Opportunity Bound: Transport and Access to College in a Megacity

Fabiola M. Alba-Vivar ⋆

(Job Market Paper)

November 4, 2023

Click here for the most recent version

Abstract

This paper examines the causal effects of new transportation infrastructure on college enrollment, choice, completion, and early labor market outcomes. I use novel geolocated administrative data to estimate a difference-in-differences model that exploits the rollout of two new public transportation lines in Lima, a megacity of 12 million people. My findings indicate that at the neighborhood level, a 17% reduction in commuting time to college increases enrollment rates by 6%, primarily driven by private college enrollment. Moreover, female students influenced by this policy tend to enroll in low-quality private colleges, which are also connected to the new lines. In contrast, male students are more likely to enroll in public colleges, which are more dispersed throughout the city. Using a model of college choice, I find that for one standard deviation increase in wage returns, male students are willing to commute up to 55% more minutes than female students. In the medium and long run, access to transport increases an individual's likelihood of graduating from college by 12% and access to white-collar jobs by 6%. These results suggest that while improved transportation can increase human capital accumulation, the increase in opportunities is limited by gender differences in willingness to travel.

Keywords: College Access, College Choice, Transport
JEL Codes: I24, I25, O18, R41

I thank Alex Eble, Cristian Pop-Eleches, Sandra Black, and Micaela Sviatschi for their constant guidance and support. I also gratefully acknowledge all the comments and suggestions from Sofia Amaral, Mark Anderson, Sam Asher, Andy de Barros, Michael Best, Daniel Bjorkegren, Girija Borker, Kirill Borusyak, Carolina-Concha Arriagada, Sarah Cohodes, Kevin Donovan, Paul Goldsmith-Pinkham, Jose Flor-Toro, Carolina Lopez, Stephen Keating, Bentley MacLeod, Maria Adelaida Martinez, Matteo Magnaricotte, Stephen Ross, Sebastian Otero, Nishith Prakash, Randall Reback, Maria Rosales-Rueda, Sebastian Tello-Trillo, Anja Tolonen, Daniel Velasquez, Eric Verhoogen, and Jack Willis. I also thank Florian Grosset, Parijat Lal, Seung-hun Lee, and Daniel Sparks, who provided invaluable support at different stages of this project. I am also thankful to all participants at the WEAI Graduate Student Workshop, AEFP, UEA, Seminario MAP, the Development Colloquium at CU, and the Economics of Education Colloquium at TC. This project was possible thanks to the generous funding of the Center for Development Economics and Policy at CU and the Department of Education Policy & Social Analysis at TC. I am deeply grateful to Roberto Mendoza, who provided excellent research assistance. I have benefited from data provided by the Peruvian Ministry of Education (MINEDU) and the National Institute of Statistics (INEI). I also thank members of the Dirección General de Educación Superior Universitaria at MINEDU for their support in this project.

⋆ Teachers College (TC) at Columbia University, *fma2139@tc.columbia.edu*, New York, NY.

1 Introduction

The majority of the world's college students, numbering over 500 million, live in cities, relying on urban infrastructure that often proves slow, unsafe, and uncertain. This challenge is especially pronounced in low- and middle-income countries, where travel speed and safety heavily rely on this infrastructure (Borker, 2020; Kondylis et al., 2020; Kreindler et al., 2023; Kreindler and Miyauchi, 2021). Despite a large body of evidence suggesting that distance to school affects various educational outcomes (Agarwal and Somaini, 2019; Card, 2001; Kane and Rouse, 1995), far less is known about how place-based policies that reduce commuting time can directly enhance college access,[1] a key driver of economic mobility (Chetty et al., 2017). In this sense, improving public transportation could not only increases access to better employment prospects and increase welfare in the long run (Balboni et al., 2020; Tsivanidis, 2022; Zarate, 2022) but also increases access to educational opportunities in the short run.

This paper examines the causal impacts of dramatic improvements to public transportation infrastructure on college access and choice. The study occurs in the context of Lima, the capital of Peru and a megacity of 12 million people. For identification, I use the opening of two new mass public transportation systems in the early 2010s—a rapid bus transit line and a new train line, which together reduced commuting time to college for thousands of college students each year. Before these new transit lines were rolled out, students in Lima spent an average of almost two hours per day commuting to and from school.[2]

I create a novel dataset of college and student location and transit options that captures variance in commuting time and educational choice over time. I geocode the locations of both colleges and students' households and then link these to data on the locations and opening dates of the public transit stations from the two lines. To generate causal estimates

[1] Most of the studies related to policies that increase college enrollment focus on targeting financial constraints or granting affirmative action to secure equal opportunities for minorities (Page and Scott-Clayton, 2016).

[2] According to the 2010 National University Census data, students living in the outskirts of the city travel, on average, 1.5 hours from home to college, whereas those living in downtown Lima travel 40 minutes, on average. This commuting time is similar to college students in Delhi as documented by Borker (2020).

of how access to improved public transportation affects college enrollment and choice, I use a difference-in-differences (DiD) approach. This approach exploits variation by cohort and neighborhoods as well as the staggered nature of the station openings, comparing educational choices and outcomes of cohorts in neighborhoods exposed to new stations to the same cohorts in neighborhoods that were exposed to planned-but-not-executed stations,[3] similar to the placebo strategy implemented by Donaldson (2018).

I find that students in neighborhoods that are connected to new lines reduce their time commuting to college by 17%,[4] which leads to a 6.3% increase in college enrollment rates. This increase is mostly driven by private colleges, while the results for public colleges remain positive but significantly lower. In Peru, public colleges are tuition free, are highly selective, and have limited ability to expand enrollment, which may explain why the effects are low.[5] These effects are also driven by non-poor neighborhoods, and I do not observe significant differences in the effects by gender.

I also find that a decrease in commuting time to college affects college choice. Lacking data on preferences and abilities, I use machine learning techniques to identify students who, in the policy's absence, were very likely or unlikely to attend college. I further determine if they were likely to enroll in a low-quality private college, a high-quality private college, or a public college. This approach allows me to disentangle the overall effects between the extensive margin (sample of new students) and the intensive margin (sample of typical students, who were likely to go to college either way). I find that new students are more likely to enroll in low-quality private and public colleges, while typical students show an increased likelihood of enrolling in a public college.

This change in college choice varies significantly by gender. For the new student sample,

[3] I collect information on Lima's strategic plans for the city as well as several transportation studies that were used to build the Metro de Lima plan, which included eight lines connecting several areas of the city. Only six of them were properly studied and examined, and only one was actually built.

[4] This reduction in average commuting time to any college is equivalent to almost 30 minutes per day.

[5] Flor-Toro and Magnaricotte (2021) document the disparities among the admissions systems for both private and public colleges in Peru. They highlight the different probabilities of admissions that students face, where the likelihood of being admitted to a public college is less than a private one.

I find that women are more likely to attend low-quality private colleges, whereas men are more likely to enroll in public colleges. Conversely, for the typical student sample, women show no change in their college choice, while men forgo low-quality private colleges in favor of public ones. The fact that women choose low-quality private colleges while men gravitate toward free public ones—which yield higher labor market returns—is noteworthy given the large gender wage gap in the Peruvian labor market.

Motivated by the differential results by gender in terms of college choice, I use a simple random utility model to explore the relationship between commuting time to college and college quality, measured by the expected wage returns of recent graduates. I estimate a mixed logit model separately by gender for two types of students, those who are likely to attend public colleges and those who are likely to attend private ones. For students likely to choose private colleges, I find that men are willing to travel 55% more than women for one additional standard deviation in post-graduation wage returns. For public college, the gap narrows: on average, men are willing to travel 7% more than women. These differentials in willingness to travel align with my reduced-form findings, which suggest that women tend to enroll in lower-quality colleges near the new stations compared to men.

Additionally, I explore medium-term results using the 2017 Peruvian Census, seven years after the first station's opening. Using a cohort-exposure analysis, I find that access to improved public transportation increases the likelihood of college completion by 12% among students living in affected areas. This implies a major reallocation of time and resources given that most Peruvian college students tend to take more than five years to complete their education. These effects are particularly higher for women and low-income students. I also examine longer-term outcomes such as employment rates. My findings suggest a positive effect on employment rates for those who enrolled in college. More importantly, these effects arise from these attendees securing in white-collar positions.

This paper contributes to several strands of the literature. First, it contributes to the literature studying the impact of improving or building new transportation in large cities in developing countries. The existing evidence shows positive effects on labor market opportunities and welfare in the medium and long run (Balboni et al., 2021; Tsivanidis, 2022; Zarate, 2022).[6] Nevertheless, less is known about the direct causal effects on human capital investment. More specifically, in terms of higher education, and to the best of my knowledge, this paper is the first to study the impact of improved transportation on college access and choice. I document how the ways in which neighborhoods experience reduced commuting time to any college in the city and how this boosts college enrollment rates.[7]

Despite the lack of data and identification challenges, evidence indicates that reduced transportation costs affect human capital investment in K-12 settings in developing countries. Bicycles, for instance, bolster women's access to education and improve schooling outcomes and aspirations (Fiala et al., 2022; Muralidharan and Prakash, 2017).[8] Moreover, reduced transportation costs can reduce gaps between low- and high-skill students (Asahi and Pinto, 2022), as families with reduced transportation costs often travel further (Herskovic, 2020). However, in places with meritocratic systems, only high-achieving students with highly educated parents take advantage of reduced transportation costs (Dustan and Ngo, 2018). My results show that living in a neighborhood connected to new lines not only increases college enrollment rates but also influences college selection. I also find differential responses by gender: male students become more likely to enroll in public colleges spread across the city, while female students enroll in low-quality private colleges that now connect

[6]Research on different types of transportation such as railroads and roads (Brooks and Donovan, 2020; Donaldson, 2018; Donaldson and Hornbeck, 2016) shows long-run, positive effects on trade and economic growth. However, Severen (2023) finds no impact on local productivity or amenities for the Los Angeles Metro Rail. Studies also find positive results for commuters on buses (BRTs) (Balboni et al., 2021; Tsivanidis, 2022) and metros (Zarate, 2022). Warnes (2021) shows that the BRT line in Buenos Aires increased segregation between high- and low-skill workers.

[7]I also find not only reduced commuting time but also that the new system provides better labor market opportunities, as it connects people from the outskirts of the city to downtown Lima (where most white-collar jobs are located). In this sense, potential college students might also anticipate better job prospects after graduating. Adukia et al. (2020) find that children stay in school longer and perform better on standardized exams in rural areas that get connected to roads. In this paper, I show reduced-form results that combine both channels: increased labor market opportunities (both for current jobs and those obtained after graduation) and reduced commuting times.

[8]Both papers study a similar policy: providing bicycles to female students so they can mobilize easily and therefore reducing transportation costs.

to the new lines.

Travel safety is another factor highly correlated to transportation costs. In Lima, informal transportation is seen as risky due to frequent accidents, sexual harassment, and muggings (Dominguez Gonzalez et al., 2020). The new transportation systems I study provide a safer ride compared to the informal buses that circulate the city. In this sense, I expect the reduction in transportation costs to also reflect an increase in travel safety for women. There is a small but growing literature on this matter, where most results show that women have a higher demand for safe transportation, which can directly affect their labor supply (Field and Vyborny, 2022; Kondylis et al., 2020).[9].

Additionally, travel safety can directly impact human capital investment. In Delhi, Borker (2020) documents how women are willing to choose a lower-quality college over a prestigious but less safe college. I focus on Lima, which is considered one of the worst large cities in the world for women's mobility and transportation, comparable to Delhi, Mexico City, and Jakarta.[10] I provide causal effects of reduced commuting time that induces women to enroll in lesser-quality private colleges connected to the new stations, thereby minimizing their exposure to harassment and facilitating their commute. In contrast, men enroll in public colleges in general and not just the ones that get connected to the new lines, suggesting that they are moving more across the city. What is more, my model estimates also suggest that men are willing to travel more minutes than women for identical expected wage returns.

The rest of the paper is organized as follows. Section 2 describes the background, Section 3 describes the data, and Section 4 presents the empirical strategy. Section 5 presents the short-term results, and Section 6 introduces a college choice model to explore the trade-off between college quality and commuting time to college. Section 7 shows the medium- and long-term results, and Section 8 concludes.

[9]In Lima, the new stations positively impacted women's labor supply (Martinez et al., 2020)

[10]See the ranking on women's safety in transportation in a Reuters study here.

2 Background

In this section, I describe the college education system in Peru along with the state of transportation for students in the city.

2.1 College Education in Peru

The Peruvian Education system is based on three levels: primary education (6 years), secondary education (5 years), and the higher education level, which often lasts from 2 (technical school) to 10 years (medical school). On average, college students graduate in 5 to 7 years. According to the 2017 Peruvian Census, 4 out of 10 recent high school graduates (between 17 and 21 years old) have access to some type of higher education. More specifically, 15% enroll in a technical school or community college, 22% enroll in college, and the remaining 63% lack access to any type of higher education (Alba-Vivar et al., 2020). Following trends similar to those in middle- and high-income countries, Peruvian women access college at slightly higher rates than men.

Several additional aspects of the college education system are worth highlighting, as they are relevant to understanding students' college choices. Unlike the US, Peru has no general standardized exams like the SAT, making it difficult to assess students' abilities. Also, unlike Chile or Brazil, there is no centralized admission system to enroll in college, as each college maintains its own distinct procedure to admit students. Students typically face a decision tree of college choice as depicted in Figure 1. A student might choose to attend college, community college, or no college at all (and either stay home or work, which accounts for approximately 60% of recent high school graduates). If a student chooses to attend college, they might choose either a private or public college.

There are several differences between private and public colleges. On the one hand, public colleges offer free tuition and have a decentralized admission system (i.e., each university organizes its own admission exam). They also receive a significant number of applicants each year, making them more selective (about 20% of applications at public colleges are successful Flor-Toro and Magnaricotte, 2021). This makes the process more uncertain for students, but it is compensated by the prestige that these public schools have, especially the STEM-oriented institutions.

On the other hand, private colleges have a greater variance in price and quality. Their admission process is very straightforward, and students face less or no uncertainty about the probability of enrolling, as only a few elite private colleges tend to have more selective admission exams. For simplification, I assume there are two types of private institutions: high- and low-quality private colleges. High-quality colleges tend to charge higher tuition fees compared to low-quality colleges (college fees range from 150 USD monthly up to 1,300 USD. For reference, the minimum monthly wage in Peru was 180 USD in 2010). For private universities, prestige is correlated with prices given the selection of their students. Graduates from low-cost and low-quality colleges enjoy fewer returns in the labor market compared to their peers in high-cost and high-quality institutions. Appendix Figure A.4 shows the wage returns for all available choices, which provides evidence that higher quality colleges yield in significant higher returns in the labor market.

Another relevant fact is that 90% of college students attend a college located in their province of birth, suggesting that most do not move to go to college. In Lima, housing for students is almost nonexistent, and the housing that is available is reserved exclusively for out-of-state students. This means that most students live at home with their parents and commute to college. Appendix Figure A.1 shows the average travel time from home to their college campus in minutes using data from the 2010 National University Census. Students living on the outskirts of the city travel an average of 80 minutes, whereas those living in downtown Lima (city center) travel an average of 40 minutes. In this setting, where students spend a significant part of their day commuting to college, it is expected that reduced commuting time can switch both their decisions to attend college and which college to attend.

2.2 Public Transportation in Lima

Lima's population is comparable to other large cities around the world, such as New York City, Paris, Xi'an, Chennai, Jakarta, Bogota, and Los Angeles. However, Lima is not nearly as dense (8,000 hab/mi2 compared to New York City's 29,302 hab/mi2), and commuting across the city can take up to three hours during rush hour. During the 1990s, market liberation policies allowed used cars and mini-buses to be imported, which became the basis of the new transportation system for commuters. These privately operated mini-buses, known as *combis*, partially alleviated the demand for transportation across the city. However, their poor quality and the lack of transit regulations made this mode of transportation unsafe for commuters and even more for young students. In 2010, 57% of people who were in a traffic accident were under 25 years of age.

In July 2010, the city opened Metropolitano, a single bus rapid transit line that connected the north and south of the city. The Metropolitano was the very first mass transportation public system in Peru, connecting 12 city districts out of the 44. Regular commuters had to pay a flat fee of 1.50 PEN, approximately 0.50 USD for regular commuters, but college students received a 50% discount.[11] A year after the Metropolitano opened, the Peruvian president inaugurated the first line of the Metro de Lima, which was built on an elevated viaduct.[12] This train connected the northeast side of the city with the southeast side and over two million people.

The Metro de Lima took almost 40 years to complete. During the 1970s, the Peruvian Ministry of Transport designed a complete metro system for the city, aiming to connect multiple districts, especially emerging neighborhoods in the outskirts. The original plans for the city included eight lines, but as years passed, only six were properly studied and evaluated (Appendix Figure A.2 shows these lines). A very small part of the project was initiated in the 1980s during the first government of President Alan Garcia, but it remained

[11] The original fee was 1.50 PEN, but it was raised in December 2012 to 2.00 PEN and then raised again by February 2015 to 2.50 PEN.

[12] It was the longest metro-type train viaduct in the world for six years, until it was overtaken by the Wuhan Metro in 2017.

incomplete for more than 20 years and never opened to the public. Peru's major economic crash avoided future developments of this project until it was revived in 2006, during Garcia's second presidential term (Campos et al., 2021).

Due to budgetary restrictions and President Garcia's wish to inaugurate the project he had promised in the 1980s, the first line (Linea 1) was rushed to open before it was completed, before the end of Garcia's term in 2011. The remaining half was subsequently inaugurated in 2014. Until today, no other lines have opened to the public, and the second line has been under construction since 2014. These delays are mostly due to several corruption scandals involving Garcia's government and Odebrecht, the consortium in charge of the first line's construction, which allegedly paid more than 20 million USD in bribes for this project. Appendix C provides more information on the history of these projects.

In this paper, I focus on the first line of the Metro de Lima and the Metropolitano's new stations. Both provided a faster and safer service compared to combis, reducing transportation costs for thousands of students in Peru's capital. Notably, both systems crossed the city from north to south, connecting several neighborhoods to downtown Lima, which is the hub of several college campuses as seen in Figure 2.

3 Data

This paper uses multiple sources of data that include administrative data from college records, geocoded stations, and the Peruvian census at the block level.

College Enrollment. Data on college enrollment come from the Peruvian Ministry of Education, which annually compiles enrollment data for every college in the country. These records contain information about students' year of enrollment, college, home address, major, age, and gender. I restrict my sample to students whose home addresses are located within the Lima and Callao region boundaries (Lima's metropolitan area). I use Google Maps API to collect GPS coordinates for their home addresses. For less than 5% of the total cases, where the algorithm failed, I impute GPS coordinates at the block or neighborhood level.[13] Additionally, I refine the sample by including only recent high school graduates or students under 20 years old for the analysis. The study's time frame ranges from 2006 to 2014.[14]

Geocoded College Campuses. I manually collect and geocode the locations of 44 college campuses in Lima's metro area. The addresses are obtained from the 2010 university census compiled by the Ministry of Education and the National Institute of Statistics and Informatics (INEI), verified using Google Maps API. Figure 2 plots the resulting GPS coordinates in red dots.

Peruvian Census. The 2007–2017 Peruvian Census data come from the INEI. Both datasets are geocoded at the block level.[15] I use the data from 2007 to obtain block-level counts by age and use this as the denominator for college enrollment rates at the block level. Then, I use the 2017 data to explore long-term outcomes such as college completion and employment status. Here, I restrict my sample to individuals living in Lima's metro area and between 17 and 28 years old.

Transportation. All data on stations from the new transportation systems come from the Autoridad de Transporte Urbano para Lima y Callao (ATU). This includes the GPS location and address of all stations. Data on Metro de Lima's planned-but-not-executed lines come from the national government's (Ministry of Transport) multiple technical records (for details, see Appendix C). I geocode all planned stations from six routes as seen in Figure A.2.

Commuting Time. A key variable in this paper is how much time students travel when commuting to college. I first calculate the average commuting time from a student's household to any college in the city, with and without the new systems in place. I then use the road

[13]Students self-declare their home address at age 18, when they obtain their national ID. This is typically validated with utility bills by the National Identification Agency in Peru (RENIEC).

[14]I do not include information after 2014 since it marks the beginning of a significant higher education reform in Peru. During the following years, the Ministry of Education changed the format in which they collected enrollment data. It also started a licensing process in 2016 that denied operational licenses to one-third of colleges in the country for failing to meet basic quality standards (Alba-Vivar et al., 2023). This could have changed the way students make their college decisions.

[15]Specifically, I use the *manzana* level, which is a unit bigger than a block but smaller than a ZIP code.

network data from OpenStreetMap API,[16] which includes information on road type (highway, motorway, etc.). Then, I calculate the optimal route, defined as the shortest possible route from households to each college in the city. I follow Velásquez (2023)'s procedure and data to impute velocities for major highways and the new lines, and then complement it with the Google Maps API data to obtain primary and secondary highway speeds. With this information, I compute commute times with and without the new lines.[17]

Labor Market Returns. I use labor market outcomes compiled by the Ministry of Labor (Planilla Electronica). This panel dataset spans from January 2014 to November 2019, and it includes monthly labor market outcomes such as wages and hours work.[18] It also includes college information such as major, gender, and college. I restrict my sample to students who graduated in 2014 and 2015, and collapse the data at the college $\times$ major $\times$ gender cell.

3.1 Descriptive Statistics

Table 1 shows the summary statistics for the main sample using information before the stations' openings. Panel A shows the college enrollment rates using two measures: (1) the logarithm of enrollment counts and (2) the rate of students of a year-cohort that enroll in any college at the block level. In my sample, fewer than one student per block enrolls in college. By accounting for the population of the same age, it becomes apparent that an average of 17% of students under 19 years old enroll in college. The college enrollment rates are defined as the following: $Rate_{it} = \frac{TotalEnroll_{it}^{16-19}}{TotalPop_{it}^{16-19}}$. Furthermore, the table indicates that, on average, women enroll in college at higher rates than men and that private college enrollment is higher than public college enrollment. Importantly, the average distance from students' homes to college is quite similar for both neighborhoods, whether they are connected to new lines or to the planned-but-not-executed lines.

[16] These data are publicly available; I use the package *osmnx* available on Python.

[17] Given computational restrictions, I compute this travel time for a random subset of households across the city that are representative at the district level, and I use it for the entire sample.

[18] For more details about this dataset, see Alba-Vivar et al. (2023).

Panel B shows average statistics using the 2007 census, which includes the total population and education levels obtained for individuals over 25 years old. There is no significant difference in terms of the population size for affected versus non-affected neighborhoods. However, the population over 25 seems to be slightly more educated in the treatment group [19].

4 Empirical Strategy

In this section, I lay out the empirical strategy which leverages a difference-in-differences methodology that exploits neighborhood exposure to new lines and variation across student cohorts. I also explain in detail the estimation for both short-term and medium-term outcomes. Subsequently, I outline the empirical challenges associated with quantifying the impact of new transit lines on students' college access and choices. Finally, I explain the methodology I use to identify students who were very likely to attend college regardless of the policy as well as their anticipated likelihood of enrolling in a specific college type.

4.1 Analysis of Short-Term Outcomes

The empirical strategy I use is a difference-in-differences approach and a flexible event study framework, to exploit neighborhood exposure to new lines and variation across student cohorts. One concern comes from selection and establishing a proper control group of never-treated neighborhoods. In this sense, simply comparing connected neighborhoods to non-connected ones within the city might overestimate my results since the allocation of the new routes is not completely random. For example, poor individuals in the city live in remote areas where the implementation of a new line is unlikely. One way to address this concern is using a placebo group as in Donaldson (2018). In this paper, the control group comes from the neighborhoods that could have been affected by the new transportation system due

[19] For this reason, I control for the share of the population with higher education in main estimates.

to planned-but-not-executed lines, as explained in Section 2.2. Figure 3 shows the neighborhoods in the treatment and never-treated groups. I define a neighborhood as exposed to the executed lines if it is within 1.5 kilometers (about a 20-minute walk) from the nearest station, as seen in Figure 3. The never-treated group is defined as neighborhoods within 1.5 kilometers of planned-but-not-executed stations. I also exclude neighborhoods that are simultaneously exposed to both opened and planned-but-not-executed lines, as indicated by the yellow shaded areas in 3. By restricting the never-treated neighborhoods connected to planned-but-not-executed lines, I reduce the selection bias that might arise from potential correlations between the placement of new lines and unobserved shifts in access to college. The event study specification is the following:

$$y_{t,i} = \sum_{\tau=-4}^{-1} \alpha_\tau D_i^{pre} \mathbb{1}(\tau = t - T^*) + \sum_{\tau=1}^{4} \phi_\tau D_i^{post} \mathbb{1}(\tau = t - T^*) + X\beta_{t,i} + \psi_t + \mu_i + e_{t,i}, \quad (1)$$

where Y_{it} represents the outcome of interest, such as college enrollment rates, at the block level i in year t. The binary treatment variable, D_i, equals one if the block is connected to a newly executed line and zero if the neighborhood is connected to the planned-but-not-executed line. $\mathbb{1}(\tau = t - T^*)$ consists of event-year dummies that represent the four years before and after the new routes opened. The coefficients of interest, ϕ_τ, show how the outcomes evolve over time after new stations are opened, allowing for the possibility of heterogeneous effects. α_τ indicates the pre-treatment effects in eventually treated neighborhoods relative to untreated ones, enabling me to test for the presence of pre-trends. μ_i are the block fixed effects, and ψ_t are the year fixed effects.

A few additional empirical challenges arise when using this strategy. First, the staggered nature of the treatment might raise some concerns given the potential heterogeneous and dynamic effects. The very first opening was the Metropolitano in 2010, the second opening was half of the Linea 1 in 2011, and the other half was opened in 2014. In this setting, heterogeneous treatment effects are likely to arise from heterogeneity in how the Metropolitano and Metro de Lima connect to different colleges in the city. To address these potential issues, I draw upon recent advancements in the DiD literature,[20] following Borusyak et al. (2023). I implement their imputation estimator, which allows for treatment-effect heterogeneity and dynamic effects. The estimation process begins by employing Ordinary Least Squares (OLS) regression exclusively on the untreated observations, accounting for any fixed effects at both the unit and time period levels. This step helps establish a baseline of the data before introducing any treatment variables. Next, in the second step, the estimator extrapolates the data to impute the potential outcome in the absence of the treatment, denoted as $Y_it(0)$. This allows for a comparison between the observed outcome and the imputed counterfactual, providing a treatment effect estimate for each treated unit. Finally, in the third step, the outcome of interest is estimated by combining these treatment effect estimates. This is achieved through a weighted summation of each estimated effect. I estimate the ATT using the administrative data and Census data at the block level.

4.2 Analysis of Medium-Term Outcomes

For medium-term outcomes, I use individual-level data from the 2017 census and estimate an exposure DiD as in Equation 2. As in the previous section, I also rely on the Borusyak et al. (2023) imputation estimator. Age cohorts are considered treated if their residency block was exposed to the executed lines by the time they were 17 years old, the age at which most high school students graduate in Peru. I also restrict the analysis to those individuals born between 1991 and 2000.[21]

[20] Several papers show that using the two-way fixed effects estimator in a staggered design might yield biased estimates given the presence of both heterogeneous and dynamic effects. See Borusyak et al. (2023); Callaway and Sant'Anna (2021); de Chaisemartin and D'Haultfœuille (2020); Sun and Abraham (2020).

[21] For this analysis, I also exclude the opening of the second half line in 2014 since students affected by this event are not on time to graduate college by 2017.

$$y_{c,i} = \sum_{\tau=-4}^{-1} \alpha_\tau D_i^{pre} \mathbb{1}(\tau = c - T^*) + \sum_{\tau=1}^{4} \phi_\tau D_i^{post} \mathbb{1}(\tau = c - T^*) + X\beta_{c,i} + \psi_c + \mu_i + e_{c,i}. \quad (2)$$

4.3 Identification Challenges

In an ideal experimental scenario, one would randomly allocate a reduction in commuting time for students. Yet, determining the impacts of policies that universally reduce commuting times for all individuals can be notably challenging.

First, even when using the placebo lines to reduce selection bias, another potential issue can arise in this setting. For example, neighborhoods located in the city center (which is both economically and geographically central in Lima) might experience higher enrollment growth since they are more likely to have a new line than those in the outskirts. In this sense, other determinants of the outcome of interest (college enrollment rates) are still not random (recent work by Borusyak and Hull, 2023 highlights this issue). This could happen since families living in these areas are also much more educated, which is a strong predictor of college enrollment (see Appendix Figure A.7).[22] In this paper, I use variation not only in the location of new lines but also in the timing of the opening. Nevertheless, to avoid issues related to neighborhoods located in the geographical center of the city, that were going to receive the treatment regardless of the allocation of newly opened lines exclude those located in downtown Lima.[23]

Another concern arises from having general equilibrium effects, which refer to the broader economic impacts that result from families having improved access to better job opportunities, potentially leading to an increased ability to afford college education. However, these effects are typically observed over an extended period. In my analysis, I focus on short-term effects, specifically within a time frame of up to four years. Additionally, a study by Velásquez

[22] Borusyak and Hull (2023) propose using a recentered treatment as an instrument that removes bias from the non-random shock exposure.

[23] I define downtown Lima as all neighborhoods in the following districts: Lima (historical center), Lince, Jesus Maria, San Isidro, Miraflores, Breña, La Victoria, and Rímac.

(2023) did not find significant effects on factors like rent or household income for the same shock.

Finally, another concern arises from the fact that the new lines might also increase labor market opportunities, as the trade-off between the long-term advantages of education and the more immediate gains from engaging in the labor market might arise Adukia et al. (2020). A student who gets connected to new lines might experience both of these effects simultaneously and this can bias my estimates of a reduction of commuting time to college on college enrollment. Even when I am not able to clearly disentangle both effects, I show evidence that this bias is relatively small in my setting. First, in Peru, students are only legally authorized to work at 18 years old, partially ruling out the option of working immediately in the formal sector after high school graduation (16 or 17 years old in this context). Second, most formal jobs (high-skill) are located in the geographical center of the city while most informal jobs (low-skill) are typically located in the outskirts. However, it could be the case that the long-term advantages of education become more salient as students can now access better formal jobs that will hire high-skill workers. In Section 5.2.1, I show how the effects vary on how close students are to where most high-skill jobs are located, suggesting most of the effects are actually coming from students who already had access to these formal jobs and not those who might have experienced an increase in future labor market opportunities.

4.4 Identifying College Access and Choice

I expect that a reduction in commuting time to college will both impact college enrollment and, as some colleges become more accessible, change college choices. In an ideal setting, I would have information about students' abilities and preferences for college, which is crucial information that could help identify those who will likely to go to college and which college they will attend. For example, this information is typically available in places with centralized admissions systems, where students often take admission exams (ability proxy) and report their college preferences. However, as previously mentioned, Peru has a decentralized admission system, and there is no standardized testing, meaning there is no measure of a student's ability when applying to college.

To address this limitation, I use machine learning as a second-best tool to recover students' choices. I first measure whether a student is likely to go to college using rich data on neighborhood characteristics at the block level. Then, I estimate the probability of attending a certain type of college. Using the k-nearest neighbors algorithm, I predict the probability of attending college for each neighborhood based on the pre-treatment data, before the opening of the new transit lines. I further refine the algorithm using a k-fold validation strategy and choose an optimal k parameter using GridSearch.[24] I predict whether students are likely to enroll in a high-quality private college, a low-quality private college, or a public college at the individual level.[25] High-quality private colleges are those whose graduates in 2014 earned more than 2,200 PEN (25% top of the wage distribution), or approximately twice the minimum monthly wage, while low-quality private colleges are those whose graduates earned less than 1,450 PEN (25% bottom of the wage distribution). All other private universities are considered medium-quality private colleges. Appendix Figure A.5 shows the distribution of college wage returns and the thresholds used.

5 Impact of Transportation Upgrade on College Education

In this section, I examine the impact of the new lines on commuting time to college, college enrollment, and college choice. I also provide evidence of the mechanisms behind these results.

[24]See Appendix D for details.

[25]I also predict if a student is likely to enroll in an elite college, a licensed college, or a non-licensed college and if a student is likely to study a STEM major. Licensed colleges are ones that obtained an operational license between 2016 and 2021. The Ministry of Education closed one-third of colleges that did not comply with basic quality standards (Alba-Vivar et al., 2023). Most of these these colleges are located in downtown Lima and are connected to the new stations.

5.1 Effects on Commuting Time to College

In this section, I investigate whether the new stations reduce the average commuting time from students' households to any college in the city when a student's neighborhood is connected. To do this, I use a simple 2-by-2 DiD model that leverages the opening of new stations and the treatment status of being connected to either the executed or planned-but-not-executed lines at the neighborhood level.[26] This estimate can be considered a lower bound and very conservative since 1) informal routes (such as combis) are not included in the data and 2) I assume students are commuting by car, which is an overestimate of their actual transportation time.

The average commuting time to any college in these neighborhoods, before the new systems are introduced, is estimated to be around one hour. This aligns with the self-reported data from the 2010 National University Census, as illustrated in Appendix Figure A.1, which shows a similar number.

The results in Table 2 suggest that the new system reduced the average commuting time to any city college by 17% (a 13-minute reduction from a baseline commute of 66 minutes per trip), which translates to almost 30 minutes per day saved on commuting. The results for private colleges are even higher, a 20% reduction in commuting time in contrast with a 14% reduction in commuting time for public colleges relative to the baseline. Note that these findings are based on the most conservative estimate, and the actual impact on commuting time could be even higher.

5.2 Effects on College Enrollment

Using the enrollment administrative data at the block level, I estimate the post-treatment results from Equation 1 for up to four years after the new lines opened. I find positive effects on college enrollment rates at the block level (Table 3, Column (1)). The coefficient indicates

[26] Appendix Figure A.6 visualizes the before-and-after variation. Notably, people living in the northeast of the city seem to have the most benefits of the new transportation system.

a 1 percentage point (pp) increase from a baseline enrollment of 14% (pre-treatment),[27] measured as the number of students enrolled in a block divided by the number of people of the same age living in that block. Column (2) shows higher effects when enrollment is measured in logarithms. Nevertheless, the magnitude of the effects is significantly smaller compared to policies that aimed to directly increase college enrollment rates. For instance, initiatives like providing information (which resulted in an 8 percentage point increase in FAFSA applications, as demonstrated by Bettinger et al. (2012)), early commitment to free tuition (leading to a 15 percentage point increase for high-achieving low-income students in a flagship university, as studied by Dynarski et al. (2021)), or offering scholarships (resulting in a 10 percentage point increase in Colombia's Ser Pilo Paga program, as outlined by Londoño-Vélez et al. (2023)) have shown more substantial effects.

I also explore the effects by type of institution since in this setting, public and private colleges have different admission systems. Column (3) shows the impact on private enrollment, which is almost entirely driving the overall effects, while Column (4) shows the impact on public enrollment, which is significantly lower. This aligns with the fact that public colleges in Peru tend to be more competitive than private colleges. Consequently, reduced commuting time may have a limited impact on enrollment, given the role of students' abilities in admission and decision-making. In contrast, private universities, with their more streamlined admission and ability to quickly adapt, can accommodate an influx of new students more readily. Columns (5) and (6) show no significant difference in college enrollment for women and men, with both showing positive and significant results.

When examining the dynamics of the effects using the event study specification, I observe that college enrollment rates increase since the first year after station openings, and the magnitude of the effects doubles up to a 2.5 percentage point increase after three years (see Figure 4a). Additionally, the pre-treatment coefficients validate my findings as they show no prior trends before the implementation of the new lines.

[27] The sample for this estimate is restricted to recent high school graduates or individuals up to 19 years old.

Examining the dynamic effects within sub-groups reveals that most of the effects are driven by enrollment in private colleges, whose enrollment keeps increasing over time, as seen in Figure 5a. The effects for public college are also positive but significantly smaller in magnitudes. When exploring the differential enrollment rates by gender, I find that the positive impacts on women intensify more rapidly over time in contrast to those on men, as illustrated in Figure 5b. This indicates a lag for women in fully capitalizing the benefits of the new system. It also underscores that the advantages of the new systems extend beyond merely reducing transportation costs but also improve travel safety, which is especially crucial for women in this city.

5.2.1 Robustness Checks

Alternative Specifications and Estimators. Figure 4a presents the event study results using the Borusyak and Hull (2023) estimator and shows no significant trends in the pre-treatment period. I test another functional form, using the hyperbolic sine transformation (Figure 4b), and the results are similar. Furthermore, in line with recent advances in the DiD literature, I use an estimator distinct from that of Borusyak and Hull (2023). Using the Callaway and Sant'Anna (2021) estimator, as seen in Appendix Figure A.8, I find similar results.

Sample Including Downtown Lima. In Section 4, I highlighted that the main specification excludes districts in downtown Lima since they were very likely to get treated by any potential transportation line. This part of the city also has a higher level of market access compared to the outskirts. Similar to the results in Table 3, adding these districts does not significantly change the results, as seen in Appendix Table A.3. I still find positive effects on college enrollment rates at the block level (1 percentage point increase). The results from Columns (2)–(6) reveal a similar pattern to the main results in Table 3 but with a slightly lower magnitude, suggesting that the neighborhoods in downtown Lima experience almost no impact.

Distance to Stations and Distance to High-Skill Jobs. It is expected that the effects are higher for students living close to the new stations. Appendix Figure A.12 confirms that this is the case. Additionally, in Section 4, I highlighted a potential confounder could be that students are motivated to go to college because they might access high-skill jobs. In this sense, I would expect higher effects of students living the furthest from the location of high-skill jobs. Appendix Figure A.13 shows that this is not the case, and that in fact, most of the results are coming from students who already lived closer to such jobs.

5.3 Effects on College Choice

I expect that the effects of the new lines are on both the extensive margin (college enrollment) and the intensive margin (college choice). This is mainly motivated by earlier findings suggesting that students are more likely to enroll in private institutions rather than public ones, coupled with the additional variation stemming from which college becomes connected to the new lines. Appendix Figure A.9a shows these patterns over time, which suggest that colleges connected to new stations enjoy a significant increase in their enrollment rates, while the rest of the colleges suffer a decrease, implying that students might be trading off between them and changing their choices. What is more, Appendix Figure A.9b shows that connected colleges might have also increase the number of major they offer. But the increase in enrollment is also different by gender, as women are drawn to connected colleges at higher rates than men (see Appendix Figure A.10).

As mentioned before, in an ideal setting, I would know students' ability and college preferences.[28] However, in this context, as with many education systems worldwide, this information is unknown give the lack of standardized testing or centralized admission systems.

To address this limitation, I leverage the big data I have collected and use machine learning techniques (described in Section 4.4 and Appendix D in more detail) to predict students' college choices, using a non-parametric algorithm, k-nearest neighbors (KNN). Using this

[28]For example, a setting where students reveal their preferences and their ability is known is in a centralized admission system.

tool enables me to exploit non-linear relationships, unlike other similar methods. I feed this algorithm with all the data available before the policy takes place (both neighborhood and students characteristics), and then I obtain the predicted probability θ_i^m that a student i will go to college (any college) and college type m, where $m =$ [High Quality Private, Medium Quality Private, Low Quality Private, Public] for the whole sample. Then, I use θ_i^m to help me identify students who are more or less likely to enroll in a college type m.

A close example of the use of machine learning to identify groups can be found in (Black et al., 2023), where a random forest method was employed to distinguish between students who met and did not meet the eligibility criteria for the Texas Top 10 policy given data limitations that prevented them to directly identify eligibility. In my case, I use the predicted probability of attending college to identify results for two samples: students who were unlikely to enroll in college (*new student sample*) or by students who were very likely to attend college regardless of the policy (*typical student sample*) but are now switching their choices. I also use it to explore whether students who are likely to attend a college type m are driving the effects on the probability of enrolling in such colleges.

First, I estimate the results on the likelihood of enrolling in college type m using the enrollment data at the individual level. Table 4 shows the results for different types of colleges. Column (1) shows that students connected to the new stations are less likely to enroll in a high-quality college. I find an statistically significant 11% reduction in the probability of enrolling in high-quality colleges relative to the baseline. Students are also less likely to enroll in a medium-quality private college by 10% (Column 2). In contrast, students become more likely to enroll in a low-quality college by 2.6 percentage points (16% increase from the baseline) as seen in Column (3). When examining the likelihood of enrolling in a public college, I also find positive results. Column (4) shows an increase of 3 percentage points in the likelihood of enrolling in public colleges (7% increase from the baseline).

Second, I investigate the extent to which the previous results are driven by the new student sample or the typical student sample. Figure 6 shows this effects. In the case of the typical sample, there exists a preference for substituting medium-quality private college options with their public counterparts. Conversely, among new students, there is trade-off between both high and medium-quality private alternatives for lower-quality private and public institutions.[29]

Third, I also explore the effects on a student's likelihood of enrolling in a college based on their probability, θ, of attending a specific type m college. I assess the heterogeneous impacts by quintiles of θ, where θ^{q1} represents the lowest quintile and θ^{q5} the highest for enrolling in school m. Figure 7a panel (a) shows the results for high-quality private colleges by quintile and also broken down by either the new or typical student sample. In the typical student sample, no noticeable effects are observed, even among those who are very likely to go to a high-quality college. In the new students sample, only those who were unlikely to enroll are showing negative effects. In Panel (b), only those who were very likely to enroll in a medium-quality private college are less likely to enroll in one, both in the typical and new students sample. Conversely, Figure 7c Panel (c) shows the breakdown for the probability of enrolling in a low-quality private college. While there are no effects for the typical students, it is interesting that all types of new students (both with high and low probability of enrolling) are attracted to these colleges. Turning to public institutions as in Panel (d), the figure shows no significant differences between the typical and new student sample. The increase in the likelihood of enrolling comes from both groups. This suggests that not only are students who were already likely to attend a public college incentivized, but even some who were initially less inclined opt for a public college.

5.3.1 Heterogeneous Effects by Gender and Socioeconomic Status

I next explore whether these effects are heterogeneous for different populations. Figures 8 and 9 illustrate these diverse impacts for different types of colleges by gender and socioeconomic status. The results suggests that male and female students respond differently to the same

[29] Appendix Table A.1 shows the results in a Table format

shock of reduced transportation costs. However, the results are also different for the typical and new student sample. Figure 8a reveals that women in the typical student sample show no changes in terms of their college of choice. However, women in the new student sample show positive and significant effects in their likelihood to attend a low-quality college and a negative and significant effect on their likelihood to attend public or a high and medium-quality private college. In contrast, men in the typical student sample are more likely to attend public college and less likely to attend low and medium-quality colleges. Those in the new student sample also show a higher likelihood of enrolling in public college and are less likely to enroll in high or medium-quality colleges. These divergent reactions carry significant implications for the gender gap, as male students appear to be gravitating toward more lucrative educational pathways than women.

The results by socioeconomic status—where poor students live in neighborhoods with household incomes below the median and non-poor are above the median—show that income does not seem to explain college choice among students in the typical student sample, as seen in Figure 9a. In contrast, the results for those in the new student sample suggest that poor students are less likely to enroll in high-quality colleges and more likely to gravitate toward public schools. The increase in the likelihood of enrolling in low-quality colleges is similar for both poor and non-poor students.

5.4 Mechanisms

I first explore whether the students who enroll in college are now opting for colleges that are located farther away in terms of distance. The rationale behind this is that since transportation becomes less of an issue, students have higher incentives to travel further and select better colleges that are more distant from their homes. Table 5, Column (1) evaluates this hypothesis and shows the impact of the new routes on the distance between home and college. Contrary to what is expected, I find that students connected to new lines are more likely to enroll in colleges that are 9% closer to them, suggesting they are not taking advantage of traveling longer distances within the city.

I also explore whether students enroll in a college that is connected to the new lines. Table 5, Column (2) shows that students connected to new lines are 10% more likely to enroll in a college that is connected to the new lines as well. However, when breaking these results down by whether the college is public or private, I find that students are 14% more likely to enroll in private colleges that become connected to new lines (and 16% more likely to enroll in a low-quality private college that is connected), while I find no significant effects for enrolling in public colleges that become similarly connected. These results are consistent with previous results suggesting that most of the effects on college enrollment rates are coming from enrollment in private colleges.

Additionally, I explore the heterogeneous effects by sub-groups for these estimates. Appendix Figure A.11 shows that all types of students are less likely to enroll in colleges that are further away from their homes. The effects are significantly driven by students living in poor neighborhoods. Panel B shows that all types of students except those living in poor neighborhoods are more likely to enroll in colleges connected to the new lines.

However, when looking at the results by whether the college is public or private, women are more likely to enroll in private colleges that get connected, while men are more likely to enroll in public ones. These results confirm that the effects on college choice are substantially different by gender. On the one hand, women choose private colleges with improved commute (both their neighborhood and chosen college get connected, significantly reducing commuting times), while men are more inclined to enroll in public colleges with similar connectivity benefits. Note that one of the few public colleges connected to the new lines is the National University of Engineering, where 80% of enrolled students are male.

An additional mechanism is major choice, particularly if a student opts for a STEM field. Appendix Figure A.14 illustrates that for both, typical and new students, women become less likely to enroll in STEM while men become more likely. The implications for STEM majors are significant, as they tend to lead to the highest post-graduation earnings. In this context, public schools exhibit greater selectivity in admitting students into STEM programs, making them relatively more competitive and potentially riskier choices. It is worth noting that since connected public schools are predominantly STEM-focused, creating a male-dominated environment, they may primarily attract male students. Conversely, private institutions have the flexibility to expand non-STEM enrollment, become relatively less selective, and consequently, have the potential to attract more female students.

5.4.1 Robustness Checks

I use an alternative specification to estimate the effects on college choice. Instead of estimating the likelihood to enroll in college type m, I use it as an outcome of a dummy variable equal to 1 if a student is likely to enroll in a college type m relative to what is predicted at baseline (probability $\theta_m > 0.5$). The effects in Appendix Table A.4 show a similar pattern to the main results.

I also test alternative definitions of college quality, such as whether a college is deemed elite or is part of an elite college consortium,[30] or if it received an operational license after the 2014 Higher Education Reform. The results consistently show that students are less likely to enroll in these institutions. Appendix Table A.2 shows these results, by testing the probability of enrolling in an elite or a licensed university (Columns 1 and 3). In the same table, I show if students changed their choices relative to the probability of attending these colleges (Columns 2 and 4), and results suggest that students who were likely to attend an elite or licensed college are less likely to enroll in one.

[30] I define elite colleges in Peru as those that are affiliated with the Consorcio de Universidades.

6 Trade-Off Between College Quality and Travel Time to College

The reduced-form results for college choice suggest discernible gender-specific responses. Specifically, reduced travel time to college leads to a higher probability of women opting for low-quality private institutions, while men tend to lean toward opting for public colleges in response to the same change. Understanding the valuation of commuting time by gender holds significant importance for policymakers since these differences can shape the impact of any public transportation upgrade on students' choices of educational institutions. This in turn can influence the distribution of students in different colleges, affecting educational outcomes and labor market opportunities in the future.

To understand such differences, I estimate key parameters such as a student's willingness to pay for commuting time in terms of college returns (or the wage premium of attending school c). I outline a simple model of college choice to recover students' preferences,[31] using a random utility model, which recognizes that individuals have different preferences and utilities for different options. This allows for a more realistic representation of decision-making in situations where students have diverse tastes and priorities. In this model, each student i maximizes an indirect utility function denoted as

$$\begin{aligned} U_{ic} = &\quad \beta_i V_{ic} + \epsilon_{ic} \\ = &\quad \beta_i^w W_{ic} + \beta_i^t T_{ic} + \delta_c + \epsilon_{ic}. \end{aligned} \tag{3}$$

Each student i chooses over a choice set of mutually exclusive colleges available in the city, $C_i = C_{i1}, C_{i2}, ...C_{iN}$. V_{ic} captures the part of the utility that varies with the student's observed characteristics, and ϵ_{ic} captures part of the utility explained by unobserved variables. V_{ic} can take a linear combination of W_{ic} wage premiums of graduating from college c; T_{ic},

[31] I follow Borker (2020), who implements a similar model, although she incorporates more sophistication by including a safety component and her estimation is at the college route level.

which represents travel time to college c; and δ_c college fixed effects. A student i will choose a college c that maximizes their utility over the set of colleges available. The probability that a student i chooses college c, as in follows:

$$
\begin{aligned}
P_{ic} = \quad & Pr(U_{ij} > U_{ik}) \quad \forall j \neq k \\
& Pr(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}) \quad \forall j \neq k \\
& Pr(\epsilon_{ik} - \epsilon_{ij} > V_{ij} - V_{ik}) \quad \forall j \neq k.
\end{aligned}
\tag{4}
$$

I am interested in measuring the trade-off between college quality and travel time to college. Similar to Borker (2020), as a proxy for college quality I use the marginal rate of substitution to measure how much travel time to college (measured in minutes) a student is willing to give up for an additional unit of college wage returns. This relationship is represented with the following equation:

$$
MRS_i^{WT} = \frac{\Delta T_{ic}}{\Delta W_{ic}} = \frac{\beta_i^w}{\beta_i^t}. \tag{5}
$$

To obtain β_i^w and β_i^t, I estimate the model using a mixed logit framework with random coefficients. I assume that ϵ_{ic} is i.i.d extreme value and also assume normally distributed coefficients for both T_{ic} and W_{ic}. This mixed logit model is highly flexible and can approximate any random utility model (McFadden and Train, 2000), as it relaxes the independence of irrelevant alternatives property. In that sense, this model overcomes the limitations of standard logit models by allowing for random taste variation and unrestricted substitution patterns (Train, 2009). In particular, having random taste variation is useful since I expect heterogeneity in students' observables. The mixed logit probabilities are integrals of standard logit probabilities over a density of parameters, as seen in the following equation, where f is the mixing distribution:

$$P_{ic}(\delta) = \int \frac{exp(V_{ic}\beta_i)}{\sum_{c=1}^{N} exp(V_{ic}\beta_i)} f(\beta_i|\delta)d\beta. \tag{6}$$

Given my interest in assessing distinct responses from both female and male students as suggested by the reduced-form results, I proceed to estimate this model separately by gender. Additionally, I conduct separate estimations based on the likelihood that students go to either private or public colleges since these two types of colleges attract different types of students. Public colleges in Peru typically impose more stringent admission criteria, attracting high-ability students who cannot afford private universities. In contrast, selection into private universities is mostly explained by the household's ability to cover the tuition fees rather than student ability. To help with interpretation, I standardize college returns (wage premium) to have a mean of zero and a standard deviation of one. I measure travel time from home to college in minutes.

6.1 Results

Table 6 shows the results of estimating the mixed logit model. Column (1) shows the results for female students enrolling in private colleges. As expected, they are less likely to choose a college that demands a longer commute and are more likely to choose one that provides higher wage returns. Column (2) shows the results for women enrolling in public colleges, and similar to Column (1), they dislike long commutes to college. However, in this case, they are less likely to choose public colleges that provide higher returns. This not surprising for this context, where the National University of Engineering provides the highest wages in the market (19% female enrollment in 2019), while the National University of Education, a majority-female college, provides some of the lowest wages. The results for male students are similar for both public and private colleges: they dislike long commutes, especially public colleges, and they are more likely to enroll in colleges that provide higher returns, especially private ones.

To evaluate the trade-off between college quality measured by wage returns and commuting time to college, I compute Equation 5 after calculating individual-level parameters corresponding to the coefficients of T and W, using the method proposed by Revelt and Train. (2000). In general, I find that students are less willing to travel to attend public colleges than private ones, and the gender differences are even more striking. On the one hand, Figure 10a illustrates the marginal rate of substitution for students who are likely to attend private colleges, categorized by gender. As anticipated, both male and female students are willing to travel differently for a one standard deviation increase in returns. Specifically, men are willing to travel 55% more minutes per one standard deviation increase in returns compared to women. It is then expected that reduced commuting time will yield differential results by gender, as seen in Section 5.3, suggesting that when new lines open, men are less likely to attend low-quality colleges compared to women. In this case, since most low-quality colleges become connected, they become relatively more attractive for women.

On the other hand, Figure 10b shows that the Marginal Rate of Substitution (MRS) for public colleges is lower for both male and female students compared to those likely to attend private colleges. The difference by gender is statistically significant but in a lower magnitude. Male students show a 7% higher willingness to travel compared to women, measured in minutes per one standard deviation in returns. With this, it is expected that reduced transportation costs have similar impacts for men and women. However, the results in Section 5.3 show that only men take advantage of the reduced costs. Is it important to highlight that among the public colleges that become connected is the National University of Engineering, which mostly attracts male students, while others like the National University of Education do not become connected and attract more female students. In this sense, since public colleges are more specialized, they are also only attracting students who are likely to enroll in such majors. The STEM differences emerge quite strongly as women tend to stay away from such majors.

7 Medium- and Long-Term Effects

7.1 College Completion

When assessing the medium-term impacts, the new transportation lines are anticipated to enhance college access, improve students' overall college experience, and increase their likelihood of graduating on time. These effects can materialize through two key channels: (i) reduced commuting time can positively impact academic performance (as documented in Tigre et al., 2017), and (ii) the new lines can increase access to internship opportunities, which is a crucial requirement for graduation in several programs.

Given the lack of data on each channel, I estimate the overall reduced-form effect using the 2017 census (seven years after the stations opened) and a DiD model that exploits cohort-exposure variation. The results show a positive impact on college completion rates (12%) compared to the baseline rates, as shown in Table 7, Column (1). The estimated coefficients are similar to whether downtown Lima is included or not (Column (2)). The dynamic effects on the event study are presented in Appendix Figure A.15, which shows that the more time a student is exposed to the new transportation lines, the higher the likelihood they will complete college by 2017.

Leveraging the comprehensive individual-level data available in the census, I can delve into the varied impacts across different groups. Figure 11 displays the heterogeneous effects by sub-groups. Women and low-income students enjoy the benefits of the new lines in terms of college completion. The effects are driven by women and students living in neighborhoods where the average income is below the national median. I do not find significant differences between students who self-declare being part of a minority group (indigenous or Afro-Peruvian) and the majority ethnic group (mestizos). Surprisingly, non-poor students show negative effects. These are the students who are enrolling in both public and low-quality colleges. Two concurrent effects could be at play: (i) enrolling in a low-quality might have been disappointing and induced students to drop out or take longer, and (ii) an increase in labor market opportunities might have also encouraged students to work instead of study.

7.2 Employment Rates

Using the same cohort-exposure strategy detailed previously, I explore how the new stations affect employment rates captured in the 2017 Peruvian Census for students who enroll in college. Table 8 shows a 10 percentage point increase in employment rates relative to the baseline. However, when breaking down these effects by job quality measured as blue- or white-collar jobs, I find a statistically significant 17 percentage point increase, or a 6% increase.

Figure 12 shows the heterogeneous effects of the impact on white-collar jobs by subgroups. All populations, except the non-poor, enjoy these benefits. Women are twice as much likely to be working in a white-collar job than men. I do not find significant differences between students who self-declare being part of a minority group and the majority ethnic group. These results suggest that even when female and low-income students are attending low-quality institutions, they are still more likely to graduate and obtain a higher-quality job compared to peers who are not connected to the new lines.

8 Conclusion

This paper studies the relationship between urban features and higher education, making it the first exploration of the causal effects of improved public transportation on both college access and college choice. The focus on college education is particularly relevant in the developing world, where access to K-12 education is becoming universally available and there is an increasing demand for higher education. The results also resonate with large cities in the developed world, where access to college, especially access to high-quality colleges for poor students, is still limited. As nations actively revamp and reimagine their transportation systems, it is important to understand the economic consequences of inefficient, unsafe, or unreliable public transportation. Recent studies in the literature highlight the importance of having the most efficient transportation routes within urban areas. For instance, Kreindler et al. (2023) show that adopting a less concentrated network can improve commuter well-being in cities like Jakarta. However, incorporating access to college education, a crucial driver of an individual's productivity, can be integrated into the planning of optimal transportation networks within cities.

The results of this paper can guide on future research in this matter. I show that access to improved public transit significantly increases college enrollment rates among recent high school graduates living in connected neighborhoods. Specifically, a 17% reduction in commuting time to any city-based college can increase college enrollment rates by 1 percentage point in connected neighborhoods. This increase is mostly driven by private colleges, which in this context include low-cost and low-quality institutions. Students connected to the new lines choose to travel less distances and opt for colleges connected to the new system, even when these colleges are of lesser quality and will affect their future wages.

The results also hold particular significance in the context of large cities where disparities in access by gender are especially pronounced. I show that men and women have different responses in terms of their college choices. Women appear to be more inclined toward choosing lower-quality colleges that are conveniently located along new transportation routes, suggesting a willingness to compromise on the quality of education for the sake of a shorter commute. Using a random utility model, I quantify these trade-offs and show that compared to women, men are willing to travel up to 55% more minutes to attend a college where graduates earn salaries that are one standard deviation higher. These findings suggest that the advantages of attending college may be limited when students opt for institutions that do not offer the best possible career prospects post-graduation. Even though women may be more likely to graduate and secure employment, their choice to attend lower-return colleges can impede their progress in breaking through the existing glass ceiling.

References

Adukia, A., Asher, S., and Novosad, P. (2020). Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. *American Economic Journal: Applied Economics*, 12(1):348–76.

Agarwal, N. and Somaini, P. J. (2019). Revealed preference analysis of school choice models. Working Paper 26568, National Bureau of Economic Research.

Alba-Vivar, F., Flor-Toro, J., and Magnaricotte, M. (2023). College Licensing and Reputation Effects on the Labor Market. *Working Paper*.

Alba-Vivar, F., Flor-Toro, J. L., and Magnaricotte, M. (2020). Los factores que limitan la transición a la educación superior situación actual y recomendaciones de política pública. Technical report, Ministerio de Educación del Peru.

Asahi, K. and Pinto, I. (2022). Transit, academic achievement and equalisation: evidence from a subway expansion. *Journal of Economic Geography*, 22(5):1045–1071.

Balboni, C., Bryan, G., Morten, M., and Siddiqi, B. (2020). Transportation, Gentrification, and Urban Mobility: The Inequality Effects of Place-Based Policies. *Working Paper*.

Balboni, C., Bryan, G., Morten, M., and Siddiqi, B. (2021). Could Gentrification Stop the Poor from Benefiting from Urban Improvements? *AEA Papers and Proceedings*, 111:532–37.

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (2012). The Role of Application Assistance and Information in College Decisions: Results from the H amp;R Block Fafsa Experiment*. *The Quarterly Journal of Economics*, 127(3):1205–1242.

Black, S. E., Denning, J. T., and Rothstein, J. (2023). Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes. *American Economic Journal: Applied Economics*, 15(1):26–67.

Borker, G. (2020). Safety First: Perceived Risk of Street Harassment and Educational Choices of Women. Technical report, The World Bank.

Borusyak, K. and Hull, P. (2023). Non-Random Exposure to Exogenous Shocks. *(Forthcoming, Econometrica).*

Borusyak, K., Jaravel, X., and Spiess, J. (2023). Revisiting Event Study Designs: Robust and Efficient Estimation. *Accepted,Review of Economic Studies*.

Brooks, W. and Donovan, K. (2020). Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua. *Econometrica*, 88(5):1965–1997.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with Multiple Time periods. *Journal of Econometrics*.

Campos, N., Engel, E., Fischer, R. D., and Galetovic, A. (2021). The ways of corruption in infrastructure: Lessons from the odebrecht case. *Journal of Economic Perspectives*, 35(2):171–90.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2017). Mobility report cards: The role of colleges in intergenerational mobility. Working Paper 23618, National Bureau of Economic Research.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–96.

Dominguez Gonzalez, K., Machado, A. L., Bianchi Alves, B., Raffo, V., Guerrero, S., and Portabales, I. (2020). Why does she move? A Study of Women's Mobility in Latin American Cities. Technical report, World Bank. Washington, DC.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and American Economic Growth: A Market Access Approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Dustan, A. and Ngo, D. K. (2018). Commuting to educational opportunity? School choice effects of mass transit expansion in Mexico City. *Economics of Education Review*, 63:116–133.

Dynarski, S., Libassi, C., Michelmore, K., and Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, 111(6):1721–56.

Fiala, N., Hernandez, A. G., Narula, K., and Prakash, N. (2022). Wheels of Change: Transforming Girls' Lives with Bicycles. (IZA DP No. 15076).

Field, E. and Vyborny, K. (2022). Women's Mobility and Labor Supply: Experimental Evidence from Pakistan . *Asian Development Bank working paper*.

Flor-Toro, J. and Magnaricotte, M. (2021). College Expansion and Unequal Access to Education in Peru. *Job Market Paper*.

Herskovic, L. (2020). The Effect of Subway Access on School Choice. *Economics of Education Review*, 78:102021.

Kane, T. J. and Rouse, C. E. (1995). Labor-market returns to two- and four-year college. *The American Economic Review*, 85(3):600–614.

Kohon, J. (2016). Metro de Lima: el caso de la Línea 1. CAF.

Kondylis, F., Legovini, A., Vyborny, K., Zwager, A. M. T., and Cardoso De Andrade, L. (2020). Demand for Safe Space: Avoiding Harassment and Stigma. *Policy Research working paper WPS 9269*.

Kreindler, G., Gaduh, A., Graff, T., Hanna, R., and Olken, B. A. (2023). Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta. Working Paper 31369, National Bureau of Economic Research.

Kreindler, G. E. and Miyauchi, Y. (2021). Measuring Commuting and Economic Activity Inside Cities with Cell Phone Records. *The Review of Economics and Statistics*, pages 1–48.

Londoño-Vélez, J., Rodriguez, C., Sanchez, F., and Álvarez Arango, L. E. (2023). Financial aid and social mobility: Evidence from colombia's ser pilo paga. Working Paper 31737, National Bureau of Economic Research.

Martinez, D., Mitnik, O., Salgado, E., Scholl, L., and Yanez-Pagans, P. (2020). Connecting to Economic Opportunity: the Role of Public Transport in Promoting Women's Employment in Lima. *Journal of Economics, Race, and Politics*, 3:1–23.

McFadden, D. and Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15(5):447–470.

MTC (2005). Plan Maestro de Transporte Urbano para el área Metropolitana de Lima y Callao en la República del Perú. *Agencia de Cooperación Internacional de Japón–JICA*.

Muralidharan, K. and Prakash, N. (2017). Cycling to School: Increasing Secondary School Enrollment for Girls in India. *American Economic Journal: Applied Economics*, 9(3):321–50.

Narrea, O. (2017). ¿mega problemas o megaproyectos? el reto del metro de lima y callao. Nota de política N°3. Escuela de Gestión Pública. Universidad del Pacífico.

Page, L. C. and Scott-Clayton, J. (2016). Improving college access in the United States: Barriers and policy responses. *Economics of Education Review*, 51:4–22.

Revelt, D. and Train., K. (2000). Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. Economics Working Papers E00-274, University of California at Berkeley.

Sallo, K. and Hickman, R. (2021). Implementing a metro project: a political economy perspective from Lima. *In Transport in Human Scale Cities*, Edward Elgar Publishing.

Severen, C. (2023). Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification. *The Review of Economics and Statistics*, 105(5):1073–1091.

Sun, L. and Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Tigre, R., Sampaio, B., and Menezes, T. (2017). The Impact of Commuting Time on Youth's School Performance. *Journal of Regional Science*, 57(1):28–47.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, 2 edition.

Tsivanidis, N. (2022). Evaluating the impact of urban transit infrastructure: evidence from Bogota's TransMilenio.

Velásquez, D. (2023). Transit Infrastructure, Couples' Commuting Choices, and Gender Inequality.

Warnes, P. (2021). Transport infrastructure improvements and spatial sorting: Evidence from buenos aires. *Job Market Paper*.

Zarate, R. D. (2022). Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City. *Policy Research working paper WPS 999 World Bank Group*.

Main Figures and Tables

Figure 1: Decision Tree of College Choice

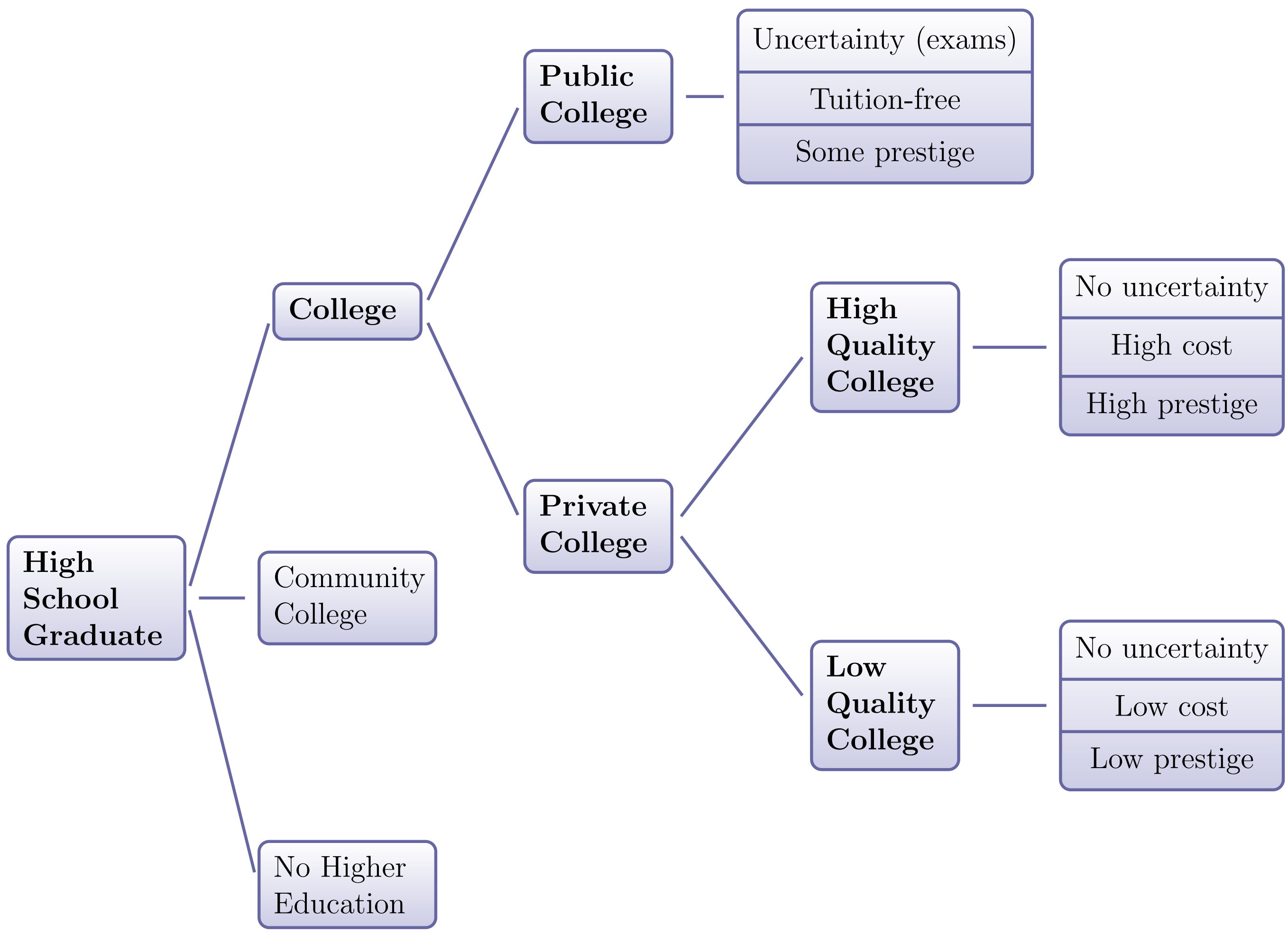
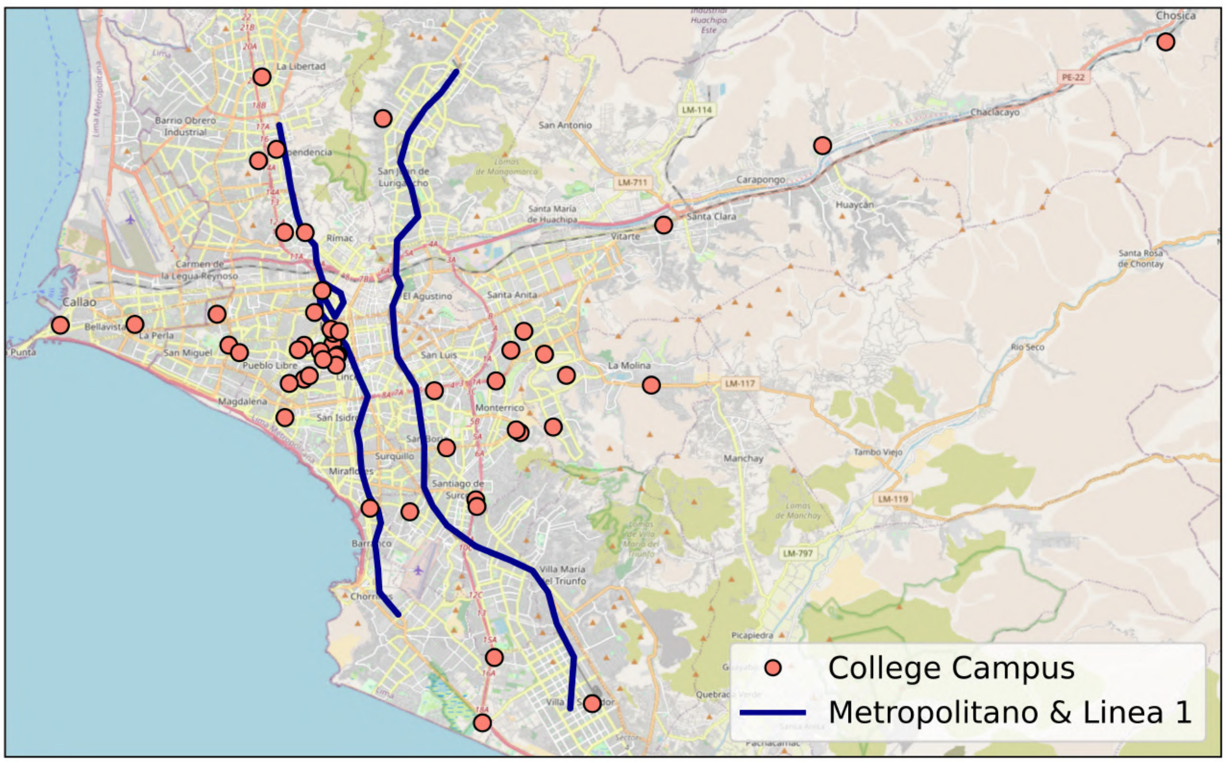
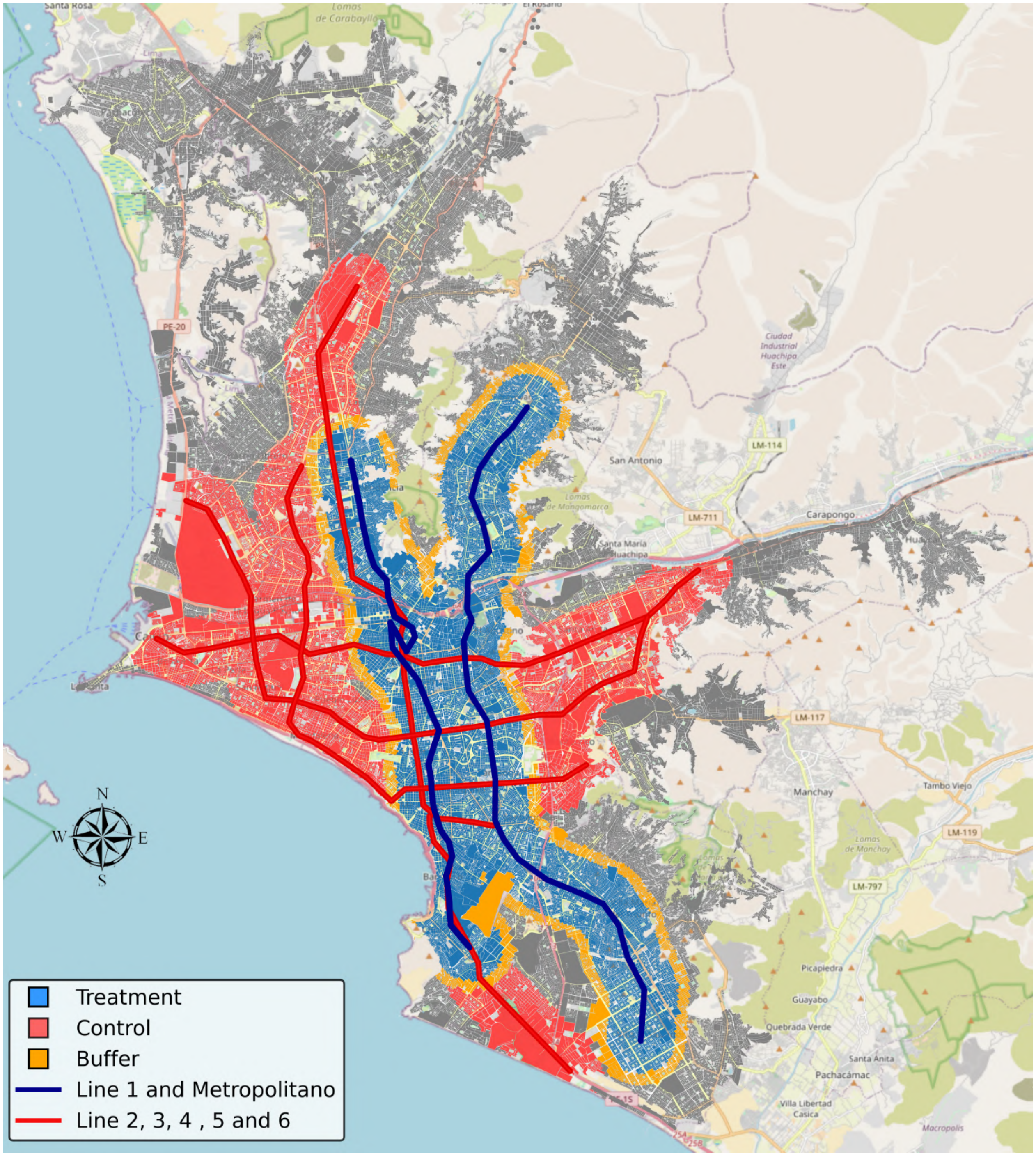


Figure 2: College Location and New Stations Across Lima



Notes. The blue lines show the new routes, to the left is the *Metropolitano*, and to the right is the *Linea No.1*. The pink dots show the location of college campuses across the city.

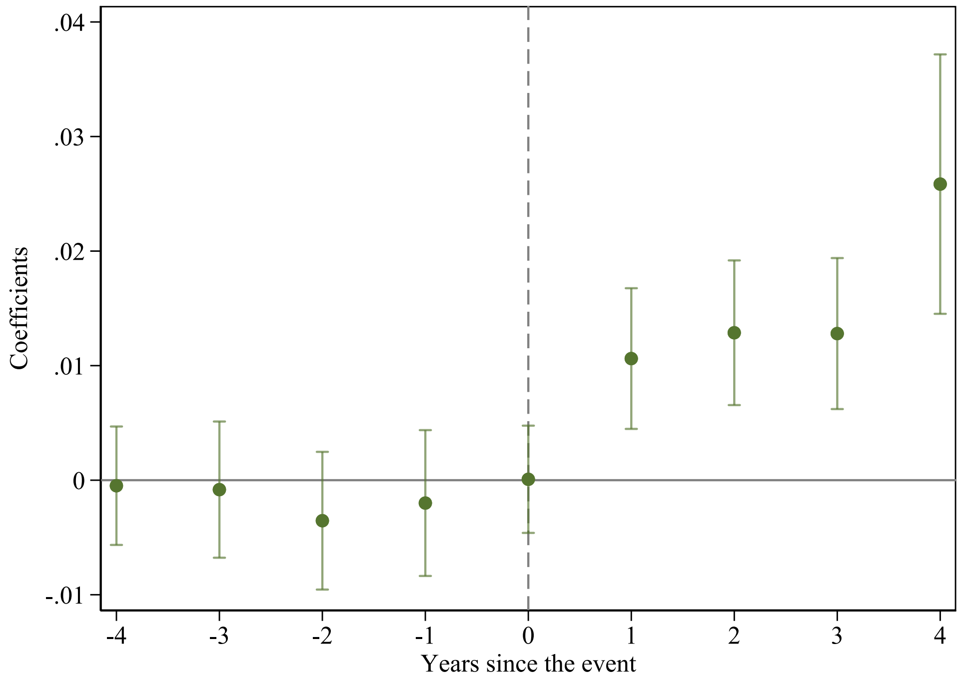
Figure 3: Neighborhoods Exposed to the Executed and Planned but Not-Executed Lines



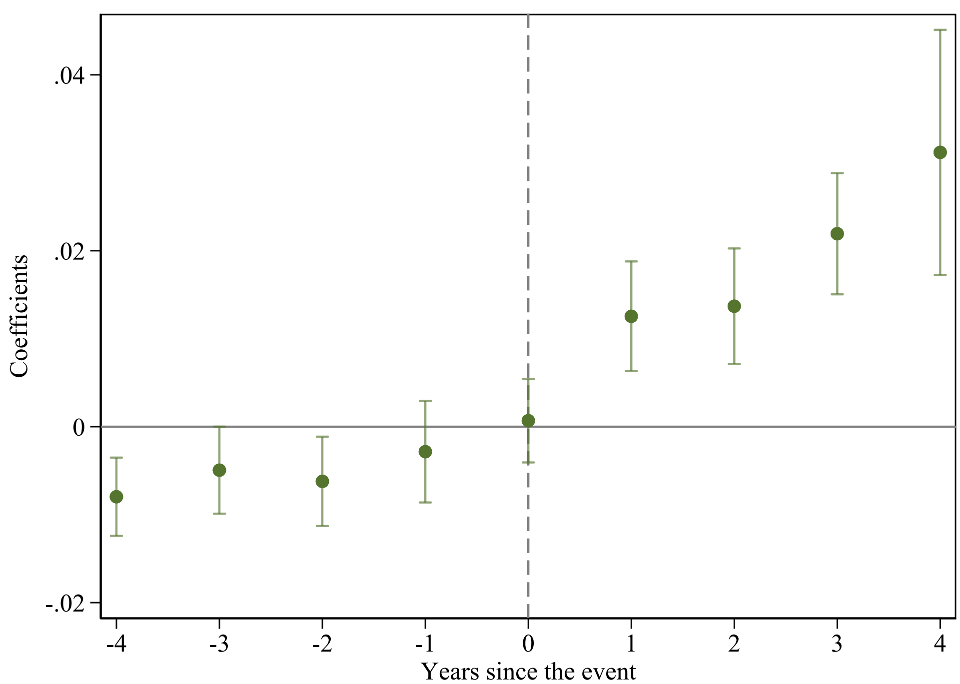
Notes. Blue-shaded areas are neighborhoods within 1.5 km distance from the nearest executed stations while red-shaded areas are neighborhoods within 1.5 km distance from the planned-but-not-executed stations. Yellow-shaded areas are 0.5 km buffer zones that are excluded from the main sample.

Figure 4: Dynamic Effects on College Enrollment Rates

(a) Enrollment Rates



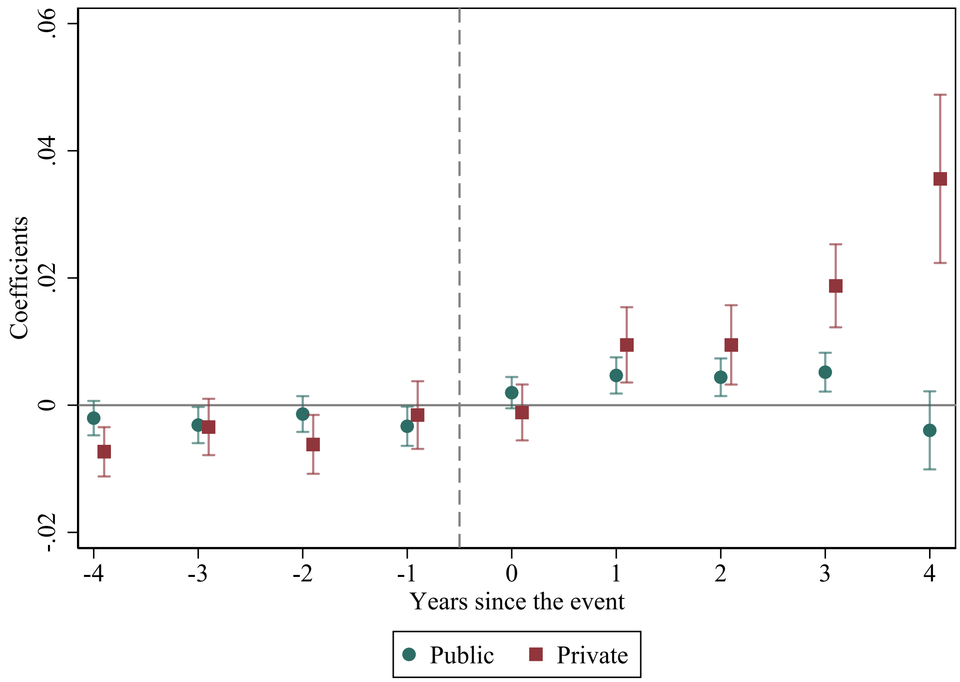
(b) Log(Enrollment Counts)



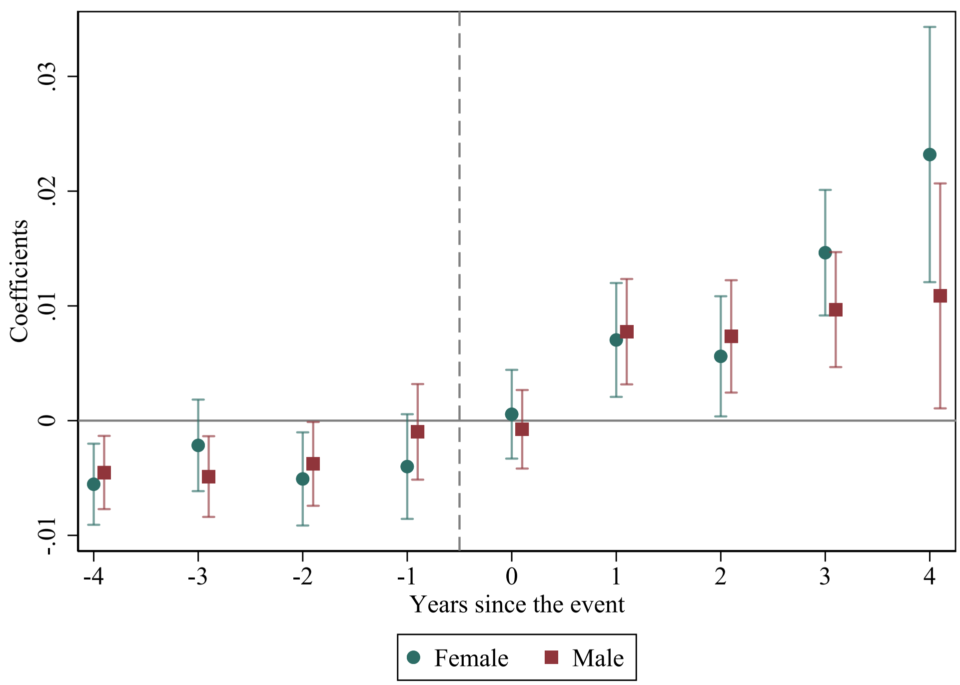
Notes. Regressions include block and cohort-fixed effects. The event study estimates are calculated using the Borusyak et al. (2023) estimator. Panel (a) shows the enrollment rates where the denominator is the total count of students enrolled divided by the total population of (potential) students of the same cohort. Panel (b) shows the logarithm transformation of enrollment counts using the hyperbolic sine.

Figure 5: Dynamic Effects of New Lines on College Enrollment Rates by Groups

(a) By College Type

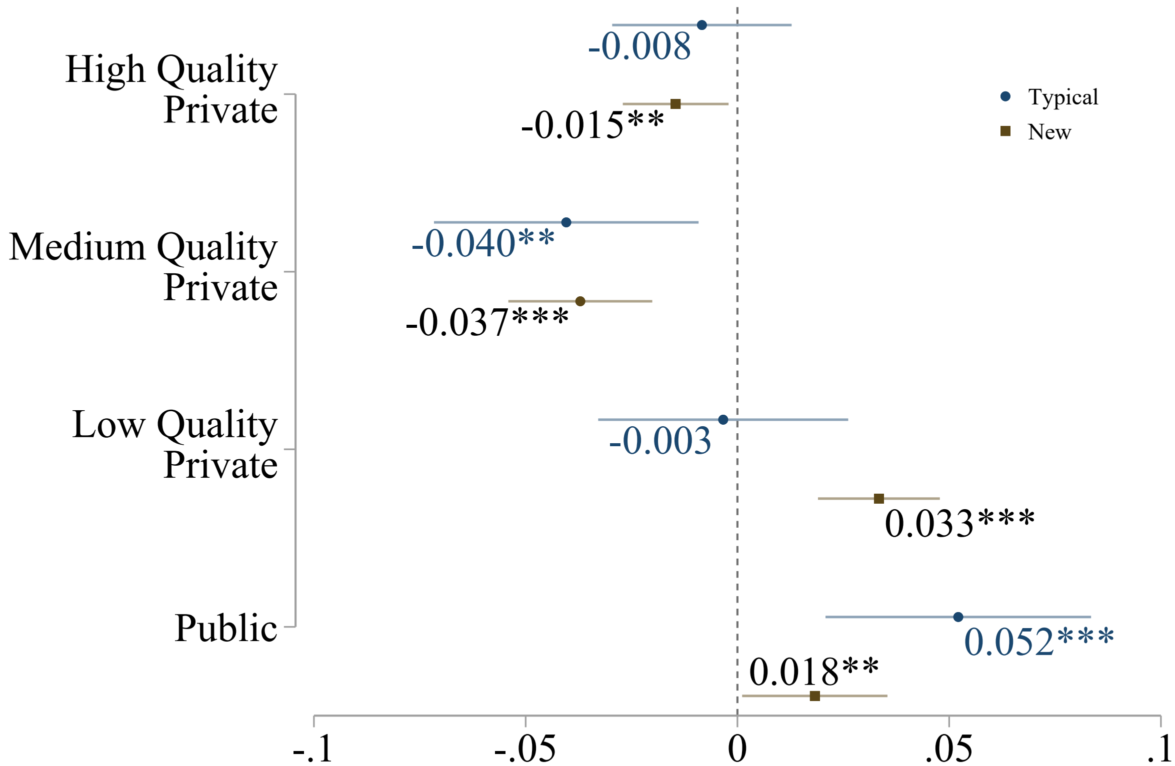


(b) By Gender



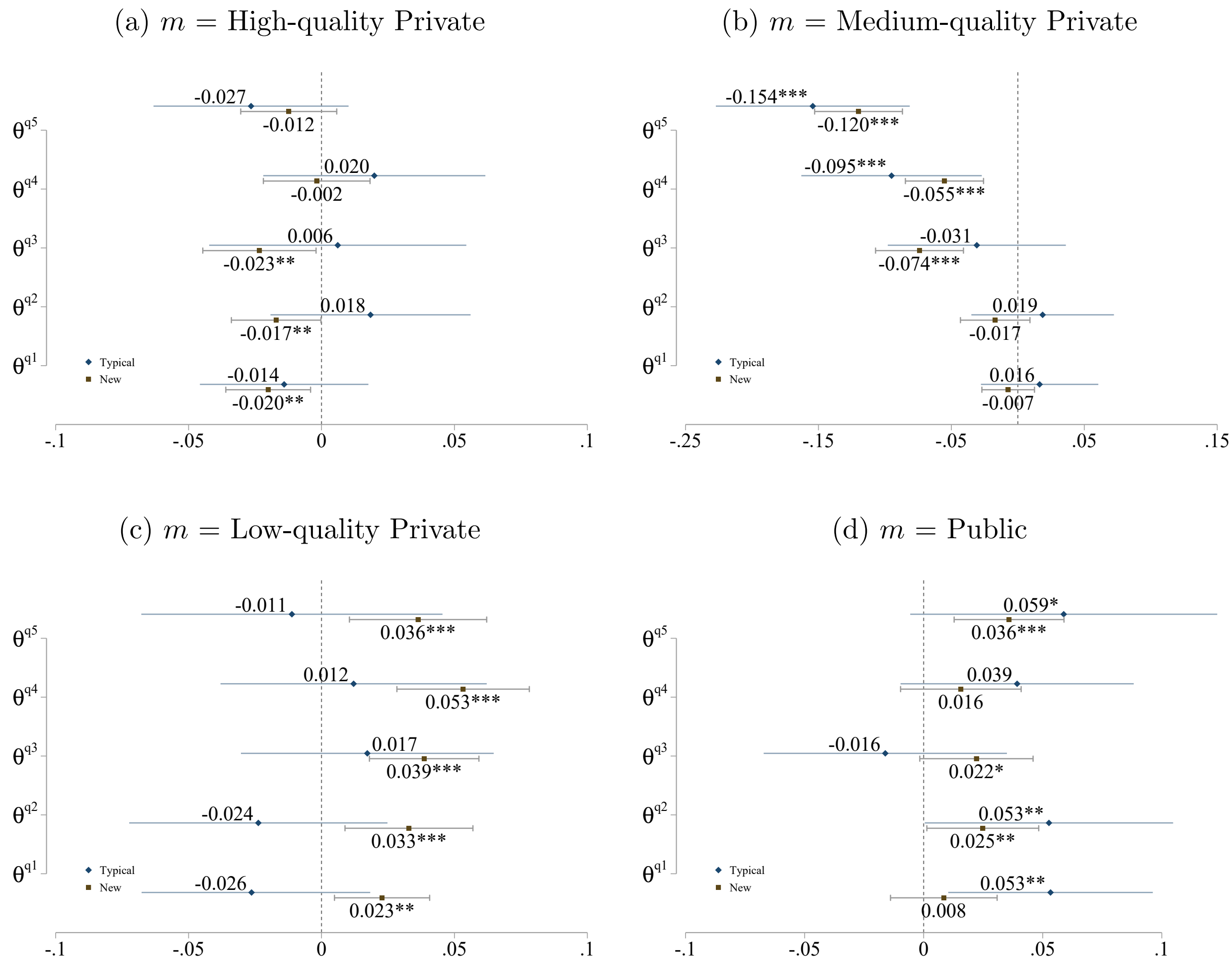
Notes. Regressions include block and cohort-fixed effects. The event study is calculated using the Borusyak et al. (2023) estimator. Panel (a) and Panel (b) shows the logarithm transformation of enrollment counts.

Figure 6: Effects on the Likelihood of Enrolling in College by Access to College



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. High-quality private colleges are all private colleges whose graduates earn more than 2200 PEN while Low-quality private colleges are the ones whose graduates earn less than 1450 PEN using administrative data from wage records in 2014. *Typical* is the sample of students whose baseline probability of attending college was higher relative to the *New* students' sample, whose baseline probability of attending college was low.

Figure 7: Heterogeneous Effects by Predicted Probability θ (quintiles) of Enrolling in College Type m by Access to College



(a) m = High-quality Private

(b) m = Medium-quality Private

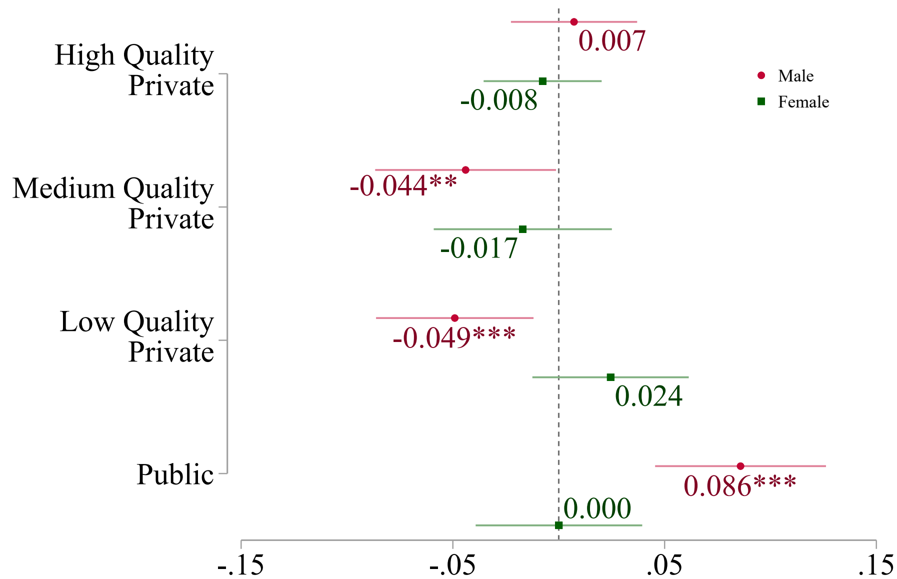
(c) m = Low-quality Private

(d) m = Public

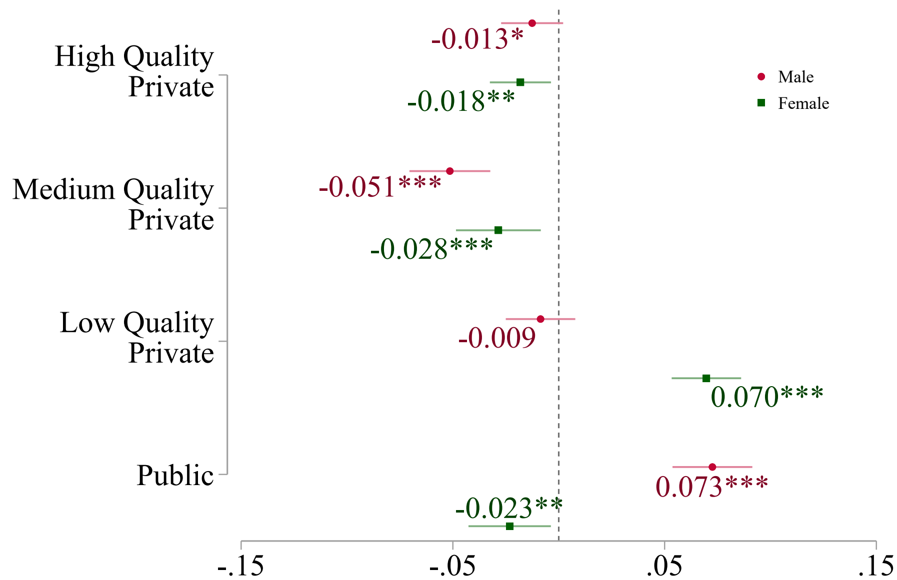
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year-fixed effects. High-quality colleges are all private colleges whose graduates earn more than 2200 PEN while low-quality private colleges are the ones whose graduates earn less than 1450 PEN using administrative data from wage records in 2014. θ^{qN} represents the quintile N of the predicted probability of enrolling in college type m. *Typical students* is the sample of students who were very likely to attend college while *new students* is the sample of students with a low probability of attending college.

Figure 8: Effects on the Likelihood of Enrolling in College Type m by Gender

(a) *Typical Students*



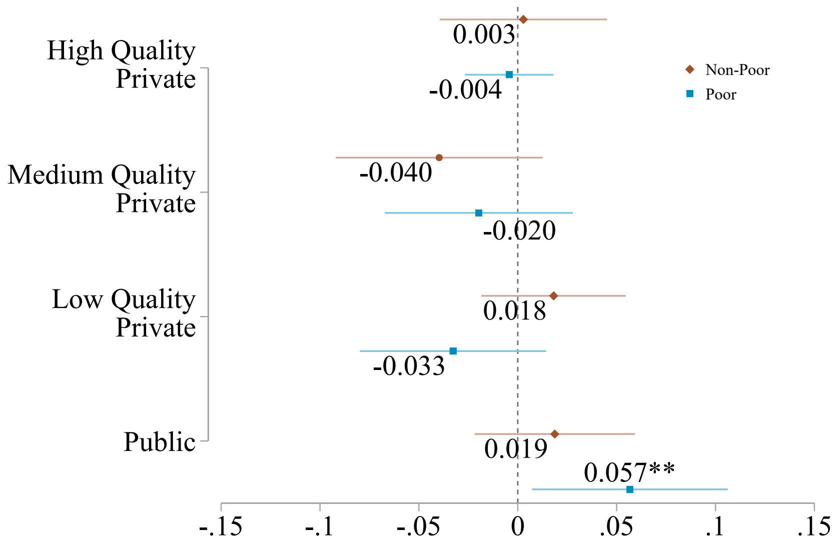
(b) *New Students*



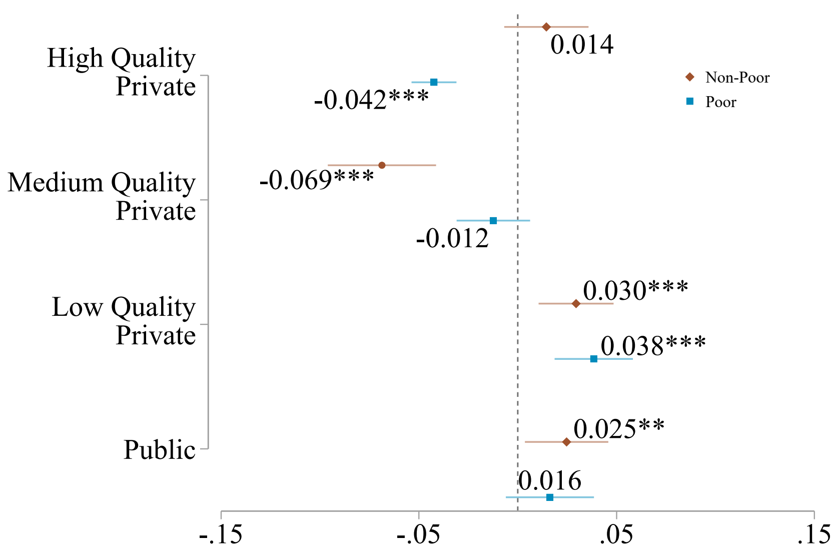
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year-fixed effects. High-quality colleges are all private colleges whose graduates earn more than 2200 PEN, low-quality private colleges are the ones whose graduates earn less than 1450 PEN, and medium-quality private colleges are those whose graduates earn in between. *Typical students* is the sample of students who were very likely to attend college while *new students* is the sample of students with a low probability of attending college.

Figure 9: Effects on the Likelihood of Enrolling in College Type m by SES

(a) *Typical Students*



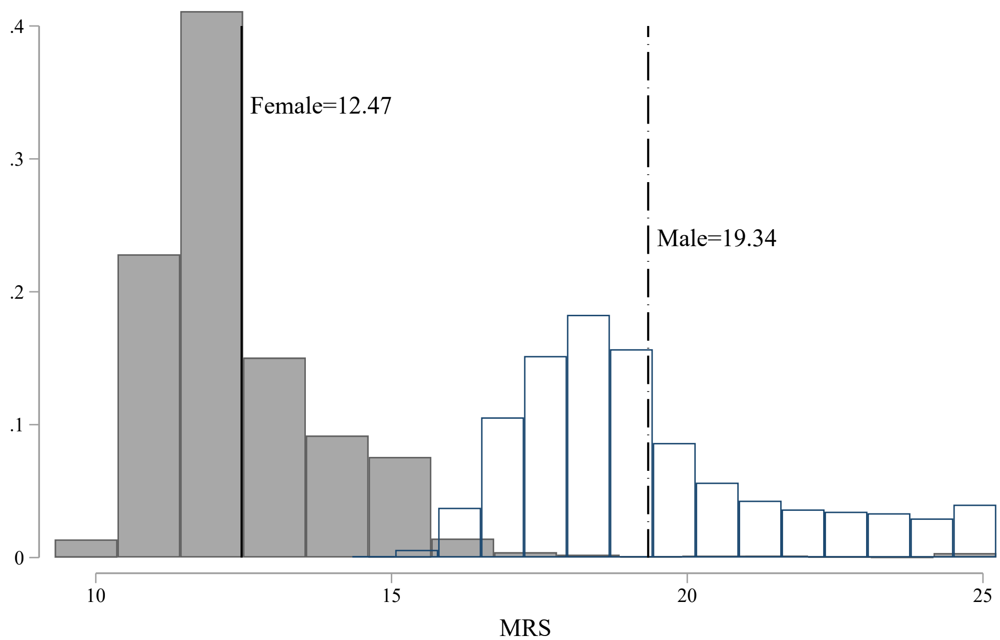
(b) *New Students*



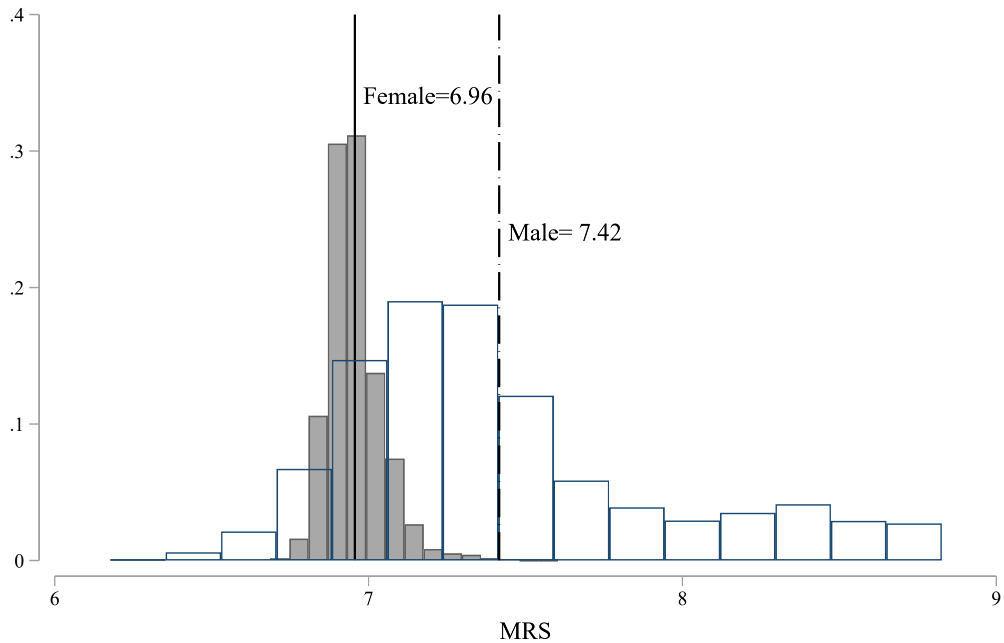
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year-fixed effects. High-quality colleges are all private colleges whose graduates earn more than 2200 PEN, low-quality private colleges are the ones whose graduates earn less than 1450 PEN and medium-quality private colleges are those whose graduates earn in between. *Typical students* is the sample of students who were very likely to attend college while *new students* is the sample of students with a low probability of attending college.

Figure 10: Marginal Rate of Substitution by Gender

(a) Students with High Prob. of Attending Private Colleges

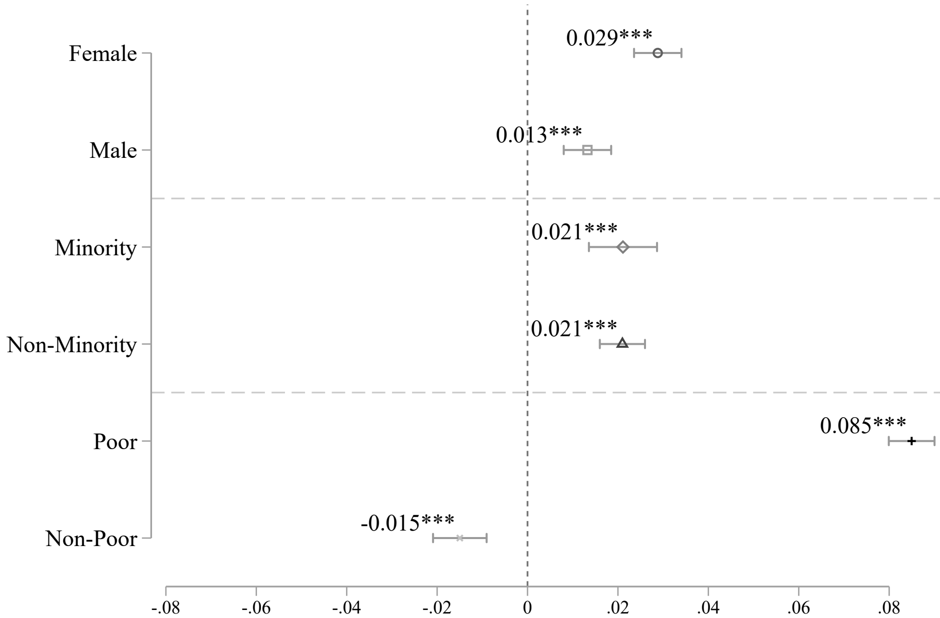


(b) Students with High Prob. of Attending Public Colleges



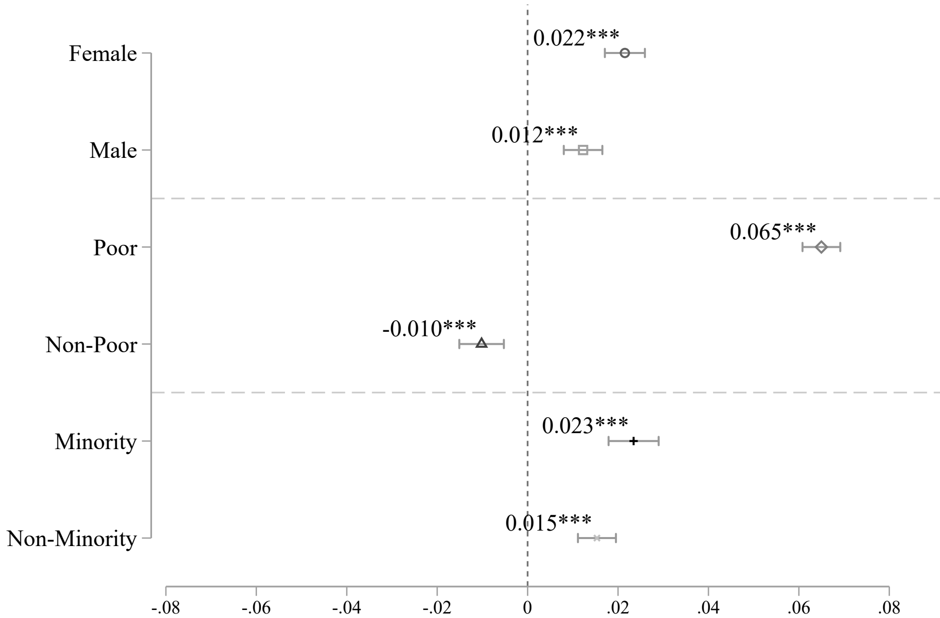
Notes. MRS distributions are winsorized at the 1% and 99% percent.

Figure 11: Heterogeneous Treatment Effects of College Completion



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years of exposure. *Source.* National Census 2017.

Figure 12: Heterogeneous Treatment Effects of White Collar Employment



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years of exposure. *Source.* National Census 2017.

Table 1: Summary Statistics at the Block Level (Pre-treatment)

	(1) Total	(2) Treatment	(3) Control
Panel A :	College Admin. Records Data		
Enrollment (rates)	0.172	0.163	0.176
	(0.363)	(0.357)	(0.368)
Enrollment (log)	0.0911	0.0701	0.0842
	(0.331)	(0.285)	(0.325)
Female Enroll. (logs)	0.0543	0.0416	0.0503
	(0.251)	(0.216)	(0.247)
Male Enroll. (logs)	0.0446	0.0335	0.0422
	(0.225)	(0.192)	(0.223)
Public Enroll. (logs)	0.0281	0.0223	0.0285
	(0.179)	(0.157)	(0.184)
Private Enroll. (logs)	0.0690	0.0517	0.0622
	(0.285)	(0.242)	(0.277)
Distance HH to College	9.695	11.24	10.10
	(5.557)	(5.156)	(6.047)
Panel B:	2007 Census		
Total Population	128.2	123.0	123.6
	(123.8)	(122.2)	(111.2)
Primary School	0.197	0.215	0.201
Pop. Share	(0.143)	(0.141)	(0.146)
Secondary School	0.394	0.415	0.394
Pop. Share	(0.175)	(0.175)	(0.175)
Higher Ed.	0.409	0.370	0.405
Pop. Share	(0.240)	(0.230)	(0.238)
Observations	211,824	78,788	110,284

Notes. This table shows the means at the block (*manzana*) level before 2010. Panel A shows the college enrollment rates using the administrative data from MINEDU. Logarithmic transformations using the inverse hyperbolic sine. Panel B shows summary statistics using the 2007 Census. Total population is the average count by block. Population shares consider people above 25 years old. Higher education includes college and community college (*institutos*).

Table 2: Effects of New Stations on Commuting Time (mins) to College

	All Colleges	Private	Public
	(1)	(2)	(3)
Treatment*Open	-13.33***	-11.81***	-11.66***
	(1.297)	(1.136)	(1.081)
Mean Control	66.54	59.91	79.07
N	582	582	582

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include district and year-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening. Column (2) excludes districts in downtown Lima. Column (3) shows the effects on commuting time to private colleges only while Column (4) shows the effects on public institutions.

Table 3: Effects of New Stations on College Enrollment Rates

	Rates	Log(All)	Log(Private)	Log(Public)	Log(Female)	Log(Men)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment*Opening	0.009***	0.012***	0.010***	0.003***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Dep. Var. Mean	0.142	0.080	0.059	0.026	0.048	0.039
N	411147	411147	411147	411147	411147	411147
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and year-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening. Column 1 shows the effects on enrollment rates. Columns 2-6 show estimates of the effects on the logarithmic transformation of enrollment adjusted by the hyperbolic sine function.

Table 4: Effects on College Choice

	High Quality	Medium Quality	Low Quality	Public
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.013**	-0.038***	0.026***	0.025***
	(0.006)	(0.008)	(0.007)	(0.008)
Dep. Var. Mean	0.112	0.346	0.165	0.377
N	101189	101189	101189	101189
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and year-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening. Columns 1-3 show the results for private colleges.

Table 5: Effects on Distance to College and College Connectivity

	Distance (km.)	Connected	Priv.	Low Q. Priv.	Public
	(1)	(2)	(3)	(4)	(5)
Treatment*Opening	-0.854***	0.038***	0.033***	0.020***	0.005
	(0.094)	(0.008)	(0.007)	(0.006)	(0.005)
Dep. Var. Mean	10.783	0.371	0.224	0.123	0.147
N	101189	101189	101189	101189	101189
Block FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Distance is measured in kilometers using the Euclidean distance from a block centroid to a college location. Columns 2-5 measure the probability of enrolling in a college connected to the new lines, with the closest station less than 3km away. Columns 3 and 4 measure the probability of enrolling in a private and low-quality private college connected to new lines, respectively. Column 5 measures the probability of enrolling in a public college connected to a new line. All regressions include block and year-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening.

Table 6: Mixed Logit Results

	All	Female		Male	
	(1)	(2)	(3)	(4)	(5)
Mean					
Commuting Time	-0.018***	-0.017***	-0.013***	-0.019***	-0.018***
	(0.00)	(0.000)	(0.000)	(0.000)	(0.000)
Wage Premium	0.234***	0.205***	-0.094***	0.377***	0.136***
	(0.002)	(0.003)	(0.007)	(0.004)	(0.005)
SD					
Commuting Time	0.0091***	0.009***	-0.002	0.010***	0.007***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
Wage Premium	0.000**	0.000	-0.000	-0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Students	102,445	42,308	13,180	31,055	15,902
Observations	3585575	1651475	290605	1115730	527765
Log-Likelihood	-350251.8	-162206.6	-28876.0	-106375.1	-51676.5

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Standard errors in parentheses. Wage premium is standardized to mean zero. Robust standard errors.

Table 7: Effects on College Completion

	Excl. DT	All
	(1)	(2)
Treatment*Open	0.021***	0.021***
	(0.002)	(0.002)
Dep. Var. Mean	0.173	0.182
N	497962	607928
Block FE	Yes	Yes
Year FE	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years of exposure. *Source.* National Census 2017.

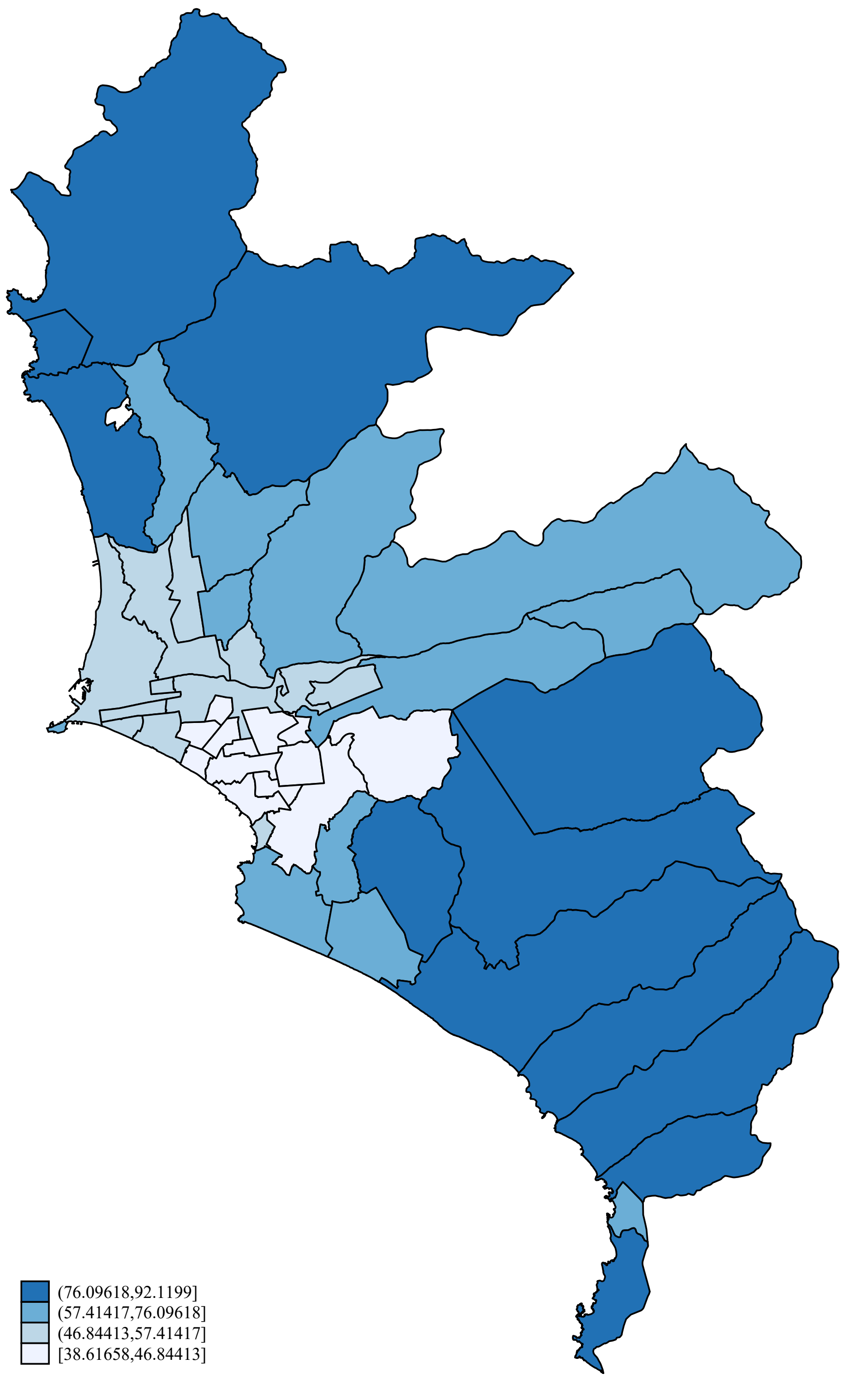
Table 8: Effects on Employment Rates

	Employ.	Blue C.	White C.
	(1)	(2)	(3)
Treatment*Open	0.010*	-0.007	0.017***
	(0.005)	(0.005)	(0.004)
Mean	0.669	0.404	0.265
N	192014	192014	192014
Block FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years of exposure. *Source.* National Census 2017.

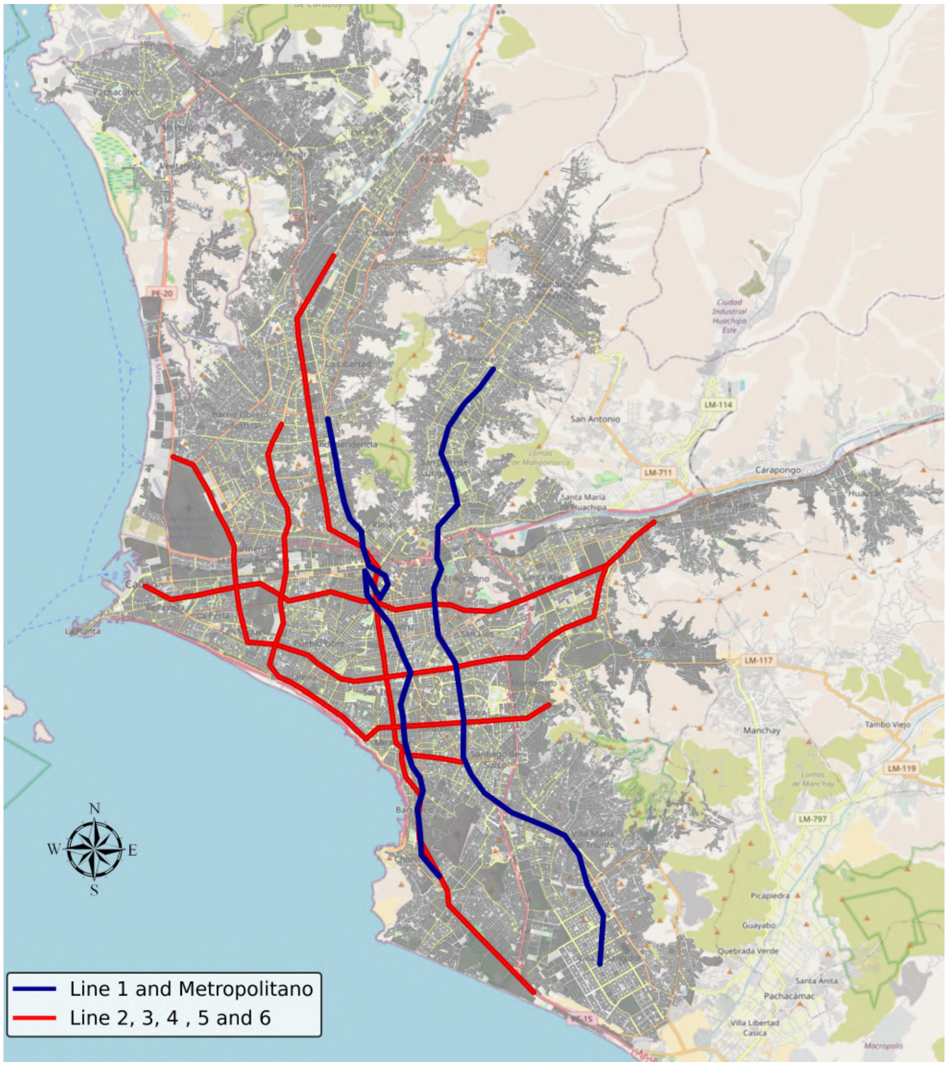
A Additional Figures and Tables

Figure A.1: Average Commuting Time from Home to College (in min)



Source. CENAUN 2010. Travel time is self-reported in minutes.

Figure A.2: Planned-but-not-executed Metro lines in Lima



Notes. Red lines are the planned lines according to the Peruvian Ministry of Transport. Blue lines represent Linea 1 (to the left) and the Metropolitano (to the right).

Figure A.3: College Campus Location

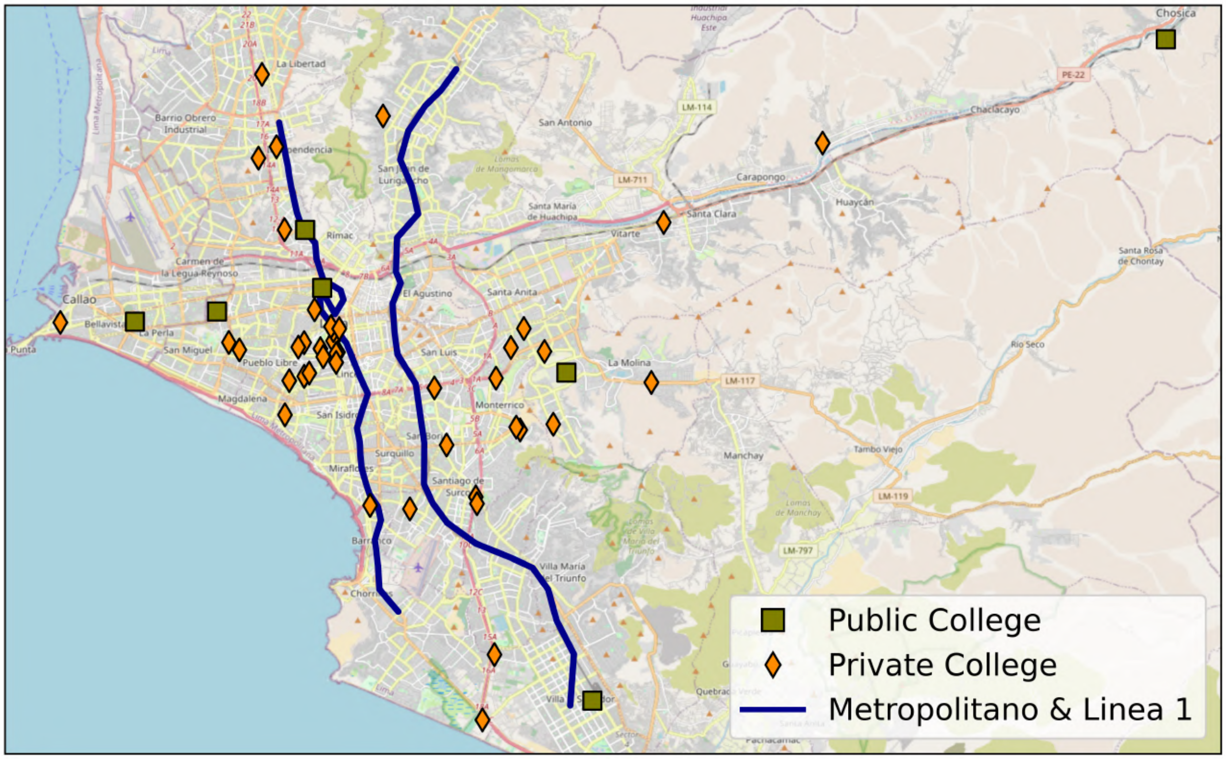
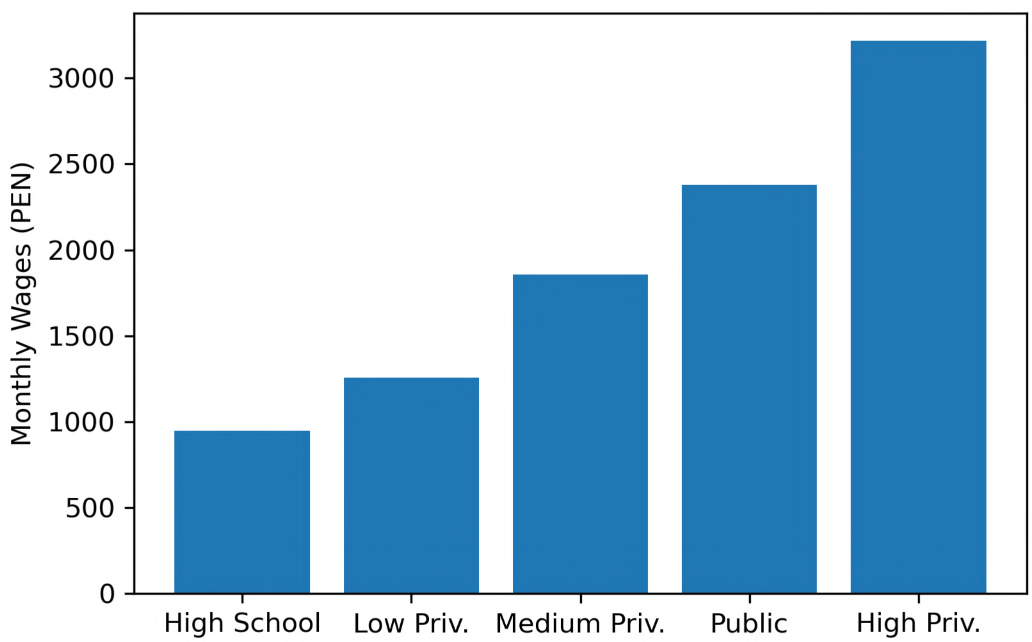
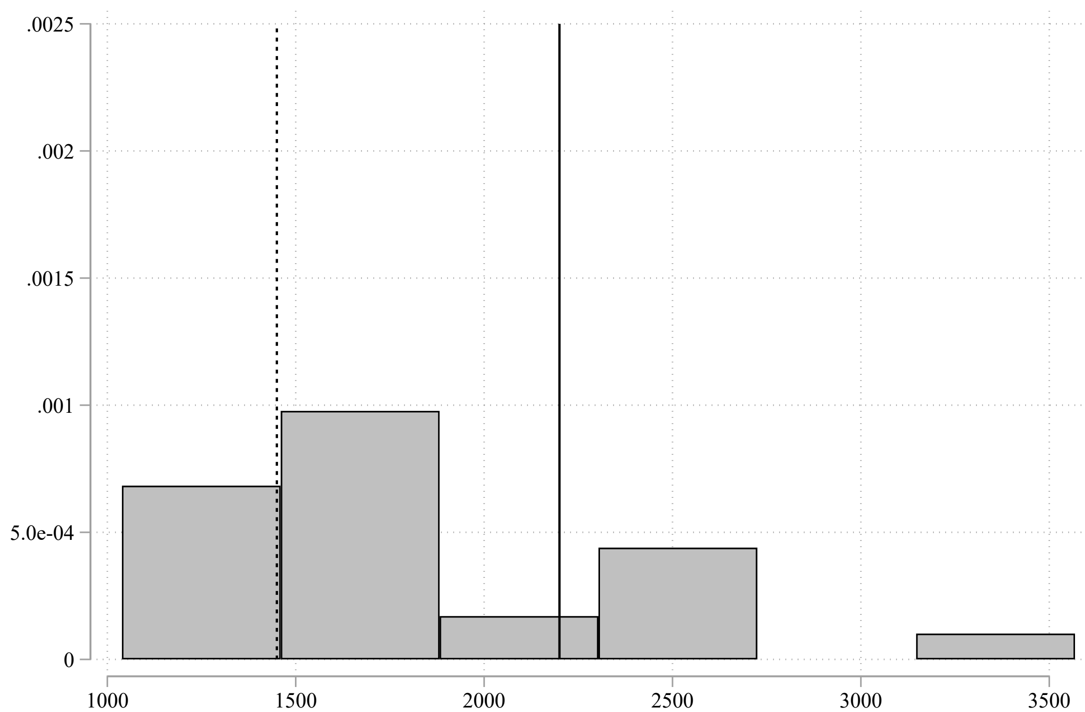


Figure A.4: Wage Returns by Educational Level



Source. ENAHO 2013-2024. Planilla Electronica, details on data are described in Alba-Vivar et al. (2023). I include wage returns for individuals between 25-35 years old living in Lima Metro Area.

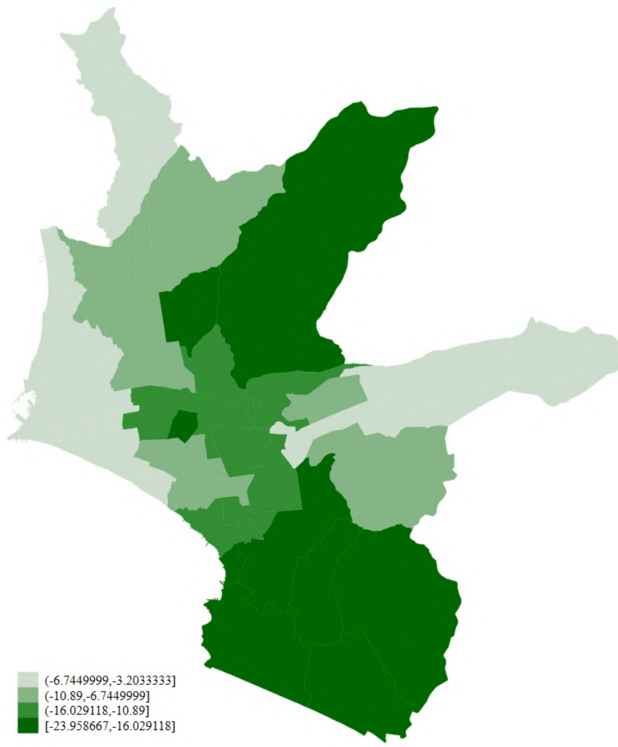
Figure A.5: College Wage Distribution (Average by College)



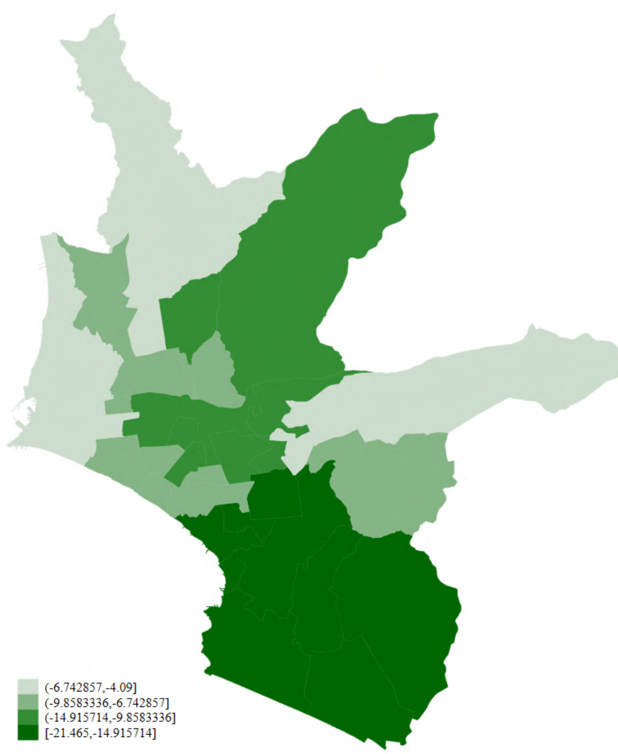
Source. Planilla Electronica, details on data are described in Alba-Vivar et al. (2023). The dashed line shows the limit for the 25% bottom of the distribution below 1450 PEN while the plain line marks the 25% top of distribution above 2200 PEN.

Figure A.6: Commuting Time to College (with/without new stations)

(a) All colleges



(b) Private colleges



Notes. Maps show the variation in commuting time to college by districts with/without new stations. Panel (a) shows variation in average commuting time to all colleges while Panel (b) shows variation in average commuting time to private colleges only.

Figure A.7: Share of population who access higher education, Census 2007

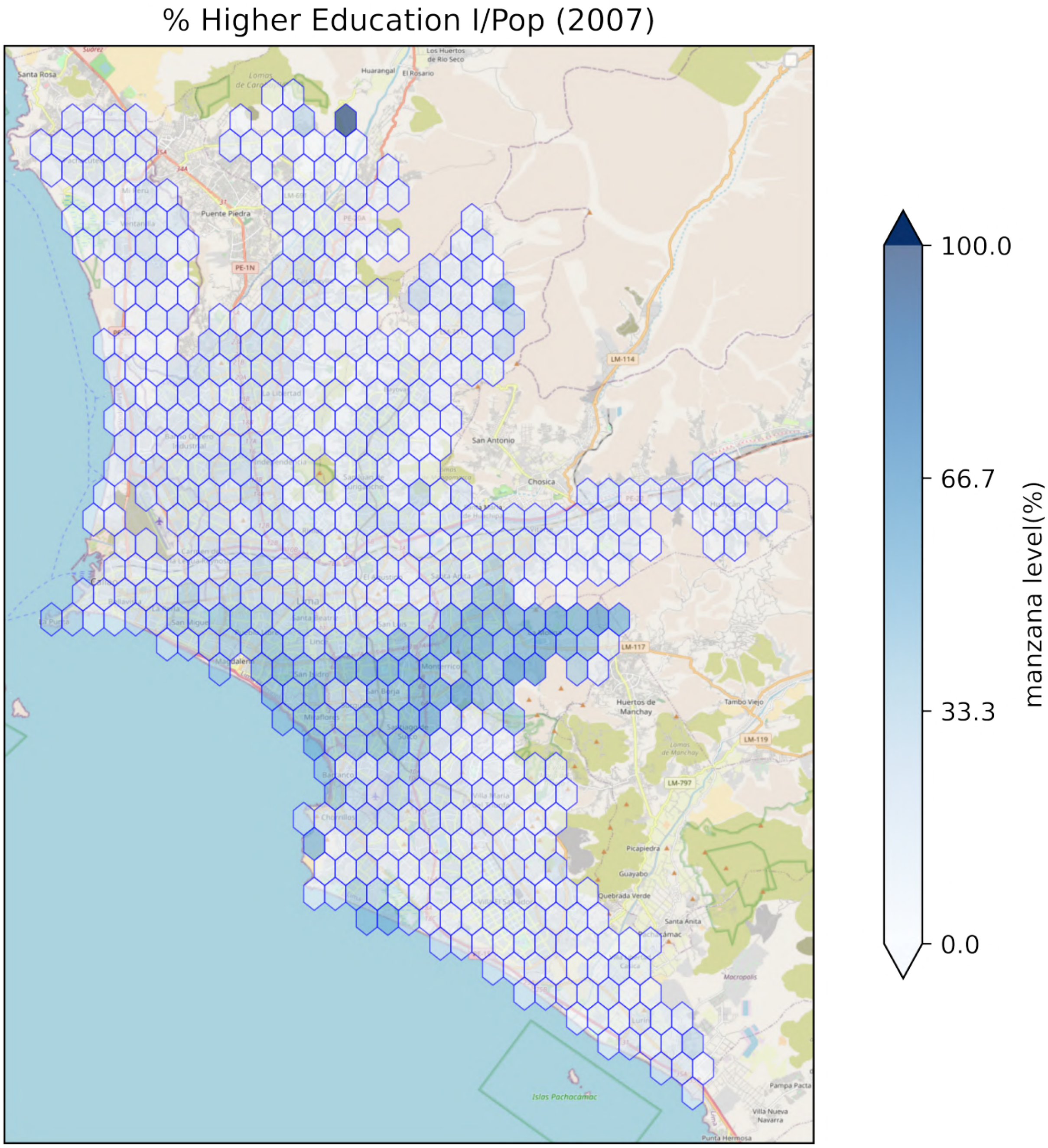


Figure A.8: Effects on College Enrollment Rates using Callaway and Sant'Anna (2021) estimator

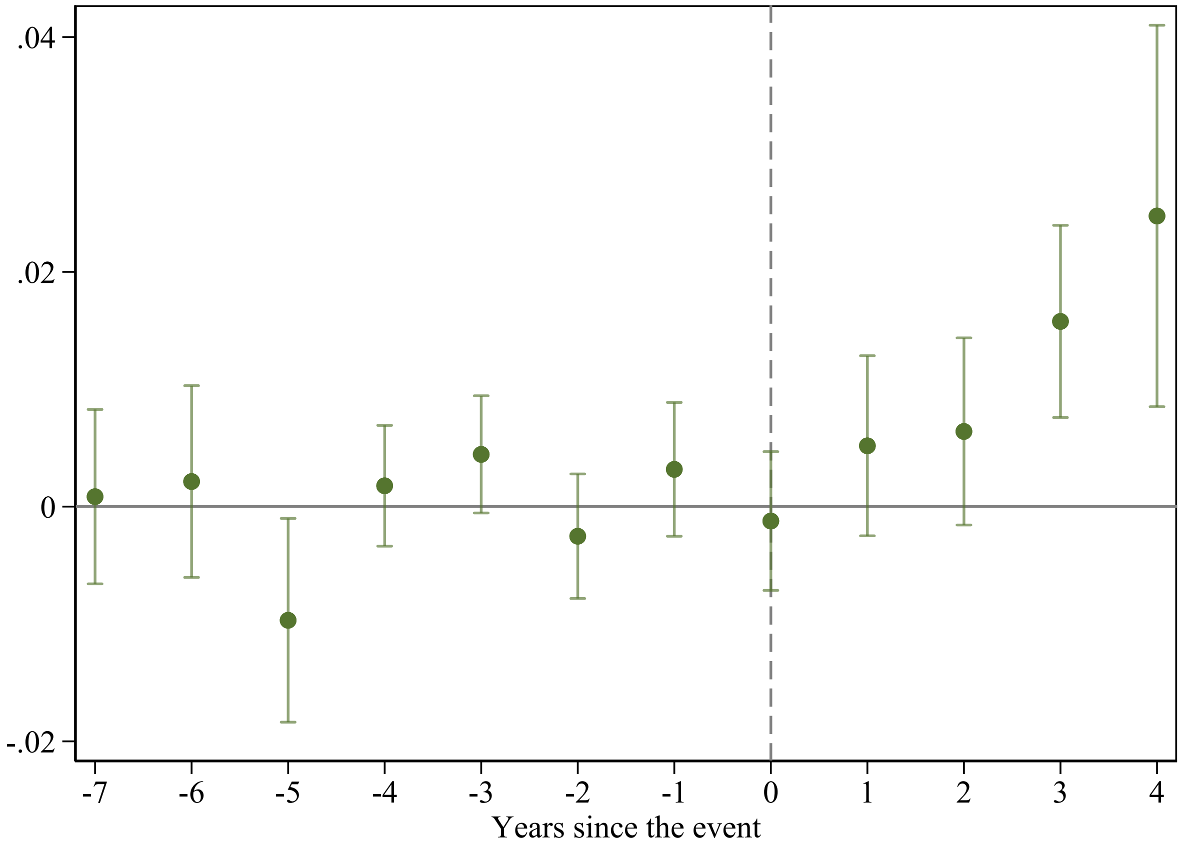
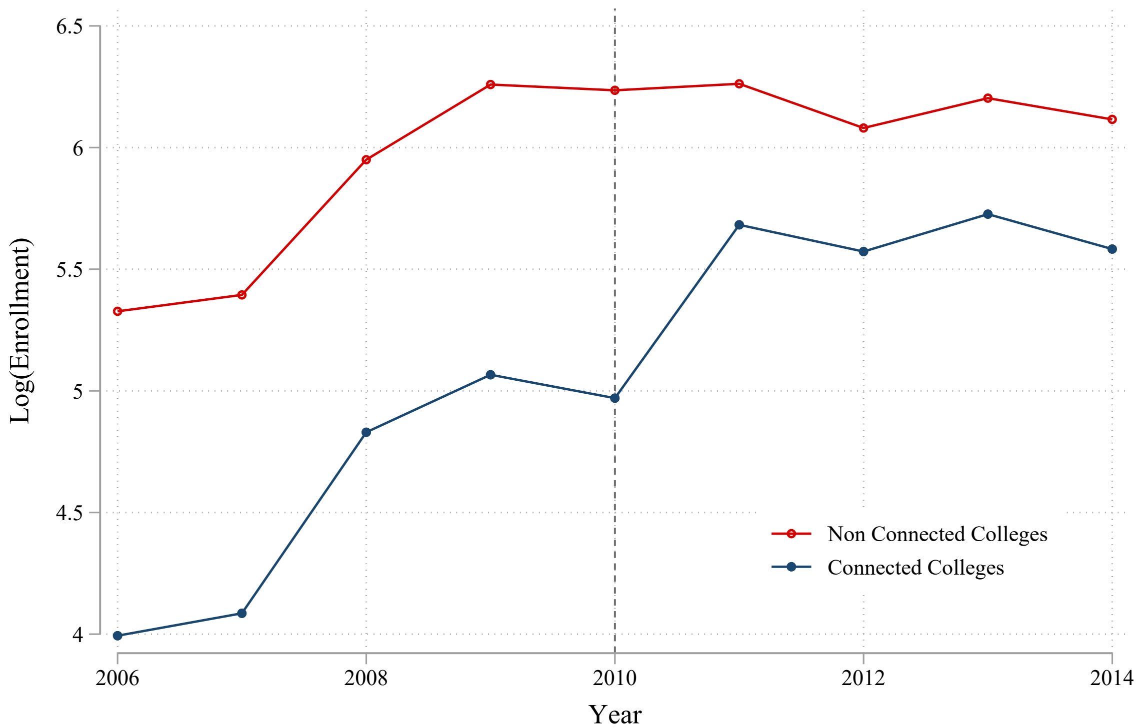


Figure A.9: Enrollment and Major Trends Over Time

(a) Log(enrollment) for connected and non-connected colleges



(b) Log(major counts) for connected and non-connected colleges

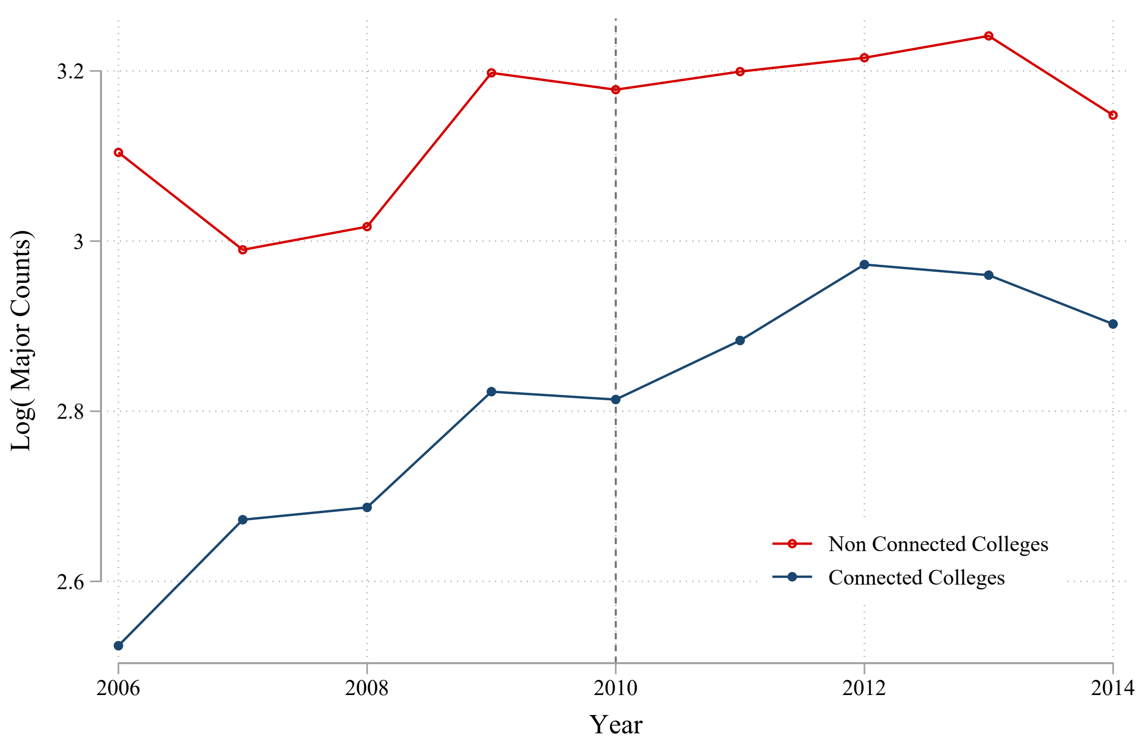


Figure A.10: Log(enrollment) by Gender for connected and non-connected colleges

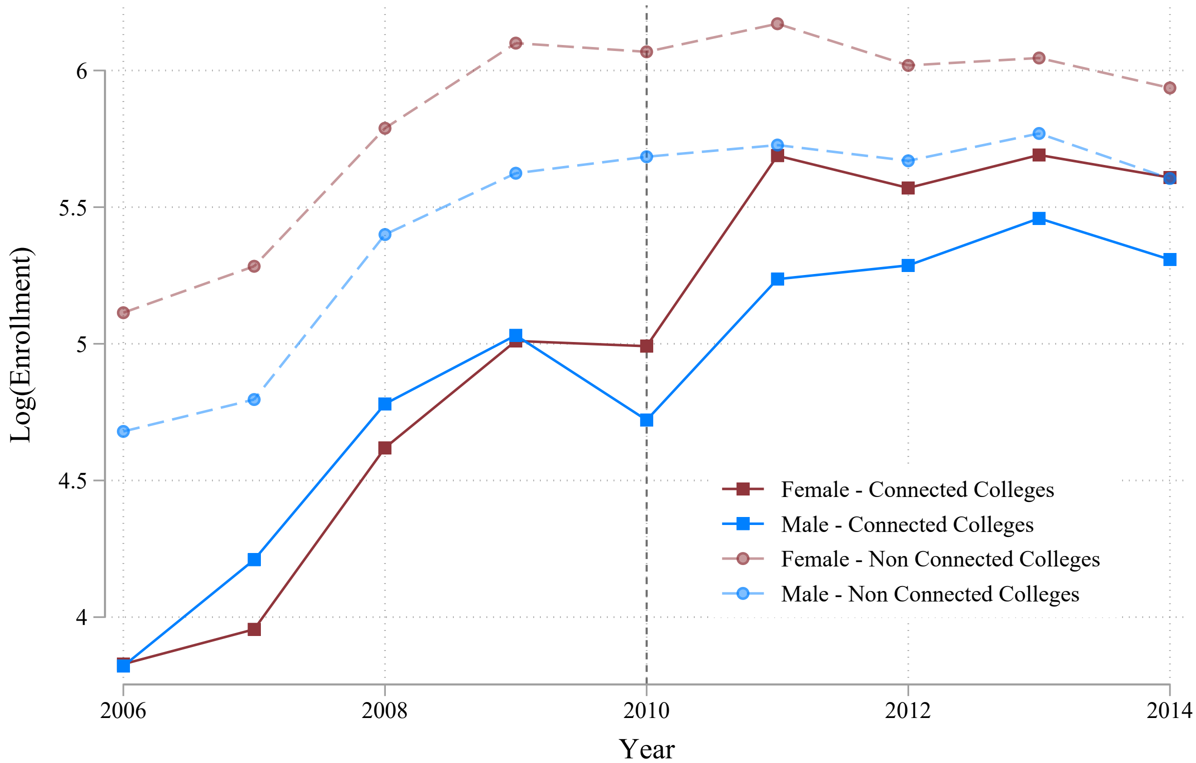
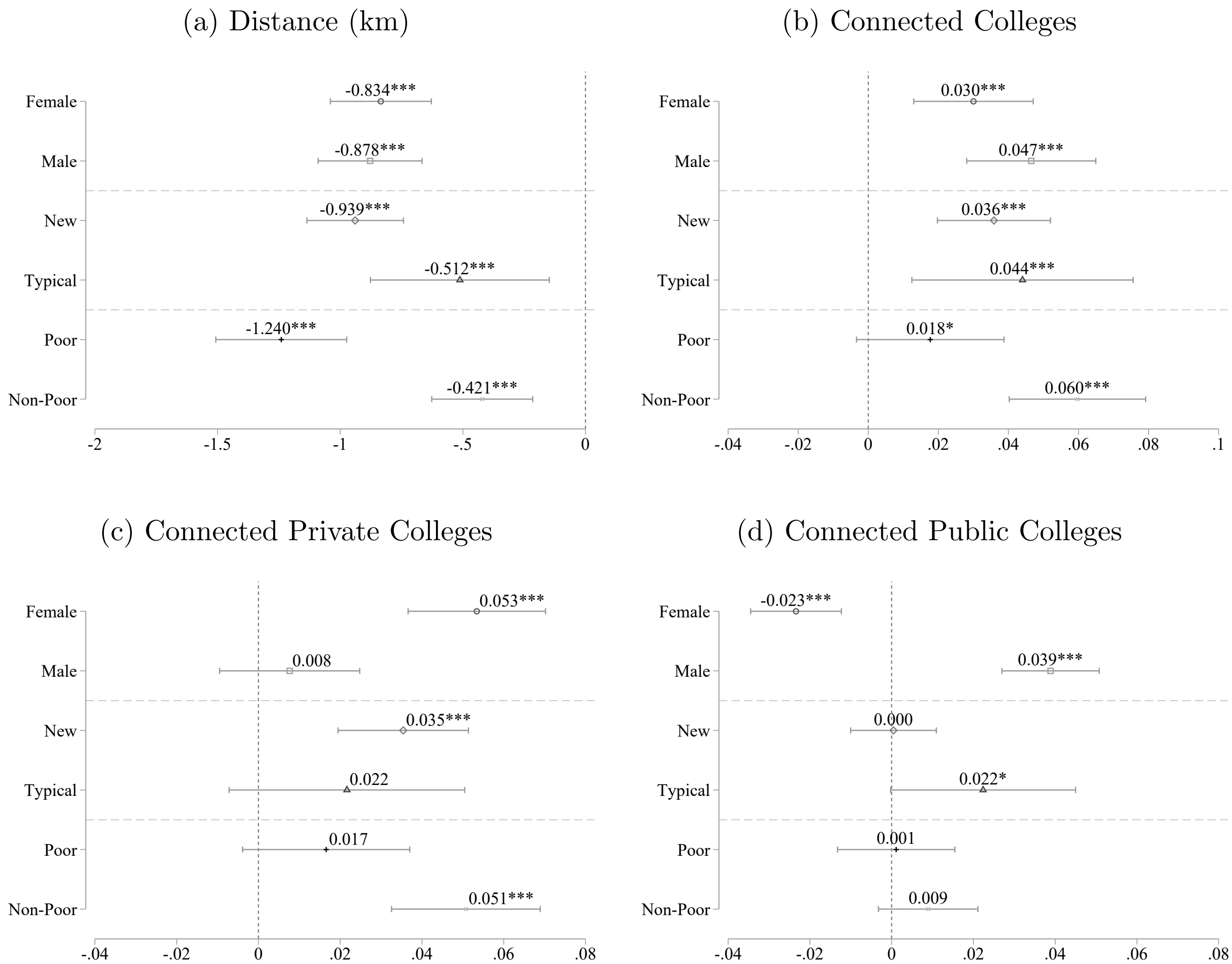
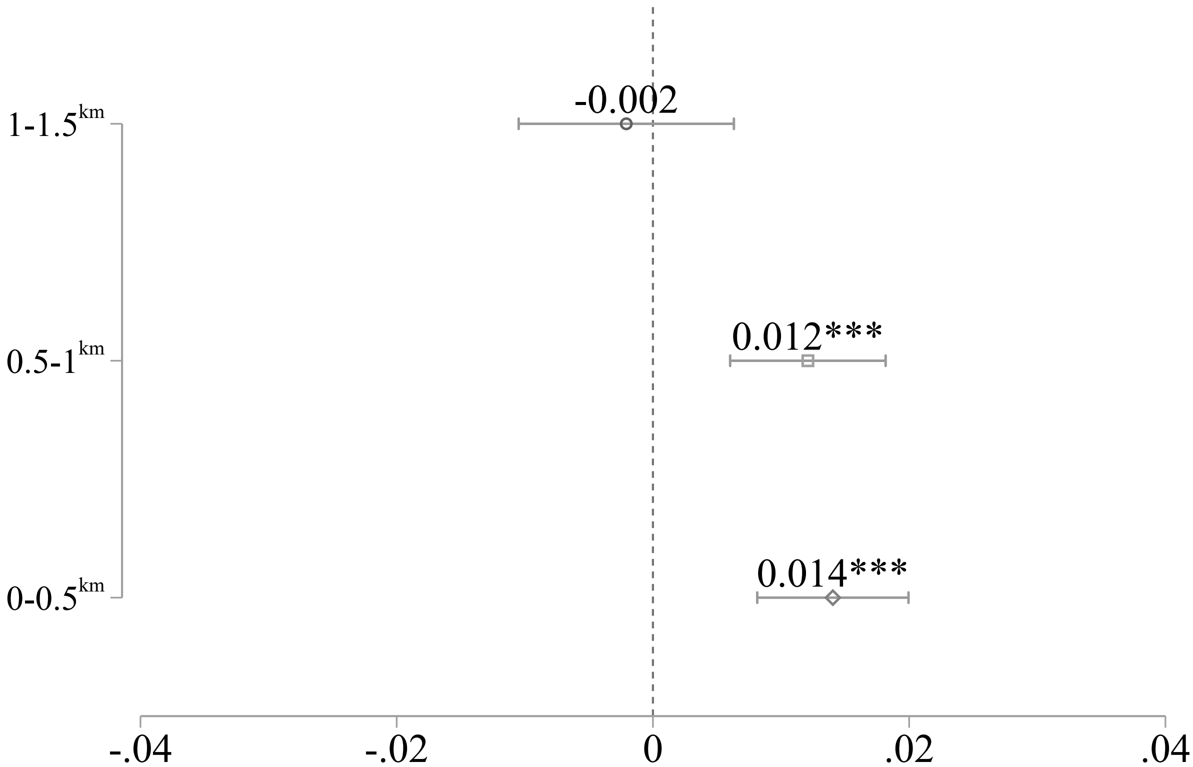


Figure A.11: Heterogeneous Effects of Distance to College and College Connectivity



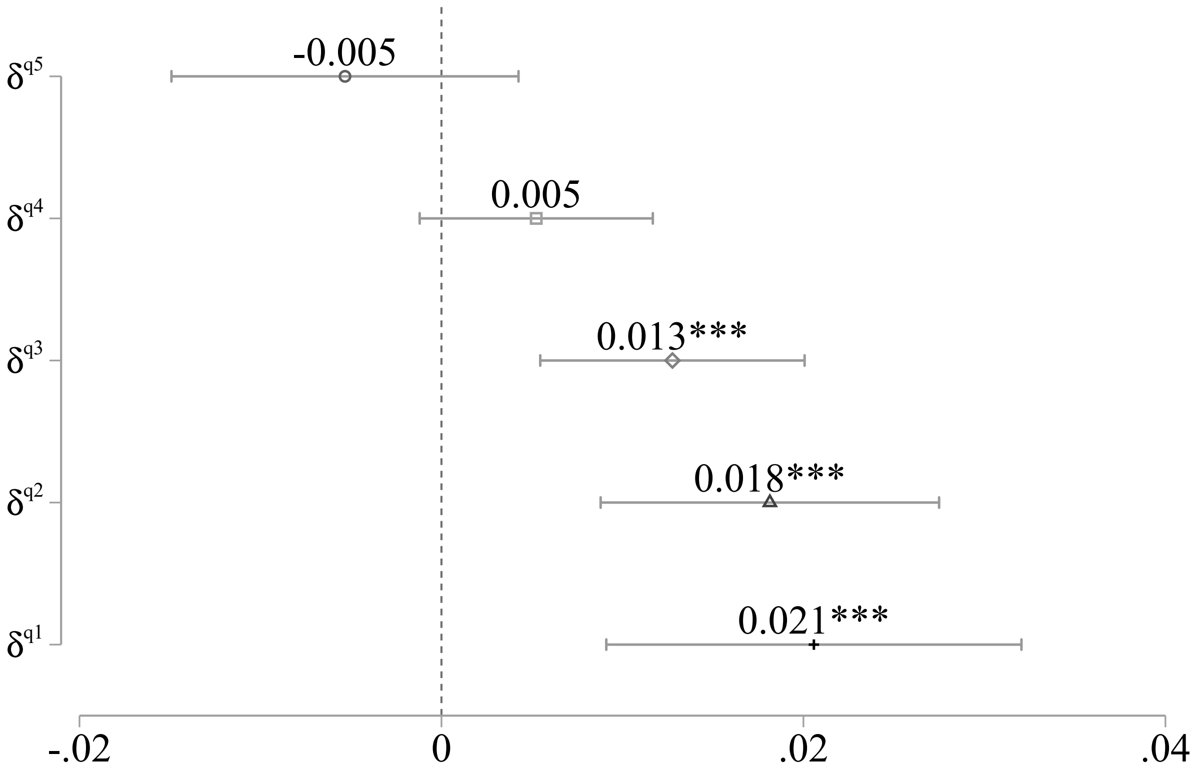
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Distance is measured in kilometers using Euclidean distance from block centroid to college location. Panels (b-d) measure the probability of enrolling in a college connected to new lines, with the closest station less than 3km away. Panel (c) and (d) measure the probability of enrolling in a private and a public college connected to the new lines, respectively. Regressions include block and year-fixed effects.

Figure A.12: Impact on College Enrollment Rates by Distance to Stations



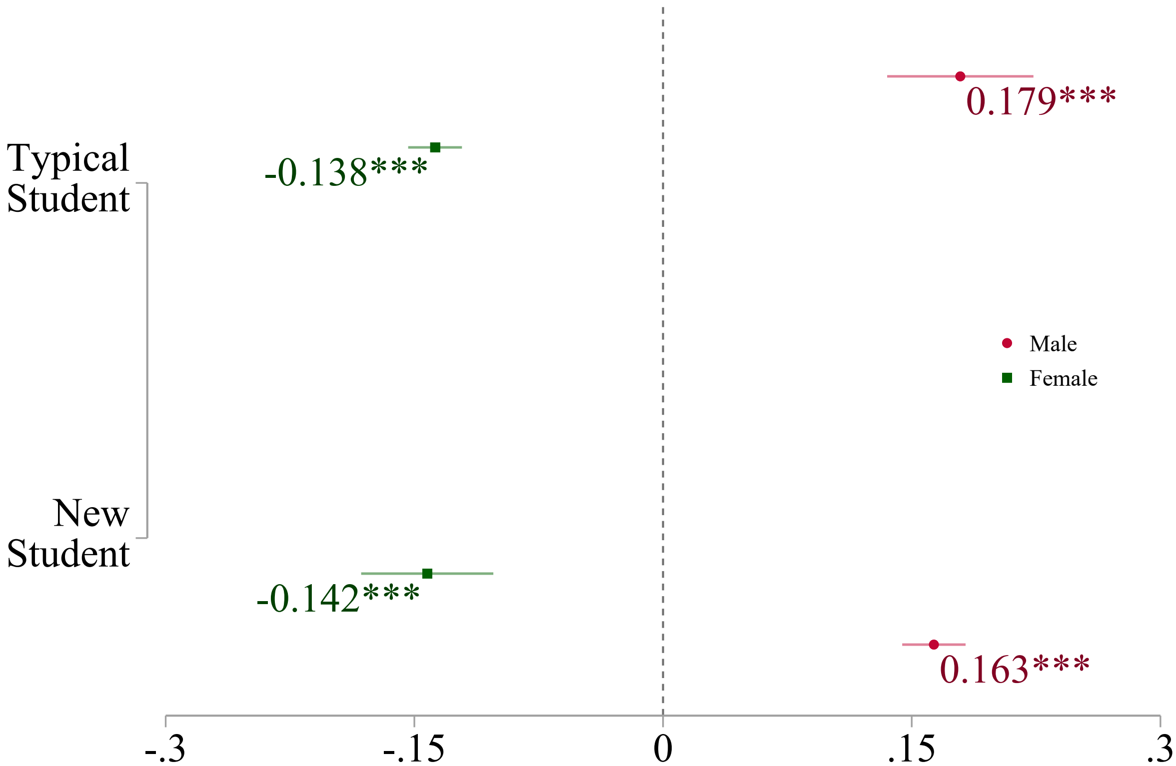
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and year fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening.

Figure A.13: Impact on College Enrollment Rates by Distance to Downtown



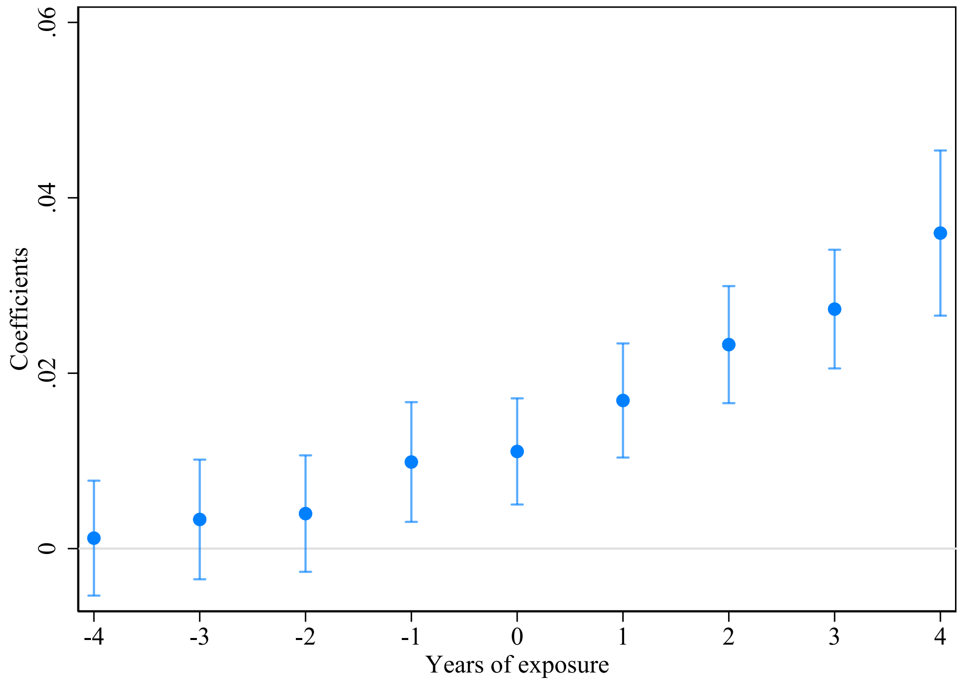
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. δ^{qN} is the quintile N of the distance to downtown distribution measured in kilometers. All regressions include block and year fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening.

Figure A.14: Effects on the Likelihood of Enrolling in a STEM Major



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year-fixed effects. *Typical students* is the sample of students who were very likely to attend college while *new students* is the sample of students with a low probability of attending college.

Figure A.15: Dynamic Effects of the New Lines on College Completion Rates



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and cohort fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening.

Table A.1: Effects on College Choice by Access

	High Quality		Medium Quality		Low Quality		Public	
	New	Typical	New	Typical	New	Typical	New	Typical
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment*Opening	-0.016**	-0.001	-0.039***	-0.029	0.034***	-0.009	0.020**	0.039**
	(0.007)	(0.012)	(0.009)	(0.019)	(0.007)	(0.017)	(0.009)	(0.018)
Dep. Var. Mean	0.112	0.109	0.347	0.342	0.163	0.177	0.378	0.373
N	79965	21224	79965	21224	79965	21224	79965	21224
Block FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and year fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening. *Typical students* is the sample of students who were very likely to attend college while *new students* is the sample of students with a low probability of attending college.

Table A.2: Effects on the Likelihood of Enrolling in a High-Quality College (Alternative Definitions)

	Elite	Diff. Elite	Licensed	Diff. Licensed
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.032***	-0.035***	-0.016***	0.025***
	(0.007)	(0.007)	(0.006)	(0.007)
Mean	0.077	0.075	0.882	-0.835
N	101187	101187	101187	101187
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$.Regressions include block and year-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening.

Table A.3: Effects of New Lines on College Enrollment Rates including Downtown Lima

	Rates	Log(All)	Log(Private)	Log(Public)	Log(Female)	Log(Men)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment*Opening	0.009***	0.011***	0.008***	0.003***	0.006***	0.005***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Dep. Var. Mean	0.144	0.093	0.071	0.029	0.056	0.046
N	461151	461151	461151	461151	461151	461151
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$.Regressions include block and year-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening. Column (1) shows the effects on enrollment rates. Columns (2-6) show estimates of the effects on the logarithmic transformation of enrollment adjusted by the hyperbolic sine function.

Table A.4: Effects of the New Lines on College Choice (relative to Predicted)

	High Quality	Medium Quality	Low Quality	Public
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.015**	-0.017	0.027***	0.016
	(0.006)	(0.011)	(0.007)	(0.010)
Dep. Var. Mean	0.112	0.346	0.165	0.377
N	101189	101189	101189	101189
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year-fixed effects. ATTs are calculated using the Borusyak et al. (2023) estimator for the first 4 years after a station opening.

B Conceptual Framework

In this simple model, students are utility maximizers. A student will choose college c from the choice set $C = [c_1, c_2, ...c_N]$ based on the utility coming from Equation 7. Other important factors like college reputation, non-monetary preferences, etc. are omitted for simplification.

$$U_i^C = f(W_i^C, D_i^C, X_i^C) \tag{7}$$

Where W_i denoted the potential earnings a student i gets when attending college c, D is the distance to college and X summarizes the students' taste for degrees and college amenities. I assume a linear utility function where β^n represents weights for each variable. A student i will choose a college c that yields the highest net utility subject to being able to afford a budget that covers tuition and transportation costs, where P is the tuition cost to attend college c and TC is the transportation costs. The outside option is to not attend college which yields a net utility of zero, $U^{NC} = 0$.

$$\begin{aligned} \max \sum_{n=1}^{N} U_i^C \quad &= \sum_{n=1}^{N} f(W_i^C, D_i^C, X_i^C) = \sum_{n=1}^{N} \beta^1 W_i^C + \beta^2 D_i^C + \beta^3 X_i^C \\ s.t. \quad B_i^C &= P_i^C + TC_i^C \end{aligned} \tag{8}$$

In the case where we have 2 colleges available for student i, colleges A and B, a student i will choose as follow:

$$\begin{aligned} U^A - TC^A - P^A > U^B - TC^B - P^B, \quad C^* = A \\ U^B - TC^B - P^B > U^A - TC^A - P^A, \quad C^* = B \end{aligned} \tag{9}$$

This simple model highlights the trade-off between future wages, distance, and personal preferences. Importantly, a change in transportation costs TC can increase students' net utility for colleges that have easier or worse access when everything else remains constant.

C Brief History of Public Transportation in Lima

Back in the mid-19th century, the Peruvian government the inaugurated first railway: the Lima-Callao line (in 1851). It connected the capital city with the nearby port city of Callao, providing a much-needed transportation link between the two cities. This line closed in was closed in the 1970s as the government shifted its resources toward the construction of highways and other means of transportation.

During these years, the idea of a train transportation system in Lima became popular. In 1973 the first Technical-Economic Feasibility Study and Preliminary Project for the Massive Passenger Transportation System in the Lima and Callao Metropolitan Area was completed. This study proposed 4 underground trains: Line 1 which connected Comas and Villa el Salvador (37km), Line 2 which connected San Borja and Maranga (13 km), Line 3 which connected Rimac and San Isidro (10 km) and Line 4 that connected La Victoria with Carmen de la Legua (10 km). These four lines added a total of 125 kilometers in total (Narrea, 2017). However, the execution was postponed in 3 governments due to economic and technical factors. It is only in the late 80s when President Alan Garcia announced the construction of the first line. During this government, the construction of an electric mass transportation system for Lima and Callao was declared of national interest. For this purpose, the Autonomous Authority of the Lima and Callao Mass Rapid Transit Electric System Special Project (AATE) were created in 1986. The agency was in charge of planning, coordinating, supervising, controlling and executing the mass transit electric system. The new system proposed 5 lines of electric trains, but only Line 1 (22 km) (Villa El Salvador - Av. Grau) was prioritized. According to Kohon (2016), this could be explained by three reasons: i) the north-south axis was prioritized to avoid the excessive growth of the east side of Lima, ii) the available surface area on a main avenue to build a viaduct instead of an underground network, which meant save costs and technical problems, and iii) connect emerging sectors such as Villa El Salvador.

The construction of the project began in 1986, but by 1990 only 1.5 km was built. The economic and political crisis stopped the project from continuing. In the 1990s, AATE was part of the Council of Ministers and proposed a complementary study of the Lima Metro Network in 1998 (MTC, 2005). Then, at the beginning of 2000's, the Municipality of Metropolitan Lima developed the COSAC study, a preliminary study of a BRT (a Metropolitano antecedent). As a consequence, the AATE changed the route of line 1 (tramo II) from Av. Grau to San Juan de Lurigancho instead of Comas. [32] In 2004, during the government of Alejandro Toledo, a new Law No. 28253 was published and once again, declared the execution of the Lima and Callao mass transportation electric system to be a public necessity. In 2006, through Law No. 28670, the extension of Line 1 of the Lima Metro, from the Atocongo Bridge to Grau Avenue (tramo I), was declared of national interest. However, the public-private concession attempts failed, as there were no bidders. According to Kohon (2016), this is explained by issues on the main feasibility study that did not consider the demand risk in this project. In this sense, contract terms were reconsidered: from a conventional public project to a public-private project, the construction and operation of the train system were separated (Campos et al., 2021). Finally, Linea 1 project was bidded and granted to the concession holder *Consorcio Metro de Lima*: the union of two major private companies: Odebrecht and Graña y Montero.

The construction of Line 2 of the Lima Metro was also granted to the *Consorcio Metro de Lima*. The project began in 2014, but it stopped two years later. The government failed to meet the deadline for the expropriation of properties that would provide the required land for the execution of the project. As a result, the government and the concession holder filed claims with the International Centre for Settlement of Investment Disputes. Additionally, this project had weak political support from public opinion in the face of the bad experiences from Linea 1 (Sallo and Hickman, 2021). Additionally, the Linea 1 project was involved in several corruption cases regarding political bribes for presidential campaigns which are currently under investigation.

[32] The AATE also left the Council of Ministers and became part of the Municipality of Metropolitan Lima.

To sum up, the delay in the implementation of the Metro is explained by several factors. On the institutional side, the AATE [33] was sensitive to changes in public administration. In addition, many other public agencies are involved in the execution of megaprojects, increasing the bureaucracy for permits and approvals. Since different government agencies have different objectives, priorities, and visions, it also generated strong coordination problems. Another limiting factor was the lack of political support from public opinion due to cases of corruption, distrust in politicians, lack of information about the social benefits of the project, and among others.

[33] In 2019, ATTE became the Urban Transport Authority (ATU, in Spanish)

D Predicted College Access and Predicted College Choice

In this section, I describe the procedures to calculate both the predicted access to college and the predicted college choice for certain types of colleges. For both analyses, I use the k-nearest neighborhood (kNN) algorithm, which is particularly good for a setting like this.

The kNN is a non-parametric algorithm that uses proximity to make a prediction. One key advantage is that it is able to capture non-linearity. To avoid its sensitivity to the choice of a distance metric, I use the grid search technique to find the optimal value of k, the number of nearest neighbors to consider when making predictions. This optimal k maximizes the test dataset score using cross-validation. Table D.1 shows the results of this procedure.

The algorithm proceeds as follows:

1. I normalized and standardized all data using the StandardScaler option from the *sklearn* package in Python to avoid having extra sensitivity to data errors.

2. I trained the algorithm using all data from the Peruvian Census of 2017 at the block level and individual-level characteristics such as sex and age (this only when predicting college choice since this information is conditional on enrolling in college). I do not include any information after 2010, the year when the first line opened, or any information regarding the treatment status.

3. After calculating the optimal k value using the grid search method with k-fold validation, I calculate the key parameters on the training data. Then I calculate the predicted value for the whole data set (before and after the new lines).

Table D.1: kNN Scores

	Access to College	Private			Public
		Low Quality	High Quality	Elite	
Best mean cross-validation score	0.995	0.896	0.887	0.902	0.678
Best k parameters	19	19	19	19	19
Test-set score	0.995	0.895	0.888	0.905	0.678