5-2019

Exploring Neighborhood Effects and Socioeconomic Background in College Enrollment: A Longitudinal Analysis in St. Cloud, Minnesota

Diego Alejandro Guerrero
St. Cloud State University, daguerrero@stcloudstate.edu

Exploring Neighborhood Effects and Socioeconomic Background in College Enrollment: A Longitudinal Analysis in St. Cloud, Minnesota

by

Diego A. Guerrero

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Applied Economics

May, 2018

Thesis Committee:
Mónica García-Pérez, Chairperson
Kenneth Rebeck
Robert C. Johnson

Abstract

We follow the transition from high school to college and the characteristics of college enrollment from 2009 to 2017 in four cohorts of high school graduates in Saint Cloud Minnesota, using student records from the school district administrative system and the National Student Clearinghouse data on college registration. Residential addresses are geocoded at the census block group level to incorporate neighborhood effects. Logistic model, Two Step Least Squares, and survival analysis are applied to explore the effects of socioeconomic determinants in college enrollment, timing of enrollment and postsecondary education choices. Logistic models fail to reflect neighborhood effects across most specifications. High school grades, sex and family background have robust effects in these models. When GPA is considered endogenous to socioeconomic determinants, findings show neighborhood effects are robust and have a large impact on high school performance and college enrollment. Neighborhood educational attainment, unemployment, and income are strong predictors of enrollment and offset individual characteristics. Racial segregation is insignificant across most specifications. Evidence from survival models suggests that GPA, sex, and socioeconomic background are related to early enrollment. Students with better high school grades are more likely to enroll in 4 Year institutions and less likely to enroll in 2 Year institution, and have lower odds to enroll into For-Profit institutions. Results highlight the importance of neighborhood effects to explain educational outcomes and heterogeneous educational choices. It also stresses dynamic complementarities in education.

JEL classification: J24; I24

Keywords: Educational Gap, College Attendance, Neighborhood Effects, Human Capital.

Acknowledgements

This research has been possible due to financial support by the Pre-College Programs department in St. Cloud State University and the Department of Economics. I wish to express my appreciation to Dr. Robert Johnson and Dr. Monica Garcia-Perez, who led the research on the Access and Opportunity Program that made possible this project.

This is also an opportunity to express thanks to the Faculty at the Department of Economics in SCSU, and to the financial support by the Michael D. White Economic Fellowship. The Faculty has been unique in kindness and warmth. Professor Garcia-Perez, in particular, has my wholehearted respect and gratefulness: she is a professional and human role model.

The author is responsible for errors, interpretations and/or omissions. Views, thoughts, interpretations, and opinions expressed in this text belong solely to the author and not necessarily reflect the institutions, committee, group or individuals in St. Cloud State University.

Table of Contents

List of Tables

List of Figures

1. Introduction

The Minnesota Statewide Longitudinal Data System reports that 71 percent of white students that graduated from high school enrolled in college during 2017 but only 60 percent of students of color had such an achievement. That is a 9 percent point difference yet an improvement compared with an 18 percent point gap in 2007. The achievement gap in college enrollment appears to have decreased in the state. Educational disparities remain, however, as 56 percent of white students met math proficiency in high school against 32 percent of comparable students of color. That gap is 25 percent points in reading tests and 27 percent points in science. There are also differences in college persistence and completion rates. As such, academic performance and achievement disparities are reported across ethnicity and income levels and are subject to policies.

In 2008, St. Cloud State University launched the Access and Opportunity Program (AOP) in partnership with other actors in the school district to improve college readiness standards in minorities and students at risk of underachievement. A thorough evaluation of AOP's impact in high school graduation is found in Garcia-Perez and Johnson (2017).

Continuing efforts to understand the academic achievement gap, we explore college enrollment and characteristics of enrollment across a sample of high school graduates in the St. Cloud school district 742 that follows from AOP's impact analysis. Using rich longitudinal data that follows high school students and their transition from graduation to college, we analyze socioeconomic determinants of college enrollment and explore neighborhood effects as determinant in educational choices. Educational choices cover from enrollment in postsecondary education to characteristics of enrollment: timing after high school graduation and institutional type.

Educational choices are of interest to study labor outcomes, human capital formation, and social inequality. An array of literature focuses on returns to education and years of schooling (Card, 2001; Card & Krueger, 1992) and looks at how returns to skills influence wage gaps and intergenerational mobility (Grogger & Eide, 1995; Juhn, Murphy, & Pierce, 1993). The decision to invest in education is important but characterizing the type of skills developed in college is also of importance. Blair et al. (1981) and Kane and Rouse (1995) focus their interest in returns to schooling across community colleges and the differences among high school graduates and bachelor graduates. In this sense, the educational pathway is not homogeneous and the literature finds relevant to study its heterogeneity and dynamics (Altonji, Blom, & Meghir, 2012; Arcidiacono, Hotz, & Kang, 2012).

Research in economics of education finds that family background is key in explaining academic performance and decision-making (for instance, in college enrollment vid. Black and Sufi (2002)). However, the importance of social interactions and networks is less explored and there is growing interest in its effects (Durlauf, 2004). Neighborhood effects, understood as spillovers and externalities from interactions in a social space, often defined geographically, emphasize the importance of socioeconomic determinants in outcomes and behavior. Empirical analysis show that exposure to highly educated neighbors or better socioeconomic context influences educational aspirations (Sewell & Armer, 1966) and improve labor and educational outcomes (Chetty & Hendren, 2018a, 2018b; Durlauf, 2004).

We explore neighborhood effects and socioeconomic determinants of college enrollment and analyze educational choices in postsecondary education in a sample of 2,332 high school graduates from St. Cloud school district from 2009 to 2017. Merging administrative data from the high school population in the Minnesota Automated Reporting Student System (MARSS), and students' college records in the National Student Clearinghouse (NSC), we construct a longitudinal dataset that follows students in the transition from high school to college. Geocoded residential addresses from high school students, latitude and longitude, is matched with IPUMS National Historical Geographic Information System (NHGIS) socioeconomic data at the census block group level. These neighborhood socioeconomic characteristics are introduced in models controlling for individual level characteristics as proxies for family socioeconomic background.

We use three models to assess neighborhood effects in college attendance. First, regression and logistic models are used to analyze college enrollment as dependent variable. Pooling the data and using timing of enrollment since high school graduation allows us to explore differences in early and delayed enrollment across the sample. Similarly, Two Step Least Squares is implemented to examine correlations between high school GPA and socioeconomic determinants. Finally, survival analysis allows us to consider enrollment as a function of time. Non-parametric, semi-parametric and parametric estimates are compared to analyze timing of enrollment.

We expect that GPA will be the predictor with the largest impact on the likelihood to enroll in college. Our hypothesis is that neighborhood effects, captured by percentile income, unemployment, education level and racial segregation in the census block group will have a large effect explaining college enrollment. Affluent neighborhoods are likely to have more resources available for students to improve their academic performance, thus increasing the likelihood to attend college. Role model effects that positively influence students' aspirations on college attendance could be captured by the proportion of the population with college degrees or more.

Results suggest that neighborhood effects are significant in determining postsecondary education enrollment and high school performance seems to be the mechanism channel of influence. Educational attainment in the neighborhood has a robust impact on high school performance and in the choice between 2 Year institutions (negative relationship) and 4 Year institutions (positive). That is, students living in neighborhoods with larger proportion of the population with college degrees, get better grades in high school and are more likely to go to college, suggesting role model effects.

Racial segregation in the census block group is insignificant across specifications, probably due to lack of variability in the neighborhoods. Neighborhood income is complicated: the percentile rank of median family income is a weak explanatory variable but income per capita percentile rank is robust. Including neighborhood unemployment in the model boosts robustness of income effects. A better understanding of causal mechanism that explains the relationship between neighborhood income and academic achievement, holding everything else constant, would increase our understanding and help disentangle these effects.

Students with better high school grades are less likely to delay enrollment in postsecondary education, more likely to enroll in 4 Year institutions and less likely to enroll in 2 Year institutions, and have lower odds to enroll into for-profit institutions. Remarkably, GPA is not the strongest predictor for attendance into for profit private institutions but socioeconomic proxies and sex explain the decision.

This project adds to the literature on heterogeneous educational choices by including neighborhood effects. It also suggests the relevance of early intervention in students to boost their outcomes, as high school performance remains the strongest determinant of college enrollment. Further research could examine persistence and performance in college, or follow similar methods to study choice of major and persistence in the major. Additionally, this rich dataset may offer opportunities to analyze labor markets and human capital choices.

The document is structured in chapters. Chapter 2 discusses the literature review on neighborhood effects and college enrollment, focusing on educational choices and outcomes in the economic literature. Chapter 3 explains data and methodology used through this research. We present an extensive description of summary statistics and the construction of all variables. In Chapter 4 we present results for all models, first with findings from models on postsecondary enrollment, and continuing with results on 2 Year institutions enrollment, 4 Year institutions enrollment, Public college enrollment, Private college enrollment, and For Profit college enrollment. Finally, we discuss main findings, implications, and further research opportunities in Chapter 5.

2. Literature Review

This chapter reviews literature on neighborhood effects and college enrollment. First we present the theoretical framework and findings in economic research to explain neighborhood effects. Although causal mechanisms and empirical findings are reviewed, we focus on the relationship between neighborhood effects and education. Furthermore, the section details some of the variables used to assess neighborhood effects. Later, we review research on economics of education that explains optimal decisions to schooling with a focus on postsecondary enrollment. This chapter emphasizes educational choices after high school graduation.

Neighborhood Effects

Neighborhood effects are part of an interdisciplinary research agenda to understand social interactions and networks from sociological, psychological, and economic perspectives. Neighborhoods are defined as some proximity in a “social space”, developed in Akerlof (1997) and Jackson (2010), allowing interactions with economic spillovers on behavior. In empirical literature, geographical boundaries defined by administrative agencies are fundamental to assess these effects, particularly residential neighborhoods and school neighborhoods as spaces where spillovers manifest.

The economic literature incorporates interdisciplinary mechanisms to explain neighborhood effects. Psychological factors, information interdependence, and interdependence on costs and benefits may explain development of peer effects, role model effects, and networks. These are imitative types of behavior: a role model with certain education may influence a child to follow similar paths. Sacerdote (2011) defines peer effects as externalities that arise from a given background, behavior, or outcome that affects outcomes from other individuals, and provides evidence on peer effects, though its magnitude across levels of education is debated. Peer effects have been studied in the context of class assignment, alongside role model effects, in Lyle (2007) but also relating neighborhoods to academic decisions (Ainsworth, 2002; Bobonis & Finan, 2009; Calvó-Armengol, Patacchini, & Zenou, 2009).

Social interactions and networks have growing relevance among theories that study social inequality, particularly neighborhood effects. For example, Chetty et al. (2014) uses tax records to study intergenerational mobility across commuting zones, finding correlations between mobility and income inequality, primary school quality, social capital, segregation, and family stability. Neighborhood effects also impact health outcomes, behavioral issues and crime (Glaeser, Sacerdote, & Scheinkman, 1996; Lee et al., 2017; Sampson, Morenoff, & Gannon-Rowley, 2002). From the literature on social interactions, Ioannides and Loury (2004) develop a theoretical framework to analyze informal contacts and network effects in wages and employment.

Racial segregation in neighborhoods may play a key role. Cutler and Glaeser (1997) estimate that reducing segregation by one standard deviation would decrease a third of white-black gaps in high school graduation rates, earnings, employment, and single motherhood. Although the causal mechanism is not yet understood, lower exposure to better role models, studied with an index of exposure in the census tract, may explain these outcomes. Segregation measures use indexes that rely on the proportion of the population in an area compared with the distribution from a larger area. Dissimilarity indexes (Cutler & Glaeser, 1997; Fryer, 2011) and Theil indexes (Chetty & Hendren, 2018b; Chetty et al., 2014; Iceland & Weinberg, 2002) are typical measurements. A different approach is followed in Hellerstein and Neumark (2008), capturing segregation with an exposure index that compares the actual distribution of race with random distributions generated with Monte Carlo Simulation.

Empirical strategies to assess neighborhood effects frequently rely on following children that move to new neighborhoods. The Moving To Opportunity program is an important case study that allowed randomly selected families to move into better localities. Evidence from this program suggests that better neighborhoods increase college attendance and decrease single parenthood (Chetty, Hendren, & Katz, 2016; Kling & Liebman, 2004). A similar approach was used in Chetty and Hendren (2018a, 2018b) designing a quasi-experiment that follows administrative tax records in a large sample and finds long term impacts of neighborhoods on adult earnings, college attendance, fertility, and marriage.

In the study of neighborhood effects, two surveys of literature stand out. Durlauf (2004) presents an extensive overview on neighborhood effects from theoretical to empirical nuances explored in the economic literature. On the other hand, a comprehensive summary of results and challenges, including results from the Moving to Opportunity experiment, is found in Sampson et al. (2002).

Regarding academic outcomes, neighborhood effects on education are a topic of interest. For instance, controlling for abilities, family background and schooling, Garner and Raudenbush (1991) find that neighborhood deprivation is negatively associated with educational attainment. Deprivation is measured as an index at the census tract equivalent in Scotland, weighting unemployment rate, youth unemployment, single parent families, earnings, overcrowding, and health. Educational achievements are also analyzed in Ainsworth (2002) using longitudinal data and finding that neighborhood effects offset family background. In this case, neighborhood variables include the proportion of college graduates in the population, the proportion of professionals, householders that lived in the same house for the last five years, unemployment, poverty rate, and a measurement of racial/ethnic diversity based on the sum of squared proportions of each race.

Sewel and Armer (1966) examine students' college aspirations –self reported intention to enroll– as determined by neighborhood socioeconomic status, controlling by sex, percentile level of socioeconomic status and percentile rank of abilities. The study is based on a random sample of public school students in Milwaukee Metropolitan Area in 1957. Students' socioeconomic status is defined by the parents' educational level, estimated funds to attend college, and approximate wealth. For neighborhoods, the unit of analysis is high school enrollment districts –an approximate combination of census tracts–; and neighborhood socioeconomic status measured with a categorical variable on the proportion of males in white-collar occupations. Correlations and conditional probabilities are analyzed to conclude that neighborhoods associate with educational aspirations in females rather than males. An issue that was not addressed but we do account for, is the correlation between explanatory variables: that is, the possibility that neighborhoods and socioeconomic status determine individual abilities.

Several variables are frequent in the literature, sometimes constructing one or more weighted indexes. First, there are measurements of economic performance in the area using earnings, employment, and the ratio of professional occupations. Also, some indexes use educational advantages or educational exposure to educated role models with college education or more. Segregation has been associated with academic outcomes (Cutler & Glaeser, 1997) and integrated as part of neighborhood effects (Chetty & Hendren, 2018b; Chetty et al., 2014;

Entwisle, Alexander, & Steffel Olson, 1994). Fryer (2011), however, argues that peer dynamics and identity models are more consistent explaining academic achievements than geographic segregation. Finally, research also uses measurements of family stability such as proportion of single-parent households.

College Enrollment

The framework of human capital models is built on Becker (1964) where individuals optimize schooling by maximizing life-time earnings against the costs of education, including foregone income during the years of education and tuition costs. Estimating returns to education is a major part of research on economics of education (Card, 2001; Card & Krueger, 1992) but there is more to analyze on schooling decisions.

Returns to college and type of degree has been a subject of debate. For instance, there exists attempts to measure college characteristics and earnings (James, Alsalam, Conaty, & To, 1989). Specifically, Blair et al. (1981) estimates that, across employed technicians, the return to an Associate degree is 13.9 percent larger than high school graduates without a degree, without accounting for any additional schooling after receiving the 2 Years degree. More recent research found community colleges have a 10 percent return larger than no college education, and similar estimated returns per credit than a Bachelor's degree (Kane & Rouse, 1995).

The decision to attend college has received much attention and research points out the importance of family background and race. The literature has a wide scope: the determinants of the decision, earnings, returns to different majors, racial disparities, and policy impact are all studied (Angrist, Autor, Hudson, & Pallais, 2016; Black & Sufi, 2002; Cameron & Heckman, 2007; Catsiapis, 1987; Hauser, 1993a). Kane (1994) and Hauser (1993b) use time series on enrollment rate using the Current Population Survey (CPS) to analyze the effects of college tuition and family background. With the same cross-section, Black and Sufi (2002) document behavioral differences in college enrollment across races. Meanwhile, Cameron and Heckman (2007) analyze race and educational attainment as influenced by family background, income, tuition costs, and cognitive abilities, using the National Longitudinal Survey of Youth (NLSY). Hyman (2018) notes that the probability to enroll in college is susceptible to nudges and policies, but students with low high school GPA are less likely to persist through college, pointing out that individuals with low performance require support and mentorship through college.

The use of cross-sectional and longitudinal surveys allowed research to omit nuances on the definition of college enrollment using educational attainment as the variable that defines college attendance. Educational attainment is a self-reported variable that shows if the individual attended or completed a degree by asking the last year of education completed. Administrative data, however, allows to study college enrollment going beyond attendance and self-reported answers. However, this has nuances: for example, the National Student ClearingHouse (NSC) tracks and reports college enrollment rates, but faces issues estimating enrollment in a given term as institutions have flexibility in determining periods. The NSC uses a set of brackets on the start and end dates of records to construct enrollment during a given period (National Student ClearingHouse Research Center, 2018).

Traditional college enrollment measurements consider students as enrolled in postsecondary education 2 years after high school graduation (Cabrera & La Nasa, 2001; Perna, 2000; Rowan-Kenyon, 2007). This approach implies that college enrollment is considered only among the population that is eligible to attend after completing previous degrees. Additionally, it highlights the importance of timing, given that students may enroll in any time of their life but delayed enrollment has consequences. Horn and Carroll (1996) find that non-traditional students have larger risk of dropping out of college, and Bozick and DeLuca (2005) document that 16 percent of high school graduates delay enrollment by more than 7 months and these students are more likely to attend 2 year institutions and drop out of college. Low-income individuals, GED recipients, and students that delay enrollment tend to enroll in public 2 year institutions (Horn & Carroll, 1996).

In this sense, it is relevant to consider human capital formation choices to be dynamic and heterogeneous. Transition from high school to college and the labor market provides many choices to individuals. Some students engage in education linearly (K-12, 4 Year degree enrollment and graduation, and then enter the labor market) but others take diverse paths: they may enter in the labor force and never continue their education; entering college may be done immediately after graduating or delayed; and post-secondary options are diverse in types of degree, major choices, quality, and more. Furthermore students may enroll and update their choices any time.

Most literature treats schooling as homogeneous but heterogeneity in human capital decisions is the topic of some research. Altonji et al. (2012) surveys the literature on heterogeneous educational choices. It also develops a theoretical model of educational and occupational choice that allows for preferences to be updated and wage uncertainty, and stresses the relevance of understanding types of human capital investment. Altonji (1993) presents empirical evidence on the probability to attend a 2 Year degree, Less than 2 Year degree, and career choices, and the effects of ex-post and estimated ex-ante labor outcomes, family background, and high school curriculum on this decision. Movement from community colleges to 4-Year degrees is analyzed in Doyle (2009) using semi-parametric survival models, and finds that students that attended to 2-year degrees are less likely to complete a Bachelor degree in time, after correcting for self-selection issues. Furthermore, the literature considers career choices and transfers across major, suggesting that these choices are influenced by expected earnings and individual abilities (Arcidiacono, 2004; Arcidiacono et al., 2012; Keane et al., 1997). Almost all of this literature consists on structural models with discrete choice dynamics and empirical models with OLS or IV approaches.

Research on educational outcomes is not unfamiliar with survival analysis. With this methodology, economic research studies drop-out rates and persistence (Bahi, Higgins, & Staley, 2015; Mangold, Bean, Adams, Schwab, & Lynch, 2002; Min, Zhang, Long, Anderson, & Ohland, 2011; Murtaugh, Burns, & Schuster, 1999) and college graduation (Chimka & Lowe, 2008; Ishitani, 2006; Juan Carlos, Crosta, Bailey, & Jenkins, 2006; Laugerman, Ii, Rover, & Mickelson, 2015). Regarding college enrollment, Ligh (1995) analyzes dynamics of enrollment in a sample of males from the NLSY using state-level tuition costs and foregone wages, while Bozick and DeLuca (2005), as previously noted, study the impact of timely or delayed enrollment as predictor in the probability to graduate with a bachelor degree.

Our work adds to the literature on college attendance and heterogeneity in human capital formation analyzing socioeconomic determinants and neighborhood effects on educational choices. We also expand on Garcia-Perez and Johnson (2017) evaluation on high school graduation and college readiness in the school district, which used propensity scores in a panel of high school students to evaluate program effects in retention rate and graduation. Following that study, our project merges new data sets to further characterize students and analyze socioeconomic determinants that explain the transition of students from high school into college, the timing of such decisions, and educational choices in the institution and degree pursued.

3. Data and Methodology

This chapter describes the data and methodology used to analyze college enrollment. Three datasets were combined in this project: Minnesota Automated Reporting Student System (MARSS), National Student Clearinghouse (NSC), and IPUMS National Historical GIS (NHGIS). As different sources are involved, we present two sections: first, a detailed description of all relevant variables and trends across them; next, a description of models estimated to explain college enrollment.

The St. Cloud State University Pre-College Programs has a history of cooperation with School District 742, servicing students to increase college readiness and academic achievement in St. Cloud. The Center for Access and Opportunity had the opportunity to access the School District data, MARSS, with rich individual-level information on the population of high school students, including their ethnicity, program participation, English proficiency, residential address, and some performance records. Names were de-identified to protect privacy. MARSS was previously used in Garcia-Perez and Johnson (2017) to evaluate high school graduation in the school district.

The NSC stores national records on postsecondary education. Data was requested by Pre-College Programs, matching full names and birthdates from senior students. Each student may have no record found or at least one record, including time periods, programs, and institutions.

By merging MARSS and NSC into a longitudinal panel, with yearly frequency, we follow the transition from high school to post-secondary education. Five cohorts, from 2009 to 2013, are defined by the period of high school graduation. Meanwhile, NSC data extends to early 2017. That is, the later cohorts are followed by 5 periods while early high school graduates had 9 years to enroll in college. Periods are defined as Academic Years from August to July. Furthermore, student's address were geocoded using ArcGIS, matched with a Census Block using the Federal Communications Commission API, and finally combined with NHGIS. NHGIS provides harmonized data from the 5-Year American Community Survey (ACS) on socioeconomic characteristics. We use the 5-Year ACS from 2008-2012 as it most closely matches periods when students' from the 2009-2013 cohorts were likely to live in the address provided before attending college. The 2009 cohort is dropped from the sample, as there is no address data available.

Table 3-1.

Summary Statistics in the School District Senior Student Population.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
LEP	4,005	0.112	0.316	0	1
Female	4,005	0.478	0.500	0	1
FRL	4,005	0.430	0.495	0	1
HS Graduated	4,026	0.775	0.418	0	1
4-YR Cohort HS Graduated	4,026	0.755	0.430	0	1
SOC	4,005	0.243	0.429	0	1

Among the population of 4,030 senior students in the school district, described in Table 3-1, only 77.5 percent graduated from high school. There are missing values for 25 students. To keep consistency, following García-Pérez and Johnson (2017), high school graduation is defined as students that graduated high school in the same 4-year cohort. This represents the most traditional students that remained in the school district during high school and advanced every grade in a timely manner. Under this definition, 75.5 percent of seniors achieved high school graduation.

The proper definition of graduation is fundamental and limits the sample of students able to attend college. In our data, this sample consists of 2,332 students. In the next section we will describe the variables extracted from the datasets and the definitions used to construct our analysis. Finally, the econometric models estimated are described, emphasizing a pooled cross-sectional approach and a longitudinal analysis using survival functions.

Variables

There are three sets of variables: dependent variables, individual characteristics, and the neighborhoods socioeconomic characteristics. The dependent variable is college enrollment, defined as a function of time and type of institution. Explanatory variables are either individual or neighborhood variables. The former is constructed with MARSS, and consists of ethnicity, sex, and proxies for socioeconomic background. The latter is a set of socioeconomic characteristics in the census block group the individual lives in, such as median income, educational attainment, ethnicity, and employment rate. We explain the construction of variables and provide summary statistics.

College enrollment. Enrollment is defined as a record found in NSC, lasting more than 30 days, by a 4-Year-Cohort High School graduate during the academic year beginning in August and ending in July. This definition excludes records from students that did not graduate from high school or were part of Post-Secondary Enrollment Option (PSEO). The 30-day period excludes records with short duration. The NSC (2018) uses alternative definitions of college enrollment, for example Fall Enrollment includes a record of any length during Fall but fails to account for students enrolled in Spring and Summer. We study enrollment on a yearly basis instead of a higher frequency because institutions in the U.S. are not homogenous in their programs: some may be quarterly, yearly, or hour-based programs.

Four types of enrollment are defined based on time since high school graduation: 1 Year Enrollment follows students that enroll at least one year after high school graduation, 2 Year Enrollment, 4 Year Enrollment, and Ever Enrollment. Ever Enrollment shows if individuals are observed in post-secondary education at some point, independent of time, and is a traditional measure that follows from the U.S. Census Bureau (Carter & Wilson, 1994).

Enrollment is a comprehensive measure encompassing post-secondary enrollment, but we also characterize it by the type of institution and program. NSC allows to discern public and private institutions, 4 Year institutions, 2 Year institutions, and Less than 2 Year institutions. Additionally, by matching institutions with the Carnegie Classification, private institutions are classified in either non-for-profit private or private for-profit.

Table 3-2.

Summary Statistics Enrollment in Sample.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ever Enrollment	2,332	0.751	0.432	0	1
1YR Enrollment	2,332	0.673	0.469	0	1
2YR Enrollment	2,332	0.713	0.452	0	1
4YR Enrollment	2,332	0.746	0.436	0	1

Summary statistics for different enrollment measures are provided in Table 3-2 to Table 3-8. Our sample consists of 2,332 students after dropping non-high school graduates and observations lost due to missing values. It must be highlighted that, given the proposed definitions, mean enrollment is expected to increase as the time window is expanded: the more time students had to enroll, it is likely there will be more enrollment. Indeed, Table 3-2 shows college enrollment among high school graduates, 67 percent of the sample enroll within the first year after high school graduation and 75 percent are enrolled at any point in time.

Table 3-3.

Summary Statistics Enrollment in Less than 2 Year Institutions.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ever Enrollment	2,332	0.00343	0.0585	0	1
1YR Enrollment	2,332	0.00472	0.0685	0	1
2YR Enrollment	2,332	0.00515	0.0716	0	1
4YR Enrollment	2,332	0.00600	0.0773	0	1

Table 3-4.

Summary Statistics Enrollment in 2 Year Institutions.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ever Enrollment	2,332	0.350	0.477	0	1
1YR Enrollment	2,332	0.221	0.415	0	1
2YR Enrollment	2,332	0.273	0.446	0	1
4YR Enrollment	2,332	0.328	0.470	0	1

Table 3-5.

Summary Statistics Enrollment in 4 Year Institutions.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ever Enrollment	2,332	0.558	0.497	0	1
1YR Enrollment	2,332	0.497	0.500	0	1
2YR Enrollment	2,332	0.518	0.500	0	1
4YR Enrollment	2,332	0.549	0.498	0	1

Table 3-3 to Table 3-5 show enrollment by institutional type: Less than 2 year institutions, 2 year institutions, and 4 year institutions. The former has very small instances: less than 1 percent of the sample is found at any point in this kind of programs. But 2 year institutions and 4 Year institutions are more frequent: up to 35 percent of enrollment and up to 56 percent, respectively. Furthermore, the possibility exists for students to enroll in more than one degree at a time and withdraw or graduate to continue in another kind of institution.

Table 3-6.

Summary Statistics Enrollment in Public Institutions.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ever Enrollment	2,332	0.689	0.463	0	1
1YR Enrollment	2,332	0.602	0.489	0	1
2YR Enrollment	2,332	0.643	0.479	0	1
4YR Enrollment	2,332	0.678	0.468	0	1

Table 3-7.

Summary Statistics Enrollment in Private Non-Profit Institutions.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ever Enrollment	2,332	0.115	0.320	0	1
1YR Enrollment	2,332	0.0943	0.292	0	1
2YR Enrollment	2,332	0.102	0.302	0	1
4YR Enrollment	2,332	0.107	0.309	0	1

Table 3-8.

Summary Statistics Enrollment in Private For-Profit Institutions.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ever Enrollment	2,332	0.0334	0.180	0	1
1YR Enrollment	2,332	0.0180	0.133	0	1
2YR Enrollment	2,332	0.0227	0.149	0	1
4YR Enrollment	2,332	0.0317	0.175	0	1
					.

Private Non-Profit, For Profit, and Public enrollment are presented in Table 3-6 to Table 3-8. For Profit institutions are private but not every private is for profit; thus we define private institutions as non-for profit alone. Enrollment in public institutions comprises from 60 percent to 67 percent of the sample.

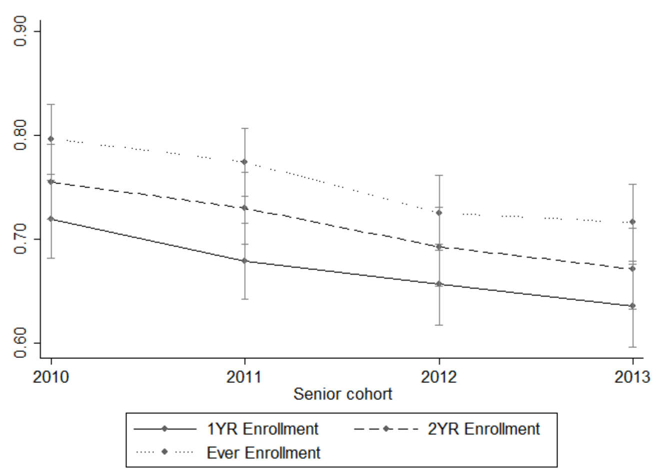


Figure 3-1. College enrollment, 2010-2013.

Using year of high school graduation, we observe trends in enrollment per time window. *Figure* 3-1 suggests that college enrollment is decreasing across cohorts in the school district. Enrollment during the first year after graduation fell almost 10 percent points, and Ever-Enrollment fell approximately 5 percent points.

Individual characteristics. MARSS provides individual characteristics of the sample, exhibited in Table 3-9. White students compose 80 percent of the sample, and among students of color (SOC) the majority is Black or African American. Following Garcia-Perez and Johnson (2017), Free and Reduced Lunch (FRL), a program that attends low income population, serves as proxy for socioeconomic status; and Limited English Proficiency (LEP) is a proxy for immigrant background. If we consider SOC, FRL, or LEP as an indicator of risk of academic underachievement, the subsample represents approximately two fifths of the sample.

Enrollment differences by sex and ethnicity are documented in A. College Enrollment, by Race and Sex. In every measure, females are more likely to enroll than males, and these differences are statistically significant in a two-group t-test. In fact, female enrollment is 10 percent points larger than male enrollment and it holds from 1 year enrollment to ever enrollment. When it comes to racial disparities, white students outperform SOC in almost every measurement, and college enrollment for whites is also 10 percent points larger. It is interesting that this pattern is not absolute among every kind of program and institution: whites are 5 to 10 percent points less likely to attend 2 Year institutions.

Table 3-9.

Summary Statistics, Students.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
Ethnicity	2,332			0	4
White	2,332	0.796	0.402	0	1
Black	2,332	0.131	0.338	0	1
Asian	2,332	0.0403	0.196	0	1
Hispanic	2,332	0.0244	0.154	0	1
Native	2,332	0.00686	0.0825	0	1
LEP	2,332	0.101	0.301	0	1
Female	2,332	0.489	0.500	0	1
FRL	2,332	0.380	0.486	0	1
Cohort	2,326			2,010	2,013
2010	2,326	0.237	0.425	0	1
2011	2,326	0.2704	0.444	0	1
2012	2,326	0.248	0.432	0	1
2013	2,326	0.242	0.428	0	1
SOC	2,332	0.203	0.403	0	1
CGPA	2,099	3.099	0.754	0.342	4.656

High School performance. Academic performance is a measure frequently explored in the literature as it relates with college enrollment (Angrist et al., 2016; Hyman, 2018; Lyle, 2007). Data on cumulative GPA is available for 2,099 students, following a near-normal distribution with mean 3.099 and standard deviation 0.754 exhibited in Table 3-10 and Figure 3-2. Though academic performance is expected to be the strongest predictive variable in college enrollment, it introduces some issues: students without data available are expected to be self-selecting not to attend college. In total, 10 percent of the sample has missing values in GPA.

Table 3-10.

Summary of Cumulative GPA.

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
CGPA	2,099	3.099	0.754	0.342	4.656

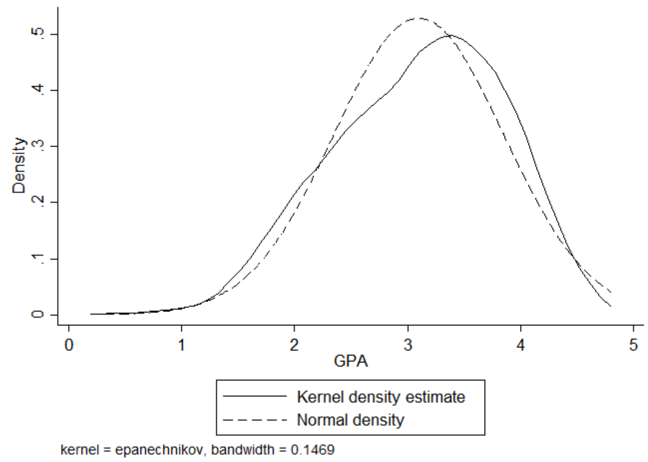


Figure 3-2. Distribution of Cumulative GPA.

Cumulative GPA in the sample ranges from 0.342 to 4.656. These values are odd in the sense that represent students with underperformance and questions remain on the nature of values over 4 (the typical maximum grade point average in the U.S.). We avoid replacing these values because the distribution shows that a small part of the sample has outlier grades.

Table 3-11 presents missing GPA values and its correlation with individual characteristics of students. It shows that although correlation is negative for FRL and LEP status, it is not a strong relationship. Furthermore, there is no correlation between missing GPA and sex.

Table 3-11.

Correlation between Missing GPA and Socioeconomic Characteristics.

	FRL	LEP	Female	Has GPA
FRL	1			
LEP	0.35	1		
Female	0.02	-0.03	1	
Has GPA	-0.16	-0.01	0	1

Regarding the relationship between cumulative GPA and socioeconomic status, Figure 3-3 to Figure 3-5 show that high school performance may be influenced by socioeconomic characteristics. If GPA is suspected to determine college attendance, socioeconomic variables may have an effect on postsecondary education through high school performance. This correlation between explanatory variables implies that regressions will be biased, and the effect of explanatory variables will be captured by the correlated variables. In particular, Figure 3-3 shows that students participating in FRL program underperform compared to those from higher socioeconomic status, while white students have better high school performance than students of color, and the GPA distribution of LEP students (proxy for immigrant status) implies lower academic performance than traditional students.

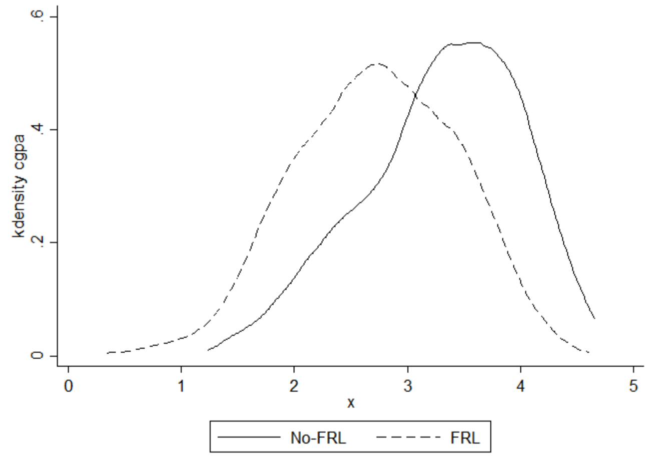


Figure 3-3. Distribution of GPA, by FRL.

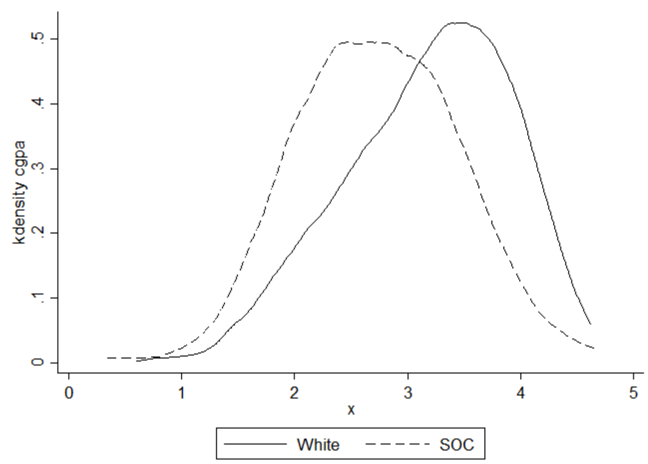


Figure 3-4. Distribution of GPA, by ethnicity.

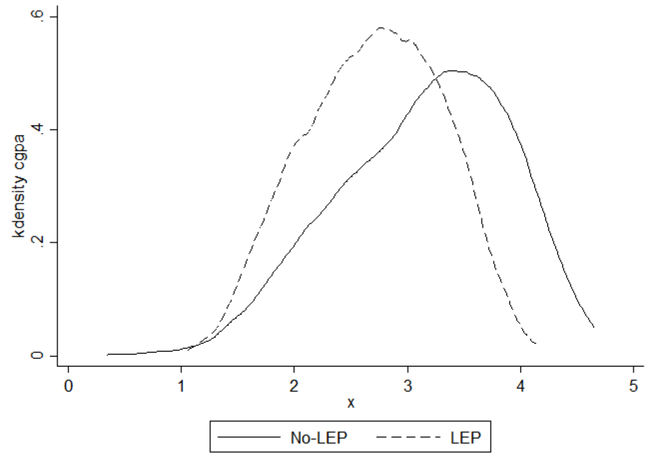


Figure 3-5. Distribution of GPA, by LEP.

Neighborhoods. Geographic definitions of neighborhoods, ranging from counties to census tracts, are common units of analysis (Chetty & Hendren, 2018a, 2018b). MARSS dataset allows us to define residential neighborhoods under the narrow definition of census block groups. Residential addresses were geocoded using ArcGIS in Python 3.6, and latitude and longitude matched with a census block group using publicly available data from the Federal Communications Commission API. Students are located in 104 block groups in Minnesota, although some outliers registered addresses in Missouri and Wisconsin; not surprisingly, 95 percent of the population is located in Sherburne, Benton and Stearns, with 82 percent of the sample concentrated in Stearns County. The distribution of students in block groups in Minnesota is observed in Figure 3-6 and a detailed distribution in St. Cloud counties is available in C. Geographic Distribution, by County.

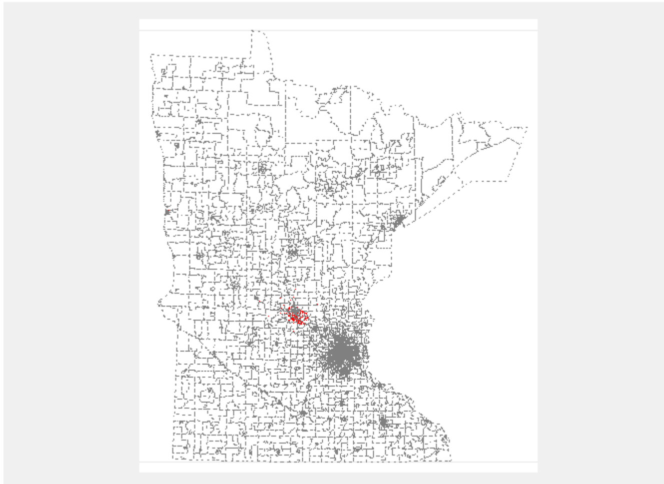


Figure 3-6. Distribution of students in Minnesota.

Table 3-12.

Summary Statistics, Neighborhoods.

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Population	104	1,595	735.9	480	3,620
Family Income, Median	104	66,003	23,715	2,499	151,413
Income per capita	104	25,668	7,777	6,494	46,457
Median Rent	98	694.9	197.8	280	1,302
Median Value	102	167,466	52,526	11,900	301,200
Proportion White	104	0.894	0.141	0.0600	1
Unemployment	104	0.0576	0.0367	0	0.187
Education Low	104	0.104	0.0777	0	0.430
Education Mid	104	0.549	0.0987	0.226	0.721
Education High	104	0.357	0.119	0.118	0.739

Every census block group has socioeconomic characteristics retrieved from the 5-Year ACS using IPUMS NHGIS and merged into the data. Income variables shows that the average block group has a median family income of USD 66,003 while mean per capita income is USD 25,668. Data also includes the median value and median rent of houses in the area. These income variables highly correlate with each other, so we choose to use median Family Income as our main neighborhood income variable and other variables serve to test robustness. Ainsworth (2002) proposes to use the proportion of college graduates among adults over 24 years old as a measurement of neighborhood advantage. We follow this idea by considering three mutually exclusive categories of education level: low education is the proportion of population without a High School or GED Diploma; Medium education considers the population with High School or GED degree, and some college experience; and, finally, High education includes population with an Associate's Degree or more. The unemployment rate was constructed with the proportion of unemployed over the whole population 15 years and over.

Income variables are redefined at percentile levels: larger ranks represent highest income in the sample. *Figure* 3-7 and *Figure* 3-8 display enrollment by median family income rank in their neighborhood. Chetty et al. (2014) points out that income measurements in percentile ranks has statistical advantages, thus, as suggested in Chetty and Hendren (2018b), we use three percentile rank categories at the bottom 25th, inclusive; top 75th, and a middle-income between the 25 and 75 percentile rank. The gap in enrollment based on income decreases as the time window is expanded from 1-year enrollment to ever-enrollment. This gap decreases by increasing enrollment in low-income neighborhoods. That is, students in higher income neighborhoods outperform students with a low socioeconomic background.

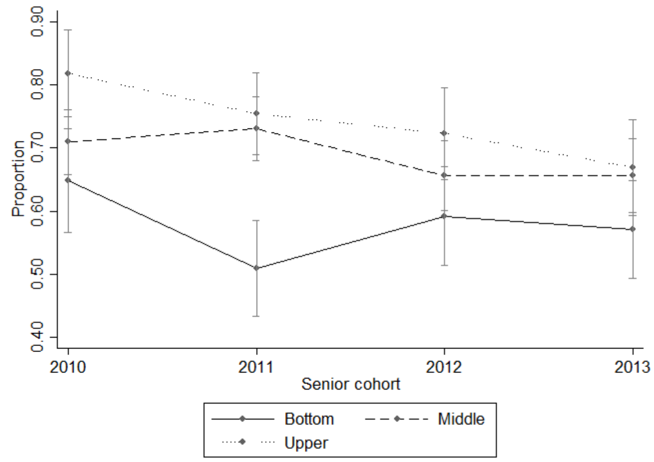


Figure 3-7. 1-year enrollment, by median family income percentile.

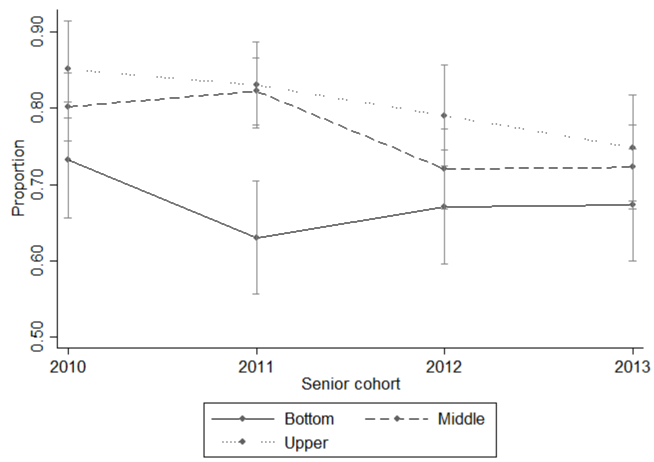


Figure 3-8. Ever enrollment, by median family income percentile.

Enrollment in 2 Year institutions, 4 Year institutions, Public and Private Institutions at each income level is exhibited at Appendix B. College Enrollment, by Neighborhood Income Level. Interestingly, 2 Year institutions do not attract students from high income background but students from low income neighborhoods are more likely to Ever Enrollment in this kind of institutions. Meanwhile, a significant disparity in enrollment into 4 Year institutions by income level of the neighborhood is also observed.

Segregation. Racial segregation and its effects on economic outcomes is also analyzed, following the literature (Chetty & Hendren, 2018b; Chetty et al., 2014; Cutler & Glaeser, 1997; Iceland & Weinberg, 2002). Measures of segregation commonly use a Theil Index to assess concentration or diversification, but this index aggregates deviations in the racial distribution at a geographic level (census tracts, usually) compared with a larger area (metropolitan areas or commuting zone). That approach is not applicable to our analysis that concentrates in a limited area. Instead, following Hellerstein and Neumark (2008), an exposure index is created to compare the observed racial distribution in each neighborhood with Monte Carlo simulations of racial distribution. That is, we randomly distribute the population across neighborhoods. The order in which the population is assigned to each neighborhood is based in a normal distribution, while the probability of assigning a non-white person to the area is a binomial distribution with mean the proportion of people of color in the population, limited to the total population actually living in that area. The simulation is performed 100 times and we estimate the proportion of non-white population per neighborhood.

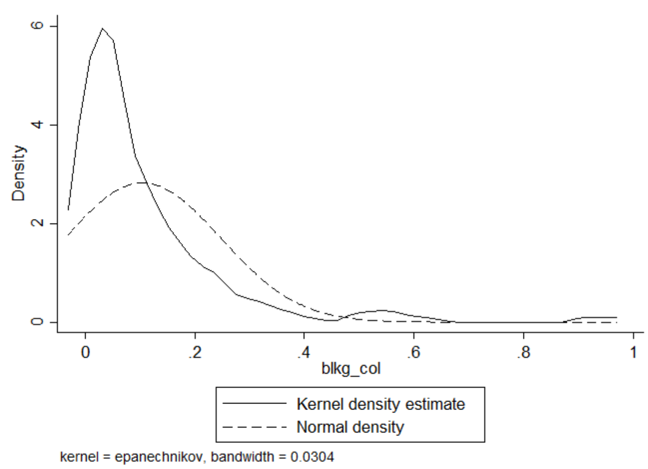


Figure 3-9. Distribution of observed proportion of population of color in neighborhoods.

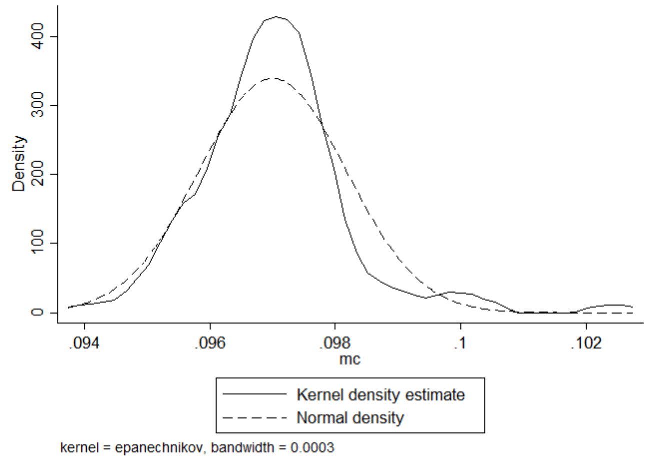


Figure 3-10. Distribution of simulated proportion of population of color in neighborhoods.

Figure 3-9 and *Figure* 3-10 display density plots with the proportion of non-white people observed *vis a vis* a hundred Monte Carlo simulations. Observed distribution is left-skewed, as people of color do not tend to live in traditionally white populated neighborhoods. Large tails suggest segregated neighborhoods with over 50 percent non-white population. The simulation exhibits a normal distribution: no neighborhood is totally-white and there does not exist a larger than 11 percent concentration of non-white population.

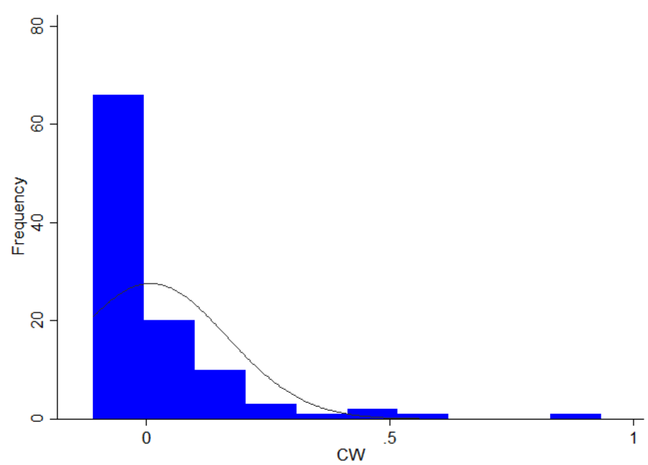


Figure 3-11. Distribution of segregation in census block groups.

To compare both distributions the exposure index CW subtracts the average of white population living in the neighborhood W_C from the proportion of non-white C_C, $CW^0 = C_C - W_C$ measures the observed exposure. If there is perfect segregation, C_C equals 1 and W_C is zero. $CW^0 = 1$ represent a neighborhood where the whole population is non-white. Next, we estimate random segregation $CW^R = C_R - W_R$ and, finally, calculate the exposure index:

$$CW = \frac{CW^0 - CW^R}{1 - CW^R} \tag{1}$$

The index represents the scale to which actual segregation deviates from maximum expected random segregation. Our study differs from Hellerstein and Neumark (2008) as we use proportion of population instead of fraction of an individual's co-workers of specific race. It can be shown that CW is equivalent to $\frac{C_C^0 - C_C^R}{1 - C_C^R}$. Furthermore, the case $CW^0 \leq CW^R$ is unexplored by Hellerstein and Neumark (2008) but it happens in our application, representing census block groups with very high concentration of white population (over 0.9).

Correlations. It is important to observe correlations between independent and dependent variables, thus they are presented in Table 3-13. Main variables of interest were selected: enrollment, income variables, GPA, race, and education level in the neighborhood. Over 0.35 positive correlation between cumulative GPA and college enrollment suggests the variable may be fundamental to analyze college enrollment. High school performance, however, also shows correlation with socioeconomic variables in neighborhoods and individual levels. Forty one percent correlation between FRL and students of color is also relevant: that race appears to be associated with income, a trend not unknown in the U.S. The segregation index has a positive correlation with individual proxies for socioeconomic status, immigrant family, and student of color: the former relates to racial correlation with income, while correlation with immigrant families and students of color may be related to the ethnic composition of immigrants.

Table 3-13.

Correlation Matrix.

	Ever Enrollment	1YR Enrollment	2YR Enrollment	4YR Enrollment	Female	FRL	LEP	SOC	CGPA	Segregation	Family Income, Median	Unemployment	Education High
Ever Enrollment	1												
1YR Enrollment	0.81	1											
2YR Enrollment	0.89	0.91	1										
4YR Enrollment	0.97	0.84	0.92	1									
Female	0.11	0.12	0.12	0.11	1								
FRL	-0.18	-0.2	-0.19	-0.18	0.02	1							
LEP	-0.06	-0.06	-0.05	-0.06	-0.03	0.35	1						
SOC	-0.07	-0.09	-0.08	-0.06	-0.02	0.41	0.64	1					
CGPA	0.35	0.4	0.38	0.35	0.2	-0.37	-0.18	-0.23	1				
Segregation	-0.08	-0.09	-0.08	-0.08	-0.03	0.29	0.33	0.32	-0.18	1			
Family Income, Median	0.12	0.13	0.12	0.13	0.02	-0.37	-0.24	-0.28	0.27	-0.53	1		
Unemployment	-0.08	-0.08	-0.07	-0.08	0.01	0.12	0.09	0.06	-0.1	0.37	-0.3	1	
Education High	0.09	0.09	0.09	0.09	0	-0.22	-0.09	-0.1	0.18	-0.25	0.65	-0.16	1

Model

In this section we explain the models selected to analyze college enrollment. First, we follow a cross-sectional approach by estimating the likelihood of enrolling in college during a given time period. This framework omits the full longitudinal property of the data but provides a reliable approach using Linear Probability Model, Two-Stage regression, and Logistic models. Next, we explain the use of non-parametric, parametric, and semi-parametric Survival models that thoroughly analyze the timing of the event and the effects of variables in explaining such probabilities.

Pooled Cross Section. Linear and logistic regressions in a quasi pooled cross-section analyze differences in college enrollment. Although the panel follows students through years, we pool observations by using the time window of the outcome variable resembling a traditional cross-section analysis. Given a time period after high school graduation T, any ith individual has an enrollment dummy variable E_i^T, a vector of individual characteristics X_i and socioeconomic characteristics of the neighborhood N_i. The Linear Probability Model follows the equation

$$E_i^T = \beta + \gamma X_i + \alpha N_i + \epsilon_i \tag{2}$$

While Equation 2 shows the Logistic model,

$$P(E^T = 1)_i = F(\beta + \gamma X_i + \alpha N_i + \epsilon) \tag{3}$$

The vector of individual characteristics includes cumulative GPA, sex, ethnicity and cohort effects, as well as proxies for socioeconomic status represented by FRL and LEP. Neighborhood effects are captured by income, racial composition and education level in the census block group.

Aforementioned correlations between explanatory variables and GPA make viable a Two Step approach to explore the impact of socioeconomic characteristics in high school performance and, later, the effect in college enrollment. The proposed two-stage model is represented in 4 and 5, where the first regression is

$$GPA_i = \beta + \gamma X_i + \alpha N_i + \mu \tag{4}$$

While the second stage regression analyzes college enrollment as explained by GPA,

$$E_i^T = \lambda + \pi GPA_i + \epsilon \tag{5}$$

Similar results are expected in OLS and Logistic methods. Equation 2, however, poses interpretative difficulties as most explanatory variables are dichotomous; an issue solved by logistic regressions when estimating odds-ratios. Equations 4 and 3 have similar problems to 1, given that the Two-Stage method is also a least squares procedure, but it successfully captures endogeneity that is omitted and violates OLS and logistic models. Henceforth, comparing results will allow a thorough analysis.

Survival Analysis. Survival analysis uses longitudinal properties in the data to explore time after graduation that students take in college enrollment. This analytical framework follows the probability density function of an event across periods. Non-parametric models estimate the probability of an event happening in a given period, while parametric and semi-parametric analysis studies the conditional probability subject to a set of explanatory variables. This section will explain these three approaches using college enrollment as event and yearly time after graduation as periods.

The survival function is a measure of the cumulative density function as a function of time. That is, let t represent time of failure (when enrollment happens), $f(t)$ is the probability density function of enrollment, and $F(t) = P(T < t) = \int_{\infty}^{t} f(x)dx$ is the cumulative density function of the distribution (the number of known enrollments at time t over the total sample). The survival function is expressed as the complement of the cdf,

$$S(t) = P(T > t) = \int_{t}^{\infty} f(x)dx = 1 - F(t) \tag{6}$$

Equation 6 is known as Kaplan-Meier estimator, usually represented graphically. From this estimator, we can derive the hazard function or failure rate: the probability over the survival function, interpreted as the instantaneous rate at which the subjects experience the event given that it has not happened up to period t. Hazard functions are also non-parametric and shown graphically with Nelson-Aalen plots following the equation:

$$\lambda = h(t) = \frac{f(t)}{S(t)} \tag{7}$$

Besides non-parametric estimates, parametric and semi-parametric estimated by maximum likelihood are of interest and allow to assess the effect of explanatory variables. Parametric survival models assume a functional form to the pdf. Cox proportional hazard or semi-parametric does a similar function without assumptions on the pdf. In our case, it is arguable that enrollment on time follows an exponential distribution, as more time after high school graduation will likely allow students to take a professional path. Under this assumption we will estimate the coefficients in

$$h\{(t), (X_i, N_i\} = h_0(t) * \exp(\gamma X_i + \alpha N_i) \tag{8}$$

Here, γ shows the increase in the probability of enrollment by each unit increase in the individual-level socioeconomic variables. Likewise α estimates the increase in the probability of enrollment followed by marginal changes in the neighborhood.

In this regard, contrasting cross-sectional and longitudinal analysis in this data should allow fulfillment of two goals. On one hand, estimating parameters of individual and neighborhoods characteristics will show the magnitude and relevance of background in high education attendance and decisions. On the other hand, our time-window approach and survival analysis will explore delayed enrollment in college and educational choices.

4. Results

In this chapter we analyze results from regressions and test robustness. First, we explore estimates on college enrollment models, both regressions and Survival Analysis. In the same manner, the next sections analyzes enrollment in 2 Year institutions and 4 Year institutions, and enrollment in public, private not-for profit and for profit institutions. All models use robust standard errors. Analysis in the first section is detailed in a broader extent than that in the following sections, as the model specification remains similar but the dependent variable is changed to specific instances of enrollment.

Findings highlight differences in postsecondary enrollment and differences across type institutions. In particular, GPA is robust and consistently explains different career and university choices. Results also highlight that individual and neighborhood effects explain high school performance. These effects are likely to influence college attendance through their impact during high school. Educational attainment in the neighborhood, in particular, is robust to individual socioeconomic proxies to explain GPA. Further robustness tests follow in D. OLS Regression. Survival Analysis explore college attendance through time. High School performance continues to have strong effects explaining time of enrollment. Furthermore, some key socioeconomic characteristics explaining the type of institution are pointed out. GPA and neighborhood exposure to high educational attainment have a negative relation with the probability of attending 2 Year institution2 Year institutions, but positive relationship to 4 Year institution4 Year institutions. Attendance in public, private non-for profit, and for profit institutions is also explored, and findings suggest that females are more likely to attend for profit institutions while GPA deters attendance in these programs.

College Enrollment

Table 4-1 to Table 4-3 show odds-ratio on 1 Year College Enrollment, 2 Year College Enrollment, and Ever Enrollment as estimated by logistic regression, respectively. Models 1 to 6 compare different specifications: columns 1 and 2 include individual variables alone, excluding all neighborhood effects; model 3 and 4 include all variables but test regression's robustness to dropping cumulative GPA; and, finally, columns 5 and 6, exclude socioeconomic proxies FRL and LEP. These columns are constant across all tables.

Sex is robust across specifications: females have 23 to 72 percent higher odds than males to enroll in college in the first year after high school graduation. The odds of females enrolling is larger in 2-year and Ever Enrollment. Furthermore, the coefficient is significant at 5 percent in model 4, which includes all neighborhood and GPA effects.

Socioeconomic proxies represented by FRL and LEP require discussion. On one hand, LEP is statistically insignificant across all models. On the other hand, FRL is strongly significant but its correlation with students' ethnicity and mean family income in the neighborhood may explain why SOC and Neighborhood income are not significant except in model 5. This would imply that, ceteris paribus, students of color are equally likely to enroll in college.

Table 4-1.

1 YR College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) Individual	(2) Individual	(3) Neighborhood	(4) Neighborhood	(5) Neighborhood	(6) Neighborhood
Female	1.722***	1.258**	1.716***	1.258**	1.652***	1.229*
	(0.158)	(0.137)	(0.158)	(0.138)	(0.150)	(0.134)
FRL	0.409***	0.722***	0.446***	0.731**		
	(0.0423)	(0.0884)	(0.0480)	(0.0921)		
LEP	1.242	1.026	1.306	1.045		
	(0.242)	(0.239)	(0.259)	(0.247)		
SOC	0.884	1.151	0.911	1.158	0.746**	1.057
	(0.134)	(0.223)	(0.139)	(0.227)	(0.0865)	(0.152)
Cohort = 2011	0.878	1.030	0.868	1.026	0.810	1.007
	(0.115)	(0.155)	(0.114)	(0.155)	(0.104)	(0.152)
Cohort = 2012	0.831	0.814	0.824	0.812	0.743**	0.778*
	(0.111)	(0.123)	(0.110)	(0.123)	(0.0972)	(0.116)
Cohort = 2013	0.749**	0.805	0.736**	0.801	0.670***	0.771*
	(0.0999)	(0.123)	(0.0986)	(0.122)	(0.0883)	(0.117)
Family Income pct = 1, 25 pct			0.862	1.000	0.737**	0.947
			(0.120)	(0.167)	(0.101)	(0.157)
Family Income pct = 2, 75 pct			1.093	1.000	1.143	1.011
			(0.148)	(0.155)	(0.153)	(0.156)
Segregation			0.766	0.831	0.738	0.812
			(0.345)	(0.444)	(0.322)	(0.433)
Education High			1.675	1.190	2.132	1.266
			(0.964)	(0.795)	(1.216)	(0.847)
CGPA		3.397***		3.382***		3.540***
		(0.299)		(0.298)		(0.306)
Constant	2.694***	0.0723***	2.181***	0.0683***	1.724**	0.0553***
	(0.294)	(0.0205)	(0.543)	(0.0260)	(0.420)	(0.0206)
Observations	2,326	2,098	2,326	2,098	2,326	2,098

*** p<0.01, ** p<0.05, * p<0.1

Table 4-2.

2 YR College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) Individual	(2) Individual	(3) Neighborhood	(4) Neighborhood	(5) Neighborhood	(6) Neighborhood
Female	1.767***	1.312**	1.765***	1.314**	1.701***	1.278**
	(0.168)	(0.148)	(0.169)	(0.149)	(0.161)	(0.144)
FRL	0.409***	0.697***	0.442***	0.699***		
	(0.0438)	(0.0885)	(0.0494)	(0.0912)		
LEP	1.198	1.059	1.236	1.055		
	(0.240)	(0.256)	(0.253)	(0.258)		
SOC	0.955	1.201	0.979	1.196	0.779**	1.080
	(0.149)	(0.239)	(0.155)	(0.240)	(0.0939)	(0.161)
Cohort = 2011	0.933	1.070	0.918	1.062	0.855	1.040
	(0.128)	(0.171)	(0.126)	(0.170)	(0.115)	(0.166)
Cohort = 2012	0.813	0.782	0.804	0.777	0.725**	0.738**
	(0.112)	(0.122)	(0.111)	(0.122)	(0.0982)	(0.114)
Cohort = 2013	0.725**	0.784	0.708**	0.776	0.644***	0.742*
	(0.0997)	(0.123)	(0.0981)	(0.123)	(0.0877)	(0.117)
Family Income pct = 1, 25 pct			0.937	1.085	0.797	1.020
			(0.135)	(0.187)	(0.113)	(0.174)
Family Income pct = 2, 75 pct			1.143	1.033	1.197	1.047
			(0.162)	(0.168)	(0.168)	(0.170)
Segregation			0.834	0.947	0.792	0.928
			(0.386)	(0.524)	(0.356)	(0.511)
Education High			1.989	1.475	2.539	1.583
			(1.183)	(1.010)	(1.502)	(1.088)
CGPA		3.280***		3.272***		3.446***
		(0.292)		(0.294)		(0.304)
Constant	3.202***	0.0979***	2.368***	0.0829***	1.856**	0.0651***
	(0.367)	(0.0282)	(0.610)	(0.0319)	(0.469)	(0.0245)
Observations	2,326	2,098	2,326	2,098	2,326	2,098

*** p<0.01, ** p<0.05, * p<0.1

Table 4-3.

Ever College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) Individual	(2) Individual	(3) Neighborhood	(4) Neighborhood	(5) Neighborhood	(6) Neighborhood
Female	1.798***	1.319**	1.794***	1.321**	1.738***	1.288**
	(0.180)	(0.156)	(0.180)	(0.156)	(0.173)	(0.152)
FRL	0.428***	0.701***	0.467***	0.714**		
	(0.0477)	(0.0924)	(0.0547)	(0.0970)		
LEP	0.964	0.895	0.995	0.898		
	(0.200)	(0.223)	(0.211)	(0.227)		
SOC	1.098	1.308	1.118	1.296	0.823	1.088
	(0.181)	(0.270)	(0.187)	(0.270)	(0.104)	(0.168)
Cohort = 2011	0.925	1.128	0.918	1.126	0.861	1.107
	(0.134)	(0.192)	(0.133)	(0.192)	(0.124)	(0.189)
Cohort = 2012	0.741**	0.703**	0.736**	0.700**	0.672***	0.670**
	(0.107)	(0.114)	(0.107)	(0.114)	(0.0958)	(0.108)
Cohort = 2013	0.698**	0.755*	0.687**	0.751*	0.628***	0.720**
	(0.101)	(0.125)	(0.100)	(0.125)	(0.0903)	(0.119)
Family Income pct = 1, 25 pct			0.850	0.949	0.730**	0.892
			(0.128)	(0.169)	(0.108)	(0.158)
Family Income pct = 2, 75 pct			1.042	0.975	1.093	0.988
			(0.155)	(0.167)	(0.161)	(0.169)
Segregation			0.959	1.149	0.866	1.079
			(0.459)	(0.666)	(0.402)	(0.622)
Education High			2.215	1.517	2.742	1.616
			(1.384)	(1.090)	(1.708)	(1.167)
CGPA		3.051***		3.034***		3.182***
		(0.284)		(0.283)		(0.291)
Constant	3.998***	0.155***	2.953***	0.137***	2.348***	0.109***
	(0.487)	(0.0469)	(0.798)	(0.0540)	(0.623)	(0.0420)
Observations	2,326	2,098	2,326	2,098	2,326	2,098

*** p<0.01, ** p<0.05, * p<0.1

Academic performance in high school has the strongest magnitude explaining college enrollment in all dimensions, and it is significant at 1 percent. As there exists correlation across explanatory variables, including GPA in the models decreases the effect of other variables. Results imply that a unit increase in GPA will increase odds of enrollment threefold.

Neighborhood effects in Table 4-1 to Table 4-3 are negligible. There appears to be no difference in enrollment between students living in neighborhoods of different income when individual socioeconomic background is accounted for. Model 5, however, does show that low-income neighborhoods underperform college enrollment in the first 1 Year and Ever-Enrollment but not in 2 Year Enrollment. These results are conflicting but it is the only neighborhood characteristic that shows explanatory power in these results. The reasons behind these non-significant effects are cumbersome.

Two-Step Least Squares estimates are shown in Table 4-4 and Table 4-5. By definition, given that GPA does not vary across college enrollment time-window, the first step regression is the same for all college enrollment variables. In the set of individual variables, FRL and SOC have a strongly significant and negative relationship to GPA. Females have approximately 0.304 larger GPA than males. Racial segregation is not significant but the remaining neighborhood variables are more promising. Neighborhood income shows the expected sign but weak significance at the 10 percent level, that is, students living in neighborhoods in the bottom 25 percentile of family income have 0.086 less GPA points than those between the 25 and 75 percentile; while students in the top income distribution have larger cumulative GPA. Education attainment in the neighborhood has the second strongest magnitude, second to socioeconomic status (FRL), and shows that students living in neighborhoods with highly educated population have advantageous results in their school performance. This coefficient is significant at 5 percent level. Finally, Table 4-5 shows the second stage regression, where GPA increases the likelihood of attending college. In fact, GPA has a larger effect in early college enrollment than in delayed enrollment and results are significant.

As expected, results are consistent across logistic and regression models exhibited in Appendix 4. GPA coefficients in the Two Step procedure and OLS signal a downward bias in our regression when it does not account for endogeneity: a point increase in GPA is expected to increase the probability of college enrollment by 0.263 to 0.314, in the Two-Step Regression, versus 0.176 to 0.233 in OLS.

Table 4-4.

High School Performance, First Step.

VARIABLES	(1) GPA
Cohort = 2011	-0.0326
	(0.0420)
Cohort = 2012	-0.0243
	(0.0421)
Cohort = 2013	-0.0934**
	(0.0426)
FRL	-0.458***
	(0.0355)
LEP	0.111
	(0.0678)
SOC	-0.220***
	(0.0525)
Female	0.304***
	(0.0295)
Segregation	-0.0777
	(0.153)
Family Income pct = 1, 25 pct	-0.0864*
	(0.0474)
Family Income pct = 2, 75 pct	0.0821*
	(0.0424)
Education High	0.381**
	(0.182)
Constant	3.031***
	(0.0779)
Observations	2,098
R-squared	0.200
F-stat	47.41

*** p<0.01, ** p<0.05, * p<0.1

Table 4-5.

College Enrollment, Two-Step Regression.

VARIABLES	(1) 1YR Enrollment	(2) 2YR Enrollment	(3) Ever Enrollment
CGPA	0.306***	0.285***	0.253***
	(0.0275)	(0.0266)	(0.0255)
Cohort = 2011	0.00480	0.00980	0.0162
	(0.0259)	(0.0251)	(0.0240)
Cohort = 2012	-0.0401	-0.0457*	-0.0579**
	(0.0259)	(0.0251)	(0.0241)
Cohort = 2013	-0.0345	-0.0368	-0.0388
	(0.0264)	(0.0255)	(0.0245)
Constant	-0.217**	-0.117	0.0203
	(0.0890)	(0.0861)	(0.0827)
Observations	2,098	2,098	2,098
R-squared	0.152	0.138	0.118

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Non-Parametric estimates in *Figure* 4-1 to *Figure* 4-8 present survival analysis by characteristic: sex, race, individual income proxy, and neighborhood percentile rank. Figure 12 presents survival estimates by sex, that is, the proportion of population that have not enrolled by that period, and it suggests that females enroll earlier than males into college: by the first year, approximately 70 percent of females are enrolled while 58 percent of males did. Cumulative hazard is presented in Figure 13, and shows the aforementioned probabilities. These results are consistent with the odds-ratio coefficient for females in logistic regression.

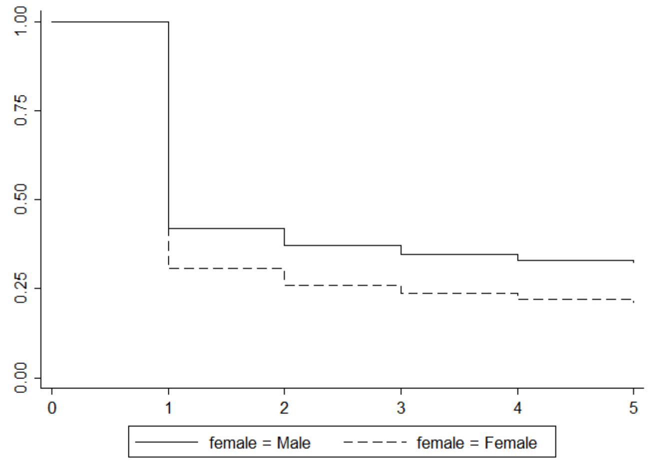


Figure 4-1. Kaplan-Meier Survival Function, by sex.

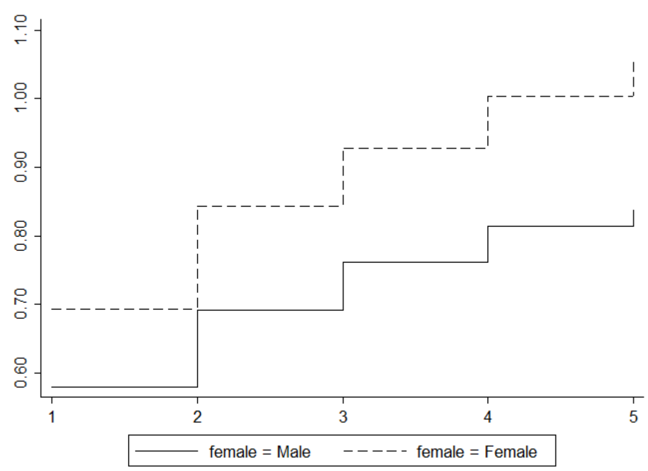


Figure 4-2. Nelson-Aalen Cumulative Hazard, by sex.

Similarly, the survival function and cumulative hazard by FRL is presented in Figures 14 and 15. By the first year, 50 percent of FRL recipients have not enrolled in college against 28 percent of non-FRL participants. That is, students from low socioeconomic background are approximately 22 percent points less likely to be seen in college during the first year and 17 percent points less likely by the fifth period.

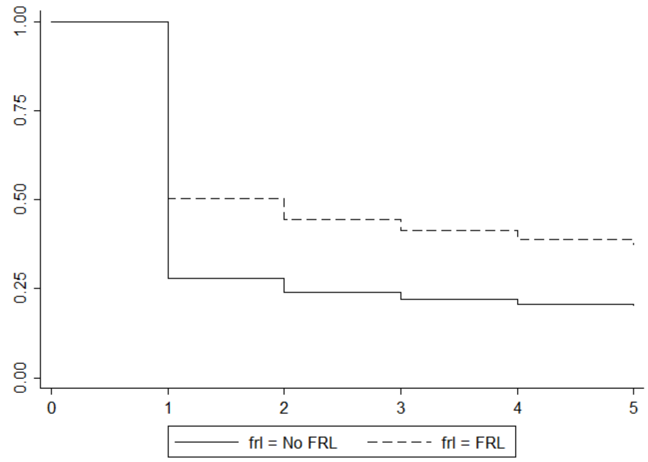


Figure 4-3. Kaplan-Meier Survival Function, by FRL.

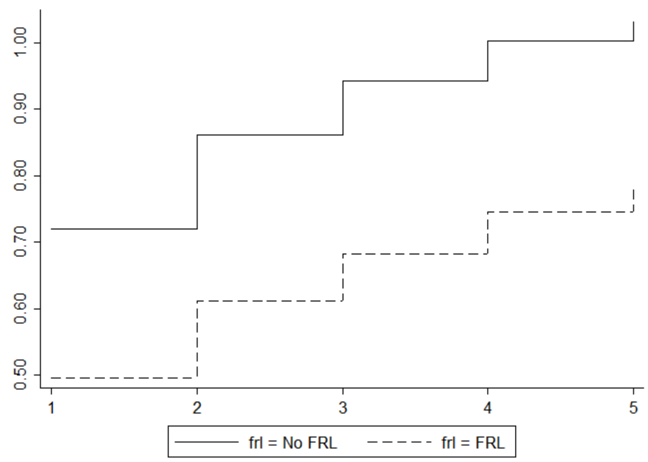


Figure 4-4. Nelson-Aalen Cumulative Hazard, by FRL.

Results in *Figure* 4-5 and *Figure* 4-6 show the survival function by family income percentile rank. Students living in neighborhoods in the bottom 25 income percentile rank have a 52 percent likelihood to be enrolled during the first year, while in the medium income category 65 percent is enrolled in the first year, and 72 percent is enrolled in postsecondary education among those living in high income neighborhoods.

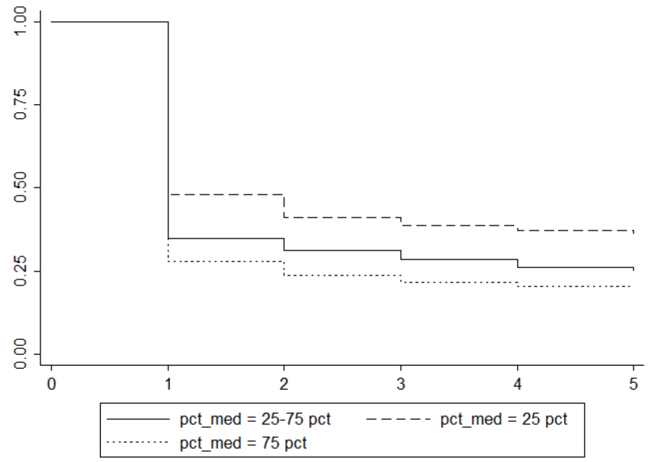


Figure 4-5. Kaplan-Meier Survival Function, by neighborhood income percentile rank.

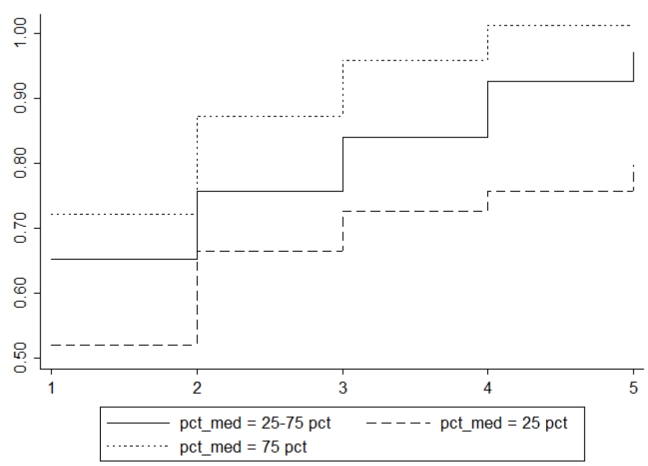


Figure 4-6. Nelson-Aalen Cumulative Hazard, by neighborhood income percentile rank.

The method is replicated by ethnicity across white students and students of color in *Figure* 4-7 and *Figure* 4-8. These plots show that two thirds of white students are enrolled by the first year while only 58 percent of students of color successfully enrolled early in college. By the end of a 5 years period, a third of students of color remains without enrolling, while only 25 percent of white students have yet to enroll.

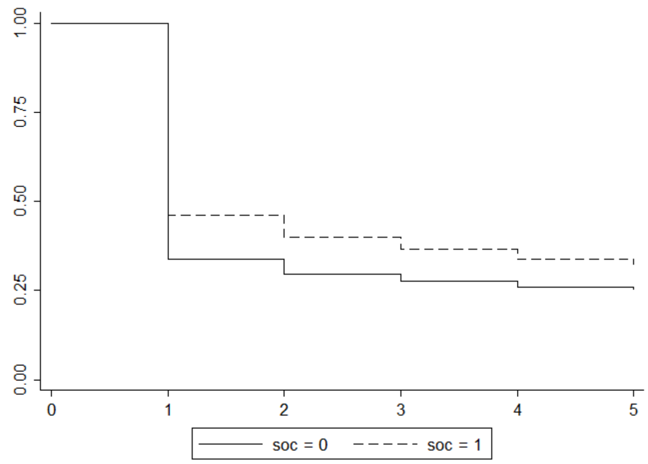


Figure 4-7. Kaplan-Meier Survival Function, by ethnicity.

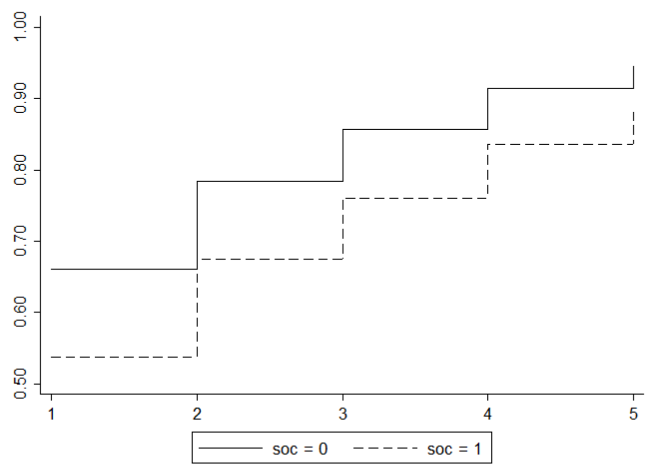


Figure 4-8. Nelson-Aalen Cumulative Hazard, by ethnicity.

Parametric and semi-parametric analysis is presented in Table 4-6. Consistent with neighborhood models above, median family income percentile in neighborhoods, segregation, and the proportion of highly educated population in the neighborhood are not statistically significant. Females, however, are 9.2 to 18.4 percent more likely to enroll early in college, following models 1 and 3 that control for high school performance. Socioeconomic status is strongly significant at 1 percent and suggests that students from low socioeconomic status are 10 to 17 percent less likely to enroll. It is important to point out that race appears to have no effect to explain college enrollment, when controlling for socioeconomic status.

Table 4-6.

College Enrollment, Survival Estimates.

VARIABLES	(1) Parametric	(2) Parametric	(3) Semi-Parametric	(4) Semi-Parametric
analysis time when record ends				
Female	1.184***	1.414***	1.092***	1.222***
	(0.0610)	(0.0692)	(0.0333)	(0.0386)
FRL	0.837***	0.598***	0.895***	0.740***
	(0.0543)	(0.0355)	(0.0368)	(0.0315)
LEP	0.952	0.959	0.980	0.979
	(0.118)	(0.109)	(0.0809)	(0.0796)
SOC	1.056	1.057	1.061	1.047
	(0.0963)	(0.0878)	(0.0622)	(0.0576)
CGPA	1.729***		1.400***	
	(0.0640)		(0.0357)	
Cohort = 2011	0.992	0.933	1.012	0.965
	(0.0696)	(0.0631)	(0.0400)	(0.0408)
Cohort = 2012	0.933	0.959	0.922*	0.930
	(0.0672)	(0.0672)	(0.0389)	(0.0417)
Cohort = 2013	1.090	1.087	0.976	0.947
	(0.0790)	(0.0767)	(0.0412)	(0.0420)
Family Income pct = 1, 25 pct	0.976	0.881	0.981	0.930
	(0.0802)	(0.0687)	(0.0528)	(0.0518)
Family Income pct = 2, 75 pct	1.026	1.087	1.002	1.036
	(0.0730)	(0.0758)	(0.0405)	(0.0439)
Segregation	0.901	0.913	0.953	0.954
	(0.246)	(0.233)	(0.174)	(0.177)
Education High	1.295	1.558	1.092	1.308
	(0.401)	(0.465)	(0.203)	(0.253)
Constant	0.0646***	0.312***		
	(0.0114)	(0.0397)		
Observations	4,039	4,784	4,039	4,784
Subjects	2098	2326	2098	2326
Failures	1586	1687	1586	1687

*** p<0.01, ** p<0.05, * p<0.1

College Enrollment: 2 Year and 4 Year Institutions

This section analyzes college enrollment in 2 Year and 4 Year institutions. Given that the specification remains the same, we compare only selected models for 1 Year College Enrollment, 2 Year College Enrollment, and Ever Enrollment in each kind of institution. We focus on Neighborhood models with and without cumulative GPA, corresponding to columns 3 and 4 in previous results. Results for Two-Step estimates are also presented, but the first-step regression is omitted as it is identical to Table 4-4. Furthermore, the analysis omits enrollment in Less than 2 Year institutions, as only one percent of the sample is ever found in this kind of institutions and such a small sample would require different methodologies. Finally, we omit non parametric estimates in survival models and present semi-parametric and parametric results.

Table 4-7 to Table 4-9 detail determinants of enrollment in 2 Year institutions. In logistic regressions in Table 4-7, Cumulative GPA is found significant at less than 1 percent and its effect is negative: students with larger GPA are less likely to enroll in these institutions with approximately half odds of enrolling by each additional GPA point. High educational attainment in the neighborhood has a significant effect at less than 5 percent during the 1 Year after high school graduation, and suggests that students from neighborhoods with high educational attainment are less likely to enroll in 2 Year institutions. As the time window is increased, however, it cannot be rejected that the effect of neighborhood educational attainment is zero although it remains significant at 10 percent. No further variables are found to significantly explain this outcome across logistic models, and the Two-Step regression is consistent: cumulative GPA has a negative and significant effect explaining the decision to enroll in 2 Year institutions. Students with successful achievements in high school appear to be less likely to attend these institutions.

Survival estimates exhibit slightly different results. When GPA is included in the model, in columns 1 and 3 in Table 4-9, females have 18 to 19 percent increase in the hazard rate of college enrollment in 2 Year institution. The coefficient is significant at less than 5 percent. GPA and Neighborhood educational attainment is consistently found to explain students being less likely to enroll in 2 Year institutions. When GPA is excluded from the models, in columns 2 and 4, students from low socioeconomic status have a 17 to 20 percent increase in their hazard ratio to enroll in 2 Year institutions, being more likely to enroll early. Students of color are also more likely to enroll early: in fact the hazard rate increase has the largest magnitude reported (27 percent) and is significant at less than 5 percent.

Table 4-7.

2 Year Institutions College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) 1YR Enrollment	(2) 1YR Enrollment	(3) 2YR Enrollment	(4) 2YR Enrollment	(5) Ever Enrollment	(6) Ever Enrollment
Female	0.882	1.070	0.968	1.189	1.018	1.167
	(0.0896)	(0.122)	(0.0913)	(0.127)	(0.0897)	(0.114)
FRL	1.176	0.865	1.234*	0.895	1.303**	0.987
	(0.140)	(0.119)	(0.137)	(0.116)	(0.136)	(0.119)
LEP	0.904	1.085	0.911	1.175	0.994	1.259
	(0.191)	(0.253)	(0.182)	(0.263)	(0.188)	(0.263)
SOC	1.232	0.964	1.188	0.881	1.254	0.979
	(0.196)	(0.177)	(0.180)	(0.156)	(0.180)	(0.161)
CGPA		0.458***		0.464***		0.537***
		(0.0354)		(0.0347)		(0.0378)
Cohort = 2011	1.101	1.116	1.132	1.101	0.989	0.981
	(0.157)	(0.174)	(0.149)	(0.159)	(0.121)	(0.130)
Cohort = 2012	0.972	0.959	0.933	0.917	0.780**	0.763**
	(0.142)	(0.152)	(0.127)	(0.135)	(0.0983)	(0.103)
Cohort = 2013	0.977	0.958	0.946	0.944	0.762**	0.734**
	(0.145)	(0.153)	(0.130)	(0.140)	(0.0973)	(0.100)
Family Income pct = 1, 25 pct	0.874	0.742*	0.905	0.767	0.903	0.798
	(0.133)	(0.128)	(0.128)	(0.124)	(0.121)	(0.121)
Family Income pct = 2, 75 pct	0.851	0.928	0.824	0.873	0.826	0.903
	(0.131)	(0.150)	(0.117)	(0.131)	(0.109)	(0.124)
Segregation	1.276	1.496	1.115	1.230	1.346	1.576
	(0.626)	(0.852)	(0.506)	(0.655)	(0.573)	(0.768)
Education High	0.248**	0.250**	0.324*	0.311*	0.402*	0.355*
	(0.154)	(0.167)	(0.187)	(0.195)	(0.219)	(0.208)
Constant	0.487***	5.409***	0.558**	6.208***	0.799	6.035***
	(0.127)	(1.983)	(0.137)	(2.169)	(0.186)	(2.009)
Observations	2,326	2,098	2,326	2,098	2,326	2,098

Table 4-8.

2 Year Institutions College Enrollment, Second Step.

VARIABLES	(1) 1YR Enrollment	(2) 2YR Enrollment	(3) Ever Enrollment
CGPA	-0.108***	-0.109***	-0.149***
	(0.0265)	(0.0284)	(0.0306)
Cohort = 2011	0.0152	0.0162	-0.00774
	(0.0250)	(0.0267)	(0.0288)
Cohort = 2012	-0.00842	-0.0171	-0.0622**
	(0.0250)	(0.0268)	(0.0288)
Cohort = 2013	-0.00896	-0.0105	-0.0696**
	(0.0254)	(0.0272)	(0.0294)
Constant	0.554***	0.609***	0.845***
	(0.0858)	(0.0919)	(0.0990)
Observations	2,098	2,098	2,098
R-squared	0.052	0.053	0.050

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-10 analyzes college enrollment in 4 Year institutions using logistic regression. Cumulative GPA and high educational attainment in the neighborhood are both robust through all models, and the coefficient have large magnitude. A one-point increase in cumulative GPA is associated with 7.7 to 8.9 times larger odds of college enrollment in a 4 Year institution. Similarly, the odds ratio coefficient for high educational attainment ranges from 5.8 to 8.3. Females also have larger odds to enroll in 4 Year institutions but the coefficient is not robust in all specifications. Students from low socioeconomic background show 55 to 60 percent smaller odds of enrollment and socioeconomic background is strongly significant for 1 Year, 2 Year and Ever Enrolled. It is noticeable that LEP, the proxy for immigrant family background, has larger odds of enrollment than non-immigrants. Socioeconomic characteristics, however, are not robust to including GPA in columns 2, 4, and 6 and we cannot reject the hypothesis that students are equally likely to enroll after controlling for high school performance.

Table 4-9.

2 Year Institutions College Enrollment, Survival Estimates.

VARIABLES	(1) Parametric	(2) Parametric	(3) Semi-Parametric	(4) Semi-Parametric
analysis time when record ends				
Female	1.198**	1.038	1.184**	1.045
	(0.0988)	(0.0783)	(0.0908)	(0.0735)
FRL	0.923	1.198**	0.936	1.169*
	(0.0910)	(0.106)	(0.0880)	(0.0965)
LEP	1.335*	1.023	1.306*	1.042
	(0.222)	(0.153)	(0.198)	(0.143)
SOC	1.026	1.274**	1.005	1.232*
	(0.142)	(0.152)	(0.128)	(0.135)
CGPA	0.571***		0.610***	
	(0.0320)		(0.0298)	
Cohort = 2011	0.984	0.992	0.989	0.996
	(0.110)	(0.103)	(0.103)	(0.0968)
Cohort = 2012	0.886	0.905	0.869	0.882
	(0.102)	(0.0988)	(0.0940)	(0.0899)
Cohort = 2013	1.130	1.162	0.960	0.986
	(0.130)	(0.127)	(0.103)	(0.100)
Family Income pct = 1, 25 pct	0.794*	0.902	0.824*	0.915
	(0.0975)	(0.101)	(0.0948)	(0.0944)
Family Income pct = 2, 75 pct	0.918	0.852	0.920	0.864
	(0.110)	(0.0986)	(0.103)	(0.0951)
Segregation	1.371	1.382	1.338	1.313
	(0.517)	(0.476)	(0.469)	(0.411)
Education High	0.346**	0.399**	0.405*	0.444*
	(0.172)	(0.187)	(0.187)	(0.194)
Constant	0.646	0.106***		
	(0.172)	(0.0210)		
Observations	7,862	8,679	7,862	8,679
Subjects	2098	2326	2098	2326
Failures	628	709	628	709

seEform in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4-10.

4 Year Institutions College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) 1YR Enrollment	(2) 1YR Enrollment	(3) 2YR Enrollment	(4) 2YR Enrollment	(5) Ever Enrollment	(6) Ever Enrollment
Female	1.816***	1.150	1.802***	1.139	1.816***	1.124
	(0.159)	(0.127)	(0.158)	(0.127)	(0.160)	(0.125)
FRL	0.439***	0.962	0.429***	0.922	0.408***	0.830
	(0.0456)	(0.124)	(0.0445)	(0.120)	(0.0424)	(0.107)
LEP	1.509**	1.087	1.493**	1.044	1.499**	1.117
	(0.305)	(0.269)	(0.298)	(0.256)	(0.294)	(0.271)
SOC	0.766*	1.299	0.778*	1.411*	0.922	1.616**
	(0.113)	(0.269)	(0.114)	(0.284)	(0.135)	(0.323)
CGPA		8.764***		8.980***		7.696***
		(0.957)		(0.997)		(0.828)
Cohort = 2011	0.779**	0.840	0.769**	0.810	0.790*	0.923
	(0.0956)	(0.123)	(0.0948)	(0.122)	(0.0986)	(0.139)
Cohort = 2012	0.765**	0.670***	0.760**	0.655***	0.759**	0.696**
	(0.0961)	(0.104)	(0.0958)	(0.102)	(0.0967)	(0.108)
Cohort = 2013	0.674***	0.652***	0.672***	0.647***	0.639***	0.658***
	(0.0858)	(0.101)	(0.0856)	(0.102)	(0.0818)	(0.102)
Family Income pct = 1, 25 pct	0.934	1.327	0.972	1.430**	0.983	1.418**
	(0.128)	(0.233)	(0.131)	(0.251)	(0.132)	(0.244)
Family Income pct = 2, 75 pct	1.117	0.953	1.172	1.001	1.052	0.888
	(0.141)	(0.149)	(0.150)	(0.162)	(0.136)	(0.144)
Segregation	0.491	0.387*	0.523	0.375*	0.502	0.375*
	(0.229)	(0.223)	(0.239)	(0.217)	(0.228)	(0.221)
Education High	5.971***	6.453***	5.873***	7.331***	6.546***	8.382***
	(3.228)	(4.392)	(3.169)	(5.006)	(3.575)	(5.723)
Constant	0.633**	0.000765***	0.693	0.000760***	0.814	0.00147***
	(0.146)	(0.000351)	(0.159)	(0.000352)	(0.189)	(0.000656)
Observations	2,326	2,098	2,326	2,098	2,326	2,098

Table 4-11.

4 Year Institutions College Enrollment, Second Step.

VARIABLES	(1) 1YR Enrollment	(2) 2YR Enrollment	(3) Ever Enrollment
CGPA	0.401***	0.401***	0.374***
	(0.0268)	(0.0266)	(0.0268)
Cohort = 2011	-0.0236	-0.0285	-0.0106
	(0.0253)	(0.0251)	(0.0252)
Cohort = 2012	-0.0627**	-0.0658***	-0.0626**
	(0.0253)	(0.0251)	(0.0253)
Cohort = 2013	-0.0601**	-0.0607**	-0.0638**
	(0.0257)	(0.0256)	(0.0257)
Constant	-0.665***	-0.644***	-0.524***
	(0.0868)	(0.0862)	(0.0867)
Observations	2,098	2,098	2,098
R-squared	0.333	0.338	0.312

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Two Step results in Table 4-11 differ from findings in 2 Year institutions. GPA shows a positive and robust coefficient to explain college enrollment in a 4 Year institution. Furthermore, in absolute values, the coefficient is larger than that of 2 Year institutions.

Survival estimates from Table 4-12 are consistent with logistic models. GPA remains as the main variable explaining enrollment in 4 Year institutions. In columns 2 and 4, low socioeconomic background appears to explain lower odds of enrollment but this effect is not robust across specifications 1 and 2 which include high school performance.

Table 4-12.

4 Year Institutions College Enrollment, Survival Estimates.

VARIABLES	(1) Parametric	(2) Parametric	(3) Semi-Parametric	(4) Semi-Parametric
analysis time when record ends				
Female	0.946	1.178*	0.928	1.147
	(0.0918)	(0.111)	(0.0811)	(0.0986)
FRL	1.012	0.645***	1.021	0.645***
	(0.129)	(0.0785)	(0.125)	(0.0742)
LEP	0.873	0.918	0.864	0.917
	(0.220)	(0.217)	(0.202)	(0.199)
SOC	1.246	1.090	1.248	1.104
	(0.207)	(0.173)	(0.198)	(0.150)
CGPA	2.391***		2.402***	
	(0.194)		(0.172)	
Cohort = 2011	0.923	0.903	0.916	0.904
	(0.119)	(0.115)	(0.112)	(0.107)
Cohort = 2012	0.930	1.031	0.950	1.069
	(0.120)	(0.131)	(0.110)	(0.124)
Cohort = 2013	1.167	1.150	1.338**	1.342**
	(0.165)	(0.161)	(0.157)	(0.164)
Family Income pct = 1, 25 pct	0.893	0.825	0.878	0.821
	(0.156)	(0.136)	(0.139)	(0.126)
Family Income pct = 2, 75 pct	0.960	1.059	0.969	1.077
	(0.127)	(0.139)	(0.120)	(0.133)
Segregation	0.850	0.658	0.889	0.654
	(0.482)	(0.352)	(0.417)	(0.296)
Education High	1.276	2.130	1.255	2.136
	(0.770)	(1.247)	(0.671)	(1.152)
Constant	0.00775***	0.108***		
	(0.00283)	(0.0272)		
Observations	1,070	1,189	1,070	1,189
Subjects	1070	1189	1070	1189
Failures	450	462	450	462

seEform in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In summary, characteristics of students attending 2 Year and 4 Year institutions appear to be different. In part, high school performance explains the decision to enroll in different kind of programs: high achievers are drawn towards 4 Year institutions, otherwise 2 Year institutions represent an alternative. Correlations between background and GPA in Two Step regressions maintain robust coefficients and similar results. Neighborhoods also appear to have a role explaining these decisions through exposure to high educational attainment.

College Enrollment: Public, Private Non For-Profit, and For Profit Institutions

As in the previous sections, we analyze enrollment in public, private, and for profit institutions using Logistic, Two-Step, and survival analysis. Results from attending to public colleges is presented first, followed by private and for-profit. Two-Step models omit tables for GPA estimates, which are equal to results in Table 4-4.

Results in Table 4-13 present college enrollment odd-ratios in public institutions. Neighborhood variables cannot be rejected to have a coefficient equal to zero across all specifications, suggesting that people are equally likely to attend public college independently of the neighborhood they live in, ceteris paribus. Although socioeconomic status is not robust to including GPA, results suggest that receiving FRL is associated with having approximately half the odds of enrolling in college than a non-FRL recipient. Socioeconomic status does appear to be significant at 5 percent in 2 Year Enrollment, controlling for GPA. An additional point in GPA is significantly associated with having 2.055 to 2.132 larger odds to attend a public institution, keeping everything else constant. Sex also appears to play a role in models 1, 3, and 5 but the effect is captured by GPA when included. Interestingly, LEP status in columns 1 and 3 may indicate that immigrants, ceteris paribus, are more likely to enroll early in public institutions, but the effect is not robust across specifications.

Table 4-13.

Public College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) 1YR Enrollment	(2) 1YR Enrollment	(3) 2YR Enrollment	(4) 2YR Enrollment	(5) Ever Enrollment	(6) Ever Enrollment
Female	1.334***	1.067	1.365***	1.118	1.433***	1.135
	(0.116)	(0.104)	(0.121)	(0.111)	(0.132)	(0.118)
FRL	0.539***	0.810*	0.550***	0.786**	0.584***	0.829
	(0.0552)	(0.0923)	(0.0575)	(0.0909)	(0.0628)	(0.0988)
LEP	1.587**	1.378	1.492**	1.355	1.189	1.091
	(0.308)	(0.300)	(0.295)	(0.302)	(0.242)	(0.254)
SOC	0.972	1.138	1.053	1.223	1.196	1.385*
	(0.143)	(0.198)	(0.157)	(0.214)	(0.185)	(0.255)
CGPA		2.132***		2.055***		2.067***
		(0.161)		(0.158)		(0.165)
Cohort = 2011	0.875	0.985	0.931	1.042	0.959	1.160
	(0.108)	(0.134)	(0.118)	(0.147)	(0.128)	(0.174)
Cohort = 2012	0.732**	0.695***	0.736**	0.696***	0.692***	0.667***
	(0.0921)	(0.0944)	(0.0940)	(0.0962)	(0.0916)	(0.0956)
Cohort = 2013	0.682***	0.713**	0.695***	0.737**	0.700***	0.739**
	(0.0861)	(0.0974)	(0.0894)	(0.103)	(0.0939)	(0.108)
Family Income pct = 1, 25 pct	0.830	0.900	0.896	0.973	0.843	0.928
	(0.111)	(0.137)	(0.122)	(0.151)	(0.119)	(0.149)
Family Income pct = 2, 75 pct	1.070	0.993	1.075	0.989	1.000	0.936
	(0.136)	(0.136)	(0.139)	(0.139)	(0.135)	(0.139)
Segregation	0.619	0.685	0.633	0.690	0.724	0.832
	(0.268)	(0.329)	(0.276)	(0.338)	(0.323)	(0.427)
Education High	0.907	0.670	0.985	0.712	1.368	0.999
	(0.488)	(0.397)	(0.542)	(0.430)	(0.785)	(0.634)
Constant	2.139***	0.253***	2.303***	0.310***	2.515***	0.329***
	(0.498)	(0.0856)	(0.548)	(0.106)	(0.623)	(0.116)
Observations	2,326	2,098	2,326	2,098	2,326	2,098

Table 4-14.

Public College Enrollment, Second Step.

VARIABLES	(1) 1YR Enrollment	(2) 2YR Enrollment	(3) Ever Enrollment
CGPA	0.202***	0.183***	0.164***
	(0.0303)	(0.0297)	(0.0285)
Cohort = 2011	-0.00510	0.00580	0.0238
	(0.0286)	(0.0280)	(0.0269)
Cohort = 2012	-0.0842***	-0.0819***	-0.0839***
	(0.0286)	(0.0280)	(0.0269)
Cohort = 2013	-0.0720**	-0.0623**	-0.0567**
	(0.0291)	(0.0285)	(0.0274)
Constant	0.0524	0.143	0.242***
	(0.0983)	(0.0962)	(0.0923)
Observations	2,098	2,098	2,098
R-squared	0.081	0.073	0.071

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4-14 provides results for the Two-Step regression on GPA and college enrollment in public institutions. GPA remains robust at less than 1 percent. The coefficient decreases from 0.202 to 0.164 when the time window is expanded from 1-Year Enrollment to Ever-Enrollment, respectively, a decrease in 17 percent that suggests a slight convergence in the probability to attend college, independent of grades.

Survival estimates from Table 4-15 suggest no difference in the timing of enrollment to public colleges across individual and neighborhood variables, with the exception of cumulative GPA. A unit increase in GPA increases the expected hazard by 37 to 38 percent.

Table 4-16 provides logistic odds-ratio estimates for enrollment in private institutions. Columns 2, 4 and 5 must be pointed out as the odds of enrollment in private college for a student with an additional unit in GPA increase fourfold, a large and significant effect at less than 1 percent.

Table 4-15.

Public College Enrollment, Survival Estimates.

VARIABLES	(1) Parametric	(2) Parametric	(3) Semi-Parametric	(4) Semi-Parametric
analysis time when record ends				
Female	0.991	1.091	0.974	1.066
	(0.0918)	(0.0978)	(0.0807)	(0.0865)
FRL	1.048	0.845	1.050	0.845*
	(0.122)	(0.0932)	(0.112)	(0.0845)
LEP	1.141	1.286	1.127	1.280
	(0.255)	(0.272)	(0.225)	(0.243)
SOC	1.047	0.915	1.061	0.932
	(0.174)	(0.146)	(0.161)	(0.133)
CGPA	1.377***		1.380***	
	(0.0970)		(0.0877)	
Cohort = 2011	0.862	0.835	0.857	0.834
	(0.109)	(0.103)	(0.101)	(0.0968)
Cohort = 2012	0.966	0.979	0.996	1.013
	(0.120)	(0.120)	(0.111)	(0.112)
Cohort = 2013	1.298**	1.279*	1.496***	1.489***
	(0.170)	(0.165)	(0.169)	(0.168)
Family Income pct = 1, 25 pct	0.794	0.782	0.786*	0.776*
	(0.127)	(0.118)	(0.111)	(0.105)
Family Income pct = 2, 75 pct	0.990	1.041	1.001	1.059
	(0.127)	(0.131)	(0.123)	(0.128)
Segregation	0.772	0.623	0.776	0.619
	(0.392)	(0.304)	(0.346)	(0.264)
Education High	1.292	1.396	1.299	1.398
	(0.740)	(0.781)	(0.686)	(0.727)
Constant	0.0532***	0.139***		
	(0.0173)	(0.0332)		
Observations	1,070	1,189	1,070	1,189
Subjects	1070	1189	1070	1189
Failures	489	506	489	506

seEform in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4-16.

Private College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) 1YR Enrollment	(2) 1YR Enrollment	(3) 2YR Enrollment	(4) 2YR Enrollment	(5) Ever Enrollment	(6) Ever Enrollment
Female	1.584***	1.111	1.556***	1.101	1.595***	1.119
	(0.234)	(0.178)	(0.223)	(0.171)	(0.214)	(0.163)
FRL	0.340***	0.689	0.331***	0.682*	0.368***	0.725
	(0.0754)	(0.160)	(0.0715)	(0.153)	(0.0716)	(0.148)
LEP	0.480	0.561	0.405*	0.484	0.416**	0.473
	(0.246)	(0.312)	(0.206)	(0.266)	(0.179)	(0.217)
SOC	0.923	1.071	1.032	1.141	1.128	1.267
	(0.255)	(0.343)	(0.268)	(0.351)	(0.264)	(0.334)
CGPA		4.740***		4.491***		4.178***
		(0.673)		(0.616)		(0.534)
Cohort = 2011	1.133	1.231	1.080	1.138	0.929	0.987
	(0.227)	(0.260)	(0.211)	(0.236)	(0.165)	(0.188)
Cohort = 2012	1.033	0.982	1.058	0.985	0.836	0.790
	(0.218)	(0.213)	(0.215)	(0.206)	(0.157)	(0.154)
Cohort = 2013	0.999	1.010	1.052	1.070	0.753	0.767
	(0.214)	(0.231)	(0.217)	(0.237)	(0.146)	(0.162)
Family Income pct = 1, 25 pct	0.978	1.085	0.924	0.998	0.860	0.991
	(0.262)	(0.300)	(0.244)	(0.274)	(0.209)	(0.250)
Family Income pct = 2, 75 pct	1.097	0.958	1.078	0.936	1.184	1.034
	(0.212)	(0.199)	(0.201)	(0.188)	(0.208)	(0.197)
Segregation	0.522	0.541	0.611	0.661	0.926	0.952
	(0.460)	(0.510)	(0.527)	(0.618)	(0.705)	(0.786)
Education High	4.891*	2.205	6.194**	3.454	3.858*	2.388
	(4.179)	(2.045)	(5.185)	(3.132)	(3.005)	(2.011)
Constant	0.0545***	0.000394***	0.0550***	0.000451***	0.0896***	0.000915***
	(0.0204)	(0.000238)	(0.0199)	(0.000263)	(0.0300)	(0.000500)
Observations	2,326	2,098	2,326	2,098	2,326	2,098

Table 4-17.

Private Non-For Profit College Enrollment, Second Step.

VARIABLES	(1) 1YR Enrollment	(2) 2YR Enrollment	(3) Ever Enrollment
CGPA	0.152***	0.165***	0.174***
	(0.0191)	(0.0197)	(0.0209)
Cohort = 2011	0.0224	0.0168	0.00412
	(0.0180)	(0.0185)	(0.0197)
Cohort = 2012	0.00643	0.00701	-0.0147
	(0.0180)	(0.0186)	(0.0197)
Cohort = 2013	0.0111	0.0167	-0.0135
	(0.0183)	(0.0189)	(0.0200)
Constant	-0.380***	-0.413***	-0.409***
	(0.0618)	(0.0637)	(0.0676)
Observations	2,098	2,098	2,098
R-squared	0.073	0.072	0.079

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results from the Second Step equation presented in Table 4-17 differ from that of public institutions: the increase in the probability of enrollment in private institutions relative to GPA increases as the time window is expanded from 1 Year enrollment to delayed enrollment. However, this behavior is not consistent with that in Table 4-16.

Survival estimates give a better understanding of changes in the rate of enrollment on time. Cumulative GPA has strong effect, suggesting that increases in GPA are associated with earlier enrollment and the likelihood of the event increases dramatically as time passes. FRL is significant at 10 percent when GPA is included but significant at one percent if GPA is not controlled for. The coefficient shows that the hazard rate decreases for students from low socioeconomic background, who are less likely to enroll in private schools.

Attendance at for profit institutions is modelled in Table 4-19 to Table 4-21. Summary statistics from previous chapters already suggested that attendance at for profit institutions is infrequent, as only 3 percent of the sample attend at these colleges.

Table 4-18.

Private Non-For Profit College Enrollment, Survival Estimates.

VARIABLES	(1) Parametric	(2) Parametric	(3) Semi-Parametric	(4) Semi-Parametric
analysis time when record ends				
Female	0.858	1.243	0.826	1.198
	(0.182)	(0.257)	(0.172)	(0.246)
FRL	0.505*	0.234***	0.511*	0.236***
	(0.179)	(0.0818)	(0.181)	(0.0845)
LEP	0.311	0.337	0.299	0.333
	(0.248)	(0.264)	(0.246)	(0.259)
SOC	1.709	1.418	1.697	1.426
	(0.564)	(0.466)	(0.577)	(0.448)
CGPA	4.530***		4.568***	
	(0.927)		(0.848)	
Cohort = 2011	1.133	1.150	1.164	1.189
	(0.302)	(0.302)	(0.302)	(0.303)
Cohort = 2012	0.848	1.035	0.850	1.071
	(0.239)	(0.289)	(0.217)	(0.289)
Cohort = 2013	0.776	0.865	0.849	0.973
	(0.270)	(0.293)	(0.299)	(0.326)
Family Income pct = 1, 25 pct	1.201	1.114	1.178	1.112
	(0.486)	(0.424)	(0.452)	(0.410)
Family Income pct = 2, 75 pct	0.967	1.159	0.957	1.168
	(0.280)	(0.329)	(0.268)	(0.328)
Segregation	1.473	1.084	1.486	1.042
	(1.959)	(1.352)	(1.436)	(0.904)
Education High	0.747	3.730	0.677	3.751
	(0.999)	(4.761)	(0.854)	(4.661)
Constant	0.000230***	0.0188***		
	(0.000202)	(0.0105)		
Observations	1,070	1,189	1,070	1,189
Subjects	1070	1189	1070	1189
Failures	95	97	95	97

*** p<0.01, ** p<0.05, * p<0.1

Table 4-19.

For Profit College Enrollment, Logistic Odds-Ratio.

VARIABLES	(1) 1YR Enrollment	(2) 1YR Enrollment	(3) 2YR Enrollment	(4) 2YR Enrollment	(5) Ever Enrollment	(6) Ever Enrollment
Female	2.834***	3.328***	2.507***	2.706***	2.258***	2.647***
	(1.019)	(1.421)	(0.791)	(1.038)	(0.551)	(0.771)
FRL	3.203***	1.720	3.152***	1.768	2.111***	1.171
	(1.254)	(0.801)	(1.070)	(0.777)	(0.556)	(0.389)
SOC	0.873	0.717	0.825	0.619	1.245	1.102
	(0.419)	(0.464)	(0.358)	(0.376)	(0.427)	(0.496)
CGPA		0.439***		0.418***		0.465***
		(0.106)		(0.0975)		(0.0755)
Cohort = 2011	0.603	0.478	0.859	0.691	0.738	0.658
	(0.269)	(0.272)	(0.333)	(0.337)	(0.215)	(0.222)
Cohort = 2012	0.618	0.697	0.586	0.608	0.555*	0.513*
	(0.282)	(0.356)	(0.249)	(0.301)	(0.173)	(0.177)
Cohort = 2013	0.534	0.471	0.568	0.562	0.403**	0.389**
	(0.270)	(0.294)	(0.256)	(0.306)	(0.147)	(0.159)
Family Income pct = 1, 25 pct	1.603	1.026	1.666	1.127	1.289	0.831
	(0.795)	(0.657)	(0.761)	(0.683)	(0.441)	(0.354)
Family Income pct = 2, 75 pct	1.028	1.188	1.314	1.689	1.232	1.502
	(0.616)	(0.707)	(0.651)	(0.895)	(0.456)	(0.574)
Segregation	9.022	30.94**	6.481	18.22*	1.509	1.858
	(12.23)	(48.35)	(8.719)	(27.76)	(1.793)	(2.784)
Education High	3.430	2.406	6.087	4.838	0.973	0.466
	(7.663)	(5.582)	(11.37)	(10.06)	(1.354)	(0.701)
LEP	0.102**		0.188**	0.165	0.257**	0.334
	(0.114)		(0.155)	(0.190)	(0.156)	(0.245)
Constant	0.00396***	0.0602**	0.00393***	0.0605**	0.0217***	0.316
	(0.00417)	(0.0820)	(0.00348)	(0.0793)	(0.0138)	(0.260)
Observations	2,326	1,888	2,326	2,098	2,326	2,098

Table 4-20.

For Profit College Enrollment, Second Step.

VARIABLES	(1) 1YR Enrollment	(2) 2YR Enrollment	(3) Ever Enrollment
CGPA	-0.00592	-0.00557	0.00148
	(0.00727)	(0.00806)	(0.0109)
Cohort = 2011	-0.00665	-0.00300	-0.00951
	(0.00685)	(0.00759)	(0.0103)
Cohort = 2012	-0.00142	-0.00337	-0.0152
	(0.00686)	(0.00760)	(0.0103)
Cohort = 2013	-0.00711	-0.00519	-0.0201*
	(0.00698)	(0.00774)	(0.0105)
Constant	0.0346	0.0354	0.0348
	(0.0236)	(0.0261)	(0.0353)
Observations	2,098	2,098	2,098
R-squared	0.005	0.004	0.001

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From Table 4-19, females appear to be twice to three times more likely to attend at these institutions. The coefficient is robust to including GPA. Low socioeconomic background also has a positive effect increasing the likelihood to attend at for-profit college, but this is not robust to GPA controls. Segregation has a remarkable coefficient in models 2 and 4, suggesting that students from more segregated neighborhoods may have up to 30 times more likelihood of enrolling in for profit institutions. This effect, however, is not robust to GPA. Furthermore, some models show that LEP students are less likely to enroll in these institutions but, once again, the effect is not robust to controlling for high school performance. Finally, GPA suggests that students with high academic achievements are less likely to attend these institutions; however, when possible bias in GPA due to correlation with explanatory variables is accounted for, in Table 4-20, GPA is not statistically different from zero.

Table 4-21.

For Profit College Enrollment, Survival Estimates.

VARIABLES	(1) Parametric	(2) Parametric	(3) Semi-Parametric	(4) Semi-Parametric
analysis time when record ends				
Female	1.944	1.701	1.924	1.651
	(1.434)	(1.060)	(1.219)	(1.047)
FRL	0.349	0.987	0.323	0.985
	(0.334)	(0.668)	(0.434)	(0.778)
LEP	3.36e-08	0.319	0***	0.330
	(0.000148)	(0.373)	(0)	(0.402)
SOC	2.646	4.582**	2.457	4.490**
	(2.367)	(3.067)	(2.405)	(2.993)
CGPA	0.374**		0.376**	
	(0.175)		(0.144)	
Cohort = 2011	1.148	0.932	1.118	0.890
	(0.905)	(0.637)	(0.823)	(0.638)
Cohort = 2012	5.66e-08	0.228	0***	0.237
	(0.000139)	(0.257)	(0)	(0.259)
Cohort = 2013	1.009	0.754	1.174	0.859
	(0.937)	(0.659)	(1.183)	(0.793)
Family Income pct = 1, 25 pct	1.009	0.662	0.982	0.653
	(1.073)	(0.570)	(1.073)	(0.660)
Family Income pct = 2, 75 pct	2.403	1.604	2.489	1.663
	(2.194)	(1.375)	(2.398)	(1.448)
Segregation	0.0847	0.287	0.0835	0.307
	(0.363)	(0.897)	(0.217)	(0.806)
Education High	0.0123	0.0234	0.0108	0.0202
	(0.0535)	(0.0850)	(0.0357)	(0.0537)
Constant	0.222	0.0103***		
	(0.475)	(0.0155)		
Observations	1,070	1,189	1,070	1,189
Subjects	1070	1189	1070	1189
Failures	9	12	9	12

seEform in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Survival Analysis in Table 4-21 is consistent with previous findings. Students with better high school performance are less likely to enroll in for-profit institutions, and the hazard rate lower than one suggests that enrollment is less likely as time increases. Other variables are not robust controlling for GPA.

Results in this section highlight the impact of high school performance in explaining the decision to enroll in public or private colleges. Neighborhood effects appear insignificant in this decision when endogeneity is not accounted for. Females are more likely to enroll at for-profit institutions but this result is not robust in private non-for-profit and public colleges.

5. Discussion and Conclusions

We explored neighborhood effects and socioeconomic determinants of college enrollment and educational choices in a sample of 2,332 high school graduates from St. Cloud school district. Logistic regressions, two-step least squares, and survival analysis were used in the analysis to test neighborhood effects. Two Step Least Squares may be the most reliable method to study college enrollment, accounting for endogeneity in high school outcomes and socioeconomic determinants.

Regression and logistic models fail to provide robust neighborhood effects across most specifications. High school grades, sex and FRL have strong effects in these models and offset neighborhood variables.

When GPA is considered endogenous to socioeconomic determinants, findings show high educational attainment in the neighborhood has a large and robust effect in high school performance. Although family income percentile is not strongly significant, introducing unemployment in the model increases robustness. Furthermore, using income per capita percentile improves the model: consistent with findings in Ainsworth (2002), neighborhood effects are larger than the socioeconomic background (FRL), sex, and race.

Survival models allowed to study the timing of enrollment. Although some evidence suggests that neighborhood income may be related to early enrollment, parametric and non-parametric estimates fail to properly incorporate these results. The fact that endogeneity is not accounted as in the Two-Step approach could explain these results. We do find that GPA, sex, and socioeconomic background explains differences in delayed enrollment.

Racial segregation in the census block group is insignificant across almost every specification. An issue that has yet to be properly understood, given the relationship between segregation, income, and education. This result, however, is not completely inconsistent as the literature stresses that segregation may not be a robust variable to explain educational outcomes (Echenique, Fryer, & Kaufman, 2006; Fryer, 2011).

Findings on the choice between institutional types is of utmost importance, because there are different returns to each decision. Students with better high school grades are more likely to enroll in 4 Year institutions, less likely to enroll in 2 Year institution, and have lower odds to enroll into for-profit institutions. Further research should consider career choice and control by expected returns to properly assess neighborhood effects. Logistic results do favor the idea that exposure to high education attainment in the neighborhood is strongly related with the decision to enroll into a 2 Year institutions or invest in 4 Year education.

This project adds to the literature on heterogeneous educational choices by including neighborhood effects. It also suggests the relevance of early intervention in students to boost their outcomes, as high school performance remains the strongest determinant of college enrollment. Further research could examine persistence and performance in college, or follow similar methods to study choice of major and persistence in the major. Additionally, this rich dataset may offer opportunities to analyze labor markets and human capital choices.

References

Ainsworth, J. W. (2002). Why Does It Take a Village? The Mediation of Neighborhood Effects on Educational Achievement. *Social Forces*, *81*(1), 117–152. http://doi.org/10.1353/sof.2002.0038

Akerlof, G. A. (1997). Social Distance and Social Decisions. *Econometrica*, *65*(5), 1005–1027.

Altonji, J. G. (1993). The Demand for and Return to Education When Education Outcomes are Uncertain. *Journal of Labor Economics*, *11*(1), 48–83.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *NBER Working Papers*, 58. http://doi.org/10.1146/annurev-economics-080511-110908.We

Angrist, J., Autor, D., Hudson, S., & Pallais, A. (2016). *Evaluating Post-Secondary Aid: Enrollment, Persistence, and Projected Completion Effects* (Working Paper No. 23015). http://doi.org/10.3386/w23015

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, *121*(1–2), 343–375. http://doi.org/10.1016/j.jeconom.2003.10.010

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, *166*(1), 3–16. http://doi.org/10.1016/J.JECONOM.2011.06.002

Bahi, S., Higgins, D., & Staley, P. (2015). A time hazard analysis of students persistence: a US university graduate mathematics major experience. *International Journal of Science and Mathematics Education*, *13*(5), 1139–1160.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special*

Reference to Education. *The Chicago University Press*. Chicago: The University of Chicago Press. http://doi.org/10.2307/2229541

Black, S. E., & Sufi, A. (2002). *Who goes to college? Differential enrollment by race and family background* (Working Paper No. 9310).

Blair, L. M., Finn, M. G., & Stevenson, W. (1981). The Returns to the Associate Degree for Technicians. *The Journal of Human Resources*, *16*(3), 449–458.

Bobonis, G. J., & Finan, F. (2009). Neighborhood Peer Effects in Secondary School Enrollment Decisions. *The Review of Economics and Statistics*, *91*(4), 695–716.

Bozick, R., & Deluca, S. (2005). Better Late than Never? Delayed Enrollment in the High School to College Transition. *Social Forces*, *84*(1), 531–554.

Bozick, R., & DeLuca, S. (2005). Better Late Than Never? Delayed Enrollment in the High School to College Transition. *Social Forces*, *84*(1), 531–554. http://doi.org/10.1353/sof.2005.0089

Cabrera, A. F., & La Nasa, S. M. (2001). On the Path to College: Three Critical Tasks Facing America's Disadvantaged. *Research in Higher Education*, *42*(2), 119–149. http://doi.org/10.1023/A:1026520002362

Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer Effects and Social Networks in Education. *The Review of Economics Studies*, *76*(14), 1239–1267.

Cameron, S. V, & Heckman, J. J. (2007). The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy*, *109*(3), 455–499.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, *69*(5), 1127–1160. http://doi.org/10.1111/1468-0262.00237

Card, D., & Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, *100*(1), 1–40.

Carter, D. J., & Wilson, R. (1994). *Minorities in Higher Education. 1994 Thirteenth Annual Status Report*. Washington.

Catsiapis, G. (1987). A Model of Educational Investment Decisions. *The Review of Economics and Statistics*, *69*(1), 33–41.

Chetty, R., & Hendren, N. (2018a). Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quaterly Journal of Economics*, *133*(3), 1107–1162. http://doi.org/10.1093/qje/qjy007.Advance

Chetty, R., & Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility II: County-level Estimates. *The Quarterly Journal of Economics*, *133*(July), 1163–1228. http://doi.org/10.1093/qje/qjy006.Advance

Chetty, R., Hendren, N., & Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *Journal of School Choice, 106*(4), 855–902. http://doi.org/10.1080/15582159.2016.1172911

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quaterly Journal of Economics*, *129*(4), 1553–1623. http://doi.org/10.1093/qje/qju022.Advance

Chimka, J. R., & Lowe, L. H. (2008). Interaction and survival analysis of graduation data. *Educational Research and Review*, *3*(1), 29–32.

Cutler, D. M., & Glaeser, E. L. (1997). Are Ghettos Good or Bad? *The Quarterly Journal of*

Economics, *112*(3), 827–872. http://doi.org/10.1162/003355397555361

Doyle, W. R. (2009). The effect of community college enrollment on bachelor's degree completion. *Economics of Education Review*, *28*(2), 199–206. http://doi.org/10.1016/j.econedurev.2008.01.006

Durlauf, S. N. (2004). Neighborhood Effects. In *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2173–2242). Elsevier. http://doi.org/10.1016/S1574-0080(04)80007-5

Echenique, F., Fryer, R. G., & Kaufman, A. (2006). Is School Segregation Good or Bad? *American Economic Review*, *96*(2), 265–269. http://doi.org/10.1257/000282806777212198

Entwisle, D. R., Alexander, K. L., & Steffel Olson, L. (1994). The Gender Gap in Math: Its Possible Origins in Neighborhood Effects. *American Sociological Association*, *59*(6), 822–838.

Fryer, R. G. (2011). *The importance of segregation, discrimination, peer dynamics, and identity in explaining trends in the racial achievement gap*. *Handbook of Social Economics* (1st ed., Vol. 1). Elsevier B.V. http://doi.org/10.1016/B978-0-444-53707-2.00004-9

García-Pérez, M., & Johnson, R. C. (2017). Understanding the Effect of an Intervention Program on High School Graduation Rates: the Access and Opportunity Program in St. Cloud, Minnesota. *Educational Planning*, *24*(2).

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood Effects on Educational Attainment: A Multilevel Analysis. *Sociology of Education*, *64*(4), 251. http://doi.org/10.2307/2112706

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and Social Interactions. *The Quaterly Journal of Economics*, *111*(2), 507–548.

Grogger, J., & Eide, E. (1995). Changes in College Skills and the Rise in the College Wage

Premium. *The Journal of Human Resources*, *30*(2), 280–310. http://doi.org/10.2307/146120

Hauser, R. M. (1993a). The Decline in College Entry among African Americans: Findings in Search of Explanations. In P. M. Sniderman, P. E. Tetlock, & E. G. Carmines (Eds.), *Prejudice, politics, and the American dilemma* (pp. 271–306). Stanford: Stanford University Press.

Hauser, R. M. (1993b). Trends in College Entry among Whites, Blacks, and Hispanics. In C. T. Clotfelter & M. Rothschild (Eds.), *Studies of Supply and Demand in Higher Education*. Chicago: University of Chicago Press.

Hellerstein, J. K., & Neumark, D. (2008). Workplace Segregation in the United States: Race, Ethnicity, and Skill. *Review of Economics and Statistics*, *90*(3), 459–477. http://doi.org/10.1162/rest.90.3.459

Horn, L. J., & Carroll, D. C. (1996). *Nontraditional Undergraduates: Trends in Enrollment from 1986 to 1992 and Persistence and Attainment among 1989-90 Beginning Postsecondary Students. Postsecondary Education Descriptive Analysis Reports. Statistical Analysis Report. Reports Research*. Washington.

Hyman, J. (2018). Nudges, College Enrollment, and College Persistence: Evidence from a Statewide Experiment in Michigan. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.3198881

Iceland, J., & Weinberg, D. H. (2002). *Racial and Ethnic Residential Segregation in the United States: 1980-2000*.

Ioannides, Y. M., & Loury, L. D. (2004). Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature*, *42*(4), 1056–1093.

http://doi.org/10.1257/0022051043004595

Ishitani, T. T. (2006). Studying Attrition and Degree Completion Behavior among First-Generation College Students in the United States. *Journal of Higher Education*, *77*(5), 861–885. http://doi.org/10.1353/jhe.2006.0042

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton: Princeton University Press.

James, E., Alsalam, N., Conaty, J. C., & To, D.-L. (1989). College Quality and Future Earnings: Where Should You Send Your Child to College? *The American Economic Review*, *79*(2), 247–252.

Juan Carlos, C., Crosta, P., Bailey, T., & Jenkins, D. (2006). *Stepping Stones To a Degree: the Impact of Enrollment Pathways and Milestones on Community College Student Outcomes* (CCRC Working Paper No. 4). http://doi.org/10.1007/sl

Juhn, C., Murphy, K. M., & Pierce, B. (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy*, *101*(3), 410–442. http://doi.org/10.2307/2138770

Kane, T. J. (1994). College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education. *Journal of Political Economy*, *102*(5), 878–911. http://doi.org/10.2307/2138651

Kane, T. J., & Rouse, C. E. (1995). Labor-Market Returns to Two- and Four-Year College. *The American Economic Review*. American Economic Association. http://doi.org/10.2307/2118190

Keane, M. P., Wolpin, K. I., Journal, S., June, N., Keane, M. P., & Wolpin, K. I. (1997). The Career Decisions of Young Men Published. *Journal of Political Economy*, *105*(3), 473–522.

Kling, J. R., & Liebman, J. B. (2004). *Experimental Analysis of Neighborhood Effects on Youth*.

SSRN (Vol. 75). http://doi.org/10.2139/ssrn.600596

Laugerman, M., Ii, M. C. S., Rover, D., & Mickelson, S. K. (2015). Estimating Survival Rates in Engineering for Community College Transfer Students Using Grades in Calculus and Physics. *International Journal of Education in Mathematics, Science, and Technology*, *3*(4), 313–321.

Lee, J. O., Jones, T. M., Kosterman, R., Rhew, I. C., Lovasi, G. S., Hill, K. G., … Hawkins, J. D. (2017). The association of unemployment from age 21 to 33 with substance use disorder symptoms at age 39: The role of childhood neighborhood characteristics. *Drug and Alcohol Dependence*, *174*, 1–8. http://doi.org/10.1016/j.drugalcdep.2017.01.005

Light, A. (1995). Hazard model estimates of the decision to reenroll in school. *Labour Economics*, *2*(4), 381–406. http://doi.org/10.1016/0927-5371(95)80042-V

Lyle, D. S. (2007). Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *The Review of Economic and Statistics*, *89*(2), 289–299.

Mangold, W. D., Bean, L. G., Adams, D. J., Schwab, W. A., & Lynch, S. M. (2002). Who Goes Who Stays: An Assessment of the Effect of a Freshman Mentoring and Unit Registration Program on College Persistence. *Journal of College Student Retention: Research, Theory & Practice*, *4*(2), 95–122. http://doi.org/10.2190/CVET-TMDM-CTE4-AFE3

Min, Y., Zhang, G., Long, R., Anderson, T., & Ohland, M. (2011). Nonparametric Survival Analysis of the Loss Rate of Undergraduate Engineering Students. *Journal of Engineering Education*, *100*(2), 349–373.

Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the Retention of University

Students. *Research in Higher Education*, *40*(3), 355–371.

National Student ClearingHouse Research Center. (2018). *Current Term Enrollment Estimates: Fall*. http://doi.org/10.1%

Perna, L. W. (2000). Differences in the Decision to Attend College among African Americans, Hispanics, and Whites. *The Journal of Higher Education*, *71*(2), 117. http://doi.org/10.2307/2649245

Rowan-Kenyon, H. T. (2007). Predictors of Delayed College Enrollment and Impact of Socioeconomic Status. *The Journal of Higher Education*, *78*(2), 188–214.

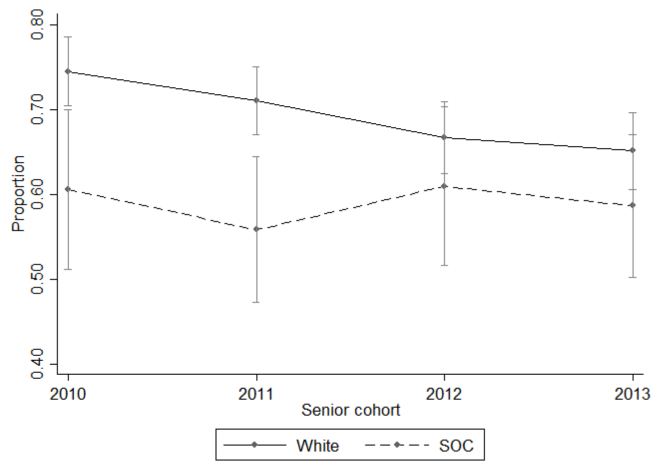
Sacerdote, B. (2011). *Peer Effects in Education: How might they work, how big are they and how much do we know Thus Far? Handbook of the Economics of Education* (1st ed., Vol. 3). Elsevier B.V. http://doi.org/10.1016/B978-0-444-53429-3.00004-1

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "Neighborhood Effects": Social Processes and New Directions in Research. *Annual Review of Sociology*, *28*(1), 443–478. http://doi.org/10.1146/annurev.soc.28.110601.141114

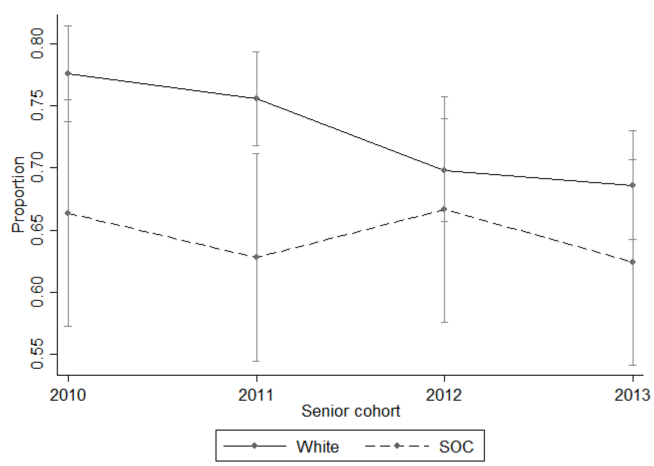
Sewell, W. H., & Armer, J. M. (1966). Neighborhood Context and College Plans. *American Sociological Review*, *31*(2), 159–168.

Appendices

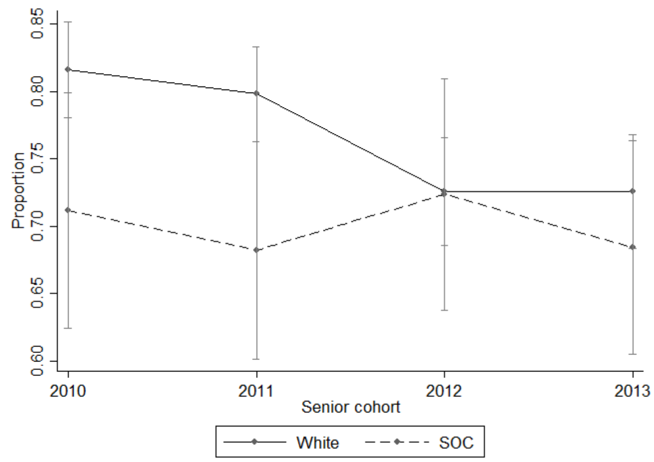
A. College Enrollment, by Race and Sex



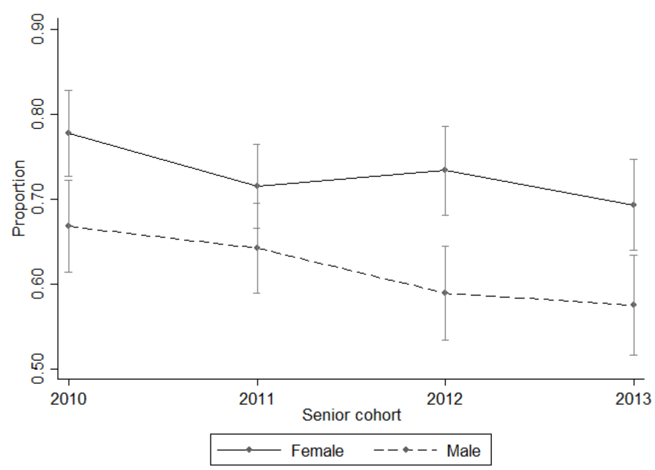
1 year enrollment, by race.



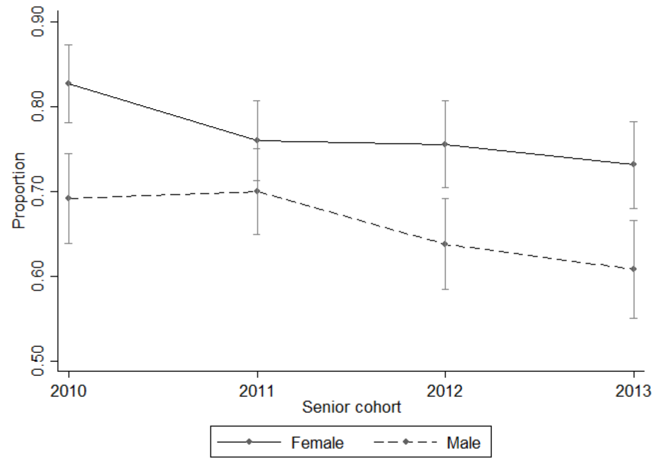
2 year enrollment, by race.



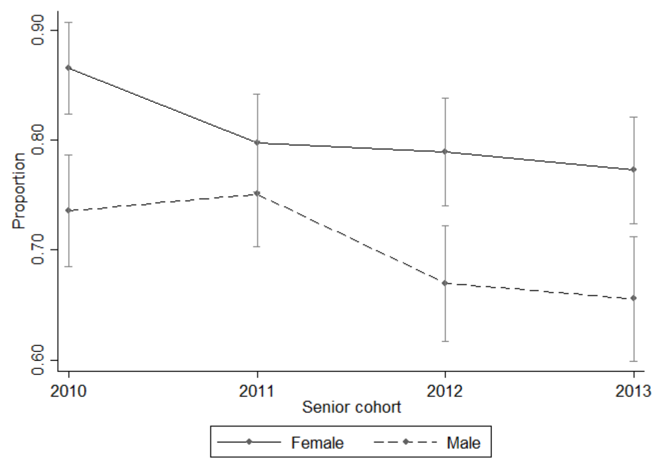
Ever-enrollment, by race.



1 year enrollment, by sex.

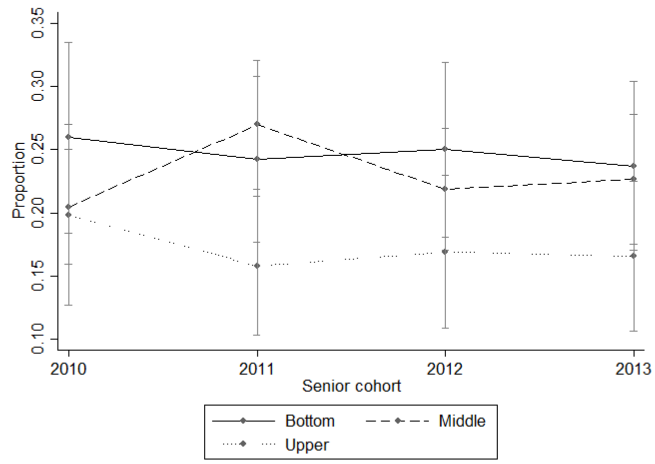


2 year enrollment, by sex.

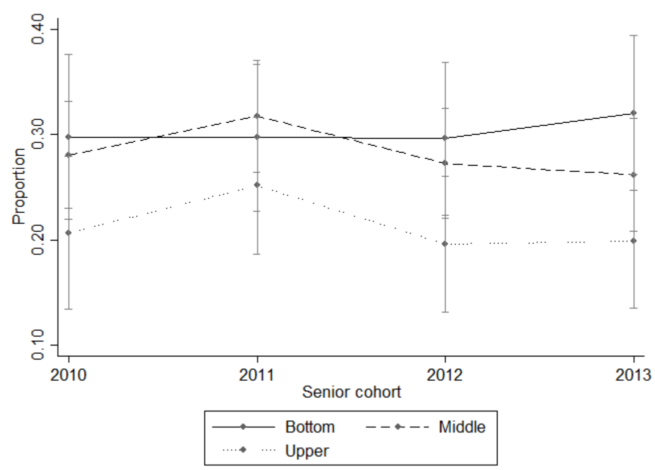


Ever enrollment, by sex.

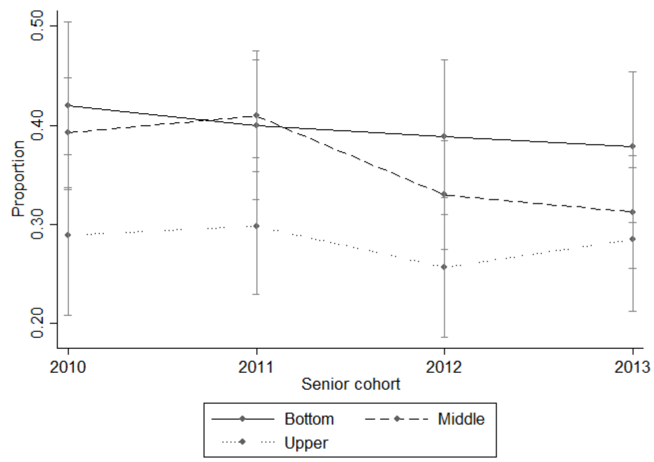
B. College Enrollment, by Neighborhood Income Level



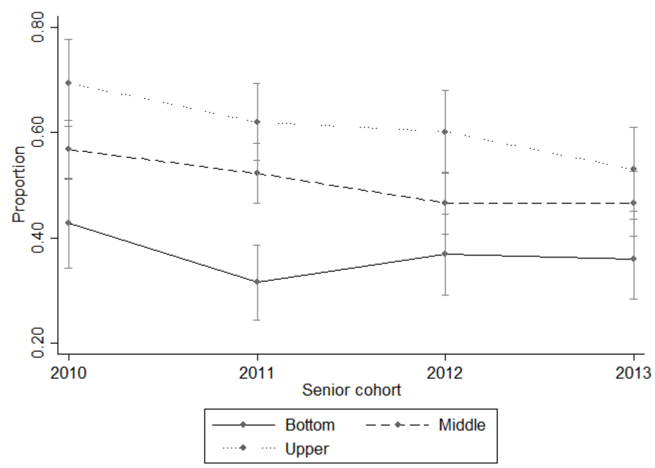
1 year enrollment in 2 year institutions, by median family income in neighborhood.



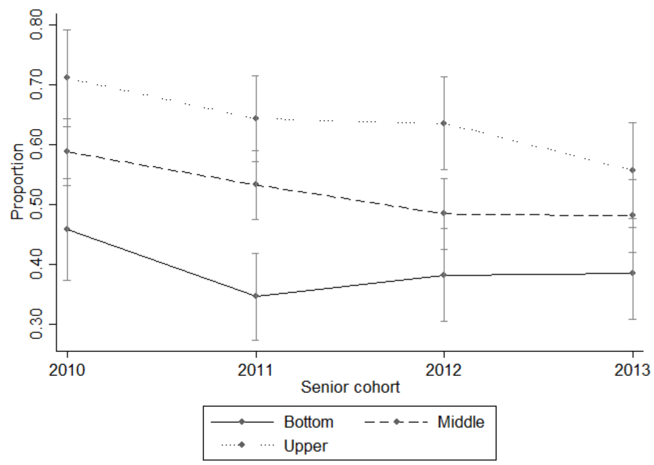
2 year enrollment in 2 year institutions, by median family income in neighborhood.



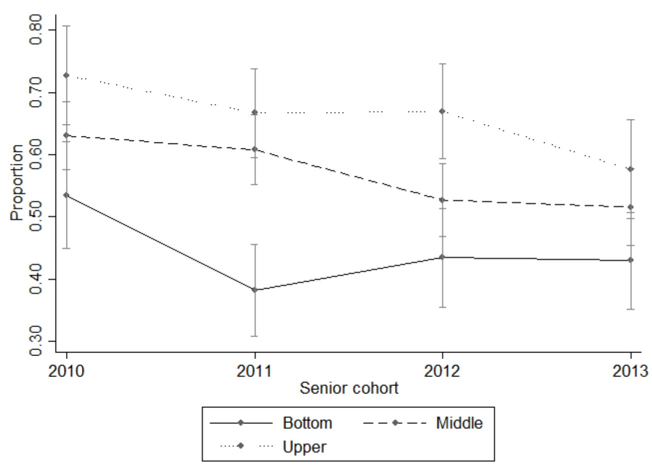
Ever enrollment in 2 year institutions, by median family income in neighborhood.



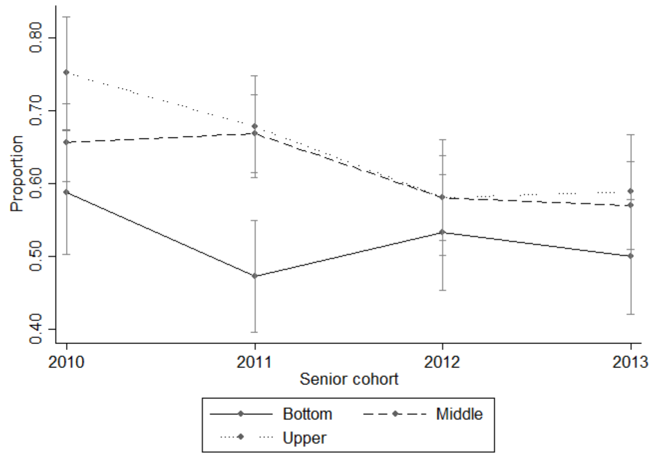
1 year enrollment in 4 year institutions, by median family income in neighborhood.



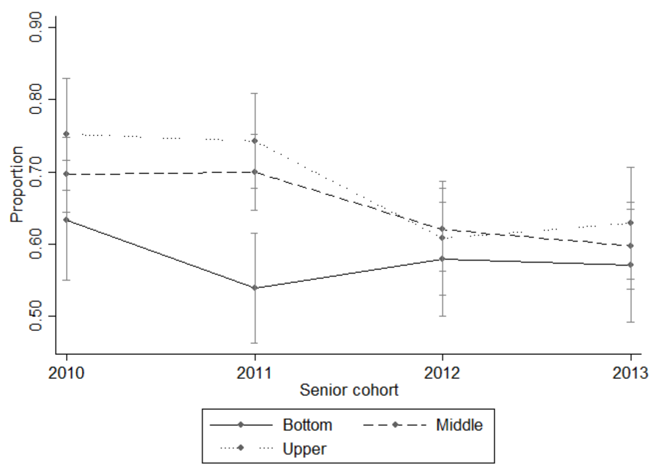
2 year enrollment in 4 year institutions, by median family income in neighborhood.



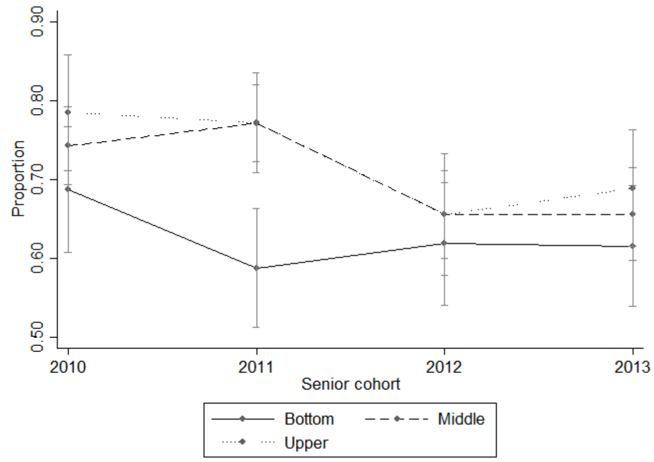
Ever enrollment in 4 year institutions, by median family income in neighborhood.



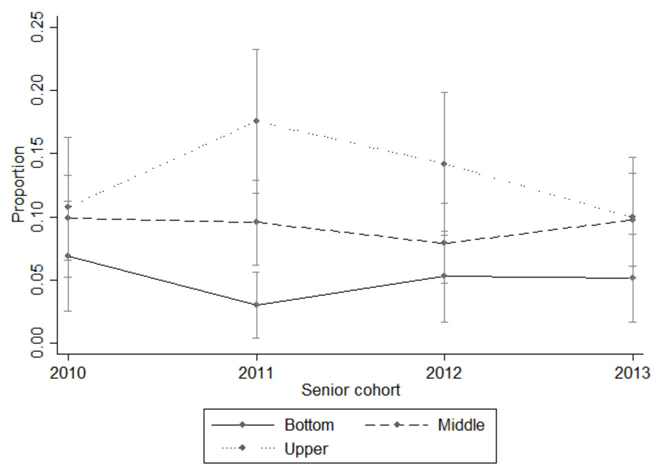
1 year enrollment in public institutions, by median family income in neighborhood.



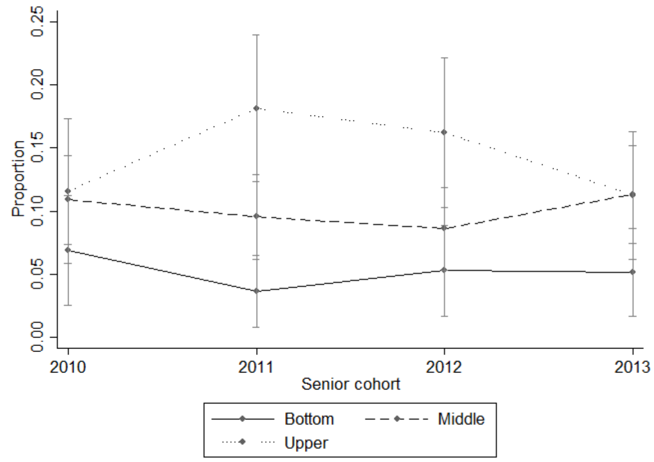
2 year enrollment in public institutions, by median family income in neighborhood.



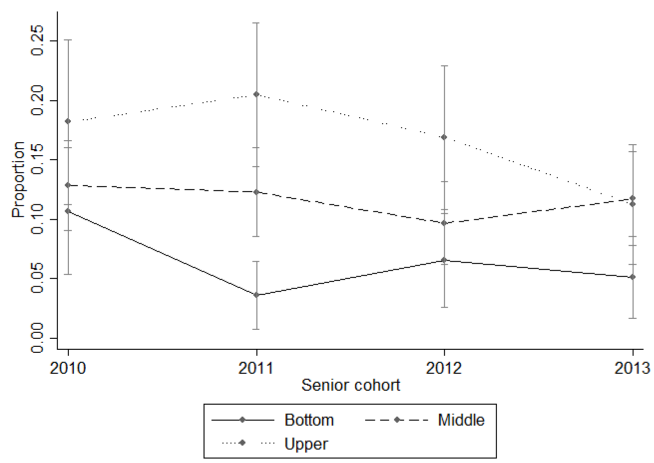
Ever enrollment public institutions, by median family income in neighborhood.



1 year enrollment in private institutions, by median family income in neighborhood.

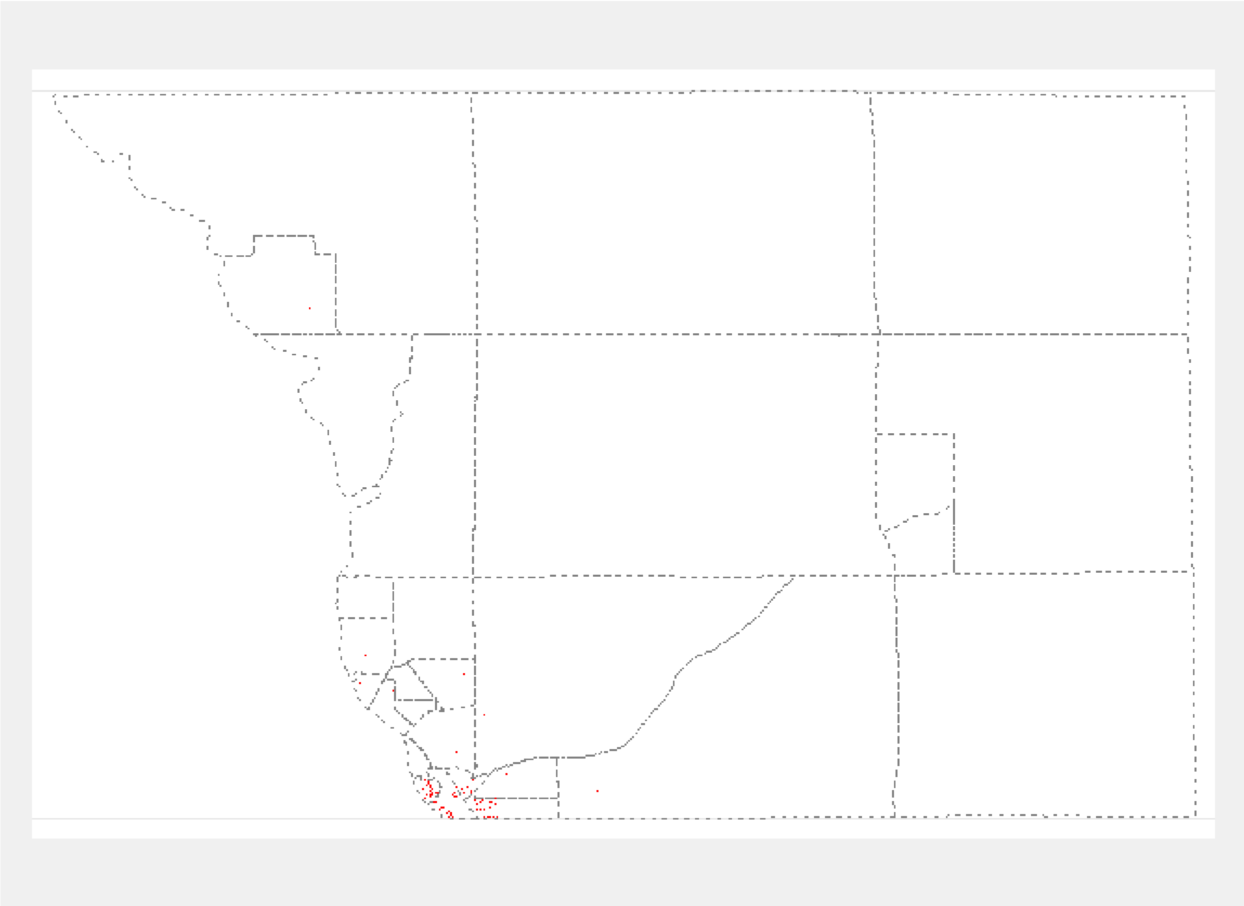


2 year enrollment in private institutions, by median family income in neighborhood.

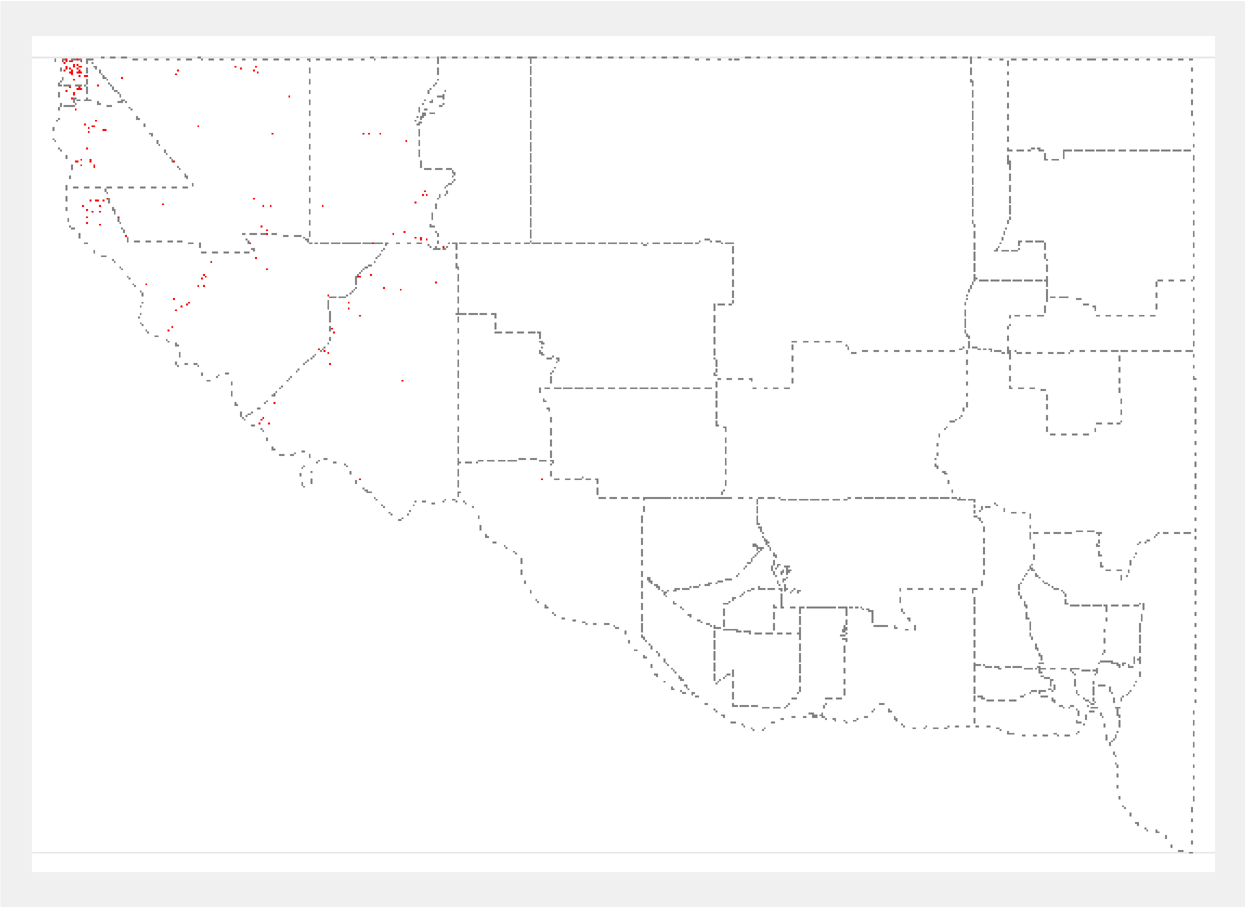


Ever enrollment in private institutions, by median family income in neighborhood.

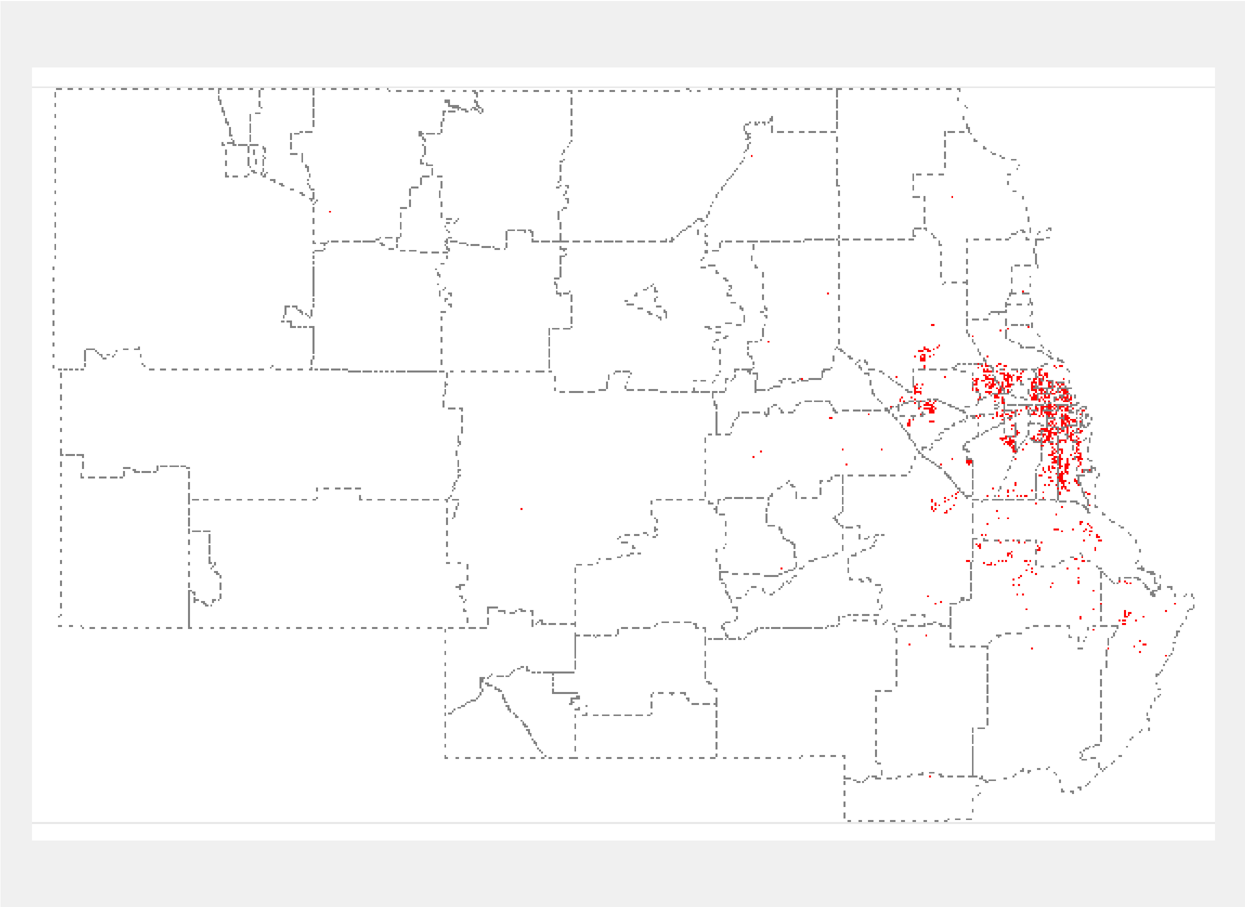
C. Geographic Distribution, by County



Geographic distribution in Benton County.



Geographic distribution in Sherburne County.



Geographic distribution in Stearns County.

D. OLS Regression

This appendix considers OLS models comparable to logistic regressions in Table 4-1, Table 4-2 and Table 4-3. Although logistic models have interpretative advantages when dealing with dichotomous variables, OLS is more easily compared in magnitude with the Two Step regressions. We check robustness of our estimates by introducing different variables.

1 YR College Enrollment, Linear Regression.

VARIABLES	(1) Individual	(2) Individual	(3) Neighborhood	(4) Neighborhood	(5) Neighborhood	(6) Neighborhood
Female	0.113***	0.0404**	0.112***	0.0403**	0.107***	0.0364*
	(0.0189)	(0.0187)	(0.0189)	(0.0188)	(0.0191)	(0.0187)
FRL	-0.196***	-0.0597**	-0.177***	-0.0578**		
	(0.0229)	(0.0233)	(0.0239)	(0.0240)		
LEP	0.0508	0.0175	0.0621	0.0214		
	(0.0449)	(0.0459)	(0.0453)	(0.0465)		
SOC	-0.0287	0.0183	-0.0224	0.0193	-0.0657**	0.00972
	(0.0338)	(0.0356)	(0.0338)	(0.0359)	(0.0265)	(0.0273)
Cohort = 2011	-0.0254	0.00356	-0.0279	0.00353	-0.0431*	0.000300
	(0.0260)	(0.0244)	(0.0259)	(0.0244)	(0.0262)	(0.0244)
Cohort = 2012	-0.0372	-0.0386	-0.0386	-0.0383	-0.0614**	-0.0454*
	(0.0270)	(0.0255)	(0.0270)	(0.0256)	(0.0271)	(0.0254)
Cohort = 2013	-0.0597**	-0.0414	-0.0629**	-0.0413	-0.0839***	-0.0474*
	(0.0274)	(0.0262)	(0.0274)	(0.0263)	(0.0276)	(0.0263)
Family Income pct = 1, 25 pct			-0.0337	-0.00410	-0.0697**	-0.0131
			(0.0311)	(0.0311)	(0.0311)	(0.0308)
Family Income pct = 2, 75 pct			0.0159	-0.00585	0.0249	-0.00428
			(0.0265)	(0.0251)	(0.0269)	(0.0251)
Segregation			-0.0637	-0.0372	-0.0746	-0.0398
			(0.102)	(0.102)	(0.101)	(0.101)
Education High			0.106	0.0119	0.160	0.0256
			(0.119)	(0.115)	(0.120)	(0.115)
CGPA		0.224***		0.223***		0.233***
		(0.0137)		(0.0138)		(0.0133)
Constant	0.723***	0.0343	0.680***	0.0326	0.631***	-0.0103
	(0.0218)	(0.0500)	(0.0515)	(0.0666)	(0.0514)	(0.0642)
Observations	2,326	2,098	2,326	2,098	2,326	2,098
R-squared	0.059	0.168	0.063	0.168	0.037	0.166

2 YR College Enrollment, Linear Regression.

VARIABLES	(1) Individual	(2) Individual	(3) Neighborhood	(4) Neighborhood	(5) Neighborhood	(6) Neighborhood
Female	0.111***	0.0440**	0.110***	0.0443**	0.105***	0.0400**
	(0.0182)	(0.0180)	(0.0183)	(0.0181)	(0.0185)	(0.0181)
FRL	-0.184***	-0.0627***	-0.168***	-0.0628***		
	(0.0224)	(0.0229)	(0.0234)	(0.0235)		
LEP	0.0414	0.0259	0.0484	0.0257		
	(0.0439)	(0.0450)	(0.0445)	(0.0457)		
SOC	-0.0113	0.0238	-0.00646	0.0234	-0.0523**	0.0141
	(0.0328)	(0.0344)	(0.0329)	(0.0347)	(0.0259)	(0.0266)
Cohort = 2011	-0.0119	0.00869	-0.0150	0.00823	-0.0293	0.00470
	(0.0248)	(0.0235)	(0.0248)	(0.0236)	(0.0250)	(0.0236)
Cohort = 2012	-0.0386	-0.0435*	-0.0405	-0.0438*	-0.0619**	-0.0515**
	(0.0260)	(0.0246)	(0.0260)	(0.0247)	(0.0262)	(0.0246)
Cohort = 2013	-0.0625**	-0.0437*	-0.0663**	-0.0442*	-0.0863***	-0.0508**
	(0.0265)	(0.0254)	(0.0265)	(0.0255)	(0.0267)	(0.0255)
Family Income pct = 1, 25 pct			-0.0147	0.0102	-0.0492	0.000475
			(0.0305)	(0.0303)	(0.0305)	(0.0301)
Family Income pct = 2, 75 pct			0.0220	-0.00116	0.0305	0.000547
			(0.0255)	(0.0244)	(0.0259)	(0.0245)
Segregation			-0.0436	-0.0143	-0.0571	-0.0165
			(0.0997)	(0.0996)	(0.0992)	(0.0989)
Education High			0.132	0.0415	0.182	0.0565
			(0.114)	(0.110)	(0.115)	(0.111)
CGPA		0.204***		0.204***		0.214***
		(0.0135)		(0.0136)		(0.0132)
Constant	0.754***	0.128***	0.697***	0.110*	0.651***	0.0639
	(0.0211)	(0.0494)	(0.0497)	(0.0645)	(0.0497)	(0.0623)
Observations	2,326	2,098	2,326	2,098	2,326	2,098
R-squared	0.056	0.157	0.059	0.157	0.035	0.153

Ever College Enrollment, Linear Regression

VARIABLES	(1) Individual	(2) Individual	(3) Neighborhood	(4) Neighborhood	(5) Neighborhood	(6) Neighborhood
Female	0.104***	0.0406**	0.103***	0.0405**	0.0990***	0.0370**
	(0.0174)	(0.0172)	(0.0174)	(0.0173)	(0.0176)	(0.0173)
FRL	-0.159***	-0.0570***	-0.142***	-0.0548**		
	(0.0215)	(0.0220)	(0.0225)	(0.0226)		
LEP	-0.00467	-0.00338	0.00214	-0.00221		
	(0.0421)	(0.0435)	(0.0428)	(0.0442)		
SOC	0.0160	0.0357	0.0191	0.0344	-0.0370	0.0146
	(0.0309)	(0.0327)	(0.0310)	(0.0330)	(0.0247)	(0.0257)
Cohort = 2011	-0.0114	0.0147	-0.0131	0.0149	-0.0249	0.0120
	(0.0234)	(0.0222)	(0.0234)	(0.0222)	(0.0235)	(0.0222)
Cohort = 2012	-0.0513**	-0.0567**	-0.0523**	-0.0565**	-0.0695***	-0.0626***
	(0.0248)	(0.0237)	(0.0248)	(0.0238)	(0.0249)	(0.0237)
Cohort = 2013	-0.0630**	-0.0457*	-0.0653***	-0.0456*	-0.0823***	-0.0512**
	(0.0252)	(0.0244)	(0.0253)	(0.0245)	(0.0254)	(0.0245)
Family Income pct = 1, 25 pct			-0.0325	-0.0115	-0.0626**	-0.0209
			(0.0296)	(0.0296)	(0.0294)	(0.0293)
Family Income pct = 2, 75 pct			0.00480	-0.00911	0.0120	-0.00766
			(0.0243)	(0.0234)	(0.0246)	(0.0235)
Segregation			-0.0130	0.0161	-0.0361	0.00738
			(0.0967)	(0.0974)	(0.0958)	(0.0968)
Education High			0.137	0.0387	0.178	0.0517
			(0.110)	(0.107)	(0.110)	(0.107)
CGPA		0.176***		0.176***		0.185***
		(0.0135)		(0.0136)		(0.0132)
Constant	0.791***	0.252***	0.739***	0.243***	0.701***	0.202***
	(0.0200)	(0.0493)	(0.0476)	(0.0624)	(0.0474)	(0.0605)
Observations	2,326	2,098	2,326	2,098	2,326	2,098
R-squared	0.050	0.136	0.053	0.136	0.034	0.133

Across all specifications, females have higher likelihood of enrollment and students from low socioeconomic background (FRL recipients) are less likely to enroll. In both cases coefficients are significant at less than 5 percent. GPA has the expected sign, similar magnitude to previous models, and is significant at less than 1 percent. Notably, neighborhood income rank continues to be statistically insignificant, even when census block group unemployment is accounted for in column 1 and when it is substituted by per capita income percentile (instead of median family income) in column 4 and 5. Column 2 excludes high education and uses low level education instead, which accounts for the population that does not have a high school degree, but the variable remains statistically insignificant. Column 3 substitutes our segregation index, described in Data and Methodology, by the proportion of population of color. Finally, column 6 studies the interaction between family economic background (FRL) and race, with interesting results: after controlling by the interaction between income and race, students of color are less likely to enroll in college and the effect is statistically significant but it also shows that students of color from low socioeconomic background are more likely to go into college than white students from low socioeconomic background.

1 Year College Enrollment tests.

VARIABLES	(1) Unemployment	(2) Education	(3) Segregation	(4) Income	(5) Income	(6) Interaction
Female	0.0414**	0.0403**	0.0403**	0.0404**	0.0413**	0.0403**
	(0.0187)	(0.0188)	(0.0188)	(0.0188)	(0.0187)	(0.0187)
FRL	-0.0583**	-0.0578**	-0.0578**	-0.0585**	-0.0589**	
	(0.0239)	(0.0240)	(0.0240)	(0.0238)	(0.0238)	
LEP	0.0212	0.0215	0.0215	0.0205	0.0210	-0.0238
	(0.0465)	(0.0465)	(0.0465)	(0.0466)	(0.0466)	(0.0476)
SOC	0.0181	0.0194	0.0194	0.0199	0.0171	
	(0.0358)	(0.0358)	(0.0359)	(0.0359)	(0.0359)	
CGPA	0.222***	0.223***	0.223***	0.223***	0.223***	0.225***
	(0.0138)	(0.0138)	(0.0138)	(0.0138)	(0.0138)	(0.0138)
Cohort = 2011	0.00282	0.00355	0.00352	0.00342	0.00351	-0.000421
	(0.0245)	(0.0244)	(0.0244)	(0.0245)	(0.0245)	(0.0245)
Cohort = 2012	-0.0388	-0.0384	-0.0383	-0.0386	-0.0383	-0.0385
	(0.0256)	(0.0256)	(0.0256)	(0.0255)	(0.0256)	(0.0255)
Cohort = 2013	-0.0408	-0.0413	-0.0413	-0.0416	-0.0401	-0.0445*
	(0.0263)	(0.0263)	(0.0263)	(0.0263)	(0.0263)	(0.0263)
Family Income pct = 1, 25 pct	0.00358	-0.00449	-0.00378			-0.00570
	(0.0315)	(0.0299)	(0.0311)			(0.0309)
Family Income pct = 2, 75 pct	0.00619	-0.00484	-0.00588			-0.00886
	(0.0264)	(0.0215)	(0.0251)			(0.0250)
Segregation	-0.00353	-0.0305		-0.0496	-0.0175	-0.0720
	(0.104)	(0.119)		(0.0955)	(0.0970)	(0.102)
Education High	-0.0141		0.0120	0.00790	0.0115	0.00663
	(0.115)		(0.115)	(0.0929)	(0.0931)	(0.115)
Unemployment	-0.402				-0.407	
	(0.283)				(0.275)	
Education Low		-0.0177				
		(0.195)				
Non-White/Population			-0.0434			

	(1)	(2)	(3)	(4)	(5)	(6)
			(0.113)			
Income Per Capita pct = 1, 25 pct				0.00703	0.0104	
				(0.0274)	(0.0274)	
Income Per Capita pct = 2, 75 pct				0.00169	-0.00594	
				(0.0217)	(0.0225)	
FRL = 1, FRL						-0.0925***
						(0.0253)
SOC = 1						-0.102**
						(0.0486)
0b.frl#0b.soc						0
						(0)
0b.frl#1o.soc						0
						(0)
1o.frl#0b.soc						0
						(0)
1.frl#1.soc						0.221***
						(0.0611)
Constant	0.0637	0.0384	0.0367	0.0296	0.0535	0.0412
	(0.0694)	(0.0532)	(0.0665)	(0.0621)	(0.0638)	(0.0665)
Observations	2,098	2,098	2,098	2,098	2,098	2,098
R-squared	0.169	0.168	0.168	0.168	0.169	0.174

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Similarly, we test robustness in the GPA First-Step regression in OLS results bellow. Second-Step results are omitted as they are similar to previous findings. Column 1 is the same model found in Table 4-4, as reference. Introducing unemployment in the original model, in Column 2, increases the robustness of income at the top percentile rank and decreases the magnitude of educational attainment. The effect of unemployment, besides, is fairly large and robust. Model 3 explores a different income variable, the percentile rank of income per capita, that is more robust than median family income. This model suggests than neighborhood income does have an effect but inappropriate measurement may distort these results. The effect is robust to introducing unemployment but the overall significance of the model is slightly less according to F-statistics, yet still significant. Finally, column 5 tests the introduction of a different variable to control for race in the neighborhood, with no effect in the model.

High School Performance, First-Step tests.

VARIABLES	(1) GPA M1	(2) GPA M2	(3) GPA M3	(4) GPA M4	(5) GPA M5
Cohort = 2011	-0.0326	-0.0342	-0.0337	-0.0336	-0.0325
	(0.0420)	(0.0420)	(0.0419)	(0.0419)	(0.0420)
Cohort = 2012	-0.0243	-0.0255	-0.0239	-0.0237	-0.0243
	(0.0421)	(0.0421)	(0.0420)	(0.0420)	(0.0421)
Cohort = 2013	-0.0934**	-0.0919**	-0.0935**	-0.0923**	-0.0934**
	(0.0426)	(0.0425)	(0.0424)	(0.0424)	(0.0426)
Female	0.304***	0.306***	0.305***	0.305***	0.304***
	(0.0295)	(0.0295)	(0.0295)	(0.0295)	(0.0295)
FRL	-0.458***	-0.458***	-0.456***	-0.456***	-0.458***
	(0.0355)	(0.0355)	(0.0353)	(0.0353)	(0.0355)
LEP	0.111	0.111	0.109	0.109	0.111
	(0.0678)	(0.0678)	(0.0677)	(0.0677)	(0.0678)
SOC	-0.220***	-0.223***	-0.219***	-0.221***	-0.220***
	(0.0525)	(0.0525)	(0.0523)	(0.0524)	(0.0525)
Family Income pct = 1, 25 pct	-0.0864*	-0.0676			-0.0863*
	(0.0474)	(0.0482)			(0.0474)
Family Income pct = 2, 75 pct	0.0821*	0.111**			0.0821*
	(0.0424)	(0.0446)			(0.0424)
Segregation	-0.0777	0.00408	-0.0320	-0.00832	
	(0.153)	(0.158)	(0.144)	(0.148)	
Education High	0.381**	0.318*	0.455***	0.458***	0.381**
	(0.182)	(0.185)	(0.152)	(0.152)	(0.182)
Unemployment		-0.976**		-0.300	
		(0.474)		(0.456)	
Income Per Capita pct = 1, 25 pct			-0.0990**	-0.0965**	
			(0.0415)	(0.0417)	
Income Per Capita pct = 2, 75 pct			0.111***	0.105***	
			(0.0374)	(0.0384)	
Non-White/Population					-0.0863
					(0.170)
Constant	3.031***	3.100***	2.998***	3.015***	3.040***
	(0.0779)	(0.0848)	(0.0698)	(0.0745)	(0.0775)
Observations	2,098	2,098	2,098	2,098	2,098
R-squared	0.200	0.202	0.204	0.204	0.200
F-stat	47.41	43.88	48.54	44.52	47.41

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

E. Summary of Variables

Summary of Variables.

Variable	Definition	Source
1 YR Enrollment in College	Record in a postsecondary institution lasting more than 30 days at least one year after high school graduation	NSC
2 YR Enrollment in College	Record in a postsecondary institution lasting more than 30 days at least two years after high school graduation	NSC
Ever Enrolled	Record in a postsecondary institution lasting more than 30 days in any Academic Year for a High School 4-Year Cohort graduate.	NSC
Cohort	Year of high school graduation.	MARSS
Female	Binary variable for the sex of the student.	MARSS
FRL	Free and Reduce Lunch. Proxy for family socioeconomic background.	MARSS
LEP	Limited English Proficiency. Proxy for immigrant background.	MARSS
SOC	Student of Color (Non-White)	MARSS
CGPA	Cumulative GPA in High School	MARSS
Family Income	Percentile rank of Median Family Income in the Census Block Group. It was categorized in three percentile levels: bottom 25 percentile, between 25 and 75 percentiles, and upper 75 percentile.	NHGIS
Income Per Capita	Percentile rank of Income per capita in the Census Block Group. It was categorized in three percentile levels: bottom 25 percentile, between 25 and 75 percentiles, and upper 75 percentile.	NHGIS
Non-White Proportion	Proportion of Non-White Population in the census block group.	NHGIS
Segregation	Exposure Index using proportion of non-white population.	
Education High	Proportion of the Population with College degree or more as educational attainment	NHGIS