allows us to quantify the constraints families face in departing from their default school option. While Epple and Romano (2003) assume a fixed cost of attending a school located in any neighborhood other than the residential one, we show that home–school distance and transportation provision dramatically influence the equilibrium outcome of a school choice expansion policy.

This paper also contributes to the growing urban literature that studies how the interaction between agglomeration and congestion forces shape city structure and individual outcomes (Ahlfeldt et al., 2015). Recent contributions have focused on the effect of transportation infrastructure on commuting patterns (Heblich et al., 2020) and neighborhood sorting (Tsivanidis, 2019). Our paper also explores how transportation provision, along with local prices, affects access to desired locations within a city. However, we depart from much of the urban literature that deals with labor market access and explore heterogeneity across neighborhoods in their measure of educational access. We believe that our results, in particular those concerning the extent to which zoning regulations restrict access to certain neighborhoods—and neighborhood-specific amenities—extend beyond this paper's focus on education inequality. Related to our work, Fogli and Guerrieri (2018) and Eckert and Kleineberg (2021) build dynamic models focused on how neighborhood choice affects children's human capital formation. Different from us, they abstract from transportation provision and school choice, which are key to our estimation of the determinants of educational access.

Finally, a large body of literature studies the formulation and identification of social interactions models with neighborhood and peer effects (see, e.g., Calvó-Armengol et al., 2009; Durlauf and Ioannides, 2010; Blume et al., 2015). A recent development in this literature has linked parental investments and parenting style to the incentives created by the local socioeconomic environments that families and children are exposed to (Doepke et al., 2019; Agostinelli et al., 2020a). Heckman and Mosso (2014) and Mogstad and Torsvik (2022) review the literature on the effects of families and neighborhoods on child development and intergenerational mobility, respectively. Our paper contributes to this literature by endogenizing both the residential decisions of families and their schooling (investment) choices within a spatial equilibrium model with peer effects. While abstracting from neighborhood choice, Allende (2020) demonstrates the importance of families' preferences for peers in the study of school competition, which is beyond the scope of this paper.

2 Data, Institutional Background, and Empirical Facts

We use data from the Wake County Public School System (WCPSS), which is the fourteenth largest school district in the United States.[3] We restrict our attention to the stu-

[3] As of 2018–19; see, National Center for Education Statistics (NCES), https://nces.ed.gov/programs/digest/d18/tables/dt18_215.30.asp?current=yes, accessed August 2021.

dents attending public school.[4] In this section we provide direct empirical evidence that the demand for neighborhoods and schools responds to changes in educational services, such as school quality and school transportation. These empirical facts will guide us in the development of our equilibrium theory of residential sorting and school choice.

2.1 Institutional Background

We provide a brief overview of the institutional elements that characterize public school choice in the WCPSS during our period of interest, focusing on the features essential for our empirical analysis. More details can be found in Appendix A. Each residential address in Wake County is associated with a *base* school at which the child is guaranteed a seat and transportation. Residential address also determines the menu of *option* schools to which parents can seek admission for their child (e.g., magnet programs, calendar transfer programs).[5] Schools that are option schools for families in one location are typically base school for families in another location. When the number of applications to an option school exceeds the number of seats, assignment is made (in the years covered in this paper) using the Boston mechanism.[6]. The family's residential address also determines whether school transportation is provided to each available option school.[7] The Wake County board of education has changed the assignment of residential addresses to base and option schools at various times. We take advantage of these institutional changes to provide empirical evidence of families' valuation of school quality and transportation and to identify structural parameters governing their neighborhood and school choices.

School boundaries and neighborhood definition. We formalize these institutional changes using the following definitions. In any school year t, each residential address in the county, which we denote by its latitude and longitude coordinates $z = (z_x, z_y)$, is as-

[4]We exclude charter schools and private schools from our analysis. Between 2003–04 and 2005–06, the period of interest, nine charter schools and 32 private schools offered kindergarten, enrolling on average 474 and 1,236 kindergartners per year, respectively, against an average of 9,720 kindergartners a year in traditional public schools. Sources: NCERDC data for charter schools, and `https://ncadmin.nc.gov/public/private-school-information/nc-directory-private-schools` and `https://ncadmin.nc.gov/public/private-school-information/state-north-carolina-private-grade-k-12-school-statistics` for private schools.

[5]We use the term *option school* to designate the different types of public programs into which WCPSS parents can enroll their child as an alternative to their base school. The different types of programs are described in Appendix A.1.1.

[6]See footnote 23 in Section 3 for more details about the mechanism and how our model relates to it.

[7]Figure 1 in Dur et al. (2018) shows a screenshot of the online platform parents can use to apply; the fourth column in the table illustrates the variation of transportation provision across schools and residential addresses. Entering an address in the WCPSS's address look-up tool (`https://wwwgis2.wcpss.net/addressLookup/`, accessed August 2021) further illustrates how school and transportation eligibility are determined by the family's residential address.

sociated with a *portfolio* of schools $\mathcal{L}(z;t) = (\mathcal{B}(z;t), \mathcal{T}(z;t), \mathcal{NT}(z;t))$, where $\mathcal{B}(z;t)$ is the base school associated with z in year t; $\mathcal{T}(z;t)$ is the set of option schools providing transportation to z in year t; and $\mathcal{NT}(z;t)$ is the set of option schools in the choice set of z that do not provide transportation to z in year t. We define *school boundaries* for year t to be the mapping from $\mathbb{R}^2$ to the set of all possible school portfolios that assign residential address z to its portfolio at t, $\mathcal{L}(z;t)$. We define a (base) school s's *catchment area* for year t, denoted as $\mathcal{C}_{s,t}$, as the set of all addresses z whose base school at t is s, that is: $\mathcal{C}_{s,t} = \{z \mid \mathcal{B}(z;t) = s\}$. Finally, we define a *neighborhood* as the union of all *contiguous* points sharing a common school portfolio across all years in our sample. That is, we call neighborhood n and denote neighborhood n's school portfolio as $\{ (\mathcal{B}_{n,t}, \mathcal{T}_{n,t}, \mathcal{NT}_{n,t}) \mid t = 2003, \ldots, 2009\}$, the set all contiguous points z such that for each t, $\mathcal{B}(z;t) = \mathcal{B}_{n,t}$, $\mathcal{T}(z;t) = \mathcal{T}_{n,t}$, and $\mathcal{NT}(z;t) = \mathcal{NT}_{n,t}$. We will use these definitions to derive empirical evidence of our mechanisms of interest in Section 2.3 and to characterize the environment of our model in Section 3.

Institutional changes. Two considerations drive changes in school boundaries. In theory, from the 2000–01 academic year and until 2011–12, the WCPSS had the goal of ensuring socioeconomic balance within and across schools. Assignment of addresses to schools was supposed to guarantee that no school had more than 40 percent of students eligible for free or reduced-price lunch (now designated as economically disadvantaged, or ED) or had more than 25 percent of students below the state's reading standards for their grade. However, in practice, the main reason residential addresses were reassigned "from school to school [was] because of population growth, and that is what it was. The busing was not intended primarily for diversity but just to fill in ... schools" (Parcel and Taylor, 2015, p. 53).[8] Over our sample period, 24 percent of base schools experienced a change in their catchment area across years (see bottom panel of Table A-2 and Figure A-2 in Appendix A.1.1 for an illustration), and 24 percent of option schools saw a change in their set of eligible neighborhoods and/or in their transportation provision (see Figure A-3 in A.1.1 and the bottom panel of Table A-2). In Appendix A.1.2 we show that this variation does not correlate with predetermined differences in the local composition of children and families.

[8]The census shows that total population in Wake County increased by 42 percent from 2000 to 2010. Authors' calculations from NCERDC data show that this translated into an increase in the public school student population of 48 percent.

2.2 Data

Data sources. The data come from four main sources. Student-level data, including school attended, sociodemographic characteristics, yearly test scores, and residential address, were obtained from the NCERDC. The WCPSS directly provided us with yearly data (maps) characterizing the assignment of residential addresses to base schools and menus of options as well as the availability of school transportation between each residential address and each option school. The WCPSS also shared school-level information such as the number of applications accepted and denied by each option school every year. We combined publicly available datasets from Wake County to gather information on house prices, house characteristics, and minimum lot size regulations.[9] Finally, we use Census tract-level information on population counts and household income from the American Community Survey ACS) five-year estimates (2006–10).

Sample description. We refer to student cohorts by the year they enter kindergarten—the year at which we assume parents make neighborhood and school decisions in the model. Cohort 2003–04 is the earliest one for which address information is available in the NCERDC data; Cohort 2008–09 is the last one to enter school before a number of significant changes in the student assignment plan, brought about by a change of political majority in the WCPSS School Board, took place in 2009. Estimation of the model requires information about admission probabilities to schools, which we only obtained for Cohorts 2003–04 to 2005–06, so we restrict our attention to these three cohorts to bring our model to the data.Next, we provide key descriptive statistics about our samples of neighborhoods, students, and schools. We refer the reader to Appendix A for further details about the combination of the different data sources, sample restrictions, and missing observations.

Neighborhoods. We construct neighborhoods as defined in Section 2.1, rank them by decreasing order of their student population, and exclude the least populated neighborhoods so as to keep 90 percent of the students. Figure A-5 in Appendix A.3 shows the final partition of Wake County into our neighborhoods, with the 2000 census tracts boundaries for comparison. The top panel of Table A-2 in Appendix A.5 describes our final sample of 305 neighborhoods. On average, each neighborhood is associated with a choice set of 13 option schools in addition to its base. Option schools tend to be significantly farther away from the neighborhood compared with the base (11 versus 3.7 miles on average). About 15 percent of neighborhoods experienced a change in base school over the sample

[9] https://www.wakegov.com/departments-government/geographic-information-services-gis/maps-apps-data, accessed August 2021.

period; almost all of them experienced a change in eligibility and/or transportation to some option school. Regarding zoning, there is significant heterogeneity in minimum lot size (MLS) regulations across neighborhoods (63 percent have some MLS regulation, and higher-density neighborhoods are concentrated in the urban center of the county—the city of Raleigh—as reported in Figure A-4 in Appendix A.2), with an average minimum lot size of 0.15 acres and a standard deviation of 0.24 acres per lot. Accordingly, there is significant variation in actual average lot size (mean of 0.45 acres, standard deviation of 0.46 acres) and house size (mean of 2,107 square feet, standard deviate of 653 square feet) across neighborhoods.

Students. For Cohorts 2003–04 through 2006–07, we drop student observations for which school attended, address, end-of-grade test scores, or economically disadvantaged (ED, hereafter) status (measured by eligibility for free or reduced-price lunch) is missing. We also exclude from the sample students attending a school outside the menu of schools attached to their residential address.[10] Our final sample contains 16,445 students over three cohorts. Thirty percent of students in the sample qualify as ED students. ED students are more likely to attend their base school than non-ED ones (92 percent versus 79 percent) while non-ED students are more likely to attend option schools with transportation than ED. Attendance at option schools without transportation is equally low among ED and non-ED students (3.5 percent). There are about twice as many applicants to option schools as attendees, since all option schools are oversubscribed and have an average admission probability of 50 percent. We use test scores at the end of third grade (standardized by cohort) as a measure of student skills. The average skill gap between ED and non-ED students is about a standard deviation of the skill distribution (mean of -0.58 for ED versus 0.40 for non-ED students).

Schools. There are 87 elementary schools in our sample. The bottom panel of Table A-2 shows descriptive statistics for our sample of schools. While all 87 schools are base schools, 27 (33 percent) are also option schools for some neighborhoods. There is also significant variance in the share of ED students across schools.

2.3 Motivating Empirical Findings

We provide direct evidence of our main hypotheses: (i) the demand for schools and neighborhoods responds to changes in the quality of the educational services, namely the

[10]Students may attend schools outside of the choice set attached to their residential address for multiple reasons. For instance, they may attend the same school as an older sibling (which was a part of the choice set in the past) or they may attend a school at which one of their parents is employed.

schools' peer composition and the availability of school transportation; and (ii) changes in school quality have a direct impact on children's learning. The motivating evidence shown in this section is based on Cohorts 2003–04 through 2008–09.

Neighborhood Demand and School Quality. We provide quasi-experimental evidence on the effect of the quality of school peers on house prices.[11] If families value good schools for their children, an increase in the quality of a school will be capitalized into the house prices of the school's catchment area. We consider the following regression model:

$$\ln r_{j,n,t} = \beta_0 + \beta_1 \ln \text{School Quality}_{n,t} + \lambda_n + \gamma_t + \epsilon_{j,n,t}\,, \tag{2.1}$$

where $\ln r_{j,n,t}$ is the (log) price of house j in neighborhood n in year t; $\ln \text{School Quality}_{n,t}$ represents the log of the mean test score of children living in the catchment area of the base school associated with neighborhood n in period t; γ_t and λ_n are year and neighborhood fixed effects, respectively; and $\epsilon_{j,n,t}$ is an error term.

In this empirical exercise, we are interested in the coefficient β_1, which we interpret as the average responsiveness of house prices with respect to changes in the quality of the base school. Estimating directly the regression in (2.3) via ordinary least squares (OLS) would lead to biased estimates for β_1 as changes in school peer composition in a given neighborhood n depend on: (i) institutional changes in the base school catchment area, and (ii) endogenous changes in the composition of families/children who live in the neighborhood. We aim to isolate the first type of variation, namely the longitudinal variation in the base school catchment areas to create exogenous variation in the pool of potential school attendees.[12] For this reason, our empirical design is based on the constructed exogenous longitudinal changes for each neighborhood n of the peer quality induced by the change in its base school catchment area $\mathcal{C}_{s,t}$ *only*:

$$\Delta_n \ln \text{School Quality}_{n,t} = \ln \widehat{\text{School Quality}}_{n,t} - \ln \text{School Quality}_{n,t-1}\,, \tag{2.2}$$

where $\widehat{\text{School Quality}}_{n,t}$ represents the mean test score at period t of children living in the catchment area designed at time t while keeping family residential location constant to its configuration at time $t-1$. In other words, our constructed variable $\Delta_n \ln \text{School Quality}_{n,t}$ measures the change in the composition of base school potential attendees, from one school year to the next, induced by changes in the base school catchment areas if nothing else had changed in between.

[11] Following Epple and Romano (2003), Bayer et al. (2007), and Avery and Pathak (2021) among others, we use average student skills (proxied by test scores) as our measure of school quality.

[12] In our structural model, we will allow for residential sorting as an endogenous family choice.

Table 1: Changes in School Quality and Effects on House Prices and Children's Learning

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: HOUSE PRICES						
	House Prices Psf (Log)					
Changes In School Quality (Log)	0.33***	0.02***	0.03***			
	(0.01)	(0.01)	(0.01)			
Observations	92,849	92,849	75,527			
Model	Pooled	Exog. Longit. (Fixed Effects)	Exog. Longit. (First Diff)			
Sale-Year F.E.	Yes	Yes	Yes			
House Characteristics	Yes	Yes	Yes			
PANEL B: CHILDREN'S TEST SCORES						
	4th-Grade Scores (Log)			5th-Grade Scores (Log)		
Changes In School Quality (Log)	0.13***	0.03**	0.10***	0.14***	0.06***	0.11***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.03)
Observations	63,127	63,127	55,684	49,391	49,391	42,821
School-Specific Trends	No	Yes	Yes	No	Yes	Yes
Model	Pooled	Exog. Longit. (Fixed Effects)	Exog. Longit. (First Diff)	Pooled	Exog. Longit. (Fixed Effects)	Exog. Longit. (First Diff)
	4th-Grade Scores (Log) Heterogeneity by 3rd-Grade Scores			5th-Grade Scores (Log) Heterogeneity by 3rd-Grade Scores		
Changes In School Quality (Log) × Below Median 3rd-Grade Score	0.14***	0.05***	0.21***	0.15***	0.08***	0.23***
	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.04)
Changes In School Quality (Log) × Above Median 3rd-Grade Score	0.10***	-0.02	0.09**	0.12***	0.02	-0.03
	(0.01)	(0.01)	(0.04)	(0.01)	(0.02)	(0.04)
Observations	63,127	63,127	55,684	49,391	49,391	42,821
School-Specific Trends	No	Yes	Yes	No	Yes	Yes
Model	Pooled	Exog. Longit. (Fixed Effects)	Exog. Longit. (First Diff)	Pooled	Exog. Longit. (Fixed Effects)	Exog. Longit. (First Diff)

The table shows the effect of school quality on house prices (Panel A) and on children's learning (Panel B). School quality is measured as the average test score of children attending the base school associated with the neighborhood in which the house is located. Details on the construction of the variable for changes in school quality are provided in Section 2.3. Test score outcomes are fourth and fifth grade math standardized scores. Standard errors are robust to heteroskedasticity and reported in parentheses. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Our regression model, which now removes unobserved spatial heterogeneity between neighborhoods as well as potential endogenous sorting, is:

$$\Delta_n \ln r_{n,t} = \beta_1 \Delta_n \ln \text{School Quality}_{n,t} + \delta_t + \Delta_n \epsilon_{n,t}\,, \tag{2.3}$$

where $\Delta_n \ln r_{n,t}$ and $\Delta_n \ln \text{School Quality}_{n,t}$ denote the within-neighborhood longitudinal variation in average house (log) prices and average (log) school quality, respectively. The parameter δ_t captures any year-specific aggregate change in prices.

Table 1 presents the results of this analysis. In column (1) we estimate the pooled version of the regression Model (2.1) without neighborhood fixed effects. In column (3), we instead estimate regression Model (2.3) using the exogenous changes in school quality defined in Equation (2.2), once accounting for neighborhood-specific effects (neighborhood fixed effects). Finally, in column (2) we replicate our preferred specification of column (3) by exploiting the entire within-neighborhood longitudinal changes of base school quality ($\Delta_n \ln \text{School Quality}_{n,t} = \ln \widehat{\text{School Quality}}_{n,t} - E[\ln \widehat{\text{School Quality}}_{n,t}|n]$).

We find that a 10 percent change in school quality is associated with an increase in house prices by 3.3 percent (column (1)). As this pooled regression exploits both cross-sectional and longitudinal changes in school quality, the estimated effect confounds unobserved differences in neighborhoods capitalized in prices with differences in the quality of their associated schools. For this reason, the size of the coefficient is considerably lower for columns (2)–(3), which eliminates the unobserved neighborhood heterogeneity and endogeneity in residential sorting. Column (3) shows that a 10 percent change in school quality induces neighborhood demand to increase, with an associated 0.3 percent change in housing prices. The results for the fixed-effect specification in column (2) are in line with this result (0.2 percent).[13] The result from our preferred specifications in column (3) of Table 1 needs to be interpreted as an intent-to-treat. Changes in elementary school quality only affect neighborhood demand for a fraction of the entire population—namely, families with children of elementary school age. We will reinterpret these estimates through the lens of our structural model that includes both affected and unaffected households.[14]

School Quality and the Effect on Children's Learning. To study the effect of changes in school boundaries and associated changes in school quality on children's learning, we consider the following value-added regression model:

$$\ln a^{'}_{i,s,t+1} = \beta_1 \ln \text{School Quality}_{n(s),t} + \beta_2 \ln a_{i,s,t} + \lambda_s + \gamma_t + \epsilon_{i,s,t+1}\,, \tag{2.4}$$

where $a^{'}_{i,s,t+1}$ represents the fourth-grade test score of child i who is enrolled in school s in school year t, while $a_{i,s,t}$ represents her third-grade test score, which is the first grade in which test scores data are available.[15] As before, our measure of base school quality $\ln \text{School Quality}_{n(s),t}$ represents the (log) average third-grade test scores of children living in s's (base) school catchment area. The regression model also includes school fixed effects (λ_s), as well as year fixed effects (γ_t). After a within-school transformation, the regression models is:

$$\Delta_s \ln a^{'}_{s,t+1} = \beta_1 \Delta_s \ln \text{School Quality}_{n(s),t} + \beta_2 \Delta_s \ln a_{s,t} + \delta_t + \Delta_s \epsilon_{s,t+1}\,, \tag{2.5}$$

[13]The drop in observations from column (1) to column (3) is because the model in first difference loses one year of data.

[14]Using the 2010 American Community Survey data for Wake County, we calculate that these families represent approximately 4 percent of households in Wake County, North Carolina. Along the same lines, our structural model features a set of households whose housing demand is affected by school quality and another set, which we label "non-families," whose demand is not.

[15]Ideally, we would like to use kindergarten or first-grade test scores. However, this information is not available in our data. We still believe our year-over-year analysis to be informative about the effects of differential peer exposure on skills given the cumulative nature of children's skill formation process.

where $\Delta_s \ln a^{'}_{s,t+1}$, $\Delta_s \ln a_{s,t}$, and Δ_s ln School Quality$_{n(s),t}$ represent the longitudinal change in the average fourth-grade scores in the school (between cohorts), the longitudinal change in the average in the school (between cohorts), and the change in the base school quality induced by changes in the base school catchment areas (as in Equation 2.2), respectively. The main parameter of interest in (2.5) is β_1, which represents the effect on children's learning of the changes in the neighborhood base school quality.

Panel B of Table 1 shows our results. In columns (1)–(3) we display the results for the fourth-grade test scores, while in columns (4)–(6) we report the results for fifth-grade test scores. For the pooled specifications (columns (1) and (4)), we find that a 10 percent increase in the base school quality is associated with an increase in fourth-grade and fifth-grade scores of 1.3 and 1.4 percent, respectively. Once we account for school-specific unobserved factors (school fixed effects), and we instead only exploit longitudinal variation in school quality induced by changes in catchment areas (columns (2)–(3) and (5)–(6)), we find that a 10 percent increase in base school quality induces an increase in fourth-grade and fifth-grade scores of 0.3–1 percent and 0.6–1.1 percent, depending on the specification. All our results should be interpreted as an intent-to-treat, as the change in the catchment areas of the neighborhood base school directly influences only children attending the base school. Finally, the second part of Panel B shows our results when we look at heterogeneous impacts of school quality by children's third-grade test scores. Our preferred specifications in columns (2)–(3) and (5)–(6) suggest that the largest benefits of improving school quality are observed among underperforming students (below median). This result is in line with previous findings on the effect of school quality on children's skills (see, e.g., Agostinelli et al., 2020b)

Transportation and the Demand for Schools. Although North Carolina law requires transportation to and from school for students who attend their base schools, option schools are not required to provide transportation for their students.[16]

We first consider whether changes in transportation services for local option schools affect the demand for base schools. The goal of this analysis is to shed light on whether families substitute away from base schools once option schools provide transportation. We test our hypothesis via the following regression model:

$$\pi_{n,t} = \alpha_0 + \alpha_1 \text{Transportation}_{n,t} + \lambda_n + \gamma_t + \eta_{n,t} \,, \tag{2.6}$$

where $\pi_{n,t}$ represents the fraction of children living in neighborhood n in year t who attend their assigned base school, while Transportation$_{n,t}$ is the number of option schools

[16]In our sample, 22 percent of option schools do not provide transportation.

that offer transportation to neighborhood n during the school year t. We also include neighborhood and time fixed effects to account for unobserved heterogeneity across space and time in the demand for schools. The error term is defined by $\eta_{n,t}$.

Specification (2.6) enables us to exploit longitudinal changes in school transportation services within neighborhoods to identify our main coefficient of interest, α_1, which is the elasticity of demand for base schools with respect to the provision of transportation to option schools. Panel A of Table 2 shows the results of our analysis. The even columns of the table include controls for family characteristics, such as distance from home to the city center, ED status, and total number of option schools available to the family's residential address. In Appendix A.1.2 we show that these changes in school transportation provision do not predict predetermined differences in school compositions, in terms of both children and families characteristics. Although this is not a formal test for anticipation effects and/or targeting policy effects based on unobservable characteristics, it reassures us that this source of identifying variation is balanced with respect to observable characteristics.

Column (1) shows that if an additional option school starts offering transportation, the probability that children living in that neighborhood attend their base school declines by 5 percentage points (approximately 4 percent of the mean base school enrollment rate). The results are robust to the inclusion of controls for family characteristics as shown in column (2).

Columns (3) and (4) of Panel A show the results when we also allow the effect of transportation on the demand for schools to vary by geographical distance. In particular, we are interested in understanding whether the substitution pattern between base schools and option schools caused by the new transportation options depends on the geographical distance of the additional option schools that offer transportation. We test this hypothesis by interacting our measure of local school transportation service with their (average) distance from the neighborhood.[17]

Our results highlight that geographical distance matters: our baseline results in column (3) show that one additional option school that provides transportation would cause the base school enrollment to decrease by 4.5 percentage points or 2.9 percentage points, depending on whether the option school is located five miles or ten miles away from home, respectively. The effect of a new option school becomes zero once that school is approximately twenty miles away. The results in column (4) resemble our baseline results.

We also analyze how the demand for option schools is affected by the their own trans-

[17]In the data, we measure the average distance from home addresses to schools in each school year.

Table 2: School Attended and Transportation Provision

	(1)	(2)	(3)	(4)
	PANEL A: BASE SCHOOL ENROLLMENT			
N Application Schools w Transportation	-0.023^{***}	-0.021^{***}	-0.061^{***}	-0.066^{***}
	(0.007)	(0.007)	(0.015)	(0.015)
N Application Schools w Transportation × Distance Schools (10 Miles)			0.032^{***}	0.039^{***}
			(0.011)	(0.011)
Observations	3,310	3,308	3,310	3,308
Neighborhood F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
	PANEL B: OPTION SCHOOL ENROLLMENT			
Transportation is Provided	0.068^{**}	0.064^{**}	0.130^{**}	0.121^{**}
	(0.029)	(0.029)	(0.049)	(0.051)
Transportation is Provided × Distance to School (10 Miles)			-0.051^{*}	-0.047^{*}
			(0.026)	(0.026)
Observations	8,340	8,340	8,340	8,340
Neighborhood F.E. × School F.E.	Yes	Yes	Yes	Yes
Year F.E. × School F.E.	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes

The table shows the effect of school transportation provision on school enrollment. Standard errors are robust to heteroskedasticity and reported in parentheses. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

portation provision. To do so, we specify the following regression model:

$$\pi_{s,n,t} = \beta_0 + \beta_1 \mathbb{1}(\text{Transportation})_{s,n,t} + \lambda_{s,n} + \gamma_{s,t} + \eta_{s,n,t}\,, \tag{2.7}$$

where $\pi_{s,n,t}$ represents the fraction of children living in neighborhood n in year t who attend option school s, while $\mathbb{1}(\text{Transportation})_{s,n,t}$ is a dummy variable for whether option school s offers transportation to neighborhood n during the school year t. We also include school-neighborhood and school-time fixed effects to account for unobserved heterogeneity across space and time for each option school.[18] $\eta_{s,n,t}$ is an error term.

We show the results in Panel B of Table 2. Column (1) shows that once an option school starts providing transportation, that school experiences an enrollment increase of 6.8 percentage points from the neighborhoods that newly received transportation. The results

[18] This specification is identified in our setting because option schools are associated with multiple neighborhoods while providing transportation only to a few of them. Within a neighborhood, the transportation arrangements of option schools change over time.

are robust to the inclusion of controls (column 2). Moreover—consistent with our previous results—geographical distance affects the demand for option schools when transportation is available. Columns (3) and (4) show that the attractiveness of transportation for option schools vanishes as the distance to that school increases, although the interaction term is only marginally significant.

Overall, we interpret the results in Table 2 as direct evidence of the role of transportation in facilitating access to educational opportunities for children. The effectiveness of transportation is mediated by geography, as providing transportation to schools that are further and further away is decreasingly effective at boosting enrollment.

3 Model

3.1 Overview

We model a city populated by a measure of *families* defined here as households with a child about to enter kindergarten. The model is static so we think of each time period as the year in which a new cohort of families decides where to live and enroll their children in school. Families are heterogeneous in terms of both their income and the children's skills. Families choose their neighborhood taking into account house prices, neighborhood-specific minimum housing constraints, and exogenous and endogenous amenities. The endogenous amenity is given by the portfolio of schools families associated with the neighborhood of choice. Schools are exogenously characterized by their capacity and location in the city, and they endogenously differ in quality and admission probability. Home–school distance determines the disutility cost from commuting, which may vary according to whether transportation is provided. Capacity constraints generate equilibrium admission probabilities to oversubscribed schools. School quality is determined by the skill composition of enrolled children, which in turn affects families' valuation for schools.

The city is also populated by *non-families*, by which we mean childless households and households with children outside the relevant age group. Non-families—in contrast to families—are only heterogeneous with respect to their income and do not value schools when choosing where to live. Although families are the main focus of our analysis, introducing non-families into the model allows us to consistently replicate in the model the empirical intention-to-treat analysis of the capitalization of school quality in housing prices given that families represent only a fraction of the housing demand in the city. Moreover, the presence of non-families is also relevant for the study of counterfactual experiments that trigger changes in the demand for neighborhoods and their consequences on house prices.

3.2 Environment

Demographics. The school district, or city, is populated by a measure $m+1$ of households. A measure 1 of households has one child who is about to start elementary school (families). A measure m of households does not have children in the relevant age group (non-families). Families are of type (w_p, a_k), where w_p is the household income and a_k is the child's skills. The joint distribution over family types is exogenously given by $\phi(w_p, a_k)$. For simplicity, we index a family type by (p, k). Non-family households have income w_p^* with distribution $\phi^*(w_p^*)$.

Geography. The city is made of N neighborhoods located on a two-dimensional surface. Neighborhood n has coordinates (n_x, n_y). Schools are denoted by s; they are also located inside the city and have geographic coordinates (s_x, s_y).

Housing Supply and Zoning. Each neighborhood is characterized by an exogenous total housing supply H_n. In addition, we model zoning restrictions on housing (e.g., minimum lot size) by allowing neighborhoods to differ with respect to the minimum housing size that can be built on them, h_n^{mls}.

School Boundaries. Each neighborhood is associated with a portfolio of schools, $\mathcal{L}_n$, that comprises three sets: $\mathcal{B}_n$,$\mathcal{T}_n$, and $\mathcal{NT}_n$.[19] The set $\mathcal{B}_n$ is a singleton that corresponds to the base school, in which admission is guaranteed and to which transportation is provided. The set $\mathcal{T}_n$ includes those schools that provide transportation to children who live in n, but admit them with probability p_s. Last, schools in $\mathcal{NT}_n$ do not provide transportation and also offer admission with probability p_s. We also define the catchment area of school s, $\mathcal{C}_s$, to be the set of neighborhoods n such that $s = \mathcal{B}_n$.

Admission Probability. Each school s is endowed with a certain number of seats for children in its catchment area. We assume that the number of available seats is sufficiently large to guarantee admission to all children that wish to attend their base school and all children that lose their application lottery described below. In addition, school s may offer a limited number of seats q_s to children who apply to it and are allowed to do so given their residential location but who are not in the school's catchment area. Specifically, the set of potential applicants is given by those children who live in neighborhood n such that

[19]Formally, $\mathcal{L}_n \equiv \mathcal{B}_n \cup \mathcal{T}_n \cup \mathcal{NT}_n$. See the discussion in Section 2.1.

$s \in \mathcal{T}_n \cup \mathcal{NT}_n$. If the number of applicants to school s exceeds its capacity, q_s, applicants are rationed through a lottery that determines who is admitted. Hence, the admission probability for a child in neighborhood n is equal to 1 if $s = \mathcal{B}_n$ and to $p_s \leq 1$ otherwise. If children are not admitted to the school they apply to, they are automatically enrolled in their base school.

Families' Preferences and the Technology of Skill Formation. We divide the choice of neighborhood and school into two sequential steps. Conditional on living in a given neighborhood n, families with children of skill k obtain the following utility from attending school s,

$$v_{k,s|n} = \widetilde{\gamma}_k \ln a'_{ks} - d_{ns}(\tau_{ns}) + \varepsilon_s,$$

where a'_{ks} is the next-period skills of a child attending school s, $d_{ns}(\cdot)$ is a utility cost from commuting, and ε_s is an idiosyncratic preference shock that follows a standard EV-Type 1 distribution. The disutility from commuting is given by the following function of the home–school road distance (in miles), τ_{ns},

$$d_{ns}(\tau_{ns}) = \begin{cases} \kappa_{1,T}\tau_{ns} & \text{if } s \in \mathcal{T}_n \cup \mathcal{B}_n \\ \kappa_{0,NT} + \kappa_{1,NT}\tau_{ns} & \text{o/w.} \end{cases}$$

That is, commuting to school entails a per mile cost, which may vary according to whether transportation is provided to that particular neighborhood, and a (relative) fixed cost of attending a school without transportation (e.g., car use).

In line with the previous child development literature, as well as our reduced-form evidence in Section 2.3, we define the technology of skill formation as follows:

$$\ln a'_{ks} = \zeta_0 + \zeta_1 \ln a_k + \zeta_2 \ln \bar{a}_s + \zeta_3 \ln a_k \ln \bar{a}_s, \tag{3.1}$$

where a_k represents the current stock of a child's skills, and $\bar{a}_s$ is the average peer skills of the school that she attends.[20] The advantage of this trans-log specification over some other commonly used functional forms in the literature—namely the constant elasticity of substitution and Cobb-Douglas specifications—is that it imposes no sign restriction on the complementarity or substitutability between peer quality and the initial stock of skills in the development of a child ($\frac{\partial^2 a'_{ks}}{\partial \bar{a}_s \partial a_k}$).[21] Using Equation in (3.1), we redefine the

[20]Our assumption that school quality is given by its children's skill composition is in line with previous work including Bayer et al. (2007), Epple and Romano (2003), and Avery and Pathak (2021), among others. We think of this measure as the policy-relevant one, since other determinants of school quality (e.g., good teachers) tend to be disproportionally attracted to schools with a positively selected pool of children (e.g., Jackson, 2009).

[21]Both Cobb-Douglas and CES production functions generate a non-negative cross-partial derivative,

value from attending school s as

$$v_{k,s|n} = \gamma_{0,k} + \gamma_k \ln \bar{a}_s - d_{ns}(\tau_{ns}) + \varepsilon_s\,,$$

where γ_k=$\widetilde{\gamma}_k\,(\zeta_2 + \zeta_3 \ln a_k)$, while $\gamma_{0,k}$= $\widetilde{\gamma}_k\,(\zeta_0 + \zeta_1 \ln a_k)$ represents a family specific constant that does not influence school choice.[22]

At the time of choosing between neighborhoods, families do not know the realization of their idiosyncratic preferences for schools. We assume that each family can apply to at most one school, and that they take admission probabilities p_s as given.[23] Recall that those who do not win the admission lottery are assigned to their base school $\mathcal{B}_n$. Therefore, the expected value of the portfolio of schools available to families in neighborhood n is equal to

$$\bar{v}_k(\mathcal{L}_n) = \mathbb{E}_{\{\varepsilon_s\}}\left[\max_{s\in\mathcal{L}_n}\{\hat{v}_{k,s|n}\}\right] = \mathbb{E}_{\{\varepsilon_s\}}\left[\max_{s\in\mathcal{L}_n}\{p_s v_{k,s|n} + (1-p_s)v_{k,\mathcal{B}_n|n}\}\right], \tag{3.2}$$

where $\hat{v}_{k,s|n} = p_s v_{k,s|n} + (1-p_s)v_{k,\mathcal{B}_n|n}$ and the expectation is taken with respect to the realization of idiosyncratic preferences.[24] This neighborhood-specific measure of educational access is reminiscent of value from commuting market access in Ahlfeldt et al. (2015).

Families' utility over neighborhoods is then represented by the utility function

$$U_{p,k,n}(c,h) = \psi\left[(1-\beta)\ln\left(\frac{c}{1-\beta}\right) + \beta\ln\left(\frac{h}{\beta}\right)\right] + \eta_p\alpha_n + \bar{v}_k(\mathcal{L}_n) + \varepsilon_n. \tag{3.3}$$

The first term, in square brackets, represents utility from consumption of a tradeable good c, that acts as numeraire, and housing h. The price of c is normalized to one, while house price is denote by r_n. The parameter ψ captures the relative importance of this bundle with respect to the other neighborhood attributes. The second term is the product of ex-

$\frac{\partial^2 a'_{ks}}{\partial \bar{a}_s \partial a_k} \geq 0$, which is a property that we reject in the data (see Section 4).

[22]For a given family, the difference in the values of two schools is: $v_{k,s'|n} - v_{k,s|n} = \gamma_k(\ln \bar{a}_{s'} - \ln \bar{a}_s) - (d_{ns'}(\tau_{ns'}) - d_{ns}(\tau_{ns})) + (\varepsilon_{s'} - \varepsilon_s)$.

[23]In the WCPSS, families submit an ordered ranking of at most three options and are assigned via the Boston mechanisms. A key feature of the Boston mechanism is that students who rank a school first get higher priority for that school than all other applicants. For most schools in the WCPSS, the number of first-choice applicants exceeds the number of seats (Dur et al., 2018), and therefore, most applicants are only considered for admission to their first choice (or the base school as their outside option), rendering all choices ranked below the first one irrelevant. For further discussion, see Appendix A.1.1.

[24]To avoid introducing additional notation, Equation (3.2) includes the trivial lottery in which a child applies to the base school and obtains value $\hat{v}_{k,\mathcal{B}_n|n} = p_s v_{k,\mathcal{B}_n|n} + (1-p_s)v_{k,\mathcal{B}_n|n} = v_{k,\mathcal{B}_n|n}$.

ogenous neighborhood amenities, α_n, and an income-specific valuation η_p.[25] The third term is the expected value from the portfolio of schools associated with neighborhood n and described in Equation (3.2). The last term is a neighborhood-specific standard EV-type 1 idiosyncratic preference, observed by families at the time of choosing their neighborhood.

Families maximize $U_{p,k,n}(c,h)$ subject to the budget and minimum housing constraints

$$\begin{aligned} \hat{w}_p = (1+\tau)w_p &\geq c + r_n h \\ h &\geq \underline{h}_n = \max\{h_0, h_n^{\text{mls}}\}. \end{aligned} \tag{3.4}$$

Families divide their income between consumption of the numeraire good and housing. The endogenous transfer τ corresponds to the land share of total housing expenditures, which we assume is rebated to households in proportion to their labor income w_p (see details below). The minimum housing constraint states that, absent regulation, house size must be larger than a certain amount h_0. We think of this minimum size as originating from essential space needs. In addition, the demanded house must also be at least as large as what the zoning restrictions dictate, h_n^{mls}.

Non-family Preferences. Non-family households have the same preferences and are subject to the same constraints as families, except for two differences. First, we allow exogenous amenities for a given neighborhood to be different between families and non-families—the latter being denoted by α_n^*. Second, we exclude all school-related variables from the utility of non-families. The resulting utility is then given by

$$\psi\left[(1-\beta)\ln\left(\frac{c}{1-\beta}\right) + \beta\ln\left(\frac{h}{\beta}\right)\right] + \eta_p\alpha_n^* + \varepsilon_n. \tag{3.5}$$

3.3 Equilibrium

The probability of choosing a lottery s for families of type k in neighborhood n is:

$$\pi_{s|n,k} = Pr\left[\hat{v}_{k,s|n} \geq \hat{v}_{k,\tilde{s}|n}\ \forall \tilde{s} \in \mathcal{L}_n\right],$$

[25] It is plausible that our measure of neighborhood amenity α_n includes both an exogenous (e.g., parks, bodies of water) and an endogenous component that is correlated with neighborhood income composition. Disentangling the relative importance of exogenous and endogenous amenities is challenging due to the need for exogenous sources of variation in—and detailed time-varying data on—neighborhood composition. We plan to explore the sensitivity of our results to the introduction of endogenous amenities in the future.

which does not have a closed-form expression.[26] [27]

Next, we consider the households' problem of allocating their income between consumption and housing. Such choice differs from the canonical constant expenditure share due to the presence of minimum housing constraints that vary by neighborhood. Therefore, the indirect utility over consumption and housing is given by

$$u_{np} = \begin{cases} \ln \hat{w}_p - \beta \ln r_n & \text{if } \beta \hat{w}_p \geq r_n \underline{h}_n \\ (1-\beta) \ln \left(\frac{\hat{w}_p - r_n \underline{h}_n}{1-\beta} \right) + \beta \ln \left(\frac{\underline{h}_n}{\beta} \right) & \text{if } \hat{w}_p > r_n \underline{h}_n > \beta \hat{w}_p \\ -\infty & \text{o/w} \end{cases} .$$

This indirect utility, different from most of the empirical random utility models, is derived from our micro-founded model accounting for both budget and housing regulation constraints. For this reason, the demand for housing in a particular neighborhood n displays non-linearities in the elasticity with respect to prices:

$$h_{np} = \begin{cases} \beta \hat{w}_p / r_n & \text{if } \beta \hat{w}_p \geq r_n \underline{h}_n \\ \underline{h}_n & \text{if } \hat{w}_p > r_n \underline{h}_n > \beta \hat{w}_p \\ 0 & \text{o/w} , \end{cases} \tag{3.6}$$

where in a given neighborhood n certain families spend a constant fraction of their income in housing ($\beta \hat{w}_p / r_n$), other families can only afford to live in the smallest dwelling permitted in this neighborhood according to housing regulations ($\hat{w}_p > r_n \underline{h}_n > \beta \hat{w}_p$), while other families are priced out entirely ($\hat{w}_p < r_n \underline{h}_n$). The shares of families in each of the three categories depend upon the equilibrium house price and local regulation. For given house prices, tighter zoning restrictions generate a higher fraction of families that are either priced out from the neighborhood or constrained in terms of house size. Our approach contrasts with standard random utility models with (log-)linear preferences over local prices (see for example Bayer et al., 2007; Ahlfeldt et al., 2015) since it features a price elasticity of housing demand that is a function of the level of prices, household income, and zoning regulations. This elasticity will differentially affect neighborhood demand across income groups in our counterfactual analyses of housing vouchers and regulation.

[26]Conditional on neighborhood n and child type k, preferences for schools do not vary with family income.

[27]The absence of a closed-form expression for choice probabilities when agents choose the lottery that maximizes their expected utility is a well-known feature of the empirical mechanism design literature. It escalates the computational burden of solving and estimating the model, as the choice probability (here, $\pi_{s|n,k}$) needs to be recovered by simulations. See, for instance, Agarwal and Somaini (2018); Calsamiglia et al. (2020); Luflade (2018); also see Agarwal and Somaini (2020) for a review of this literature.

The indirect utility over neighborhood n for families of type (p,k) is given by

$$x_{npk} = u_{np} + \eta_p \alpha_n + \bar{v}_k(\mathcal{L}_n).$$

It follows that the probability of living in neighborhood n is

$$\pi_{n|pk} = \frac{\exp(x_{npk})}{\sum_{\tilde{n}} \exp(x_{\tilde{n}pk})},$$

and the probability of living in neighborhood n and applying to school s is $\pi_{n,s|pk} = \pi_{n|pk}\pi_{s|n,k}$. Similarly, define the indirect utility for non-families as

$$x^*_{np} = u_{np} + \eta_p \alpha^*_n.$$

The choice probability for non-families is equal to

$$\pi^*_{n|p} = \frac{\exp(x^*_{np})}{\sum_{\tilde{n}} \exp(x^*_{\tilde{n}p})}.$$

In equilibrium, households' choices must be consistent with the endogenous value of house prices, r_n, transfer rate τ, admission probabilities p_s, and school peers $\bar{a}_s$. In particular, let

$$\begin{aligned}
\Pi_{np} &= \Sigma_k \pi_{n|pk}\phi(p,k) + m\pi^*_{n|p}\phi^*(p) \\
\Pi_s &= \Sigma_{p,k}\Sigma_{n\notin \mathcal{C}_s}\pi_{n,s|pk}\phi(p,k) \\
\Pi_{sk} &= \begin{cases} \Sigma_{p,n}\pi_{ns|pk}p_s\phi(p,k) & \text{if} \quad s \notin \mathcal{B}_n \\ \Sigma_{p,n}[\pi_{nB_n|pk} + \sum_{s\in\{T_n \cup NT_n\}} \pi_{ns|pk}(1-p_s)]\phi(p,k) & \text{if} \quad s \in \mathcal{B}_n \end{cases}
\end{aligned}$$

be the measure of households of type p who live in neighborhood n, the measure of applicants to school s, and the measure of children of type k that attend school s, respectively. While the first two variables are a straightforward aggregation of individual choices, the third one is a combination of individual choices and school lottery outcomes.

The market clearing condition for housing in neighborhood n reads

$$H_n = \sum_p \Pi_{np} h_{np}. \tag{3.7}$$

Let I denote total income in the economy. If a fraction μ of housing expenditure constitutes land rent that we assume to be distributed proportionally to household income, we obtain

$$\tau = \frac{\mu(r_n \cdot H_n)}{I}. \tag{3.8}$$

Admission probabilities are either equal to 1 if the school has enough seats to accommodate all applicants, or some value less than 1 if there is rationing due to oversubscription. Formally,

$$p_s = \min\left\{\frac{q_s}{\Pi_s}, 1\right\}. \tag{3.9}$$

The quality of peers is given by the average quality of children that attend a certain school,

$$\bar{a}_s = \mathbb{E}_{\Pi_{k|s}}[a_k], \tag{3.10}$$

where $\Pi_{k|s} = \Pi_{sk}/\Sigma_{\tilde{k}}\Pi_{s\tilde{k}}$ is the conditional distribution of children of type k who attend school s.

We are now ready to define an equilibrium for this economy.

Definition. An equilibrium for this economy is a set of choice probabilities for families, $\{\pi_{ns|pk}\}$, and non-families, $\{\pi^*_{n|p}\}$, aggregate variables $\{r_n\}$, $\{\bar{a}_s\}$, $\{p_s\}$, and transfer rate $\{\tau\}$ such that:

- given aggregate variables, $\{\pi_{ns|pk}\}$ and $\{\pi^*_{n|p}\}$ are choice probabilities induced by i) families' solutions to the school choice problem (3.2), and choice of consumption, housing, and neighborhood to maximize the objective function (3.3), subject to constraints (3.4); and ii) non-families' maximization of the objective function (3.5), subject to constraints; (3.4)
- the aggregate variables satisfy housing market clearing (3.7), school capacity constraints (3.9), and consistency of school composition, (3.10);
- the transfer rate is consistent with the land share of housing expenditure as in (3.8).

4 Model Estimation

We divide our set of parameters into three groups. The first one is calibrated following the existing literature, reduced-form evidence, or direct empirical observations. We also clarify some necessary normalizations in households' utility functions. The second group of parameters estimated a first step, before the estimation of the rest of the model. The third group of parameters is estimated inside the model by the method of simulated moments. The latter group is the main focus of our analysis as it sheds light on the determinants of school and neighborhood choice and their heterogeneity across family types.

Parameters Set Outside the Model. We set the housing utility parameter, β, and the land share in housing expenditure, μ, both equal to 0.25, in line with traditional estimates from the urban economics literature (e.g., Ahlfeldt et al., 2015). We map our measure of housing regulation in the data into the minimum housing constraint in the model. To do so, we regress the empirical minimum house size (in square feet) in each neighborhood on the ordinance-prescribed neighborhood-level measure of minimum lot size (in acres), mls_n.[28] The implied minimum house size is $\underline{h}_n = h_0 + 892 \times \text{mls}_n$, where $h_0 = 641$. Since mls_n spans from 0 to about 1 acre, the implied minimum housing constraint is up to about 1,500 sq. ft. in some neighborhoods. The relative measure of non-families is set to $m = 24$. We think of families in the model as those households with children who are four or five years old, which represent 4 percent of households in Wake County in the ACS.[29] We divide children types into deciles (of the skills distribution, $T_k = 10$) and household income into $T_p = 16$ bins, given the level of disaggregation available in the ACS. We also use the ACS to compute the empirical distribution of non-family types, $\phi^*(w_p)$, and we combine the administrative school data with the ACS to compute the distribution of family types, $\phi(w_p, a_k)$. We report details on the construction of our measures of households in Appendix A.6.

Parameters Estimated Outside the Model. We estimate the technology of skill formation in (3.1) in a first step outside our simulation-based estimation algorithm. This choice has at least two advantages: (i) it simplifies our estimation procedure while reducing its computational burden; and (ii) it is robust to model misspecification, as it is less sensitive to specific assumptions on preferences, timing, and expectations/information sets. We estimate the technology in (3.1) using an instrumental variable estimator. In particular, we exploit the previously defined exogenous variation in potential school peers induced by the changes in school catchment areas as an instrument for the endogenous quality of students attending a given school (see Section 2.3). The implied exclusion restriction within this IV framework is consistent with our model: changes in school boundaries alter school composition, which in turn affects children's learning via peer effects. Our outcome of interest is children's fourth-grade test scores.

[28] For each neighborhood, we calculate the minimum house size for newly constructed houses in the data, and we regress it on the ordinance-prescribed minimum lot size observed in each neighborhood. We use the regression results to predict the implied minimum house size for each neighborhood. We use data on newly constructed houses only to avoid the problem that certain dwelling units were constructed before the zoning regulations were in place.

[29] We calculate this statistic using the 2010 ACS data. We use the variable "Number of own children under age 5 in household" to calculate the percentage of households in Wake County with at least one child under age five.

Parameter estimates for the technology of skill formation are shown in column (2) of Table B-2. First, we find that the process of skill formation in children is highly persistent, with a coefficient on own past skill, ζ_1, of 0.84. This implies that, for the average level of log-Peers' Skills in our sample (0.39), a 1 percent increase in one's current stock of skills translates into an increase of 0.82 percent in one's next-period skills. Moreover, we find a positive and significant effect of peer quality: the estimated parameter ζ_2 is 0.40, which implies that a 1 percent increase in peer quality for a child with an average initial endowment of log-skills (normalized to zero) translates into an increase of 0.4 percent of her next-period skills. Finally, the estimated interaction term of the trans-log specification (ζ_3) is -0.064, which suggests that peer effects are stronger for disadvantaged children (lower initial skill levels), consistently with our reduced-form results in Panel B of Table 1. Quantitatively, this means that the same 1 percent increase in peer quality would translate into an increase of next-period skills of 0.52 percent and 0.28 percent for children who are two standard deviations below and above the mean of the initial skill distribution, respectively (see Figure B-1).

4.1 Targeted Moments and Identification

The set of parameters we estimate is given by transportation cost, $\{\kappa_{1,T}, \kappa_{0,NT}, \kappa_{1,NT}\}$; preferences over school peers, γ_k; valuation for housing consumption ψ; exogenous neighborhood amenities for families and non-families, α_n and α_n^*; marginal valuation of amenities by income, η_p; housing supply $H_{n,t}$ for each neighborhood n and year t; and capacity $q_{s,t}$ for each school s and year t.

Although in our equilibrium model there is no one-to-one mapping between parameters and moments, we present the intuition behind the identification argument that guides our estimation procedure. Details about the computation of moments are shown in Appendix B. Standard errors are computed using the delta method. We target the (i) average share of children who attend schools that do not provide transportation, (ii) the distance to school conditional on transportation being provided, or (iii) transportation not being provided. These three moments are particularly informative of the (relative) fixed cost of attending a school without transportation and the per mile cost of attending a school with or without transportation, $\{\kappa_{0,NT}, \kappa_{1,T}, \kappa_{1,NT}\}$.

The parameters γ_k characterize preferences over school peer composition by child type k. We estimate the vector γ_k to replicate the average school (peer) quality attended by children of type k. Notice that differences in preference for school quality across children types is only responsible for the residual heterogeneity in school quality that is not accounted for by neighborhood sorting along family income, combined with the empirical correlation between family income and children types. Although relative valuations for school quality can be identified from variations in school composition, the overall level of the γ_ks cannot. Intuitively, appropriately increasing (or decreasing) the value of all γ_ks would preserve the heterogeneity in school valuations that allows us to match the observed variation in school composition. Therefore, we complement our targeted moments by replicating the response of house prices to changes in school quality induced by the redesign of the school portfolio, as reported in column (2) of Table 1.

The parameter ψ is identified by the correlation across neighborhoods between minimum lot size, mls_n, and the share of neighborhood households with less than median income—that is, those households for whom minimum housing restrictions are binding in at least some neighborhoods at the observed prices. Intuitively, stricter zoning is particularly costly for low-income households because they are more likely to be constrained in their housing choice.[30] Therefore, the higher the valuation for housing consumption, captured by ψ, the more zoning restrictions reduce neighborhood choice by constrained families. The amenities α_n and α_n^* are set to replicate the empirical distribution of families and non-families across neighborhoods. Intuitively, amenities capture the incentives to reside in certain locations after accounting for observable neighborhood attributes like school quality, house prices, and zoning restrictions.

The vector of parameters η_p captures heterogeneity in valuation for amenities by income type and generates (residual) income sorting across neighborhoods. Analogously to γ_k, the vector η_p is set to match the average neighborhood income for households of type p. Parameter η_1, for the first income bracket, is normalized to one. The mean amenities over neighborhoods—$\mathbb{E}[\alpha]$ and $\mathbb{E}[\alpha^*]$—are normalized to zero. The estimated housing supply $H_{n,t}$ allows the model to replicate the average equilibrium house prices in each neighborhood over the estimation sample and the average house price across neighborhoods in each year.[31] Last, application school capacity $q_{s,t}$ is such that the model matches the empirical admission probability in each application school s and year t.

4.2 Estimates

Estimates for commuting cost parameters, the utility weight on housing, families' valuation of school quality as a function of their child's skills (decile), and families' valuation of neighborhood amenities as a function of their income are shown in Table 3. Figure

[30] For households that are unconstrained everywhere, higher values of ψ generate lower demand for more-expensive neighborhoods, but with the same elasticity across income levels.

[31] Formally, the housing supply satisfies $\ln H_{n,t} = \ln H_n + \ln H_t$. Equilibrium house prices in the model are matched to per period housing payments in the data. See Appendix A.6 for the construction of per period payments from housing sales data.

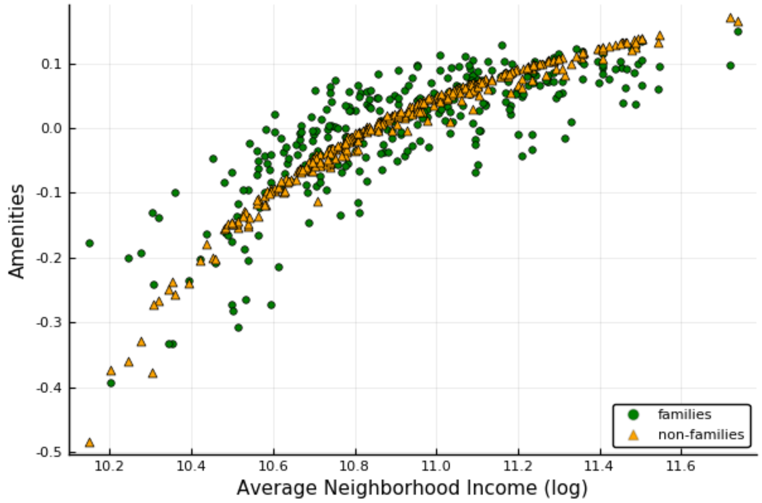
Table 3: Estimated Parameters

Parameter	(1) Estimate	(2) Std. Error	Parameter	(3) Estimate	(4) Std. Error
Disutility From Commuting			*Preference For Nbhd Amenities*		
$\kappa_{0,NT}$	2.22	0.13	η_2	0.99	0.55
$\kappa_{1,NT}$	0.26	0.01	η_3	1.11	0.28
$\kappa_{1,T}$	0.25	0.01	η_4	1.06	0.26
Preference For School Quality			η_5	0.94	0.45
γ_1	0.02	0.41	η_6	1.50	0.30
γ_2	0.73	0.40	η_7	1.57	0.31
γ_3	0.91	0.41	η_8	1.18	0.43
γ_4	1.08	0.41	η_9	1.50	0.26
γ_5	1.46	0.42	η_{10}	1.46	0.32
γ_6	2.12	0.43	η_{11}	1.68	0.25
γ_7	2.45	0.43	η_{12}	1.94	0.27
γ_8	2.87	0.45	η_{13}	2.52	0.18
γ_9	3.60	0.44	η_{14}	2.88	0.12
γ_{10}	4.96	0.39	η_{15}	3.74	0.06
Utility From Housing Consumption			η_{16}	5.59	0.07
ψ	9.42	39.12			

Column (1) shows estimated values for commuting cost parameters, the utility weight on housing consumption. and preferences for school quality by children's skill type. Associated standard errors are shown in column (2). Column (3) shows estimated valued for households' preferences for neighborhood amenities by income type. Associated standard errors are shown in column (4). Recall that the fixed cost of commuting with transportation ($\kappa_{0,T}$) is normalized to zero and that the preference for neighborhood amenities is normalized to one for the lowest income group (η_1).

1 shows estimated neighborhood amenities and plots them against average income in the neighborhood. We estimate that the absence of school transportation entails a large fixed cost, equivalent to the disutility of about ten additional miles. The relationship between families' valuation for school quality and their child's skills level is monotonically increasing, meaning that families with higher-skilled children value better schools more. Families' valuation for neighborhood amenities is estimated to be increasing and convex in family (log) income. Neighborhoods not only differ from each other in terms of exogenous neighborhood amenities and school portfolios but also in terms of zoning regulations. Tighter minimum housing constraints at the neighborhood level induce a smaller presence of low-income families, with an elasticity that is governed by the utility weight on housing. High-income households' strong preference for neighborhood amenities, along with minimum house size regulations, disproportionately restrict location choice for low-income households and contribute to income segregation across neighborhoods. In line with this heterogeneity, neighborhoods with higher estimated amenities tend to be

Figure 1: Estimates of Neighborhood Amenities by Neighborhood Income



The figure shows estimated neighborhood amenities (α_n for families, and α_n^* for non-families) on the y-axis, plotted against neighborhood (log) income (x-axis).

disproportionately chosen by high-income families.

To help understand the magnitude of the estimated parameters, we show in Table 4 how demand for schools and neighborhoods is affected by variations in the key determinants of choices. Specifically, Panel A shows how, conditional on residential neighborhood, demand for the base school changes for families with a high- or low-skills child, as the characteristics of schools in the neighborhood's portfolio change, holding everything else equal. Panel B shows how demand for the median-amenity neighborhood changes for high- and low-income families, with a high- or low-skilled child, as its school portfolio and zoning restrictions change, holding everything else equal.

Panel A in Table 4 illustrates the effect of distance, availability of school transportation, and school quality on school choice as residential neighborhood is held fixed. Given transportation provision to the base school, increasing the distance from the neighborhood to its base by one mile (a 25 percent increase from the average neighborhood-to-base distance in the sample) decreases the probability that the residents of a neighborhood choose the base as their most preferred school by 3.7 percent.[32] Removing transportation to option schools, while providing it to the base, on average increases the probability that residents of a neighborhood prefer their base school over all other schools in their portfolio by 4–7 percent. Indeed, as a consequence of removing transportation to options schools, children with skills below the median mainly substitute toward the base school,

[32]The constant decrease across skill groups is expected, given that commuting-cost parameters are set to be the same across groups in the model (see estimates in Table 4). Estimating a model in which commuting cost are allowed to differ by income (ED/non-ED) yields similar estimates for both groups.

Table 4: Changes in Choice Probabilities as School and Neighborhood Characteristics Change

	Below-median child skills		Above-median child skills	
	ED families	Non-ED families	ED families	Non-ED families
PANEL A: DEMAND FOR (BASE) SCHOOL				
Dist. to base increased by 1 mile	−3.7		−3.7	
Transp. to options removed	+6.5		+4.1	
Base quality decreased by 10%	−1.4		−4.5	
PANEL B: DEMAND FOR (THE MEDIAN-AMENITY) NEIGHBORHOOD				
Dist. to base increased by 1 mile	−16.5	−16.6	−16.9	−16.9
Transp. to options removed	−9.9	−9.6	−6.8	−6.2
Base quality decreased by 10%	−5.6	−6.9	−18.4	−20.9
Adm. prob. reduced by half	−10.0	−10.5	−9.4	−10.8
Zoning restriction increased by 10%	−5.6	0	−5.6	0

Panel A shows the percentage change in baseline choice probabilities, given residential neighborhood, conditional on a child's skills relative to the median, for the base school associated with the residential neighborhood, after and before the change described in the left-most column, holding all other things constant. Panel B shows the percentage change in baseline choice probabilities, conditional on family income (qualifying as ED or not) and a child's skills relative to the median, for the median-amenity neighborhood after and before the change described in the left-most column, holding all other things constant.

while children with higher skills, whose valuation for school quality is higher (Table 3), are more likely to seek higher-quality, non-transportation option schools. In line with this idea, decreasing the quality of the base school by 10 percent translates into a decrease in choice probability by 1.4 percent for low-skilled and 4.5 percent for high-skilled children. Panel A of Table 4 only illustrates part of families' responses to changes in school characteristics when families choose their school and neighborhood jointly because changing the characteristics of the schools in a neighborhood's portfolio also affects families' probability of choosing that neighborhood in the first place. Comparing Panel A to the first three rows of Panel B illustrates how a significant part of families' response to policies affecting school characteristics would be missing from a model that takes residential location as given. In terms of neighborhood characteristics, the last row of Panel B illustrates the role of constraints imposed by zoning regulations. Increasing the minimum lot size by 10 percent decreases demand for the neighborhood only for ED families (by 5.6 percent) while leaving high-income families unaffected.

4.3 Model Fit and Validation

4.3.1 In-Sample Fit

Table B-1 in the Appendix shows the model in-sample fit. Notice first that the model is able to replicate the residential income segregation in the city. Although the model replicates patterns of residential income segregation, the heterogeneity in neighborhood school portfolios and the presence of heterogeneous child skills for a given family income create a wedge between neighborhood and school composition. However, thanks to the estimated heterogeneity in valuation for school quality, the model also matches the empirical evidence on school peer composition. For example, children in the highest decile of the skill distribution are on average attending schools where peers' skills are 47 percent higher than the schools attended by children in the lowest decile. In terms of school choice and transportation, we observe that both distance and transportation play a role in determining school choice: both in the model and in the data, the average distance to the school attended is approximately 3.5 miles and 6.9 miles for children with and without transportation, respectively. In addition, the model replicates the negative correlation (-0.23) between zoning and the shares of families in the neighborhood who are below the median income of the economy. This negative correlation is explained by the fact that zoning effectively distorts the housing demand for low-income families by generating barriers to entry in highly regulated neighborhoods.

Finally, we replicate within the model, as part of our targeted moments in the estimation, the regression in (2.3). We regress longitudinal changes in house prices on the constructed measure of policy-induced changes in school quality. These exogenous changes in school quality are constructed—as we did in the data—by using changes in peer quality in the assigned base schools that were induced by changes in the base school catchment areas only. Panel A in Table 5 shows that the model exhibits similar behavioral responses of housing demand to changes in school quality: a 10 percent change in school quality induced only by changes in the school catchment areas causes—both in the model and in the data—an increase in house prices of about 0.3 percent.

4.3.2 Validation

Our final goal is to validate the behavioral responses of the estimated model in terms of demand for schools and neighborhoods, which are the key endogenous margins that will drive the results of our subsequent counterfactual analyses. For this reason, we assess quantitatively the model's ability to replicate some of the informative quasi-experimental reduced-form estimates on demand for transportation we presented in Section 2.3, as well as some of the evidence on neighborhood choice from the MTO experiment.

Replicating Elasticity of Demand for School with Respect to Transportation Provision. We want to validate a key model statistics—the elasticity of demand for schools with respect to transportation. This is an important margin to validate for our counterfactual analysis, as one of the policy tools to reduce inequality in educational access is to expand

Table 5: Quasi-Experimental Estimates: Model vs. Data

	(1)	(2)	(3)	(4)
	Model		Data	
PANEL A: HOUSE PRICES (TARGETED)				
Changes In School Quality (Log)	0.03		0.03	
PANEL B: BASE SCHOOL ENROLLMENT (UNTARGETED)				
N Application Schools w Transp	−0.026	−0.047	−0.023	−0.061
N Application Schools w Transp × Distance to Application Schools (10 Miles)		0.024		0.032
PANEL C: APPLICATION SCHOOLS ENROLLMENT (UNTARGETED)				
Transportation is Provided	0.046	0.056	0.068	0.130
Transportation is Provided × Distance to Application Schools (10 Miles)		−0.016		−0.051

The table compares quasi-experimental estimates obtained from the data (columns (3)–(4), reproduced from Tables 1 and 2 and analogous coefficients obtained from regression of model-predicted house prices on model-predicted changes in school quality (Panel A), and of model-predicted enrollments on transportation provision indicators (Panels B and C).

transportation provision within the city. For this reason, we compare the results in our model to the same regressions we run in Table 2.

The model broadly replicates—without being targeted in estimation—the reduced-form elasticities of school demand with respect to transportation observed in the data. Panels B and C in Table 5 show the comparison of the regression coefficients in the model and in the data. As previously discussed, the estimated effects are identified by exploiting the within-neighborhood (Panel A) and within-school (Panel B) longitudinal variation in transportation provision of available option schools. Consistent with the data, the model displays the following properties for the demand for schools: (i) families substitute away from their base school once local option schools start providing transportation; (ii) this substitution pattern is stronger when the option school is geographically closer; (iii) the demand for an option school increases once that school starts providing transportation in the neighborhood; and (iv), this increase in demand depends (negatively) on the geographical distance to that school.

Replicating MTO Experimental Findings. The second key model response we want to validate relates to the elasticity of neighborhood choices with respect to income. This feature of the model is important both because it affects the residential income segregation implied by the model and because it directly affects the residential responses to the counterfactual analysis of the voucher policy we discuss in the next section.

The Moving to Opportunity (MTO) experiment represents a valuable data source to validate our model. The Department of Housing and Urban Development, in collaboration with the public housing authorities of five major cities (Baltimore, Boston, Chicago, Los Angeles, and New York), created two treatment arms by randomly assigning housing subsidies to eligible households living in poor neighborhoods. The first group received an unrestricted (in terms of residential constraints) Section 8 housing voucher. We refer to this group as *Section 8*. The second group (*Experimental*) received instead a constrained housing voucher that could be used only in neighborhoods with low poverty rates (below 10 percent). The goal of this experiment was to evaluate the effects on neighborhood change on various households' socioeconomic outcomes.

We base our validation exercise on the results in Galiani et al. (2015) for the MTO experiment in Boston. The authors show the experimental estimates on the neighborhood poverty rates for both *Section 8* and *Experimental* groups. We run the same experiment in our model and find that it replicates well the behavioral responses they document: when families are restricted in terms of destination neighborhoods, they move to neighborhoods that display on average a 6 percent poverty rate (7 percent in our model). On the other hand, when families are free to choose, they decide to use the voucher to live in neighborhoods that display on average 20 percent poverty rate (19 percent in our model).

Overall, we find that the model displays behavioral responses in terms of the demand for housing and schools in line with experimental and quasi-experimental evidence in the data. We interpret this as a validation of some of the key elasticities in the model that will govern our counterfactual analyses. Equipped with these results, we now turn to the analysis of our three policy counterfactual exercises.

5 Policy Counterfactuals

Our structural estimates highlight the key determinants of households' neighborhood and school choices. They also stress the importance of costs and constraints in shaping the heterogeneity in choices made by high- and low-income households when it comes to residential neighborhood and schools for their children. In particular, commuting costs in the absence of school transportation, while the same across income levels, can be a source of unequal access to high-quality schools if these tend to be located closer to high-income neighborhoods or if no transportation is provided to lower-income neighborhoods. Our estimates also show that zoning regulations that mandate a minimum house size may prevent low-income households from accessing certain neighborhoods and therefore the schools these neighborhoods provide exclusive or convenient access to.

In this section, we further explore the role of local institutions in shaping inequality of educational opportunities. In light of our structural estimates, we focus on two key margins. On the one hand, we explore the extent to which expanding school choice enables low-income families to access high-quality schools without the need to change their residential location. On the other hand, we provide low-income families with the opportunity to reside in neighborhoods that grant access to high-quality schools by addressing the constraints that prevent them from living in these neighborhoods: high house prices and tight (highly restrictive) zoning. Toward this aim, we evaluate three distinct policies. The first policy focuses on the first margin and expands school choice in neighborhoods in the catchment area of low-quality base schools to allow residents of these neighborhoods to attend one of the highest-quality schools in the district. To highlight how transportation provision limits access to these additional options, we implement two versions of this policy in which transportation to top schools is or is not provided. The last two policies consider the second margin. Our second counterfactual policy, designed after the MTO program, is a family-based policy that grants a voucher to low-income families in high-poverty neighborhoods and covers the fair market rent for them to live in low-poverty neighborhoods. Last, we implement a place-based policy by relaxing minimum housing regulations in tightly regulated high-income neighborhoods. In each of the counterfactual policy exercises, we assume that the base schools capacity is adjusted as to guarantee admission to all students in their catchment area. Every other institutional feature of the city (e.g., school boundaries) is kept constant as in 2006, our benchmark year.

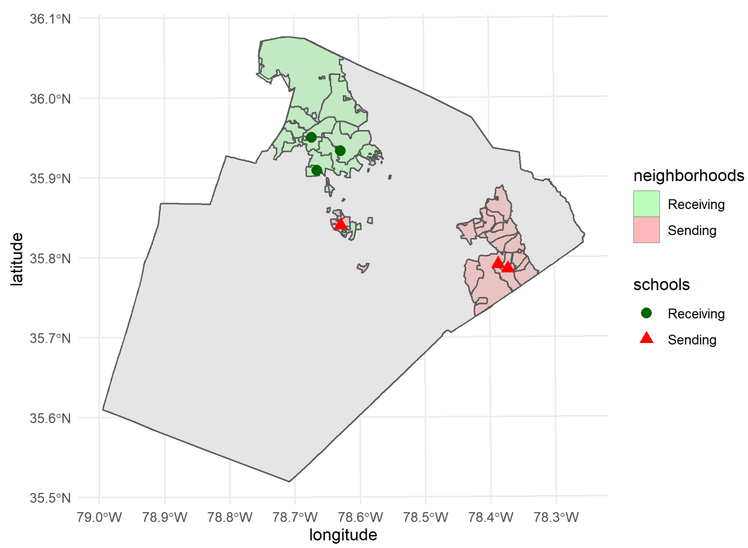
5.1 School Choice Expansion

5.1.1 Policy Design

We identify the three lowest-quality schools and the three highest-quality schools in the district. We pair all the neighborhoods that share the three lowest-quality schools as their base school to one of the high-quality schools, and allow the residents of the low-quality school neighborhoods to apply to the newly offered high-quality school. For convenience, we refer to the three low-quality base schools as *sending* schools and to the neighborhoods in their respective catchment areas as *sending* neighborhoods. Conversely, we refer to the three higher-quality base schools as *receiving* schools and to the neighborhoods in their catchment areas as *receiving* neighborhoods. Figure 2 shows the locations of sending and receiving schools and neighborhoods.[33]

[33] Table C-1 shows the key characteristics of the sending and receiving schools and neighborhoods. By construction, receiving neighborhoods have significantly better schools and children skills. They also feature around 40 percent higher neighborhood income house prices than sending neighborhoods.

Figure 2: School Choice Expansion: Sending and Receiving Schools and Neighborhoods



The figure shows the distribution in the map of Wake County of the sending (red areas) and receiving (green areas) neighborhoods that are part of the school choice expansion counterfactual policy. The red triangles and the green dots represent the sending and receiving schools, respectively.

We assume that the capacity of each of the three receiving schools is increased by 10 percent and that the additional capacity is reserved for applicants from the associated sending neighborhoods. If the receiving school is oversubscribed—that is, if the number of applicants from sending neighborhoods exceeds the number of seats—then applicants from sending neighborhoods are selected by lottery (while students from receiving neighborhoods are always guaranteed admission to their base school).[34] We analyze this policy under two scenarios, specifically, with and without transportation between the sending neighborhoods and their receiving school. Finally, to assess the importance of endogenous residential sorting for this type of educational policies, we replicate our counterfactual analysis while holding neighborhood choices constant at their baseline values.

5.1.2 Results

Key results under our model of endogenous neighborhood choice are shown in Panel A of Table 6. The first two columns show results when including transportation provision in the implementation of the school choice expansion; the last two columns show results when school transportation is not offered. Of the three receiving schools, one is much

[34]To keep the size of each neighborhood's school portfolio the same as in the baseline economy, we remove the option school with lowest average valuation from the portfolios of sending neighborhoods. The removed schools have minimal school enrollment from sending neighborhoods in the baseline economy.

Table 6: Effects of School Choice Expansion on Sending and Receiving Neighborhoods

	Transportation provided		No transportation	
	Close Pair	Farther Pair	Close Pair	Farther Pair
Dist. to receiving school (miles)	7.17	24.47	7.17	24.47
PANEL A: NEIGHBORHOOD CHOICE				
Sending schools and neighborhoods				
Student take-up rate (%)	19.2	3.7	5.8	0.3
Share attending receiving base (%)	13.2[a]			
Change in base quality (%)	+14.6	+0.5	0	0.1
Receiving schools and neighborhoods				
Admission probability	0.686	1	1	1
Change in base quality (%)	−20.3	0	−1.2	0
PANEL B: FIXED NEIGHBORHOODS				
Sending schools and neighborhoods				
Student take-up rate (%)	22.0	3.1	5.2	0.3
Change in base quality (%)	−6.2	−5.7	−7.4	−0.6
Receiving schools and neighborhoods				
Admission probability (%)	0.744	1	1	1
Change in base quality (%)	−4.1	−0.1	−0.7	0

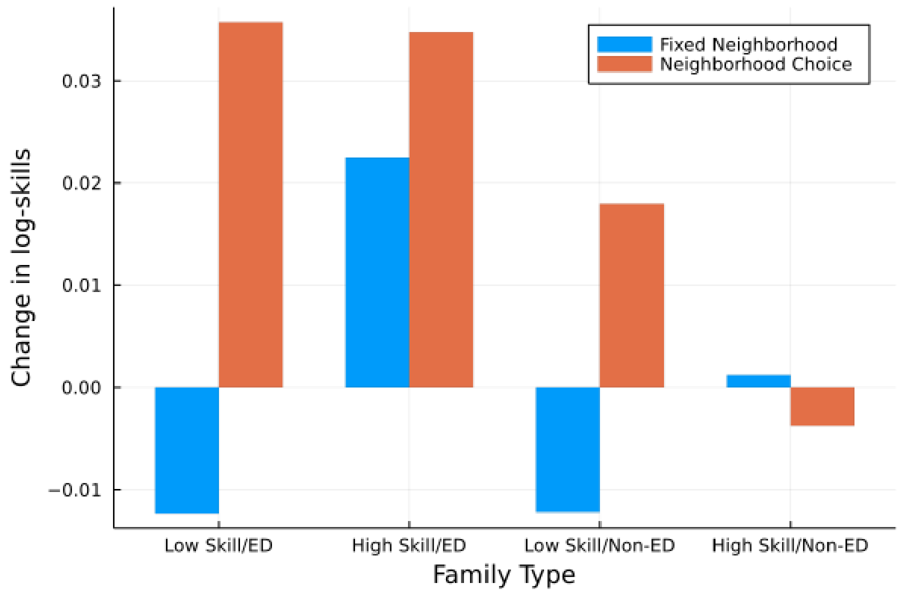
The table shows the changes (from the baseline equilibrium) induced by the school choice expansion policy on sending and receiving neighborhoods. "Close Pair" refers to the pair of sending and receiving base schools located relatively close to each other (about seven miles); "Farther Pair" refers to the pair of sending and receiving base schools located far each other (about 25 miles).

[a] The share of students attending the receiving base is the product of the share applying (student take-up rate) and the admission probability; here: $13.3 = 19.4 \times 0.69$.

closer to the sending neighborhoods than the other two (seven miles versus more than 20 miles; see top row in Table 6). Under each implementation scenario, results are shown for the closest pair ("Close Pair") in the first column, and the farthest pair ("Farther Pair") in the second column (results for the omitted pair are similar to the farthest). Table C-2 in the appendix provides additional details about the results.

Without Transportation. Focusing first on the last two columns of Panel A of Table 6, we find low take-up rates when school transportation is not provided to the newly offered schools. None of the receiving schools is oversubscribed (admission probabilities all equal to 1), and less than 6 percent of children living in sending neighborhoods attend the high-quality receiving school. Note that this rate is highest for the Close Pair, in which the receiving school is relatively close to the sending neighborhoods. As a consequence of the low value placed by households on the school choice expansion, neighborhood choice is virtually unaffected. Symmetrically, because of the very low take-up in sending neighborhoods, the quality of the receiving base schools and, in turn, the composition of the receiving neighborhoods remain roughly constant.

Figure 3: School Choice Expansion (with Transportation): Effects on children skills



The graph shows the changes (from the baseline equilibrium) induced by the school choice expansion policy (with transportation) on children skills under two model's specifications: with endogenous neighborhood choice (orange) or with fixed neighborhood choice as at baseline (blue). The results are broken down by economically disadvantaged (ED) status and baseline skill level.

With Transportation. Offering school transportation along with expanded school choice results in a very different picture, especially for our Close Pair (first column). More children from these sending neighborhoods apply to their receiving school than there are seats available for them under the policy, resulting in a 0.64 admission probability conditional on applying; therefore, 19 percent of children from these sending neighborhoods apply for admission while only two-thirds of them end up attending the receiving school. The high value of the added school option increases the demand for sending neighborhoods in the Close Pair (average house prices rise by 0.3 percent), particularly from families with high-skilled children who value school quality more (average children's skills in the sending neighborhood increase by 17.3 percent) and which tend to have higher household income (average income increases by 3.3 percent). In that regard, our model generates results that resemble the so-called gentrification documented by Billings et al. (2018): when the catchment area of low-performing base schools is given higher priority to attend high-quality option schools, neighborhood average income and house prices increase while the probability of attending the initial base school decreases. In the other two pairs of neighborhoods, results are more muted. While take-up is higher than in the no-transportation experiment, it remains low (receiving schools remain undersubscribed) because of the disutility from distance between the receiving school and sending neighborhoods.

In Figure 3 we show the effects of this policy on children's skills. Since we are analyzing a place-based policy, we compute the average effect on skills by family type, where the average is taken with respect to the endogenous distribution of families of a given type over neighborhoods and schools in equilibrium. To relate the magnitude of the outcome to the exposure of different family types to the policy, we normalize (divide) the change in skills for families of a given type by the share of those families that live in either sending or receiving neighborhoods at baseline.

The orange bars show the policy results when we compute the new equilibrium of the economy, where individuals adjust their neighborhood and school choices. Overall, the policy positively affects children of disadvantaged households, whose skills increase approximately by 3.6 percent of a standard deviation. The school choice policy is effective in improving educational outcomes for disadvantaged children not despite gentrification—one of the major concerns in the public debate—but because of it. In our model, gentrification amplifies the positive effects of the place-based policy by attracting high-skilled children to low-achieving neighborhoods, some of whom attend the local base school upon losing the admission lottery for the high-quality receiving school.[35]

The effects on children of higher-income families are mixed. On the one hand, children who originally lag behind (below-median initial skill) benefit from the policy (+1.9 percent). On the other hand, children who are originally above the median skill distribution experience a negative impact of 0.37 percent. This heterogeneity of the policy impacts is explained by differences in neighborhood choices as the latter families are relatively more likely to be in receiving neighborhoods where the policy has a detrimental effect on school peer composition.

The Role of Neighborhood Choice. To assess the importance of the endogeneity of residential choices, which is typically ignored in the education literature (see, e.g., Abdulkadiroğlu et al., 2017; Agarwal and Somaini, 2018) when analyzing school choice policies, we repeat our counterfactual analysis but hold households' residential choices constant at their baseline values. That is, families cannot relocate after the change in portfolios in sending neighborhoods; they need to choose a school for their child given the school portfolio attached to their baseline neighborhood.

Results are shown in Panel B of Table 6 and Figure 3 (blue bars). Table 6 shows that under the school choice expansion, only those who highly value the added receiving school

[35]We rule out the hypothesis that the positive effect on the local base school is due to the timing of the shock, that is, to families with high-skilled children applying to the local base schools due to idiosyncratic preferences that are realized after the neighborhood is chosen. We find that while the composition of *attendees* at the local base school significantly improves, the composition of *applicants* does not.

take up the opportunity to apply. These families predominantly have higher-skilled children and end up switching away from their base school in the sending neighborhood. In contrast to the experiment under endogenous residential choice, there is no net inflow of high-skilled children into the sending neighborhood. For all these reasons, the base school of the sending neighborhoods displays a decline in the peer quality of 24 percent, a conclusion that is opposite of the results we observe when families are allowed to choose their residential location.

Figure 3 highlights that, in this exercise, the negative consequences of the policy on children's skills are borne by lower-skill children, in line with the compositional effect previously discussed: on the one hand, lower-skill children of ED households (-1.2 percent) keep attending the local base school after its quality dropped because of their higher-skill neighbors who started attending the newly offered high-quality school ($+2.2$ percent). On the other hand, lower-skill children of non-ED households also experience a decline in skills (-1.2 percent) caused by a deterioration in school quality. Both the inflow of new peers from the sending school as well as the outflow of higher-skill children toward alternative option schools in the neighborhood contribute to this negative effect. Overall, the results of this exercise highlight the importance of accounting for endogenous neighborhood sorting when considering placed-based policies: expanding school choices generates a positive effect on disadvantaged children, even though we would conclude otherwise if the residential sorting responses to this policy were ignored.

5.2 Housing Vouchers

5.2.1 Policy Design

In the second counterfactual, we consider a policy reminiscent of the Moving to Opportunity (MTO) experimental voucher program that was implemented in the United States in the mid-1990s—although with a few key differences. We define as *eligible* the ED families who live in a neighborhood where 40 percent of households are ED in the baseline equilibrium.[36] Housing vouchers are granted to eligible families, with the requirement to locate in a neighborhood where the share of ED households is less than 10 percent. Conditional on eligibility, the voucher represents a housing subsidy that equals the fair market rent.[37] We assume the policy is financed by a proportional income tax on all households in the county. We analyze the MTO experiment in our framework under two implementation

[36] As a reminder for the reader, we previously defined economically disadvantaged (ED) households those who are eligible for free or reduced-price lunch at school.

[37] The fair market rent (FMR) is the 40th percentile of the rent distribution in the metropolitan area (Galiani et al., 2015), that is, \$9,600 per year in our data.

Table 7: Housing vouchers: Effects on Eligible Families

	Single-Family Voucher Policy		Voucher Policy at Large Scale		
	All eligible	Takers only	All eligible	Takers only	Non-Takers only
Families' take-up rate (%)		63		58	
Chg. in school quality (%)	33	53	22	28	13
Chg. in avg. nbhd income (%)	88	123	78	116	25
Chg, in avg. house price (%)	51	81	49	83	2

The table shows take-up rates, as well as the changes (from the baseline equilibrium) induced by the housing voucher policy on eligible families.

scales. In the first case, we implement the MTO experiment on a large scale, where all eligible families are offered participation in the program, and we compute the new equilibrium in the city. In the second case, we shut down any general equilibrium response to this policy, and we look at the average treatment effect for an eligible family, given their endogenous choice of program take-up and subsequent neighborhood and school choices. The purpose of this exercise is to study the equilibrium outcomes of MTO-like policies as well as to understand the external validity and the policy relevance of the estimated treatment effects of the MTO program in the literature (see, e.g., Kling et al., 2007; Chetty et al., 2016).

5.2.2 Results

Results are shown in Table 7 and Figure 4. The first two columns of Table 7 and the purple triangles in Panel A of Figure 4 refer to the single-family implementation of the policy, while the last three columns in Table 7 and the rest of the information in Figure 4 refer to the large-scale implementation.

Starting with the single family treatment effect of the vouchers, we find that sixty-three percent of eligible families take up the offered voucher and end up living in a house whose price is 81 percent higher than it would be at baseline (Table 7). They also live in a neighborhood with an average income that is 123 percent higher and that has a much higher quality base school (+53 percent). This overall improvement in the neighborhood and school quality where families live translates into a positive average effect on children's skills of 22.1 percent.

In the large-scale implementation, families who take up the voucher (58 percent) are still exposed to higher-income, more-expensive neighborhoods, and to better schools. However, once general equilibrium effects are taken into account, the improvement in the base school quality of the neighborhood where they live is significantly lower (+28 percent). As a consequence, the general equilibrium effect of the policy on children's skills is 40 percent lower than what the single family treatment would predict (+13.2 percent vs.

+22.1 percent). The difference between the treatment effect and the general equilibrium analyses of the MTO program on children is explained by the endogenous equilibrium change in receiving school quality generated by both the inflow of children from the poor neighborhoods (who on average are lagging behind in the skill distribution) as well as residential flight from the receiving neighborhoods by higher-income households with higher-skilled children. Panel B in Figure 4 shows that children who were living in the receiving neighborhood at baseline (below 10 percent poverty rate) experience an average negative impact on their skill development of 6.6 percent of a standard deviation, even after accounting for endogenous residential resorting. In equilibrium, the significant outflow of families from the sending neighborhoods also affects eligible families living in these neighborhoods who decided not to participate in the program. Table 7 shows that average neighborhood income for these families increases by 25 percent as a result of takers leaving their neighborhood of origin. By similar logic, base school quality in high-poverty neighborhoods increases by 13 percent. Finally, the rest of the children in the city, who were not living in either sending or receiving neighborhoods, experience a small average increase in skills (+1.7 percent) because of the inflow of higher-skill children who move away from receiving neighborhoods.

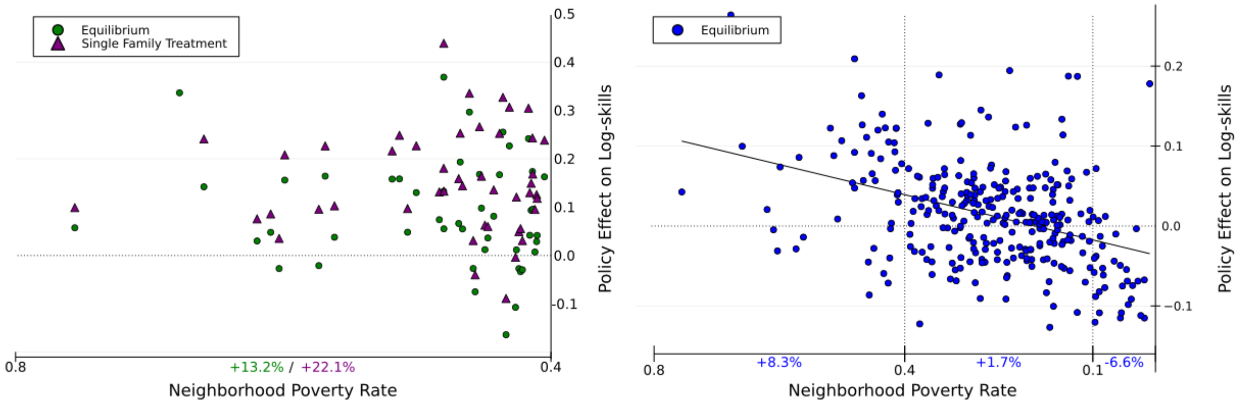
Overall, our quantitative exercise highlights that accounting for spatial general equilibrium effects is key for the evaluation of large-scale MTO-like policies. Because of general equilibrium effects, the benefits one would predict using individual treatment analysis in terms of socioeconomic mobility and opportunities for children living in disadvantaged neighborhoods are substantially attenuated when the policy is implemented at larger scale.

5.3 Upzoning

In this last section we return to a place-based approach and explore the role of zoning regulations on the inequality of educational access. We use the receiving (high-income) neighborhoods selected in the school choice expansion policy as the target for our upzoning experiment. In these receiving neighborhoods, two-thirds of which have the highest minimum lot size in the county, we reduce the regulated minimum house size to 800 square feet per dwelling unit, which represents approximately the average minimum house-size constraint in the economy.[38] The aim of this experiment is to analyze the effects of loosening zoning regulations and hence expanding the set of neighborhoods available to low-income families. On the one hand, this policy can change the socioeconomic com-

[38] In our sample, only three of the 15 receiving neighborhoods have a minimum lot size that we map into a minimum house size of less than 800 square feet. In this case, we maintain the original regulation.

Figure 4: Voucher Effects on Children's Skills by Poverty of Neighborhoods of Origin



(a) TE and GE Effects for Voucher Families (b) Equilibrium Effects and Neighb. of Origin

Plot A shows the equilibrium effects of the large-scale voucher implementation on children's skills by poverty rate of the neighborhood of origin (baseline). Plot B shows the voucher policy effects both in terms of the treatment effect (TE, single-family implementation), as well as general equilibrium effects (large-scale implementation) for families who participate in the program.

position of the neighborhood and effectively expand the educational opportunities for disadvantaged families. On the other hand, it can generate equilibrium responses with high-income families fleeing from the receiving neighborhoods, which would jeopardize neighborhood and school integration.

Tables C-3 and C-4 show the results. The policy does mitigate income segregation by changing neighborhood socioeconomic composition in the (high-income) neighborhoods directly targeted by the policy: on average, neighborhood family income decreases by 17 percent as a result of upzoning. At the same time, the new equilibrium displays a sizable reduction in the attractiveness of the receiving neighborhoods in terms of school children composition, with a reduction of the average children's skills by 39 percent. This drop in school quality causes housing prices to drop (-0.4 percent). Importantly, the last row of Table C-3 shows that upzoning would have instead caused house prices to increase by 0.7 percent in the targeted neighborhoods if the quality of the receiving base schools were kept exogenously constant at their baseline values. We see this result as direct evidence of how zoning reduces competition for houses in attractive neighborhoods, lowering the cost of accessing high-quality schools for high-income families.

6 Conclusion

We build and estimate a spatial equilibrium model of neighborhood and school choice that accounts for the key institutional determinants on educational access in the United

States: school attendance boundaries, heterogeneity in school transportation provision, and housing zoning regulations. The model replicates empirical facts of residential sorting and endogenous school choices in the data as well as quasi-experimental estimates on how demand for neighborhoods and schools respond to changes in school quality and transportation.

The model sheds light on the complexity of designing at-scale policies that aim to reduce inequality of educational opportunities. On the one hand, we find that expanding school choice is ineffective if not paired with school transportation and that the outcome of the policy crucially depends on the ensuing changes in *both* neighborhood and school choice. On the other hand, housing vouchers can positively affect targeted families, but such policy loses impact at larger scale because of residential equilibrium responses of families living in the receiving high-income neighborhoods.

Finally, we study how residential zoning regulations affect neighborhood and school composition. We show that zoning regulations reduce competition for houses in affluent neighborhoods by forming barriers to entry for low-income families, hence magnifying dispersion in school quality.

All in all, our equilibrium model highlights how individual incentives impose constraints on the outcomes policymakers aim to achieve (e.g., reducing education inequality). We see the study of optimal policies to achieve those outcomes through the design of local institutions like school boundaries and zoning regulations as a fruitful avenue for future research.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag Pathak, "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match," *American Economic Review*, 2017, *107* (12), 3635–3689.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism," *Annual Review of Economics*, 2018, *86* (2), 391–444.

_ **and** _ , "Revealed Preference Analysis of School Choice Models," *Annual Review of Economics*, 2020, *12* (1), 471–501.

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti, "It Takes a Village: The Economics of Parenting with Neighborhood and Peer Effects," 2020. NBER Working Paper 27050.

__ , **Morteza Saharkhiz, and Matthew Wiswall**, "Home and School in the Development of Children," 2020. NBER Working Paper 26037.

Ahlfeldt, Gabriel, Stephen Redding, Daniel Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence From the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

Allende, Claudia, "Competition Under Social Interactions and the Design of Education Policies," 2020. Working Paper.

Avery, Christopher and Parag Pathak, "The Distributional Consequences of Public School Choice," *American Economic Review*, 2021, *111* (1), 129–152.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 2007, *115* (4), 588–638.

Billings, Stephen B., Eric J. Brunner, and Stephen L. Ross, "Gentrification and Failing Schools: The Unintended Consequences of School Choice under NCLB," *The Review of Economics and Statistics*, 2018, *100* (1), 65–77.

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

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman, "Linear Social Interactions Models," *Journal of Political Economy*, 2015, *123* (2), 444–496.

Calsamiglia, Caterina, Chao Fu, and Maya Güell, "Structural Estimation of a Model of School Choices: The Boston Mechanism versus Its Alternatives," *Journal of Political Economy*, 2020, *128* (2), 642–680.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer Effects and Social Networks in Education," *Review of Economic Studies*, 2009, *76* (4), 1239–1267.

Chetty, Raj, 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*, 2016, *106* (4), 855–902.

Doepke, Matthias, Giuseppe Sorrenti, and Fabrizio Zilibotti, "The Economics of Parenting," *Annual Review of Economics*, 2019, *11*, 55–84.

Dur, Umut, Robert Hammond, and Thayer Morrill, "Identifying the Harm of Manipulable School-Choice Mechanisms," *American Economic Journal: Economic Policy*, 2018, *10* (1), 187–213.

Durlauf, Steven N. and Yannis M. Ioannides, "Social Interactions," *Annual Review of Economics*, 2010, *2* (1), 451–478.

Eckert, Fabian and Tatjana Kleineberg, "Saving the American Dream? Education Policies in Spatial General Equilibrium," 2021. Unpublished Manuscript, Yale University.

Epple, Dennis and Holger Sieg, "Estimating Equilibrium Models of Local Jurisdictions," *Journal of Political Economy*, 1999, *107* (4), 645–681.

_ **and Richard Romano**, "Neighborhood Schools, Choice, and the Distribution of Educational Benefits," in Caroline M. Hoxby, ed., *The Economics of School Choice*, University of Chicago Press, 2003, pp. 227–286.

Fogli, Alessandra and Veronica Guerrieri, "The End of the American Dream? Inequality and Segregation in US Cities," 2018. Unpublished Manuscript, University of Chicago.

Galiani, Sebastian, Alvin Murphy, and Juan Pantano, "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment," *American Economic Review*, 2015, *105* (11), 3385–3415.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger, "Heterogeneous Preferences and the Efficacy of Public School Choice," 2009. Unpublished Manuscript.

Heblich, Stephan, Stephen Redding, and Daniel Sturm, "The Making of the Modern Metropolis: Evidence from London," *The Quarterly Journal of Economics*, 2020, *135* (4), 2059–2133.

Heckman, James J. and Stefano Mosso, "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 2014, *6* (1), 689–733.

Hill, Daryl, Rodney Hugues, Matthew Lenard, David Leibowitz, and Lindsay Page, "New Schools and New Classmates: The Disruption and Peer Group Effects of School Reassignment," 2021. EdWorkingPaper No. 21-412 (Annenberg Institute at Brown University).

Jackson, Kirabo, "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation," *Journal of Labor Economics*, 2009, *27* (2), 213–256.

Kapor, Adam, Christopher Neilson, and Seth Zimmerman, "Heterogeneous Beliefs and School Choice Assignment Mechanisms," *American Economic Review*, 2020, *100* (5), 1274–1315.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Laverde, Mariana, "Unequal Assignments to Public Schools and the Limits of Schools Choice," 2020. Unpublished Manuscript.

Luflade, Margaux, "The Value of Information in Centralized School Choice Systems," 2018. Unpublished Manuscript.

Mogstad, Magne and Gaute Torsvik, "Family Background, Neighborhoods, and Intergenerational Mobility," 2022. Forthcoming in the Handbook of Family Economics.

Nechyba, Thomas, "Mobility, Targeting, and Private-School Vouchers," *American Economic Review*, 2000, *90* (1), 130–146.

Parcel, Toby L. and Andrew J. Taylor, *The End of Consensus: Diversity, Neighborhoods, and the Politics of Public School Assignments*, Chapel Hill, NC: University of North Carolina Press, 2015.

Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, and Randy Walsh, "ESTIMATING THE GENERAL EQUILIBRIUM BENEFITS OF LARGE CHANGES IN SPATIALLY DELINEATED PUBLIC GOODS*," *International Economic Review*, 2004, *45* (4), 1047–1077.

Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's Transmilenio," 2019. Unpublished Manuscript.

Online Appendix for "On the Spatial Determinants of Educational Access" by Francesco Agostinelli, Margaux Luflade and Paolo Martellini

A Data appendix

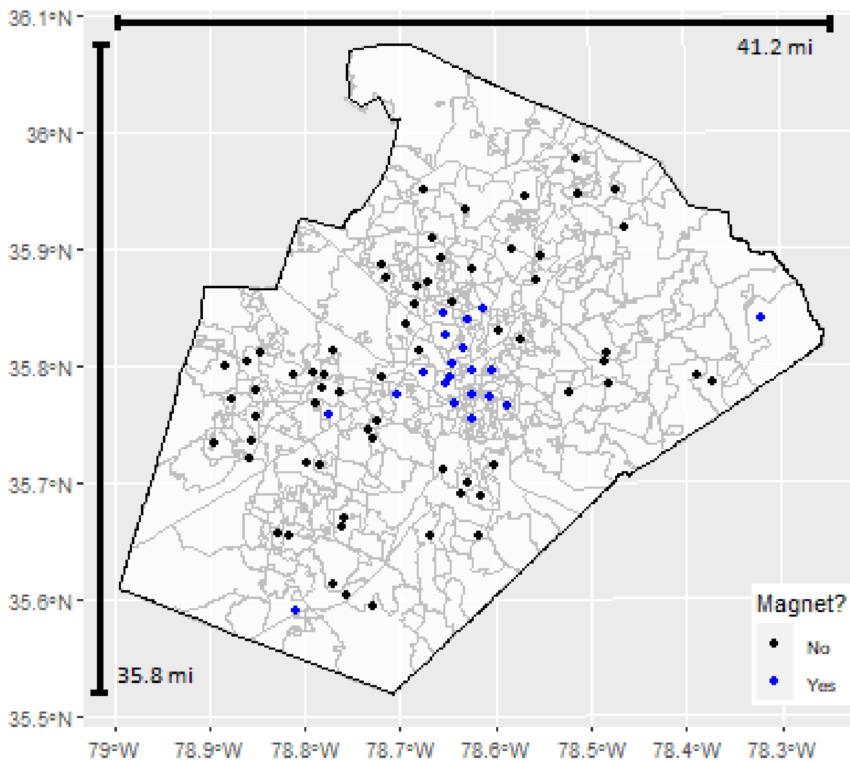
A.1 Additional institutional details

The WCPSS is the county-wide school covering Wake County, North Carolina, which is the county of the state capital, Raleigh. The WCPSS was, in 2019–20, the fourteenth largest school district in the United States, with more than 161,000 students. Over the 2000–10 decade, the public school population in th WCPSS increased from about 95,000 to more than 140,000. Figure A-1 illustrates the geography of the county—in particular its size—and the locations of the elementary schools open during our sample period (2003–04 to 2006–07).

A.1.1 Public school choice in the WCPSS

Each address in Wake County is associated with a *base* school at which the child is guaranteed a seat and transportation. The school district offers two main ways for parents to have their child attend a public school other than their base: magnet programs and calendar transfers, each of which we describe below —although when we bring our structural model to the data, the two types of options are not differentiated and are pooled under the umbrella category of "option school."

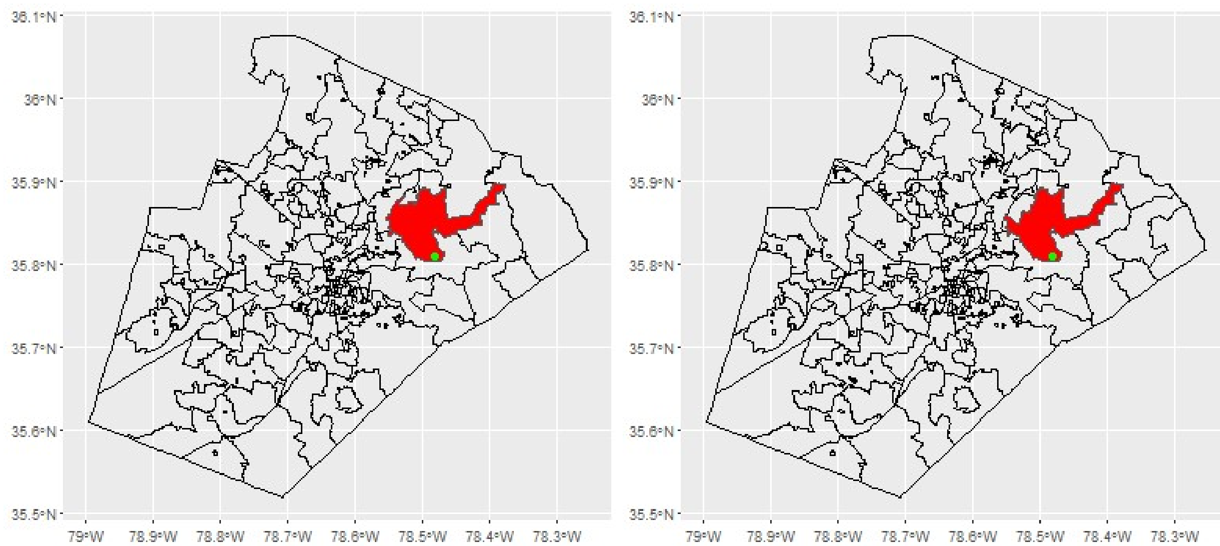
Figure A-1: Elementary School Locations in Wake County



The figure shows the location of elementary schools in Wake County in 2006.

Historically, from the creation of the district in 1976 until 2000, the student assignment policy was driven by the goal of promoting racial diversity in schools. Residential addresses were assigned to base schools so that each school would have 15–45 percent of Black students. Magnet programs were created as a second instrument to facilitate racial integration in schools: a number of urban schools were endowed with special educational programs (e.g., arts, foreign languages) that were expected to draw white suburban students. Starting from the 2000–01 academic year and until 2011–12, the WCPSS moved from the goal of ensuring racial diversity in schools to that of ensuring socioeconomic balance. Assignments of addresses to base schools was then supposed to serve the goal that no school had more than 40 percent of students eligible for free or reduced-price lunch (FRPL) nor more than 25 percent of students below the state's reading standards for their grade. While socioeconomic balance in schools was a target for the school board until the early 2010s, pressure to accommodate unequal population growth across the county has been the main driver of school reassignments as illustrated by this quote from Parcel and Taylor (2015, p. 53) who said reassignment "from school to school [was] because of population growth, and that is what it was. The busing was not intended primarily for diversity but just to fill in . . . schools." As an illustration of changes in catchment area boundaries, Figure A-2 shows base schools' catchment areas for school years 2004–05 (left panel) and 2005–06 (right panel). The green dot shows the location of Forrestville Road Elementary School (school code 920413), and the area shaded in red shows its catchment area. Comparing left and right panels shows that the northwest part of the catchment area was reassigned to another base school between the two years.

Figure A-2: Elementary School Catchment Areas, 2004–05 (left) and 2005–06 (right)



(a) 2004–05

(b) 2005–06

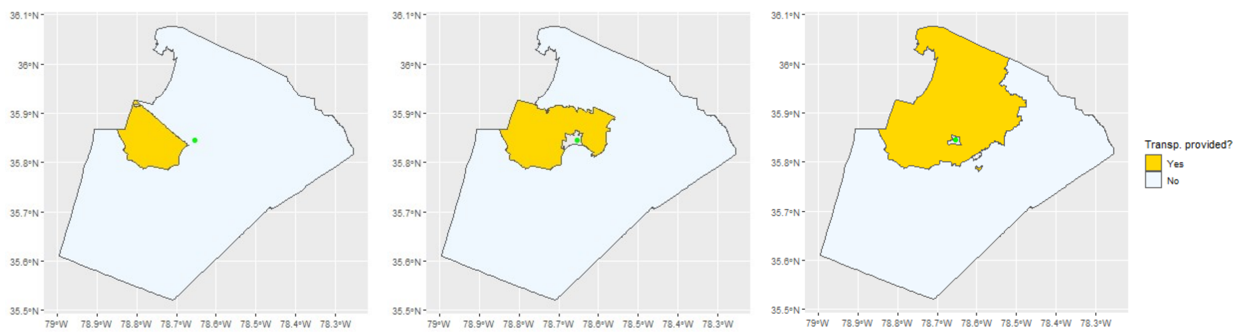
The figure shows elementary schools catchment areas for 2004–05 (left) and 2005–06 (right). The catchment area for Forrestville Road Elementary School is highlighted to show changes from one year to the other.

Magnet programs were created as a second instrument to facilitate racial and, then, socioeconomic integration in schools. Through these, a number of urban schools were endowed with special educational programs (e.g., arts, foreign language immersion, etc.) that were expected to draw white suburban students. In our period of interest, the WCPSS had 17 magnet programs at the elementary school level. Based on their residential address, parents can apply to a subset of these programs for their child. Also based on their residential address, parents may or may not be offered school transportation to the magnet program.[39] Families can apply to up to three magnet programs, and assignment is made according the Boston Mechanism (for 90 percent seats in each school) or a pure lottery (for the remaining 10 percent of seats in each schools). Magnet choice set and transportation provision do not only change cross-sectionally, they also change over time

[39]Figure 1 in Dur et al. (2018) shows a screenshot of the online platform parents can use to apply; the fourth column in the table illustrates the variation of transportation provision across schools and residential addresses.

during the period of interest, with several magnets expanding and/or changing their transportation provision. Figure A-3 provides an illustration of transportation changes over time. Twenty-four magnet programs saw a change in the set of neighborhoods eligible to apply and/or in their transportation provision over the sample period (see bottom panel of Table A-2).

Figure A-3: Changes in Transportation Provision to Magnet Programs—An Example: Brooks Elementary



(a) 2004–05 (b) 2006–07 (c) 2009–10

The figure shows the areas of the county from which school transportation to the magnet program at Brooks Elementary was provided in various school years.

Calendar transfers allow students to attend a school running on a different calendar than their base school. Schools in the WCPSS operate following one of two calendars—the traditional September to June academic calendar or a year-round calendar designed as a response to the rapid population growth to allow schools to accommodate more students at a time.[40] Each base school is paired with one alternative calendar school to which assigned families can apply and to which transportation will be provided.

Assignment to magnet and calendar options is centralized. As reported by Dur et al. (2018, p. 192), "90 percent of magnet seats are assigned via the Boston Mechanism ... For elementary schools, priority points at school s depend on whether the student's sibling will attend school s next year (highest priority), whether the student lives in a high-performing [area] based on historical test score data (second highest), and whether the student's base school is overcrowded (third highest). ... Finally, 10 percent of magnet seats are assigned through a pure lottery; specifically, a lottery that is independent of a student's priority points. The district introduced the 10 percent lottery to encourage more students to participate in the magnet application process."

[40] In year-round schools, students are placed on four different tracks, each of them alternating year-round between nine weeks of class and three weeks of break. At any point in time, one of the four tracks is on break, allowing the school to serve a larger number of students.

To recap, here are the key simplifying assumptions we make in the model that depart from the institutional choice choice setting of the WCPSS: (i) no distinction between calendar transfers and magnet applications; both are considered to be a single type of "option school"; (ii) applications to at most one option school, either magnet or calendar, in contrast to two distinct application procedures, and up to three choices in the magnet application procedure (one in calendar application); (iii) assignment by pure lottery, with equal probability of admission among those eligible who apply, in contrast to a Boston mechanism (for 90 percent of magnet seats) with priorities described above. Dur et al. (2018, p. 192) note that the "WCPSS used the Boston Mechanism for the reason that Boston and many other districts used it: it is intuitive, easy to explain, and maximizes the number of students assigned to their reported first choice." Indeed, a key feature of the Boston mechanism is that students who rank a school first get higher priority for that school than all other applicants. In the case in which, for each school, the number of first-choice applicants exceeds the number of seats, each applicant will only be considered for admission to their first choice, rendering all choices ranked below the first one irrelevant. In that regard, our single-application assumption is a reasonable approximation of the actual assignment process.

A.1.2 On the exogeneity of institutional changes

The identification of β_1 in Equation (2.3) (Section 2.3) and of the structural parameter γ (Section 4) relies on within-neighborhood variation over time, and requires the changes in school quality induced by changes in catchment areas to be unanticipated by households. The parameters α_1 and β_1 in Equations (2.6) and (2.7) (Section 2.3) provides empirical evidence on the elasticity of school demand to transportation provision. Its estimation relies on within-neighborhood changes over time in the transportation provision to the different option schools in the portfolio of the neighborhood, and supposes that changes are not anticipated by households.

Changes in base schools' catchment areas. While the school board targeted socioeconomic balance in schools was a target until the early 2010s, pressure to accommodate unequal population growth across the county has been the main driver of base school reassignments as illustrated by this quote from Parcel and Taylor (2015, p. 53) who said reassignment "from school to school [was] because of population growth, and that is what it was. The busing was not intended primarily for diversity but just to fill in . . . schools." In addition, while the fact that changes in catchment areas were likely well-known to families over the period of interest (Parcel and Taylor, 2015), Hill et al. (2021, p. 7) argue that "the selection of any given geographic node for reassignment was, conditional on observable traits of the node, essentially random and not manipulable or anticipated by [neighborhood] residents. ... As a result of the reassignment plan, geographically proximal and observationally similar [neighborhoods] were treated differently. Students from the same geographic area but different assignment nodes, who had been assigned to attend the same school in one year, would be assigned to attend different schools the following year."

Changes in magnet programs' transportation provision. The reasons underlying changes in transportation provision are not as well documented in the literature (nor in the minutes of school board meetings) as those underlying changes in base schools' catchment areas. The argument for the exogeneity of policy changes can therefore not be made in the same way as it was for changes in base schools' catchment areas. Instead, to assess whether neighborhoods chosen for the changes could be predicted based on their observable characteristics, we test whether neighborhoods experiencing changes in school transportation differ from neighborhoods that do not. We do so by estimating the following regression:

$$\begin{aligned} y_{nst} = {} & a_1\mathbf{1}[\text{Bus Provided (1 Year Ahead)}]_{nst} \\ & + a_2\mathbf{1}[\text{Bus Provided (2 Years Ahead)}]_{nst} + \delta_{n,s} + \lambda_t + \epsilon_{nst}, \end{aligned} \tag{A-1}$$

where y_{nst} is a neighborhood-level characteristic at time t (e.g., share of economically disadvantaged families) and $\mathbf{1}[\text{Bus Provided (T Year(s) Ahead)}]_{nst}$, for $T = 1, 2$, is an indicator variable equal to 1 if school transportation is added between neighborhood n and school s at the beginning of school year $t + T$. $\delta_{n,s}$ and λ_t are, respectively, a neighborhood-school fixed effect and a time fixed effect. Results can be seen in Table A-1. Results show that neighborhoods affected by transportation changes were not significantly different from unaffected neighborhoods in terms of family characteristics (Panel A) or children's initial achievements (Panel B).

A.2 Data sources

Student data. Student-level data were obtained from the North Carolina Education Research Data Center[41] (NCERDC). The data show, for each year and each student enrolled in a North Carolina public school in grades 3–8, the school the child is enrolled in, end-of-grade test scores in math and reading, and a set of demographic variables (gender, race, economically disadvantaged status). Starting in 2006, the data also show the student's

[41] https://childandfamilypolicy.duke.edu/research/nc-education-data-center/, accessed August 2021.

Table A-1: Pre-Transportation-Change Neighborhood Characteristics

PANEL A: FAMILY CHARACTERISTICS

	Econ. disadv.			Black			White			Hispanic			Asian		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Bus Provided (1 Yr Ahead)	.0050		.0098	.0085		.0054	−.0027		.0055	−.0049		−.0060	.0031		.0048
	(.0076)		(.0099)	(.0072)		(.0093)	(.0081)		(.0104)	(.0063)		(.0079)	(.0041)		(.0049)
Bus Provided (2 Yrs Ahead)		−.0004	.0009		.0013	.0058		.0027	.0082		−.0022	−.0059		.0006	.0010
		(.0086)	(.0010)		(.0080)	(.0091)		(.0091)	(.0101)		(.0070)	(.0077)		(.0042)	(.0048)
Mean Dep.Var.	0.400	0.396	0.400	0.278	0.282	0.286	0.445	0.453	0.448	0.166	0.160	0.161	0.060	0.054	0.055
SD Dep. Var.	0.363	0.363	0.360	0.307	0.310	0.308	0.355	0.359	0.354	0.233	0.233	0.229	0.133	0.125	0.124
Observations	30218	23531	22153	30243	23541	22163	30243	23541	22163	30243	23541	22163	30243	23541	22163

PANEL B: CHILDREN'S TEST SCORES

	Math score (stdized)			Read. score (stdized)		
Bus Provided (1 Yr Ahead)	−.0262		−.0530**	−.0187		−.0099
	(.0181)		(.0234)	(.0182)		(.0240)
Bus Provided (2 Yrs Ahead)		−.0071	−.0291		.0076	.00745
		(.0204)	(.0291)		(.0210)	(.0235)
Mean Dep.Var.	−0.100	−0.102	−0.105	−0.088	−0.093	−0.092
SD Dep. Var.	0.666	0.671	0.666	0.669	0.678	0.670
Observations	30182	23504	22126	30158	23492	22114

This table shows test results for whether neighborhoods experiencing changes in school transportation changes differ from neighborhoods that do not. Column (3) shows results for the specification shown in Equation (A-1). Column (1) shows results for a similar specification focusing on the prediction of demographic changes one year ahead, thus omitting $\mathbf{1}[\text{Bus Provided (2 Years Ahead)}]_{nst}$; column (2) shows results for a similar specification focusing on the prediction of demographic changes two years ahead, thus omitting $\mathbf{1}[\text{Bus Provided (1 Year Ahead)}]_{nst}$. Standard errors are robust to heteroskedasticity and reported in parentheses. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

residential census block group and (a noisy version of) residential coordinates.

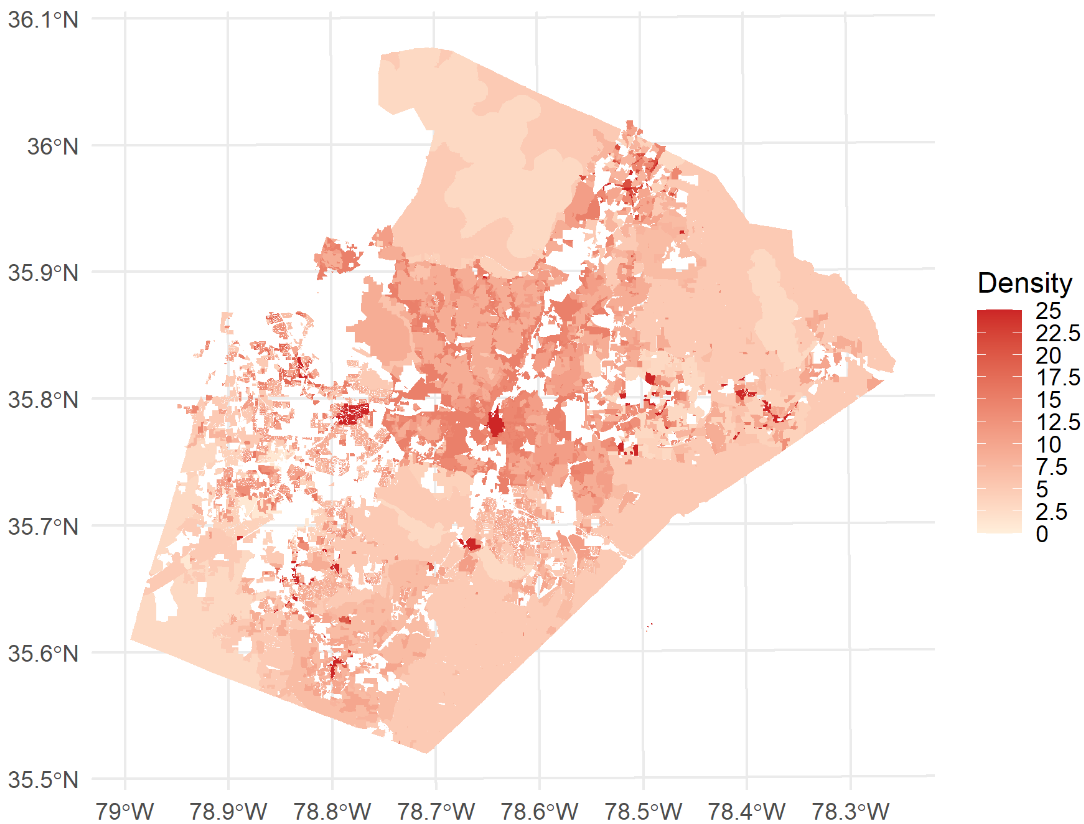
Catchment areas, transportation provision, admission probabilities. Choice sets of schools were created from data shared by the Wake County Public Schools System—namely, maps showing yearly and for every address in Wake County, the base school associated with the address point, calendar options for the address, as well as the choice set of magnet programs the address can apply to. For each magnet in the choice set and each year, the data also show whether school transportation is provided between the magnet program and the address point.

Real estate data. Publicly available records from Wake County show details about all real estate transactions in Wake County starting from 1956.[42] For each property sold, these data show the sale price and date, exact address of the property, characteristics of the lot and of the buildings/units, if any. In particular, we use the following characteristics in the analysis: sale date, sale price, acreage of the lot, year the building was built, whether the building is for residential use, and its type (single-family house, apartment, etc.), and heated area. We use heated area as our measure of house size.

Zoning data. Multiple entities are in charge of zoning regulations in the county. While part of county land is regulated by the county itself, the zoning in other areas is done by a number of different local municipalities and/or unincorporated areas—namely: Raleigh, Apex, Cary, Fuquay-Varina, Garner, Holly Springs, Knightdale, Morrisville, Rolesville, Wake Forest, Wendell, and Zebulon. Geographic data on the zoning regulations for each entity is publicly available at: `https://data-wake.opendata.arcgis.com/` (accessed August 2021). Each entity uses its own zoning categories and labels. By harmonizing regulation categories and labels across entities, we create a geographical dataset that gives, for any (residential) point in the county, the associated minimum lot size (MLS) regulation. Figure A-4 represents MLS regulations over (residential land in) Wake County. Density regulations are typically expressed in dwelling units (du) per acre —the stronger the regulation, the lower the density allowed. Lighter areas in Figure A-4 are zoned for lower density, meaning that fewer dwelling units are allowed to be built on one acre of land. The inverse of density gives the more intuitive measure for MLS, which is acre per lot. There is a relatively wide range of MLS regulations throughout Wake County—from more than 25 du/acre in the urban center of the county, to less that 1 du/acre in the western periphery.

[42] `https://www.wakegov.com/departments-government/tax-administration/real-estate`, accessed August 2021

Figure A-4: Minimum Lot Size Restrictions (in Dwelling Units Per Acre) in Wake County



The figure shows density regulations throughout Wake County.

American Community Survey (ACS) data. We use the following (tract- and county-level) variables from the ACS five-year estimates (2006–10): "Family Type by Presence of Own Children Under 18 Years by Family Income in the Past 12 Months (in 2010 Inflation-Adjusted Dollars)" (NHGIS Code J5A) and "Own Children Under 18 Years by Family Type and Age" (NHGIS Code JM3). Data were downloaded from https://www.nhgis.org/ (accessed August 2021).

A.3 Construction of neighborhoods

Each neighborhood n is characterized by a sequence of base schools and school choice sets from school year 2003–04 to school year 2009–10: $\{ (\mathcal{B}_{n,t}, \mathcal{T}_{n,t}, \mathcal{NT}_{n,t}) \mid t = 2003, \ldots, 2009\}$ is the base school associated with n in year t, $\mathcal{T}_{n,t}$ is the set of option schools providing transportation to neighborhood n in year t, $\mathcal{NT}_{n,t}$ is the set of option schools in the choice set of neighborhood t but not providing transportation. Neighborhood n is the union of all *contiguous* points with school choice menu $\{ (\mathcal{B}_{n,t}, \mathcal{T}_{n,t}, \mathcal{NT}_{n,t}) \mid t = 2003, \ldots, 2009\}$. Formally, let us denote each residential address by its coordinates (x, y). $(x, y) \in n$ only if the following three points are satisfied:

1. (x, y) has base school $\mathcal{B}_{n,t}$ in school year t, for each t.
2. $\mathcal{T}_{n,t}$ is the set of all schools (except for $\mathcal{B}_{n,t}$) providing transportation to (x, y) in school year t.
3. $\mathcal{NT}_{n,t}$ is the set of all schools open for application to (x, y) but not providing transportation to (x, y) in school year t.

In addition, we require neighborhoods to consist of fully contiguous points so if two regions share the same portfolio $\{ (\mathcal{B}_{n,t}, \mathcal{T}_{n,t}, \mathcal{NT}_{n,t}) \mid t = 2003, \ldots, 2009\}$ but are not touching, they make up distinct neighborhoods. Our definition of neighborhoods implies that at any point in our sample period, two addresses in the same neighborhood share the same portfolio of schools—base and options with and without transportation. Conversely, two addresses can be in distinct neighborhoods for two reasons. Either their respective portfolios of schools differ at some point in the sample period or, if they share the same portfolio of schools, they are part of two geographic regions with no common border. We match third graders from the NCERDC data to the constructed neighborhoods based on their address information. We then rank neighborhoods by decreasing order of their student populations and exclude the lowest ranked neighborhoods so as to keep 90 percent of the students. This ensures that (i) computations remain manageable, and (ii) all neighborhoods in the sample contain students every year.

Figure A-5 shows the obtained partition of Wake County into neighborhoods, with 2000 census tract boundaries for comparison.

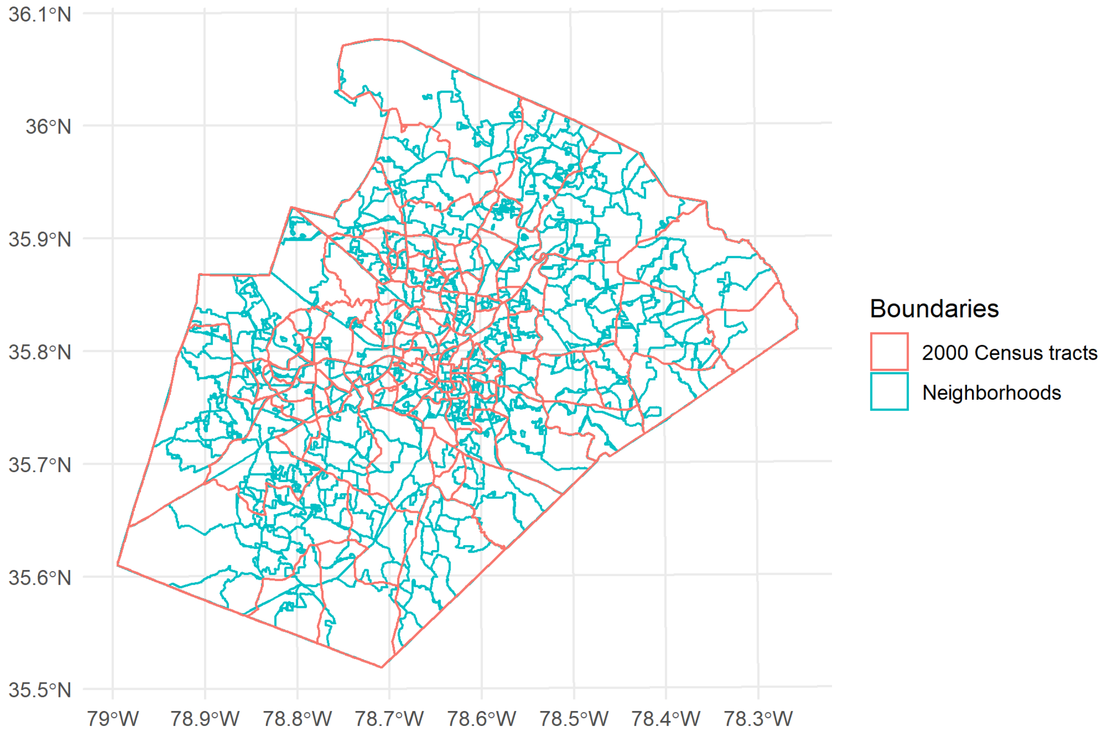
A.4 Final sample construction

The final estimation sample is obtained after three successive sample restrictions:

1. We restrict the sample to students enrolled in third grade in a Wake County public school in school years 2006–07 to 2011–12 and with the following information not missing for the third grade year: residential address, school attended, economically disadvantaged status, end-of-grade test scores.

2. After matching students to their neighborhood, we count the number of students assigned to each neighborhood, rank neighborhoods by decreasing order of their student population, and exclude the lowest ranked neighborhoods so as to keep 90 percent of the students. This ensures that (i) computations remain manageable, and (ii) all neighborhoods in the sample contain students every year. Our final sample consists of 305 neighborhoods.

Figure A-5: Comparison of Constructed Neighborhoods vs. 2000 Census Tracts



The figure shows the boundaries of our constructed neighborhoods (in blue) and of 2000 census tracts (in red).

3. Given their residential neighborhood and detailed administrative information about catchment areas, we are able to determine whether each student attends a school that is indeed in that student's choice set. More precisely, under the assumption that the student has been living at the same address since his kindergarten year, we observe three sets of students: children attending a school assigned to their neighborhood as a base or option school when they entered kindergarten; children attending a school that is not in the choice set attached to their neighborhood in their kindergarten year, but assigned to their neighborhood as a base or option school when they entered first, second, or third grade; and children attending a school that has never been part of their choice set since the year they entered kindergarten. Since their choices cannot be explained given the choice set, we exclude the latter set of students from our sample.

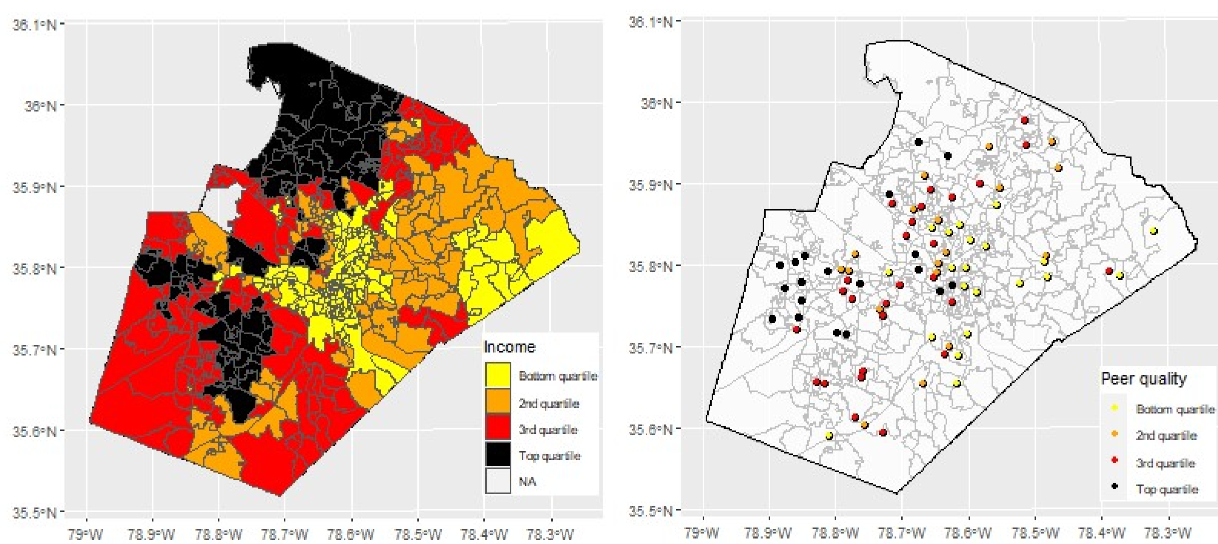
Students are observed for the first time in their third grade school year. Most students in Wake County start school in kindergarten. We assume that residential and school choices were made by the family at the time the child entered school for the first time, that is, when the child entered kindergarten. We therefore need to impute the neighborhood and school chosen by the family as the child entered kindergarten. To do this, we make two assumptions:

- Regarding neighborhood choice at school entry, we assume the residential address chosen at kindergarten entry is the same as the address observed in third grade.
- Regarding school choice at school entry:
  - If, in third grade, the child attends a school assigned to her neighborhood as a base or option school when she entered kindergarten, then we assume the child attended that school in kindergarten.
  - If, in third grade, the child attends a school that was not in her neighborhood's choice set when she entered kindergarten, but that was in her neighborhood's choice set in a later year (i.e., in the child's first, second, or third grade year), then we assume the child entered kindergarten attending the base school attached to her neighborhood at the time.

A.5 Descriptive statistics

Table A-2 is the main table supporting the sample description in Section 2.2. Figure A-6 shows the distribution of household income and school quality across the county.

Figure A-6: Distribution of Household Income (Left) and School Quality (Right)



(a) Household income (b) School quality

The map on the left shows average household income by neighborhood (source: ACS 5-year estimates 2006–10). The map on the right shows school quality for each school in the sample. Our measure of school quality is average (standardized) third-grade math test scores (source: NCERDC, 2006).

Table A-2: Descriptive Statistics

	Mean	Stdev	Min	Max
PANEL A: NEIGHBORHOOD SAMPLE				
# transactions obs. /yr	25.06	24.39	1	175
Avg house size (sqft)	2107	652.8	784	4054
Avg sale price by sqft	97.64	22.93	13.05	323.2
Avg user cost by sqft	5.44	1.93	0.83	13.09
Avg lot size (acre)	0.45	0.46	0	4.50
Has MLS regulation	0.63	0.48	0	1
Avg MLS regulation (acre)	0.15	0.24	0	0.92
Avg # of school options (excl. base)	12.78	0.61	11	14
Avg # of school options w/ transp. (excl. base)	3.67	0.95	1	6
Distance to base sch. (miles)	3.71	3.11	0.01	16.82
Avg. distance to option sch.	10.73	5.62	0.34	30.66
Has base change during period	0.15	0.36	0	1
Has change in option set during period	0.99	0.08	0	1
Avg # of student obs. /yr	17.97	18.31	1	128
Share of econ. disadv. (ED) students	0.37	0.30	0	1
# of neighborhoods in sample	*305*			
# neighborhood-year obs.	*915*			
PANEL B: STUDENT SAMPLE				
Is econ. disadv. (ED)	0.30	0.46	0	1
Attends base, cond. on being ED	0.92	0.27	0	1
Attends base, cond. on being non-ED	0.81	0.39	0	1
Attends option w/ transp., cond. on ED	0.05	0.21	0	1
Attends option w/ transp., cond. on non-ED	0.16	0.37	0	1
Attends option w/o transp., cond. on ED	0.03	0.18	0	1
Attends option w/o transp., cond. on non-ED	0.03	0.17	0	1
Ability (standardized test score) cond. on ED	-0.58	0.87	-3.20	2.23
Ability cond. on non-ED	0.40	0.88	-3.07	2.33
# of student-yr obs.	*16,445*			
PANEL C: SCHOOL SAMPLE				
Avg peer quality	1.49	0.50	0.29	2.99
Share econ. disadv. students	0.34	0.16	0.04	0.72
# of student obs. in sample	58.81	44.58	0	232
Is option school for some address	0.33	0.47	0	1
Has catchm. area change during period (base)	0.24	0.43	0	1
Has elig./transp. change during period (opt. sch.)	0.24	0.44	0	1
# of schools in sample	*87*			
# school-year obs.	*261*			

In the top (respectively middle, bottom) panel, the mean, standard deviation, minimum, and maximum are taken over the sample of neighborhood-year (respectively student-year, school-year) observations.

A.6 Mapping the model to the data

Variables are listed roughly in the order of their introduction in Section 3.2.

Household income w_p and average neighborhood income $\bar{w}_n$. The ACS five-year estimates (2006–10) table "Family Type by Presence of Own Children Under 18 Years by Family Income in the Past 12 Months (in 2010 Inflation-Adjusted Dollars)" (NHGIS Code J5A) gives household counts by census tracts for families with and without children and for 16 brackets of household income. We use variables J5AE004–J5AE019, J5AE040–J5AE055, and J5AE075–J5AE090 to characterize the income distribution of our "families," and J5AE021–J5AE036, J5AE057–J5AE072, and J5AE092–J5AE107 for non-families. We use the 16 ACS brackets as our discrete values for household type p in the model. Household income w_p for a household in bracket p of the ACS is constructed in three steps. First, gross income is assumed to be the middle point of the ACS bracket p (and $250,000 for the top bracket "more than $200,000"). Next, net income is obtained from gross income using the NBER TAXSIM program,[43] assuming the following household characteristics: married couple, spending 28 percent of their income on a mortgage, and with one dependent younger than 13. We use these household characteristics for all households in the model, that is families and non-families. To aggregate the ACS household income levels into the ED and non-ED categories available at the student level in the NCERDC data, we assign the seven lower ACS brackets (that is, with family income in the past 12 months below $39,999 in 2010 inflation-adjusted dollars) to ED, and the nine higher brackets to non-ED. ED status in the NCERDC is determined by eligibility for free or reduced-price lunch. Income levels for eligibility to the programs are determined annually by the USDA.[44] For reference, the eligibility thresholds for the school year 2007–08 for reduced-price lunch (below 185 percent of the federal poverty line) were $31,165 annual income for a household of three, and $38,203 for a household of four.[45]
Average income in neighborhood n is obtained as: $\bar{w}_n = \sum_p \text{mid}_p \times Pr(p \mid n)$, where mid_p is the middle value of ACS income bracket p, and $Pr(p \mid n)$ is the share of households with income in bracket p in neighborhood n.

Child skills a_k and school peer quality $\bar{a}_s$. We use end-of-third-grade (math) test scores as a measure of a student's skills. Test scores are standardized by grade and cohort. For the structural estimation, we consider ten discrete skills bins corresponding to the deciles of the continuous standardized test score distribution. a_k is set at the average skill level in bin k, $k = 1, \ldots, 10$. School peer quality for school s and year t is measured as the average

[43] https://users.nber.org/~taxsim/, accessed August 2021.
[44] https://www.fns.usda.gov/cn/income-eligibility-guidelines, accessed August 2021.
[45] https://www.govinfo.gov/content/pkg/FR-2007-02-27/pdf/07-883.pdf, accessed August 2021.

standardized test score for third grade students enrolled in school s in year t. All third grade students with non-missing test scores (and school attended information) are used to compute school peer quality (while only those with non-missing ED status and address information are kept in the structural sample of students).

Joint distribution of parental income and child skills. The joint distribution of parental income and child skills is not directly observable in the data. On the one hand, the NCERDC data, which contain information about child skills, only report ED and non-ED as measures of socioeconomic status. On the other hand, the ACS, which shows population counts by income brackets, does not contain any information about children skills. We infer the joint distribution as follows. Note that:

$$Pr(k,s,p,n) = Pr(k,s,p \mid n)Pr(n) = Pr(k,s \mid p,n)Pr(p \mid n)Pr(n).$$

There:

- $Pr(n)$ is obtained from the ACS as the probability that a family with a child aged four to five lives in neighborhood n.
- $Pr(p \mid n)$ is the share of families with a child younger than 18 years old (smallest level of aggregation available for income data in the ACS) conditional on living in neighborhood n.
- $Pr(k,s \mid p,n)$ is not observed since the NCERDC data only contain information about ED status, that is, only gives $Pr(k,s \mid ED,n)$ and $Pr(k,s \mid non-ED,n)$. We assume $Pr(k,s \mid p,n) = Pr(k,s \mid ED,n)$ for all p that belong to the ED category; and $Pr(k,s \mid p,n) = Pr(k,s \mid non-ED,n)$ for all p that belong to the non-ED category.

From there, we derive $Pr(k,p) = \sum_s \sum_p Pr(k,s,p,n)$.

Neighborhoods and schools coordinates; school assignments to neighborhoods. For each neighborhood, we use its centroid coordinates as the coordinates of the neighborhood. Schools coordinates and school portfolios $\mathcal{B}_{nt}, \mathcal{T}_{nt}, \mathcal{NT}_{ntt}$ associated with (defining) each neighborhood are taken directly from the WCPSS data.

Zoning restrictions, minimum house size $h_n^{\mathbf{mls}}$, and essential minimum house size h_0. To each neighborhood, we attach a MLS. For neighborhood that overlap multiple zoning areas with distinct MLS restrictions, the neighborhood-level MLS restriction is constructed as the least constraining MLS in the neighborhood. Formally, $\underline{\text{mls}}_n = \min\{\underline{\text{mls}}(x,y) \mid (x,y) \in n\}$, where (x,y) simply denotes the coordinate of any point in Wake County zoned for residential use, and $\underline{\text{mls}}(x,y)$ is the MLS restriction in place at that point. In the model though, we assume households choose and are constrained in their choice of *house* size, rather than *lot* size. We map neighborhood restrictions on minimum lot size ($\underline{\text{mls}}_n$) into a minimum house size available (h_n^{mls}). Regressing observed house sizes (in square feet) in the data on our measure (in acres) of minimum lot size ($\underline{\text{mls}}_n$) yields the mapping: $\underline{h}_n = 641 + 892 \times \underline{\text{mls}}_n$. From this mapping, we deduce h_n^{mls} for each neighborhood n, as well as the essential minimum housing $h_0 = 641$ (minimum house size in the absence of regulation).

Admission probabilities p_s. To estimate the model, we use information about admission probability in each option school for children entering kindergarten in Fall 2003, Fall 2004, and Fall 2005. The WCPSS provided five types of historical data that we use to infer the needed admission probabilities: (i) the number of applications received, accepted, and denied by grade and by year for each magnet program in Fall of 2007 and Fall of 2008; (ii) the number of applications received, accepted, and denied by year for each magnet school from Fall 2003 to Fall 2011; (iii) the number of applications received, accepted, and denied by grade and by year for each calendar transfer program in Fall of 2007 and Fall of 2008; (iv) the number of applications received, accepted, and denied by year for each calendar transfer program in Fall 2006; (v) the number of applications received, accepted, and denied by year overall by calendar transfer programs from Fall 2003 to Fall 2011. For program s in year t, we set admission probability p_{st} to one of the following:

- if the number of applications received $\text{appli}_{st}^{kinder}$ and accepted ($\text{accept}_{st}^{kinder}$) or denied ($\text{appli}_{st}^{kinder}$ - $\text{accept}_{st}^{kinder}$) for kindergarten entry in year t are observed, then we set $p_{st} = \frac{\text{accept}_{st}^{kinder}}{\text{appli}_{st}^{kinder}}$
- otherwise, if the number of applications received appli_{st}^{all} and accepted (accept_{st}^{all}) or denied (appli_{st}^{all} - accept_{st}^{all}) in year t are observed only overall in all grades, then we set $p_{st} = \frac{\widehat{\text{accept}}_{st}^{kinder}}{\widehat{\text{appli}}_{st}^{kinder}}$, where we infer $\widehat{\text{appli}}_{st}^{kinder} = \text{appli}_{st}^{all} \times \frac{\text{appli}_{s,2006}^{kinder}}{\text{appli}_{s,2006}^{all}}$, and $\widehat{\text{accept}}_{st}^{kinder} = \text{accept}_{st}^{all} \times \frac{\text{accept}_{s,2006}^{kinder}}{\text{accept}_{s,2006}^{all}}$ (using $t = 2006$ because $\text{appli}_{st}^{kinder}$ and $\text{accept}_{st}^{kinder}$ are both available for that year).

Distance between schools and neighborhoods τ_{ns}. As the distance between neighborhood n and school s, we use the road distance between the centroid of n and s. The road distance between any two points is computed using the OSRM package, which is an interface between R and the OSRM API. OSRM is a routing service based on OpenStreetMap data.[46]

[46] https://www.openstreetmap.org/, accessed August 2021.

House prices. We use average house price by neighborhood and year. We proceed in three steps to construct these average prices from the Wake County real estate data described in A.2. First, we convert all prices into 2010 dollars to be consistent with household income provided in 2010 dollars in the ACS. Second, we derive the average price per square foot for each neighborhood and year. Finally, we convert this average sale price per square foot into a per-period housing payment. Per-period payments R are derived from sale prices P by: $R = K \times P$ where the constant K is chosen so that, given the income distribution in the data, the average house size (in square feet) demanded by households matches the average house size in the data when households spend a share $\beta = .25$ of their income on housing. In practice $K = 1/15$.

B Estimation appendix

Let T denote the number of years used in estimation ($T = 3$), N the number of neighborhoods ($N = 305$), K the number of children's skills bins ($K = 10$), and P the number of household income bins ($P = 16$). Let $\tilde{S}$ be the number of option schools. We estimate model parameters using $4 + K + P + (N-1) \times 2 + N + T + \tilde{S} \times T + 1 = 1,028$ moments, which we define formally here. Below, we use $Pr(\cdot)$ to denote empirical probabilities that are obtained directly from the data—in particular from the ACS and the NCERDC data, as described in Appendix Section A.6.

B.1 Data moments

1. Average (over years) share of children that attend schools that do not provide transportation,

$$\frac{1}{T}\sum_t \frac{\#\mathcal{A}_t}{\#\mathcal{M}_t} \quad \text{where} \quad \mathcal{M}_t \text{ is the set of all students in year } t$$
$$\text{and} \quad \mathcal{A}_t = \bigcup_n \{i \in \mathcal{M}_t \mid i \in n \text{ and } s(i) \in \mathcal{NT}_{nt}\}$$

2. Average (over years and neighborhoods) distance to school attended conditional on transportation being provided

$$\frac{1}{\#\mathcal{A}}\sum_{i\in\mathcal{A}} \tau_{n(i)s(i)} \quad \text{where} \quad \mathcal{A} = \bigcup_t \bigcup_n \{i \mid i \in n \text{ and } s(i) \in \mathcal{B}_{nt} \cup \mathcal{T}_{nt}\}$$

3. Average (over years and neighborhoods) distance to school attended conditional on

transportation not being provided

$$\frac{1}{\#\mathcal{A}} \sum_{i \in \mathcal{A}} \tau_{n(i)s(i)} \quad \text{where} \quad \mathcal{A} = \bigcup_t \bigcup_n \{i \mid i \in n \text{ and } s(i) \in \mathcal{NT}_{nt}\}$$

4. Average peer quality in the school attended by a child with skills type k, for all k [K moments]

$$\frac{1}{\#\mathcal{A}_k} \sum_{i \in \mathcal{A}_k} \bar{a}_{s(i)} \quad \text{where} \quad \mathcal{A}_k = \{i \mid a(i) = a_k\}, \text{ for each } k$$

5. Average neighborhood income for households of type p, for each p [P moments]

$$\sum_n \bar{w}_n \times Pr(n \mid p), \text{ for each } p, \text{ where } Pr(n \mid p) \text{ is obtained from ACS data}$$

6. Empirical distribution of families and non-families across neighborhoods [$(N-1) \times 2$ moments]

$$\begin{aligned} & Pr(n \mid F) \text{ for each } n, \text{ where } F \text{ denotes families} \\ \text{and } & Pr(n \mid NF) \text{ for each } n, \text{ where } NF \text{ denotes non-families} \end{aligned}$$

7. Correlation across neighborhoods between minimum lot size and the share of neighborhood households with less than median income

$$\begin{aligned} \frac{1}{SD_{\text{mls}} SD_{\text{med}}} \frac{1}{N} \sum_n & \left\{ \underline{mls}_n - \left[\frac{1}{N} \sum_n \underline{mls}_n \right] \right\} \\ & \times \left\{ Pr(w \leq \text{med}(w) \mid n) - \left[\frac{1}{N} \sum_n Pr(w \leq \text{med}(w) \mid n) \right] \right\} \end{aligned}$$

where $\text{med}(w)$ is the median household income in the county,

$$SD_{\text{mls}} = \sqrt{\frac{1}{N} \sum_n \left\{ \underline{mls}_n - \left[\frac{1}{N} \sum_n \underline{mls}_n \right] \right\}^2},$$

$$\text{and} \quad SD_{\text{med}} = \sqrt{\frac{1}{N} \sum_n \left\{ Pr(w \leq \text{med}(w) \mid n) - \left[\frac{1}{N} \sum_n Pr(w \leq \text{med}(w) \mid n) \right] \right\}^2}$$

8. Average (over time) house prices in each neighborhood and average (over neigh-

borhood) house prices in each year [$N+T$ moments]

$$\frac{1}{T}\sum_t \text{price}_{nt} \ \text{ for each } n \quad \text{and} \quad \frac{1}{N}\sum_n \text{price}_{nt} \ \text{ for each } t$$

9. Admission probabilities to application schools [$\tilde{S} \times T$ moments]; these are directly available in the data (see A.6)

10. Regression coefficient of changes in house prices on changes in associated school quality; see Equation (2.3) and Table 1, column (2) in Section 2.3.

B.2 Model moments

Model-generated moments can be written as a function of the model parameters. Recall from Section 3.3 that:

$$\pi_{n|pk} = \frac{\exp(x_{npk})}{\sum_{\tilde{n}} \exp(x_{\tilde{n}pk})} \quad \text{with} \quad x_{npk} = u_{np} + \eta_p \alpha_n + \bar{v}_k(\mathcal{L}_n),$$

where $\bar{v}_k(\mathcal{L}_n) = \mathbb{E}_{\{\varepsilon_s\}}\left[\max_{s\in\mathcal{L}_n}\{\hat{v}_{k,s|n}\}\right]$, with $= \hat{v}_{k,s|n} = p_s v_{k,s|n} + (1-p_s)v_{k,\mathcal{B}_n|n}$. The school year subscript t is dropped to simplify exposition. The probability of choosing lottery s conditional on neighborhood n and child skills k,

$$\pi_{s|nk} = Pr\left[\hat{v}_{k,s|n} \geq \hat{v}_{k,\tilde{s}|n} \ \forall \tilde{s} \in \mathcal{L}_n\right],$$

does not have a closed-form solution and is estimated by simulation.
It follows that the probability that a family of type (p,k) chooses neighborhood n and applies to school $s \in \mathcal{L}_n$ is:

$$\pi_{ns|pk} = \pi_{s|nk} \times \pi_{n|pk}.$$

Note that if p_s is the admission probability to school s conditional on applying, then the probability that a family of type (p,k) chooses neighborhood n and attends school $s \in \mathcal{L}_n$ is:

$$\pi^{\text{att}}_{ns|pk} = \pi_{ns|pk} \times p_s.$$

Again, $Pr(\cdot)$ is used to denote empirical probabilities that are obtained directly from the data—in particular from the ACS and the NCERDC, as described in Appendix Section

A.6.
Then:

1. Average (over years) share of children that attend schools that do not provide transportation

$$\frac{1}{T}\sum_t \left\{ \sum_p \sum_k \left(\sum_n \sum_{s \in \mathcal{NT}_{nt}} \pi^{\text{att}}_{ns|pk} \right) \times Pr(p,k) \right\}$$

2. Average (over years and neighborhoods) distance to school attended conditional on transportation being provided

$$\frac{1}{T}\sum_t \left\{ \sum_p \sum_k \left(\sum_n \sum_{s \in \mathcal{B}_{nt} \cup \mathcal{T}_{nt}} \pi^{\text{att}}_{ns|pk} \times \tau_{ns} \right) \times Pr(p,k) \right\}$$

3. Average (over years and neighborhoods) distance to school attended conditional on transportation not being provided

$$\frac{1}{T}\sum_t \left\{ \sum_p \sum_k \left(\sum_n \sum_{s \in \mathcal{NT}_{nt}} \pi^{\text{att}}_{ns|pk} \times \tau_{ns} \right) \times Pr(p,k) \right\}$$

4. Average peer quality in the school attended by a child with skills type k, for all k's [K moments]

$$\frac{1}{T}\sum_t \left\{ \sum_p \left(\sum_n \sum_{s \in \mathcal{L}_{nt}} \pi^{\text{att}}_{ns|pk} \times \bar{a}_s \right) \times Pr(p \mid k) \right\}$$

5. Average neighborhood income for households of type p, for each p [P moments]

$$\frac{1}{T}\sum_t \sum_n \left[\sum_{\tilde{p}} w_{\tilde{p}} \times \tilde{\pi}_{n|\tilde{p}} \right] \times \tilde{\pi}_{n|p}$$

$$\text{with } \tilde{\pi}_{n|p} = \left(\sum_k \pi_{n|pk} \times Pr(k \mid p) \times Pr(F \mid p) + \pi^*_{n|p} \times Pr(NF \mid p) \right)$$

where F denotes families and NF denotes non-families

6. Empirical distribution of families and non-families across neighborhoods [(N −

1) $\times$ 2 moments]

$$\frac{1}{T}\sum_t\sum_p\sum_k \pi_{n|pk} \times Pr(p,k \mid F), \;\; \text{for each } n, \;\text{where } F \text{ denotes families,}$$

$$\text{and } \; \frac{1}{T}\sum_t\sum_p \pi^*_{n|p} \times Pr(p \mid NF), \;\; \text{for each } n, \;\text{where } NF \text{ denotes non-families}$$

7. Correlation across neighborhoods between minimum lot size and the share of neighborhood households with less than median income

$$\frac{1}{SD_{\text{mls}}SD_{\text{med}}}\frac{1}{N}\sum_n\left\{\underline{mls}_n - \left[\frac{1}{N}\sum_n \underline{mls}_n\right]\right\}$$
$$\times\left\{\sum_p \tilde{\pi}_{p|n} \times \mathbf{1}[p \leq \text{med}(w)] - \left[\frac{1}{N}\sum_n\sum_p \tilde{\pi}_{p|n} \times \mathbf{1}[p \leq \text{med}(w)]\right]\right\},$$

where $\text{med}(w)$ is the median household income in the county,

$$SD_{\text{mls}} = \sqrt{\frac{1}{N}\sum_n\left\{\underline{mls}_n - \left[\frac{1}{N}\sum_n \underline{mls}_n\right]\right\}^2},$$

$$\text{and} \quad SD_{\text{med}} = \sqrt{\frac{1}{N}\sum_n\left\{\sum_p \tilde{\pi}_{p|n} \times \mathbf{1}[p \leq \text{med}(w)] - \left[\frac{1}{N}\sum_n\sum_p \tilde{\pi}_{p|n} \times \mathbf{1}[p \leq \text{med}(w)]\right]\right\}^2}$$

8. Average (over time) equilibrium house prices in each neighborhood and average (over neighborhood) equilibrium house prices in each year [$N + T$ moments]

$$\frac{1}{T}\sum_t r_{nt} \;\; \text{for each } n \quad \text{and} \quad \frac{1}{N}\sum_n r_{nt} \;\; \text{for each } t$$

9. Admission probabilities to application schools [$\tilde{S} \times T$ moments], using q_s to denote the measure of students school s can accommodate (i.e., its capacity)

$$\frac{1}{q_{s,t}} \times \sum_k\sum_p\sum_n \pi_{s|n,k} \times \pi_{n|p} \times Pr(k,p)$$

10. Regression coefficient of changes in house prices on changes in associated school quality—obtained by estimating regression Equation (2.3) using model-predicted analogues of the right-hand and left-hand sides variables.

B.3 Moments values

Moment values obtained from the data and predicted by the model are shown in Table B-1.

Table B-1: Data vs. Model Moments Values

Parameter	Key moment	Data value	Model value	Dataset
$\kappa_{0,NT}$	share of children in schools w/o transp.	3.51%	3.51%	NCERDC
$\kappa_{1,T}$	average distance to school cond. on transp. (miles)	3.47	3.47	NCERDC
$\kappa_{1,NT}$	average distance to school cond. on no transp. (miles)	6.93	6.93	NCERDC
γ_k	avg. peer ability in sch. attended by type-k child (decile)			NCERDC
	k=1	1.63	1.62	
	k=2	1.72	1.72	
	k=3	1.76	1.76	
	k=4	1.79	1.79	
	k=5	1.85	1.86	
	k=6	1.95	1.96	
	k=7	2.02	2.02	
	k=8	2.09	2.10	
	k=9	2.21	2.21	
	k=10	2.41	2.41	
γ_1	regression coeff. of chg. in house prices on chg. in school quality	0.030	0.030	NCERDC/RE
η_p	average nbhd income for type-p household (1000$)			ACS
	w_1=11.7	59.2	59.1	
	w_2=14.2	61.7	60.4	
	w_3=18.7	63.7	62.8	
	w_4=21.9	63.7	62.8	
	w_5=24.5	62.8	61.9	
	w_6=27.2	67.8	67.0	
	w_7=30.2	68.7	67.8	
	w_8=33.7	65.6	64.7	
	w_9=37.2	68.4	67.6	
	w_{10}=42.4	68.3	67.4	
	w_{11}=51.1	70.0	69.2	
	w_{12}=65.0	72.0	71.3	
	w_{13}=83.4	75.8	75.2	
	w_{14}=101.3	77.9	77.4	
	w_{15}=129.4	82.1	81.8	
	w_{16}=184.6	88.8	88.9	
ψ	corr. MLS and share of hh w/ income $\leq$ median	-0.234	-0.231	ACS
H_n	avg equilibrium house prices over nbhd each year t ($/sqft)			RE
	t=2006	7.42	7.42	
	t=2007	7.33	7.33	
	t=2008	7.47	7.47	
Values not shown because of space (matched exactly)				
α_n	average neighborhood n share among families			ACS
α_n^*	average neighborhood n share among non-families			ACS
H_n	average equilibrium house prices in over years in each nbhd n			RE
q_s	admission probability p_s			WCPSS

The table shows the values taken by the moments in the data and the values predicted by the model. It also shows what main parameter each moment is informative about, and the main dataset the moment is computed from. RE stands for Real Estate transactions data; WCPSS for the school-level data provided directly by the WCPSS.

B.4 Inference

Estimates are obtained by solving:

$$\hat{\Theta} = \text{argmin}_{\Theta} \left[m^{\text{model}}(\Theta) - m^{\text{data}} \right]' \left[m^{\text{model}}(\Theta) - m^{\text{data}} \right].$$

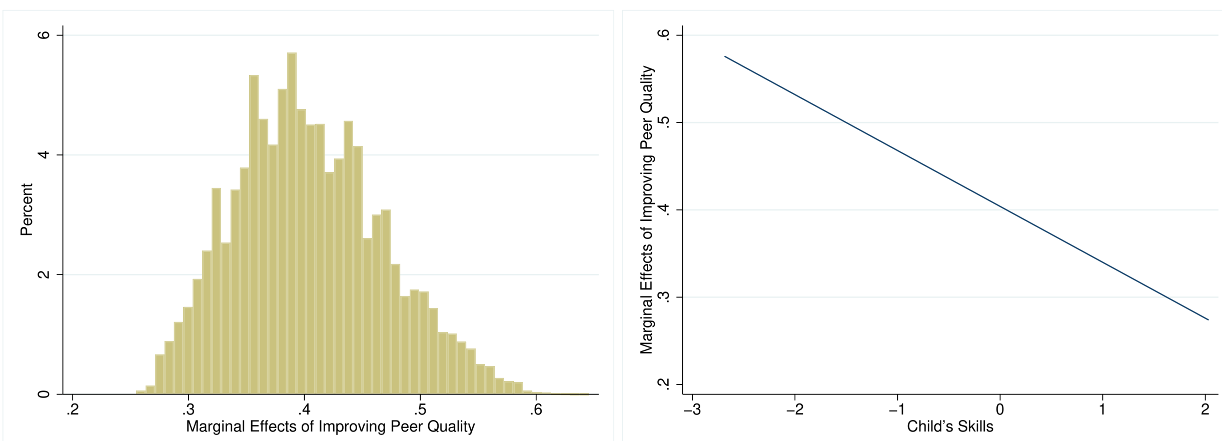
Standard errors are computed using the delta method. Denote h the mapping between estimates and data moments: $\hat{\Theta} = h(m^{\text{data}})$. If m^{data} is asymptotically normally distributed with variance matrix V^{data}, then $\hat{\Theta}$ is also asymptotically normal, with variance:

$$\Sigma_{\hat{\Theta}} = \nabla h \left(m^{\text{data}} \right)' \cdot V^{\text{data}} \cdot \nabla h \left(m^{\text{data}} \right),$$

where the element (i, j) of $\nabla h \left(m^{\text{data}} \right)$ is the partial derivative of the i element of $\hat{\Theta}$ with respect to the jth data moment: $\frac{\partial \hat{\Theta}_i}{\partial m_j^{\text{data}}}$. We compute $\nabla h \left(m^{\text{data}} \right)$ numerically, and derive V^{data} by the bootstrap.

B.5 Other Estimates

Figure B-1: Estimated Marginal Effects of Peer Quality on Children's Learning



(a) (b)

This figure shows the heterogeneity of peer effects according to the estimated regression model in column (2) of Table B-2. Panel (a) shows the cross-sectional distribution of the marginal effects of peer quality in our sample. Panel (b) shows the heterogeneity of the marginal effects of peer quality by initial children's skills.

C Counterfactual appendix

Table B-2: The Technology of Skill Formation

	(1)	(2)
	Outcome: Next-Period Log-Skills (t+1)	
Log-Child's Skills (ζ_1)	0.817	0.842
	(0.005)	(0.010)
Log-Peers' Skills (ζ_2)	0.402	0.404
	(0.140)	(0.141)
Log-Peers' Skills $\times$ Log-Child's Skills (ζ_3)		-0.064
		(0.019)
Intercept (ζ_0)	-0.125	-0.117
	(0.053)	(0.053)
Cragg-Donald Wald F-stat (First Stage Excl. Instruments)	961.55	478.87

The table shows the estimates for the technology of skill formation. In column (1) we estimate a simple technology where peer effects are linear. In column (2) we estimate the same model as in Equation (3.1), our preferred specification. Both models are estimated via instrumental variable estimators, where we allow both *Peers' Skills* and *Log-Peers' Skills* $\times$ *Log-Child's Skills* to be endogenous. We construct our instruments based on the exogenous variation in the potential pool of peers generated by variation in the school catchment areas (see Section 2.3 for the description of $\Delta_n \ln \textit{School Quality}_{n,t}$, and Panel B of Table 1 for the reduced-form effects on children's learning.). In the model in column (2) we interact $\Delta_n \ln \textit{School Quality}_{n,t}$ with the child's own skills (*Log-Child's Skills*) as an instrument for *Log-Peers' Skills* $\times$ *Log-Child's Skills*. All the regressions include both year and school fixed effects. Standard errors are robust to heteroskedasticity and reported in parentheses.

Table C-1: School Choice Expansion: Sending and Receiving Schools and Neighborhoods at Baseline

	Sending schools and neighborhoods			Receiving schools and neighborhoods		
	(1)	(2)	(3)	(1)	(2)	(3)
Total student share (%)	0.6	0.5	0.3			
Base school quality	1.04	1.05	0.94	3.12	3.11	3.36
Avg. income (in 1,000 $)	40.3	55.4	47.5	72.5	73.0	87.2
Avg. child ability	1.08	1.06	0.95	3.04	3.00	3.30
Avg. house price (in $/sqft)	7.18	6.65	6.53	9.10	9.85	9.57
Avg. zoning restriction (in sqft)	810	742	641	1365	1283	1417

The table shows baseline statistics about each of the three (one per column) sending and receiving base schools and their associated neighborhoods.

Table C-2: School Choice Expansion: Complete Results and Setting

	Transportation provided			No transportation		
	Closest	Middle	Furthest	Closest	Middle	Furthest
Dist. to receiving school (miles)	7.17	21.25	24.47	7.17	21.25	24.47
PANEL A: ENDOGENOUS NEIGHBORHOOD CHOICE						
Sending schools and neighborhoods						
Student take-up rate (%)	19.2	10.7	3.7	5.8	1.1	0.3
Change in base quality (%)	+14.6	+1.6	+0.5	-0.1	+0.6	+0.1
Change in avg. income (%)	+3.3	+1.5	+1.4	+1.4	+0.2	+0.1
Change in avg. child ability (%)	+17.3	+18.3	+8.7	+9.0	+2.2	+0.8
Change in avg. house price (%)	+0.3	+0.3	+0.1	+0.1	0	0
Receiving schools and neighborhoods						
Admission probability	0.686	1	1	1	1	1
Change in base quality (%)	-20.3	-1.2	0	-1.2	-0.2	0
Change in avg. income (%)	-2.7	-0.2	-0.1	-0.1	0	0
Change in avg. child ability (%)	-17.9	-0.9	0	-1.0	-0.1	0
Change in avg. house price (%)	-0.4	-0.1	0	0	0	0
PANEL B: FIXED NEIGHBORHOODS						
Sending schools and neighborhoods						
Student take-up rate (%)	22.0	8.5	3.1	5.2	1.0	0.3
Change in base quality (%)	−6.2	−10.8	−5.7	−7.4	−1.4	−0.6
Receiving schools and neighborhoods						
Admission probability	0.744	1	1	1	1	1
Change in base quality (%)	−4.1	−0.9	−0.1	−0.7	−0.1	0.0

The table shows the changes (from the baseline equilibrium) induced by the school choice expansion policy on sending and receiving neighborhoods. Pair 1 refers the pair of sending and receiving base schools located relatively close to each other (about seven miles, "Close Pair" in the main text); Pair 2 refers to the pair of sending and receiving base schools located far each other (about 25 miles, "Farther Pair" in the main text).

Table C-3: Effects of Upzoning on Targeted (Receiving) Neighborhoods and Non-Targeted Neighborhoods

	Targeted neighborhoods	Non-targeted neighborhoods
Share of families at baseline (%)	1.5	98.5
Chg. in neighborhood family income (%)	−17.0	+0.7
Chg. in base school quality (%)	−38.5	+2.8
Chg. in house prices (%)	−0.4	+0.1
Chg. in house prices—exog. school quality (%)	+0.7	−0.1

The table shows the changes (from the baseline equilibrium) induced by the upzoning policy in receiving neighborhoods. Effects are shown for neighborhoods directly targeted by the zoning changes (first column), and the rest of the county (second column).

Table C-4: Effects of Upzoning on Children's Skills

	ED households	Non-ED households
Child w/ baseline skills below median	.075	.024
Child w/ baseline skills above median	.146	-.011

The table shows the changes (from the baseline equilibrium) induced by the upzoning policy on children's skills, broken down by baseline skill level.