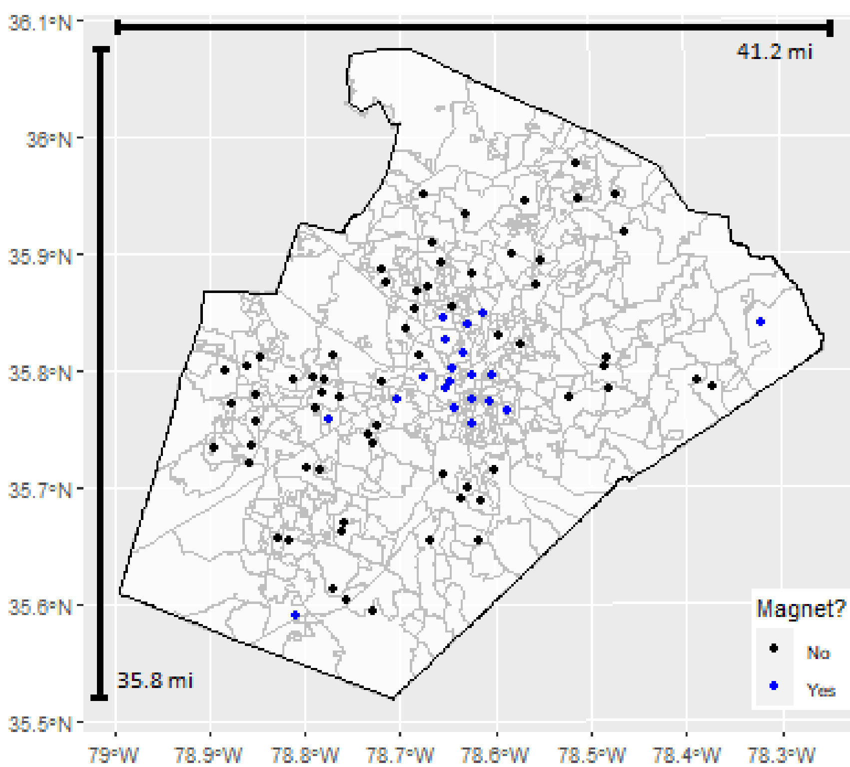
# Online Appendix for "On the Spatial Determinants of Educational Access" by Francesco Agostinelli, Margaux Luflade and Paolo Martellini

## A Data appendix

### A.1 Additional institutional details

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

Figure A-1: Elementary School Locations in Wake County




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

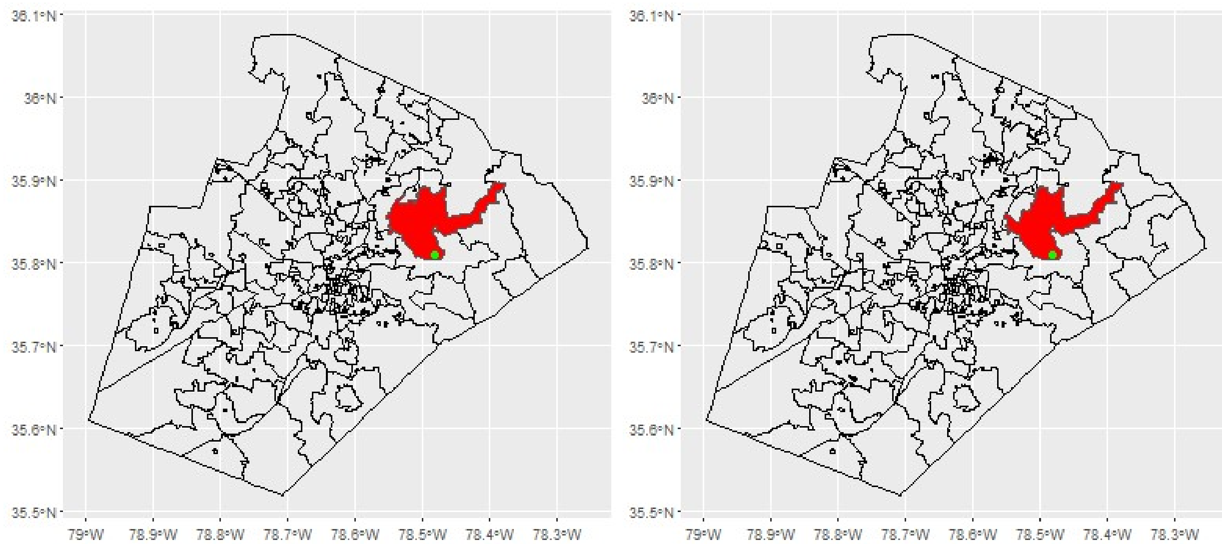
### A.1.1 Public school choice in the WCPSS

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

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

**Magnet programs** were created as a second instrument to facilitate racial and, then, socioeconomic integration in schools. Through these, a number of urban schools were endowed with special educational programs (e.g., arts, foreign language immersion, etc.) that were expected to draw white suburban students. In our period of interest, the WCPSS had 17 magnet programs at the elementary school level. Based on their residential address, parents can apply to a subset of these programs for their child. Also based on their residential address, parents may or may not be offered school transportation to the

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




(a) 2004–05

(b) 2005–06

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

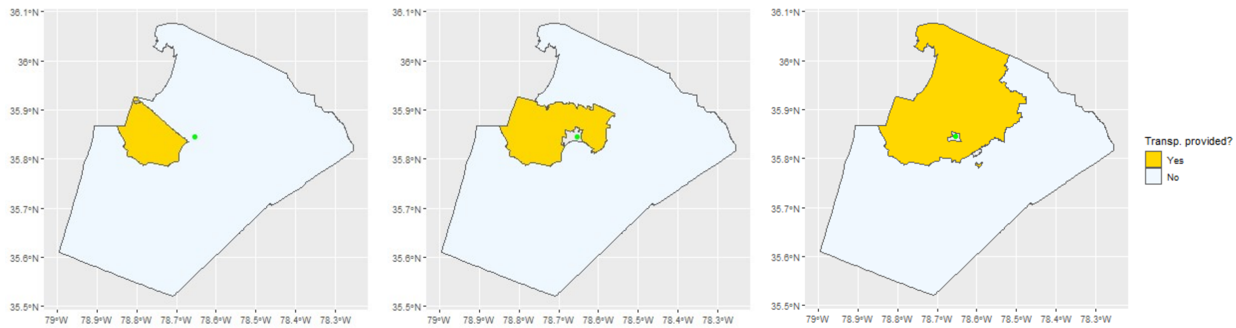
magnet program.[38] Families can apply to up to three magnet programs, and assignment is made according the Boston Mechanism (for 90 percent seats in each school) or a pure lottery (for the remaining 10 percent of seats in each schools). Magnet choice set and transportation provision do not only change cross-sectionally, they also change over time during the period of interest, with several magnets expanding and/or changing their transportation provision. Figure A-3 provides an illustration of transportation changes over time. Twenty-four magnet programs saw a change in the set of neighborhoods eligible to apply and/or in their transportation provision over the sample period (see bottom panel of Table A-2).

**Calendar transfers** allow students to attend a school running on a different calendar than their base school. Schools in the WCPSS operate following one of two calendars—the traditional September to June academic calendar or a year-round calendar designed as a response to the rapid population growth to allow schools to accommodate more students at a time.[39] Each base school is paired with one alternative calendar school to which

[38]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.

[39]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.

assigned families can apply and to which transportation will be provided.

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

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

### A.1.2 On the exogeneity of institutional changes

The identification of $\beta_1$ in Equation (2.3) (Section 2.3) and of the structural parameter $\gamma$ (Section 4) relies on within-neighborhood variation over time, and requires the changes in school quality induced by changes in catchment areas to be unanticipated by households. The parameters $\alpha_1$ and $\beta_1$ in Equations (2.6) and (2.7) (Section 2.3) provides empirical evidence on the elasticity of school demand to transportation provision. Its estimation relies on within-neighborhood changes over time in the transportation provision to the different option schools in the portfolio of the neighborhood, and supposes that changes are not anticipated by households.

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

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

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

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

## A.2 Data sources

**Student data.** Student-level data were obtained from the North Carolina Education Research Data Center[40] (NCERDC). The data show, for each year and each student enrolled in a North Carolina public school in grades 3–8, the school the child is enrolled in, end-of-grade test scores in math and reading, and a set of demographic variables (gender, race, economically disadvantaged status). Starting in 2006, the data also show the student's residential census block group and (a noisy version of) residential coordinates.

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

**Real estate data.** Publicly available records from Wake County show details about all real estate transactions in Wake County starting from 1956.[41] For each property sold, these

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

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

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

PANEL A: FAMILY CHARACTERISTICS

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

PANEL B: CHILDREN'S TEST SCORES

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

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

data show the sale price and date, exact address of the property, characteristics of the lot and of the buildings/units, if any. In particular, we use the following characteristics in the analysis: sale date, sale price, acreage of the lot, year the building was built, whether the building is for residential use, and its type (single-family house, apartment, etc.), and heated area. We use heated area as our measure of house size.

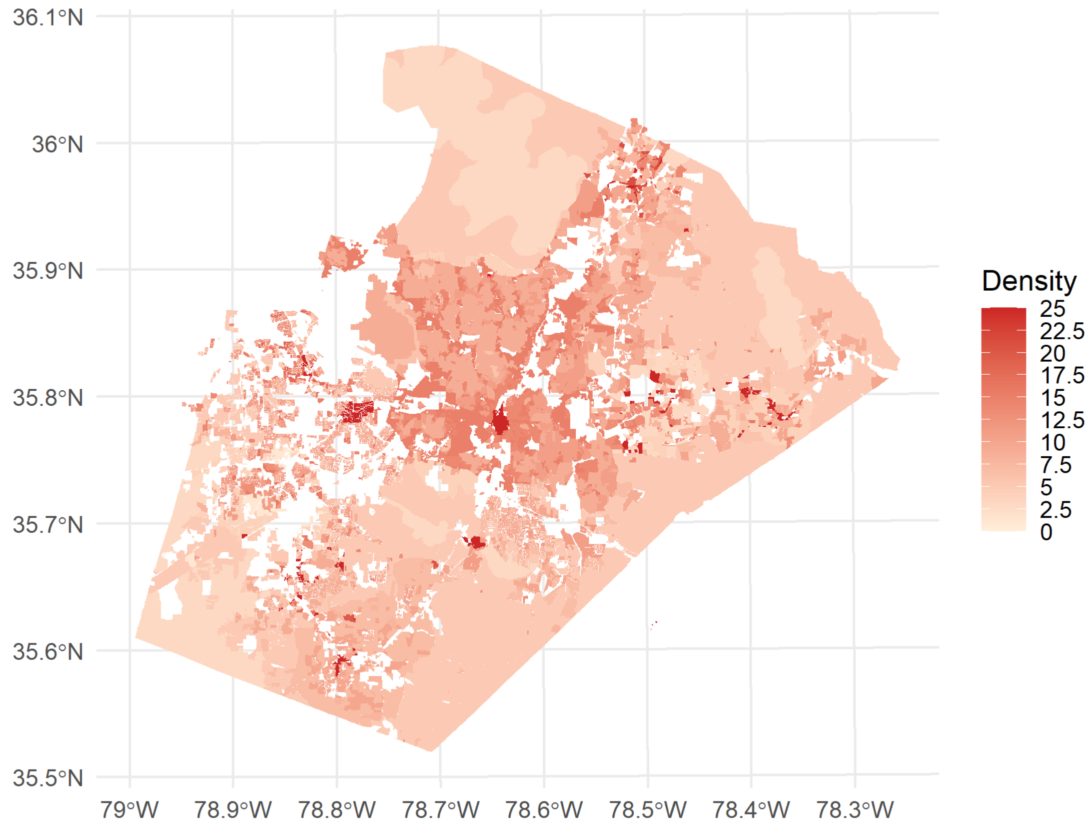
**Zoning data.** Multiple entities are in charge of zoning regulations in the county. While part of county land is regulated by the county itself, the zoning in other areas is done by a number of different local municipalities and/or unincorporated areas—namely: Raleigh, Apex, Cary, Fuquay-Varina, Garner,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

**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

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




The figure shows density regulations throughout Wake County.

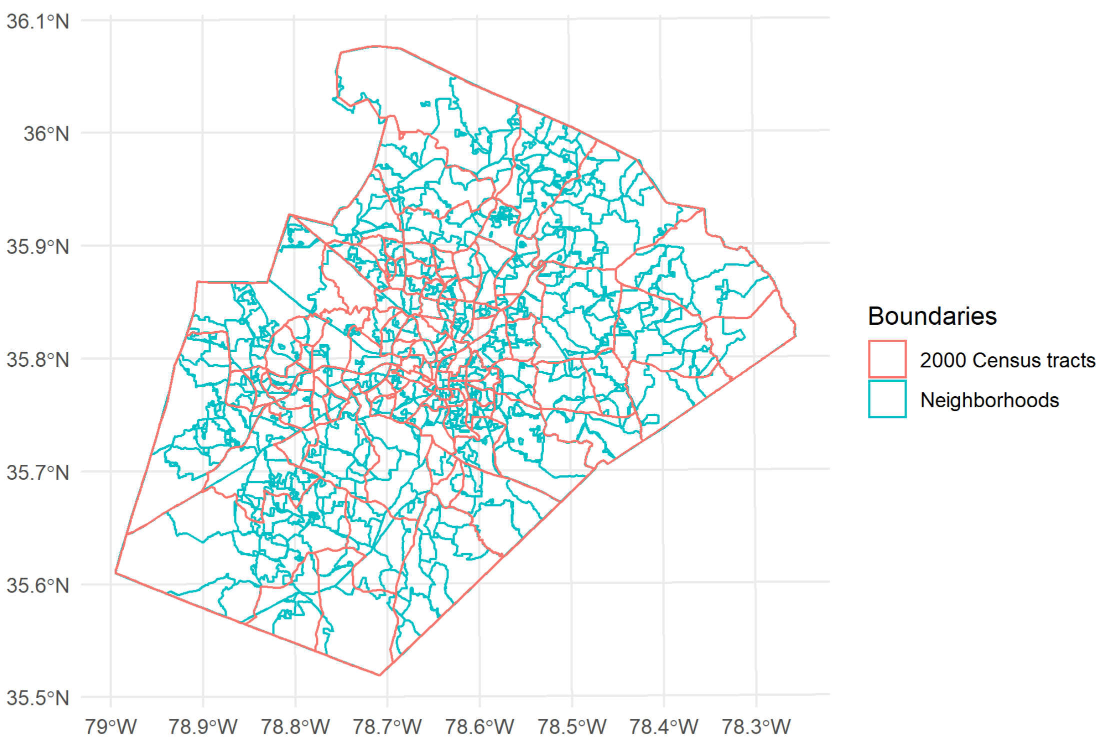
all *contiguous* points with school choice menu $\{(\mathcal{B}_{n,t}, \mathcal{T}_{n,t}, \mathcal{NT}_{n,t}) \mid t = 2003, \ldots, 2009\}$. Formally, let us denote each residential address by its coordinates $(x, y)$. $(x, y) \in n$ only if the following three points are satisfied:

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

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

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

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).

## A.4 Final sample construction

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

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

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

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

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

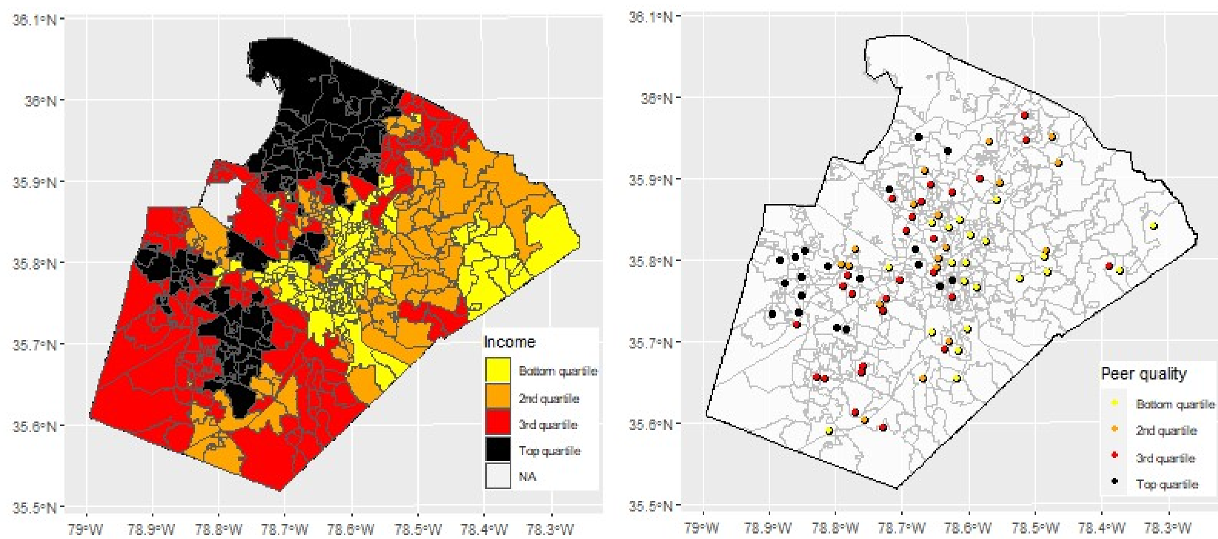
- Regarding neighborhood choice at school entry, we assume the residential address chosen at kindergarten entry is the same as the address observed in third grade.
- Regarding school choice at school entry:
  - If, in third grade, the child attends a school assigned to her neighborhood as a base or option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

then we assume the child entered kindergarten attending the base school attached to her neighborhood at the time.

## A.5 Descriptive statistics

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

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




(a) Household income (b) School quality

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

## 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

Table A-2: Descriptive Statistics

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

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

use the 16 ACS brackets as our discrete values for household type $p$ in the model. Household income $w_p$ for a household in bracket $p$ of the ACS is constructed in three steps. First, gross income is assumed to be the middle point of the ACS bracket $p$ (and \$250,000 for the top bracket "more than \$200,000"). Next, net income is obtained from gross income using the NBER TAXSIM program,[42] assuming the following household characteristics: married couple, spending 28 percent of their income on a mortgage, and with one dependent younger than 13. We use these household characteristics for all households in the model, that is families and non-families. To aggregate the ACS household income levels into the ED and non-ED categories available at the student level in the NCERDC data, we assign the seven lower ACS brackets (that is, with family income in the past 12 months below \$39,999 in 2010 inflation-adjusted dollars) to ED, and the nine higher brackets to non-ED. ED status in the NCERDC is determined by eligibility for free or reduced-price lunch. Income levels for eligibility to the programs are determined annually by the USDA.[43] For reference, the eligibility thresholds for the school year 2007–08 for reduced-price lunch (below 185 percent of the federal poverty line) were \$31,165 annual income for a household of three, and \$38,203 for a household of four.[44]
Average income in neighborhood $n$ is obtained as: $\bar{w}_n = \sum_p \text{mid}_p \times Pr(p \mid n)$, where $\text{mid}_p$ 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 report ED and non-ED as measures of socioeconomic status. On the other hand, the ACS, which shows population counts by income brackets, does not contain any information about children skills.

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

We infer the joint distribution as follows. Note that:

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

There:

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

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

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

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

**Admission probabilities $p_s$.** To estimate the model, we use information about admission probability in each option school for children entering kindergarten in Fall 2003, Fall

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

- if the number of applications received $\text{appli}_{st}^{kinder}$ and accepted ($\text{accept}_{st}^{kinder}$) or denied ($\text{appli}_{st}^{kinder}$ - $\text{accept}_{st}^{kinder}$) for kindergarten entry in year $t$ are observed, then we set $p_{st} = \frac{\text{accept}_{st}^{kinder}}{\text{appli}_{st}^{kinder}}$
- otherwise, if the number of applications received $\text{appli}_{st}^{all}$ and accepted ($\text{accept}_{st}^{all}$) or denied ($\text{appli}_{st}^{all}$ - $\text{accept}_{st}^{all}$) 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 $\tau_{ns}$.** 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.[45]

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

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

matches the average house size in the data when households spend a share $\beta = .25$ of their income on housing. In practice $K = 1/15$.

# B Estimation appendix

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

## B.1 Data moments

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

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

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

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

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

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

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

moments]

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

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

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

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

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

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

$$\frac{1}{SD_{\text{mls}}SD_{\text{med}}}\frac{1}{N}\sum_n\left\{mls_n-\left[\frac{1}{N}\sum_n mls_n\right]\right\} \times\left\{Pr(w\leq\text{med}(w)\mid n)-\left[\frac{1}{N}\sum_n Pr(w\leq\text{med}(w)\mid n)\right]\right\}$$

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

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

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

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

borhood 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,

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

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

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

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

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

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

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

## B.3 Moments values

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

Table B-1: Data vs. Model Moments Values

| Parameter | Key moment | Data value | Model value | Dataset |
|---|---|---|---|---|
| $\kappa_{0,NT}$ | share of children in schools w/o transp. | 3.51% | 3.51% | NCERDC |
| $\kappa_{1,T}$ | average distance to school cond. on transp. (miles) | 3.47 | 3.47 | NCERDC |
| $\kappa_{1,NT}$ | average distance to school cond. on no transp. (miles) | 6.93 | 6.93 | NCERDC |
| $\gamma_k$ | avg. peer ability in sch. attended by type-$k$ child (decile) | | | NCERDC |
| | k=1 | 1.63 | 1.62 | |
| | k=2 | 1.72 | 1.72 | |
| | k=3 | 1.76 | 1.76 | |
| | k=4 | 1.79 | 1.79 | |
| | k=5 | 1.85 | 1.86 | |
| | k=6 | 1.95 | 1.96 | |
| | k=7 | 2.02 | 2.02 | |
| | k=8 | 2.09 | 2.10 | |
| | k=9 | 2.21 | 2.21 | |
| | k=10 | 2.41 | 2.41 | |
| $\gamma_1$ | regression coeff. of chg. in house prices on chg. in school quality | 0.030 | 0.030 | NCERDC/RE |
| $\eta_p$ | average nbhd income for type-$p$ household (1000$) | | | ACS |
| | $w_1$=11.7 | 59.2 | 59.1 | |
| | $w_2$=14.2 | 61.7 | 60.4 | |
| | $w_3$=18.7 | 63.7 | 62.8 | |
| | $w_4$=21.9 | 63.7 | 62.8 | |
| | $w_5$=24.5 | 62.8 | 61.9 | |
| | $w_6$=27.2 | 67.8 | 67.0 | |
| | $w_7$=30.2 | 68.7 | 67.8 | |
| | $w_8$=33.7 | 65.6 | 64.7 | |
| | $w_9$=37.2 | 68.4 | 67.6 | |
| | $w_{10}$=42.4 | 68.3 | 67.4 | |
| | $w_{11}$=51.1 | 70.0 | 69.2 | |
| | $w_{12}$=65.0 | 72.0 | 71.3 | |
| | $w_{13}$=83.4 | 75.8 | 75.2 | |
| | $w_{14}$=101.3 | 77.9 | 77.4 | |
| | $w_{15}$=129.4 | 82.1 | 81.8 | |
| | $w_{16}$=184.6 | 88.8 | 88.9 | |
| $\psi$ | corr. MLS and share of hh w/ income $\leq$ median | $-0.234$ | $-0.231$ | ACS |
| $H_n$ | avg equilibrium house prices over nbhd each year $t$ ($/sqft) | | | RE |
| | t=2006 | 7.42 | 7.42 | |
| | t=2007 | 7.33 | 7.33 | |
| | t=2008 | 7.47 | 7.47 | |
| *Values not shown because of space (matched exactly)* | | | | |
| $\alpha_n$ | average neighborhood $n$ share among families | | | ACS |
| $\alpha_n^*$ | average neighborhood $n$ share among non-families | | | ACS |
| $H_n$ | average equilibrium house prices in over years in each nbhd $n$ | | | RE |
| $q_s$ | admission probability $p_s$ | | | WCPSS |

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

## B.4 Inference

Estimates are obtained by solving:

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

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

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

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

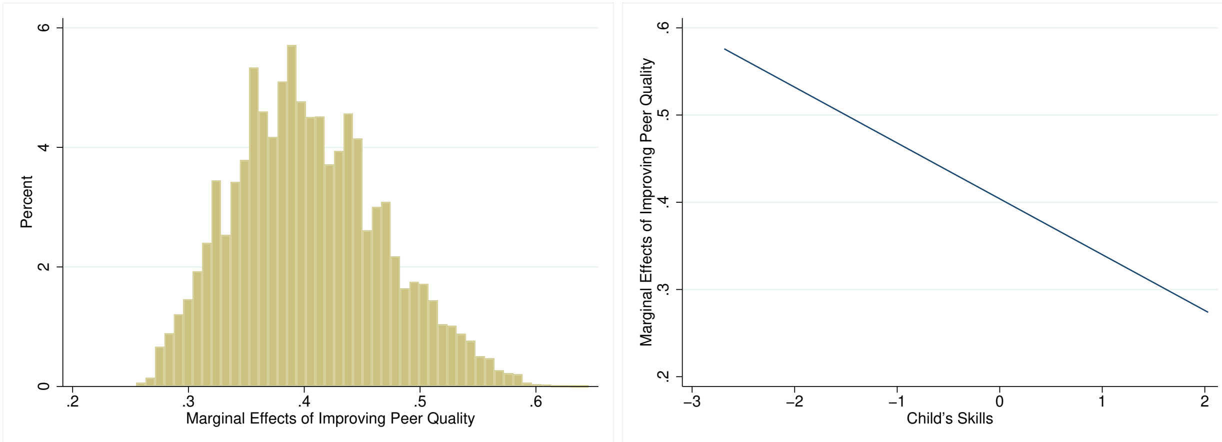
## B.5 Other Estimates

Table B-2: The Technology of Skill Formation

| | (1) | (2) |
|---|---|---|
| | Outcome: Next-Period Log-Skills (t+1) | |
| Log-Child's Skills ($\zeta_1$) | 0.817<br>(0.005) | 0.842<br>(0.010) |
| Log-Peers' Skills ($\zeta_2$) | 0.402<br>(0.140) | 0.404<br>(0.141) |
| Log-Peers' Skills $\times$ Log-Child's Skills ($\zeta_3$) | | -0.064<br>(0.019) |
| Intercept ($\zeta_0$) | -0.125<br>(0.053) | -0.117<br>(0.053) |
| Cragg-Donald Wald F-stat (First Stage Excl. Instruments) | 961.55 | 478.87 |

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

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




(a) (b)

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

# C Counterfactual appendix

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

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

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

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

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

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

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

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

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

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

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

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