On the Spatial Determinants of Educational Access *

Francesco Agostinelli † Margaux Luflade ‡ Paolo Martellini §

August 2021

Abstract

We study the role of local institutions—that is, 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 use the estimated model to analyze three policies that aim to improve educational access to economically disadvantaged children: expanding school choice, providing housing vouchers, and upzoning residential neighborhoods. We find that the success of school choice expansion is contingent on integrating transportation services, and that the common assumption in the school choice literature of policy-invariant residential location would lead to opposite implications for the equilibrium change in school composition. The voucher program benefits eligible families, but the benefits fade in equilibrium as the policy is implemented on a large scale. Finally, upzoning is an effective policy in lowering inequality in school composition via a reduction in neighborhood income segregation.

JEL Classification: I24, R23, R31.

*We thank Ornella Darova for outstanding research assistance. We thank seminar participants at the Saieh Family Fellows Workshop (Becker Friedman Institute), University of Minnesota, University of Pennsylvania, and Urban Economics Association Conference. We also thank the Office of Student Assignment of the Wake County Public School District, in particular the office of Student Information, Demographic and Spatial Analytics, for sharing school boundary maps and helping us understand the WCPSS institutional setting. All errors remain 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), as well as the geography of cities (e.g., the geographical distance between homes and schools). Moreover, the ultimate outcomes of these policies are determined by the equilibrium response of households to the altered incentives with respect to neighborhood and school choice.

In this paper we study how local institutions—neighborhood school portfolios, school transportation services, and local housing regulations—interact with residential and school choices of families in the formation of the observed inequality in educational access. To 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 that comes from differential access to both neighborhood base schools and other local option schools; (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 to what extent the demand for neighborhoods and schools respond to changes in the school quality and the availability of school transportation, the key mechanisms motivating our theory. 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 (Wake County Public School System, WCPSS) spanning roughly 850 square miles, making 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 data sources. 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, their schooling attendance, test scores performance, and residential information. We merge this information with (i) information about school geographical enrollment boundaries, the choice set of option schools for each neighborhood, and admission probabilities for these option schools; (ii) the map of the school transportation system; and (iii), data on house prices and residential zoning regulations for the entire county.

Our first set of results documents how neighborhood demand responds to school quality. We exploit longitudinal variation in the school enrollment boundaries at the neighborhood level 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% change in elementary school quality induces a 0.3% 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.

We also show that the demand of families for schools depends on the availability of transportation services. By exploiting longitudinal changes in transportation arrangements for option schools within neighborhoods, our results highlight that once option schools start providing transportation to and from a specific neighborhood, families substitute away from their assigned neighborhood (base) schools, and they increase their demand for those option schools. This pattern of substitution between schools depends on the geographical location of schools: as distance from these option schools increases, the effect of the newly offered 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 their residential location. 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, 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 and school composition, distance between home and schools attended, and admission probabilities to oversubscribed schools. Our model replicates the segregation in neighborhood income as well as the heterogeneity in school composition in terms of children's skills. Moreover, the model is able to replicate the quasi-experimental reduced-form estimates on how the demand for neighborhoods and schools responds to changes in school quality and transportation. We find that both transportation and school quality affect school choice. To give a sense of the magnitude, an increase in distance to base school by one mile reduces attendance by 6%, conditional on neighborhood choice, which is around the same as a decline in quality by 10%. In addition, the value of school options also affects neighborhood choice. A decline in school quality by 10% lowers—depending on the child's skills—the attendance rate (for a given neighborhood choice) and the probability of choosing a given neighborhood by 2–6% and by 5–17%, respectively.

The estimated model is used to consider three different policies that are aimed at expanding educational access for economically disadvantaged children. In the first counterfactual policy, we study the consequences for residential sorting and school enrollment of expanding school choice for the poorest neighborhoods. We perform this by including the highest-quality schools as additional school options for children living in 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% of children starting to attend the newly available schools when transportation is not available—the policy becomes more successful once targeted families are provided with transportation service (take-up rate up to 20% for a school seven miles away from home).

We find this policy also has a positive effect on the local base school quality: by making the originally low-income neighborhoods more appealing, the policy generates an inflow of higher-income families (gentrification) that causes an overall improvement in the base schools of the originally disadvantaged neighborhoods. The latter outcome is a distinct feature of our model in which families choose both schools and neighborhoods. In contrast, we find biased policy predictions on the take-up rates of newly offered schools when residential location is held fixed, as it is commonly assumed in the school choice literature. Strikingly, a policy-invariant neighborhood choice would lead to a change in equilibrium school quality in the originally disadvantaged neighborhoods of opposite sign compared with the equilibrium of our model.

Our second counterfactual policy focuses on expanding educational opportunities for children via housing vouchers, which is a policy comparable to the Moving to Opportunity (MTO) program. In this exercise, we provide rent subsidies to low-income families to live in higher-income (high-quality schools) neighborhoods. This exercise sheds light on the importance of accounting for families' equilibrium response to the policy. While a single-family implementation of this policy generates positive results for low-income families that are in line with the previous results on the evaluation of the small-scale MTO program (see for example Katz et al., 2001; Kling et al., 2007; Chetty et al., 2016), a scaled-up version of the program induces significant equilibrium changes in residential sorting. We estimate that an additional 1% inflow of voucher-recipient families into eligible neighborhoods would cause an outflow of 0.83% of higher-income families from those same neighborhoods. As a consequence, these general equilibrium responses reduce the estimated returns of the voucher program.

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 housing size. The results of our upzoning policy highlight how housing-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 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). In a seminal paper, Bayer et al. (2007) were the first to propose an empirical framework to study the role of neighborhood school quality in families' residential choices. Our paper departs from Bayer et al. (2007) 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; Abdulkadiroglu et al., 2017; 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' where 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.[1] 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 distance home-school, school transportation, and zoning regulations.

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, together 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 about the extent to which zoning regulations restrict access to certain neighborhoods—and neighborhood-specific amenities—extend beyond this papers' focus on education inequality.[2] Closely 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 do not focus on transportation provision and school choice, which are key to our estimation of the determinants of educational access.

[1] 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.

[2] Kulka (2019) augments the model of Bayer et al. (2007) with zoning restrictions that limit low-income households' neighborhood choice. However, she abstracts from endogenous amenities (school quality) and school choice.

Finally, a large literature studies the formulation and identification of social interactions models with neighborhood and peer effects (see for example Brock and Durlauf, 2002; Durlauf, 2004; Calvó-Armengol et al., 2009; Durlauf and Ioannides, 2010; Blume et al., 2011, 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 (Agostinelli, 2018; Doepke et al., 2019; Agostinelli et al., 2020). Our paper builds upon this literature by endogenizing both the residential decisions of families and their schooling (investment) choices within a spatial equilibrium model with peer effects.

2 Data, Institutional Background, and Empirical Facts

We use data from the Wake County Public School System (WCPSS), which is the 14th largest school district in the United States.[3] We restrict our attention to the students attending public school. 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 provide guidance in the development of our equilibrium theory of residential sorting and school choice.

2.1 Institutional Background

Here, we provide a brief overview of public school choice in 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. Each address also determines the menu of *option* schools to which parents can seek admission for their child (e.g., magnet programs, calendar transfer programs).[4] Schools that are option schools for families in one location are typically base school for families in another location. In addition to determining the menu of schools their child is eligible to attend, the family's residential address also determines whether school transportation is provided to each available school.[5]

[3] In 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.

[4] 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, as well as the assignment mechanism in case of oversubscription, are described in Appendix A.1.

[5] 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 WCPSS's address look-up tool (https://wwwgis2.wcpss.net/addressL

Assignment of residential addresses to base and option schools is, during our period of interest, driven by two considerations. In theory, from the 2000-01 academic year and until 2011-12, WCPSS had the goal of ensuring socioeconomic balance within and across schools. Assignments of addresses to schools was supposed to serve the goal that no school have more than 40% of students eligible for free or reduced-price lunch (now designated as economically disadvantaged, or ED) nor more than 25% 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). Indeed, the Census shows that total population in Wake County increased by 42% from 2000 to 2010, translating into an increase in the public school student population of 48%.[6] Over our sample period, 24% of base schools experienced a change in their catchment area across years (see bottom panel of Table A-1 and Figure A-2 in Appendix A.1 for an illustration), and 24% 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 and the bottom panel of Table A-1).

2.2 Data

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 North Carolina Education Research Data Center (NCERDC). 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. WCPSS also shared school-level information such as the number of applications accepted and denied by each option school every year. We use data publicly available on zoning and real estate transactions from Wake County.[7] Finally, we use Census tract-level information on population counts and household income from the American Community Survey ACS) five-year estimates (2006–10).

We refer to student cohorts by the year they enter kindergarten—the year at which we assume in the model parents make neighborhood and school decisions. To provide motivating evidence in Section 2.3, we take full advantage of Cohorts 2003–04 to 2008–

ookup/, accessed August 2021) further illustrates the determination of school and transportation eligibility by the family's residential address.

[6]Authors' calculations from NCERDC data.

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

09. 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 by a change of majority in the WCPSS School Board, took place in 2009. We refer to this set of cohorts as our *empirical evidence sample*. The structural estimation requires information about admission probabilities to schools, which we only obtained for a subset of three years in our period of interest. Therefore, our *structural sample* (Section 4), is restricted to Cohorts 2003–04 to 2006–07. Here, we provide basic descriptive statistics about our samples of neighborhoods, students, and schools (focusing on the structural sample). We refer the reader to Appendix A for further details about the combination of the different data sources as well as sample restrictions and missing observations.

School boundaries and neighborhoods. In any school year t, each residential address point in the county, which we denote by its latitude and longitude coordinates (x, y), is associated with a *portfolio* of schools $\mathcal{L}(x, y; t) = (\mathcal{B}(x, y; t), \mathcal{T}(x, y; t), \mathcal{NT}(x, y; t))$, where $\mathcal{B}(x, y; t)$ is the base school associated with (x, y) in year t; $\mathcal{T}(x, y; t)$ is the set of option schools providing transportation to (x, y) in year t; and $\mathcal{NT}(x, y; t)$ is the set of option schools in the choice set of (x, y) that do not provide transportation to (x, y). We define *school boundaries* for year t to be the mapping from $\mathbb{R}^2$ to the set of all possible school portfolios that assigns residential address (x, y) to its portfolio at t, $\mathcal{L}(x, y; t)$. We define a (base) school s's *catchment area* for year t, denoted $\mathcal{C}_{s,t}$ as the set of all points (x, y) whose base school at t is s, that is: $\mathcal{C}_{s,t} = \{(x, y) \mid \mathcal{B}(x, y; t) = s\}$. Finally, we define a *neighborhood* as the union of all *contiguous* points sharing a common school portfolio across years. 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 (x, y) such that for each t, $\mathcal{B}(x, y; t) = \mathcal{B}_{n,t}$, $\mathcal{T}(x, y; t) = \mathcal{T}_{n,t}$, and $\mathcal{NT}(x, y; t) = \mathcal{NT}_{n,t}$. We will employ these definitions to describe the changes in the institutional setting that guide our motivating empirical findings and to characterize the environment of our model in Section 3.

We rank neighborhoods thus constructed by decreasing order of their student population and exclude the lowest ranked neighborhoods so as to keep 90% of the students. We also link our neighborhood sample to data on zoning regulations, real estate transactions, and with Census tracts characteristics. Figure A-5 in Appendix A.3 shows the final partition of Wake County into our neighborhoods, with 2000 Census tracts boundaries for comparison. The top panel of Table A-1 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 than the base (11 versus 3.7 miles on average). Of our neighborhoods, 15% experienced a change in base school over the sample period; almost all of them experience a change in eligibility and/or transportation to some option school. There is significant heterogeneity in minimum lot size (MLS) regulations across neighborhoods (63% 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.2.4), with an average lot size of .15 acres and a standard deviation of .24 acres per lot over our neighborhood sample. Accordingly, there is significant variation in actual average lot size (mean of .45 acres, standard deviation of .46 acres) and house size (mean of 2,107 sq. ft., standard deviate of 653 sq. ft.) across neighborhoods.

Students. From Cohorts 2003–04 to 2006–07, we drop student observations for which school attended, address, end-of-grade test score, or economically disadvantaged (ED) status (measured by eligibility for free and 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.[8] Our final sample contains 16,445 students over the three cohorts. Thirty percent of students in sample qualify as ED. ED students are significantly more likely to attend their base school than their non-ED counterparts (92% versus 81%). Conversely, non-ED students are much more likely to attend option schools with transportation than ED students. Attendance at option schools without transportation is equally low among ED and non-ED students (3.5%). 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 slightly lower than a standard deviation (-0.58 versus 0.40).

Schools. There are 87 elementary schools in our sample. The bottom panel of Table A-1 shows descriptive statistics for our sample of schools. While all 87 schools are base schools, 27 (33%) host an option program. Consistent with population growth being the main driver of school reassignments, there is significant variance in the share of economically disadvantaged students across schools.

2.3 Motivating Empirical Findings

Here we provide direct evidence in favor of our main hypothesis: that demand for schools and neighborhoods responds to changes in the quality of the educational services, namely the schools' peer composition and the availability of school transportation.

Neighborhood Demand and School Quality. We provide quasi-experimental evidence

[8] 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.

for the effect of the quality of school peers on house prices. If families value good schools for their children, an increase in the quality of a school will capitalize in the house prices of the school's catchment area. Alternatively, if parents do not value good schools, house prices should not respond to changes in school quality. 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, while $\ln \text{School Quality}_{n,t}$ 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. $\epsilon_{j,n,t}$ is an error term.

In this empirical exercise, we are particularly interested in the coefficient β_1, which we interpret as the average responsiveness of house prices with respect to marginal changes in the quality of the base school. Estimating directly the regression in (2.3) via OLS would lead to biased estimates for β_1 as changes in school peer composition in a given neighborhood n depends on two elements: (i) 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. 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 the school catchment area $\mathcal{C}_{s,t}$ *only*:

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

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 the previous (t-1) family composition. In other words, our constructed variable $\Delta\text{School Quality}_{n,t}$ measures the change in the composition of base school attendees, from one school year to the next, induced by changes in the base school catchment areas if nothing else had changed in between.

The newly specified regression model, which accounts for geographical unobserved heterogeneity between neighborhoods as well as potential endogenous sorting, is:

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

where $\Delta \ln r_{n,t}$ and $\Delta \ln \text{School Quality}_{n,t}$ denote the within-neighborhood longitudinal variation in average house (log) prices and average (log) school quality, respectively. The

Table 1: House Prices and School Quality

	(1)	(2)
	House Price Psf (Log)	
Changes In School Quality (Log)	0.33***	0.03***
	(0.01)	(0.01)
Observations	92,849	75,527
Model	Pooled	Exog. Longitudinal (First Diff)
Sale-Year F.E.	Yes	Yes
House Characteristics	Yes	Yes

The table shows the effect of school quality on house prices. 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. Standard errors are robust to heteroskedasticity and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

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 model (2.1) without neighborhood fixed effects. In Column (2), we instead estimate regression Model (2.3) using the exogenous changes in school quality defined in Eq. (2.2).

According to Column (1), a 10% change in school quality is associated with an increase in house prices by 3.3%. As this pooled regression exploits both cross-sectional and longitudinal changes in school quality, the estimated effect could confound 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 Column (2), which accounts for unobserved neighborhood heterogeneity and endogeneity in residential sorting. Column (2) shows that a 10% change in school quality induces neighborhood demand to increase, with an associated 0.3% change in housing prices, respectively.[9]

The results from our preferred specifications in Column (2) of Table 1 need to be interpreted as an intent-to-treat. As the change in elementary school quality only affects a fraction of the entire population—namely, families with children of elementary school

[9]The drop in observations from Column (1) to Column (2) is because the model in first difference loses the first year of data.

age—our results are most likely a lower bound for the actual average effect of school quality on neighborhood demand from families with pre-elementary school children.[10]

Transportation and the Demand for Schools. Although North Carolina law requires transportation to and from school for students who attend base schools, transportation is not required for students who attend option schools.[11]

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:

$$\pi_{n,t} = \alpha_0 + \alpha_1 \text{Transportation}_{n,t} + \lambda_n + \gamma_t + \eta_{n,t}\,, \tag{2.4}$$

where $\pi_{n,t}$ represents the fraction of children living in neighborhood n in year t who attend their assigned base school, while $\text{Transportation}_{n,t}$ is the number of local option schools 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 (2.4) 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 given the family's residential address.

Column (1) shows that if an additional option school starts offering transportation, the probability that children living in that neighborhood keep attending their base school declines by 5 percentage points (approximately 4% 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 ge-

[10] Using the 2010 American Community Survey data for Wake County, we calculated that these families represent approximately 4% of households in Wake County, North Carolina. Along the same line, 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 is not.

[11] In our sample, 22% of option schools do not provide transportation.

Table 2: School Attendance 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. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

ographical 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.[12]

The results highlight that the elasticity of base school demand with respect to option schools transportation depends upon the geographical distance to the additional schools offering transportation services. Column (3) shows our baseline results. An 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 file 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.

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

We also analyze how the demand for option schools is affected by the their own transportation provision. To do so, we specify the following regression model:

$$\pi_{s,n,t} = \beta_0 + \beta_1 \mathbb{1}(\text{Transportation})_{s,n,t} + \lambda_{s,n} + \gamma_{s,t} + \eta_{s,n,t}\,, \tag{2.5}$$

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.[13] $\eta_{s,n,t}$ 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. The results 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 on accessing educational opportunities for children. Geography also plays a role in the access to educational services in the city, as our results show that providing free transportation to faraway schools is not effective in boosting enrollments.[14]

3 Model

3.1 Overview

We model a city populated by a measure of families (with children) and non-families (without children). 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 can apply to conditional on their neighborhood choice. Schools are exogenously characterized by their capacity and location in the city, and endogenously differ in quality and admission probability. Home-school distance determines the disutility cost from commuting,

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

[14] When we replicate our analysis by families' ED status, we find similar quantitative estimates.

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.

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 recover family preferences for school quality from observed variations in house prices given that families represent only a fraction of the housing demand in the city. The presence of non-families is also relevant for the study of counterfactual experiments that trigger changes in neighborhood choice.

3.2 Environment

Demographics. The school district, or city, is populated by a measure $m + 1$ of households. A measure 1 of households have one child who is about to start elementary school. We refer to these households as families. A measure m of households do not have children in the relevant age group and will be referred to as 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 $\phi(w_p, a_k)$. 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 housing supply H_n. In addition, we capture 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, h_n^{mls}.

School Boundaries. Each neighborhood is associated with a portfolio of schools, $\mathcal{L}_n$, that comprises three sets: $\mathcal{B}_n, \mathcal{T}_n, \mathcal{NT}_n$.[15] 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 $s = \mathcal{B}_n$.

[15]Formally, $\mathcal{L}_n \equiv \mathcal{B}_n \cup \mathcal{T}_n \cup \mathcal{NT}_n$. See the discussion in Section 2.2.

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 $s \in \mathcal{T}_n \cup \mathcal{NT}_n$. 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.

Family Preferences and Timing of Choices. We divide the neighborhood-school choice 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,

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

where $\bar{a}_s$ is the average skill composition of the school, $d_{ns}(\cdot)$ is a utility cost from commuting, and ε_s is an idiosyncratic preference shock that follows a standard EV-Type 1 distribution.[16] 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).

At the moment of choosing among 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. 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

[16]Our assumption that school quality is given by its children's skill composition is in line with previous work including Epple and Romano (1998) 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)).

n is equal to

$$\bar{v}_k(\mathcal{L}_n) = \mathbb{E}_{\{\varepsilon_s\}}\left[\max_{s\in\mathcal{L}_n}\{\hat{v}_{k,s|n}\}\right] = \mathbb{E}_{\{\varepsilon_s\}}\left[\max_{s\in\mathcal{L}_n}\{p_s v_{k,s|n} + (1-p_s)v_{k,\mathcal{B}_n|n}\}\right], \tag{3.1}$$

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.[17] This neighborhood-specific measure of educational access is reminiscent of the commuting market access in Ahlfeldt et al. (2015).

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

$$U_{p,k,n}(c,h) = \psi\left[(1-\beta)\ln\left(\frac{c}{1-\beta}\right) + \beta\ln\left(\frac{h}{\beta}\right)\right] + \eta_p\alpha_n + \bar{v}_k(\mathcal{L}_n) + \varepsilon_n. \tag{3.2}$$

The first term, in square brackets, represents utility from consumption of a tradeable good c and housing h. The price of c is normalized to 1, while house price r_n is paid to absentee landlords outside the economy. The parameter ψ captures the relative importance of this bundle with respect to the other neighborhood attributes. The second term is the product of exogenous neighborhood amenities, α_n, and an income-specific valuation η_p.[18] The third term is the expected value from the portfolio of schools associated with neighborhood n and described in equation (3.1). 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

$$\begin{aligned} w_p &\geq c + r_n h \\ h &\geq \underline{h}_n = \max\{h_0, h_n^{\text{mls}}\}. \end{aligned} \tag{3.3}$$

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

[17] To avoid introducing additional notation, equation (3.1) includes the trivial lottery in which a child applies to the base school and obtains value $\hat{v}_{k,\mathcal{B}_n|n} = p_s v_{k,\mathcal{B}_n|n} + (1-p_s)v_{k,\mathcal{B}_n|n} = v_{k,\mathcal{B}_n|n}$.

[18] 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.

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. Second, we exclude all school-related variables from the utility of non-families. The resulting utility is then given by

$$\psi\left[(1-\beta)\ln\left(\frac{c}{1-\beta}\right)+\beta\ln\left(\frac{h}{\beta}\right)\right]+\eta_p\alpha_n^*+\varepsilon_n. \tag{3.4}$$

3.3 Equilibrium

The probability of choosing a lottery s for families of type k[19] in neighborhood n is:

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

which does not have a closed-form expression.[20]

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

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

with the associated housing demand

$$h_{np}=\begin{cases}\beta\theta_p/r_n & \text{if } \beta w_p\geq r_n\underline{h}_n\\ \underline{h}_n & \text{if } w_p>r_n\underline{h}_n>\beta w_p\\ 0 & \text{o/w.}\end{cases}$$

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

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

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

[20]The absence of 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.

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 and applying to school s is $\pi_{n,s|pk} = \pi_{n|pk}\pi_{s|n,k}$. Similarly, define

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

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, admission probabilities p_s, and school peers $\bar{a}_s$. In particular, let

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

be the measure of household of type p who lives 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 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

$$H_n = \sum_p \Pi_{np} h_{np}. \tag{3.5}$$

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,

$$p_s = \min\left\{\frac{q_s}{\Pi_s}, 1\right\}. \tag{3.6}$$

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

$$\bar{a}_s = \mathbb{E}_{\Pi_{k|s}}[a_k], \tag{3.7}$$

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 in a position 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, $\{\pi^*_{n|p}\}$, and vectors of aggregate variables $\{r_n\}$, $\{\bar{a}_s\}$, $\{p_s\}$ 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 (3.1), and choice of consumption, housing, and neighborhood to maximize the objective function (3.2), subject to constraints (3.3); and ii) non-families' maximization of the objective function (3.4), subject to constraints (3.3)

- the aggregate variables satisfy housing market clearing (3.5), school capacity constraints (3.6), and consistency of school composition, (3.7).

4 Model Estimation

We divide our set of parameters into two groups. The first one is calibrated following the existing literature, reduced-form evidence, or direct empirical observations. We also clarify some necessary normalizations in the characterization of households' utility functions. The second group of parameters is estimated inside the model by 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.

4.1 Parameters Set Outside the Model

We set the housing utility parameter to $\beta = 0.25$, in line with traditional estimates of the share of income dedicated to housing expenditure. 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 sq. ft.) in each neighborhood on the statutory neighborhood-level measure of minimum lot size (in acres), mls_n.[21] The implied minimum housing size is $\underline{h}_n = h_0 + 892 \times \text{mls}_n$, where $h_0 = 641$. Since mls_n spans from

[21] For each neighborhood, we calculate the minimum house size for newly constructed houses in the data, and we regress it on the statutory measure of 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.

0 to about 1 acre, the implied minimum housing constraint is up to about 1,500 sq. ft. in certain neighborhoods.

The relative measure of non-families is set to $m = 24$. We think of families in the model as those households with children that are four or five years old, which represent 4% of households in Wake County in the American Community Survey (ACS).[22] 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.

4.2 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 amenities for families and non-families, α_n and α_n^*; marginal valuation of amenities by income type, η_p; housing supply $H_{n,t}$ for each year t; and school capacity $q_{s,t}$ for each 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. We target the (i) average share of children that attend schools who 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 in order 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

[22] We calculate this statistics using the 2010 American Community Survey data. We used 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. These families represent the relevant cohorts in our model as their children are about to start elementary school, but are not yet enrolled—at which point they would not be subject to later changes in school assignment.

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.[23] 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 1. The mean amenities over neighborhoods—$\mathbb{E}[\alpha]$ and $\mathbb{E}[\alpha^*]$—are normalized to 0. 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.[24] [25] 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.

4.3 Estimates

Estimates for commuting cost parameters and the utility weight on housing and consumption are shown in Table 3. Panel (a) of Figure 1 shows estimates for families' valuation of school quality as a function of their child's skills (decile). Panel (b) shows estimates for families' valuation of neighborhood amenities as a function of their income. Figure 2 shows estimated neighborhood amenities and plots them against average income in the

[23]For households that are unconstrained everywhere, higher values of ψ generate lower demand for more expensive neighborhoods, but proportionally so across income levels.

[24]Formally, the housing supply satisfies $\ln H_{n,t} = \ln H_n + \ln H_t$.

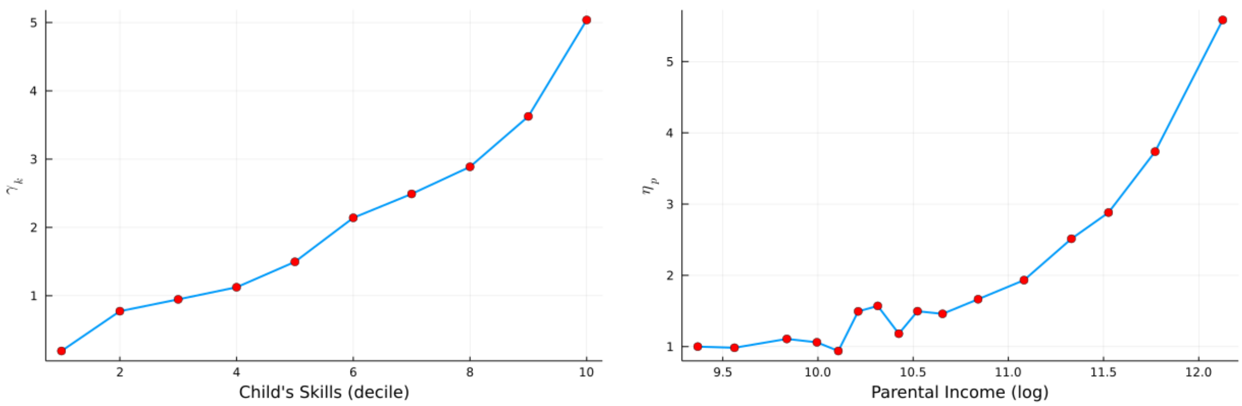
[25]Equilibrium house prices in the model are matched to user costs of housing in the data. See Appendix A.6 for the construction of user costs from housing sales data.

Table 3: Estimated Commuting Cost Parameters and Utility Weight on Housing Consumption

	Disutility From Commuting			Utility from Housing Consumption
	$\kappa_{0,NT}$	$\kappa_{1,NT}$	$\kappa_{1,T}$	ψ
Estimate	2.16	0.27	0.25	8.18

The table shows estimated values for commuting cost parameters and the utility weight on housing and consumption. Recall that the fixed cost of commuting with transportation ($\kappa_{0,T}$) is normalized to 0.

Figure 1: Heterogeneity in Preferences for School and Neighborhood



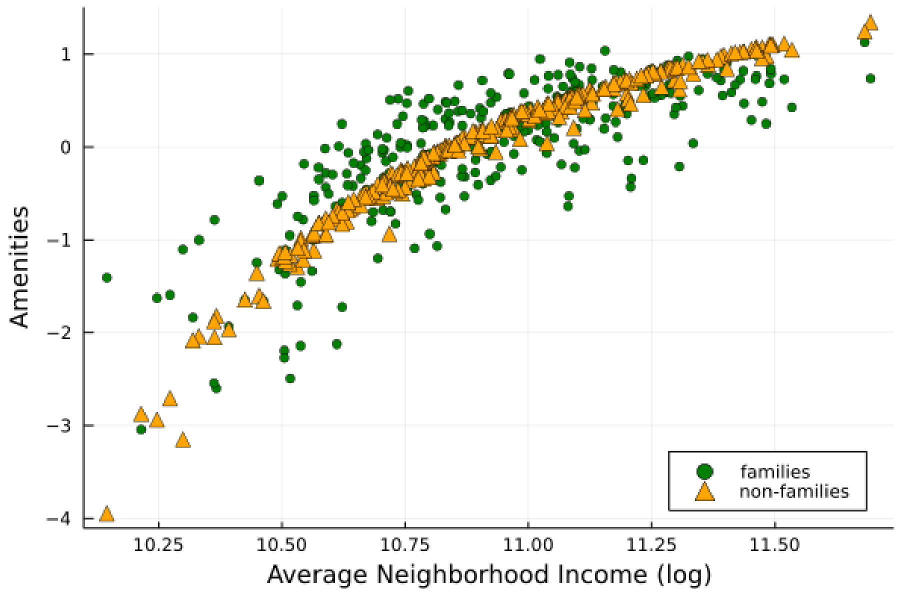
(a) Preferences for School Quality (b) Preferences for Neighborhood Amenities

Panel (a) shows estimates for families' valuation of school quality as a function of their child's skills (decile). Panel (b) shows estimates for families' valuation of neighborhood amenities as a function of their income. Note that η_1 for the lowest income bracket is normalized to 1.

neighborhood. We estimate that the absence of school transportation entails a large fixed cost, equivalent to the disutility of about ten additional miles (Table 3). The relationship between families' valuation for school quality and their child's skills level is monotonically increasing (Figure 1, Panel (a)), meaning that families with higher-skilled children value better schools more. Family's valuation for neighborhood amenities is estimated to be increasing (and convex) in family (log) income (Figure 1, Panel (b)). This heterogeneity contributes to generating income segregation across neighborhoods. In line with this heterogeneity, neighborhoods with higher estimated amenities tend to be disproportionately chosen by high-income families (Figure 2). 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 consumption and housing reported (Table 3).

In order to interpret the estimated parameters we show how demand for schools and

Figure 2: 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).

neighborhoods are affected by variations in the key determinants of choices. Specifically, Table 4 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. Table 5 shows how demand for the median-amenity neighborhood changes for high- and low-income families, with a high- or low-skilled child, as the characteristics of the neighborhood and its school portfolio change, holding everything else equal.

Table 4: Elasticity of School Demand as School Characteristics Change

	Below-median child skills	Above-median child skills
Dist. to base increased by 1 mile	0.94	0.94
Transp. to options removed	1.2	1.14
Base quality decreased by 10%	0.98	0.94

The table shows the ratio of the 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.

Table 4 illustrates the effect of distance, availability of school transportation, and school quality on school choice as residential neighborhood is being held fixed. Given transportation provision to the base school, increasing the distance from the neighborhood to its base by one mile (a 25% increase from the average neighborhood-to-base distance in the sample) decreases the probability that the residents of the neighborhood choose the base as their most preferred school decreases, on average, by 6%. Note that 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 3).[26] In contrast, 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 the other schools in their portfolio by 14–20%. To understand the larger increase in attendance at the base school for children with skills below the median, it is useful to know that non-transportation option schools (before the removal of transportation) are 24% better than the base and to note that school quality valuation is estimated to be increasing in the child's skills (Figure 1, Panel (a)). As a consequence of removing transportation to options schools, children with skills below the median mainly substitute toward the base school, while children with skills above the median are more likely to substitute toward high-quality, non-transportation option schools. In line with this idea, decreasing the quality of the base school by 10% translates into a decrease in choice probability by 2% for low- and 6% for high-skilled children.

While Table 4 helps show the relative importance of school quality, distance, and transportation in families' choice of schools, it only illustrates part of families' responses to changes in school characteristics. In the model, families choose school and neighborhood jointly, so changing the characteristics of the schools in a neighborhood's portfolio also affects families' probability of choosing the neighborhood in the first place. The first three rows of Table 5 illustrate this point, still taking as an example the median-amenity neighborhood in the county. Increasing the distance from the neighborhood to its base by one mile decreases demand for the neighborhood similarly across all income and skill groups, by about 15%. Looking at the extensive margin, keeping transportation to the base school but removing transportation from the neighborhood to all other school options decreases demand for the neighborhood by 15–20%. Regarding school quality, a decrease in the quality of the base school associated with the median-amenity neighborhood by 10% (bringing the school from the 76th to the 62nd percentile of the school quality distribution in the county), triggers a decrease in the demand for the neighborhood by about 5% for families with a child below the median of the skill distribution and 17% for families with a child above the median of the skill distribution, conditional on both low and high incomes. In the model, children's skills are the main driver of heterogeneity in family valuation of school quality. It is therefore not surprising that conditional on children's skills, demand responds similarly across income groups. Note however that

[26]Estimating a model in which commuting cost are allowed to differ by income (ED/non-ED) yields similar estimates for both groups.

Table 5: Elasticity of Neighborhood Demand as Neighborhood and School Characteristics Change

	ED families		Non-ED families	
	Below-median child skills	Above-median child skills	Below-median child skills	Above-median child skills
Dist. to base increased by 1 mile	0.85	0.85	0.85	0.85
Transp. to options removed	0.80	0.85	0.81	0.85
Base quality decreased by 10%	0.95	0.83	0.94	0.82
Adm. prob. reduced by half	0.91	0.88	0.89	0.86
Nbhd amenity decreased by 10 pctls	0.79	0.79	0.66	0.65
Zoning restriction increased by 10%	0.95	0.95	1.00	1.00

The table shows the ratio of 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.

because of the strong positive correlation between parental income and a child's skills, high-income families, on average, respond to this type of change more than low-income families.

Turning to the role of neighborhood characteristics in the determination of neighborhood demand, the fourth row of Table 5 shows that moving down the median-amenity neighborhood by 10 percentiles in the amenity distribution (bringing it to the 40th percentile) decreases demand for the neighborhood by 20% for low-income families and by 35% for high-income families. The heterogeneity in the response to changes in amenities reflects the fact that estimates for neighborhood amenity valuation increase with family income (Figure 1, Panel (b)).

Finally, the last row of Table 5 illustrates the role of constraints imposed by zoning regulations. Increasing the minimum lot size in the median-amenity neighborhood by 10% decreases demand for the neighborhood only for ED families (by 5%). This is because, as the minimum lot size increases, constraints (3.3) become binding for a larger share of low-income families, while leaving high-income families unaffected.

4.4 Model Fit

Table B-1 in the Appendix shows the in-sample fit. The model fits the data well. The model is able to replicate the residential income segregation in the city. In line with the data, non-ED families live in neighborhoods with an average household income that is approximately 20% higher than that of neighborhoods where ED families live.

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

Table 6: Quasi-Experimental Estimates: Model vs. Data

	(1)	(2)	(3)	(4)
	Model		Data	
Panel A: House Prices (Targeted)				
Changes In School Quality (Log)	0.036		0.03	
Panel B: Base School Enrollment (Untargeted)				
N Application Schools w Transp	−0.024	−0.046	−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 coeffients 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).

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% higher than the schools attended by children in the lowest decile of skill distribution.

In terms of school choice and transportation, we observe that the 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.

Finally, the model consistently predicts the level of housing prices in the city, as well as 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.

Replicating Quasi-Experimental Estimates within the Model. We additionally show how the model replicates some of the informative quasi-experimental reduced-form estimates we presented in Section 2.3. The 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.

First, 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 exogenous changes in school quality. The 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 6 shows that the model exhibits similar behavioral responses of housing demand to changes in school quality: a 10% change in school quality induced only by changes in the school catchment areas causes—both in the model and in the data—a surge in house prices of about 0.3%.

We want to validate a second important model response—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 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. 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 as 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. Panels B and C in Table 6 show the comparison of the regression coefficients in the model and in the data.

Overall, the model displays behavioral responses in terms of the demand for housing and schools similar to the one estimated in the data via quasi-experimental regression design in Section 2.3. 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, the high cost of commuting 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 mandating a minimal quantity of housing 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 families' access to high-quality 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 zoning. Toward this aim, we evaluate three distinct policies.

The first policy 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. In a second counterfactual analysis, we implement a voucher program that targets low-income families in neighborhoods with low-quality base schools and covers the difference between rhe cost of housing and 25% of the family's income (up to $8,000), conditional on moving to a neighborhood with a high-quality base school. Finally, our third policy counterfactual considers the effects on neighborhood and school quality of relaxing minimum house size regulations in tightly regulated high-income neighborhoods.

To facilitate comparison of results, we target the same sets of neighborhoods across all three policies. Specifically, we identify the three lowest-quality schools and the three highest-quality schools in the district. 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. Table 7 describes key characteristics of our three sets of sending and receiving neighborhoods (one for each base school). Receiving neighborhoods have significantly better schools and children skills, by construction, and around 40% higher neighborhood income. Particularly relevant for our counterfactual analysis, they also feature higher house prices and about 80% more-stringent housing regulations than sending neighborhoods.

In each of the counterfactual policy exercises, we assume that the capacity of base schools adjust to always guaranteed admission to all students in their catchment area.

Table 7: 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.

Every other exogenous element of the city (e.g., school boundaries) is kept to its value in a benchmark year, 2006. When discussing results, we contrast the outcomes of the proposed policies with the baseline equilibrium in that year.

5.1 School Choice Expansion

5.1.1 Policy Design

We consider a policy that allows residents of neighborhoods with low-quality base schools to attend one of the top schools in the district. Specifically, we pair each of the sending schools to one receiving school and assume that the capacity of each of the three receiving schools is increased by 10% and that the 10% are 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 available to them—then applicants from sending neighborhoods are selected by lottery (while students from receiving neighborhoods are always guaranteed admission to their base). We analyze this policy under two implementation scenarios: with and without transportation between the sending neighborhoods and their receiving school. In addition, to understand the importance of accounting for the joint choice of school and neighborhood when analyzing school choice policies, we compare results under two frameworks: (i) the framework of our model in which both residential neighborhood and school are chosen by households; and (ii) a framework in which the school choice expansion takes place (unexpectedly) after households have chosen their neighborhood and residential choice cannot be updated.

Table 8: 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: ENDOGENOUS NEIGHBORHOOD CHOICE				
Sending schools and neighborhoods				
Student take-up rate (%)	20.0	3.9	6.5	0.3
Share attending receiving base (%)	12.9[a]			
Change in base quality (%)	+17.6	+0.4	+1.0	0
Receiving schools and neighborhoods				
Admission probability	0.64	1	1	1
Change in base quality (%)	−14.9	0	−0.5	0
PANEL B: FIXED NEIGHBORHOODS				
Sending schools and neighborhoods				
Student take-up rate (%)	16.6	3.2	5.6	0.3
Change in base quality (%)	−23.7	−5.8	−6.5	−0.7
Receiving schools and neighborhoods				
Admission probability	1	1	1	1
Change in base quality (%)	−7.0	−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: $12.9 = 20.0 \times 0.64$.

5.1.2 Results

Key results under our model of endogenous neighborhood choice are shown in Panel A of Table 8. 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 neighborhoods, one is much closer to the sending neighborhoods than the other two. As a consequence, one of the expansion pairs is characterized by a much smaller distance between sending and receiving neighborhoods than the other two (seven miles versus more than 20 miles; see top row in Table 8). 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 of far-apart neighborhoods are similar to the latter). Table C-1 in the appendix provides more details about the results.

Without Transportation. Focusing first on the last two columns of Panel A of Table 8, 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 1–6% of children living in sending neighborhoods attend their assigned 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, the composition of the sending neighborhoods remains virtually unchanged—very few households are steered into choosing one of the sending neighborhoods because of the expanded school choice associated with it. 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 relatively constant.

With Transportation. Offering school transportation along with expanded school choice results in a very different picture, especially for our Close Pair (shown in Column 1). 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, while 12.9% of children from these sending neighborhoods attend their receiving school, (12.9/0.64 =) 20% take up the offer and apply for admission. Because households choose schools and neighborhoods jointly, the policy generates a response that shares some features with what is commonly referred to as gentrification. The high value of the added school option increases the demand for sending neighborhoods in the Close Pair (average house prices rise by 0.4%), particularly from families with high-skilled children who value school quality more highly (average children's skills in the sending neighborhood increases by 19.0%) and which tend to have higher household incomes (average income increases by 3.6%). Billings et al. (2018) document a similar phenomenon in Charlotte, North Carolina, where they find that, as the consequence of the fact that students whose base school fails to achieve its No Child Left Behind (NCLB) adequate yearly progress (AYP) are given higher priority to attend high-quality schools in the district, the composition of the failing school's attendance zone changes. Average income and house prices increase in the catchment area, while the probability of attending the initial base school decreases.

Interestingly, because the high-quality additional option school draws higher-skilled residents (children) to the neighborhood but admission to the option school is not guaranteed (0.64 admission probability), the average quality of the sending base school increases significantly (+17.6%).[27] While applicants are relatively high-skilled in their respective

[27]The nested assumption made in the model that idiosyncratic utility shocks for schools (ε_s) realize after households choose their residential neighborhood also contributes to the increase in the base school's quality. While families with higher-skilled children are drawn to the sending neighborhoods by the higher

sending neighborhoods, they have lower skills on average than children in the receiving neighborhoods. Consequently, because they now represent a significant share of the receiving school's student body (they fill up all of the extra 10% of seats), the average quality of the receiving school in the Close Pair decreases significantly (−14.9%). This decrease in base school quality triggers an opposite response in receiving neighborhoods to the one observed in sending neighborhoods: house prices decrease (−0.4% on average) and so do average neighborhood child skills (−13.1%) and income (−2.0%). 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 distance between the receiving school and sending neighborhoods; this shows that transportation provision does not erase all of the disutility families experience from having their child attend a faraway school.

5.1.3 The Role of Neighborhood Choice

To understand the importance of accounting for the joint choice of school and neighborhood when analyzing school choice policies, we turn to a framework in which families choose their residential neighborhood before (and not expecting) the school choice expansion policy, that is, based on the initial portfolio of schools attached to the sending neighborhood. Results are shown in Panel B of Table 8. Under school choice expansion, only those who highly value the added receiving school take up the opportunity to apply, and those who highly value the added higher-quality school tend to have higher-skilled children. Because under its initial school portfolio, the neighborhood attracted fewer high-skilled children than it would under its expanded choice portfolio, fewer families take up the policy than in the endogenous neighborhood setting. The receiving school is not oversubscribed, so all applicants are admitted. As a consequence, a very large share of the higher-skilled children in the sending neighborhoods leave their base school to attend the receiving school. In contrast to what happens under the endogenous neighborhood framework, when neighborhood choice is fixed, there is no inflow to the sending neighborhoods of new high-skilled children who would potentially end up attending the base school. The cream-skimming effect is severe: the base school quality in the sending neighborhoods decreases by a large amount (−24%), while it increases in the endogenous neighborhood setting. Symmetrically, the receiving school quality decreases less, both because of the lower take-up and the absence of a flight response from residents of the receiving neighborhood.

---

expected value of school choice associated with these neighborhoods (given by Eq. (3.1)), a share of them end up preferring the base school to the added high-quality option school ex post, once the ε_s's realize.

Table 9: Effects of the Voucher Policy 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 (%)		17		16	
Chg. in school quality (%)	28	167	25	152	2
Chg. in avg. nbhd income (%)	10	58	12	53	5
Chg, in avg. house price (%)	7	39	6	38	0.1

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

5.2 Housing Vouchers

5.2.1 Policy Design.

In the second counterfactual, we consider a policy reminiscent of the Moving to Opportunity (MTO) program that was implemented in the United States in the mid-1990s—although with a few key differences. We define as *eligible* the families who, in the baseline equilibrium, would live in one of the sending neighborhoods defined above—that is, a neighborhood in the catchment area of one of the three lowest-quality base schools in the district. Housing vouchers are granted to eligible families, conditional on their choosing to locate in one of the receiving neighborhoods—that is, a neighborhood in the catchment area of one the three highest-quality base schools. Conditional on eligibility, the amount of the voucher is to be spent on housing and is equal to the difference between the cost of housing and 25% of the family's income, up to $8,000.[28] We assume the policy is financed by a proportional income tax on all households in the county. We analyze this policy under two implementation scales. In the single-family experiment, only one of the eligible families, chosen at random, is offered participation in the program, and that family makes its neighborhood and school choices taking the baseline equilibrium as given. In this scenario, and as compared to the baseline, only the eligible family, if it does take up the voucher, experiences any change in its chosen neighborhood and school characteristics. Under the large-scale policy, all eligible families are simultaneously offered participation in the program, all households make housing and school choices, and a new equilibrium in the school district is obtained.

5.2.2 Results

Results are shown in Table 9, with the first two columns referring to the single-family implementation and the last three to the large-scale implementation.

[28]An average eligible family needs $8000 to be able to live in the average receiving neighborhood while spending a fraction $\beta = 0.25$ of their income on housing.

Single-family implementation. Seventeen percent of eligible families take up the offered voucher and, as a consequence, live in a house whose price is 39% higher than it would be at baseline. They also get to experience a neighborhood with an average income 58% higher and with a much higher quality base school (+167%).

Large-scale implementation. While average effects over the whole eligible population (intent-to-treat) are of similar magnitude as in the single-family implementation, gains are distributed differently when the policy is implemented at a large scale. Families who take up (16%) still experience the largest part of the gains, but these gains are partially muted by the general equilibrium response to the policy. The receiving neighborhoods have, in equilibrium, 53% higher average income and give access to base schools with quality 152% higher than the sending neighborhoods at baseline. The difference with the single-family effects is due to both a significant share of eligible families (with lower incomes and, on average, lower-skilled children) settling in the receiving neighborhoods, and flight by higher-income households with higher-skilled children who live there at baseline. We make this point precise by computing the elasticity of the share of non-ED families who live in receiving neighborhoods to the share of ED families in those same neighborhoods. We find a value of -0.83, which implies that a 1% increase in the probability that ED families choose to live in receiving neighborhoods due to housing vouchers is associated with a 0.83% decline in the probability that non-ED families live in those same neighborhoods.

Mostly due to prices in the receiving neighborhoods adjusting down, takers end up living in houses 38% more expensive than at baseline (compared to 39% in the single-family scenario). To gauge the size of this price change, it is useful to note that takers represent 0.1% of families in the county. The significant outflow of eligible families also affects sending neighborhoods, and therefore non-takers, in equilibrium. Average neighborhood income for non-takers increases by 5% as a result of takers not living in the sending neighborhood and the share of higher-income families in the neighborhood consequently increasing. By similar logic, base school quality in the sending neighborhood increases (+2%). This is because families who take up the voucher and move to a receiving neighborhood are economically disadvantaged and tend to have lower-skilled children. The resulting decrease in school quality in receiving neighborhoods and increase in sending neighborhoods mirrors the outcome of the school choice expansion policy with transportation.

Housing vouchers and zoning regulations. It is interesting to note that there is a positive correlation between the tightness of zoning regulations in receiving neighborhoods and the increase in their share of low-income families induced by the policy. Zoning regulations imply a minimum housing size ranging from about 750 sq. ft. in the bottom third of receiving neighborhoods ("loose neighborhoods") to roughly 1,500 sq. ft. in the top two thirds ("tight neighborhoods"). As a consequence of the voucher policy, the share of ED families increases by 1–10% in loose neighborhoods, and by 25–45% in tight neighborhoods. To understand the mechanism underlying these results, it is useful to look at the marginal utility of a dollar of a housing voucher. From Section 3.3, the indirect utility from consumption and housing when voucher v is offered is:

$$u_n(v) = \begin{cases} \log(w) - \beta\log(r) + \beta\log\left(1 + \frac{v}{\beta w}\right) & \text{if } \beta(w+v) \geq r\underline{h}_n \\ (1-\beta)\log\left(\frac{w+v-r\underline{h}_n}{1-\beta}\right) + \beta\log\left(\frac{\underline{h}_n}{\beta}\right) & \text{if } \beta(w+v) < r\underline{h}_n < w+v, \\ -\infty & \text{o/w} \end{cases}$$

and the marginal utility of the first dollar of a voucher is therefore:

$$\left.\frac{\partial u_n}{\partial v}\right|_{v=0} = \begin{cases} \frac{1}{w} & \text{if } \beta w > r\underline{h}_n \\ \frac{1-\beta}{w-r\underline{h}_n} & \text{if } \beta w < r\underline{h}_n < w. \end{cases}$$

Intuitively, the marginal utility of housing vouchers spent in a given neighborhood is higher when families are constrained, which is more likely to be the case the lower the family income, the higher the house prices, and the tighter the zoning regulations. Absent vouchers, most ED families would be constrained in their housing choice in tight receiving neighborhoods; hence, they disproportionately increase their presence in such neighborhoods once housing vouchers make them more affordable.

5.3 Upzoning

In this last section we further explore the role of zoning regulations on the inequality of educational access. Results from Section 5.2 suggest that housing vouchers are particularly effective in increasing demand for tightly zoned neighborhoods (with high-quality schools). This is because vouchers allow low-income families to afford the cost of living in these neighborhoods, where the minimum housing size is approximately 1,500 sq. ft.[29] For this reason, in this last section we explore the effects of upzoning our receiving neighborhoods.

We implement this experiment by reducing the regulated minimum housing size to 800 sq. ft. per dwelling unit in all of our receiving (high-income) neighborhoods, which represents approximately the average minimum house size constraint in the economy.[30]

[29] See Section 4.1 for details regarding the measurement of housing size regulations.

[30] In our sample, only five of the 30 receiving neighborhoods have at baseline a minimum regulated house

Table 10: Effects of Upzoning on Targeted (Receiving) Neighborhoods and Non-Targeted Neighborhoods

	Targeted neighborhoods	Non-targeted neighborhoods
Share of families at baseline (%)	4.2	95.8
Chg. in neighborhood family income (%)	−18.6	+0.7
Chg. in base school quality (%)	−39.1	+2.8
Chg. in house prices (%)	−0.5	+0.1
Chg. in house prices—exog. school quality (%)	+0.7	−0.2

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 columns).

The intention of this experiment is to analyze the counterfactual effects of loosening zoning regulations, which expands the opportunities for low-income families to live in the targeted high-income neighborhoods, and for their children to attend the local high-quality schools. On the one hand, this policy can change the socioeconomic composition 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 the original policy goal of neighborhoods and schools integration.

Table 10 shows the results. The policy is effective in changing neighborhood socioeconomic composition in the neighborhoods directly targeted by the policy: on average, neighborhood family income decreases by 18% 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 composition by 39%. This drop in school quality causes housing prices to drop (−0.5%).

Finally, the last row of Table 10 shows that upzoning would have instead caused house prices to increase by 0.7% in the targeted neighborhoods if the quality of the receiving base schools were kept exogenously constant at their baseline values. This result is a direct evidence for how zoning artificially reduces competition for houses in attractive neighborhoods, lowering the cost of accessing high-quality schools for high-income families.

6 Conclusions

We build and estimate a spatial equilibrium model of neighborhood and school choice that accounts for key institutional barriers to educational access in the United States: school attendance boundaries, heterogeneity in school transportation provision, and hous-

size lower than 800 sq. ft. In this case, we maintain the original regulation.

ing 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. On the other hand, housing vouchers can positively affect targeted families, but such policy loses impact on a 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.

Overall, our equilibrium model highlights how individual incentives impose constraints on the outcomes policymakers can achieve. Finding the efficient way to achieve certain policy objectives (e.g., reducing education inequality) through changes in local institutions like school boundaries and zoning regulations is a challenging but exciting avenue for future research.

References

Abdulkadiroglu, 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, "Investing in Children's Skills: An Equilibrium Analysis of Social Interactions and Parental Investments," 2018. Unpublished Manuscript, University of Pennsylvania.

__ , **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.

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

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., L. E., W. A. Brock, S.N. Durlauf, and Y. M. Ioannides, "Identification of Social Interactions," in J. Benhabib, A. Bisin, and M. O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, Amsterdam: Elsevier, 2011, pp. 853–964.

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

Brock, William A. and Steven N. Durlauf, "A Multinomial-Choice Model of Neighborhood Effects," *American Economic Review*, 2002, *92* (2), 298–303.

Brueckner, Jan K., *Lectures on Urban Economics*, Cambridge, MA: MIT Press, 2011.

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., "Neighborhood Effects," in Henderson J. Vernon and Jacques-François Thisse, eds., *Cities and Geography*, Vol. 4 of *Handbook of Regional and Urban Economics*, Elsevier, 2004, pp. 2173–2242.

_ **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 Richard E. Romano, "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects," *American Economic Review*, 1998, *88* (1), 33–62.

_ **and** _ , "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.

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

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.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman, "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, 2001, *116* (2), 607–654.

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

Kulka, Amrita, "Sorting into Neighborhoods: The Role of Minimum Lot Sizes," 2019. Unpublished Manuscript.

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.

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.

Poterba, James M., "Taxation and Housing: Old Questions, New Answers," *The American Economic Review*, 1992, *82* (2), 237–242.

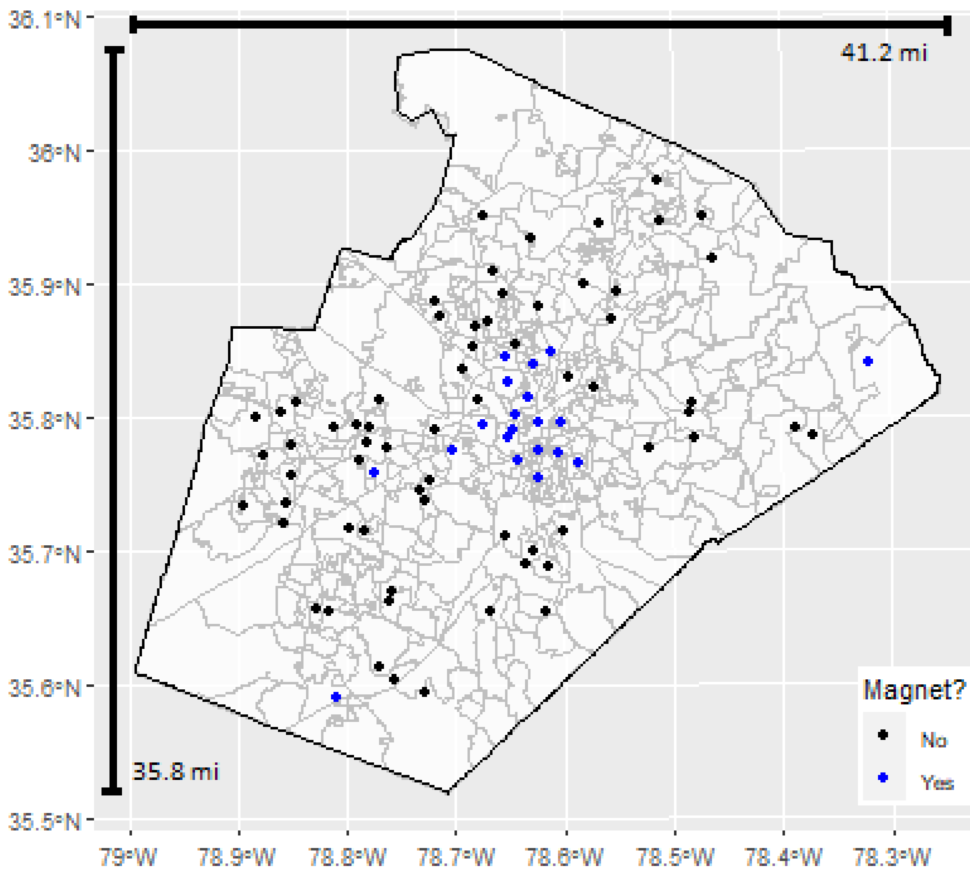
Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's Transmilenio," 2019. Unpublished Manuscript.

A Data appendix

A.1 Additional institutional details

WCPSS is the county-wide school covering Wake County, North Carolina, which is the county of the state capital, Raleigh. 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 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).

Figure A-1: Elementary School Locations in Wake County



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

Public school choice in 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

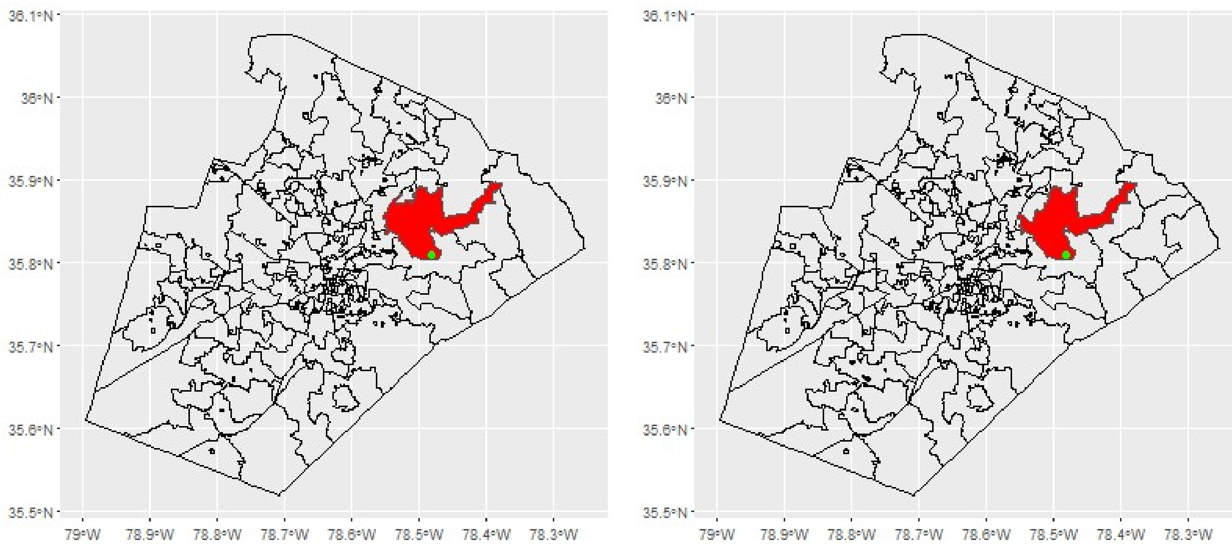
—although when we bring our structural model to the data, the two types of options are not differentiated and are pooled under the umbrella category of "option school".

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% 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, etc.) expected to draw white suburban students. Starting from the 2000–01 academic year, and until 2011–12, WCPSS moved from the goal of ensuring racial diversity in schools to that of ensuring socio-economic balance. Assignments of addresses to base schools was then supposed to serve the goal that no school had more than 40% of students eligible for free or reduced-price lunch (FRPL) nor more than 25% of students below the state's reading standards for their grade. While socioeconomic balance in schools was a target for the school board throughout 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.

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, 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.[31] Families can apply to up to three magnet programs, and assignment is made according the Boston Mechanism (for 90% seats in each school) or a pure lottery

[31] 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.

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.

(for the remaining 10% of seats in each schools). Magnet choice set and transportation provision do not only change cross-sectionally, they also change over time during the period of interest, with several magnets expanding and/or changing their transportation provision. We refer the reader to Appendix A, and Figure A-3 for more details about the magnet assignment process and changes over time. Twenty-four of 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-1).

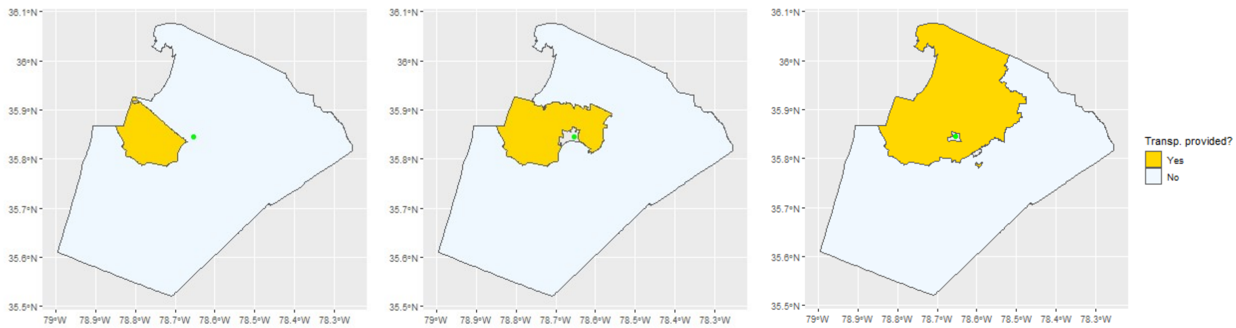
Calendar transfers allow students to attend a school running on a different calendar than their base school. Schools in WCPSS operate following one of two calendars—the traditional September to June academic calendar or a year-round calendar designed as a response to the rapid population growth to allow schools to accommodate more students at a time.[32] Each base school is paired with one alternative calendar school to which assigned families can apply and to which transportation will be provided.

Simplifying assumptions. To recap, here are the key simplifying assumptions we make in the model that depart from the institutional setting of WCPSS:

- No distinction between calendar transfers and magnet applications, both consid-

[32] 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.

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.

ered as a single type of "option school."

- 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.
- Assignment by pure lottery, with equal probability of admission among those eligible who apply, in contrast to a Boston mechanism in which priorities are determined as follows: "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)" (Dur et al., 2018, p.192).

A.2 Data sources

A.2.1 Student data

Student level data were obtained from the North Carolina Education Research Data Center[33] (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, a set of demographic variables (gender, race, economically disadvantaged status). Starting in 2006, the data also show the student's residential census block group and (a noisy version of) residential coordinates.

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

A.2.2 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 point in Wake County, the base school associated with the address point, calendar options for the address point, as well as the choice set of magnet programs the address point 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.

A.2.3 Real estate data

Publicly available records[34] from Wake County show details about all real estate transactions in Wake County starting from 1956. 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.

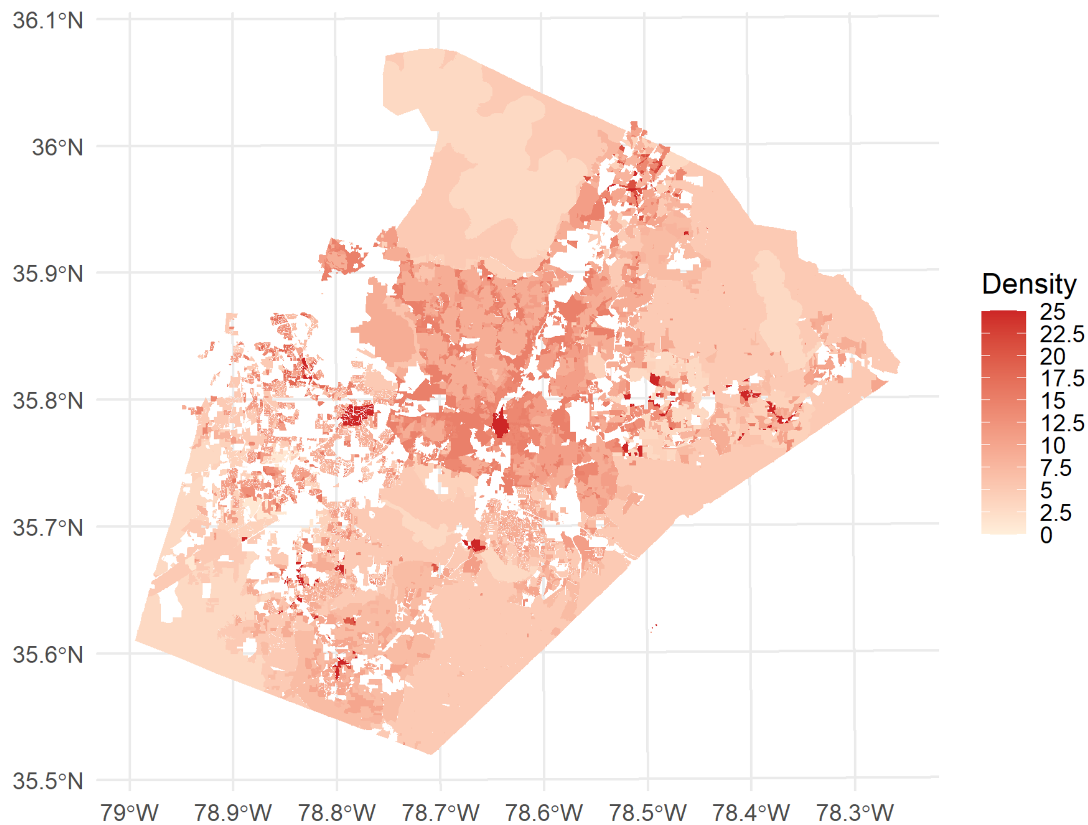
A.2.4 Zoning data

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 data set that gives, for any (residential) point in the county, the associated MLS regulation. Figure A-4 represents minimum lot size 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 minimum lot size, which is acre per lot. There is a relatively wide range of MLS regulations throughout Wake County—

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

from more than 25 du/acre in the urban center of the county, to less that 1 du/acre in the western periphery.

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



The figure shows density regulations throughout Wake County.

A.2.5 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 point 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 be made 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 exact 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% of the students. This ensures that (i) computations remain manageable, and (ii) all neighborhoods in sample contain students every year.

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

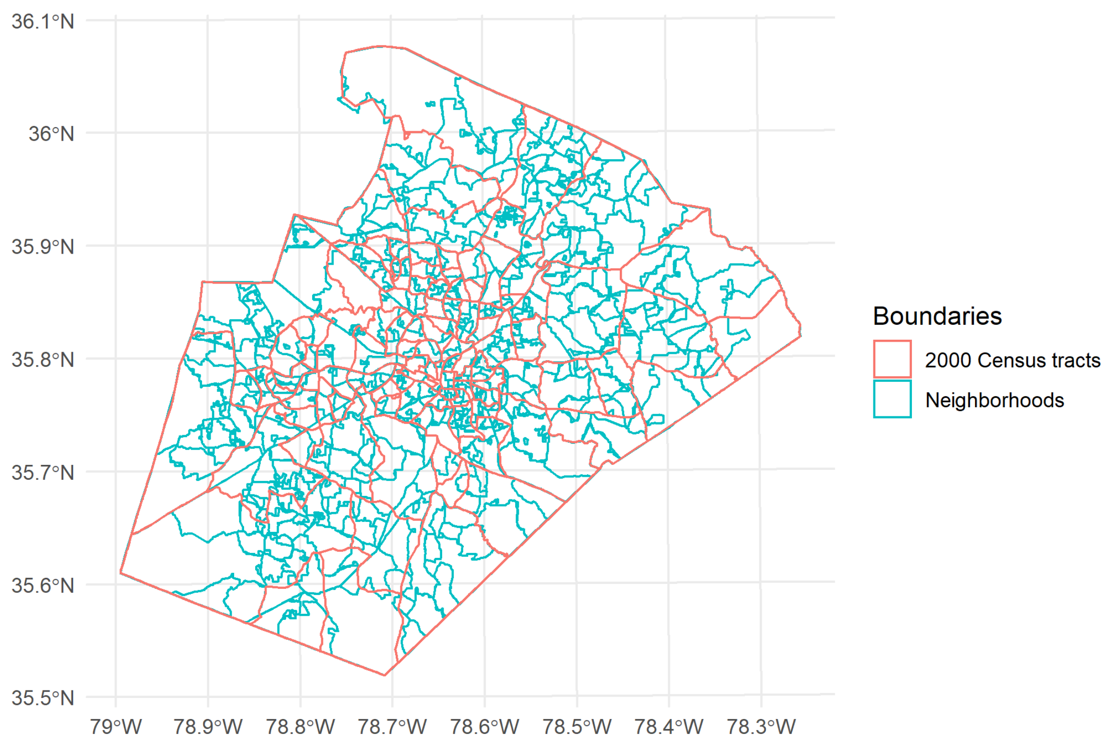
A.4 Final sample construction

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

1. First, 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

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).

student population, and exclude the lowest ranked neighborhoods so as to keep 90% of the students. This ensures that (i) computations remain manageable, and (ii) all neighborhoods in the sample contain students every year. Our final sample is consists of 305 neighborhoods.

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 Main descriptive statistics table

Table A-1 is the main table supporting the sample description in Section 2.2.

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

Table A-1: 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.

using the NBER TAXSIM program,[35] assuming the following household characteristics: married couple, spending 28% 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.[36] For reference, the eligibility thresholds for the school year 2007–08 for reduced-price lunch (below 185% of the federal poverty line) were $31,165 annual income for a household of three, and $38,203 for a household of four.[37]
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 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 reports 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:

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

- $Pr(n)$ 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 minimum lot size (MLS). For neighborhood overlapping with 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 housing 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. 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 the Falls of 2007 and 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 the Falls of 2007 and 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 in 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.[38]

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 these average sale prices per square foot into user cost of housing. For this, we follow Poterba (1992) and Brueckner (2011), and write:

$$R = (i + d + h - g) \times P,$$

where R is user cost, P is house value, i is the mortgage interest rate, d is the depreciation rate, h is the property tax rate, and g is the annual change in house values. In practice, we set the depreciation rate to 1.5 percent per year (Brueckner, 2011). We set i to the effective

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

interest rate in the FHFA's Monthly Interest Rate Survey (MIRS).[39] Property tax rates for each taxing unit within Wake County are given on the Wake County website for each year.[40] Finally, we use FRED data[41] for Wake County to set the annual change in house values, g.

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 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

[39] https://www.fhfa.gov/DataTools/Downloads/Pages/Monthly-Interest-Rate-Data.aspx (accessed August 2021) —See NC in the By State table of the Historical Summary Tables section.

[40] https://www.wakegov.com/departments-government/tax-administration/tax-bill-help/tax-rates-fees, accessed August 2021.

[41] https://fred.stlouisfed.org/series/ATNHPIUS37183A, accessed August 2021.

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]

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

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
$\kappa_{0,NT}$	share of children in schools w/o transp.	3.51%	3.52%
$\kappa_{1,T}$	average distance to school cond. on transp. (miles)	3.47	3.47
$\kappa_{1,NT}$	average distance to school cond. on no transp. (miles)	6.92	6.88
γ_k	avg. peer ability in sch. attended by type-k child (decile)		
	k=1	1.63	1.64
	k=2	1.72	1.72
	k=3	1.76	1.76
	k=4	1.79	1.79
	k=5	1.85	1.85
	k=6	1.95	1.95
	k=7	2.02	2.02
	k=8	2.09	2.08
	k=9	2.21	2.21
	k=10	2.41	2.42
γ_1	regression coeff. of chg. in house prices on chg. in school quality	0.030	0.036
η_p	average nbhd income for type-p household (1000$)		
	w_1=11.7	59.8	59.3
	w_2=14.2	61.2	60.7
	w_3=18.7	63.6	63.1
	w_4=21.9	63.7	63.1
	w_5=24.5	62.8	62.2
	w_6=27.2	67.8	67.3
	w_7=30.2	68.6	68.1
	w_8=33.7	65.6	65.0
	w_9=37.2	68.4	67.8
	w_{10}=42.4	68.3	67.7
	w_{11}=51.1	70.0	69.4
	w_{12}=65.0	72.0	71.4
	w_{13}=83.4	75.8	75.2
	w_{14}=101.3	77.9	77.3
	w_{15}=129.4	82.1	81.4
	w_{16}=184.6	88.8	88.0
ψ	corr. MLS and share of hh w/ income $\leq$ median	-0.234	-0.242
H_n	avg equilibrium house prices over nbhd each year t ($/sqft)		
	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	neighborhood n share among families		
α_n^*	neighborhood n share among non-families		
H_n	avg equilibrium house prices in over years in each nbhd n		
q_s	admission probability p_s		

The table shows the values taken by the moments in the data and the values predicted by the model. It also shows the main parameter each moment is informative about.

C Counterfactual appendix

Table C-1: School Choice Expansion: Complete Results and Setting

	Transportation provided			No transportation		
	Pair 1	Pair 2	Pair 3	Pair 1	Pair 2	Pair 3
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 (%)	12.9	11.0	3.9	6.5	1.1	0.3
Change in base quality (%)	+17.6	+1.3	+0.4	+1.0	+0.2	0
Change in avg. income (%)	+3.6	+1.4	+1.4	+1.6	+0.1	+0.1
Change in avg. child ability (%)	+19.0	+17.4	+8.5	+10.2	+1.9	+0.7
Change in avg. house price (%)	+0.4	+0.3	+0.1	+0.1	0	0
Receiving schools and neighborhoods						
Admission probability	0.64	1	1	1	1	1
Change in base quality (%)	-14.9	-1.1	0	-0.5	-0.1	0
Change in avg. income (%)	-2.0	-0.2	-0.1	-0.1	0	0
Change in avg. child ability (%)	-13.1	-0.8	+0.1	-0.5	-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 (%)	16.6	8.9	3.2	5.6	1.0	0.3
Change in base quality (%)	−23.7	−10.9	−5.8	−6.5	−1.5	−0.7
Receiving schools and neighborhoods						
Admission probability	1	1	1	1	1	1
Change in base quality (%)	−7.0	−0.9	−0.1	−0.7	−0.1	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).