Information Frictions in Education and Inequality *

Ana Figueiredo †

JOB MARKET PAPER

LINK TO THE LATEST VERSION

Abstract

Why does the place children grow up shape their opportunities in life? This paper studies the role of local information transmission as a potential explanation. Using school-district level data from Michigan, I uncover a novel empirical fact: when earnings of college graduates are sufficiently low, a higher share of college graduates living in a school-district is associated with lower college enrollment of students graduating from a high-school in that district. While this pattern is hard to reconcile with models featuring local funding of education or peer effects, I show that it is consistent with a simple model with local learning. I assume that individuals are uncertain about the skill-premium and learn about it by observing signals of the wages earned by nearby high-skill individuals. In this environment, a higher exposure to highly educated neighbors brings more information about the skill-premium. However, additional information only translates into more education if the observed wages generate a higher perception of the skill-premium. I use a calibrated version of the model to show that local information transmission is quantitatively important: it explains 60% of the inequality in college enrollment between children with different backgrounds.

JEL Classification: D80, E24, I24, J62

Keywords: Inequality, intergenerational mobility, education, information frictions

---

*I am thankful to Jan Eeckhout and Isaac Baley for their guidance and helpful feedback. I also thank Cristina Bélles, Andrea Caggese, Davide Debertoldi, Steven Durlauf, Jordi Galí, María Lombardi, Robert E. Lucas, Christopher Rauh, Cezar Santos, Nancy Stokey, Edouard Schaal, Laura Veldkamp as well as participants at CREi-UPF Macrobreakfast, BGSE PhD Jamboree 2015, SSSI Chicago 2015, XXI Workshop on Dynamic Macroeconomics, PET 16 Rio, SED Edinburgh and Lubramacro 2017 for their comments. I acknowledge financial support from Fundação para a Ciência e Tecnologia, PhD Grant SFRH/BD/94516/2013.

†PhD candidate in Economics, Universitat Pompeu Fabra, Barcelona - Spain, ana.gomesfigueiredo@upf.edu.
Most of what we know we learn from other people. (Lucas, 1988)

1 Introduction

Over the past three decades the college wage gap had a dramatic increase in the US (Autor et al., 2008). However, educational attainment still exhibits a stark socioeconomic gradient, generating persistent inequality across generations.\(^1\) Neighborhood context is key to understand this pattern (Chetty and Hendren, 2017). Potential channels include the local funding of education (Bénabou, 1996a,b; Fernández and Rogerson, 1996; Durlauf, 1996) and human capital spillovers (Bénabou, 1993; Cavalcanti and Giannitsarou, 2013; Bowles et al., 2014). Motivated by recent empirical contributions showing that inaccurate beliefs about earnings have a significant impact on educational decisions (Jensen, 2010; Kaufmann, 2014; Bleemer and Zafat, 2015; Hastings et al., 2016; Belfield et al., 2016), and that students from less advantaged backgrounds are those who are most affected by informational barriers (Hoxby and Avery, 2014), this paper proposes a novel explanation featuring imperfect information and local information transmission: individuals are uncertain about the skill-premium and learn about it by observing noisy signals of wage realizations of their neighbors. Using school-district data from Michigan, I show that this mechanism is able to explain empirical patterns that other channels in the literature cannot, and it is quantitatively important.

The analysis proceeds in three steps. I begin by uncovering a new fact: the relationship between the share of college graduates living in a school-district and college enrollment by high-school graduates from that district is only positive if earnings of those college graduates are sufficiently high; otherwise it is negative. This evidence is at odds with conventional wisdom which predicts this relationship to be positive regardless of earnings. I then show that the observed pattern can be explained through a simple model of human capital formation with imperfect information and local learning. The local nature of learning implies that the place where children grow up determines the pool of outcomes observed and, therefore, shapes their perceptions about the skill-premium. In this environment, exposure to high-skill individuals bears additional information about the skill-premium, which only translates into more educational investment if the observed wages increase the perception about the skill-premium. Finally, using a calibrated version of the model, I find that the proposed mechanism is quantitatively important: by itself, learning can explain 63% of the dispersion in college enrollment across school-districts.

To identify the heterogeneous correlation between college enrollment and neighborhood skill composition along the earnings dimension, I exploit within city variation in college enrollment, earnings of college graduates and the share of college graduates across school-districts using data from Michigan over the period from 2008 to 2014. My empirical findings take into account unobserved factors that may influence enrollment across cities and years as well as city level characteristics that might be trending over time (e.g. gentrification or deterioration of housing quality). Furthermore, I show that my result is robust across various specifications intended to address the possibility of sorting into school-districts.

\(^1\)In 2011, 83% of high school graduates who had at least one parent with a bachelor’s degree enroll in college, while the college enrollment rate of students whose parents had no more than a high school degree was only 54% (Bleemer and Zafat, 2015).
First, I find that my findings remain unchanged when I control for cohort characteristics using the share of females in the graduating cohort and the average ACT score, a proxy for students ability. Second, I find that my estimates are robust to controlling for socio-economic characteristics of the school-district (racial composition, median household income, unemployment rate, median earnings of high-school graduates, among others), as well as school quality, measured by expenditures, local revenues and teacher per student. This implies that my results are not masking the effects of better schools, credit constraints or differences in ability. Importantly, using a novel econometric estimation method to assess potential biases from unobservables developed by Oster (2016), I show that, under the assumption that selection on the observables is proportional to selection on the unobservables, my results are robust to omitted variable bias.

In the second part of the paper, I develop a theoretical framework that formally illustrates and quantifies the role of information frictions and local information transmission in explaining the observed pattern. In the model, children decide whether to invest or not in education. However, they have imperfect information about the returns to this investment and learn about it by observing noisy signals of wage realizations of their neighbors. Spatial location matters because it shapes children’s perception about the skill-premium. Neighborhoods differ in their distribution of skills through the location decision of parents, which is driven by exogenous amenities and dispersion forces in the form of inelastic house supply and taste heterogeneity. Consistent with the novel empirical evidence presented in this paper, I show that, in this environment, locations with a higher share of high-skill neighbors only have more children investing in education if the earnings of these high-skill neighbors are sufficiently high. More exposure to highly educated neighbors brings about more information, but additional information only translates into greater investment in education if it results in a higher perception about the skill-premium.

The simplicity of the model allows me to clearly highlight the mechanism at play. The distribution of amenities across neighborhoods determines the skill-mix of each neighborhood. Neighbors matter for the education decision through their level of education and earnings. In particular, the interaction between these two components determines the inference about the value of education. A higher share of high-skill neighbors reduces the cost of skill acquisition as in Bénabou (1993), Bowles et al. (2014) and Kim and Loury (2013), but it also generates additional information, which reduces uncertainty about the skill-premium, making children in that neighborhood more likely to invest in education. However, as a reaction to more information (more precise signals), children rely more on the information provided by their neighbors. Given this, the net effect of high-skill neighbors on education investment at the neighborhood level depends on the level of their wage relative to initial beliefs. If high-skill wages are sufficiently high, then children revise upwards their perceptions about the skill-premium. This generates a positive relationship between the share of high-skill neighbors and the probability of investing in education: as the share of high-skill neighbors increases, children are more certain that the value of education is higher than they previously though. In contrast, if

---

2To emphasize the potential role of the neighborhood’s skill-mix through the proposed learning mechanism, I make several assumptions that make the model simpler such as the exogeneity of amenities across locations. In section 3.6, I discuss the implications of each assumption and show that the main theoretical is robust to alternative model specifications.
high-skill neighbors’ earnings are sufficiently low, a higher share of high-skill neighbors makes children more certain that the skill-premium is low. In this case, there is a negative association between the share of high-skill neighbors and the share of children investing in education.

To evaluate the quantitative implications of imperfect information and the local information transmission mechanism, I discipline the parameters of the model by a set of moments that describe the wage distribution by educational attainment, the distribution of households, and college enrollment across neighborhoods in the city of Detroit in 2013. First, I show that the model provides a good fit for the data. Armed with the calibrated economy, I then ask the following question: by how much would the college enrollment rate change in the absence of the information disclosed by high-skill neighbors? I find that the learning mechanism plays an important role: if individuals did not observe any public signal from high-skill neighbors, instead of 42% of high-school graduates enrolling in college in Detroit, only 18% would. This result lies on the fact that learning from high-skill neighbors decreases uncertainty about the skill-premium by 28% and increases its expected value by 3%, on average.

I decompose the contribution of the different channels proposed in the literature besides local learning (human capital spillovers and school quality), and find that local learning and school resources have the largest effect on enrollment. However, the former is, by far, the most important channel in explaining inequality across school-districts. First, it accounts for 63% of the dispersion in college enrollment across school-districts. Second, it explains 60% of the difference between the probability of being high-skill for a child born to a high-skill family and the probability of being high-skill for a child born to a low-skill family. Finally, I simulate a disclosure policy that gives the correct information about the high-skill wage distribution. I find that correcting beliefs, while keeping differences in human capital spillovers and school quality across school-districts, increases college enrollment rate from 42% to 83%. Furthermore, it closes the gap in the probability of being a high-skill worker by 90%, which implies that children from low-skill households are almost as likely to become high-skill as children from high-skill families.

My results highlight the importance of imperfect information and local information transmission for the intergenerational propagation of inequality. Therefore, they have important policy implications. For policy-makers interested in addressing opportunity equality, understanding educational decisions is key to have a clear picture about the origins of persistence inequality across generations. In this context, assessing the role of information frictions is important as policies that reduce these frictions differ substantially from policies aimed at tackling liquidity constraints or school quality. Examples of the latter are the several housing reallocation programs implemented in different countries such as the well-known Moving To Opportunity in the US. In contrast, my results support broader information interventions, such as the one studied by Hoxby and Turner (2015), Bleemer and Zafat (2015) and Hastings et al. (2017), which make the returns to education more salient, especially among individuals from lower socio-economic backgrounds. Given the low cost of these information campaigns, the policy

---

3Moving To Opportunity (MTO) was a large experiment of the U.S. Department of Housing and Urban Development conducted between 1994 and 1998 in five large U.S. cities. Participant families were randomly assigned to one of the three groups: an experimental voucher group that was offered a housing voucher to move to a neighborhood with a poverty rate below 10%, a voucher group that was offered a housing voucher with no geographical restrictions, and a control group that did not receive a voucher. The vast empirical literature assessing the effects of MTO has found mixed results in several outcomes. For an overview of results from the several MTO studies see Topa and Zenou (2015).
case for implementing them is clear, specially when the success of other policies, such as subsidies or students loans, depends on whether children have have full information on education returns and costs.

Related Literature  This paper primarily relates to a theoretical literature that studies the role of residential location in determining intergenerational mobility and persistent inequality across generations. So far, the previous literature has focused on two main channels. One is the local financing of public schools (Bénabou, 1996b,a; Féméndez and Rogerson, 1996; Durlauf, 1996). Because schools are funded through property taxes, wealthier families segregate into homogenous communities and poor children attend schools with lower resources. The other channel is human capital spillovers. These spillovers have been modeled in different ways in the literature. In Akerlof (1997) and Akerlof and Kranton (2000, 2002), they correspond to the influence of identity. In locations where few parents are well educated, obtaining a high level of education may render the feeling of being alienated from those with whom one wants to share an identity. In contrast, Bénabou (1993), Bowles et al. (2014), Cavalcanti and Giannitsarou (2013) and Kim and Loury (2013) consider that either the skill acquisition technology or the cost of human capital formation depend on the human capital of the individual’s social network or neighborhood, without specifying a particular mechanism. Lastly, Mookherjee et al. (2010) suggest instead that residential location affects parents’ aspirations and, therefore, children’s occupational choice.

All these models share two common features. First, they suggest that the relationship between children’s educational outcomes and highly educated neighbors is positive, regardless of their earnings or other socio-economic characteristics of the neighborhood. Second, they assume individuals effectively know the current and future net returns of education. This is likely an important omission as recent studies show that individuals are uncertain about schooling returns and that perceptions about the value education and information constraints have significant impacts on different educational decisions (Jensen, 2010; Attanasio and Kaufmann, 2014; Kaufmann, 2014; Hoxby and Turner, 2014; Stinebrickner and Stinebrickner, 2014; Bobba and Frisancho, 2014; Wiswall and Zafar, 2015; Bleemer and Zafat, 2015; Belfield et al., 2016; Kunz and Staub, 2016). This paper fills this gap in the literature by introducing uncertainty about the skill-premium and local information into an otherwise standard model of human capital formation. Consistent with the documented empirical pattern, and in contrast with the previous theoretical literature, my results show that in this environment the association between highly educated neighbors and education outcomes is positive if and only if the labor market information they disclose leads to an increase in the perceptions about education returns. Otherwise, more exposure to highly educated neighbors translates into lower educational attainment.

---

4Bleemer and Zafat (2015) show that higher perceptions about college returns increase the probability of parents sending their child to university. The role played by perceptions and information on the decision to pursue further education is transversal to developed and developing countries. In the context of developing countries, Jensen (2010) finds that an intervention in the Dominican Republic which informs 8th grade students about returns increases school attendance. Also, Attanasio and Kaufmann (2014) and Kaufmann (2014) show that expected returns and risk perceptions are key determinants of education decisions in Mexico. In the US, Hoxby and Turner (2015) designed an intervention aimed to improve information of disadvantaged students at the college application stage and find that it made them more likely to submit applications and attend college. Using a unique survey of secondary students in the UK, Belfield et al. (2016) show that perceptions about the returns and the consumption value of education play a role in education decisions.
This paper also builds on a theoretical and empirical literature that studies environments with information frictions and social learning and shows how these affect agents’ decision making in different contexts such as technology adoption (Munshi, 2004), fertility decisions (Munshi and Mayaux, 2006), retirement savings (Duflo and Saez, 2003), female labor participation (Férrandez, 2013) and firms’ investment decisions (Fajgelbaum et al., 2016). Closely related to this paper is Fogli and Veldkamp (2011). They focus on explaining the rise of women’s labor force participation in a few locations that gradually spread to nearby areas, as information about the costs of working was transmitted locally. My model introduces a similar learning environment in a model of human capital investment with local interactions under uncertainty.

The facts presented in the paper also speak to an important and vast empirical literature aimed at studying the impact of neighborhoods’ socioeconomic environment on educational attainment of the young generation. This literature is reviewed in Durlauf (2004) and Topa and Zenou (2015). Despite being key to understanding the implications of the neighborhoods’ skill-mix, the existing literature does little to investigate heterogeneity in the effect of neighborhoods’ composition on students’ educational attainment. The exception is Gibbons et al. (2013) who finds no heterogeneous effects on test scores of students between age 11 and 14 across different location characteristics such as number of students or population density. This paper suggests, in contrast, that there are important heterogeneities along the earnings dimensions. Furthermore, while most of this literature (Oreopolous, 2003; Kling et al., 2007; Sanbomatsu et al., 2008; Gibbons et al., 2013; Chetty et al., 2016, among others) treats neighborhoods as “black box” in terms of the specific causal channels, I am able to shed some light on a specific mechanism through which the characteristics of a neighborhood affect educational attainment. My results suggest that information externalities as individuals learn about the benefits of schooling from the experiences of their neighbors seem to be part of the explanation. This mechanism is able to explain the observed regularities and seems to be quantitatively important.

Outline The paper proceeds as follows. Section 2 presents novel evidence regarding the relationship between neighborhoods and educational outcomes and Section 3 outlines a model with local information frictions to explain the observed pattern. In Section 4, I assess the quantitative importance of the proposed mechanism. Section 5 concludes.

2 Neighborhoods and Education: An Heterogeneous Relationship

In this section, I investigate the presence of heterogeneities in the relationship between neighborhoods’ skill composition and students’ education decisions along the earnings dimension.

2.1 Data Description

I use school-district level data from Michigan combining three sources of information.

College enrollment data comes from the Michigan Department of Education (CEPI). I measure college enrollment as the share of high-school graduates in public schools in a given school-district.
that enroll in a 4-year college within 6 months. This data also provides information on total number of student and cohort characteristics, namely students’ gender and race per grade and the average American College Testing (ACT) score at the school-district level.\(^5\)

**Socio-demographic data** comes from the Education Demographic and Geographic Estimates of the National Center for Education Statistics (NCES - EDGE). This data has rich information on the socio-economic characteristics of school-districts such as racial composition, family median income and unemployment rate. More important, this data has information on the median annual earnings by education level. In particular I observe the median earnings of individuals with a bachelor degree and with more than 4 years of college, and it has data on the education level of individuals over 25 years old.

**School-district financing data** comes from the Common Core of Data of the National Center for Education Statistics (NCES - CCD). Besides detailed information on expenditures and revenues, broken down by source (state, federal and local), this data has information on K-12 enrollment and the number of teachers in public schools per school-district.

My sample is an unbalanced panel of school-districts in Michigan urban areas in the 2008-2014 period. Notice that college enrollment as well as cohort characteristics and school financing data only cover public schools in each school-district. However, in Michigan a significant portion of the student body attend public schools, which mitigates concerns that the sample used is not representative of the whole student population in Michigan.\(^6\)

**Descriptive statistics** Table A.4 reports descriptive statistics for my sample, and shows that median earnings and the share of college graduates vary widely across districts, as do expenditures and revenues per student and student achievement, measured by the average ACT score and college enrollment rate. Table A.5 displays correlations among the main variables. College graduates’ earnings, the ACT score and the share of college graduates living in the school-district are highly and positively correlated with the share of high-school graduates that enroll in a 4-year college within 6 months of graduation. Local revenue per pupil is also positively correlated with college enrollment. However, note that expenditures per pupil show no correlation with this variable. Interestingly, expenditures per pupil exhibits a small, but negative, correlation with ACT score. The observed pattern seems to suggest that school resources play a small role in student achievement, as measured by college enrollment and the score in the ACT.

Panel A in Figure A.1 plots the average enrollment in each quartile of the share of college graduates’ distribution. The dashed line corresponds to school-districts in which college graduates have low

---

\(^5\)The ACT is a standardized test that measures high school students’ skills to complete college-level work in four different areas (english, math, reading, and science) and is used as a college entrance exam in the United States. There is one ACT score (1 to 36) for each test and a composite ACT score, which is an average of the four tests. In my sample I have information on the latter. More information here: [http://www.act.org/](http://www.act.org/).

\(^6\)For instance, in 2013 around 83% of total students were enrolled in a public school, the vast majority in a local neighborhood school (only 6.5% of those enrolled in public K-12 schools were in a charter or magnet school).
earnings, and the solid line represents school-districts in which college graduates have high earnings (i.e., those in the first and last quartile of the distribution of college graduates’ earnings). Panel B shows the difference between the dashed and the solid line and its 95% confidence interval. When the share of college graduates is low, the difference in college enrollment between locations where college graduates have high earnings and locations where their earnings are low is not significantly different from 0. As the share of college graduates increases, the difference between both groups of school-districts widens substantially. In particular, in the last quartile of the distribution of the share of college graduates, school-districts where college graduates have high earnings have an average college enrollment that is higher by 10 percentage points, when compared to school-districts where college graduates have low earnings. This pattern seems to suggest that exposure effects to highly educated individuals depend on the level of their labor earnings. This pattern could also be explained by liquidity constraints, in the next section I perform a formal analysis where I am able to control for this and other channels.

2.2 Empirical Methodology

To formally examine whether there is an heterogeneous relationship between the share of college graduates living in a school-district and the share of high-school graduates that enroll in a 4-year college along the earnings dimension, I estimate the following equation:

\[
Enrollment_{ijt} = \beta_0 + \beta_1 College_{ijt} + \beta_2 College_{ijt} \times Earnings_{ijt} + \delta X_{ijt} + \gamma_j + \gamma_t + \rho_j t + \epsilon_{ijt}
\] (1)

where \(Enrollment_{ijt}\) is the share of high-school graduates in public schools that enroll in a 4-year college within 6 months of graduation in school-district \(i\), city \(j\) and year \(t\), \(College_{ijt}\) is the share of individuals over 25 years old with a college degree living in school-district \(i\), city \(j\) and year \(t\), and \(Earnings_{ijt}\) corresponds to the median annual earnings of these individuals. \(X_{ijt}\) is a set of school-district controls, \(\gamma_j\) and \(\gamma_t\) are city and year fixed effects, and \(\rho_j\) is a city-specific time trend. \(\epsilon_{ijt}\) is the error term, that captures all unobserved determinants of college enrollment of school-district \(i\) in city \(j\) and year \(t\). I allow for arbitrary within-district correlation of the errors by clustering standard errors at the school-district level. Under the standard exogeneity restrictions, the effect of the share of college graduates living in the school-district on the college enrollment of high-school graduates is identified by \(\beta_1\) and \(\beta_2\),

\[
\frac{\partial Enrollment_{ijt}}{\partial College_{ijt}} = \beta_1 + \beta_2 \times Earnings_{ijt}
\] (2)

If \(\beta_1 > 0\) and \(\beta_2 = 0\), the effect of the share of college graduates is constant across different levels of earnings, in line with the theoretical literature that focuses on the role human capital spillovers as the mechanisms linking neighborhoods’ skill-mix to educational outcomes. In contrast, if \(\beta_1 < 0\) and \(\beta_2 > 0\) or \(\beta_1 > 0\) and \(\beta_2 < 0\), there is an earnings threshold above which the effect of the share of college graduates living in school-district on the college enrollment is positive, and below which is negative. Figure A.2 illustrates the effect of college graduates on college enrollment under different signs of the coefficients of interest.
OLS estimation of the effect of college graduates on college enrollment of high-school graduates using equation 2 controls for unobserved factors that may influence enrollment and are associated with the city to which the school-district belongs to as well as for time varying shocks affecting all school-districts. It further controls for unobserved city level characteristics that might be trending over time such as gentrification dynamics or deterioration in housing quality. The concern regarding the identification of $\beta_1$ and $\beta_2$ is that $\text{College}_{ijt}$ and $\text{Earnings}_{ijt}$ are correlated with $\varepsilon_{ijt}$, since individuals with different educational attainment tend to locate in systematically different neighborhoods, whose characteristics might lead to different decisions regarding college enrollment of high-school graduates. For instance, Bayer et al. (2004) and Bayer et al. (2007) find that individuals with a college degree are willing to pay $13.03 more per month than high-school graduates to live in a neighborhood with a higher school quality, as measured by average test scores. They are also willing to pay more for locations with higher population density, average income and a higher share of black residents.

The set of controls in $X_{ijt}$ mitigates these concerns. First, I control for the characteristics of the cohort that graduated from high-school in a given year by including (i) the share of females in the 12th grade, and (ii) the average ACT score of the graduating class. The latter is particularly important as it allows me to control for the fact that highly educated parents have children with higher ability, hence more likely to enroll in college. Because in 2007 Michigan implemented a mandatory ACT policy, which requires and pays for college entrance exams for all public school eleventh graders, the average ACT score is a good proxy for the ability of high-school graduates from public schools. Second, $X_{ijt}$ also includes school quality measures, namely expenditure, local revenue and teachers per pupil. Thus, the coefficients on $\text{College}_{ijt}$ and the interaction term are not capturing the effect of better schools in locations with highly educated adults as suggested by models that explore local funding of education as a mechanism that links neighborhoods to educational outcomes (Bénabou, 1996a,b; Férrandez and Rogerson, 1996; Durlauf, 1996). Equation 2 also controls for socioeconomic conditions at the school-district level such as the the share of black and white residents, the median annual family income, the unemployment rate and the median earnings of high-school graduates. Finally, I also control for location attributes by including population density.

The key assumption for causal interpretation of $\beta_1$ and $\beta_2$ is then that unobserved determinants of college enrollment are mean-independent of the share of college graduates and their earnings, conditional on the controls included. I discuss the plausibility of this interpretation further in section 2.5.

### 2.3 Results

I start by estimating a specification of equation 2 that does not include the interaction term, $\text{College}_{ijt} \times \text{Earnings}_{ijt}$. Panel A in table 1 reports the results. Column 1 shows that there is a positive relationship between the share of college graduates living in the school-district and college enrollment by high-school graduates attending public schools in the same district. As one can see in column 2, even though ACT scores seem correlated with the share of college graduates and earnings the coefficient

---

7 As a robustness check, I also run a specification that controls for time shocks affecting all school-districts within a city.

8 As it has been widely documented, nowadays females are more likely to enroll in college than males.
of interest $\hat{\beta}_1$ remains large and significant. Moreover, it remains almost unchanged if I control for socioeconomic and location characteristics (column 3), school quality measures (column 4), for time shocks that affect all school-districts within a city (column 5) and city-specific linear trends (column 6), with the latter being the preferred specification. According to the estimate in column 6, an increase in the share of college graduates living in the school-district by 10 percentage points, is associated with an increase of college enrollment at the school-district by 3.66 percentage points. This result is in line with the findings in Chetty and Hendren (2017), who find that moving to an area with higher college attendance rates at a younger age increases a child’s probability of attending college. Note that the coefficient estimate of $\text{College}_{ijt}$ remains very similar to the one reported in column 3 when I introduce the proxies for school resources. It appears then that the role played by highly educated neighbors goes beyond the school resources, in contrast to what is suggested by models of local public funding proposed by developed by Bénabou (1996b), Bénabou (1996a), Fernández and Rogerson (1996) and Durlauf (1996).  

**Heterogeneity by Earnings** The results presented so far show that the relationship between the skills of older neighbors and college enrollment is positive regardless of school-districts’ socioeconomic characteristics. However, this result might mask heterogeneities along some dimensions such as earnings. I investigate the presence of heterogeneities replicating the specifications in panel A, table 1, including now the interaction term, $\text{College}_{ijt} \times \text{Earnings}_{ijt}$ as well as $\text{Earnings}_{ijt}$ by itself.

Column 1, panel B in table 1, reports the OLS estimates of the effect of the share of college graduates on the share of high-school graduates in public schools that enroll in a 4-year college. Both the $\text{College}_{ijt}$ and the interaction coefficients are significant at the 99% confidence level, with $\hat{\beta}_1 < 0$ and $\hat{\beta}_2 > 0$, uncovering the existence of a threshold in the earnings distribution below which a higher share of college of college graduates living in the school-district is associated with a decrease in the college enrollment rate. Figure A.4 displays the average marginal effect of college graduates on college enrollment along the earnings dimension. One can see that as earnings increase, the effect on an increase in the share of college graduates in one percentage point increases. More importantly, it shows that for low values of college graduates’ earnings this effect is negative, while at high values is positive and significant.

Column 2, 3 and 4 in Panel B of table 1, present, respectively, results from specifications controlling for characteristics of the graduating class (share of females and ACT score), socioeconomic conditions of the school-district and school resources, in order to account for neighborhood traits that can be correlated with both college enrollment rate and the share of college graduates and their earnings in the school-district. I find that, in all three specifications, the sign and significance of $\hat{\beta}_1$ and $\hat{\beta}_2$ are barely affected. As before, when I include school quality controls, not only the coefficient estimates of school quality are small and statistically insignificant, but the magnitude of the coefficients of interest - $\hat{\beta}_1$ and $\hat{\beta}_2$ - are very similar to the ones reported in column 3. This evidence, as before, points against school quality as the mechanism linking neighborhoods composition and educational outcomes. Column 5 includes city-year fixed effects, so as to control for shocks affecting all school-district in a given city and

---

9 For the coefficients on school quality measures and average ACT score, see table A.6.
Table 1: College Enrollment and College Graduates

Dependent variable: Share of High-School Graduates that Enroll in a 4-year College

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: No Hetenotegenity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates</td>
<td>0.777***</td>
<td>0.437***</td>
<td>0.387***</td>
<td>0.375***</td>
<td>0.367***</td>
<td>0.366***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.035)</td>
<td>(0.050)</td>
<td>(0.053)</td>
<td>(0.054)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>Observations</td>
<td>1841</td>
<td>1839</td>
<td>1827</td>
<td>1818</td>
<td>1818</td>
<td>1818</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.703</td>
<td>0.786</td>
<td>0.798</td>
<td>0.798</td>
<td>0.803</td>
<td>0.801</td>
</tr>
<tr>
<td><strong>Panel B: Hetenotegenity by Earnings</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates</td>
<td>-5.992***</td>
<td>-5.528***</td>
<td>-4.768***</td>
<td>-4.788***</td>
<td>-4.778***</td>
<td>-4.697***</td>
</tr>
<tr>
<td></td>
<td>(1.454)</td>
<td>(1.272)</td>
<td>(1.125)</td>
<td>(1.113)</td>
<td>(1.115)</td>
<td>(1.102)</td>
</tr>
<tr>
<td>College Graduates × Earnings, College Degree</td>
<td>0.619***</td>
<td>0.552***</td>
<td>0.478***</td>
<td>0.479***</td>
<td>0.478***</td>
<td>0.472***</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.117)</td>
<td>(0.104)</td>
<td>(0.103)</td>
<td>(0.103)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>Observations</td>
<td>1841</td>
<td>1839</td>
<td>1827</td>
<td>1818</td>
<td>1818</td>
<td>1818</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.734</td>
<td>0.795</td>
<td>0.804</td>
<td>0.805</td>
<td>0.810</td>
<td>0.807</td>
</tr>
</tbody>
</table>

The table reports coefficients from an OLS regression with robust standard errors clustered at the school-district level reported in parentheses. Column 1 includes only city and year fixed effects. Column 2 to 6 control for characteristics of the graduating class (the share of females among the high-school graduates and the average ACT score). Column 3 to 6 also include socioeconomic controls, which include the share of black and white residents, the unemployment rate, the median family income, school-district size. Column 5 includes city-year fixed effects and column 6 city fixed effects and a city-specific time trend. The sample includes all school-districts within MSA’s in Michigan over the period 2008 and 2014. ***, ** and * represent statistical significance at 1%, 5% and 10% levels, respectively. Source: CEPI, NCES-EDGE and NCES-CCD.

To year, and to address possible concerns over heterogeneous trends, column 6 includes city-specific linear trends. I find that the sign, the magnitude and significance of coefficients remains nearly unchanged. Note that in all these specifications I control for the median annual earnings of high-school graduates.10

2.4 Robustness Checks

Next, I provide evidence that the results are robust to (i) alternative measures of earnings of college graduates, (ii) different samples and (iii) a different specification.

Earnings of High-skill Neighbors I replicate the estimation of column 6 in table 1 using median annual earnings of individuals with a post-graduate degree and the average between this measure and the median annual earnings of individuals with a college degree so as to capture the earnings of individuals with a college education or higher. Columns 1 and 2 in table A.8 show that using these two alternatives proxies leaves the magnitude, sign and significance of the coefficients of interest relatively unchanged.

School Choice The empirical analysis in the previous section assumes that high-school graduates live in the school-district where they go to school. However, in Michigan there is a school choice

---

10 As I include additional controls, I lose some observations due to missing variables. My results are robust to restricting the sample to school-districts with the full set of controls.
program, established in 1996, under which families can opt to move their children out of the schools they would attend by residency to neighboring districts.\footnote{According to section 105/105C of the Michigan State School Aid Act, all students in Michigan must be allowed to choose to leave their home districts, and when students move districts, the state aid funding travels with them to the destination district. Nevertheless, school-districts are allowed to choose whether to accept students from other districts (http://www.michigan.gov/mde/0,4615,7-140-6530_30334-106922--,00.html).} Between 2008 and 2014, only 19% of the 12th students were non-resident students. Nevertheless, to check the robustness of the results to the inclusion/exclusion of students that do not live in the school-district they attend, I re-estimate equation 2 focusing only on school-districts which have a low share of non-resident students attending the 12th grade (I fix the share threshold at 10%). Column 3 in table A.8 shows that the results previously found hold when we exclude from the analysis school-districts with a higher share of non-resident students attending 12th grade. More important, the coefficients estimates are relatively similar.\footnote{Data on non-resident students per grade and school-district comes from the Michigan Department of Education.}

**Great Recession** The sample used in the empirical analysis covers the period between 2008 and 2014, which includes the period of the Great Recession. According to the National Bureau of Economic Research, the Great Recession began in December 2007 and ended in June 2009. Given this, I re-estimate equation 2 restricting the sample to the years after this period, 2010-2013. I find that the results are robust to the Great Recession (columns 4 and 5 in table A.8): the sign of the estimated coefficients is the same as the one in column 6 in table 1 and their magnitude is relatively unchanged.

**Urban and Rural School-districts** So far, I have focused on school-districts that are located within MSA’s boundaries. Column 6 in table A.8 shows that the sign and significance of coefficients remains unchanged when I also take into account school-districts in rural areas, albeit their magnitude is smaller.

**Quadratic Specification** Equation 2 assumes that the effect of college graduates on college enrollment is linear in earnings. However, if the effect is instead quadratic in earnings, approximating it with a linear specification could be driving the negative sign of $\hat{\beta}_1$. Given this, I estimate a version of equation 2 where I consider the effect of college graduates on enrollment to be quadratic in earnings. Column 6 in table A.8 reports the estimated coefficients, and the right panel in figure A.3 displays the average marginal effect of college graduates on college enrollment along the earnings dimension under this specification. This figure is very similar to the left panel, thus the assumption that the effect of college graduates on enrollment is linear in earnings is suitable.

**2.5 Discussion**

All in all, I find robust evidence regarding the existence of a threshold in the earnings distribution below which a higher share of college of college graduates living in the school-district is associates with a decrease in the college enrollment rate in that district. Importantly, the estimated effect of the neighborhood’s skill composition on the college enrollment rate is net of the effect of school
resources and local amenities. I disentangle the correlation caused by neighbors’ characteristics from neighborhood amenities by taking into account (i) cohort and socioeconomic characteristics of the school-district, such as the racial composition, household income and the average ACT score, among others, and (ii) school quality, measured by expenditures and local revenues per student, in the OLS estimation. The choice of controls is governed by evidence showing that, in comparison to high-school graduates, college graduates prefer neighbors with a higher school quality, average income and different racial compositions (Bayer et al., 2007), which may have also an effect on the decision to enroll in college.

**Omitted Variable Bias**  As previously mentioned, the key assumption for causal interpretation of \( \beta_1 \) and \( \beta_2 \) is that unobserved determinants of college enrollment are not correlated with the share of college graduates, conditional on the controls included. Even though I cannot test this assumption, to assess the possible degree of omitted variable bias I follow the approach of Oster (2016). She developed a method to examine how robust estimates are to omitted variable bias by studying coefficient movements and movements in \( R^2 \) values when including additional controls. Under the assumption that selection on the observables is proportional to selection on the unobservables by a factor \( \delta \), Oster (2016)’s bias-adjusted coefficient is

\[
\beta^*_i = \hat{\beta}_i - \delta(\tilde{\beta}_i - \hat{\beta}_i)R_{\text{max}}\frac{\hat{R} - \tilde{R}}{\hat{R} - \tilde{R}}, i = 1, 2
\]  

where \( \hat{\beta}_i \) and \( \hat{R} \) are the estimated coefficients and \( R^2 \) of column 6 in Panel B of table 1 and \( \tilde{\beta}_i \) and \( \tilde{R} \) come from OLS estimation of equation 2 with no controls (i.e. not including city and year fixed effects, a city-specific trend and the vector \( X \)). \( \delta \) captures the explanatory power of unobserved variables as a proportion of the explanatory power of observed variables and \( R_{\text{max}} \) denotes the \( R^2 \) of a hypothetical OLS regression if one could control for all relevant (observed and unobserved) variables. To identify \( \beta^*_i \), I use \( \delta = 1 \) and \( R_{\text{max}} = 1 \), which yields the identified set for the coefficient estimates \( [\hat{\beta}_i, \beta^*_i] \).

The identified set for \( \beta_1 \) is \([-4.697, -3.424]\) and for \( \beta_2 \) is [0.472, 0.331]. Because both exclude zero, my results can be interpreted to be robust to omitted variable bias under the assumption that selection on the observables is proportional to selection on the unobservables by a factor \( \delta \) as argued by Oster (2016). Figure A.4 plots the average marginal effect with \( \hat{\beta}_i \) and \( \beta^*_i \), and it shows that the heterogeneity in the relationship between college graduates and enrollment along the earnings dimension remains unchanged. In the next section, I propose a novel channel that is able to explain this pattern.

To the best of my knowledge, this result is novel in the extensive empirical literature studying the link between neighborhoods and educational outcomes. Except for Gibbons et al. (2013), who finds heterogeneous effects on students’ test scores, the existing body of work does little to investigate heterogeneity in the effect of neighborhoods’ composition on children’s educational attainment, despite

---

12
this being crucial to understanding the consequences of social mixing. The fact that the relationship between neighborhoods’ skill composition and college enrollment is heterogeneous along the earnings dimension is at odds with models that explain the link between residential location and educational outcomes through human capital spillovers, which suggest this relationship to be positive, regardless of earnings.

3 Model of Education Choice with Local Learning

To explain the heterogeneous relationship between neighborhood skill composition and students’ educational outcomes along the earnings dimension, this section describes a stylized location model with an educational choice. Motivated by empirical evidence showing that individuals lack information about education returns and that neighborhoods play a role as an information source, the model makes two key assumptions. First, when deciding whether to become a high-skill worker or not, children do not know the skill-premium. Second, children learn about it by observing wage realizations of their direct neighbors. The neighborhood’s skill-mix, which is driven by exogenous amenities and dispersion forces (in the form of an inelastic supply of houses in each neighborhood and taste heterogeneity), shapes children’s perception about the skill-premium and, therefore, the education choice.

In line with the empirical evidence, the model shows that in an environment with imperfect information and local learning, there is a wage threshold below which a higher share of high-skill neighbors living in the neighborhood translate into lower investment in education. To clearly illustrate the mechanism, I abstract from other important channels in the literature and make several assumptions that make the model simpler. In section 3.6, I discuss their implication and show that they do not affect the model’s key prediction.

3.1 Environment

Population There are $M$ households living in a city. Each household is composed by a parent and a child. Parents are of two types, high-skill ($H$) and low-skill ($L$), $k \in \{H, L\}$. Each parent provides, inelastically, one unit of labor in the city, for which she is compensated with a wage. The city is closed, hence the population of high-skill and low-skill parents in the city, $M_H$ and $M_L$, respectively, are exogenously given.

City The city is composed by a set of $J$ discrete neighborhoods, indexed by $j \in \{1, ..., J\} \equiv J$. Neighborhoods differ in their attractiveness. This can be due to geographical characteristics (weather, coastal access, etc), but also due to man-made features (school quality, retail environment, distances to places of employment, recreation, noisy streets, etc.). I call amenities to all these features that influence a location attractiveness besides rental prices. As in Busso et al. (2013), each neighborhood

15Gibbons et al. (2013)’s focus and contribution differs, however, from the one of this paper as they study the effect of neighborhood peers on test scores of students and between age 11 and 14, and look to heterogeneity along distinct dimensions than the one of this paper, namely, number of students and population density, among others.

16As in Diamond (2016), I use a two skill group model because the largest group divide in wages across education is seen between college and non-college graduates, as found by Katz and Murphy (1992) and Goldin and Katz (2008).
is characterized a fixed bundle of amenities $A_j$ composed of two different attributes, labeled 1 and 2, $A_j = \{A_{1,j}, A_{2,j}\}$ and school quality $q_j$. Quality attributes of each location are taken by individuals as exogenously given.\footnote{Even though this may strike as a strong assumption, in section 3.6 I argue that introducing endogenous amenities (considering, for instance, that a component of neighborhood’s attractiveness depends on its skill-mix) would not change the main prediction of the model.} All local residents have access to these amenities. Even though the city’s high-skill and low-skill populations are exogenous, the quantity of high and low-skill parents living in a given neighborhood $j$, $M_{H,j}$ and $M_{L,j}$ respectively, are endogenously determined equilibrium outcomes. The city has sufficient capacity that everyone can reside on it, but I consider that each location $j$ is endowed with an inelastic supply of identical houses $H_j$ as in Bayer et al. (2007) and Ferreyra (2009). Houses are are owned by a zero measure of absentee landlords, who rent it to households. Families live in only one house.

**Preferences** All individuals have preferences over an homogeneous consumption good $c$ and amenities. The consumption good is a tradable numeraire good with price normalized to one. For simplicity, I consider that only parents consume. I assume that all individuals have constant absolute risk aversion (CARA) utility over consumption with risk-aversion parameter $\gamma$.\footnote{In section 3.6, I show that the model’s main prediction is qualitatively robust to this assumption.} The utility for an individual $i$ of type $k \in \{H, L\}$ living in neighborhood $j$ is given by

$$U(c^k_{i,j}, \Phi^k_{i,j}) = \frac{-\exp(-\gamma(c^k_{i,j}))}{\Phi^k_{i,j}}$$

where $c^k_{i,j}$ is consumption of individual $i$ of with skill-type $k \in \{H, L\}$ living in neighborhood $j$. $\Phi^k_{i,j}$ maps the attractiveness of neighborhood $j$ to the individual $i$’s utility value for her.

**Wages** Parents pay for consumption and one unit of housing out of their labor income. I consider wages to be exogenous. Let $w^H \equiv \log(\omega^H)$ and $w^L \equiv \log(\omega^L)$, I assume that $w^H_i = w^H + \epsilon^H_i$, with $\epsilon^H_i \sim N(0, \sigma^2_{\epsilon^H})$, and that $w^L_i = w^L + \epsilon^L_i$, with $\epsilon^L_i \sim N(0, \sigma^2_{\epsilon^L})$. $w^H > w^L$. Following empirical evidence showing that wage dispersion is substantially higher among highly educated workers (Lee et al., 2017), for simplicity, I normalize $\sigma^2_{\epsilon^L}$ to 0. Section 3.6 discusses the implications if instead $\sigma^2_{\epsilon^L} > 0$.

**Timing and Decisions** The timing of decisions in the model is the following. Parents draw a wage from the wage distribution corresponding to their skill level and then choose where to locate within the city. Children are born with identical beliefs about the high-skill wage, receive information from high-skill neighbors and update these beliefs. Based on these beliefs, children decide to invest or not in education by comparing the cost of skill acquisition with their perceptions about the skill-premium.

### 3.2 Parents’ Location Choice

At the very beginning of the period, before their children decide whether to invest or not in education, a $k$-type parent draws a wage from the $k$-type wage distribution. Then, parents simultaneously choose a neighborhood $j$ to live in such that they maximize their utility taking as given labor income.
The location choice is affected by two factors. First, an utility shock associated with living in each neighborhood in the city. This can be interpreted as the idiosyncratic utility cost or benefit of living in a given neighborhood. Second, parents compare the attractiveness of living in different neighborhoods. Taking this into consideration, a parent chooses to live in neighborhood \( j \) if either he likes location \( j \) for idiosyncratic reasons or because amenities are much better in \( j \). Note that, I make no assumption regarding altruism, therefore parents are not concerned about their children expected utility or income, and there are no bequests, so parents do not need to solve a dynamic problem. Parents \( i \) with skill level \( k \in \{H, L\} \) solves the following program:

\[
\max_j \quad U(c_{i,j}^k, \Phi_{i,j}^k) = -\exp(-\gamma(c_{i,j}^k)) \frac{\Phi_{i,j}^k(q_j, A_j, \varepsilon_{i,j})}{\Phi_{i,j}^k(q_j, A_j, \varepsilon_{i,j})} \\
\text{subject to} \quad c_{i,j}^k + r_j = w_i^k
\]

where \( w_i^k \) is the wage of parent \( i \) with skill level \( k \) and \( r_j \) is the rent payed to live in neighborhood \( j \). I consider that

\[
\Phi_{i,j}^k = (q_j A_{1,j}^{\beta_k} A_{2,j}^{1-\beta_k}) \varepsilon_{i,j}
\]

where \( \beta_k \) measures parents’ relative taste for each location attribute in the amenities vector and is skill-specific, \( 0 \leq \beta_k \leq 1 \). This aims to capture the idea that different types of individuals tend to prefer different types of amenities as in Glaeser et al. (2016) and Diamond (2016).\(^{19}\)

Individual’s \( i \) idiosyncratic taste for neighborhood \( j \) is denoted by \( \varepsilon_{i,j} \). I model this heterogeneity following McFadden (1973).\(^{20}\) For each parent \( i \), I consider that the idiosyncratic taste for neighborhood \( j \) is drawn from a Fréchet distribution (also called the Type II extreme value distribution):

\[
\Pr(\varepsilon_{i,j} \leq x) = e^{-x^{-\theta}}, \text{ for } x > 0, \text{iid, } \theta > 0
\]

where the parameter \( \theta \) reflects the amount of variation in the distribution and is treated as common across all parents.\(^{21}\) In the location choice context, \( \theta \) governs preference heterogeneity for locations across parents. For simplicity, I assume, from now on, that \( \theta = 1 \).\(^{22}\) The idiosyncratic taste shock implies that when faced with the same rental prices and neighborhood amenities equal parents, with the same skill and wage, may choose to live in different locations.

The indirect utility function of parent \( i \) of type \( k \in \{H, L\} \) living in neighborhood \( j \) can then be

\(^{19}\)Glaeser et al. (2016) assume that the income share of amenities is higher for skilled than unskilled individuals. Diamond (2016) allows for the utility value of the cities’ amenities to differ between high and low skill groups. There is empirical evidence that supports this specification. Bayer et al. (2004) and Bayer et al. (2007) document that individuals with different education levels have a different willingness-to-pay for different location attributes: for instance, when compared to high-school graduates, college graduates are slightly more willing to pay to live in locations that are further away from the workplace and characterized by a higher population density.

\(^{20}\)Following McFadden (1973), a long line of models with location decisions using preference heterogeneity has emerged, such as Bayer et al. (2007), Kennan and Walker (2011), Ferreyra (2009), Busso et al. (2013), Ahlfeldt et al. (2015), Monte et al. (2015), Diamond (2016), among others.

\(^{21}\)The general cumulative distribution function for the Fréchet distribution is \( \Pr(X \leq x) = \exp(-(\frac{x-\mu}{\beta})^{-\theta}) \) if \( x > \mu \), where \( \mu \) is the location parameter and \( \beta \) is the scale parameter. I am implicitly setting \( \beta=1 \) and \( \mu=0 \).

\(^{22}\)The larger is \( \theta \), the smaller is taste dispersion: if \( \theta \) tends to infinite, the variance of idiosyncratic shocks is zero. In that case, only amenities determines neighborhood choice.
represented as
\[ U(w^k_i, r_j, \Phi_{k,j}^j, \varepsilon_{i,j}) = \frac{-\exp(-\gamma(w^k_i - r_j))}{\Phi_{k,j}^j \varepsilon_{i,j}} \]  
(8)

where \( \Phi_{j}^k = (q_j A_{1,j} \beta_k A_{2,j}^{1-\beta_k}) \). Let \( \rho_{i,j}^k \) be the probability that, after observing the vector of \( \varepsilon_{i,j} \) (one for each location), parent \( i \) with skill level \( k \) chooses to live in location \( j \). The distributional assumption on the idiosyncratic taste allows me to derive a close-form expression for \( \rho_{i,j}^k \):
\[ \rho_{i,j}^k = \frac{\Phi_{j}^k \exp(\gamma(w^k_i - r_j))}{\sum_{j' \in J} \Phi_{k}^{j'} \exp(\gamma(w^k_i - r_{j'}))} \]  
(9)

Other things equal, a type-\( k \) parent is more likely to live in a neighborhood the more attractive are \( j \)-specific amenities and the lower are rental prices \( (r_j) \). Since migration is only allowed in the beginning of the period, \( \rho_{i,j}^k \) translate directly into the neighborhood size distribution. The equilibrium number of \( k \)-skill parents in neighborhood \( j \), \( M_{k,j} \), is given by
\[ M_{k,j} = \sum_{i=1}^{M_k} \rho_{i,j}^k = \rho_{j}^k M_k \]  
(10)

where \( M^k \) is the exogenous measure of \( k \)-type parents living in the city.\(^{23}\) Given this, the total population living in neighborhood \( j \) is \( M_j = M_{j}^H + M_{j}^L \). In order for the housing market to clear, the demand for houses in neighborhood \( j \) must equal the supply in that location and so:
\[ H_j = \rho_{j}^H M_H + \rho_{j}^L M_L, \forall j \in J \]  
(11)

The distribution of amenities across neighborhoods determines the skill-mix of each neighborhood and, therefore, whether low-skill households live more or less isolated from the high-skill ones. As shown in the example in Appendix B.1, when amenities are equal across neighborhoods the spatial equilibrium is non-sorted. In this environment, amenities do not react to the characteristics of the population that chooses to live on it. In section 3.6, I discuss the implications of relaxing this assumption.

### 3.3 Children Investment Decision

Children are born to a household of type \( k \), \( k \in \{H, L\} \), living in neighborhood \( j \). Besides family background, children differ in their innate ability \( a \), which is known. The distribution of ability is assumed to be the same across neighborhoods and households types and is given by the distribution function \( G(a) \), with support \( [a, \overline{a}] \). Innate ability determines the human capital of the child together with human capital spillovers, as in Bénabou (1993), Bowles et al. (2014) and Kim and Loury (2013), and school resources (Bénabou, 1996a,b; Durlauf, 1996):
\[ h_i = a^\varepsilon_i \cdot q_j^\kappa \cdot m_{H,j}^\rho \]  
(12)

\(^{23}\)See Appendix B for details.
where \( q_j \) is the level of expenditures per student and \( m_{H,j} \) is the share of high-skill graduates.

Given their level of human capital \( h_i \), all children have to decide whether to invest or not in education. Not investing implies the child to work as a low-skill worker, while investing, implies the payment of the investment cost and working as a high-skill worker. Following the literature, the cost function \( c \) is continuous and strictly decreasing in \( h_i \), and thus in innate ability, human capital spillovers and school resources. The key feature in this model is that at the investment stage, children are uncertain about the return to human capital investment, namely, they do not know the true value of the wage they will receive as a high-skill worker, \( w^H \). Therefore, children make their investment choice based on their perceptions about the skill-premium.

**Information Set** Spatial location matters in the model because it determines the composition of the signals in the children’s information set. Children acquire information about \( w^H \) through social learning. In particular, they learn about it by observing noisy signals of the wage realizations of high-skill parents living in the same neighborhood as them. Each signal from a high-skill neighbor \( s \) living in neighborhood \( j \) is given by

\[
    w^H_{s,j} = w^H + \epsilon^H_{s,j},
\]

where \( \epsilon^H_{s,j} \) denotes the signal noise. Following Fajgelbaum et al. (2016), I assume that the information gathered by each high-skill neighbor in neighborhood \( j \) is proportional to its size,

\[
    \epsilon^H_{s,j} \sim N(0, M_j \sigma^2_{\epsilon^H}),
\]

this means that the largest is the neighborhood, the noisier are the signals. Because of the normality assumption, a sufficient statistic for the information provided by high-skill parents living in neighborhood \( j \) is the public signal

\[
    w^H_j = \frac{1}{M^H_j} \sum_{i=1}^{M^H_j} w^H_{s,j} = w^H + \epsilon^H_j,
\]

with

\[
    \epsilon^H_j \equiv \frac{1}{M^H_j} \sum_{s=1}^{M_{H,j}} \epsilon^H_{s,j} \sim N(0, m_{H,j}^{-1} \cdot \sigma^2_{\epsilon^H}),
\]

where \( m_{H,j} \) is the fraction of high-skill parents living neighborhood \( j \). The signal is neighborhood-specific: all children born in \( j \) observe the same high-skill parents, hence a common public signal, \( w^H_j \). Important for the model’s key prediction, the signal’s precision, \( m_{H,j}^{-1} \cdot \sigma^2_{\epsilon^H} \), increases with the share of high-skill parents in the neighborhood.

**Learning** Initial beliefs are assumed to be identical across all children, \( \tilde{w}^H_i \sim N(\bar{\mu}, \bar{\sigma}^2) \). To update these beliefs, they use information gathered by the public signal \( w^H_j \). Children are passive learners and

---

24 This assumption can be relaxed by allowing the prior to be heterogeneous along the parent’s skill-type, with the prior mean and/or variance of a child born to a low-skill parent, being different than the ones of children in high-skill households. Section 3.6 discusses the implications of relaxing this assumption for the model’s main prediction.
cannot take any action to change the quality of this signal: after receiving information from high-skill parents, each child just updates her prior beliefs using Bayes’ rule. The normality assumption about the prior and the signal implies that the posterior belief about $w_i^H$ is also normally distributed with mean $\hat{\mu}_j$ and variance $\hat{\sigma}_j^2$ given by

$$\hat{\mu}_j = \frac{\sigma_j^2}{\sigma^2 + \sigma_j^2} \bar{\mu} + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \sigma_j^2} w_j^H, \quad (17)$$

$$\hat{\sigma}_j^2 = \left(\hat{\sigma}^{-2} + \sigma_j^{-2}\right)^{-1} \quad (18)$$

where $\sigma_j^2$, the signal’s variance, is equal to $m_{j,H}^{-1} \sigma_{\epsilon,H}^2$. The Bayesian estimator of the high-skill wage is an uncertainty-weighted average of the initial belief and the new information given by the public signal $w_j^H$. Uncertainty about the high-skill wage, defined as the variance of the children beliefs about $w_i^H$, does not depend on the realization of the public signal but on the fraction of high-skill neighbors $m_{j,H}$, the prior’s variance $\hat{\sigma}^2$, and wage dispersion $\hat{\sigma}_{\epsilon,H}^2$. From 17 and 18, I establish the following:

**Lemma 1.** Uncertainty about $w_i^H$, $\hat{\sigma}_j^2$, decreases in the fraction of high-skill neighbors in the neighborhood $m_{j,H}$ but increases with prior uncertainty $\hat{\sigma}^2$ and wage dispersion $\hat{\sigma}_{\epsilon,H}^2$.

**Lemma 2.** When making their estimates about $w_i^H$, children living in neighborhoods with a higher fraction of high-skill neighbors $m_{j,H}$, put relatively more weight on the public signal $w_j^H$.

Note that because children share a common prior and information neighborhood-specific, beliefs about $w_i^H$ are common across children living in the same neighborhood. Nevertheless, they may differ across neighborhoods depending on the allocation of high-skill parents across locations. The fraction of high-skill parents in neighborhood $j$, $m_{j,H}$, plays two roles. On the one hand, it determines uncertainty associated with the returns to educational investment. On the other hand, it determines the weight children put on the public signal: as the fraction of high-skill parents increases, the weight on the prior decreases relative to the weight on the public signal. This implies that those children who have more labor market information, meaning that live in a neighborhood with a higher fraction of high-skill parents, update their beliefs in response to signals to a greater extent than those that have less information: $\Delta \mu_j \equiv \hat{\mu}_j - \bar{\mu} = \frac{\hat{\sigma}_j^2}{\sigma_j^2 \hat{\sigma}^2} (w_j^H - \bar{\mu})$ increases with $m_{j,H}$ (this follows from 17).

**Educational choice**  Given the cost of skill acquisition, $c(a_i, m_{j,H}, q_j)$, and beliefs about $w_i^H$, a child chooses either to invest or not in education. Let $I_j$ be the information set of any child born to a family living in neighborhood $j$, $I_j = \{w_j^H\}$. The optimal policy of a child $i$ born in neighborhood $j$ with innate ability $a_i$ is to invest in education if and only the cost of doing so is lower than the perceived skill-premium, conditional on the information set:

$$V(w^L, \bar{\mu}_j, \hat{\sigma}_j^2, a_i) = \max \{V_j^L(w^L), V_j^H(\bar{\mu}_j, \hat{\sigma}_j^2) - c(a_i, q_j, m_{j,H})\}, \quad (19)$$
where $V_j^H(\hat{\mu}, \hat{\sigma}^2)$ is the perceived value of investing in education for a child born in neighborhood $j$,

$$V_j^H = \sum_{j' \in J} \mathbb{E}_{w_H} [U(c^H_{i,j'}, \Phi_{j'}^H)] \rho_{j'}^H$$  \hspace{1cm} (20)

with $\mathbb{E}_{w_H} [U(c^H_{i,j'}, \Phi_{j'}^H)] \rho_{j'}^H$ being the expected utility of being high-skilled and living in location $j'$, and $V_j^L(w)$ is the expected value of being a low-skill worker for a child living in neighborhood $j$ (because $V_j^L$ is equal across neighborhoods I will drop the subscript $j$ from now on),

$$V_j^L = \sum_{j' \in J} U(c^L_{i,j'}, \Phi_{j'}^L) \rho_{j'}^L$$  \hspace{1cm} (21)

where $\Phi_{j'}^H = q_j \cdot A_{1,j'} \cdot A_{2,j}^{1-\beta_H}$, $\Phi_{j'}^L = q_j \cdot A_{1,j}^{1-\beta_L}$, $\mathbb{E}$ is the expectations operator and the expectation is taken over the high-skill wage. $\rho_{j'}^H$ and $\rho_{j'}^L$ are the probability of living in neighborhood $j'$ conditional on being a high-skill worker and the probability of living in neighborhood $j'$ conditional on being an low-skill worker, respectively. I assume children are myopic in the sense that they not consider that their education decision will determine populations and rental prices, so when computing the expected skill premium, they consider that they will pay the same rent as their parents. Note that in this setting I completely abstract from credit constraints. I do it so not because I do not think they might be important for the decision to invest in education, but so I can start with the simplest model possible that allows me to isolate the the implications of the local information transmission channel in human capital formation. This choice is also supported by empirical evidence found by Carneiro and Heckman (2002), who found that credit constraints do not play a significant role in post-secondary education.

Since the skill acquisition cost is decreasing in ability, the child’s optimal investment decision takes the form of a cut-off rule $a^*_j(w, \hat{\mu}_j, \hat{\sigma}_j^2, m_H^j, q_j)$ such that a child only invests in education if $a_i \geq a^*_j$. This threshold is defined by the following indifference condition

$$V_j^H(\hat{\mu}_j, \hat{\sigma}_j^2) - V_j^L(w) = c(a^*_j).$$  \hspace{1cm} (22)

Given this threshold, for a child $i$ born to a household living in neighborhood $j$, the probability of investing in education is then given by

$$s_{i,j} = 1 - G(a^*_j).$$  \hspace{1cm} (23)

Note that $s_{i,j}$ does not depend on the parents’ type but only on the optimal threshold, $a^*_j$, which is equal across all children living in neighborhood $j$. Hence, the decision to invest in human capital is not linked to the parents’ educational attainment directly, but it is rather linked to the skill-mix of the neighborhood: all children living in same neighborhood, with an ability level higher than $a^*_j$ invest in education, independently of their parents’ type. This result lies on the fact that the driver for the investment decision is the child’s information endowment, which is common across children living within the boundaries of a neighborhood. Given this, the fraction of children investing in education
in neighborhood $j$ is

$$s_j = \frac{\sum_{i=1}^{M_j} s_{i,j}}{M_j} = 1 - \mathcal{G}(a_j^*)$$  

(24)

### 3.4 Equilibrium

Given $M_H$, $M_L$, the distribution of high-skill and low-skill wages, the distribution of ability $a_i$, a vector of school quality $q = \{q_1, ... q_J\}$, and the vector of neighborhood amenities $A = \{A_1, ... A_J\}$, the equilibrium is defined by an allocation of $M_H$ and $M_L$ over $J$ neighborhoods with an associated vector of housing rental prices $r = \{r_1, ... r_J\}$, a vector of cutoff rules $a^* = \{a^*_1, ..., a^*_J\}$, a vector of high-skill wage estimates $\hat{\mu} = \{\hat{\mu}_1, ..., \hat{\mu}_J\}$ and uncertainty $\hat{\sigma}^2 = \{\hat{\sigma}^2_1, ..., \hat{\sigma}^2_J\}$, value functions $V_j(w^L, \hat{\mu}_j, \hat{\sigma}^2_j, a_i)$, $V_j^H(w^L, \hat{\mu}_j, \hat{\sigma}^2_j)$ in each neighborhood $j$, and a vector with the fraction of children investing in education in each location $s = \{s_1, ..., s_J\}$ such that:

1. Parents choose a location $j$ within the city boundaries to maximize utility 4 subject to the budget constraint,

2. For each neighborhood $j$, the value function $V_j(w^L, \hat{\mu}_j, \hat{\sigma}^2_j, a_i)$ solves 19, yielding the cutoff rule in $a_j^*$,

3. Housing market clears in each neighborhood.

Because there are no agglomeration forces (amenities are exogenous), the dispersion forces of the model—inelastic supply of land and taste heterogeneity—ensure the existence of a unique set of rents that clears the housing market, as shown in Bayer et al. (2004). In this environment, the distribution of amenities and school quality across neighborhoods determines the skill-mix of each neighborhood and whether low-skill households live more or less isolated from the high-skill ones. In turn, the spatial allocation of families determines children’s inference about the skill-premium and, therefore, the optimal decision regarding the investment in education.

### 3.5 Comparative Statics

Taking the expectations over the unknown wage, $w^H_i$, the perceived skill-premium for a child born in neighborhood $j$, $\Delta V_j \equiv V_j^H(\hat{\mu}_j, \hat{\sigma}^2_j) - V_j^L(w^L)$, conditional on the information set, is:

$$\Delta V_j = J \left( \frac{-\exp(-\gamma(\hat{\mu}_j - \gamma(\hat{\sigma}^2_j/2))}{\sum_{j' \in J} \exp(\gamma r_{j'}^{l^2})} - \frac{-\exp(-\gamma w^L)}{\sum_{j' \in J} \exp(\gamma r_{j'}^{l^2})} \right).$$  

(25)

Note that in contrast to the literature (Bowles et al., 2014; Kim and Loury, 2013), the benefit of human capital investment is not merely the expected wage gap. Instead, in this framework, the benefit of investing in education takes into account differences in amenities and rental prices paid across different skill groups. This is important because high-skill workers tend to live in places with higher housing costs which may offset some of the consumption benefits from higher wages, but they also tend to enjoy better amenities which may compensate for higher housing costs possibly increasing their well-being. The importance of these differences is highlighted by Diamond (2016), who finds that from 1980 to 2000 changes in cities’ rents and amenities increased welfare inequality between college and high-school graduates by more than the increase suggested by the wage gap alone. For details, see Appendix B.2.
The key variable that drives the optimal investment threshold $a^*_j$ and, as a consequence, the optimal investment decision is beliefs about $w^H_i$. Combining equations 22 and 23, I begin by establishing two intuitive properties of the optimal investment decision:

**Lemma 3.** The ability threshold $a^*_j$ is strictly decreasing in $\hat{\mu}_j$ and strictly increasing in $\hat{\sigma}^2_j$. Hence, the probability of investing in education $s_j$ is strictly increasing in $\hat{\mu}_j$ and strictly decreasing in $\hat{\sigma}^2_j$.

Proof. See Appendix B.3. ■

First, a higher expected value of the high-skill wage $\hat{\mu}_j$ increases the probability that a child will invest in education, holding all else equal. Increasing the expected value of the high-skill wage $\hat{\mu}_j$ increases the perceived skill-premium (equation 25) decreasing $a^*_j$ and, therefore, the fraction of children from neighborhood $j$ that invest in education. Second, greater uncertainty about the high-skill wage ($\hat{\sigma}^2_j$) translates into a lower perception of the skill-premium (equation 25) and, as thus into a lower probability of investing in education, holding all else equal. More uncertainty makes educational investment more risky. Because I consider individuals to be risk-averse, as uncertainty increases, the ability threshold increases and the share of children investing in education decreases. Higher levels of risk aversion amplify this effect.

**High-skill Neighbors ($m_{j,H}$)** Spatial location matters for the decision to invest in education because high-skill neighbors determine the cost of skill acquisition but also because, in this environment, they shape children’s perception about the skill-premium through their estimate of the high-skill wage $\hat{\mu}_j$ and its uncertainty $\hat{\sigma}^2_j$. Regarding uncertainty, the role played by high-skill neighbors is straightforward. A higher share of highly educated neighbors living in a given location $j$ means that children born to that location observe more a precise signal ($\sigma^2_j$ is lower), holding all else equal. As a consequence, they are less uncertain about the high-skill wage. This result follows from the filtering problem (equation 18) and is established in lemma 1. Because children are risk averse, lower uncertainty associated with human capital investment translates into a widening of the mass of children that invest in education ($a^*_j$ decreases, thus $s^*_j$ increases), as established in lemma 3.

The effect of high-skill neighbors on $\hat{\mu}_j$ is, however, ambiguous. As a reaction to a more precise signal, when estimating $w^H_i$ in a Bayesian fashion (equation 17), children place a higher weight on the labor market information disclosed by their neighbors (the public signal $w^H_j$), as formalized in lemma 2. This implies that those children with more information, i.e. those that live in a neighborhood with a larger fraction of high-skill neighbors, update their beliefs in response to new information to a greater extent than those that have less information: $\Delta \mu_j \equiv \hat{\mu}_j - \hat{\mu}$ increases with $m_{j,H}$. However, having more information does not necessarily translate into a higher perception about the skill-premium: $\Delta \mu_j$ may be positive or negative depending on the size of the signal relative to the prior (equation 17). If the signal is sufficiently low, living in a neighborhood with a high fraction of high-skill neighbors translates into a lower $\hat{\mu}_j$ than the one in locations with a low share of high-skill neighbors. On the other hand, if the signal is sufficiently high, children from neighborhoods with a larger share of high-skill neighbors will have a higher $\hat{\mu}_j$. As shown in lemma 3, a lower/higher $\hat{\mu}_j$ translates into a lower/higher fraction of children investing in education.
Figure 1 illustrates how $\hat{\mu}_j$ and $\hat{\sigma}^2_j$ vary across the different spatial allocations in the example given in Appendix B.1. Panel B plots the change in uncertainty, and it shows that its decrease due to the arrival of new information is higher in the neighborhood with a higher share of high-skill parents. Panel C and D plot the change in the estimate, when the public signal $w^H_j$ is sufficiently high and low, respectively. In both panels, the magnitude of the estimate change increases in the share of high-skill neighbors. However, while at high values of $w^H_j$, the estimate is higher in the neighborhood with a higher human capital level, the opposite is true when the signal’s magnitude is small. In this case, being exposed to high-skill neighbors translates into a lower perception of the skill-premium.

The total effect of high-skill neighbors on the share of children investing in education depends then on the size of the signal. If the signal is sufficiently high such that $\Delta \mu_j > 0$, a higher of high-skill neighbors increases the perceived skill-premium increases (the estimate of $w^H_j$ is higher and its uncertainty is lower). Thus, the share of children that decide to invest in education increases. In contrast, if the signal is sufficiently low and $\Delta \mu_j < 0$, there are two opposing forces on the perceived skill-premium. High-skill neighbors decrease uncertainty and the cost of skill-acquisition, but they also decrease children’s estimate about the high-skill wage. Whether the share of children investing in education increases or decreases depends on which effect dominates. This, in turn, depends on the size of the signal relative to a threshold $w^*$.

Overall, more information about $w^H_j$ (living in a location with a high $m_{j,H}$) increases the share of children investing in education if and only if $w^H_j > w^*$. Under this condition, perceived skill-premium is increasing in the share of high-skill neighbors. Otherwise, if $w^H_j < w^*$, the expected value of $w^H_j$ decreases in the fraction of highly educated parents in the neighborhood, and this effect dominates the fact that uncertainty is lower, reducing the probability of investing in education even though the exposure to high-skill neighbors is higher. Proposition 1 formalizes this result.

**Proposition 1.** Given a spatial allocation of $M_H$ and $M_L$ over $J$ neighborhoods in the city, locations with a higher fraction of high-skill parents, $m_{j,H}$, have a higher fraction of children investing in education $s_j$ if and only if $w_j > w^*$.

**Proof.** See Appendix B.3. ■

To sum up, the skill-mix of neighborhoods and the education decision of children are connected through an information channel. The configuration of the city, namely, the distribution of high-skill parents across neighborhoods shapes the public signal $w^H_j$ children observe. Local information diffusion creates inequalities between neighborhoods as their skill-mix generates different perceptions about the skill-premium. But, more importantly, under information frictions and social learning, the effect of local interactions in the education decision is not only about being more exposed to high-skill neighbors, as suggested by previous literature, it is also about the labor market information they disclose. More exposure implies more information, but more information does not necessarily increase the probability of investing in education, this will depend on the information that children observe, namely, the magnitude of the public signal.
Figure 1: Beliefs across different spatial allocations - An Illustration

Panel A: Equilibrium share of high-skill households in neighborhoods 1 (solid line) and 2 (dotted line) for different levels $A_{1,1}$ in neighborhood 1. Panel B: Relative change in uncertainty ($\hat{\sigma}^2_{1}/\tilde{\sigma}^2 - 1$) in neighborhoods 1 (solid line) and 2 (dotted line) across different spatial equilibriums. Panel C: Relative change in the perceived expected value of $w_{H_i}$ ($\Delta \mu_{j} \equiv \hat{\mu}_{j}/\tilde{\mu} - 1$) in neighborhood 1 (solid line) and neighborhood 2 (dotted line) if $w_{H} > w^*$. Panel D: Change in the perceived expected value of $w_{H_i}$ ($\Delta \mu_{j} \equiv \hat{\mu}_{j}/\tilde{\mu} - 1$) in neighborhood 1 (solid line) and neighborhood 2 (dotted line) if $w_{j} < w^*$. Each $A_{1,1}$ value corresponds to a different spatial allocation as shown in figure B.1. $H_1 = H_2 = 75, M^H = 100, M^L = 50, \beta_H = 1, \beta_L = 0, A_{1,2} = A_{1,2} = A_{2,2}, \tilde{\mu} = 9, \hat{\sigma}^2 = 0.06, w^U = 7.6, \sigma^2_{\epsilon H} = 0.03$.

**Wage dispersion ($\sigma^2_{\epsilon H}$)** Information transmission from high-skill neighbors as a channel through which children learn about the high-skill wage $w^H_i$ depends on its dispersion $\sigma^2_{\epsilon H}$. For a given spatial equilibrium, the higher is $\sigma^2_{\epsilon H}$, the lower is the change in the estimate of $w^H_i$ and its uncertainty upon arrival of new information. Therefore, the lower is the potential to learn from high-skill neighbors. This follows from the fact that as $\sigma^2_{\epsilon H}$ increases, the public signal becomes less precise as shown in 16. The overall effect of wage dispersion on the share of children investing in education depends also on the size of the signal. If the signal is sufficiently high, such that $\Delta \mu_{j} > 0$, the share of children that decide to invest in education is decreasing in wage dispersion because the perceived skill premium decreases (the estimate of $w^H_i$ is lower and its uncertainty is higher). In contrast, if the signal is sufficiently low, such that $\Delta \mu_{j} < 0$, there are two opposing effects. Wage dispersion increases uncertainty, but it also increases the estimate. So, the total effect depends on which effect dominates.

**School quality ($q_j$)** The quality of the school at each location plays a standard role: holding all else equal, higher values of school quality translate into a lower cost of investing in human capital is lower, hence the probability of investing in education increases.

**Low-skill wage ($w^L$)** The wage of low-skill workers also plays a standard role: for lower values of the low-skill wage, holding all else equal, the perceived skill-premium is higher, hence the probability of investing in education increases.

### 3.6 Discussion of the Model’s Assumptions

I make several assumptions that make my model simpler without affecting its main qualitative result. In this section, I discuss the implications of each assumption for the model’s results.
Exogenous amenities  I consider that neighborhood’s amenities are taken to be exogenous. However, places that attract a higher share of skilled workers may endogenously become more desirable places to live in (see, for instance Diamond, 2016). In line with this, one could consider that neighborhood amenities have two distinct parts: (i) an exogenous component that is invariant to the skill-mix of the neighborhood such as the geographic characteristics, and (ii) an endogenous component that depends on the share of high-skill workers in the neighborhood. One could re-define $\Phi_{k,j}$ in 8 as $\Phi_{k,j} = q_j A_{1,j} \beta_k A_{2,j}^{1-\beta_k}$, with $A_{1,j} = m_{j,H}$ begin a location attribute that I allow to endogenously respond to the types of families living in the neighborhood, namely the share of high-skill parents. Allowing for endogenous amenities affects the spatial allocation of households across neighborhoods within the city without affecting the role of high-skill neighbors in the decision to invest in education described in proposition 1. Note that the introduction of these agglomeration forces generates the potential for multiple equilibria in the model if these agglomeration forces are sufficiently strong relative to the exogenous differences in characteristics across locations. However, within each equilibrium the main prediction of the model holds.

Uncertainty about Low-Skill Wage  If $\sigma^2_{e,L} > 0$, children will not only uncertain about the high-skill wage but also about the low-skill wage. In this case, a higher fraction of high-skill neighbors yields more information about the high-skill wage, but less information about the low-skill wage. This will amplify differences in the perceived skill-premium across neighborhoods and, therefore, in the share of children investing in education. Appendix B.4 shows that under this scenario there is also a signal threshold $w^*$ below which a higher fraction of high-skill neighbors decreases the fraction of children investing in education. However, in this setting, the magnitude of this threshold also depends on the magnitude of the signal children receive about the low-skill wage.

Common prior  In section 3.3, I assume that children share a common prior about $w^H_i$ and update this prior using information at the neighborhood level. This implies that the probability of investing in education (23) is independent of the parent’s type. This assumption can be relaxed by allowing the prior to be heterogeneous along the parent’s type, with the prior mean and/or variance of a child born to a low-skill parent, being different than the ones of children in high-skill households. If children priors depend on their parent’s type, for a neighborhood $j$, there will two ability thresholds that determine the probability of investing in education for children born to high and low-skill families, $s^H_j$ and $s^L_j$. The thresholds $a^*_{H,j}$ and $a^*_{L,j}$ are defined by the indifference condition, $V^H_{k,j}(\hat{\mu}_{k,j}, \hat{\sigma}^2_{k,j}) - V^L(w_u) = c(a^*_{k,j})$, where $V^H_{k,j}(\hat{\mu}_{k,j}, \hat{\sigma}^2_{k,j})$ is the perceived value of being a high-skill worker for a child born to a $k$-type household living in neighborhood $j$. Given this, the probability of investing in education for a child $i$ born to a household of type $k$ living in neighborhood $j$ is $s_{i,j}^k = 1 - G(a^*_{k,j})$, and the fraction of children investing in education in neighborhood $j$ is

$$s_j = \frac{\sum_{i=1}^{M_j^H} s_{i,j}^H + \sum_{i=1}^{M_j^L} s_{i,j}^L}{M_j}$$ (26)
s_j increases with the fraction of high-skill neighbors m_{j,H} if \( \frac{\partial s^H_j}{\partial m_{j,H}} + \frac{\partial s^L_j}{\partial m_{j,L}} > 0 \). Whether \( \frac{\partial s^H_j}{\partial m_{j,H}} \) and \( \frac{\partial s^L_j}{\partial m_{j,L}} \) are greater or lower than zero will depend, respectively, on the magnitude of the signal \( w^H_j \) relative to the threshold \( w_j^{*H} \) and \( w_j^{*L} \), as stated in proposition 1. I choose to assume a common prior across all children because in the data I use in section 5 to validate the model’s prediction regarding the role of high-skill neighbors, I observe the share of high-school graduates that enroll in college (which corresponds to the fraction of children investing in education in the model) at the neighborhood level \( s_j \), but not by parent’s education in each neighborhood, \( s_j^k \).

Correlated ability

**Risk-aversion**  Assuming individuals have CARA utility function over consumption with risk-aversion parameter \( \gamma \) is not crucial for the model’s prediction regarding the role of high-skill neighbors described in proposition 1. Appendix B.4 shows that if instead individuals are risk neutral with a linear utility function in consumption and amenities, there is also a threshold \( w^* \) below which the relationship between the share of high-skill neighbors and children investing in education is negative – albeit higher than the one in proposition 1 due to the fact that now individuals do not dislike uncertainty. Hence, under risk neutrality, the magnitude of the signal does have to be higher in order to trigger a positive relationship between the probability of investing in education and the share of high-skill neighbors. This is due to the fact that under risk neutrality \( a_j^* \) depends only on the posterior mean \( \hat{\mu}_j \) but not on the posterior variance \( \hat{\sigma}_j^2 \).

Overall, I show that, if one takes into account information frictions and social learning, the effect of high-skill neighbors on the fraction of children investing in education may be negative, consistent with the empirical evidence presented in section 2. In contrast with the existing literature, in this model, more exposure to high-skill neighbors brings more information, but additional does not necessarily translate into more investment in education. This will depend on the labor market information disclosed by highly educated neighbors. In the next section, I estimate the model and assess the quantitative importance of information frictions and social learning as channel through which neighborhoods affect the decision to enroll in college.

4 The Importance of Local Learning

Even though the local learning mechanism is consistent with the empirical evidence presented in section 2, it remains an open question whether this channel is quantitatively important. To tackle this issue, I calibrate the model to match 2013 data regarding the wage distribution by educational attainment, the distribution of individuals and college enrollment rates across school-districts in the city of Detroit. I choose Detroit because it is the largest city in Michigan, with 95 school-districts in 2013.

Armed with the calibrated economy, I ask the following question: by how much would college enrollment change if children did not observe any information from high-skill neighbors. It should
be noted that a more realistic analysis would nest the learning mechanism within a richer framework including other mechanisms proposed in the literature as the local funding of education. As this is the first paper to explore the contribution of the local information constraints to the accumulation of human capital and, therefore, in persistent inequality, assessing its quantitative potential in a simple model that allows both the theory and the calibration to be fairly transparent is an important first step to subsequently developing more complicated quantitative models.

4.1 Definition of Variables in the Model

City  A city corresponds to a metropolitan statistical area (MSA) that is a region consisting of a group of counties that have a high degree of economic and social integration with the core county as measured through commuting.

Neighborhoods  I define a neighborhood in the model to be a school-district. The most commonly definition of a neighborhood is a census tract, a “small, relatively permanent statistical subdivisions of a county”, which have generally a size between 1200 and 8000 people. School-districts tend to be relatively larger. I pick school-districts over census tracts, because school-districts are the smallest unit of analysis for which I observe both college enrollment by high-school graduates and socioeconomic characteristics of the location such as the % of college graduates, median family income, among others. I use the Geographic Correspondence Engine with Census 2010 from the Missouri Census Data Center to link school-districts to MSA’s.

High and Low-skill Neighbors  I use education to proxy for skills as in Acemoglu and Autor (2011) and Diamond (2016), and define “high-skill” neighbors as those individuals living in the school-district who have at least a 4-year bachelor’s degree while “low-skill” neighbors are those who have less years of education than that.

4.2 Functional Forms

The parameterization of the model is as follows: The utility function is CARA with risk aversion parameter \( \gamma \). Innate ability is assumed to be uniformly distributed between \( \overline{a} \) and \( \underline{a} \). The cost functions is given by \( C(a_i) = \overline{c} - \phi(a_i^{\overline{c}} \cdot q_j^{\overline{c}} \cdot m_{i,H}^{\overline{c}}) \), where \( a_i \) is innate ability, \( q_i \) is expenditures per student in school-district \( j \) and \( m_{i,H} \) corresponds to the share of high-skill neighbors living in the school-district.

4.3 Calibration Strategy

Calibration is proceed in two steps. In the first step, I set parameters that either have a direct counterpart in the data or that have been used in previous literature. In the second step, I use the simulated method of moments to estimate the remaining parameters.

---


29 The linking file can be download here http://mcdc.missouri.edu/websas/geocorr14.html.
First Step  I set \( \tau \) equal to one and \( q \) to zero, also I assume that high-skill individuals only care about \( A_{1,j} \) and low-skill individuals only care about \( A_{2,j} \): \( \beta_H = 1 \) and \( \beta_L = 0 \). The number of neighborhoods \( J \) equals the number of school-districts in Detroit in 2013. As in Babcock et al. (1993), I set the risk aversion parameter \( \gamma \) to 0.5.

The wage distributions in the model match the empirical distributions of labor income of full-time workers with different skills from the American Community Survey 2008-2013. I define full-time workers are defined to be individuals aged between 25 and 55 years working at least 35 hours per week, 48 weeks per year. For the low-skill wage distribution, I normalize the variance to 0 and calibrate the mean \( w^L \) to match the mean of the log monthly-wage distribution of low-skill full time workers. For the mean of the high-skill wage distribution \( w^H \), I match the mean of the distribution of the log monthly-wage distribution of high-skill full-time workers. Because I normalize the variance of the low-skill wage to 0, I set the variance of the high-skill wage distribution equal to the difference between the variance of the labor income distribution of high-skill full-time workers and the variance of the labor income distribution of low-skill full-time workers. I set the mean and variance prior (\( \tilde{\mu} \) and \( \tilde{\sigma}^2 \)) such that the difference between the true and the perceived ratio of high to low-skill wage matches the one in Bleemer and Zafat (2015).

Regarding amenities, I recover the distribution of \( A_{1,j} \) and \( A_{2,j} \) across the school-districts from the data. From NCES-EDGE, I observe for each school-district: total population \( M_j \), the number of high-skill and low-skill individuals, \( M^H_j \) and \( M^L_j \), and rents \( r_j \). Following Diamond (2016), as a measure of rents, I use the median gross rent at each school-district, which includes both the housing rent and the cost of utilities.\(^{30}\) Assuming that the current allocation of individuals across school-districts is in equilibrium, for any two neighborhoods \( j \) and \( j' \), the following holds

\[
\frac{M^k_j}{M^k_j'} = \frac{\Phi^k_j}{\exp(\gamma r_j')} \frac{\Phi^k_{j'}}{\exp(\gamma r_j)}
\]

where \( M^k_j \) is the number of type \( k \)-individuals that live in \( j \). For high-skilled individuals, \( \Phi^H_j = q_j A_{1,j} \), for low-skill individuals, \( \Phi^L_j = q_j A_{2,j} \). I set both amenities’ attribute equal to one for Detroit City school-district, and then back out the level of \( A_{1,j} \) and \( A_{2,j} \) for the other school-districts using 27.

Second Step  The parameters without observable counterparts are the cost function parameters, \( \tau, \varphi \) and \( \phi \). I estimate them using the simulated of method of moments, which picks the parameter vector \( \theta=(\tau, \varphi, \phi, \kappa, \rho) \) that minimizes the weighted sum of square deviations between data moments and their model-generated counterpart:

\[
\hat{\theta} = \text{arg min} \ (y(\theta) - y^*)'W(y(\theta) - y^*)
\]

\(^{30}\)Ideally, I would like to have school-district specific rent indices controlling for differences in the quality of housing across school-districts following the hedonic-regression approach by Eeckhout et al. (2014). However, because I cannot link individuals in the ACS to the school-districts where they live, this is not possible, thus I use the reported median gross rent in NCES-EDGE.
where \( W \) is the identity matrix, implying that each moment is equally weighted, \( y^* \) is a \( t \times 1 \) vector of moments observed in the data and \( y(\theta)^* \) is a \( t \times 1 \) of those moments from the model evaluated at a given parameter vector \( \theta \). In the estimation, I match 2013 data moments of the distribution college enrollment (mean, standard deviation and p75-p50 ratio), the correlation between college enrollment and college graduates and the correlation between college enrollment and expenditures per student. Table A.9 summarizes all parameters. An advantage of estimating the model is the understanding of what features of the data identify each parameter. The mean of college enrollment across districts identifies \( \bar{c} \). The school-district variation in enrollment identifies \( \phi \), while p75-p50 ratio identifies \( \phi \). Finally, the correlation of enrollment with the share of college graduates and expenditures per student identify \( \rho \) and \( \kappa \), respectively.

4.4 Model Fit

This section discusses the calibrated economy calibrated to Detroit in 2013. Table 2 compares the empirical targets for the calibrated parameters and the corresponding moments produced by the model. The left panel in figure 2 depicts the predicted and observed values for the school-district college enrollment rate for the 95 school-districts within Detroit, and the right panel plots the enrollment distribution in the model and observed in the data. The calibrated model reproduces reasonably well the five targeted moments, and the distribution of enrollment is highly correlated with the one observed in the data. Overall, I view these results as indicative that the model successfully captures patterns in the data.

In the calibrated economy, the average high-skill family lives in a school-district where 24.4% of its population is high-skill, while the average low-skill family lives in a location where the proportion of high-skill is 8 percentage points lower. Differences in the skill-composition of locations translate into differences in the subsequence education decisions of children. The probability of becoming a high-skill worker for a child born to a low-skill household is 8 percentage points lower than the probability of becoming a high-skill worker for a child from a high-skill family. An important question is what is driving this gap: differences in perceptions, school resources or human capital spillovers?

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>p75/p50</th>
<th>Corr. w/ ( m_{j,H} )</th>
<th>Corr. w/ ( q_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>0.38</td>
<td>0.14</td>
<td>1.28</td>
<td>0.84</td>
<td>0.12</td>
</tr>
<tr>
<td>Model</td>
<td>0.38</td>
<td>0.11</td>
<td>1.24</td>
<td>0.96</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The table reports targeted moments in the estimation. Corr w/ \( m_{j,H} \) corresponds to the correlation between the share of college graduates and college enrollment. Corr w/ \( q_j \) corresponds to the correlation between expenditures per student and college enrollment. Observations are at school-district level. The sample is composed by 95 school-districts within Detroit in the year 2013.
4.5 Quantifying the Mechanism

Armed with the calibrated economy, I ask the following question: by how much would college enrollment rate change in the absence of the public signal from high-skill neighbors? To answer this question, I simulate what would happen if individuals did not update their initial beliefs ($\hat{\mu}_j = \tilde{\mu}_j$ and $\hat{\sigma}^2_j = \tilde{\sigma}^2_j$). The left and the right panel in figure 3 plot, respectively, the perceived skill-premium and college enrollment in the baseline model (with the learning mechanism, and thus matching the data) versus the no-learning counterfactual across school-districts.

I find that high-skill neighbors play an important role in correcting initial beliefs. Before observing any information, children hold beliefs about the high-skill wage that are downward biased ($\tilde{\mu}_j < w^H_j$) and more uncertain ($\tilde{\sigma}^2_j > \sigma^2_j$). By observing the noisy public signal $w^H_j$, children’s estimate of the high-skill wage increases 3% and its uncertainty decreases 28%, on average. As a consequence, the perceived skill-premium rises 6.7%, on average (right panel in figure 3). This as a significant effect on enrollment as shown in the right panel of figure 3. In particular, I find that if individuals did not observe any public signal from high-skill neighbors, the college enrollment rate across school-districts at the would be 22 percentage points lower, on average. At the city level, this means that instead of having 42% of high-school graduates enrolling in college within 6 months of graduation, only 18% would.

**Skill persistence** In the calibrated economy, a high-skill family lives in a neighborhood with a share of college graduates that is 8 percentage points higher, on average, than the average neighborhood where low-skill families live. This difference translates into differences in the average perceived skill-premium, which in turn translate into different probabilities of investing in education. Namely, a child...
Figure 3: Full Model vs. No-Learning Counterfactual

A. Perceived Skill-Premium

B. Enrollment

The left panel plots the perceived skill-premium across school-districts in the full model (blue) versus the no-learning counterfactual (red). The right panel plots college enrollment rate across school-districts in the full model (back) versus the no-learning counterfactual. The sample is composed by the 95 school-districts within Detroit in the year 2013.

that is born to a high-skill family has a probability of becoming a high-skill worker that is 22% higher when compared to a child that is born to a low-skill family. By shutting down local learning, I find that differences in perceptions are responsible for 60% of this difference.

4.6 Decomposition: which channel matters the most?

In the model, dispersion across neighbors arises through three different mechanisms: (i) the information externalities: school-districts that initially a higher share of college graduates generate more informative signals that cause college enrollment rate to be higher; (ii) human capital spillovers and (iii) expenditures per student. Given this, a natural extension of the main counterfactual exercise is to ask which channel matters the most for college enrollment. One way to conduct this decomposition by shutting down each channel at the time. The left panel in figure 4 plots college enrollment across school-districts in the benchmark economy (which coincides with the data) as well as in three different scenarios: one with no local learning, another without human capital spillovers ($\varphi=0$), and finally a counterfactual in which the human capital does not depend on school resources ($\rho=0$). Table 3 reports college enrollment rate at the city level, dispersion of enrollment across school-districts and differences in the probability of becoming under all mentioned scenarios. While local learning and school resources have a similar effect on the enrollment rate, local learning, is by far, the most important channel in explaining inequality across school-districts. In particular, it accounts for 63% of the dispersion in college enrollment across school-districts and explains 60% of the difference between the probability of being high-skill for a child born to a high-skill family and a child born to a low-skill family.
What if there were no information frictions? Local learning plays an important role in the persistence of inequality across generations. An important question to ask is then how much would college enrollment change if individuals had the right information about the high-skill wage distribution? To answer this question, I simulate the model assuming that $\mu_j = w^H$ and $\sigma_j = \sigma^H$. The right panel in figure 4 plots college enrollment across school-districts under this scenario and the benchmark economy, and shows that correcting beliefs college enrollment increases substantially in all school-districts: 83% of high school graduates would enroll in college, which compares to 42% in the benchmark economy, as shown in table 3. More important, my results show that by simple correcting beliefs (i.e. leaving the other sources of inequalities across neighborhoods at work) one can reduce significantly inequalities across children from different backgrounds: in particular, this difference reduces in 90%. This finding points in favor of the implementation of disclosure policies aimed at giving information to children.

Figure 4: Benchmark Economy vs. Counterfactuals

A. Counterfactual Decomposition

B. Benchmark and No Info. Frictions

The left panel plots college enrollment across school-districts under the benchmark economy (blue) and three different scenarios: no local learning (red), no human capital spillovers ($\rho = 0$, yellow) and no school resources ($\kappa = 0$, green). The right panel plots college enrollment across school-districts in a scenario where individuals know the right distribution of high-skill wage ($\hat{\mu}_j = w^H$ and $\hat{\sigma}_j^2 = \sigma^2$, pink). The sample is composed by the 95 school-districts within Detroit in the year 2013.
Table 3: Benchmark Economy vs. Counterfactuals

<table>
<thead>
<tr>
<th></th>
<th>Benchmark</th>
<th>No-learning</th>
<th>$\varphi = 0$</th>
<th>$\kappa = 0$</th>
<th>No Frictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrollment rate</td>
<td>0.42</td>
<td>0.18</td>
<td>0.50</td>
<td>0.19</td>
<td>0.84</td>
</tr>
<tr>
<td>Std. Dev. Enrollment</td>
<td>0.11</td>
<td>0.04</td>
<td>0.07</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Probability of being High-skill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child from high-skill family</td>
<td>0.46</td>
<td>0.19</td>
<td>0.52</td>
<td>0.23</td>
<td>0.84</td>
</tr>
<tr>
<td>Child from low-skill family</td>
<td>0.38</td>
<td>0.16</td>
<td>0.47</td>
<td>0.14</td>
<td>0.83</td>
</tr>
<tr>
<td>Difference</td>
<td>0.08</td>
<td>0.03</td>
<td>0.05</td>
<td>0.09</td>
<td>0.01</td>
</tr>
</tbody>
</table>

The table reports several statistics under the benchmark economy and four different scenarios: no local learning, no human capital spillovers in the cost function ($\rho = 0$) and no school resources in the cost function ($\kappa = 0$), and no information frictions ($\hat{\mu}_j = w_H$ and $\hat{\sigma}^2_j = \sigma_{\epsilon H}$). Observations are at school-district level. The sample is composed by 95 school-districts within Detroit in the year 2013.

5 Conclusion

It is widely known that the neighborhood context is critical for children’s educational decisions and, therefore, for intergenerational mobility. Despite the large literature on potential causal factors, none has considered the fact that individuals do not have perfect knowledge about education returns and that neighborhoods may be a potential source of information. This paper fills this gap in the literature.

In the model, children compare the cost of acquiring skills with the expected return from investing in education. This return in unknown and children’s beliefs about it are shaped by the parents’ location decision within the city: children observe labor market earnings of high-skill neighbors and update their beliefs in a Bayesian fashion. As a theoretical contribution, I show that, in contrast to the previous literature, under imperfect information and social learning, locations with a higher level of human capital have more information, but are do not necessarily have a higher proportion of children investing in education. This is only the case if labor market earnings of high-skill neighbors is sufficiently high so that it translates into a higher perception about the skill-premium.

Using school-district data from Michigan over the period 2008 to 2014, I provide evidence supporting the model’s key prediction. In particular, I find that the relationship between the share of college graduates and the share of high-school graduates enrolling in college in a given school-district is positive only if college graduates’ earnings are sufficiently high. I also calibrate the model to capture features of college enrollment across school districts within Detroit. I show that this simple framework can replicate the data and that the proposed local information transmission mechanism is quantitatively important.

All in all, my results highlight the importance of taking into account imperfect information and the local transmission of information to understand the importance of local externalities in human capital formation and, therefore, for persistent inequality between social groups. More important, these results support the implementation of information campaigns as a policy instrument to address opportunity inequality.
References


**Table A.4: Summary statistics**

The table reports summary statistics for the main variables used in the empirical analysis. Observations are at the school-district level and cover the period from 2008 to 2014. *Enrollment in a 4-year College* measures the share of high-school graduates in public schools that enroll in a 4-year college within 6 months after graduation. *College graduates* is the share of population over 25 years old with 4 or more years of college. *Black and white residents* are measured as the share of total population in the school-district that are black and white, and the *unemployment rate* is the share of unemployed individuals. *ACT score* is the score of the American College Testing averaged over all high-school graduates in public schools. *Females* measures the share of high school graduates in public schools that are females. *Earnings by Educational Attainment* correspond to median annual earnings per education level at the school-district level and are expressed in 2010 dollars. Expenditures and revenues per pupil are also expressed in in 2010 dollars. Source: CEPI, NCES-EDGE and NCES-CCD.

<table>
<thead>
<tr>
<th>Observations</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>College Enrollment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrollment in a 4-year College</td>
<td>1847</td>
<td>0.33</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Earnings by Educational Attainment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School Degree</td>
<td>1851</td>
<td>26462.86</td>
<td>4039.23</td>
<td>12365.45</td>
</tr>
<tr>
<td>College Degree</td>
<td>1851</td>
<td>46730.47</td>
<td>9205.04</td>
<td>11230.26</td>
</tr>
<tr>
<td>Post-Graduate Degree</td>
<td>1848</td>
<td>60924.20</td>
<td>12024.40</td>
<td>15378.39</td>
</tr>
<tr>
<td><strong>Socioeconomic Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates</td>
<td>1851</td>
<td>0.23</td>
<td>0.13</td>
<td>0.04</td>
</tr>
<tr>
<td>Median Family Income</td>
<td>1851</td>
<td>62700.72</td>
<td>17822.69</td>
<td>19409.57</td>
</tr>
<tr>
<td>Black Residents</td>
<td>1839</td>
<td>0.86</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>White Residents</td>
<td>1839</td>
<td>0.08</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>Unemployment Rate</td>
<td>1851</td>
<td>0.11</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Total Population</td>
<td>1839</td>
<td>28834.78</td>
<td>53748.40</td>
<td>2145.00</td>
</tr>
<tr>
<td><strong>Cohort Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT Score</td>
<td>1851</td>
<td>19.10</td>
<td>2.01</td>
<td>12.23</td>
</tr>
<tr>
<td>Females</td>
<td>1845</td>
<td>0.51</td>
<td>0.05</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>School Quality Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure per student</td>
<td>1840</td>
<td>10820.13</td>
<td>2312.69</td>
<td>7624.06</td>
</tr>
<tr>
<td>Local revenue per student</td>
<td>1840</td>
<td>3432.45</td>
<td>2034.61</td>
<td>761.49</td>
</tr>
<tr>
<td>Teachers to student ratio</td>
<td>1842</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Table A.5: Correlations between Main Variables

The table reports the correlation pattern between the main variables used in the empirical analysis. The correlations are computed using the 1851 district-year observations in the sample over the period from 2008 to 2014. Source: CEPI, NCES-EDGE and NCES-CCD.

<table>
<thead>
<tr>
<th>Enroll in a 4-year College</th>
<th>College Graduates</th>
<th>Median Earnings, College Grad.</th>
<th>Median Family Income</th>
<th>ACT Score</th>
<th>Expenditure per student</th>
<th>Local Revenue per student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrollment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Earnings, College Grad</td>
<td>0.427</td>
<td>0.445</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Family Income</td>
<td>0.701</td>
<td>0.830</td>
<td>0.683</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT Score</td>
<td>0.749</td>
<td>0.714</td>
<td>0.509</td>
<td>0.789</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Expenditure per student</td>
<td>0.047</td>
<td>0.181</td>
<td>-0.028</td>
<td>0.061</td>
<td>-0.127</td>
<td>1</td>
</tr>
<tr>
<td>Local Revenue per student</td>
<td>0.230</td>
<td>0.424</td>
<td>0.148</td>
<td>0.307</td>
<td>0.182</td>
<td>0.587</td>
</tr>
</tbody>
</table>
The table reports coefficients from an OLS regression with robust standard errors clustered at the school-district level reported in parentheses. The dependent variable is the share of high-school graduates that enroll in a 4-year college within 6 months of graduation, with mean equal to 0.33. Column 2 to 6 control for characteristics of the graduating class (the share of females among the high-school graduates and the average ACT score). Socioeconomic controls include the share of black and white residents, the unemployment rate, the median family income, school-district size. The sample includes all school-districts within MSA’s in Michigan over the period 2008 and 2014. ***, ** and * represent statistical significance at 1%, 5% and 10% levels, respectively. Source: CEPI, NCES-EDGE and NCES-CCD.

<table>
<thead>
<tr>
<th>Dependent variable: Share of High-School Graduates that Enroll in a 4-year College</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>College Graduates</td>
<td>0.777***</td>
<td>0.437***</td>
<td>0.387***</td>
<td>0.375***</td>
<td>0.367***</td>
<td>0.366***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.035)</td>
<td>(0.050)</td>
<td>(0.053)</td>
<td>(0.054)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>ACT Score</td>
<td>0.0308***</td>
<td>0.0379***</td>
<td>0.0381***</td>
<td>0.0384***</td>
<td>0.0385***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Expenditure per student</td>
<td>0.00405</td>
<td>0.00335</td>
<td>0.00408</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Revenue per student</td>
<td>0.00237</td>
<td>0.00351</td>
<td>0.00255</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers to student ratio</td>
<td>-0.000631</td>
<td>-0.000242</td>
<td>-0.000327</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1841</td>
<td>1839</td>
<td>1827</td>
<td>1818</td>
<td>1818</td>
<td>1818</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.703</td>
<td>0.786</td>
<td>0.798</td>
<td>0.798</td>
<td>0.803</td>
<td>0.801</td>
</tr>
<tr>
<td>Socioeconomic controls</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Year FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>City FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>City-year FE</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>City trend</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
Table A.7: College Enrollment and College Graduates: Heterogeneity by Earnings

The table reports coefficients from an OLS regression with robust standard errors clustered at the school-district level reported in parentheses. The dependent variable is the share of high-school graduates that enroll in a 4-year college within 6 months of graduation, with mean equal to 0.33. Column 2 to 6 control for characteristics of the graduating class (the share of females among the high-school graduates and the average ACT score). Socioeconomic controls include the share of black and white residents, the unemployment rate, the median family income, school-district size and median annual earnings of high-school graduates. The sample includes all school-districts within MSA’s in Michigan over the period 2008 and 2014. ***, ** and * represent statistical significance at 1%, 5% and 10% levels, respectively. Source: CEPI, NCES-EDGE and NCES-CCD.

<table>
<thead>
<tr>
<th>Dependent variable: Share of High-School Graduates that Enroll in a 4-year College</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>College Graduates</td>
<td>-5.992***</td>
<td>-5.528***</td>
<td>-4.768***</td>
<td>-4.788***</td>
<td>-4.778***</td>
<td>-4.697***</td>
</tr>
<tr>
<td></td>
<td>(1.454)</td>
<td>(1.272)</td>
<td>(1.125)</td>
<td>(1.113)</td>
<td>(1.115)</td>
<td>(1.102)</td>
</tr>
<tr>
<td>College Graduates × Earnings, College Degree</td>
<td>0.619***</td>
<td>0.552***</td>
<td>0.478***</td>
<td>0.479***</td>
<td>0.478***</td>
<td>0.472***</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.117)</td>
<td>(0.104)</td>
<td>(0.103)</td>
<td>(0.103)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>Earnings, College Degree</td>
<td>0.0125</td>
<td>-0.0852***</td>
<td>-0.0625**</td>
<td>-0.0606**</td>
<td>-0.0618**</td>
<td>-0.0575**</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.026)</td>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>ACT Score</td>
<td>0.0294***</td>
<td>0.0360***</td>
<td>0.0361***</td>
<td>0.0363***</td>
<td>0.0365***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Expenditure per student</td>
<td>0.00323</td>
<td>0.00230</td>
<td>0.00310</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Revenue per student</td>
<td>0.00412</td>
<td>0.00523</td>
<td>0.00397</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers to student ratio</td>
<td>-0.00287</td>
<td>-0.00236</td>
<td>-0.00227</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1841</td>
<td>1839</td>
<td>1827</td>
<td>1818</td>
<td>1818</td>
<td>1818</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.734</td>
<td>0.795</td>
<td>0.804</td>
<td>0.805</td>
<td>0.810</td>
<td>0.807</td>
</tr>
<tr>
<td>Socioeconomic controls</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Year FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>City FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>City-year FE</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>City trend</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
Table A.8: College Enrollment and College Graduates: Robustness Checks

The table reports coefficients from an OLS regression with robust standard errors clustered at the school-district level reported in parentheses. The dependent variable is the share of high-school graduates that enroll in a 4-year college within 6 months of graduation, with mean equal to 0.33. Each column replicates column 6 in table using either a different proxy for high-skill neighbors earnings or a different sample. Column 1 and 2 use, respectively, the median annual earnings of individuals with a post-graduate degree and the average between this variable and median annual earnings of individuals with a college degree. Column 3 restricts the sample to school-districts with less than 10% of non-resident students. Column 4 and 5 report estimation results using only the post-Great Recession years (2010-2014). The former uses all the sample of urban school-districts, while the latter only uses urban school-districts with less than 10% of non-resident students. Column 6 includes school-districts in urban and rural areas. In this specification, I also include a dummy variable that equals one if the school-district belongs to an urban area. All columns include year and city fixed effects, a city-specific trend and a vector socioeconomic controls. The latter include the share of black and white residents, unemployment rate, median family income, school-district size and median annual earnings of high-school graduates. The sample includes all school-districts within MSA’s in Michigan over the period 2008 and 2014. ***, ** and * represent statistical significance at 1%, 5% and 10% levels, respectively. Source: CEPI, NCES-EDGE and NCES-CCD.

<table>
<thead>
<tr>
<th>Dependent variable: Share of High-School Graduates that Enroll in a 4-year College</th>
<th># Earnings Measures Only Resident Only College Graduates</th>
<th>Only College Graduates × Earnings, Post-college Degree</th>
<th>Only ACT Score</th>
<th>Only Expenditure per student</th>
<th>Only Local Revenue per student</th>
<th>Only Teachers to student ratio</th>
<th>Only College Graduates × Earnings, average</th>
<th>Only College Graduates × Earnings, College Degree</th>
<th>Only College Graduates × Earnings, College Degree (squared)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.240)</td>
<td>(1.338)</td>
<td>(1.433)</td>
<td>(1.195)</td>
<td>(1.433)</td>
<td>(1.034)</td>
<td>(0.554)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates × Earnings, Post-college Degree</td>
<td>0.430***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT Score</td>
<td>0.0347***</td>
<td>0.0351***</td>
<td>0.0315***</td>
<td>0.0380***</td>
<td>0.0315***</td>
<td>0.0391***</td>
<td>0.0365***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure per student</td>
<td>0.00340</td>
<td>0.00363</td>
<td>0.00713</td>
<td>0.00440</td>
<td>0.00713</td>
<td>0.00213</td>
<td>0.00310</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local Revenue per student</td>
<td>0.00142</td>
<td>0.00200</td>
<td>0.00138</td>
<td>0.00636</td>
<td>0.00138</td>
<td>-0.00160</td>
<td>0.00398</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers to student ratio</td>
<td>-0.00165</td>
<td>-0.00169</td>
<td>-0.0113</td>
<td>-0.00451</td>
<td>-0.0113</td>
<td>0.00637</td>
<td>-0.00227</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates × Earnings, average</td>
<td>0.490***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.122)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates × Earnings, College Degree</td>
<td>0.452***</td>
<td>0.483***</td>
<td>0.452***</td>
<td>0.382***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td>(0.111)</td>
<td>(0.134)</td>
<td>(0.096)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Graduates × Earnings, College Degree (squared)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0219***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>Observations</td>
<td>1815</td>
<td>1815</td>
<td>876</td>
<td>1424</td>
<td>876</td>
<td>3023</td>
<td>1818</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.811</td>
<td>0.811</td>
<td>0.792</td>
<td>0.782</td>
<td>0.792</td>
<td>0.689</td>
<td>0.807</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table A.9: Parameters

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameter</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Exogenously chosen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of neighborhoods</td>
<td>$J$</td>
<td>95</td>
<td>Number of school-districts within Detroit in 2013 (CEPI)</td>
</tr>
<tr>
<td>Risk-aversion (CARA)</td>
<td>$\gamma$</td>
<td>0.5</td>
<td>Babcock et al. (1993)</td>
</tr>
<tr>
<td>Low-skill wage’s mean</td>
<td>$w^L$</td>
<td>7.65</td>
<td>Low-skill workers earnings distribution (ACS 2008-2013)</td>
</tr>
<tr>
<td>High-skill wage’s mean</td>
<td>$w^H$</td>
<td>8.47</td>
<td>High-skill workers earnings distribution (ACS 2008-2013)</td>
</tr>
<tr>
<td>High-skill wage’s variance</td>
<td>$\sigma^H$</td>
<td>0.03</td>
<td>High-skill workers earnings distribution (ACS 2008-2013)</td>
</tr>
<tr>
<td>Prior mean</td>
<td>$\bar{\mu}^2$</td>
<td>7.7</td>
<td>Bleemer and Zafat (2015)</td>
</tr>
<tr>
<td>Prior variance</td>
<td>$\bar{\sigma}^2$</td>
<td>0.055</td>
<td>Bleemer and Zafat (2015)</td>
</tr>
<tr>
<td>Panel B: Estimated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost function parameter</td>
<td>$\bar{\sigma}$</td>
<td>8.18</td>
<td>Mean deviation of enrollment</td>
</tr>
<tr>
<td>Cost function parameter</td>
<td>$\varphi$</td>
<td>0.96</td>
<td>Std. deviation of enrollment</td>
</tr>
<tr>
<td>Cost function parameter</td>
<td>$\phi$</td>
<td>0.66</td>
<td>$p75/p50$</td>
</tr>
<tr>
<td>Cost function parameter</td>
<td>$\rho$</td>
<td>0.10</td>
<td>Corr. btw. college graduates and enrollment</td>
</tr>
<tr>
<td>Cost function parameter</td>
<td>$\kappa$</td>
<td>0.17</td>
<td>Corr. btw. college graduates and expenditures per student</td>
</tr>
</tbody>
</table>
Panel A plots the average college enrollment for each quartile of the distribution of college graduates. The dashed and the solid line represent, respectively, school-districts in the last and first quartile of the distribution of college graduates’ earnings. Average college enrollment is measured as the average % of high-school graduates that enroll in 4-year colleges within 6 months after graduation across school-districts in each quartile. Panel B plots the difference between the dashed and the solid line in panel A for each quartile of the distribution of college graduates. The dashed lines in panel B represent bootstrap confidence intervals at the 95% level. Source: CEPI, NCES-EDGE and NCES-CCD and author’s calculations (2008-2013).
Figure A.2: Interpretation of Coefficients’ Signs

The graph illustrates the effect of college graduates on college enrollment along the earnings dimension under different signs of the coefficients of interest, $\beta_1$ and $\beta_2$. The red line displays the effect under the human capital spillovers channel proposed in the literature.

Figure A.3: College graduates and Enrollment: Heterogeneity By Earnings

Both panels plot the average marginal effect of an increase in the share of college graduates by one unit on the college enrollment rate for different levels of median earnings of college graduates. The left panel plots the average marginal effect from the specification in column 4 in table A.7, while the right panel plots the average marginal effect when I consider a quadratic specification in earnings (column 7 in table A.8). The shaded area represents 95% confidence intervals. The x-axis corresponds to the log median earnings of college graduates in 2010 dollars.
Figure A.4: College graduates and Enrollment: Adjusted-bias Coefficients

This figure plots the average marginal effect when I use the estimated coefficients in column 6 in table A.7 (blue line) and the bias-adjusted coefficients, $\beta^*_1$ and $\beta^*_2$ (green line) when the influence of unobservables on the outcome variable is of similar magnitude as the impact of observable variables, $\delta = 1$. $\beta^*_i = \hat{\beta}_i - \delta(\hat{\beta}_i - \tilde{\beta}_i) \frac{1 - \hat{R}^2}{\hat{R}^2}$, where $\hat{\beta}$ are the estimated coefficients and $R^2$ of column 6 in table A.7 and $\tilde{\beta}$ and $\tilde{R}$ are the estimated coefficients and $R^2$ of OLS estimation of equation 2 with no controls (i.e. not including city and year fixed effects, a city-specific trend and the vector $X$). The x-axis corresponds to the log median earnings of college graduates in 2010 dollars.
B Theoretical Appendix

Location decisions I report additional details for the characterization of parents locations decisions, as described by 9. Given the Fréchet distribution for the idiosyncratic taste, \( \varepsilon_{i,j} \sim \text{Fréchet}(\theta, 1) \), it follows that \( \varepsilon_{i,j}^{-1} \sim \text{Weibull}(\theta, 1) \).\(^{31}\) Hence, the indirect utility function described by 8 is also Weibull distributed:

\[
v_{i,k,j} \varepsilon_{i,j} \sim \text{Weibull}(\theta, v_{i,k,j}) \tag{B.1}
\]

where \( v_{i,k,j} = -\exp(-\gamma(w_{i,k,j} - R_j)) \), with \( \Phi_{k,j} = \alpha_{s,n} \cdot \alpha_{u,n} \cdot \beta_{j} \cdot s \cdot n \), is a constant.\(^{32}\) Let \( X_1, \ldots, X_n \) be statistically independent, with each \( X_i \sim \text{Weibull}(\theta, v_i) \), for \( \theta, v_1, \ldots, v_n > 0 \). Then

\[
Pr[k \in \arg\min X_i] = \frac{v_k^{-\theta}}{\sum_i v_i^{-\theta}}, \forall k \in I
\tag{B.2}
\]

Combining B.1 and B.2, and setting \( \theta = 1 \), the probability that a parent \( i \) with skill level \( k \) chooses to live in location \( j \) out of all possible locations, \( \rho_{i,k,j} \), is:

\[
\rho_{i,k,j} = Pr[U_{i,k,j} \geq U_{i,k,n'}; \forall j' \in J],
\]

\[
= \frac{a_{j,n} \exp(\gamma(w_{i,k,j} - r_j))}{\sum_{j' \in J} \Phi_{k,j'} \exp(\gamma(w_{i,k,j} - r_{n'}))}
\]

which simplifies to

\[
\rho_{k,j} = \frac{\Phi_{j,n} \exp(\gamma(-r_j))}{\sum_{j' \in J} \Phi_{k,j'} \exp(\gamma(-r_{j'}))}
\tag{B.3}
\]

Because \( \rho_{i,k,j} \) does not depend on the wage, which is the same no matter where the family lives in the city, it is equal across individuals in the same skill group. Given this, the number of \( k \)-skill parents in each neighborhood is

\[
M_k^j = \sum_{i=1}^{M_k} \rho_{i,k,j} = \sum_{i=1}^{M_k} \rho_{k,j} = \rho_{k,j} \cdot M^k
\]

B.1 Spatial Equilibrium - An Illustration

Let’s consider the example of a city with two neighborhoods, 1 and 2, each with the same capacity, \( H_1 = H_2 \). I set \( A_{2,1} = A_{1,2} = A_{2,2} = 2.5 \), and look to the spatial equilibrium for different values of \( A_{1,1} \). Panels A, B and C in figure B.1 show, respectively, the equilibrium skill-mix in neighborhood 1 and 2 and equilibrium rents in both locations, the endogenous variables, as a function of \( A_{1,1} \). At low values of \( A_{1,1} \), the probability of choosing to live in neighborhood 2, conditional on being a high-skill parent, is high relative to the probability of choosing to live in neighborhood 1. On the other hand, the probability of choosing to live in neighborhood 2, conditional on being a low-skill parent, is very

---

\(^{31}\)The cumulative distribution function of the Weibull distribution with parameters \( \theta \) and \( \lambda \) is \( Pr(X \leq x) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^\theta\right) \) with \( x \geq 0 \). The mean is \( \lambda \Gamma(1 + 1/\theta) \) and the variance is \( \lambda^2 [\Gamma(1 + 2/\theta) - \Gamma^2(1 + 1/\theta)] \). Since \( \beta \), the scale parameter of the Fréchet distribution, is equal to 1, \( \lambda = 1 \).

\(^{32}\)If \( Y = tX \), where \( X \sim \text{Weibull}(\theta, 1) \), then \( Y \) is \( \text{Weibull}(\theta, t) \).
low due to the high rents in this location. This makes neighborhood 2 mainly composed of high-skill households. At high values of $A_{1,1}$, neighborhood 1 becomes more attractive to high-skill families, increasing housing prices in neighborhood 1. Higher rents in neighborhood 1, in turn, make this neighborhood less attractive, and low-skill households transfer to neighborhood 2. Note that when amenities are equal across neighborhoods, rents and the skill-mix of each location is also equal. In this situation, the spatial equilibrium is non-sorted.

**Figure B.1: Spatial Equilibrium - An Example**

Panel A: Equilibrium skill-mix in neighborhood 1 for different levels of $A_1$ in neighborhood 1, share of high-skill households (solid line) and share of low-skill households (dotted line). Panel B: Equilibrium skill-mix in neighborhood 2 for different levels of $A_1$ in neighborhood 1, share of skilled families (solid line) and share of unskilled families (dotted line). Panel C: Equilibrium rents in neighborhood 1 (dotted line) and neighborhood 2 (solid line) for different levels of $A_1$ in neighborhood 1. $H_1 = H_2 = 75$, $M^H = 100$, $M^L = 50$, $\beta_H = 1$, $\beta_L = 0$, $A_{1,2} = A_{1,2} = A_{2,2}$.

**B.2 Value Functions**

For a child born in neighborhood $j$, the perceived value of being a high-skill worker, $V_j^H$, is given by

$$V_j^H = \sum_{j' \in \mathcal{J}} \Gamma \left(1 + \frac{1}{\theta} \right) E_{w_i} [U(c_{i,j}, \Phi_{H,j'}) | I_j] \rho_{j'}^H$$

(B.4)

where $\Gamma \left(1 + \frac{1}{\theta} \right)$ is the expected value of the idiosyncratic component of utility and $\Gamma(.)$ the gamma function. $E$ is the expectations operator and the expectation is taken over the high-skill wage. $\rho_{j'}^H$ is the probability of living in neighborhood $j'$ conditional on being a high-skill worker.\(^{33}\) I assume $\theta = 1$

\(^{33}\)Since the idiosyncratic taste and the skilled wage are two independent random variables, it follows that $E[w_i^H \cdot \epsilon_{i,j}] = E[w_i^H] \cdot E[\epsilon_{i,j}]$
for simplicity, hence \( \Gamma\left(1 + \frac{1}{\theta}\right) = 1 \). B.4 is equal to

\[
\sum_{j' \in J} \mathbb{E}_{w_i} \left[ \frac{-\exp(-\gamma(w_i^H - r_{j'}))}{\Phi_{s,n'}} \right] \rho_{j'}^H = \sum_{j' \in J} \left[ \frac{-\exp(-\mu_j - \gamma(\hat{\sigma}_2^2/2) - r_{j'})}{\Phi_{j'}^H} \right] \rho_{j'}^H =
\]

which simplifies to

\[
V_j^H = -\exp(-\gamma(\hat{\mu}_j - \gamma(\hat{\sigma}_2^2/2))) \left( \frac{J}{\sum_{j' \in J} \exp(\gamma r_{j'})} \right) \tag{B.5}
\]

B.5 is equal all children born in neighborhood \( j \), but different across children from neighborhoods as long as the share of skilled individuals differs.

For a child born in neighborhood \( j \), the expected value of becoming an unskilled worker, \( V_j^L \), is given by

\[
V_j^L = \sum_{j' \in J} \Gamma \left(1 + \frac{1}{\theta}\right) U(c_{L,j'}, \Phi_{j'}^L) \rho_{j'}^L \tag{B.6}
\]

where \( \Gamma\left(1 + \frac{1}{\theta}\right) \) is the expected value of the idiosyncratic component of utility and \( \Gamma(.) \) the gamma function. \( \rho_{j'}^L \) is the probability of living in neighborhood \( j' \) conditional on being a high-skill worker. I assume \( \theta = 1 \) for simplicity, hence \( \Gamma\left(1 + \frac{1}{\theta}\right) = 1 \). B.6 is equal to

\[
\sum_{j' \in J} \left[ \frac{\exp(\gamma(u^L - r_{j'}))}{\Phi_{j'}^L} \right] \rho_{j'}^L
\]

which simplifies to

\[
V^L = -\exp(-\gamma u^L) \left( \frac{J}{\sum_{j' \in J} \exp(\gamma r_{j'})} \right) \tag{B.7}
\]

B.7 is equal for all children in the city, regardless of where they live. Hence I suppress \( j \).
B.3  Proofs

Proof of Lemma 3  Given $V^H_j$ (B.5) and $V^L_j$ (B.7), the perceived skill premium for a child born in neighborhood $j$, $\Delta V_j \equiv V^H_j - V^L_j$, is given by

$$\Delta V_j = J \left( \frac{- \exp(-\gamma(\hat{\mu}_j - \gamma(\hat{\sigma}_j^2/2)) - \exp(-\gamma w^L_j)}{\sum_{j' \in J} \frac{\phi_{j'}^H}{\exp(\gamma r_{j'})}} - \frac{\exp(-\gamma w^L_j)}{\sum_{j' \in J} \frac{\phi_{j'}^L}{\exp(\gamma r_{j'})}} \right)$$  (B.8)

where $j$ indexes the neighborhood where the child lives, $J$ is the number of neighborhoods in the city.

The optimal investment decision takes the form of a cut-off rule. The ability cut-off, $a^*_j$, is defined by the indifference condition $\Delta V_n = c(a^*_n)$. Defining $\varpi_j \equiv \Delta V_j - c(a^*_j)$, I establish the following:

1. $\frac{\partial s_{i,j}}{\partial \hat{\mu}_j} > 0$. The effect of $\hat{\mu}_j$ on the probability of becoming a high-skill worker, $s_{i,j}$ is given by

$$\frac{\partial s_{i,j}}{\partial \hat{\mu}_j} = \frac{\frac{\partial s_{i,j}}{\partial a^*_j}}{\frac{\partial a^*_j}{\partial \hat{\mu}_j}}$$

By the implicit function theorem, $\frac{\partial a^*_j}{\partial \hat{\mu}_j} = - \frac{\frac{\partial \varpi_j}{\partial a^*_j}}{\frac{\partial \varpi_j}{\partial \hat{\mu}_j}}$. The numerator is higher than zero, the denominator

2. $\frac{\partial s_{i,j}}{\partial \hat{\sigma}_j^2} < 0$. The effect of $\hat{\sigma}_j^2$ on the probability of becoming a high-skill worker $s_{i,j}$ is given by

$$\frac{\partial s_{i,j}}{\partial \hat{\sigma}_j^2} = \frac{\frac{\partial s_{i,j}}{\partial a^*_j}}{\frac{\partial a^*_j}{\partial \hat{\sigma}_j^2}}$$

By the implicit function theorem, $\frac{\partial a^*_j}{\partial \hat{\sigma}_j^2} = - \frac{\frac{\partial \varpi_j}{\partial a^*_j}}{\frac{\partial \varpi_j}{\partial \hat{\sigma}_j^2}}$. The numerator is higher than zero, the denominator

Proof of Proposition 1  The effect of $m^H_j$ on the probability of investing in education $s_{i,j}$ is given by

$$\frac{\partial s_{i,j}}{\partial m^H_j} = \frac{\frac{\partial s_{i,j}}{\partial a^*_j}}{\frac{\partial a^*_j}{\partial m^H_j}}$$

By the implicit function theorem, $\frac{\partial a^*_j}{\partial m^H_j} = - \frac{\frac{\partial \varpi_j}{\partial a^*_j}}{\frac{\partial \varpi_j}{\partial m^H_j}}$. The numerator is higher than zero, the denominator
is given by

\[
\frac{\partial \varpi}{\partial m_j} = \frac{\partial \varpi}{\partial \Delta V_j} \frac{\partial \Delta V_j}{\partial m_j} + \frac{\partial \varpi}{\partial c(a_j^s)} \frac{\partial c(a_j^s)}{\partial m_j} = -\gamma \left( \frac{\sigma^2\mu}{m_j^H (\bar{\sigma}^2 + \sigma^2)^2} \cdot \left[ w_j^H - \bar{\mu}_j + \frac{\gamma \sigma^2}{2} \right] + \frac{\partial \varpi}{\partial c(a_j^s)} \frac{\partial c(a_j^s)}{\partial m_j} \right)_{>0 \text{ or } <0}
\]

where \( \gamma = J \cdot \frac{\exp(-\gamma(\bar{\mu}_j - \gamma(\bar{\sigma}^2/2)))}{\sum_{j' \in J} \exp(\bar{r}_{j'j})} \).

If \( w_j^H > \bar{\mu}_j - \frac{\bar{\sigma}^2}{2} \), then \( \frac{\partial s_{i,j}}{\partial m_j} > 0 \). On the other hand if \( w_j^H < \bar{\mu}_j - \frac{\bar{\sigma}^2}{2} \) and \( w_j^H \) is sufficiently low such that the positive effect through the cost function does not compensate the negative effect through the information channel, \( \frac{\partial s_{i,j}}{\partial m_j} < 0 \). The signal threshold below which the effect \( \frac{\partial s_{i,j}}{\partial m_j} < 0 \) is lower than the one in the case with no human capital spillovers in the cost function.

**B.4 Implications of other specifications**

**Risk neutrality** Consider that individuals have a linear indirect utility function given by

\[
U(w_i^k, r_j, \Phi_j^k, \varepsilon_{i,j}) = w_i^k - r_j + \Phi_j^k + \varepsilon_{i,j}
\]

where \( \Phi_j^k = A_{1,j}^1 A_{2,j}^{1-\beta_k} \) and the utility shock \( \varepsilon_{i,j} \) follows the extreme value type 1 distribution with parameters \( \mu_\varepsilon \) and \( \sigma_\varepsilon \).\(^{34}\) The distributional assumption on the idiosyncratic taste, \( \varepsilon \), allows derive a close-form expression for \( \rho_{i,j}^k \):

\[
\rho_{i,j}^k = \frac{\exp(w_i^k - r_j + \Phi_j^k)}{\sum_{j' \in J} \exp(w_i^k - r_{j'} + \Phi_{j'}^k)}
\]

Other things equal, as before, a type-\( j \) parent is more likely to live in a neighborhood the more attractive are \( j \)-specific amenities and the lower are rental prices \( (r_j) \). Since migration is only allowed in the beginning of the period, \( \rho_{i,j}^k \) translate directly into the neighborhood size distribution. The equilibrium number of \( j \)-skill parents in neighborhood \( j \), \( M_j^k \), is given by

\[
M_{j,k} = \sum_{i=1}^M \rho_{i,j}^k = \rho_j^k M_k
\]

Using B.10 and B.11, I can compute the perceived expected value of being a high-skill worker, \( V_j^H \) and the expected value of being a low-skill worker, \( V_j^L \) functions, and the perceived skill premium for a child born in neighborhood \( j \), \( \Delta V_j \), B.12. It can be shown that:

1. \( \frac{\partial s_{i,j}}{\partial \bar{\mu}_j} > 0 \),
2. \( \frac{\partial s_{i,j}}{\partial \bar{\sigma}_j^2} = 0 \), this follows from the fact that children are risk neutral, and,
3. \( \frac{\partial s_{i,j}}{\partial m_j} > 0 \) if \( w_i^H > \bar{\mu}_j \),

as before.

\(^{34}\) The extreme value type 1 distribution is commonly used in the discrete-choice literature. The density of the extreme value type 1 distribution with parameters parameters \( \mu_\varepsilon \) and \( \sigma_\varepsilon \) is \( f(x) = \exp(-\exp(-(x - \mu_\varepsilon)/\sigma_\varepsilon)) \).
Uncertainty about Low-Skill Wage If $\sigma_{L}^{2} > 0$, B.12 can be re-written as

$$\Delta V_{j} = J \left( \frac{-\exp(-\gamma(\hat{\mu}_{j}^{H} - \gamma(\hat{\sigma}_{L,j}^{2}/2)))}{\sum_{j' \in J} \phi_{j'}^{H} \exp(\gamma_{j,j'})} - \frac{-\exp(-\gamma(\hat{\mu}_{j}^{L} - \gamma(\hat{\sigma}_{L,j}^{2}/2)))}{\sum_{j' \in J} \phi_{j'}^{L} \exp(\gamma_{j,j'})} \right)$$  \hspace{1cm} (B.12)

where $\hat{\mu}_{j}^{H}$ and $\hat{\sigma}_{L,j}^{2}$ are the posterior mean and variance of the beliefs about $w_{i}^{H}; \hat{\mu}_{j}^{L}$ and $\hat{\sigma}_{L,j}^{2}$ are the posterior mean and variance of the beliefs about $w_{i}^{L}$ for a child born in neighborhood $j$. Following the same steps as in the proof of lemma 3 above, it can be shown that:

1. $\frac{\partial s_{i,j}}{\partial \mu_{j}^{H}} > 0$ and $\frac{\partial s_{i,j}}{\partial \mu_{j}^{L}} < 0$
2. $\frac{\partial s_{i,j}}{\partial \sigma_{L,j}^{2}} < 0$ and $\frac{\partial s_{i,j}}{\partial \sigma_{L,j}^{2}} > 0$

Naturally, the higher is the expected value of the low-skill wage, the lower is the probability to invest in education, since the perceived skill-premium is lower, holding all else constant. On the other hand, because individuals are risk-averse, higher uncertainty about the low-skill wage, increases the perceived skill-premium, hence the probability of investing in education.

As before, the effect of $m_{j}^{H}$ on the probability of investing in education $s_{i,j}$ is given by

$$\frac{\partial s_{i,j}}{\partial m_{j,h}} = \frac{\partial s_{i,j}}{\partial a_{i,j}} \frac{\partial a_{i,j}}{\partial m_{j,h}}$$

By the implicit function theorem, $\frac{\partial a_{i,j}}{\partial m_{j,h}} = -\frac{\partial \omega}{\partial m_{j,h}}$. The numerator is higher than zero, the denominator is given by

$$\frac{\partial \omega}{\partial m_{j,h}} = \frac{\partial \omega}{\partial \Delta V_{j}} \left[ \frac{\partial V_{j}^{H}}{\partial m_{j,h}} + \frac{\partial V_{j}^{L}}{\partial m_{j,h}} \right] + \frac{\partial \omega}{\partial c(a_{i}^{*})} \frac{\partial c(a_{i}^{*})}{\partial m_{j,h}}$$

$$= J \cdot \left[ Y_{H} \cdot \frac{\sigma_{H}^{2}}{m_{j}^{H} (\hat{\sigma}_{L}^{2} + \hat{\sigma}_{L,j}^{2})^{2}} \cdot \left[ w_{j}^{H} - \hat{\mu}_{j}^{H} + \frac{\gamma \hat{\sigma}_{H}^{2}}{2} \right] - Y_{L} \cdot \frac{\sigma_{L}^{2}}{m_{j}^{H} (\hat{\sigma}_{L}^{2} + \hat{\sigma}_{L,j}^{2})^{2}} \cdot \left[ -w_{j}^{L} + \hat{\mu}_{j}^{L} - \frac{\gamma \hat{\sigma}_{L}^{2}}{2} \right] \right]$$

where $Y_{H} = \gamma \frac{\exp(-\gamma(\hat{\mu}_{j}^{H} - \gamma(\hat{\sigma}_{L,j}^{2}/2)))}{\sum_{j' \in J} \phi_{j'}^{H} \exp(\gamma_{j,j'})}$, and $Y_{L} = \gamma \frac{\exp(-\gamma(\hat{\mu}_{j}^{L} - \gamma(\hat{\sigma}_{L,j}^{2}/2)))}{\sum_{j' \in J} \phi_{j'}^{L} \exp(\gamma_{j,j'})}$.

1. If $w_{j}^{L} = \hat{\mu}_{j}^{L} - \frac{\gamma \hat{\sigma}_{L}^{2}}{2}$ such that $B = 0$, the results in proposition 1 hold: $\frac{\partial s_{i,j}}{\partial m_{j,h}} > 0$ if $w_{i}^{H} > \hat{\mu}_{j}^{H} + \frac{\gamma \hat{\sigma}_{H}^{2}}{2}$.

2. If $w_{j}^{L} < \hat{\mu}_{j}^{L} - \frac{\gamma \hat{\sigma}_{L}^{2}}{2}$ such that $B > 0$, then the threshold below which $\frac{\partial s_{i,j}}{\partial m_{j,h}} < 0$ is higher than the one in in proposition 1.
3. If \( w^L_j > \bar{\mu}_j^L - \frac{\gamma^2}{2} \bar{\sigma}_j^2 \) such that \( B < 0 \), then the threshold below which \( \frac{\partial s_{i,j}}{\partial m_{j,H}} < 0 \) is lower than the one in in proposition 1.