Neighborhood Impacts on STEM Choice

Jeonghyeok Kim*

September 9, 2024

Preliminary draft- Please do not circulate

Abstract

I show that the neighborhoods where students grow up significantly influence their choice of college major, focusing on students in Texas who move across school districts and counties. By using variation in the grades at which students move, I find substantial exposure effects of neighborhoods on the likelihood of choosing a STEM major. Specifically, students who move to high-STEM areas—measured by the proportion of non-moving students obtaining a STEM degree—show a linear increase in their probability of selecting a STEM major with the time spent in the new place. My estimates suggest that a move from the 25th to the 75th percentile in STEM rates in grade 0 increases the likelihood of majoring in STEM by 1.4 percentage points (34% of the mean). To distinguish the causal effects of neighborhoods from confounding factors, I analyze moves induced by displacement shocks and exploit sharp variations in predicted neighborhood effects across school cohorts, genders, races, and specific college majors. These tests also suggest that the observed exposure effects are not merely due to an overall improvement in educational attainment but due to an increase in major specific exposure. Importantly, I find that the STEM exposure measure is strongly tied to the occupational composition of the neighborhood, further providing causal evidence that neighborhood STEM occupational composition influences students' STEM career choices. Lastly, an examination of heterogeneity across demographic groups reveals that the exposure effect is consistent across genders, races, and economic statuses, indicating that the positive effects of living in high-STEM areas extend equally to students from both disadvantaged and non-disadvantaged families. Equalizing neighborhood exposure during school years could reduce disparities across those groups in STEM major choice by 20-40%.

*Jeonghyeok Kim: Department of Economics, University of Houston, Houston, Texas, USA (jkim124@uh.edu).

1 Introduction

As Science, Technology, Engineering, and Math (STEM) fields become increasingly pivotal to the U.S. economy (Carnevale, Smith, and Melton 2011; Langdon et al. 2011), concerns about a shortage of STEM workers to meet industry demand have grown (Holden and Lander 2012; National Science and Technology Council 2000; National Science Board 2003; Xue and Larson 2015; Zaza et al. 2020). A recent report from the National Science Foundation continues to warn of an accelerating STEM talent crisis that threatens U.S. economic and national security (National Science Board 2024). Moreover, the representation of different races, genders, and socio-economic statuses in STEM remains imbalanced (Maltese and Tai 2011; Rozek et al. 2019; Xie, Fang, and Shauman 2015). Despite making up 28% of the labor market, Black and Hispanic individuals hold only 17% of STEM positions. Women account for just a quarter or less of the workforce in computing and engineering sectors (Fry, Kennedy, and Funk 2021), a disparity that also contributes to wage gaps (Blau and Kahn 2017).

A body of research centers on individual or family characteristics—ability, preference, and preparedness—as explanatory factors for the choice of STEM and for disparities among demographic groups (Archer et al. 2012; Arcidiacono, Aucejo, and Hotz 2016; Hanushek et al. 2023). While some previous research documents on educational situations —including school resources, stereotypes, teachers, and school atmosphere (Beilock et al. 2010; Clotfelter, Ladd, and Vigdor 2005; Lavy and Sand 2015; Legewie and DiPrete 2014; Wang 2013), relatively little is known about the importance of non-educational environments and the importance of the overall neighborhood.

Using Texas administrative data covering students in the public education system, I present new quasi-experimental evidence on the effects of neighborhoods on college STEM choices—defined as attaining a bachelor's degree in STEM majors. My analysis is motivated by Chetty and Hendren (2018) showing that the neighborhoods in which children grow up shape their adulthood outcomes, including earnings and college attendance rates. This paper extends this by going beyond better-worse comparison to the characteristics of neighborhoods. I explore whether neighborhood exposure even shapes the choice of field of study and potential career choices. For

instance, I wonder if students living near NASA stations are more likely to study engineering while students living near medical centers are more likely to choose to study health-related majors.

In this paper, I conduct three sets of analyses. First, I estimate the causal effects of neighborhoods on STEM choices with multiple tests validating the empirical design. Second, I link these neighborhood effects to the local STEM occupational composition. Third, I assess the heterogeneity of these effects across different groups to determine whether neighborhood effects extend to underrepresented groups.

In the baseline analysis, I focus on students who enrolled in kindergarten (grade 0) between 1994 and 2003 and moved once across school districts during their school years. On average, spending an additional year in a school district where the mean STEM rate among non-moving students, who never moved across school districts in school years, is 1 percentage point higher increases a student's probability of majoring in STEM in college by approximately 0.04 percentage points. In other words, the likelihood of a student choosing a STEM major converges to the STEM rate of permanent residents in the new district at a rate of 4% per year of exposure. Conversely, moving to an area with a lower STEM rate among non-moving students reduces a student's expected probability of choosing a STEM major by 4% per year of exposure. These effects are also consistent when students move across counties.

The findings suggest that neighborhoods exert significant exposure effects on students' STEM major choices: each additional school year spent in a high-STEM environment increases the likelihood of a student choosing a STEM major. The college major choice of students who move gradually converges to non-moving students who live in their new neighborhood, with this convergence occurring at a constant rate throughout their school years. In other words, the annual exposure effect is consistent—moving to a high-STEM area in grade 3 rather than grade 4 has the same impact on STEM probability as moving in grade 10 instead of grade 11. The estimates indicate that $13\times4=52\%$ of the variation in STEM major choices across different areas can be attributed to the causal effects of place in their school years. The convergence rate of 4% per year implies that a child moving in kindergarten from the 25th percentile (0.0259) to the 75th percentile (0.0539) in STEM rates—a 1.24 standard deviation difference—would see an

increase in the probability of majoring in STEM by 1.42 percentage points. This represents 34% of the average STEM major choice rate (4.18%).

The key identification assumption for interpreting the 4% convergence rate as a causal exposure effect is that the likelihood of majoring in STEM for students who move to higher versus lower STEM areas does not vary based on the grade at which they move. To assess the validity of this assumption, I first conduct a series of placebo (overidentification) tests using the heterogeneity in outcomes among non-moving students across different subgroups, such as cohorts, genders, and races. Although STEM rates within school districts are generally stable over time, some districts experience improvement while others decline. Similarly, while STEM rates are highly correlated across genders and races, there are some places where one group does worse than others and vice versa. Using this variation, I perform multivariable regressions and find that students who move to a new neighborhood exhibit outcome convergence with non-moving students of the same school cohort and the same gender/race group in the destination, but not or less with those of the other school cohorts and other gender/race groups (while controlling for their own cohort's/group's STEM rates). This cohort/group-specific convergence supports the hypothesis that places have causal effects on students' outcomes, proportional to the time they spend in the new neighborhoods. Such a pattern would be unlikely to arise from sorting or other omitted variables, as these cohort-specific effects only manifest after a significant time lag, once students graduate college.

Although the estimated exposure effects are not driven by selection or unbiased, they might be influenced by the overall educational achievement of the area rather than the STEM exposure, even when controlling for other place characteristics, including changes in bachelor's degree rates between origin and destination. In that case, the neighborhood effects I estimated are not about choice of study but about improvement in general educational attainment. To explore this possibility, I conduct a placebo test using different majors: STEM, business, and liberal arts. I find that the major of students who move to a new area is only responsive to the rate of the same major of non-moving students of the destination, but is unrelated to the rate of other majors (conditional on the same major predictions). This major-specific convergence aligns with the expectation that places have causal effects on children's major choices based on exposure time,

but it is unlikely to result from improvements in general educational achievement. In addition, I find consistent estimation results with baseline results using a subset of moves more likely to be driven by exogenous aggregate shocks.

Thus, the tests suggest that for any omitted variable to bias the exposure effect estimates, it would need to produce similar results even with displacement shocks and closely match the STEM rates of non-moving students by cohort, gender, and race, in proportion to exposure time. Additionally, the variable would need to reflect major-specific convergence. Given the improbability of such a variable existing, I conclude that the baseline estimate of a 4% convergence in STEM major choice per year of exposure to neighborhoods is likely unbiased.

Next, I explore the reasons behind the convergence in college major choices. A neighborhood is defined in this paper to encompass various aspects of a place. My measure for exposure effects–STEM rates of non-movers–summarizes all those aspects. When it comes to STEM choice, a neighborhood can be broadly divided into school and non-school environments, which interact with each other. While the impact of school environments, such as curriculum and teachers, is well documented, non-school environments have received less attention. To explore this, I use the occupational composition of a neighborhood as a proxy for the non-school environment and examine its influence on students' choice of college majors.

I find a strong correlation between the proportion of STEM occupations in a neighborhood and the proportion of students majoring in STEM, regardless of whether the sample is restricted to bachelor's degree holders. In contrast, I do not observe a similar relationship for other majors. I then replicate the baseline analysis using the STEM occupation composition of the school district as the exposure measure. I find constant neighborhood effects of STEM occupation exposure. The findings suggest that spending an additional year in a school district where the proportion of STEM occupations is 1 percentage point higher increases a student's probability of majoring in STEM by 0.026 percent. Given that the standard deviation of the STEM occupation measure is twice as large as that of the baseline exposure measure, these results highlight the significant influence of non-educational environments on students' STEM major choices. Although individual occupations are not available in my data, estimating the proportion of STEM occupations within industries, I further find that neighborhood influences do indeed extend to

occupational choices. Additionally, an overidentification test shows that this effect is specific to students' gender (i.e., female STEM workers to female students) and STEM occupations, supporting the idea that occupational compositions causally impact children's major choices rather than merely reflecting neighborhood selection.

The analyses using STEM occupations in neighborhoods highlight the significant influence of non-educational environments on fostering future generations of STEM professionals. Given that it is unlikely that parents in STEM occupations selectively move to certain areas because of educational environments based on the results of the overidentification test, the strong correlation between high STEM rates among non-moving students and high STEM occupation rates in the neighborhoods is more likely driven by the prevalence of STEM occupations, rather than the reverse. Additionally, high-STEM areas tend to offer more robust STEM education environments, such as greater availability of science courses. This suggests that high STEM occupation rates may naturally contribute to creating educational environments that are more conducive to pursuing STEM fields and, ultimately, STEM careers.

An important question that remains is whether high-STEM neighborhoods also encourage underrepresented groups to major in STEM. For example, are students from economically disadvantaged families more likely to pursue STEM majors if they live near a NASA space station, even if their parents have no connection to NASA? This question is particularly important to understand the gaps across demographic and socioeconomic status in STEM choices. To investigate this, I examine the heterogeneity in exposure effects across economic status, gender, and race. Focusing first on economic disadvantage status, I find that convergence rates are similar regardless of economic background, indicating that the benefits of living in high-STEM areas extend equally to students from both disadvantaged and non-disadvantaged families. However, students from not-disadvantaged families exhibit larger positive selection effects when they move to higher-STEM areas compared to disadvantaged students. This suggests that while the exposure effects are similar for both groups, the choice to move to high-STEM areas may be more deliberate for non-disadvantaged families.

Next, I examine the heterogeneity across gender and race and again find comparable exposure effects. While the exposure effects are similar across groups, students of different economic

statuses and races tend to live in different neighborhoods, making the levels of exposure vary among these groups. To explore the impact of varying exposure levels, I conduct a simple counterfactual experiment: how much would the gap in STEM choice be reduced if students from different backgrounds lived in the same neighborhoods? On average, White students' neighborhood STEM exposure is 39-61% higher than that of Black students and 30-52% higher than that of Hispanic students. If Black and Hispanic students experienced the same neighborhood exposure rates as White students over school years, their likelihood of majoring in STEM would increase by 0.68-1.14 and 0.68-1.04 percentage points, respectively—representing 21-34% and 31-44% of the existing gap. Similarly, the average neighborhood exposure rate for economically not-disadvantaged students is 41-71% higher than that of disadvantaged students. Applying the same thought experiment suggests that the STEM choice among disadvantaged students could increase by 0.88-1.3 percentage points, or 23-34% of the gap. In the similar counterfactual equalizing the exposure between male and female students using own gender convergence, I find that women's STEM choice would increase by 0.43 percentage points, closing 27% of the gap.

I estimate neighborhood effects on other college majors beyond STEM, which cover more than 70% of four-year college graduates, and analyze, instead of the baseline measure, bachelor's degree completion, STEM enrollment in any college and STEM graduate degree completion. I find consistent results, which generalize the findings from STEM majors to other college majors and different levels of educational attainment. The insights from this paper suggest that the influence of neighborhood effects extends beyond STEM majors to all college majors and impacts not just bachelor's degree holders but also other levels of post-secondary education. Thus, my research highlights that students' choices of study are shaped not only by their innate or family-influenced preferences or abilities but also by their early experiences and exposures in the neighborhoods. This influence plays a significant role in their overall decisions regarding post-secondary education.

The first contribution of this study is to offer a novel perspective on the factors leading to STEM. Existing literature emphasizes individual (Stinebrickner and Stinebrickner 2014; Tai et al. 2006; Wang 2013), family (Castleman, Long, and Mabel 2018; Miller and Kimmel 2012;

Mulligan, Hastedt, and McCarroll 2012), and school characteristics (Arold 2024; Bleemer and Mehta 2021; Clotfelter, Ladd, and Vigdor 2005; Oliver et al. 2021). In their review of the literature, Xie, Fang, and Shauman (2015) find less is known about potential contextual factors, such as local labor market characteristics or proximity to science-focused industry.[1] This paper extends the literature by introducing the concept of contextual factors that causally promote students' engagement with STEM education and identifying that occupational composition of the neighborhoods is crucial in the neighborhood effects.

The second contribution is to provide an explanation for gender and racial gaps in STEM (Xie, Fang, and Shauman 2015). Previous literature provides multiple channels including abilities (Ellison and Swanson 2010; Hedges and Nowell 1995), preparedness (Arcidiacono, Aucejo, and Hotz 2016; Card and Payne 2017), cultural beliefs (Carrell, Page, and West 2010; Lavy and Sand 2015; Papageorge, Gershenson, and Kang 2020), school environment (Landaud, Ly, and Maurin 2020; Philippis 2023). I add another important layer—neighborhood—that has not been much studied. I find that students are similarly affected by neighborhoods to choose to study STEM across genders, races, and socio-economic status, which implies 20%-40% of the gaps across groups could be decreased if they had the same rate of exposure in school years.

Lastly, this study makes a contribution to the understanding of developmental neighborhood effects for children by highlighting the multidimensionality of neighborhoods. Previous research has underscored the role of neighborhoods in shaping children's outcomes (Sampson, Morenoff, and Gannon-Rowley 2002; Sharkey and Faber 2014; Wodtke, Harding, and Elwert 2011). To isolate the causal effects, recent works use unique events such as the Moving to Opportunity policy or quasi-experimental situations moving across places, and find substantial positive impacts of having better neighborhoods in developmental periods (Chetty and Hendren 2018; Chetty, Hendren, and Katz 2016; Chyn 2018; Nakamura, Sigurdsson, and Steinsson 2022). Following the empirical strategy used in Chetty and Hendren (2018), I contribute to this literature by showing that neighborhood effects extend beyond the better-worse comparison, influencing the choice of field of study and potentially career path as well. This investigation also elucidates

[1] Although some relationships between institutional characteristics and students' STEM choices or outcomes have been documented (Catsambis and Beveridge 2001; Greenman, Bodovski, and Reed 2011), it is important to note that the observed associations may be confounded by the selective sorting of students into different institutions.

the mechanisms underlying the long-term positive effects presented in earlier studies.

2 Data

I use individual-level Texas administrative data sets, which include three sources: K-12 education records in public schools starting from the academic years 1994-1995 from the Texas Education Agency (TEA), which allow me to track students' neighborhoods; all public and most private post-secondary education data in Texas from the Texas Higher Education Coordinating Board (THECB)[2], allowing me to construct the STEM measure.; quarterly individual data on employment, industry, and earnings for all workers covered by the Unemployment Insurance program from the Texas Workforce Commission (TWC)[3], to construct STEM industry measure and control variables.

One limitation of the THECB and TWC data is that the data coverage is restricted to Texas. If someone goes out of Texas, I cannot observe their out-of-state educational or workforce outcomes and thus cannot distinguish whether they have moved out of state or did not attend college. However, this is unlikely to significantly bias the results considering the relatively low out-migration of Texas (Foote and Stange 2022) and the empirical designs I implement.

2.1 Sample Definitions

The analysis sample includes all students who started kindergarten (grade 0) between 1994-2003. Students also need to be observed for at least 10 school years to be included. I divide the sample into two parts: non-movers and movers. I define the non-movers of each school district p as the subset of students who reside in a single school district in all school years. The movers sample consists of students who are not entirely residing in one school district. In my baseline analysis, I focus on the subset of individuals who attend school districts with non-mover enrollment above

[2] The THECB data contain all public community and technical colleges; all public universities and health-related institutions; almost all independent colleges and universities (available from 2003 onward); and career schools and colleges (available from 2004 onward). See http://www.txhighereddata.org/Interactive/CBMStatus/ for additional information on participating institutions.

[3] Unemployment Insurance covers workers if employers pay $1,500 or more in a calendar quarter, or have at least one employee during twenty different weeks in a calendar year. Thus, TWC data does not include earnings from independent contract work, self-employment, under-the-table payments, earnings from federal jobs, and earnings outside Texas. For more details, see https://www.twc.texas.gov/tax-law-manual-chapter-3-employer-0.

700 (i.e., on average, 70 students in each school cohort; 9.6% of the observations) to ensure that I have adequately large samples to estimate non-moving students' outcomes (the key independent variables in my analysis) precisely. There are approximately 1,612,323 students in the baseline analysis sample, of whom 1,278,084 are non-moving students.

2.2 Variable Definitions and Summary Statistics

In this section, I define the key variables I use in the analysis. I begin by defining the STEM choice of a student and then defining the two key variables I measure for STEM exposure: STEM attainment of non-moving students and neighborhood STEM occupational composition.

My baseline STEM choice measure is defined as obtaining a bachelor's degree in STEM majors from institutions in Texas. STEM majors are classified following the STEM major list from the Department of Homeland Security in 2016.[4] Later in this paper, I also define it as enrolling in STEM majors in any college or obtaining a graduate degree.

I construct a measure of STEM occupation choice based on the industry in which individuals work. While direct data on students' occupations after graduation is unavailable, I develop a proxy for STEM occupation using industry information. Leveraging the Occupational Employment Statistics (OES) Survey from the Bureau of Labor Statistics, I calculate the proportion of STEM workers within each 4-digit NAICS industry code. I then compute the average proportion of STEM workers in the industries where individuals are employed after age 24.

My primary measure of STEM exposure is defined as among non-moving students in a school district, the proportion of students obtaining a bachelor's degree in STEM majors by age 26:

$$STEM_p = \frac{\text{STEM BA}_p}{\text{All students}_p}$$

I use the occupational composition of neighborhoods as a proxy for the non-school environment. Although I do not have access to students' addresses, I can approximate neighborhood occupational composition by using the school addresses to estimate the school attendance zones. Using occupational data from the 2000 Census at the Census tract level, I first construct a

[4] See https://www.ice.gov/sevis/schools#dhs-stem-designated-degree-program-list-and-cip-code-nomination-process for the full list of STEM majors.

measure of STEM occupation exposure for each tract.[5] Based on the list of STEM occupations from ONET, I calculate the STEM proportion in each Census tract ($STEM_c$) as the ratio of STEM workers to the total population. The STEM occupation exposure measure for each school district ($STEMOcc_p$) is then defined as the average STEM proportion of the Census tracts where the schools (s) are located, along with nearby tracts, weighted by the number of students enrolled in each school:

$$STEMOcc_p = \sum_{s \in p} \left(\frac{Enroll_s}{\sum_{s \in p} Enroll_s} \times STEM_{c(s)} \right)$$

Table 1 presents the summary statistics for the exposure measures in the analysis sample. The first two columns show statistics for non-moving students who live in school districts with an average cohort size of more than 70 non-moving students. The next two columns focus on about 500,000 students who moved once between school districts with similarly sized cohorts. Although my analysis does not require movers to be comparable to non-movers, I find that non-moving students have higher STEM attainment than students who moved once. The average STEM rate is 4.7% for non-moving students, compared to 3.8% for one-time movers.

3 Empirical Strategy

This section starts by documenting how students' STEM choices vary across different places in Texas. Following this, I define the key exposure effects and explain the research design used to identify these effects using observational data.

3.1 Geographical Variation in Outcomes of Permanently Residing Students

I conceptualize "neighborhood" effects as the combined influence of various factors, including schools, teachers, peers, neighbors, and local workers. In the main text, I focus on variations across school districts. Texas has approximately a thousand school districts, each with a median of 13 schools and an enrollment of 497 students. I replicate the results presented in the main

[5] Census tracts are small, relatively permanent statistical subdivisions of a county, with an average population of about 4,000 inhabitants.

text at the county level in Appendix B.

I present the STEM rates of students who attended the same school district throughout their school years. Let $\overline{STEM}_p$ represent the average STEM rate for these students in district p. Figure 1 illustrates the average STEM rates by school district, with districts divided into seven quantiles based on their $\overline{STEM}_p$ values. Darker shades indicate districts with higher average STEM rates. There is significant variation in STEM attainment across school districts. The pattern is persistent even when I only consider students who earned a bachelor's degree. The correlation between district-level STEM rates among all students and among bachelor's degree holders is 0.66. Thus, the variation in STME choice across school districts is not just because of differences in general educational attainment. Notably, the districts with the highest STEM rates—particularly the top five—are located near major hubs like NASA Johnson Space Center or Austin, known for its concentration of tech companies.

Moreover, the variation in STEM rates across different areas is notably persistent over time. Figure 2 compares STEM rates for school districts between the 1994-1997 and 2001-2003 cohorts of non-moving students. Districts with high STEM rates in earlier cohorts tend to have higher STEM rates in later cohorts as well. This persistence is evident even when the analysis is limited to students with a bachelor's degree. This enduring trend could be attributed to two potential factors: neighborhood effects influencing students' field of study choices or specific types of students gravitating towards areas where STEM majors are more prevalent. In the following analysis, I explore how much of this difference across neighborhoods is attributed to the causal effects of neighborhoods on STEM choices.

3.2 Estimating Exposure Effects

I aim to assess the potential increase in student STEM choices if they were raised in environments where non-moving students have higher STEM rates. To investigate this, I analyze students who move between school districts with varying STEM rates in different grades to estimate the impact of exposure to neighborhoods. I generally follow Chetty and Hendren (2018) for estimation designs.

Imagine a hypothetical experiment where students are randomly assigned to different loca-

tions p at grade m and remain there for the remainder of their schooling. In this experimental setup, the predicted STEM choice $STEM_i$ for students can be expressed as a linear function of the average STEM rate $\overline{STEM}_p$ for non-moving students in school district p:

$$STEM_i = \alpha + \beta_m \overline{STEM}_p + \theta_i, \tag{1}$$

where the error term θ_i accounts for family influences and other factors affecting students' STEM choices. Because random assignment ensures that θ_i is unrelated to $\overline{STEM}p$, applying OLS to estimate the model provides a coefficient β_m that reflects the average effect of spending grade m and subsequent years in a district with higher STEM rates among non-moving students. The exposure effect γ at grade m is then defined as $\beta_{m+1} - \beta_m$.

I analyze the impact of neighborhood exposure by focusing on students who move between school districts with varying STEM rates using observational data. In this context, the error term θ_i in equation (1) is likely to be correlated with $\overline{STEM}_p$. When estimating equation (1) for students who have moved once, the regression coefficient is given by $b_m = \beta_m + \delta_m$ where δ_m represents the selection effect, defined as $\frac{cov(\theta_i, \bar{y}_d)}{var(\bar{y}_d)}$. This selection captures the degree to which factors influencing students' STEM choices for movers are correlated with the STEM choices of non-movers. However, the identification of exposure effects does not require that where people move is orthogonal to a student's future STEM choice. Instead, the key requirement is that the timing of the move to either high or low STEM areas should be unrelated to a student's STEM choices.

4 Baseline Estimates of Exposure Effects

In this section, I present the baseline estimates of exposure effects γ_m using students who moved across school districts once in their entire school years. First, I present a set of semi-parametric estimates of γ_m, which control for origin fixed effects and closely align with the hypothetical experiment described earlier. Following that, I present estimates from parametric models that show how movers' STEM choices can be parsimoniously modeled as a linear combination of

the STEM rates of non-moving students in origin and destination.[6]

First, consider the cohort of students who move at grade m from 1 to 12. I analyze the relationship between these students' STEM choices in college and those of non-moving students in their destination school district by using the following regression model:

$$STEM_i = \alpha_{osm} + \sum_{m=1}^{12} b_m I(m_i = m)\Delta STEM_{od} + \eta X_d + \varepsilon_{1i} \tag{2}$$

where $STEM_i$ denotes the students' STEM bachelor's degree attainment by age 26, α_{osm} is a fixed effect for the origin school district o by school cohort s by grade at move m, $I(m_i = m)$ is an indicator function that is 1 when $m_i = m$ and 0 otherwise, and $\Delta STEM_{od} = \overline{STEM}_d - \overline{STEM}_o$ is the difference in predicted STEM choice of non-moving students in the destination versus origin. X_d denotes a control vector including the BA rate of the destination. Equation (2) can be interpreted as a generalized observational analog of the specification in (1) that we would ideally estimate in experimental data.

Before presenting the estimation results, I first construct a non-parametric binned scatter plot for students who moved in grade 3. After demeaning both $STEM_i$ and $\Delta STEM_{od}$ within each origin (o) by birth cohort (s), I divide the residuals into twenty equal-sized groups (ventiles) and plot the mean value for each bin. Figure 3 presents that students who move to areas where non-moving students are more (less) likely to choose STEM degrees are more (less) likely to major in STEM degrees. The relationship between $STEM_i$ and $\Delta STEM_{od}$ appears linear. The estimated regression coefficient of $b = 0.7$ indicates that a 1 percentage point increase in $\Delta STEM_{od}$ is associated with a 0.7 percentage point increase in $STEM_i$.

Figure 4 (a) presents the estimates of b_m from equation (2). These estimates reveal important patterns. Notably, the values of b_m decrease consistently with the grade at which students move. Under the identification assumption, this declining trend provides evidence of an exposure effect, indicating that moving to a higher STEM area earlier in a student's school years results in a larger increase in probability of majoring in STEM. The linearity observed between b_m and the grade at move m suggests that the exposure effect, defined as $\gamma_m = b_m - b_{m+1}$, remains

[6] I show that the findings are robust to alternative cutoffs for enrollment size in Appendix Figure A.1, and present estimates that include students who move more than once in Appendix X.

approximately constant with respect to the grade at move m. By regressing $\hat{b}_m$ on m, I estimate an average annual exposure effect of $\gamma = 0.040$ (s.e. = 0.003). This implies that the outcomes of students who move converge toward those of non-movers in the destination area at a rate of 4.0% per year of exposure during the school years. Additionally, the estimates of b in the higher grades approach zero, indicating that most of the positive effects observed in moving in early grades can be attributed to exposure effects rather than selection effects.

Equation (2) includes over 30,000 fixed effects (α_{osm}), which makes estimation challenging in smaller samples. To overcome this difficulty, I estimate a model where I control for the STEM rates of the origin, rather than relying on origin fixed effects. This leads to the following regression:

$$STEM_i = \alpha + \sum_{m=1}^{12} b_m I(m_i = m)\Delta STEM_{od} + \theta_2 X_d + \sum_{m,s} I(m_i = m, s_i = s)(\alpha^1_{m,s} + \alpha^2_{m,s}\overline{STEM}_o) + \varepsilon_{2i}. \tag{3}$$

Figure 4 (b) presents the coefficients b_m from estimating equation (3). These coefficients closely resemble those shown in Figure 4 (a). By regressing the b_m coefficients on m, I estimate an average annual exposure effect of $\gamma = 0.041$ (s.e. = 0.002). This estimate is consistent with that obtained from the fixed effects specification, indicating that controlling for the STEM rates in the origin school district from non-moving students sufficiently accounts for differences between origin school districts. In other words, the STEM choices of movers can be effectively modeled as a weighted average of the STEM rates of non-moving students in both the origin and destination, with the number of school years spent in each location.

Furthermore, I estimate the same analysis using equations (2) and (3), dividing students based on their economic status and race in Appendix C. Although the exposure effect estimates using group-specific rates are noisier and slightly smaller at $\gamma = 0.038$, the overall exposure effects remain consistent. I also conduct a placebo test using students' math scores in grade 3 as the dependent variable considering it as a baseline performance measure. As shown in Appendix Figure A.2, I do not find any patterns in their test scores related to the grade in which students move.

Considering approximately constant exposure effects with respect to the grade at move m,

I parameterize the exposure effects shown in Figure 4 linearly, replacing the non-parametric $\sum_{m=1}^{12} b_m I(m_i = m)\Delta_{od}$ term in (3) with a line:

$$STEM_i = \alpha + \gamma m_i \Delta_{od} + \theta_1 \Delta_{od} + \theta_2 X_d + \sum_{m,s} I(m_i = m, s_i = s)(\alpha^1_{m,s} + \alpha^2_{m,s}\overline{STEM}_o) + \varepsilon_{3i}. \quad (4)$$

Estimating this specification yields an average annual exposure effect of $\gamma = 0.040$ (s.e. = 0.004), as reported in column (1) of Table 2. This estimate remains robust across different specifications and sample definitions. Column (2) presents the estimate obtained using semi-parametric specifications from equation (2). Column (3) includes additional controls in the parametric estimation from equation (4), incorporating average labor market outcomes for non-moving students in both the origin and destination school districts, such as yearly earnings and employment rates. Column (4) replicates the analysis from column (1) while controlling for individual-level 3rd grade math and reading scores of movers as baseline performance measures. Furthermore, I replicate the analysis at the county level in Appendix B. Although the exposure effect estimate at the county level is slightly larger at $\gamma = 0.048$, the overall exposure effects remain consistent.

The convergence rate of 4% per year of exposure between grades 1 and 12 implies that a student who moves at grade 0 would assimilate approximately (13 - 0) $\times$ 4% = 52% of the observed differences in outcomes between permanent residents of their origin and destination school districts.

4.1 Validation of Baseline Design

I evaluate the key identifying assumption—that the potential outcomes for students who move to higher versus lower STEM areas are not influenced by the grade at which they move—by conducting tests designed to address potential selection bias and omitted variable bias. First, I carry out a series of outcome-based placebo (or overidentification) tests of the exposure effect model. These tests utilize variations in STEM rates among non-moving students across different subgroups, allowing for precise, testable predictions about how students' major choices should change when they move to different areas. Additionally, I conduct a placebo test using alternative

exposure measures based on different college majors to further validate the findings. Lastly, I examine moves that are likely driven by displacement shocks, which lead many families to relocate, thereby minimizing omitted variable bias related to the decision to move.

To test for bias due to unobservable factors, I conduct placebo tests that leverage the heterogeneity in average STEM rates across different subgroups. Specifically, I examine variation across school cohorts, gender, and race/ethnicity. The causal exposure effect model predicts that a student's major choice should converge precisely to the average STEM rates of non-moving students within their own subgroup. In contrast, models based on omitted variables or selection bias would not produce such subgroup-specific convergence under reasonable assumptions about parents' information and preferences. This variation in STEM rates among non-moving students across subgroups provides a set of overidentifying restrictions. By using these restrictions, I can more effectively test whether neighborhoods truly have causal effects on students' STEM choices or if the observed patterns might be driven by unobserved factors.

I begin with variation in place effects across school cohorts. Although non-moving students' STEM choices are generally very stable over time, those in some areas have improved over time, while others have gotten worse. Such changes could occur, for instance, because of changes in the local schools or other area-level characteristics that affect students' STEM choices. I exploit this heterogeneity across school cohorts to test for confounds in the baseline research design. In the context of the causal exposure effect model, when a student moves to a new destination d, the change in STEM rates $\Delta STEM_{od,s(i)}$ within that student's own school cohort $s(i)$ should have a stronger influence on their STEM choice than the change in STEM rates $\Delta STEM_{ods}$ for other cohorts $s \neq s(i)$. The rationale is that a student's decision to pursue STEM is more likely to be influenced by the STEM rates within their own cohort in the neighborhood, rather than the rates for older or younger cohorts. Conversely, unobserved factors θ_i are unlikely to vary significantly across different school cohorts s in correlation with $\Delta STEM_{ods}$, since these cohort-specific variations emerge only later in college and are not apparent at the time of the move. Thus, by comparing the predictive power of STEM rates from a student's own cohort to those from surrounding cohorts, I can evaluate the potential bias introduced by unobserved

factors.[7]

To conduct this analysis, I estimate the baseline specification in (4) by substituting the change in STEM rates for non-moving students within the student's own cohort, $\Delta STEM_{od,s(i)}$, with corresponding STEM rates for neighboring school cohorts $s(i)+t$, $\Delta STEM_{od,s(i)+t}$ (see Appendix **??** for more details). The results, presented in Table 3 Panel A, show the exposure effect estimates (γ_t) from these regressions. The estimates for adjacent younger and older cohorts are similar to the baseline estimate of $\gamma = 0.040$, reflecting the strong correlation in STEM rates among non-moving students. In column (4), where predictions for both younger and older are included together, the coefficients $\tilde{\gamma}_t$ for placebo exposure effects (γ_t for $t \neq 0$) are smaller and statistically insignificant. Notably, the estimate for younger cohorts is near zero, indicating the limited impact on older cohorts' college major choices. However, the exposure effect estimate for the student's own cohort remains stable at 0.034. When combined, the influence of the student's own and older cohorts approximates $\gamma = 0.04$.

Next, I perform another set of placebo tests examining differences by student gender and race. I start by creating gender-specific and race-specific predictions for the mean STEM choice rate of non-moving students, denoted as $\overline{STEM}_p^D$. Here, the superscript D represents different groups: male (m) and female (f) for gender, and Black (b), Hispanic (h), and White (w) for race. Locations with high STEM rates for boys also tend to be high for girls, with a population-weighted correlation of 0.86 between $\overline{STEM}_p^m$ and $\overline{STEM}_p^f$ across districts. In the case of race, the correlation, 0.35, is not as strong as gender. I use the residual variation between genders or races to conduct placebo tests, assuming that unobserved shocks are unlikely to have gender or race-specific effects.[8]

In Table 3 Panel B, I estimate exposure effect models similar to equation (4), but with predictions divided by gender. Column (1) replicates Table 2 column (1) using gender-specific predictions $\Delta STEM_{od}^g$, rather than a combined prediction for both genders. This yields an exposure effect estimate of $\gamma = 0.032$ per year. In column (2), I use the prediction for the other gender, $\Delta STEM_{od}^{-g}$, which results in an estimate of 0.030, consistent with the high correlation

[7] In Appendix D.2, I provide a more formal presentation of the assumption underlying the overidentification test.

[8] When it comes to races, there might be sorting across racial groups within school districts. In that case, the placebo test using racial variation would be not as plausible as other placebo tests.

between outcomes for different genders. In column (3), I include predictions for both genders simultaneously, isolating the exposure effects based on differential variation across genders within school districts. Here, the coefficient for the own-gender prediction is $\gamma = 0.025$, which is 2.5 times larger than the coefficient for the other-gender prediction. In Panel C, similarly, I analyze race-specific exposure effects. The results reveal convergence within racial groups, with the coefficient for the own-race prediction at $\gamma = 0.028$, which is approximately 4-5 times larger than the coefficients for other racial group predictions.

I conduct a placebo test using change in non-moving students' major choice rates across different college majors in addition to STEM: business (ΔBIZ_{od}) and liberal arts ($\Delta LIBARTS_{od}$). I conduct this analysis by estimating the baseline specification in equation (4), adding those major-specific exposure measures to STEM exposure measure (see Appendix **??** for details). The estimates in Table 4 present the estimated exposure effects (γ) from these regressions. In column (1), the estimate of STEM exposure measure (γ) is similar to the baseline estimate of γ with other college major exposure measures. However, the estimates for exposure to business and liberal arts majors (γ_{BIZ} and $\gamma_{LIBARTS}$) are positive but not statistically significant. This suggests that increased exposure to other fields may result in a shift away from STEM. In columns (2) and (3), I find major-specific convergence with business and liberal arts majors. Additionally, Appendix Table A.1 examines two popular STEM majors (engineering and biological sciences), and the findings again indicate a pattern of major-specific convergence. This analysis also helps address concerns that the observed exposure effects might be influenced by overall educational attainment rather than being specifically driven by STEM exposure.

Lastly, focusing on moves where the underlying cause of the move is known, I account for unobservable shocks. For instance, consider identifying families forced to move from an origin o to a nearby destination d due to an exogenous shock, such as a natural disaster. These displacement shocks can lead to different changes in neighborhood STEM rates ($\Delta STEM_{od}$). If these shocks do not have direct effects on students that are correlated with $\Delta STEM_{od}$, the identification assumption holds, and I can obtain unbiased estimates of the exposure effect γ by focusing on displaced students. As presented in Appendix Figure A.15, I find consistent exposure effects even when focusing on moves potentially associated with significant displacement. These

findings support the conclusion that the baseline exposure effect estimates reflect the causal impact of neighborhood environments rather than unobserved factors associated with family decisions to move. This result implies that the exposure effects identified by the initial analysis can be generalized to a broader population beyond those families who voluntarily choose to move. See Appendix D.1 for more details.

The results in this section consistently demonstrate that the evidence for exposure effects remains strong across various refinements of the baseline design, suggesting that the observed effects are not driven by selection. This implies that any omitted variable θ_i that might bias my estimate of the exposure effect γ must meet the following specific criteria: (1) it must persist even in the presence of moves triggered by displacement shocks; (2) it must precisely mimic the outcomes of non-moving students by school cohort, gender, and race in proportion to exposure time; and (3) it must accurately replicate the specific major choices of non-moving students also in proportion to exposure time. Given the unlikelihood that a plausible omitted variable would satisfy all these conditions, I conclude that my estimate of $\gamma \simeq 0.04$ represents an unbiased estimate of the annual exposure effect.

5 Occupation Composition in the Neighborhood

In this section, I explore reasons for STEM exposure effects. A neighborhood is defined in this paper to encompass various aspects of a place. My measure for exposure effects–STEM rate of non-movers–summarizes all those aspects because this is the outcome of all the factors influencing STEM choices. A neighborhood can be broadly divided into school and non-school environments, which interact with each other. While the impact of school environments, such as curriculum and teachers, is well-documented (Beilock et al. 2010; Clotfelter, Ladd, and Vigdor 2005; Lavy and Sand 2015; Legewie and DiPrete 2014; Wang 2013), non-school environments have received less attention. To explore this, I use the occupational composition of a neighborhood as a proxy for the non-school environment and examine its influence on students' choice of college majors.

I begin by plotting the enrollment-weighted STEM occupation exposure measure ($STEMOcc_p$) across school districts in Figure 5. Importantly, the distribution of the STEM occupation rate

closely mirrors the baseline STEM exposure measure, which is based on the STEM rates of non-moving students. These two measures exhibit a correlation of 0.78. Even when the rate is calculated among BA holders (i.e., STEM/all BAs), the correlation remains relatively strong at 0.45. Moreover, the relationship appears linear as presented in Appendix Figure A.4. As STEM occupation rate increases by 1 percentage point, the baseline STEM exposure measure increases by 0.4. In contrast, I do not observe a similar relationship for other majors. For example, the correlation between liberal arts rates and STEM occupation rates is 0.6 for all students but drops to -0.1 among BA holders. This suggests that non-educational environments play a significant role in fostering future generations of STEM professionals.

To assess whether neighborhood occupational composition has a causal impact on STEM choice, I estimate equation (3), substituting the difference in baseline STEM exposure between origin and destination ($\Delta STEM_{od}$) with the difference in STEM occupational exposure ($\Delta STEMOcc_{od} = STEMOcc_d - STEMOcc_o$). As shown in Figure 6 (a), the estimates of b_m decline steadily with the grade at the time of the move. Since the exposure effects appear linear, I parameterize them linearly, following equation (4) with STEM occupation exposure measure. The results in Table A.2 present that the exposure effects are robust across different specifications. Specifically, for each additional year a student spends in a neighborhood with a 1 percent higher STEM occupational proportion, the likelihood of choosing a STEM major increases by 0.026 percent. It implies that a child moving in kindergarten from the 25th percentile (0.0463) to the 75th percentile (0.0945) in STEM rates—a 1.28 standard deviation difference—would see an increase in the probability of majoring in STEM by 0.0482*2.6*13=1.63 percentage points. This represents 39% of the average STEM major choice rate (4.18%).

Using the industry individuals work in after age 24, I further estimate whether neighborhood influence extends to their occupational choices. In Figure **??** (b), I replicate Figure 6 (a), but replace the outcome variable with a STEM industry measure—the proportion of STEM workers within 4-digit NAICS industry codes. The estimates of b_m steadily decline with the grade at which the move occurred. Similar results are found using the baseline neighborhood exposure measure, as shown in Appendix Figure A.5. Since the exposure effects appear linear, I model them parametrically following equation (4), using the STEM occupation exposure measure. The

results in Table A.3 show that the exposure effects are robust across different specifications. Although individual occupations are not available in my data, estimating the proportion of STEM occupations within industries suggests that neighborhood influences do indeed extend to occupational choices.

I conduct analogous overidentification tests using the gender of workers and other occupational compositions as examined with the baseline exposure measure. First, I test for overidentification across genders by replicating Table 3 Panel B, separately analyzing the STEM worker composition among male and female populations. As shown in Table A.5, the coefficient for the own-gender occupation is $\gamma = 0.011$, which is 2.75 times larger than the coefficient for the other-gender occupational composition. These results highlight that the exposure effects are gender-specific, emphasizing the importance of exposure to female STEM workers in increasing STEM participation among female students. Furthermore, I perform the overidentification test using another occupational measure—managerial positions—replicating Table 4. As presented in Table A.6, I find that students' STEM choice is influenced only by the STEM occupational exposure measure, while the managerial worker composition shows an insignificant positive effect. This suggests that exposure to other occupations might deter students from choosing STEM majors. In column (2), a similar pattern emerges, where the choice of a business major is more responsive to managerial occupation exposure effects. These overidentification tests alleviate concerns that the estimated occupational exposure effects are driven by neighborhood selection based on educational environments by parents. If neighborhood selection because of educational environments were responsible, we would expect similar estimates from different occupational exposure effects, given that it is unlikely only parents with STEM occupations would sort into specific neighborhoods.

The analyses in this section highlight the significant influence of non-educational environments on the choice of STEM majors. Given that it is unlikely that parents in STEM occupations selectively move to certain areas because of educational environments based on the results of the overidentification test, the strong correlation between high STEM rates among non-moving students and high STEM occupation rates in the neighborhoods is more likely driven by the prevalence of STEM occupations, rather than the reverse. Additionally, high-STEM areas tend

to offer more robust STEM education environments, such as greater availability of science courses (including AP science courses; see Appendix Figure A.6). This suggests that high STEM occupation rates may naturally contribute to creating educational environments that are more conducive to pursuing STEM fields.

6 Analyzing Heterogeneity Across Economic Status, Genders, and Races

An important question that remains is whether high-STEM neighborhoods also encourage underrepresented groups to major in STEM. For example, are students from economically disadvantaged families more likely to dream of being engineers if they live near a NASA space station, even if their parents have no connection to NASA? Or are students from disadvantaged backgrounds more likely to pursue computer science degree if they live in a neighborhood with a lot of computer scientists? This question is particularly important to understand the gaps across demographic and socioeconomic status in STEM choices. To investigate this, I examine the heterogeneity in exposure effects across economic disadvantage status, genders, and races.

Focusing first on economic disadvantage, I estimate equation (3) separately for economically disadvantaged and non-disadvantaged students. Students are classified as economically disadvantaged if they receive free or reduced-price lunch in grade 0 (see Appendix Figure A.7 for results using alternative definitions of economic disadvantage). Figure 7 presents the coefficients b_m for both groups, showing a steady decline with the grade at which the move occurs. By regressing $\hat{b}_m$ on m, I estimate an average annual exposure effect of $\gamma = 0.041$ for non-disadvantaged students and 0.32 for disadvantaged students. This suggests that the benefits of living in high-STEM areas extend similarly to students from both disadvantaged and non-disadvantaged backgrounds. However, the overall likelihood of choosing STEM majors differs between the groups. Non-disadvantaged students consistently show a larger positive level across all school years, with a gap of approximately 0.4. This suggests that while the exposure effects are similar for both groups, non-disadvantaged families may be more deliberate in moving to high-STEM areas, as factors influencing STEM choice are more strongly correlated with the STEM rates in their new location.

Next, I examine the heterogeneity across gender and race. Building on the findings related to

gender and race-specific convergence, I use the exposure rates of each demographic group to explore potential heterogeneity in the exposure effects. I estimate models similar to equation (3), but separately for each gender or race, using predictions based on their respective demographic group. As shown in Figure 8 (a) and (b), the exposure effects are similar across different genders and races. By regressing $\hat{b}_m$ on m, I estimate the average annual exposure effect (γ) for each group: -0.034 for male students, -0.042 for female students, -0.030 for White students, and -0.026 for Black and Hispanic students. These results indicate that the positive exposure effects are consistent across different economic statuses and demographic groups. Appendix Table A.7 presents linearly parameterized estimation results following equation (4) and finds similar exposure effects across groups. This highlights that the benefits of living in research hubs or areas with a concentration of STEM-related industries extend beyond immediate economic impacts on adults. They also influence the next generation's choice of field of study and potentially their future careers.

While the exposure effects are similar across groups, the level of exposure varies due to differences in the neighborhoods where students of different economic statuses and races reside. To explore the impact of these varying levels of exposure, I conduct a counterfactual experiment: how much would the gap in STEM choice be reduced if students from different backgrounds lived in the same neighborhoods? If I define a school district as a neighborhood, White students on average experience 30% higher neighborhood STEM exposure compared to Black and Hispanic students (see Appendix E for more details). If Black and Hispanic students had the same level of neighborhood exposure as White students over their school years, their probability of majoring in STEM would increase by 0.68 percentage points each, closing 21% and 31% of the STEM choice gap, respectively. Similarly, non-disadvantaged students have an average neighborhood exposure rate that is 41% higher than that of disadvantaged students. If disadvantaged students experienced the same exposure level, their likelihood of choosing STEM majors would increase by 0.88 percentage points, narrowing 23% of the gap.

Since school districts encompass larger areas than the actual neighborhoods students interact with daily, these results should be considered a lower bound. Conducting a similar counterfactual analysis using exposure levels calculated at the school level, I find that the racial gap in STEM

choice could decrease by 34% between Black and White students and by 44% between Hispanic and White students. The gap based on economic status could also decrease by 34%. Additionally, using occupational exposure measures in the same counterfactual, I find that similar percentages of the gap would be reduced if exposure levels were equal across groups (see Appendix E for more details).

Moreover, although male and female students live in the same neighborhoods, the proportion of each pursuing STEM is different, which is critical given the findings on own-gender convergence. To explore this, I conduct a counterfactual experiment: if both male and female students were equally influenced by both genders (i.e., there is no gender-specific convergence), how much of the gender gap in STEM choice would be reduced? In other words, if female students had the same exposure levels as male students during their school years, how much would the gender gap decrease? Using the difference in estimates between genders from Table 3 Panel B, I find that women's STEM choice would increase by 0.43 percentage points, closing 27% of the gap. Similarly, when using the difference in STEM occupational exposure between male and female workers, I find a comparable reduction in the gap—23%.

7 Other Outcomes

In this section, I estimate neighborhood effects on college majors beyond STEM, and I also analyze STEM enrollment or obtaining a STEM graduate degree instead of the baseline measure of STEM bachelor's degree completion. The majors I analyze in this section cover more than 70% of four-year college graduates.

Appendix Figure A.9 presents the estimation results from equation (3), replacing $\Delta STEM_{od}$ with $\Delta major_{od} = \overline{major}_d - \overline{major}_o$, where $\overline{major}_p$ is the fraction of students obtaining a bachelor's degree in the specific major among non-moving students. I find that the estimates of b_m decline steadily with the grade at move m for all college majors. By regressing $\hat{b}_m$ on m, I estimate the average annual exposure effect γ_{major} to range from -0.027 for social sciences to -0.59 for biological sciences.

Appendix Figure A.10 shows the estimation results from equation (3) using a different STEM definition. Here, STEM choice is defined as enrolling in a STEM major at any type of college

in Figure (a) and obtaining a graduate degree in a STEM major in Figure (b). STEM exposure changes ($\Delta STEM_{od}$) after moving are also defined accordingly. Once again, I find that the estimates of b_m decline steadily with the grade at move m. By regressing $\hat{b}_m$ on m, I estimate the average annual exposure effect to be $\gamma_{enroll} = -0.035$ for enrollment and $\gamma_{grad} = -0.034$ for a graduate degree which are similar to baseline estimate $\gamma = -0.041$.

This analysis generalizes the findings from STEM majors to other college majors and different levels of educational attainment. The insights from this paper suggest that the influence of neighborhood effects extends beyond STEM majors to all college majors and impacts not just bachelor's degree holders but also other levels of post-secondary education. Thus, my research highlights that students' choices of study are shaped not only by their preferences or abilities but also by their early experiences and exposures. This influence plays a significant role in their overall decisions regarding post-secondary education.

8 Conclusion

This paper has found causal impacts of neighborhoods on students' college STEM choices comparing students who move at different grades across school districts and counties. The neighborhood effect estimates imply that moving in kindergarten from an area with the 25th percentile to the 75th percentile of STEM rates increases the probability of majoring in STEM by 1.4 percentage points, which is 35% of the mean. I further found that the occupational composition of the area is a key driver of these neighborhood effects. Importantly, the exposure effects are not limited to majority groups in STEM fields; they are consistent across genders, races, and economic statuses, with similar magnitudes. Counterfactual analysis based on the estimates suggests that 20-40% of the gaps across these groups could be reduced if they had the same level of STEM exposure.

An important implication of this paper is the spillover effects of research hubs and agglomeration economies. Previous studies have analyzed the direct effects of agglomerations and local business booms (Allcott and Keniston 2018; Black, McKinnish, and Sanders 2005; Cust and Poelhekke 2015; Feyrer, Mansur, and Sacerdote 2017; Greenstone, Hornbeck, and Moretti 2010). While these hubs and economies generate numerous job opportunities for adults within

neighborhoods, an additional potential benefit lies in their spillover impact on the human capital development of the next generation. These spillover effects have the potential to transcend income and socioeconomic disparities, ultimately benefiting children from disadvantaged households. Since this research establishes causal links between the placement of research hubs and agglomeration economies and student STEM choices, the strategic selection of these locations should become a more prominent policy agenda.

References

Allcott, Hunt, and Daniel Keniston. 2018. "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America." *The Review of Economic Studies* 85, no. 2 (April): 695–731. https://doi.org/10.1093/restud/rdx042.

Archer, Louise, Jennifer DeWitt, Jonathan Osborne, Justin Dillon, Beatrice Willis, and Billy Wong. 2012. "Science Aspirations, Capital, and Family Habitus: How Families Shape Children's Engagement and Identification with Science." *American educational research journal* 49 (5): 881–908.

Arcidiacono, Peter, Esteban M Aucejo, and V Joseph Hotz. 2016. "University Differences in the Graduation of Minorities in Stem Fields: Evidence from California." *American Economic Review* 106 (3): 525–562.

Arold, Benjamin W. 2024. "Evolution Vs. Creationism in the Classroom: The Lasting Effects of Science Education*." *The Quarterly Journal of Economics* (June).

Beilock, Sian L, Elizabeth A Gunderson, Gerardo Ramirez, and Susan C Levine. 2010. "Female Teachers' Math Anxiety Affects Girls' Math Achievement." *Proceedings of the National Academy of Sciences* 107 (5): 1860–1863.

Black, Dan, Terra McKinnish, and Seth Sanders. 2005. "The Economic Impact of the Coal Boom and Bust." *The Economic Journal* 115 (503): 449–476.

Blau, Francine D, and Lawrence M Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of economic literature* 55 (3): 789–865.

Bleemer, Zachary, and Aashish Mehta. 2021. "College Major Restrictions and Student Stratification." *Available at SSRN 3981921*.

Card, David, and A Abigail Payne. 2017. *High School Choices and the Gender Gap in Stem.* Technical report. National Bureau of Economic Research.

Carnevale, Anthony P, Nicole Smith, and Michelle Melton. 2011. "Stem: Science Technology Engineering Mathematics." *Georgetown University Center on Education and the Workforce.*

Carrell, Scott E, Marianne E Page, and James E West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *The Quarterly journal of economics* 125 (3): 1101–1144.

Castleman, Benjamin L., Bridget Terry Long, and Zachary Mabel. 2018. "Can Financial Aid Help to Address the Growing Need for STEM Education? The Effects of Need-Based Grants on the Completion of Science, Technology, Engineering, and Math Courses and Degrees." *Journal of Policy Analysis and Management* 37 (1): 136–166. https://doi.org/10.1002/pam.22039.

Catsambis, Sophia, and Andrew A Beveridge. 2001. "Does Neighborhood Matter? Family, Neighborhood, and School Influences on Eighth-Grade Mathematics Achievement." *Sociological focus* 34 (4): 435–457.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *The Quarterly Journal of Economics* 133 (3): 1107–1162.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

Chyn, Eric. 2018. "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children." *American Economic Review* 108 (10): 3028–3056.

Clotfelter, Charles T, Helen F Ladd, and Jacob Vigdor. 2005. "Who Teaches Whom? Race and the Distribution of Novice Teachers." *Economics of Education review* 24 (4): 377–392.

Cust, James, and Steven Poelhekke. 2015. "The Local Economic Impacts of Natural Resource Extraction." *Annual Review of Resource Economics* 7, no. Volume 7, 2015 (October): 251–268. https://doi.org/10.1146/annurev-resource-100814-125106.

Ellison, Glenn, and Ashley Swanson. 2010. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *Journal of Economic Perspectives* 24 (2): 109–128.

Feyrer, James, Erin T. Mansur, and Bruce Sacerdote. 2017. "Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution." *American Economic Review* 107, no. 4 (April): 1313–1334. https://doi.org/10.1257/aer.20151326.

Foote, Andrew, and Kevin M Stange. 2022. "Attrition from Administrative Data: Problems and Solutions with an Application to Postsecondary Education. Working Paper 30232." *National Bureau of Economic Research.*

Fry, Richard, Brian Kennedy, and Cary Funk. 2021. "Stem Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity." *Pew Research Center,* 1–28.

Greenman, Emily, Katerina Bodovski, and Katherine Reed. 2011. "Neighborhood Characteristics, Parental Practices and Children's Math Achievement in Elementary School." *Social science research* 40 (5): 1434–1444.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of political economy* 118 (3): 536–598.

Hanushek, Eric A, Babs Jacobs, Guido Schwerdt, Rolf Van der Velden, Stan Vermeulen, and Simon Wiederhold. 2023. *Where Do Stem Graduates Stem from? the Intergenerational Transmission of Comparative Skill Advantages.* Technical report. National Bureau of Economic Research.

Hedges, Larry V, and Amy Nowell. 1995. "Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals." *Science* 269 (5220): 41–45.

Holden, J, and Eric Lander. 2012. "Report to the President: Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics." *Executive Office of the President President's Council of Advisors on Science and technology.*

Landaud, Fanny, Son Thierry Ly, and Éric Maurin. 2020. "Competitive Schools and the Gender Gap in the Choice of Field of Study." *Journal of Human Resources* 55, no. 1 (January): 278–308. https://doi.org/10.3368/jhr.55.2.0617.8864R.

Langdon, David, George McKittrick, David Beede, Beethika Khan, and Mark Doms. 2011. "Stem: Good Jobs Now and for the Future. Esa Issue Brief# 03-11." *US Department of Commerce.*

Lavy, Victor, and Edith Sand. 2015. *On the Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases.* Technical report. National Bureau of Economic Research.

Legewie, Joscha, and Thomas A DiPrete. 2014. "The High School Environment and the Gender Gap in Science and Engineering." *Sociology of education* 87 (4): 259–280.

Maltese, Adam V, and Robert H Tai. 2011. "Pipeline Persistence: Examining the Association of Educational Experiences with Earned Degrees in Stem Among Us Students." *Science education* 95 (5): 877–907.

Miller, Jon D, and Linda G Kimmel. 2012. "Pathways to a Stemm Profession." *Peabody Journal of Education* 87 (1): 26–45.

Mulligan, Gail M, Sarah Hastedt, and Jill Carlivati McCarroll. 2012. "First-Time Kindergartners in 2010-11: First Findings from the Kindergarten Rounds of the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K: 2011). Nces 2012-049." *National Center for Education Statistics.*

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson. 2022. "The Gift of Moving: Intergenerational Consequences of a Mobility Shock." *The Review of Economic Studies* 89 (3): 1557–1592.

National Science and Technology Council. 2000. *Ensuring a Strong Us Scientific, Technical, and Engineering Workforce in the 21st Century.* Executive Office of the President.

National Science Board. 2003. *The Science and Engineering Workforce: Realizing America's Potential.* National Science Foundation.

———. 2024. *The State of U.s. Science and Engineering 2024.* National Science Foundation.

Oliver, Daniel, Robert Fairlie, Glenn Millhauser, and Randa Roland. 2021. "Minority Student and Teaching Assistant Interactions in Stem." *Economics of education review* 83:102125.

Papageorge, Nicholas W., Seth Gershenson, and Kyung Min Kang. 2020. "Teacher Expectations Matter." *The Review of Economics and Statistics* 102, no. 2 (May): 234–251. https://doi.org/10.1162/rest_a_00838.

Philippis, Marta De. 2023. "STEM Graduates and Secondary School Curriculum: Does Early Exposure to Science Matter?" *Journal of Human Resources* 58, no. 6 (November): 1914–1947. https://doi.org/10.3368/jhr.1219-10624R1.

Rozek, Christopher S, Gerardo Ramirez, Rachel D Fine, and Sian L Beilock. 2019. "Reducing Socioeconomic Disparities in the Stem Pipeline Through Student Emotion Regulation." *Proceedings of the National Academy of Sciences* 116 (5): 1553–1558.

Sampson, Robert J, Jeffrey D Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing "neighborhood Effects": Social Processes and New Directions in Research." *Annual review of sociology* 28 (1): 443–478.

Sharkey, Patrick, and Jacob W Faber. 2014. "Where, When, Why, and for Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects." *Annual review of sociology* 40:559–579.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2014. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *The Review of Economic Studies* 81, no. 1 (January): 426–472. https://doi.org/10.1093/restud/rdt025.

Tai, Robert H, Christine Qi Liu, Adam V Maltese, and Xitao Fan. 2006. "Planning Early for Careers in Science." *Science* 312 (5777): 1143–1144.

Wang, Xueli. 2013. "Why Students Choose Stem Majors: Motivation, High School Learning, and Postsecondary Context of Support." *American Educational Research Journal* 50 (5): 1081–1121.

Wodtke, Geoffrey T, David J Harding, and Felix Elwert. 2011. "Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation." *American sociological review* 76 (5): 713–736.

Xie, Yu, Michael Fang, and Kimberlee Shauman. 2015. "Stem Education." *Annual review of sociology* 41:331–357.

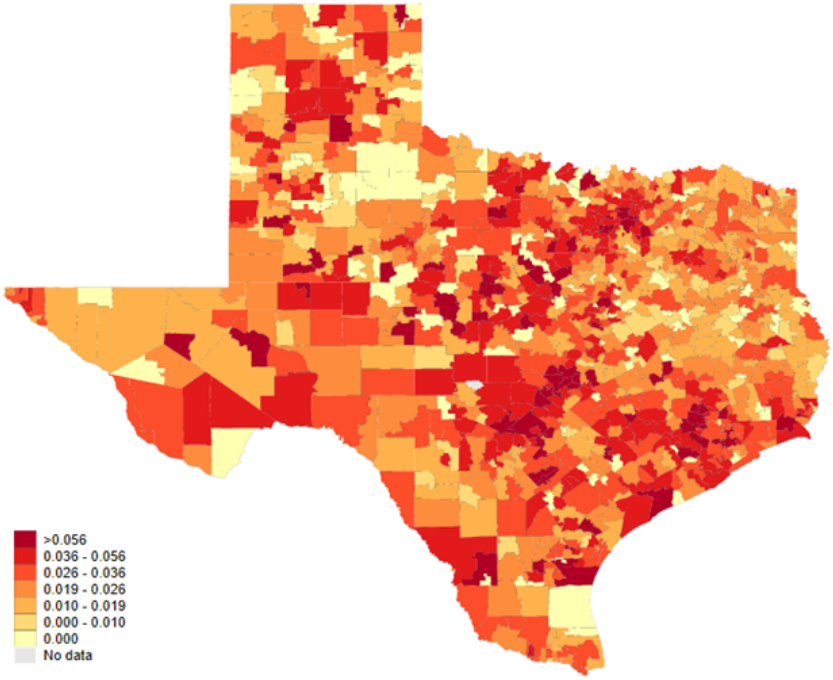
Xue, Yi, and Richard C Larson. 2015. "Stem Crisis or Stem Surplus? Yes and Yes." *Monthly labor review* 2015.

Zaza, Sam, Kristie Abston, Murat Arik, Patrick Geho, and Victor Sanchez. 2020. "What Ceos Have to Say: Insights on the Stem Workforce." *American Business Review* 23 (1): 136–155.

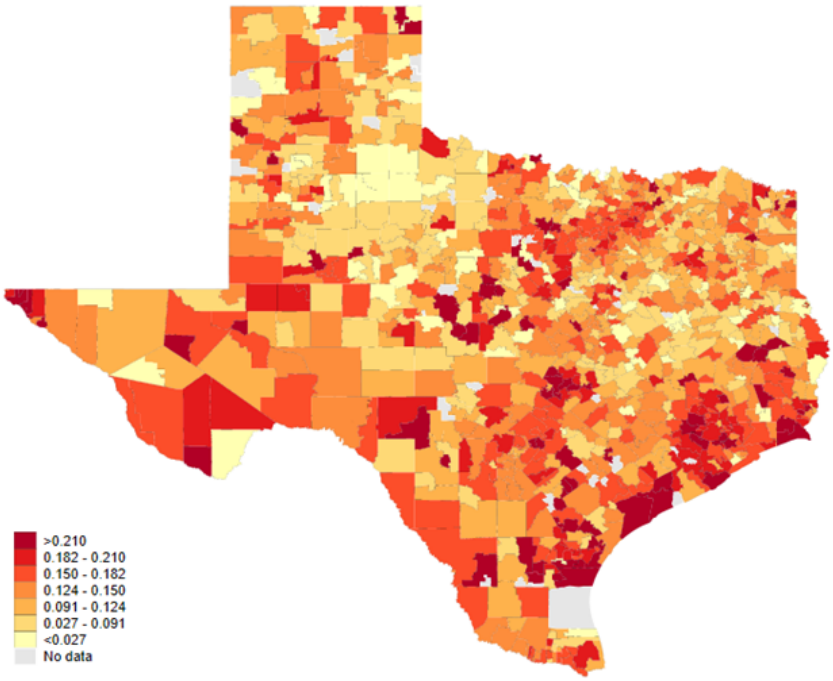
9 Figures and Tables

Fig. 1. STEM Rate Across School District

(a) STEM rate ($\frac{\text{STEM BAs}}{\text{All students}}$)



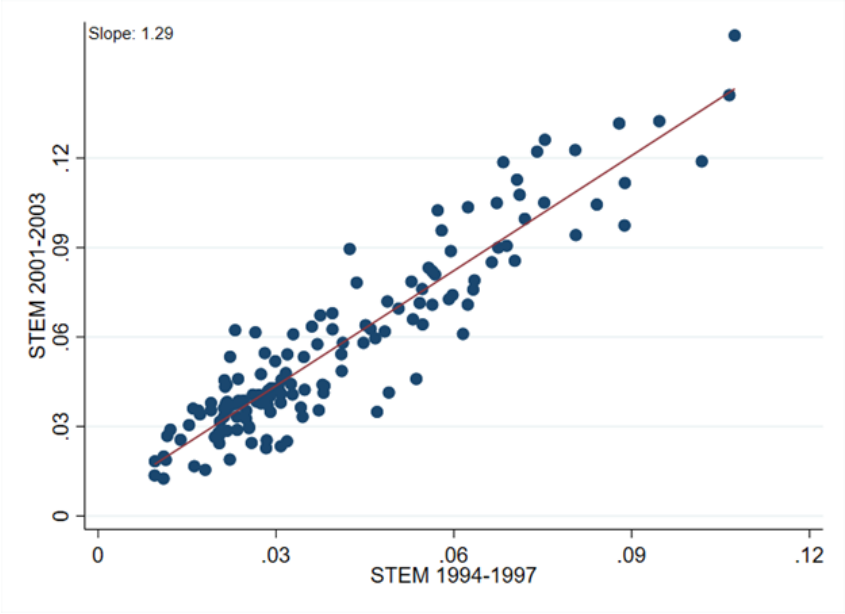
(b) STEM rate among the BAs ($\frac{\text{STEM BAs}}{\text{All BAs}}$)



Notes: These maps plot the proportion of non-moving students obtaining STEM bachelor's degrees by age 26 among all students in Figure (a) and among bachelor's degree holders in Figure (b). STEM majors are defined following the Department of Homeland Security 2016 STEM major list. The maps are constructed by grouping school districts into seven quantiles and shading the areas so that darker colors correspond to higher outcomes for students. Areas with no data are shaded with a grey color. The sample includes all students in the kindergarten in 1994-2003 who did not move across school districts in their school years.

Fig. 2. STEM Rate is Highly Persistent over Time Across School District

(a) STEM rate ($\frac{\text{STEM BAs}}{\text{All students}}$)

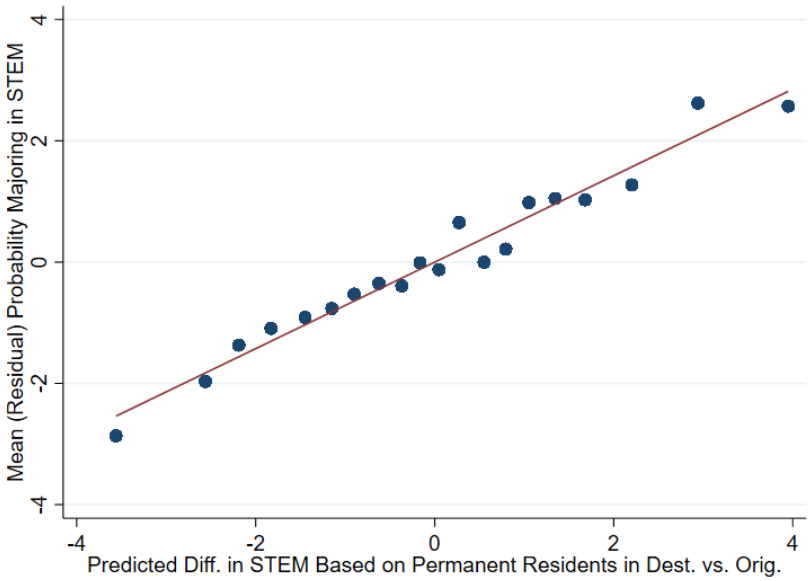


(b) STEM rate among the BAs ($\frac{\text{STEM BAs}}{\text{All BAs}}$)



Notes: The figures plot the proportion of non-moving students obtaining a bachelor's degree in STEM majors in Texas school districts from 1994-1997 school cohorts starting kindergarten in that corresponding year compared to 2001-2003 school cohorts. The proportion in Figure (a) is calculated among all students in each school district and (b) is among all bachelor's degree holders. Small school districts are excluded. STEM majors are defined according to the Department of Homeland Security's 2016 list.

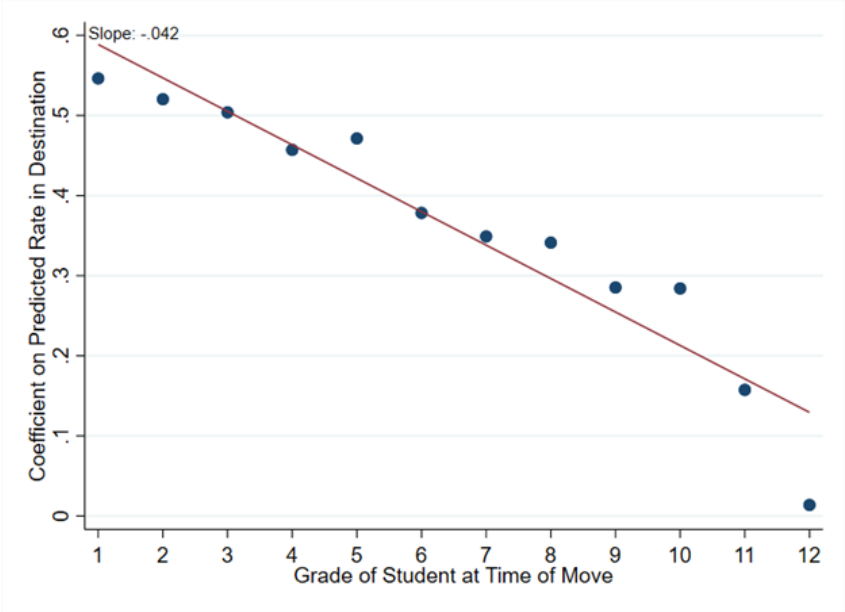
Fig. 3. Mover's Outcomes Versus Predicted Outcomes Based on Permanent Residents in Destination



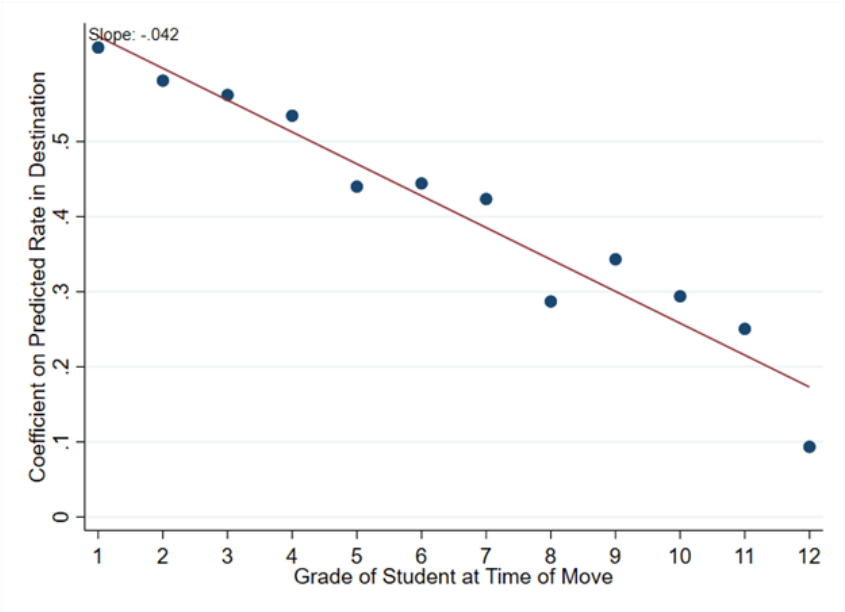
Notes: This figure presents a binned scatter plot depicting the relationship between the STEM probability of students who moved to a different school district at grade 3 and the differences in the average probability of majoring in STEM of non-moving students in the destination vs. origin school districts. The sample includes all students in kindergarten in 1994-2003 who moved when the student was in grade 3 and moved only once in school years. Students' STEM choices $STEM_i$ are measured as obtaining a STEM degree by age 26. $\overline{STEM}_p$ is the average STEM choice of non-moving students in school district p. To construct the figure, I first demean both $STEM_i$ and $\Delta_{od} = \overline{STEM}_d$ - $\overline{STEM}_o$ by origin (o) by birth cohort (s) cells in the sample of movers at grade m = 3 to construct residuals: $STEM_i^r = STEM_i$ - $E[STEM_i|o,s]$ and $\Delta_{od}^r = \Delta_{od}$ - $E[\Delta_{od}|o,s]$. I then divide the Δ_{od}^r into twenty equal-size groups (ventiles) and plot the mean value of $STEM_i^r$ vs. the mean value of Δ_{od}^r in each bin.

Fig. 4. Exposure Effects on STEM Choice in College

(a) Semi-parametric Estimates

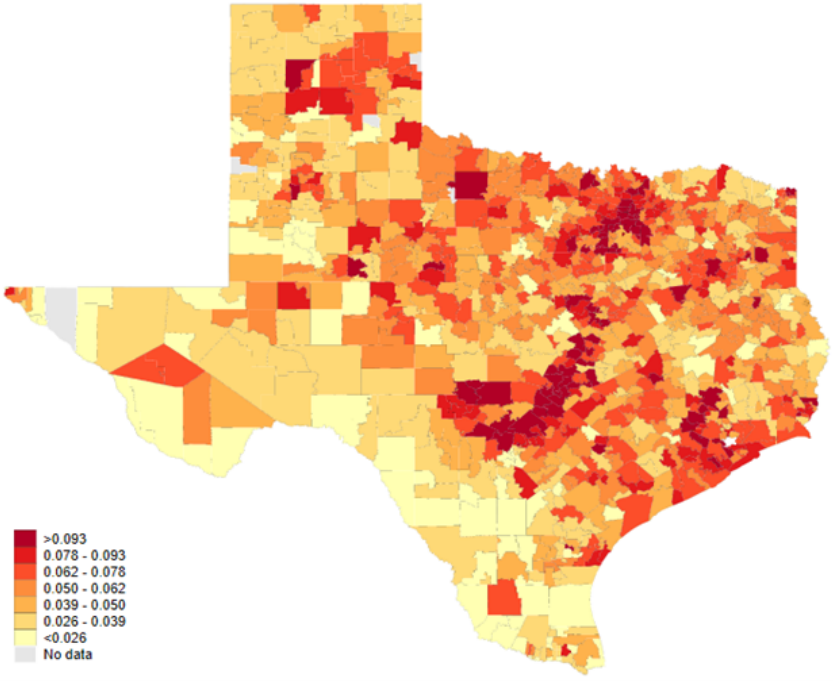


(b) Parametric Estimates



Notes: Figure (a) plots estimates of the coefficients b_m across the students' grade when they move (m) using the semi-parametric specification in equation (2), measuring students' STEM degree attainment. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' STEM choice $STEM_i$ is measured as obtaining a bachelor's degree in STEM majors by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at age m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta_{od} = \overline{STEM}_d - \overline{STEM}_o$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include origin school district-by-school cohort-by-grade at move fixed effects when estimating this specification. Figure (b) plots estimates from the parametric specification in equation (3). This specification replicates the specification used in Figure (a), replacing the origin fixed effects with the STEM rate for non-movers in the origin. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line are reported on the left side of each panel. The magnitudes of the slope represent estimates of annual exposure effects.

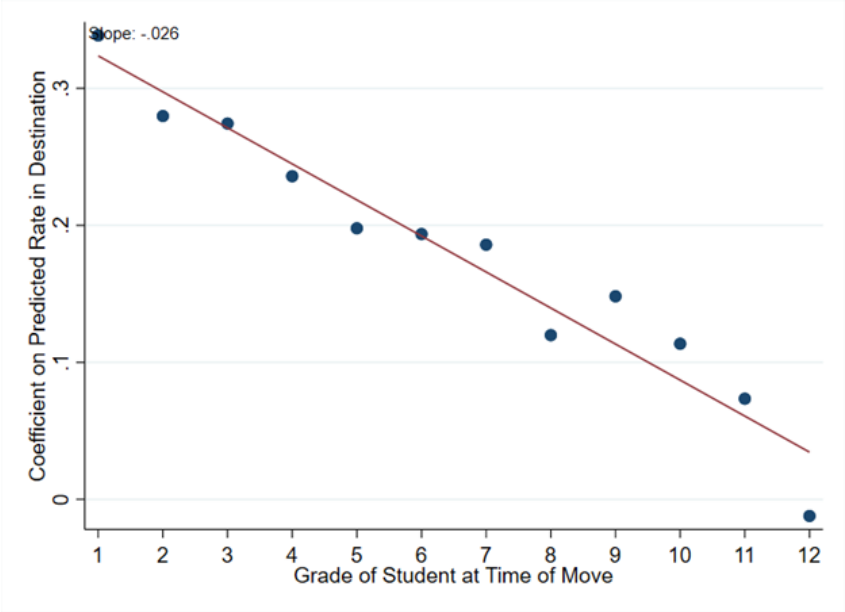
Fig. 5. STEM Occupation Exposure Measure



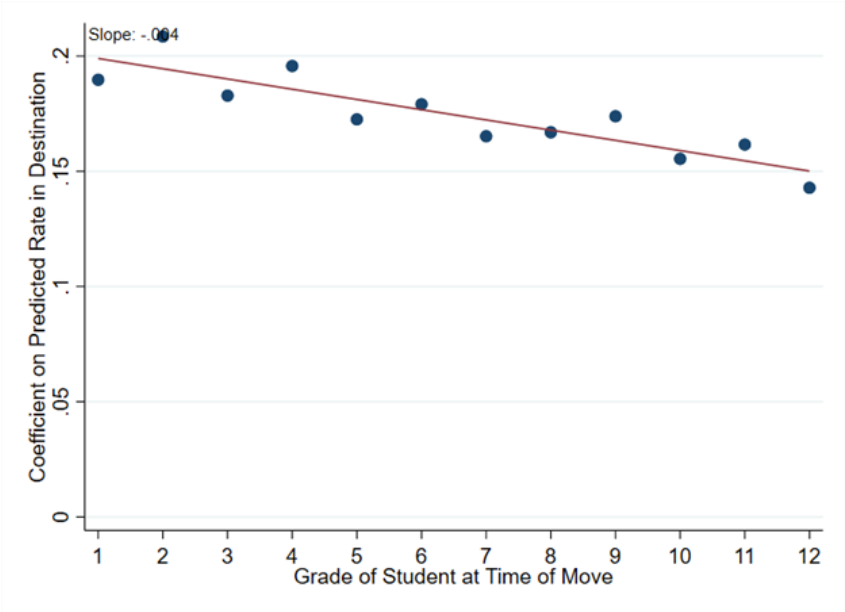
Notes: The map plots the proportion of STEM workers among the population in each school district. To construct a school district-level STEM occupation measure, I first construct a measure of STEM occupation exposure for each Census tract using occupational data from the 2000 Census at the Census tract level. Based on the list of STEM occupations from ONET, I calculate the STEM proportion in each Census tract as the ratio of STEM workers to the population 18-54. The STEM occupation exposure measure for each school district is then defined as the average STEM proportion of the Census tracts where the schools are located, along with nearby tracts, weighted by the number of students enrolled in each school. See Section 2 for more details. The map is constructed by grouping school districts into seven quantiles and shading the areas so that darker colors correspond to higher proportion of STEM workers. Areas with no data are shaded with a grey color.

Fig. 6. Exposure Effects on STEM Choice: STEM Occupation

(a) Occupation Exposure on STEM Major Choice

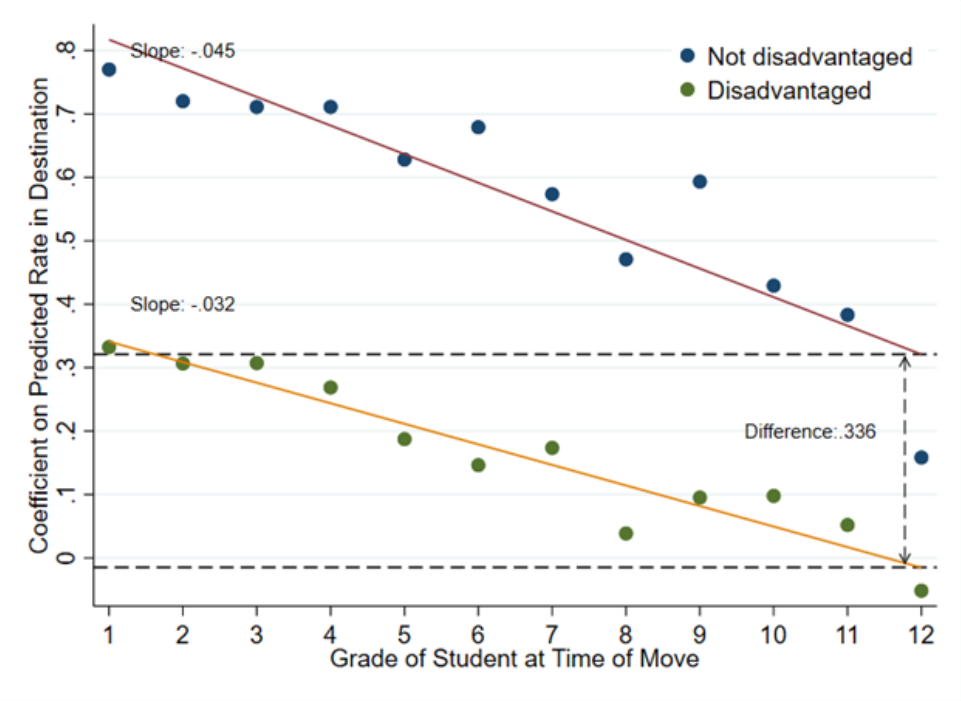


(b) Occupation Exposure on STEM Industry Choice



Notes: Figure (a) plots estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), replacing $\Delta STEM_{od}$ to $\Delta STEMOcc_{od} = \overline{STEMOcc_d}$ - $\overline{STEMOcc_o}$, the difference in STEM worker proportions between destination and the origin. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' STEM choices $STEM_i$ are measured as obtaining a STEM bachelor's degree by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where the neighborhood STEM occupation rate increases from 0 to 1 at grade m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta STEMOcc_{od}$, the difference between STEM worker proportion in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with the STEM occupation rate in the origin interacted with school cohort fixed effects. Figure (b) plots estimates of Figure (a) replacing outcome variable to STEM industry measure—proportion of STEM workers within 4 digit NAICS industry. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line are reported on the left side of each panel. The magnitude of the slope represents estimates of annual exposure effects to occupational composition.

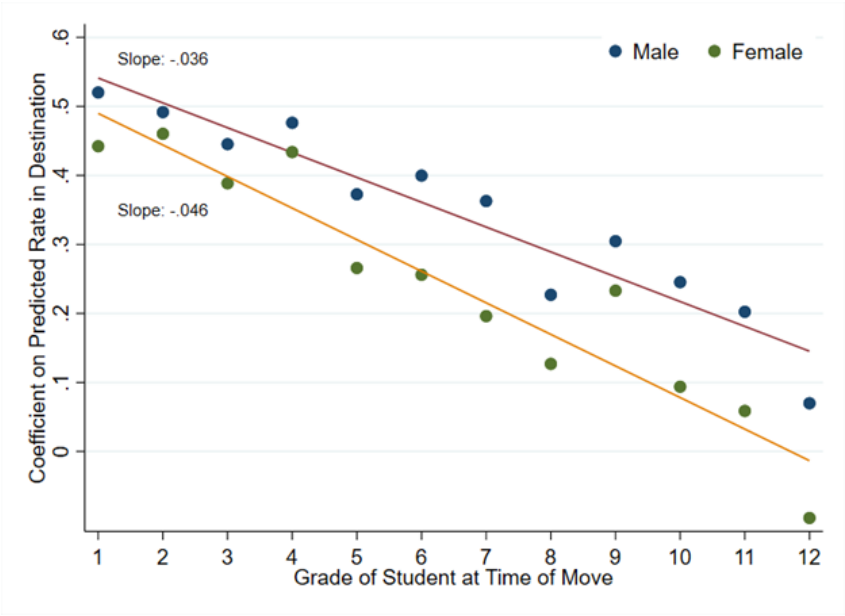
Fig. 7. Heterogeneous Exposure Effects by Economic Disadvantaged Status



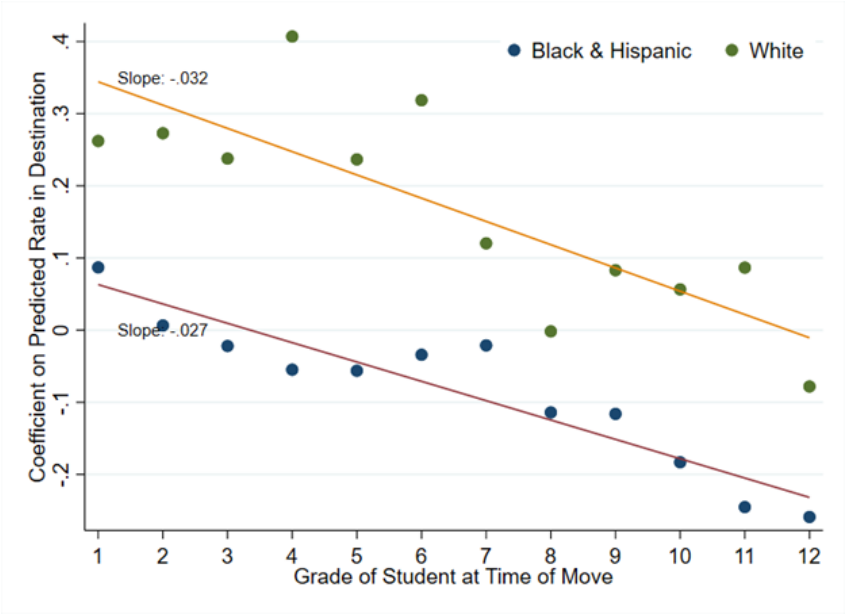
Notes: Figure plots estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), separately by the economically disadvantaged status of students. Students are defined as economically disadvantaged if they receive free or reduced-price lunch in kindergarten. The blue dots and the red line are from the sample of not economically disadvantaged students and the green dots and the orange line are from the sample of disadvantaged students. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' STEM choices $STEM_i$ are measured as obtaining a STEM bachelor's degree by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at grade m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta STEM_{od} = \overline{STEM}_d - \overline{STEM}_o$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with predicted STEM rate for non-moving students in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line is reported on the left side of each line.

Fig. 8. Heterogeneous Exposure Effects by Gender and Race

(a) Gender



(b) Race/Ethnicity



Notes: Figures plot estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), separately by gender or race/ethnicity of students. The blue dots and the red line are from the sample of male students for Figure (a) and White students for Figure (b) and the green dots and the orange line are from the sample of female students for (a) and Black/Hispanic students for (b). The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' STEM choices $STEM_i$ are measured as obtaining a STEM bachelor's degree by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at grade m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta STEM_{od} = \overline{STEM}_d$ - $\overline{STEM}_o$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with predicted STEM rate for non-moving students in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slope of the regression line is reported on the left side of each line.

Table 1: Summary Statistics for School District Non-Movers and Movers

	Non-moving Students		Moving once students	
	(1) Mean	(2) Std. Dev.	(3) Mean	(4) Std. Dev.
Attaining Bachelor's degree	0.239	0.427	0.196	0.397
Attaining STEM degree	0.047	0.212	0.038	0.191
Female students	0.49	0.50	0.50	0.50
Black students	0.11	0.31	0.14	0.35
Hispanic students	0.47	0.50	0.43	0.50
White students	0.43	0.49	0.42	0.59
Economically disadvantaged students	0.50	0.50	0.54	0.50
Observations	1,120,840		334,239	

Notes: The table presents summary statistics for the samples used in the school district-level analyses. The sample consists of all students in the data who attend kindergarten between 1994-2003 and are observed for at least 10 years within Texas public schools. I report summary statistics for two subsets of this sample. Non-moving students represent students who do not move across school districts throughout their school years–who live in school districts with more than 70 non-moving students in one cohort on average between 1994-2003. Moving-once students represent students who move once across school districts with more than 70 non-moving students in one cohort on average in their school years. Attaining a bachelor's degree is defined as having a bachelor's degree in an institution in Texas by age 26. Attaining a STEM degree is defined as having a bachelor's degree in STEM majors defined by the STEM degree list from the Department of Homeland Security in an institution in Texas by age 26. See Section 2 for further details on variable and sample definitions.

Table 2: Exposure Effects in STEM Major Choice

	Parametric			Semi-parametric	
	(1) Baseline	(2) + Labor	(3) + Test scores	(4) Baseline	(5) + Destination FE
Exposure Effects	-0.040***	-0.039***	-0.041***	-0.037***	-0.035***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Observations	334,239	334,239	286,406	334,239	334,239

Notes: This table reports estimates of annual exposure effects on students' STEM degree attainment by age 26. The estimates can be interpreted as the impact of spending an additional school year in a school district where non-moving students have a 1% higher STEM attainment rate. Standard errors are shown in parentheses and clustered at the origin level. Each column reports estimates from a regression of a student's STEM degree attainment by 26 on the difference between non-moving students' STEM rates in the destination vs. the origin, interacted with the grade of the student at the time of the move (m). Each regression also includes additional controls specified in equation (4). Column (1) reports the estimate of γ from equation (4) using all students of one-time movers, defined in the notes to Table 1. Column (2) adds to column (1) labor market outcomes including average earnings and employment rates of non-moving students in origin and destination. Column (3) adds to column (1) third-grade math reading test scores of students moving across school districts. Column (4) reports the estimate of γ from the sem-parametric estimation from equation (2). Column (5) adds to column (4) destination fixed effects.

Table 3: Exposure Effect Estimates: Group Specific Convergence

Panel A: Cohort				
	(1)	(2)	(3)	(4)
Exposure effects: own cohort	-0.039***			-0.040***
	(0.004)			(0.007)
Exposure effects: older cohort		-0.046***		-0.000
		(0.005)		(0.008)
Exposure effects: younger cohort			-0.030***	0.002
			(0.004)	(0.006)
Observations	334,239	334,239	334,239	334,239
Panel B: Gender				
	(1)	(2)		(3)
Exposure effects: own gender	-0.037***			-0.029***
	(0.004)			(0.005)
Exposure effects: other gender		-0.035***		-0.011***
		(0.004)		(0.004)
Observations	334,239	334,239		334,239
Panel B: Race/Ethnicity				
	(1)	(2)	(3)	(4)
Exposure effects: own race	-0.043***			-0.039***
	(0.004)			(0.005)
Exposure effects: other race1		-0.030***		-0.002
		(0.003)		(0.003)
Exposure effects: other race2			-0.036***	-0.003
			(0.003)	(0.004)
Observations	334,239	334,239	334,239	334,239

Notes: These tables report estimates of annual exposure effects γ using group-specific average STEM attainment of non-moving students. Standard errors are shown in parentheses and clustered at the origin level. In all columns, the dependent variable is the student's STEM degree attainment by age 26. In all Panels, the sample consists of all students of one-time movers, defined in the notes to Table 1. Panel A shows cohort-specific convergence. Column (1) of Panel A replicates column (1) of Table 2, replacing the average STEM rates in the origin and destination with the average STEM rates of students who are in the same cohort as the student who moves. See Appendix D.2 for further details on the regression specifications. Columns (2) and (3) replicate column (1), replacing the own-cohort STEM rate with the STEM rates of older and younger cohorts, respectively. Column (4) combines the variables in columns (1), (2), and (3), including all the own-, older-, and younger-cohorts (placebo) STEM rates. Panel B shows gender-specific convergence. Column (1) of Panel B replicates column (1) of Table 2, replacing the STEM rate based on all non-movers in the origin and destination with STEM rates of non-moving students who have the same gender as the student who moves. Column (2) replicates column (1), replacing the own-gender STEM rate with the STEM rate of the opposite gender. Column (3) combines the variables in columns (1) and (2), including both the own-gender and other-gender (placebo) predictions. Panel C shows race/ethnicity-specific convergence. Column (1) of Panel C replicates column (1) of Table 2, replacing the STEM rates based on non-movers in the origin and destination with the STEM rates of non-moving students who are in the same race/ethnicity group as the student who moves. Black, Hispanic, and White students are included. Columns (2) and (3) replicate column (1), replacing the own-race/ethnicity STEM rate with the STEM rates of other race/ethnicity groups, respectively. Column (4) combines the variables in columns (1), (2), and (3), including all the other race/ethnicity groups (placebo) STEM rates.

Table 4: Exposure Effect Estimates: Major Specific Convergence

	(1) STEM	(2) Business	(3) Liberal Arts
Exposure effects: STEM	-0.036***	-0.008	-0.006
	(0.012)	(0.011)	(0.008)
Exposure effects: Business	-0.001	-0.037***	0.009
	(0.013)	(0.014)	(0.009)
Exposure effects: LibArts	-0.007	0.042**	-0.049***
	(0.017)	(0.020)	(0.013)
Observations	334,239	334,239	334,239

Notes: This table reports estimates of annual exposure effects γ, adding different exposure measures based on different college majors. Standard errors are shown in parentheses and clustered at the origin level. In all columns, the dependent variable is the student's specified degree attainment by age 26. The sample consists of all students of one-time movers, defined in the notes to Table 1. Column (1) replicates column (1) of Table 2, adding two different exposure measures based on other college majors: business and liberal arts (placebo) majors. Those two measures are constructed in the same way as STEM exposure measures (e.g., $\Delta biz_{od} = \overline{biz}_d - \overline{biz}_o$, the difference between non-movers' rates of business major choice in the destination vs. the origin). Columns (2) and (3) replicate column (1), replacing the dependent variable with business majors and liberal arts majors, respectively.

ONLINE APPENDIX

Neighborhood Impacts on STEM Choice

Jeonghyeok Kim

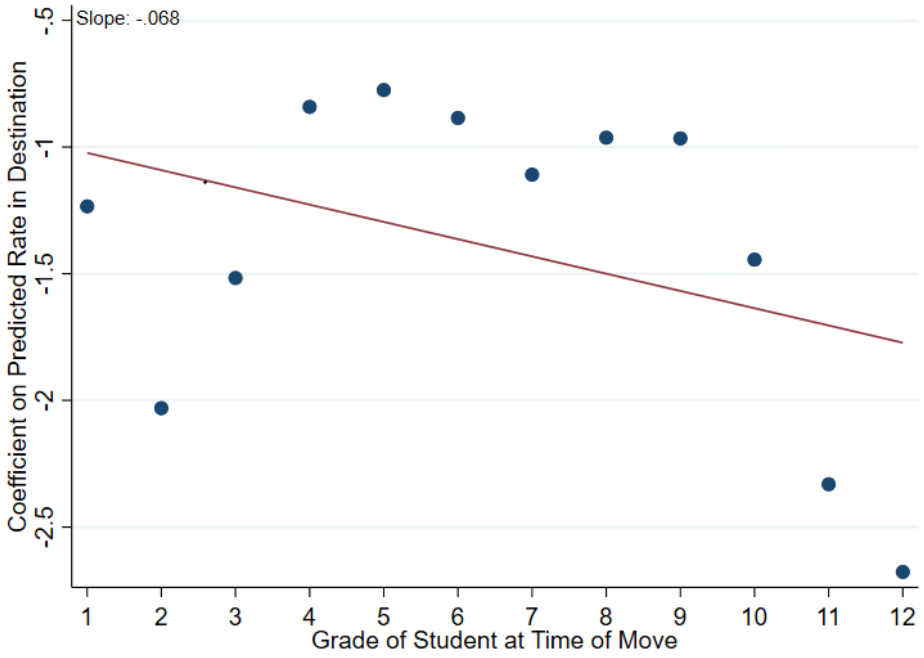
A Additional Results

Fig. A.1. Exposure Effects Estimates Using Different Enrollment Cutoffs



Notes: This table reports estimates of annual exposure effects on students' STEM degree attainment by age 26 with 95% confidence intervals, using different enrollment cutoffs ranging from no cutoff to a minimum of 2,000 non-moving students in school cohorts attending kindergarten between 1994-2003 within a school district. Each dot reports the estimate of γ from equation (4) using all students of one-time movers, defined in the notes to Table 1.

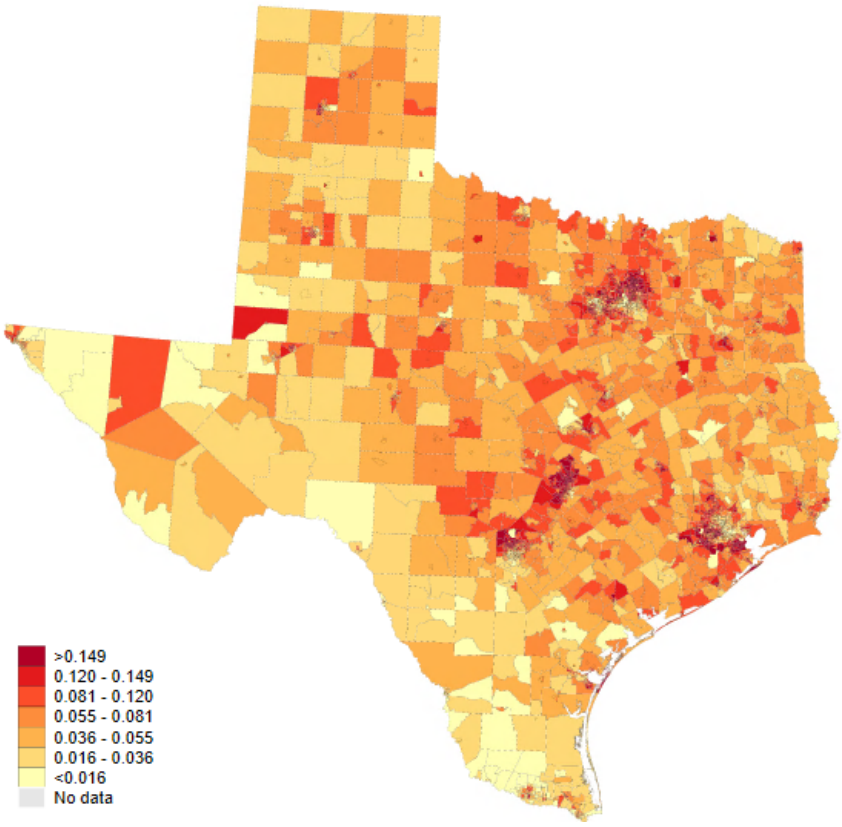
Fig. A.2. Placebo Test using Math score in Grade 3



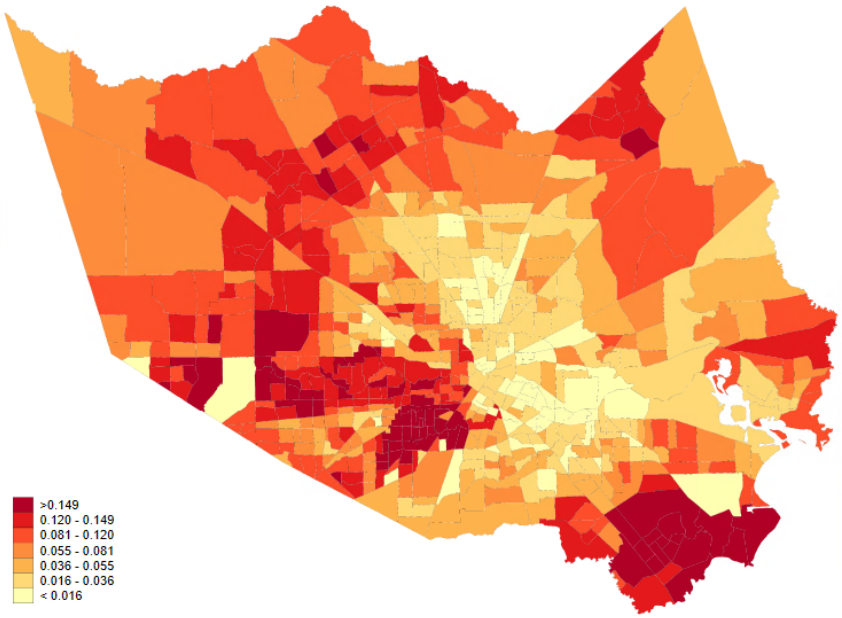
Notes: This figure plots estimates of the coefficients b_m across the students' grade when they move (m) using the same specification of Figure 4 (d), replacing the dependent variable with third-grade math score. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at age m. They are estimated by regressing the students' third-grade math score on $\Delta_{od} = \overline{STEM}_d - \overline{STEM}_o$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include origin school district-by-school cohort-by-grade at move fixed effects when estimating this specification. The slope of the regression line is reported on the left side of each panel.

Fig. A.3. STEM Occupation Proportion Across Census Tracks

(a) STEM Occupation Proportion

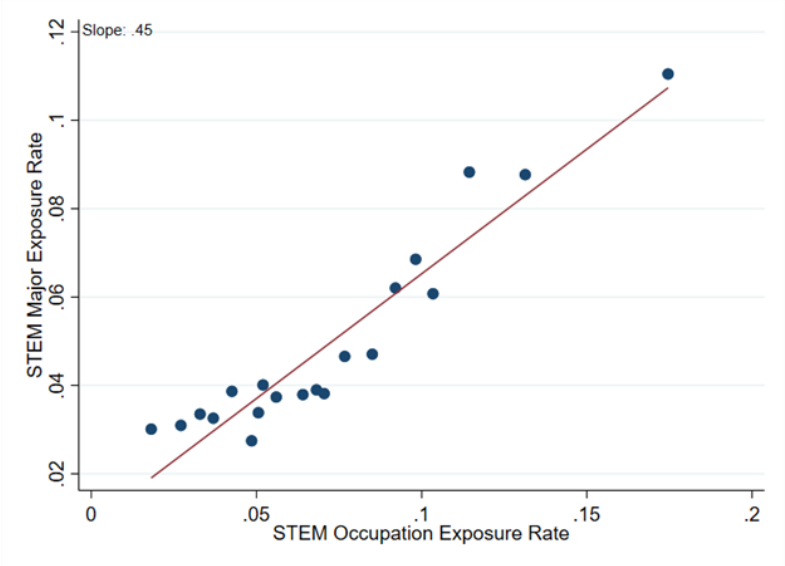


(b) STEM Occ. Proportion - Harris County



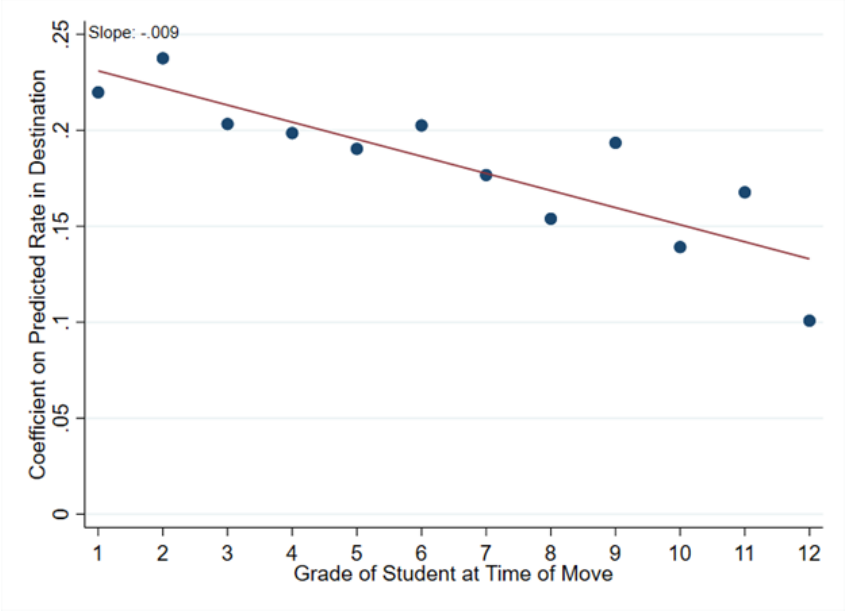
Notes: These maps plot the proportion of STEM workers in Census tracks. The maps are constructed by grouping Census tracks into seven quantiles and shading the areas so that darker colors correspond to higher proportion of STEM workers. Areas with no data are shaded with a grey color. It is based on the 2000 Census and the STEM occupation follows the list of STEM occupations from ONET.

Fig. A.4. STEM Major Rate and STEM Occupation Rate



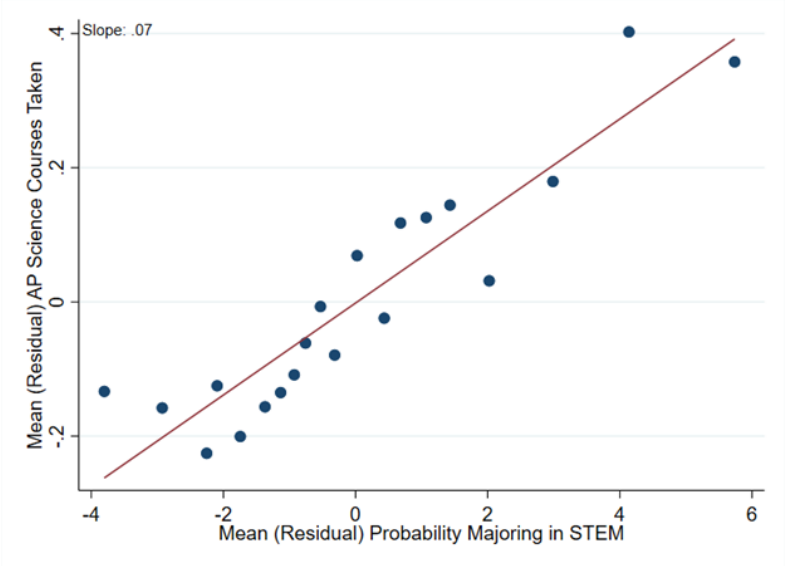
Notes: This figure presents a binned scatter plot depicting the relationship between the STEM occupation exposure rate ($STEMOcc_p$) and STEM major exposure rate ($\overline{STEM}_p$). The sample includes all students in kindergarten in 1994-2003 who never moved across school districts in school years. I divide the $STEMOcc_p$ into twenty equal-size groups (ventiles) among non-movers and plot the mean value of $\overline{STEM}_p$ vs. the mean value of $STEMOcc_p$ in each bin.

Fig. A.5. Exposure Effects on STEM Industry Choice



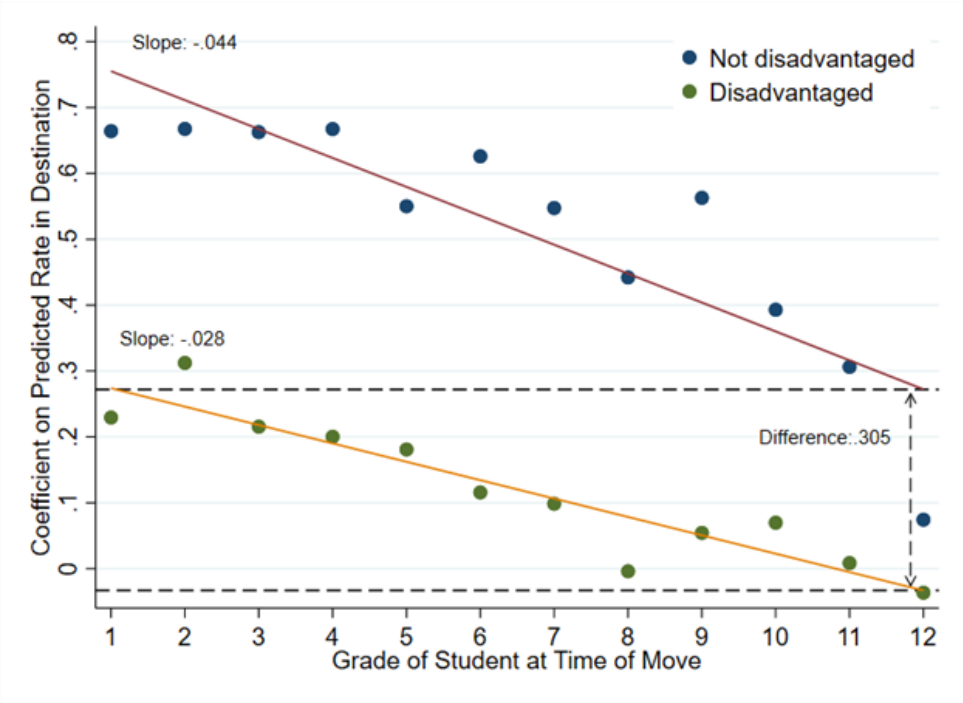
Notes: Figure plots estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), replacing outcome variable to STEM industry measure——proportion of STEM workers within 4 digit NAICS industry. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at grade m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta STEM_{od} = \overline{STEM}_d - \overline{STEM}_o$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with predicted STEM rate for non-moving students in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line is reported on the left side of each line.

Fig. A.6. STEM Choice Rates and AP Science Couses Taken by School Districts



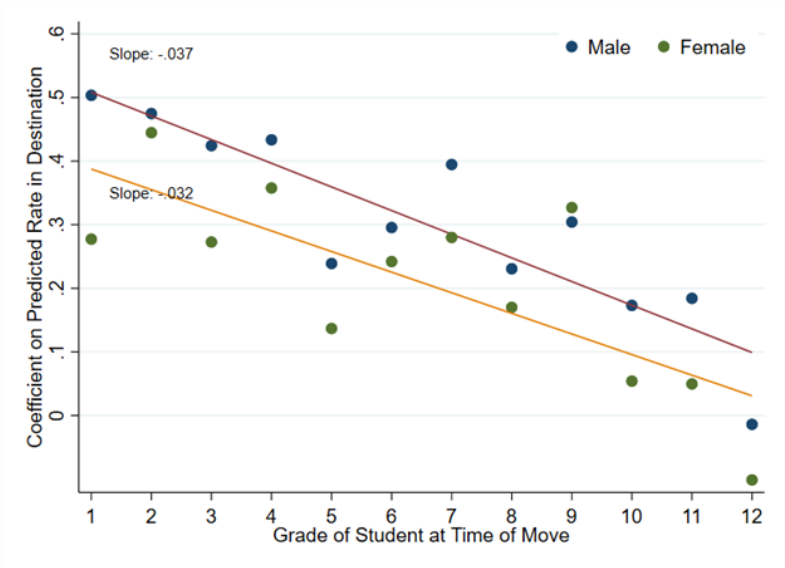
Notes: This figure presents a binned scatter plot depicting the relationship between the STEM major exposure rate ($\overline{STEM}_p$) and average AP science courses taken by one student in a school district. The sample includes all students in kindergarten in 1994-2003 who never moved across school districts in school years. To construct the figure, I first regress $\overline{STEM}_p$ on the number of non-science AP courses and obtain residual $\overline{STEM}^r_p$ and regress the number of science AP courses on the number of non-science AP courses and obtain residual AP^r_{sci}. Then, I demean both AP^r_{sci} and $\overline{STEM}^r_p$ by the school district. I then divide the demeaned $\overline{STEM}^r_p$ into twenty equal-size groups (ventiles) and plot the mean value of demeaned AP^r_{sci} vs. the mean value of demeaned $\overline{STEM}^r_p$ in each bin.

Fig. A.7. Heterogeneous Exposure Effects by Economic Disadvantaged Status: Robust



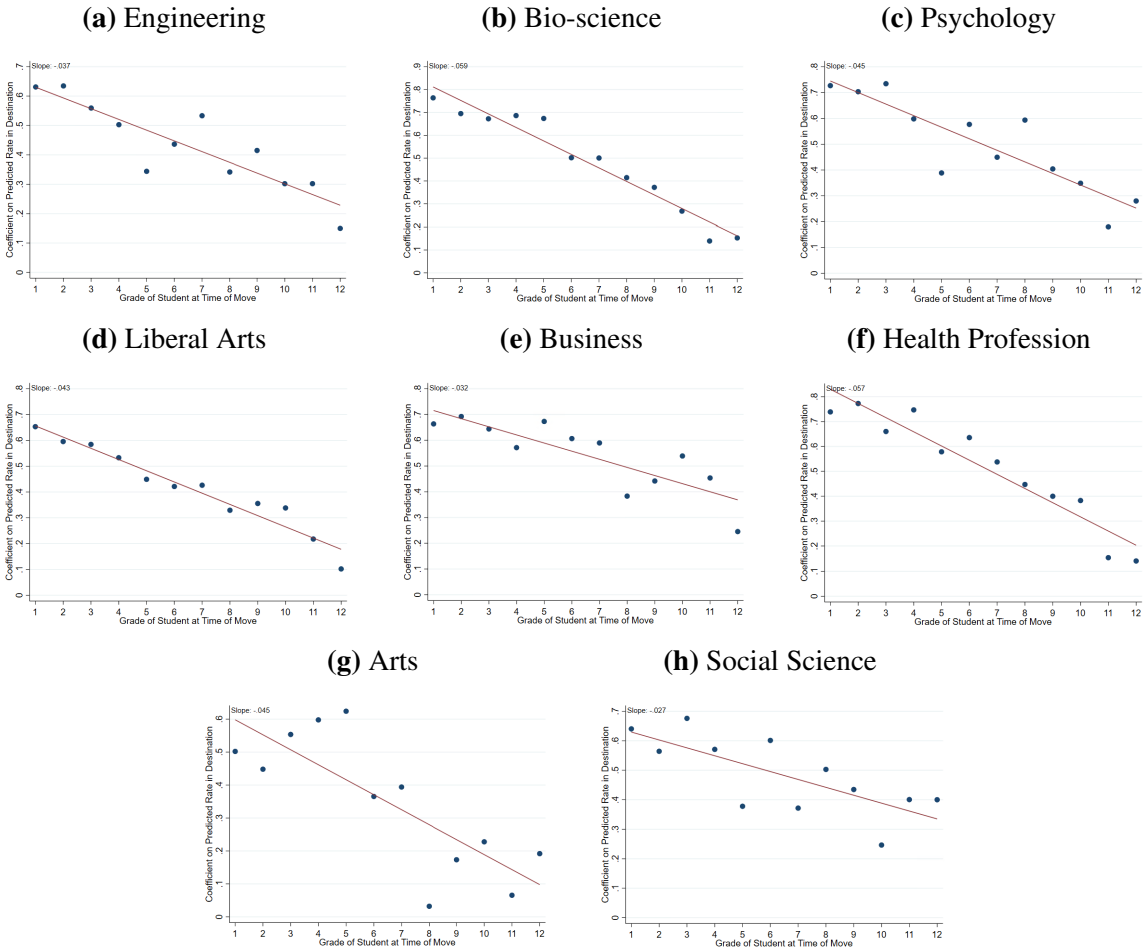
Notes: Figure plots estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), separately by the economically disadvantaged status of students. Students are defined as economically disadvantaged if they receive free or reduced-price lunch in both kindergarten and at the time of move. The blue dots and the red line are from the sample of not economically disadvantaged students and the green dots and the orange line are from the sample of disadvantaged students. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' STEM choices $STEM_i$ are measured as obtaining a STEM bachelor's degree by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at grade m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta STEM_{od} = \overline{STEM}_d - \overline{STEM}_o$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with predicted STEM rate for non-moving students in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line is reported on the left side of each line.

Fig. A.8. Heterogeneous Exposure Effects by Gender: Engineering



Notes: Figure plots estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), measuring students' engineering degree attainment, separately by gender of students. The blue dots and the red line are from the sample of male students and the green dots and the orange line are from the sample of female students. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' engineering choice is measured as obtaining an engineering bachelor's degree by age 26. They are estimated by regressing the students' engineering degree attainment in college $engin_i$ on $\Delta engin_{od} = \overline{engin}_d - \overline{engin}_o$, the difference between non-moving students' predicted rates of engineering choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with predicted engineering rate for non-moving students in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line is reported on the left side.

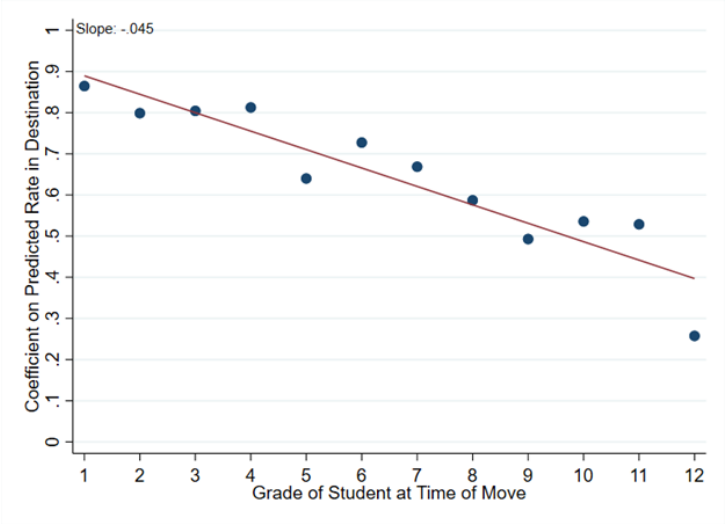
Fig. A.9. Exposure Effects on Other Major Choice in College



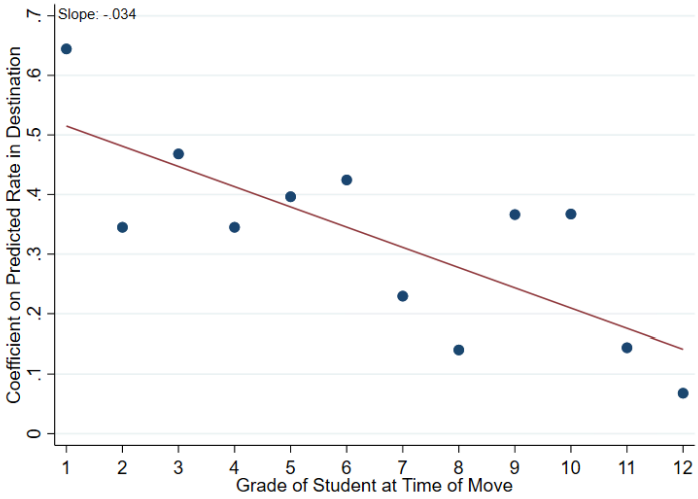
Notes: Figures plot estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), measuring students' degree attainment in the specified major on the label. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' major choice is measured as obtaining a bachelor's degree in the specified major by age 26. They are estimated by regressing the students' degree attainment of the major $major_i$ on $\Delta major_{od} = \overline{major}_d - \overline{major}_o$, the difference between non-moving students' predicted rates of the major choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with the predicted major rate for non-moving students in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line is reported on the left side.

Fig. A.10. Exposure Effects on STEM Choice: Enrollment and Graduate Degree

(a) Enrolling in Any College



(b) Obtaining in Graduate Degree



Notes: Figure plots estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), defining STEM choice differently. STEM choice is defined as enrolling in a STEM major in any college by age 26 for Figure (a) and obtaining a STEM graduate degree in Figure (b). This specification replicates the specification used in Figure 4 (b), replacing the STEM rates of school districts and STEM choice of movers at the specified levels. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at grade m. They are estimated by regressing the students' STEM enrollment or graduate degree $STEM_i$ on $\Delta STEM_{od}$, the difference between STEM enrollment rates or STEM graduation degree rates the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with the STEM occupation rate in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line is reported on the left side of each panel.

Table A.1: Exposure Effect Estimates: Major Specific Convergence-Narrower

	(1) Engineering	(2) Biological Science	(3) LibArts
Exposure effects: Engin	-0.044***	0.011*	0.002
	(0.008)	(0.006)	(0.008)
Exposure effects: Bio	0.010	-0.062***	0.007
	(0.021)	(0.015)	(0.019)
Exposure effects: LibArts	0.009	-0.009	-0.048***
	(0.012)	(0.008)	(0.012)
Observations	334,239	334,239	334,239

Notes: This table reports estimates of annual exposure effects γ, adding different exposure measures based on different college majors. Standard errors are shown in parentheses and clustered at the origin level. In all columns, the dependent variable is the student's specified degree attainment by age 26. The sample consists of all students of one-time movers, defined in the notes to Table 1. Column (1) replicates column (1) of Table 2, including three different exposure measures based on college majors: engineering, biological science, and liberal arts. Engineering and biological science are two popular majors that are included in STEM majors. Those three measures are constructed in the same way as STEM exposure measures (e.g., $\Delta engin_{od} = \overline{engin}_d - \overline{engin}_o$, the difference between non-movers' rates of engineering major choice in the destination vs. the origin). Columns (2) and (3) replicate column (1), replacing the dependent variable with business majors and liberal arts majors, respectively.

Table A.2: Exposure Effects of STEM Occupations

	Parametric			Semi-parametric	
	(1) Baseline	(2) + Labor	(3) + Test scores	(4) Baseline	(5) + Destination FE
Exposure Effects	-0.026***	-0.025***	-0.027***	-0.020***	-0.019***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Observations	334,239	334,239	286,406	334,239	334,239

Notes: This table reports estimates of annual exposure effects on students' STEM degree attainment by age 26. The estimates can be interpreted as the impact of spending an additional school year in a school district with 1% more STEM workers in a neighborhood. Standard errors are shown in parentheses and clustered at the origin level. Each column reports estimates from a regression of a student's STEM degree attainment by 26 on the difference between neighborhood STEM occupation rates in the destination vs. the origin, interacted with the grade of the student at the time of the move (m). Each regression also includes additional controls specified in equation (4). Column (1) reports the estimate of γ from equation (4) with STEM occupation exposure measure using all students of one-time movers, defined in the notes to Table 1. Column (2) adds to column (1) labor market outcomes including average earnings and employment rates of non-moving students in origin and destination. Column (3) adds to column (1) third-grade math reading test scores of students moving across school districts. Column (4) reports the estimate of γ from the sem-parametric estimation from equation (2) with STEM occupation exposure measure. Column (5) adds to column (4) destination fixed effects.

Table A.3: Exposure Effects of STEM Occupations on STEM industry

	Parametric			Semi-parametric	
	(1) Baseline	(2) + Labor	(3) + Test scores	(4) Baseline	(5) + Destination FE
Exposure Effects	-0.005***	-0.005***	-0.003***	-0.008***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Observations	334,239	334,239	286,406	334,239	334,239

Notes: This table reports estimates of annual exposure effects on students' STEM industry choice. The estimates can be interpreted as the impact of spending an additional school year in a school district with 1% more STEM workers in a neighborhood. Standard errors are shown in parentheses and clustered at the origin level. Each column reports estimates from a regression of a student's STEM industry choice on the difference between neighborhood STEM occupation rates in the destination vs. the origin, interacted with the grade of the student at the time of the move (m). Each regression also includes additional controls specified in equation (4). Column (1) reports the estimate of γ from equation (4) with STEM occupation exposure measure using all students of one-time movers, defined in the notes to Table 1. Column (2) adds to column (1) labor market outcomes including average earnings and employment rates of non-moving students in origin and destination. Column (3) adds to column (1) third-grade math reading test scores of students moving across school districts. Column (4) reports the estimate of γ from the sem-parametric estimation from equation (2) with STEM occupation exposure measure. Column (5) adds to column (4) destination fixed effects.

Table A.4: Exposure Effect Estimates to Neighborhoods Occupational Composition: Robustness

	(1) Base	(2) + $\leq$ 0.5 mile	(3) + $\leq$ 1	(4) + $\leq$ 2	(5) + $\leq$ 3	(6) + $\leq$ 5	(7) Only Ele.
Exposure	-0.026***	-0.026***	-0.026***	-0.026***	-0.025***	-0.026***	-0.026***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Obs.	423,851	423,851	423,851	423,851	423,851	423,851	423,851

Notes: This table reports estimates of annual exposure effects on students' STEM degree attainment by age 26, with different neighborhood STEM occupation measures. The estimates can be interpreted as the impact of spending an additional school year in a school district with 1% more STEM workers in a neighborhood. Standard errors are shown in parentheses and clustered at the origin level. Each column reports estimates from a regression of a student's STEM degree attainment by 26 on the difference between neighborhood STEM occupation rates in the destination vs. the origin, interacted with the grade of the student at the time of the move (m). Each regression also includes additional controls specified in equation (4). Column (1) reports the baseline estimate of γ from Table A.2 column(1) where STEM rate of school ($STEM_{c(s)}$) is defined as STEM rate of Census track school is located. Columns (2)–(6) define the school's STEM rate as the STEM rate of the census tracts within the specified radius (in miles) in each column, including the tract where the school is located. Column (7) includes only elementary schools to calculate the district occupation exposure measure.

Table A.5: Gender Specific Convergence to Neighborhood Occupational Composition

	(1)	(2)	(3)
Exposure effects: own gender	-0.019***		-0.018***
	(0.002)		(0.003)
Exposure effects: other gender		-0.017***	-0.002
		(0.002)	(0.002)
Observations	288,507	288,507	288,507

Notes: The table reports estimates of annual exposure effects γ using gender-specific STEM occupational rates. Standard errors are shown in parentheses and clustered at the origin level. In all columns, the dependent variable is the student's STEM degree attainment by age 26. The sample consists of all students of one-time movers, defined in the notes to Table 1. The table shows gender-specific convergence to own-gender STEM worker proportion. Column (1) replicates column (1) of Table A.2, replacing the STEM occupation rate based on all workers in the origin and destination with STEM occupation rates of the same gender as the student who moves. Column (2) replicates column (1), replacing the own-gender STEM occupation rate with the STEM occupation rate of the opposite gender. Column (3) combines the variables in columns (1) and (2), including both the own-gender and other-gender (placebo) predictions.

Table A.6: Major Specific Convergence to Neighborhood Occupational Composition

	(1) STEM	(3) Business
Exposure Effects-STEM Occ.	-0.019***	-0.007**
	(0.004)	(0.003)
Exposure Effects-Manager Occ.	0.010	-0.033***
	(0.010)	(0.009)
Observations	354,129	354,129

Notes: This table reports estimates of annual exposure effects γ, adding a different exposure measure based on a different occupation. Standard errors are shown in parentheses and clustered at the origin level. In all columns, the dependent variable is the student's specified degree attainment by age 26. The sample consists of all students of one-time movers, defined in the notes to Table 1. Column (1) replicates column (1) of Table A.2, adding another exposure measure based on managerial occupations. The measure is constructed in the same way as STEM exposure measures (e.g., $\Delta manager_{od} = \overline{manager}_d - \overline{manager}_o$, the difference between the proportion of managerial occupations in the destination vs. the origin). Column (2) replicates column (1), replacing the dependent variable with business majors.

Table A.7: Heterogeneous Effects Across Different Groups

	(1) Male	(2) Female	(3) Black	(4) Hispanic	(5) White	(6) Not Econ.	(7) Econ.
Exposure	-0.033***	-0.042***	-0.023***	-0.024***	-0.028***	-0.037***	-0.030***
	(0.004)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.004)
Obs.	156,522	150,789	44,400	138,454	110,999	145,990	161,321

Notes: The table reports estimates of annual exposure effects γ using a group of samples. Standard errors are shown in parentheses and clustered at the origin level. In all columns, the dependent variable is the student's STEM degree attainment by age 26. The sample consists of a group of students of one-time movers specified in the column title. So, column (1) includes only male students, and column (6) only includes not economically disadvantaged students, which is defined based on free or reduced-priced lunch in kindergarten. Columns (1)-(5) replicate column (1) of Table 2, replacing the STEM rate based on all non-movers in the origin and destination with STEM rates of the specified group of students. Columns (6) and (7) use the STEM rate based on all non-movers in the origin and destination.

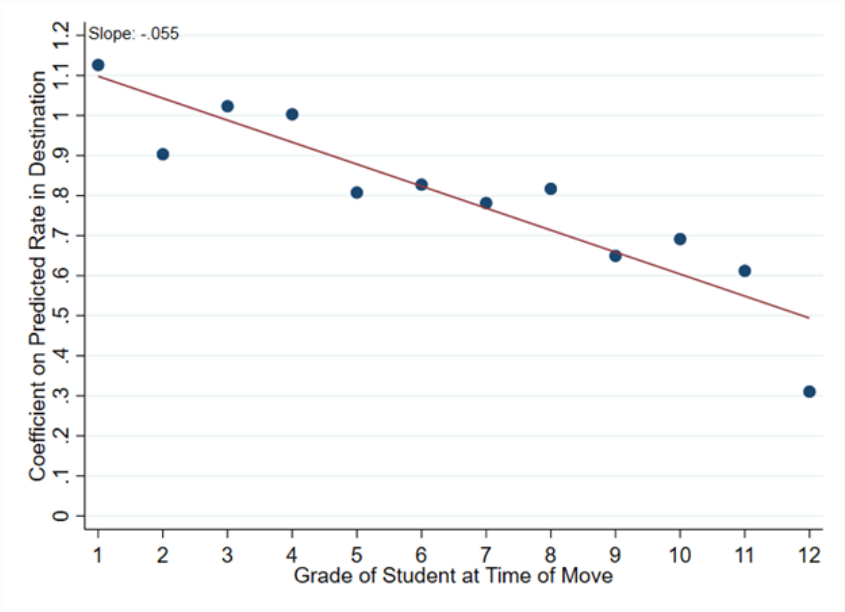
B County Level Estimates

In this appendix, I present estimates of exposure effects using counties as the geographic unit of analysis instead of school districts. The county-level STEM exposure, $\overline{STEMp}$, is calculated using the same method as in Figure 1. I then regress students' STEM choices on $\Delta STEMod = \overline{STEM}_d - \overline{STEM}_o$, which is the difference in STEM exposure between their destination and origin counties, interacted with the grade at which they moved. This analysis follows a specification similar to that of equation 2, but applied at the county level. Consistent with the baseline analysis, I focus on students who move exactly once between counties. The sample is further restricted to origin and destination counties with at least 1,000 non-moving students within a 10-year school cohort, excluding less than 5% of the population.

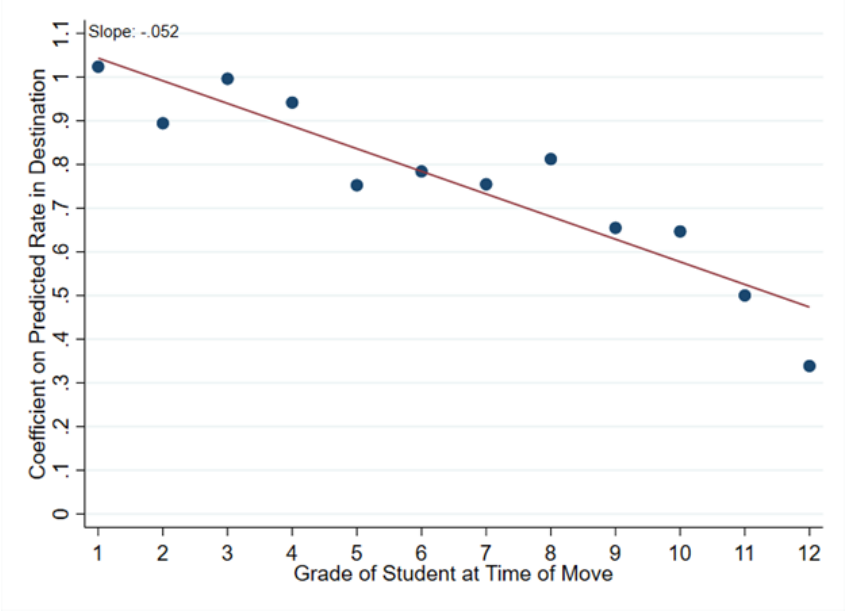
Appendix Figure A.11 and Table A.8 present estimates analogous to those in Figure 4 and Table 2, respectively. The results show a slightly larger estimate of $\gamma = 0.048$ (s.e. = 0.009), compared to the baseline estimate of $\gamma = 0.041$. Overall, the estimates are consistent with the baseline results. Additionally, Appendix Figure A.12 presents analogous estimates to Figure 6, using county-level STEM occupation rates. Once again, the estimates of γ are comparable to those in the baseline analysis.

Fig. A.11. Exposure Effects on STEM Choice: County Level

(a) Semi-parametric Estimates



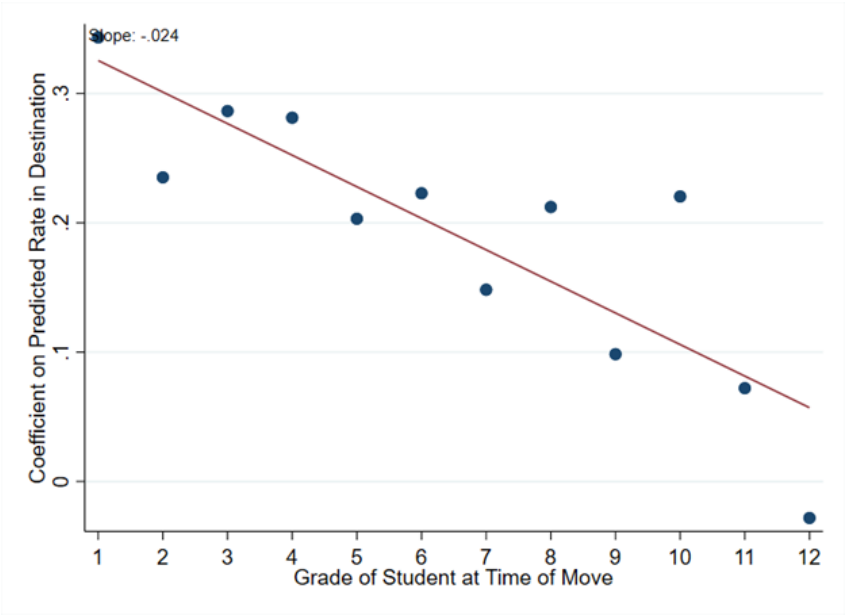
(b) Parametric Estimates



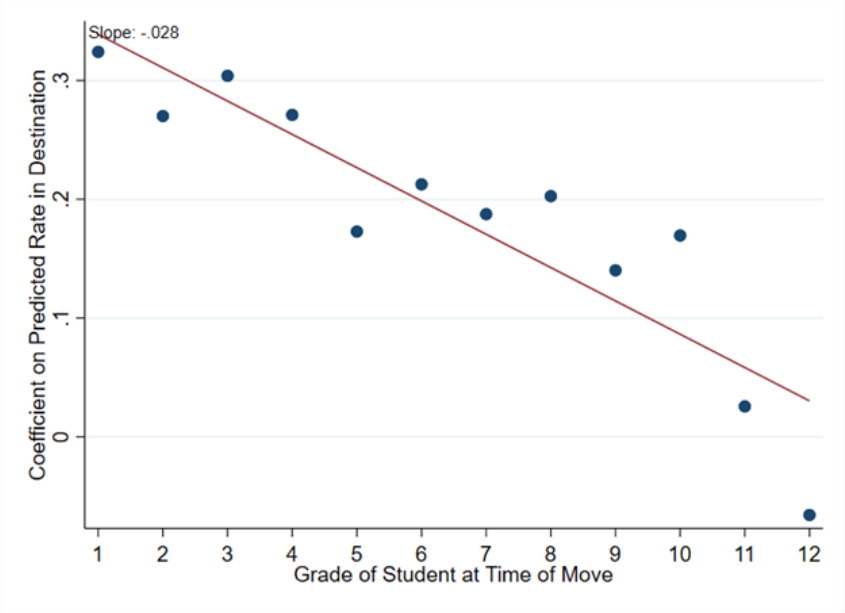
Notes: Figure (a) plots estimates of the coefficients b_m across the students' grade when they move (m) using the semi-parametric specification in equation (2), measuring students' STEM degree attainment. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years across counties. Students' STEM choice $STEM_i$ is measured as obtaining a bachelor's degree in STEM majors by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at age m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta_{od} = \overline{STEM}_d - \overline{STEM}_o$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include origin county-by-school cohort-by-grade at move fixed effects when estimating this specification. Figure (b) plots estimates from the parametric specification in equation (3). This specification replicates the specification used in Figure (a), replacing the origin fixed effects with the STEM rate for non-movers in the origin. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line are reported on the left side of each panel. The magnitudes of the slope represent estimates of annual exposure effects.

Fig. A.12. Exposure Effects on STEM Choice: STEM Occupation at County Level

(a) Semi-parametric Estimates



(b) Parametric Estimates



Notes: The figure plots estimates of the coefficients b_m across the students' grade when they move (m) using the parametric specification in equation (2), replacing $\Delta STEM_{od}$ to $\Delta STEMOcc_{od} = \overline{STEMOcc}_d$ - $\overline{STEMOcc}_o$, the difference in STEM worker proportions between destination and the origin. The sample includes all students in kindergarten in 1994-2003 who moved only once across counties in their school years. Students' STEM choices $STEM_i$ are measured as obtaining a STEM bachelor's degree by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where the neighborhood STEM occupation rate increases from 0 to 1 at grade m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta STEMOcc_{od}$, the difference between STEM worker proportion in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include indicators for the student's grade at the time of the move interacted with the STEM occupation rate in the origin interacted with school cohort fixed effects. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line are reported on the left side of each panel. The magnitude of the slope represents estimates of annual exposure effects to occupational composition.

Table A.8: Exposure Effects in STEM Major Choice: County-Level

	Parametric			Semi-parametric	
	(1) Baseline	(2) + Labor	(3) + Test scores	(4) Baseline	(5) + Destination FE
Exposure Effects	-0.048***	-0.049***	-0.049***	-0.052***	-0.049***
	(0.009)	(0.009)	(0.007)	(0.012)	(0.011)
Observations	200,567	200,567	179,607	200,567	200,567

Notes: This table reports estimates of annual childhood exposure effects on students' STEM degree attainment by age 26 at the county level. The estimates can be interpreted as the impact of spending an additional school year in a county where non-moving students have 100% higher STEM attainment rate. Standard errors are shown in parentheses. Each column reports estimates from a regression of a student's STEM degree attainment by 26 on the difference between non-movers' STEM rate in the destination vs. the origin, interacted with the grade of the student at the time of the move (m). Each regression also includes additional controls specified in equation 4. Column (1) reports the estimate of γ from equation 4 using all students in the primary analysis sample of one-time movers, defined in the notes to Table 1. Column (2) adds to column (1) labor market outcomes including average earnings and employment rates of non-moving students in origin and destination. Column (3) adds to column (1) third-grade math reading test scores of students moving across school districts. Column (4) reports the estimate of γ from the sem-parametric estimation from equation **??**. Column (5) adds to column (4) destination fixed effects.

C Economic Status and Race Specific Estimation

In this appendix, I present estimates of exposure effects using race and economic group specific level measures instead of school districts as whole. The group specific STEM exposure, $\overline{STEM}_{pg}$, is calculated using the same method as in Figure 1 but using race (Black, Hispanic, and White) and economic status (free or reduced price lunch status) specific rates. Thus, each school district has six different STEM rates. I then regress students' STEM choices on $\Delta STEModg = \overline{STEM}_{dg} - \overline{STEM}_{og}$, which is the difference in STEM exposure between their own group in destination and origin districts, interacted with the grade at which they moved. This analysis follows a specification similar to that of equations (2) and (3):

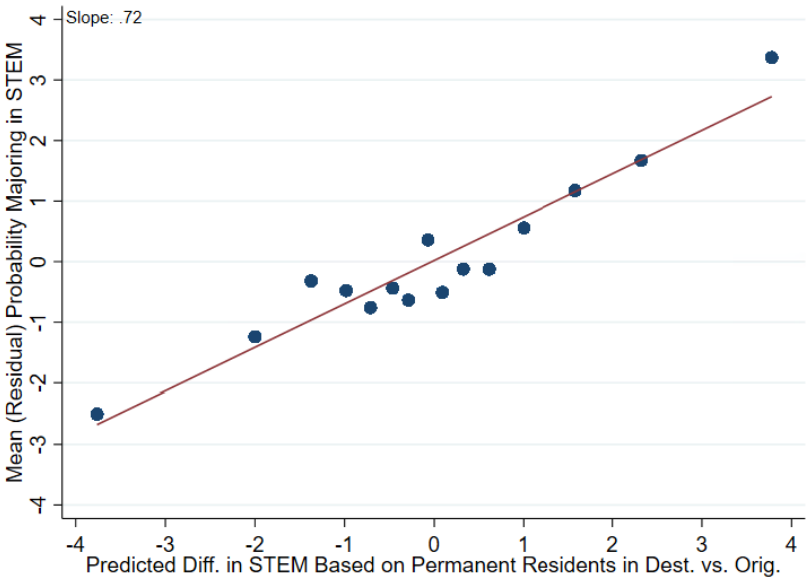
$$STEM_i = \alpha_{osmg} + \sum_{m=1}^{12} b_m I(m_i = m)\Delta STEM_{odg} + \eta X_{dg} + \varepsilon_{1i} \tag{A.1}$$

$$STEM_i = \alpha + \sum_{m=1}^{12} b_m I(m_i = m)\Delta STEM_{odg} + \theta_2 X_{dg} + \sum_{m,s,g} I(m_i = m, s_i = s, g_i = g)(\alpha^1_{m,s,g} + \alpha^2_{m,s,g}\overline{STEM}_{o,g}) + \varepsilon_{2i}. \tag{A.2}$$

Instead of α_{osm}, in equation (A.1) I include α_{osmg}, a fixed effect for the origin school district o by school cohort s by grade at move m by economic and race group g. Other variables also change to get race and economic status specific values. Consistent with the baseline analysis, I focus on students who move exactly once between counties. The sample is further restricted to origin and destination counties with at least 1,500 non-moving students within a 10-year school cohort.

Appendix Figures A.13 and A.14 present estimates analogous to those in Figures 3 and 4, respectively. While the estimates are generally consistent with the baseline results, they exhibit more noise. Additionally, they show a slightly smaller estimate of γ compared to the baseline value of $\gamma = 0.041$.

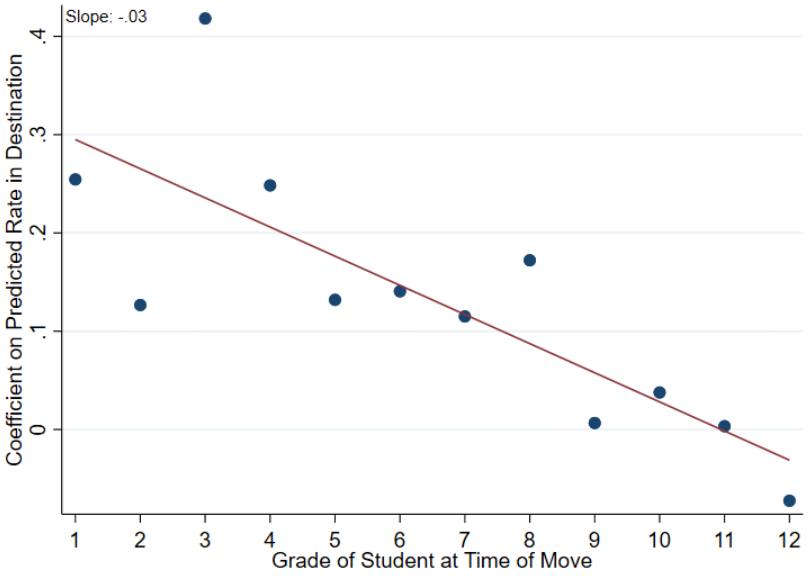
Fig. A.13. Mover's Outcomes Versus Predicted Outcomes Based on Permanent Residents in Destination



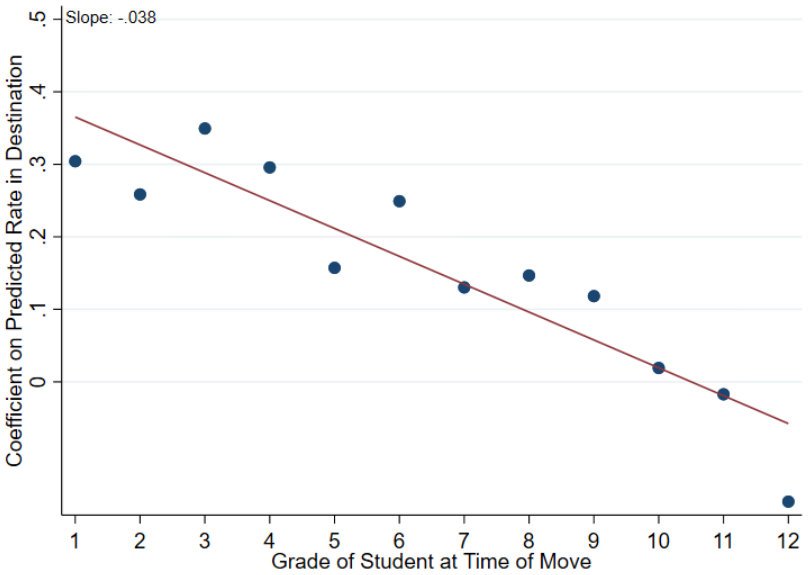
Notes: This figure presents a binned scatter plot depicting the relationship between the STEM probability of students who moved to a different school district at grade 3 and the differences in the average probability of majoring in STEM of non-moving students in the destination vs. origin school districts. The sample includes all students in kindergarten in 1994-2003 who moved when the student was in grade 3 and moved only once in school years. Students' STEM choices $STEM_i$ are measured as obtaining a STEM degree by age 26. $\overline{STEM}_{pg}$ is the average STEM choice of non-moving students in school district p in group g based on race and economic status. To construct the figure, I first demean both $STEM_i$ and $\Delta_{odg} = \overline{STEM}_{dg} - \overline{STEM}_{og}$ by origin (o) by birth cohort (s) cells in the sample of movers at grade m = 3 to construct residuals: $STEM_i^r = STEM_i - E[STEM_i|o,s,g]$ and $\Delta_{od}^r = \Delta_{odg} - E[\Delta_{od}|o,s,g]$. I then divide the Δ_{odg}^r into fifteen equal-size groups and plot the mean value of $STEM_i^r$ vs. the mean value of Δ_{odg}^r in each bin.

Fig. A.14. Exposure Effects on STEM Choice in College

(a) Semi-parametric Estimates



(b) Parametric Estimates



Notes: Figure (a) plots estimates of the coefficients b_m across the students' grade when they move (m) using the semi-parametric specification in equation (2), measuring students' STEM degree attainment. The sample includes all students in kindergarten in 1994-2003 who moved only once in their school years. Students' STEM choice $STEM_i$ is measured as obtaining a bachelor's degree in STEM majors by age 26. The b_m coefficients can be interpreted as the effect of moving to an area where non-moving students' STEM rate increases from 0 to 1 at age m. They are estimated by regressing the students' STEM degree attainment in college $STEM_i$ on $\Delta_{odg} = \overline{STEM}_{dg}$ - $\overline{STEM}_{og}$, the difference between non-movers' predicted rates of STEM choice in the destination vs. the origin, interacted with each grade of the student at the time of the move m. I include origin school district-by-school cohort-by-grade at move-by-race-by-economic status fixed effects when estimating this specification. Figure (b) plots estimates from the parametric specification in equation (3). This specification replicates the specification used in Figure (a), replacing the origin fixed effects with the STEM rate for non-movers in the same group in origin. Best-fit lines are estimated using unweighted OLS regressions of the b_m coefficients on m. The slopes of the regression line are reported on the left side of each panel. The magnitudes of the slope represent estimates of annual exposure effects.

D Details for Validation of Baseline Design

D.1 Displacement Shocks

One method to account for unobservable shocks is to focus on moves where the underlying cause of the move is known. For instance, consider identifying families forced to move from an origin o to a nearby destination d due to an exogenous shock, such as a natural disaster. These displacement shocks can lead to different changes in neighborhood STEM rates ($\Delta STEM_{od}$). If these shocks do not have direct effects on students that are correlated with $\Delta STEM_{od}$, the identification assumption holds, and I can obtain unbiased estimates of the exposure effect γ by focusing on displaced students. By isolating a subset of moves driven by exogenous shocks, this approach allows for a more credible assessment of whether changes in students' STEM choices are influenced by the new neighborhood environment, rather than by unobservable factors.

I begin by identifying displacement shocks through population outflows at the school district level. Let K_{pt} represent the number of students who leave school district p in year t within a sample of one-time movers, and let $\bar{K}_p$ denote the average outflow for that district over several years. The shock to outflows in year t for district p is then defined as $kpt = K_{pt}/\bar{K}_p$.

Figure A.15 illustrates the results of this analysis. To create this figure, I sort observations into population-weighted bins based on the magnitude of the shock k_{dt}. The first point in the figure represents the estimated annual exposure effect using all observations with k_{dt}. The second point shows the estimate γ using observations where k_{dt} is at or above the 10th percentile. Subsequent points are calculated similarly, increasing the threshold by 10 percentiles each time, with the final point showing an estimate of γ based only on districts in the top two percentiles of outflow rates. The dotted lines represent the 95% confidence intervals for these estimates.

If the initial estimates were purely driven by selection bias, I would expect γ to drop to zero as the sample is restricted to students whose moves were likely induced by exogenous displacement shocks. However, the coefficients remain stable at around $\gamma \simeq 0.04$, even when focusing on moves associated with significant displacement. This suggests that students who were in lower grades at the time of the move continue to show an increased likelihood of choosing STEM fields as adults, even when their move was likely driven by an external shock. These findings

support the conclusion that the baseline exposure effect estimates reflect the causal impact of neighborhood environments rather than unobserved factors associated with family decisions to move. This result implies that the exposure effects identified by the initial analysis can be generalized to a broader population beyond those families who voluntarily choose to move.

D.2 Specifications for Outcome-Based Placebo Tests

D.2.1 Assumption for Group-Specific Convergence

In this section, I provide more formal details of the overidentification tests. Intuitively, it is improbable that a correlated shock would coincide precisely with cohort-level differences in place effects, as reflected in the STEM rates of non-moving students. Formally, this test is based on the assumption that if unobserved factors θ_i are correlated with exposure to the place effect for a given cohort $s(i)$ (proxied by the STEM rates of non-moving students), they should also be correlated with exposure to the place effects of adjacent cohorts t:

$$Cov(\theta_i, m\Delta_{od,s(i)}|X) > 0 \Rightarrow Cov(\theta_i, m\Delta_{od,s'}|X, m\Delta_{od,s(i)}) > 0$$

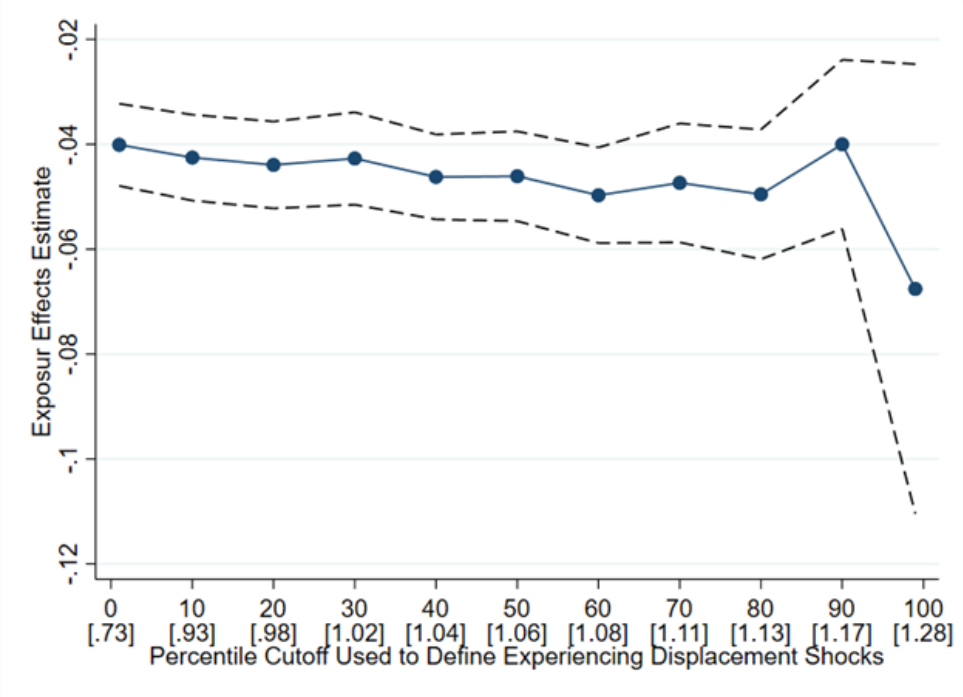
where X represents the vector of fixed effects and other controls in 4. Assuming this holds true, the results in Table 3 indicate that my estimates of γ capture causal neighborhood effects that are specific to each cohort, rather than being influenced by omitted variables, which would not exhibit cohort-specificity.

D.2.2 Construction of School Cohorts

In the section, I provide details for the construction of school cohorts in the overidentification test using school cohorts. I have 10 school cohorts from 1994-2003. Due to the size of the sample, I divide the cohorts into three groups: 1994-1997, 1998-2000, and 2001-2003 cohorts. If I cannot observe $\Delta STEM_{odt}$, I set the younger or older exposure variables to zero and I include an indicator, $I_a = I\{\text{cohort s}(i) + \text{a is missing}\}$ for each. For instance, students in the 1994-1997 cohort have their own cohort exposure effects and younger cohort exposure effects but older

cohort exposure effects.

Fig. A.15. Exposure Effects Estimates Using Displacement Shocks



Notes: This figure presents estimates of annual exposure effects for areas experiencing displacement shocks, defined as school district-by-year cells with large outflows of residents. I measure outflows by defining K_{pt} as the number of students who leave school district p in year t in my one-time movers sample, and $\bar{K}p$ as the average outflow over the years. The outflow shock in year t for school district p is then defined as $kpt = K_{pt}/\bar{K}p$. The sample is divided into 10 population-weighted bins based on the magnitude of the shock kpt. For each subset of observations where k_{pt} exceeds the percentile threshold listed on the x-axis, I estimate γ using equation (4). The dashed lines represent 95% confidence intervals for the estimates. The mean value of the outflow shock k_{pt} used in each subsample is shown in brackets below the percentile thresholds.

E Details for Counterfactual Analysis

In this section, I outline the methodology for conducting the counterfactual analysis. The analysis includes all students who began kindergarten between 1994 and 2003. To determine the exposure level, I calculate the annual exposure rate based on the proportion of students who attain a STEM degree by age 26. The mean exposure level is then computed for each group at two levels: district and campus. Although the exposure effects are estimated at the district level, it is important to note that school districts may encompass areas larger than the neighborhoods with which students regularly interact. Therefore, I consider the district-level estimates to be a lower bound for more accurate counterfactual results. In the similar way, I conduct counterfactual analysis using STEM occupation exposure levels.

Same STEM neighborhood (district level). What if the average neighborhood STEM rates for underrepresented groups increase to match those of other groups? 4.6 percent of White students choose STEM majors, compared to 1.4 percent of Black students and 2.4 percent of Hispanic students. The average neighborhood exposure rate—based on the proportion of non-moving students choosing to major in STEM—of White students (0.056% of non-moving students) is 39% higher than that of Black students (0.043) and 30% more than Hispanic students (0.043). With the same neighborhood exposure rate as neighborhoods of White students in 13 years, Black students would increase the probability of majoring in STEM by (0.056-0.043)*13*4.0=0.68 percentage points, 21% of the gap between Black and White students. Hispanic students would increase the probability of majoring in STEM by (0.056-0.043)*13*4.0=0.68 percentage points, 31% of the gap between Hispanic and White students.

The average neighborhood of economically not disadvantaged students (0.058) is 41% higher than that of disadvantaged students (0.041). 5.5 percent of economically not disadvantaged students choose STEM majors, compared to 1.7 percent of disadvantaged students. The same neighborhood exposure rate as non-disadvantaged neighborhoods in 13 years, economically disadvantaged students would increase the probability of majoring in STEM by (0.058-0.041)*13*4.0=0.88 percentage points, 23% of the gap between those two groups.

Same STEM neighborhood (Campus level). The average neighborhood exposure rate—based

on the proportion of non-moving students choosing to major in STEM—of White students (0.058) is 61% higher than that of Black students (0.036) and 52% more than Hispanic students (0.038). With the same neighborhood exposure rate as neighborhoods of White students in 13 years, Black students would increase the probability of majoring in STEM by (0.058-0.036)*13*4.0=1.14 percentage points, 34% of the gap between Black and White students. Hispanic students would increase the probability of majoring in STEM by (0.058-0.038)*13*4.0=1.04 percentage points, 44% of the gap between Hispanic and White students.

The average neighborhood of economically not disadvantaged students (0.060) is 71% higher than that of disadvantaged students (0.035). The same neighborhood exposure rate as non-disadvantaged neighborhoods in 13 years, economically disadvantaged students would increase the probability of majoring in STEM by (0.060-0.035)*13*4.0=1.3 percentage points, 34% of the gap between those two groups.

Same STEM neighborhood (STEM occupation). The average neighborhood exposure rate—based on the proportion of proportion of STEM workers in a school district—of White students (0.084) is 39% higher than that of Black students (0.073) and 30% more than Hispanic students (0.059). With the same neighborhood exposure rate as neighborhoods of White students in 13 years, Black students would increase the probability of majoring in STEM by (0.084-0.073)*13*2.6=0.37 percentage points, 12% of the gap between Black and White students. Hispanic students would increase the probability of majoring in STEM by (0.084-0.059)*13*2.6=0.85 percentage points, 39% of the gap between Hispanic and White students.

The average neighborhood of economically not disadvantaged students (0.086) is 41% higher than that of disadvantaged students (0.061). The same neighborhood exposure rate as non-disadvantaged neighborhoods in 13 years, economically disadvantaged students would increase the probability of majoring in STEM by (0.086-0.061)*13*2.6=0.85 percentage points, 22% of the gap between those two groups.