Opportunity Bound: Transport and Access to College in a Megacity

Fabiola M. Alba-Vivar ⋆

(Job Market Paper)

September, 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. First, at the neighborhood level, I find that a 17 percent reduction in commuting time to college leads to an increase by 1 *p.p.* in enrollment rates, relative to a 14 percent baseline and this is primarily driven by an increase in private college enrollment. Second, I find that female students who enrolled in college induced by the policy are more likely to enroll in low-quality private colleges that are also connected to the new lines. In contrast, men are more likely to enroll in public colleges, which are more dispersed over the city. Using a model of college choice, I show that male students are willing to travel 55 percent more time than female students, in order to enroll in a private college offering wage returns one standard deviation higher. In the medium and long run, access to transport increases a person's likelihood of graduating from college by 12 percent and access to white-collar jobs by 4 percent. These results suggest that better transportation can increase human capital accumulation but the increase in opportunities is limited by gender differences in their willingness to travel.

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

I thank Alex Eble, Cristian Pop-Eleches, Micaela Sviatschi and Sandra Black for their constant guidance and support. I also gratefully acknowledge all the comments and suggestions from Sofia Amaral, Mark Anderson, Sam Asher, Dan Bjorkegren, Girija Borker, Kirill Borusyak, Kevin Donovan, Paul Goldsmith-Pinkham, Carolina Lopez, Bentley MacLeod, Sebastian Otero, Nishith Prakash, Maria Rosales-Rueda, Sebastian Tello-Trillo, Anja Tolonen, Daniel Velasquez, Eric Verhoogen and Jack Willis. I thank Matteo Magnaricotte, Carolina-Concha Arriagada and Stephen Keating for their invaluable support on this project. I am also thankful to all participants at the WEAI Graduate Student Workshop, AEFP, 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

Hundreds of millions of people live in "megacities" – metropolitan areas with populations exceeding 10 million – in the developing world.[1] As a result, the welfare of one in eight people on the planet is determined in large part by the amenities these cities provide and how accessible they are. Improving public transportation not only amplifies access to better employment prospects and increases welfare but also holds the potential to mitigate urban inequalities (Balboni et al., 2020; Tsivanidis, 2022; Zarate, 2022). Nonetheless, less is known about how it can impact a key driver of economic mobility: human capital accumulation. In this paper, I study how a dramatic improvement in public transportation – and, as a result, a large decrease in the cost of transit – changes the face of tertiary education.

This paper examines the causal impacts of new public transportation infrastructure on both access to college and college choice in the context of Lima, the capital of Peru and a megacity of 12 million people. I study the opening of two new mass public transportation systems in the early 2010s: a rapid bus transit line and a new train line which reduced commuting time to college for thousands of college students each year. Prior to the rollout of these new transit lines, students in Lima spent almost two hours per day, on average, 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 begin with administrative college enrollment records for almost 450,000 students over 10 years. I geocode the locations of both colleges and students' households and then link these to data on the location and date of opening of the public transit stations from the two lines. I use a difference-in-differences approach to generate causal estimates of how access to improved public transportation affects

[1] A recent report by Euromonitor highlights how most of the future upcoming megacities in the world will be from developing countries. These upcoming megacities include Luanda, Dar es Salaam, Baghdad, Chennai, Bogota, and Chicago, which are currently experiencing higher economic and population growth.

[2] According to the 2010 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 New Dheli as documented by Borker (2020).

college access and choice. My 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 public transit stations to the same cohorts in neighborhoods that were exposed to planned but non-opened stations, similar to the placebo strategy implemented in Donaldson (2018).

First, I show that neighborhoods that are connected to new lines experience a reduction of 17% in commuting time to college[3] which leads to a 6.3% increase in college enrollment rates (or 1 *p.p.* relative to a 14% baseline rate). Furthermore, the results show that the increase in access is mostly driven by private college enrollment, whereas the impact on public college enrollment remains positive but economically negligible. In the Peruvian context, public colleges are tuition-free, highly selective, and have limited ability to expand enrollment which can explain why the effects are limited.[4] Notably, I do not observe significant differences in the enrollment rates effects by gender, and most of the effects are driven by neighborhoods that typically had low or zero college enrollment rates, suggesting that this increase is coming from new students who would not have gone to college in the absence of the new lines.

Second, I find that a decrease in commuting time to college also corresponds with a change in college choice. I use machine learning techniques to identify students who were induced to enroll in college, and if they were likely to enroll in a low-quality private college, a high-quality private college, or a public college. This allows me to disentangle the overall effects between (*new students*) accessing college versus (*typical students*) who were likely to go to college but are changing their decision. Overall, I find that students who live in neighborhoods connected to the new stations are more likely to enroll in low-quality private and less likely to enroll in high-quality private colleges and these effects are only for *new students*. In contrast, I find an increase in the likelihood of enrolling in a public college for both types of students, but even more for *typical* students who were predicted to go to college

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

[4]See the work of Flor-Toro and Magnaricotte (2021) who document the disparities among the admissions systems for both private and public in Peru.

regardless of the policy.

Crucially, this change in college choice is significantly different by gender. For the sample of *new students* induced by the policy, I find that females are more likely to attend a low-quality private college, and males, in contrast, increase their likelihood of enrolling in public colleges. On the other hand, for the sample of *typical students* who were likely to attend college, I find that females show no effects in terms of their college of choice while males are more likely to attend public college. The implications of women choosing low-quality private colleges compared with men taking advantage of attending free public colleges are of interest given the large wage gender gaps in the Peruvian labor market.

Third, 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 private or public school, given the systematic differences in admission systems. For students enrolling in private colleges, I find men are willing to travel 55 percent more than women for one additional standard deviation in wage returns. The patterns for public college show less of a gap, where men are willing to travel 7 percent more than women on average. Overall, the willingness to travel is on average higher for private colleges than public ones but the gender disparities are larger for men and women who attend private schools. These differentials in willingness to travel align with the results I find in the reduced form that suggest that women are more likely to enroll in lower-quality colleges connected to the new stations in comparison with men.

Additionally, I explore medium-term results using the 2017 Peruvian Census, 7 years after the first opening. I find that access to improved public transportation increases the likelihood of college completion. I estimate a 12 percent increase in college completion among students living in affected areas. Using a cohort-exposure analysis relative to the year of high school graduation, I show that these effects increase in magnitude among those who were exposed to improved transit options for the longest. This implies a major reallocation of time and resources, given that most Peruvian college students tend to complete their education over more than five years. These effects are particularly high for women and low-income students. I also examine longer-term outcomes such as employment rates. My finding suggests that there is a positive effect on employment rates for those who enrolled in college measured in the 2017 Census. More importantly, these effects are coming from students who enrolled in college finding employment in white-collar jobs.

This paper contributes to several strands of the literature. First, it contributes to the literature on the economics of transportation where most of this literature regarding the economic impact of improving or building new transportation in large cities in the development world shows positive effects on economic activity and labor market opportunities (Balboni et al., 2021; Tsivanidis, 2022; Zarate, 2022). [5] Nevertheless, less is known about the direct causal effects on human capital investment. In this paper, I document whether neighborhoods see a decrease in commuting time to *any* college in the city and how this directly affects college enrollment rates.[6]

Given the lack of data and identification challenges, there is little evidence on how city transportation policies affect college access. Current evidence in developing countries has shown positive effects on women's access to basic education as well as improving schooling outcomes and aspirations (Fiala et al., 2022; Muralidharan and Prakash, 2017)[7]. What is

[5] Currently, there is evidence on different types of transportation such as railroads and roads (Brooks and Donovan, 2020; Donaldson, 2018; Donaldson and Hornbeck, 2016) which shows positive effects on trade and economic growth. However, Severen (2021) finds no impact on local productivity or amenities in Los Angeles Metro Rail. There is also evidence on buses or BRTs (Balboni et al., 2021; Tsivanidis, 2022) and metros (Zarate, 2022) showing positive results for commuters. In the case of Peru, Velásquez (2023) finds no effects on effects in rents, income, poverty, and household expenditures for the opening of the new BRT line in Lima.

[6] It is also worth highlighting that there is not only a reduction in transportation costs but the new system is also providing 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 graduation. 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 and therefore, urban areas that offer higher returns to education. The results of this paper show reduced-form results that combine both channels: increased labor market opportunities (both for current jobs and after graduation) and reduced commuting times.

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

more, a decrease in transportation costs can reduce gaps between low-skill and high-skill students Asahi and Pinto (2022), as families who experience reduced transportation costs travel further and enroll in higher-quality schools Herskovic (2020). Not surprisingly, in certain contexts, where access to education is based on meritocratic systems, only high-achieving students with highly-educated parents take advantage of the reduction in transportation costs (Dustan and Ngo, 2018). My results show that living in a neighborhood that gets connected to new lines affects not only college enrollment rates but also can change *which* college students enroll in. In particular, I show differential responses by gender. Male students become more likely to enroll in public colleges while female students become more likely to enroll in low-quality private colleges since these institutions get connected to the new lines as well.

Another important potential mechanism related to transportation costs is travel safety. In Lima, informal transportation is seen as risky due to the high incidence of accidents, sexual harassment, and fear of muggings (Dominguez Gonzalez et al., 2020). This scenario is similar to others in large cities in the developing world, where women face harassment and crime at significantly higher rates than men, and this can yield differential responses by gender when travel costs change. There is a small but growing literature on this matter, where most results show that women have a higher demand for safe transportation and this can directly affect their labor supply (Field and Vyborny, 2022; Kondylis et al., 2020).[8] Borker (2020) explores how the perceived risk of street harassment can help explain women's college choices in Delhi. She finds that women are willing to choose a low-quality college over a top college that is perceived to be one standard deviation safer. In this paper, I focus on Lima, a city considered one of the worst large cities in the world for women's mobility and transportation, comparable to Delhi, Mexico City, and Jakarta.[9] My results are consistent with previous literature given that the new transportation systems provided a safer ride compared to the informal buses

[8]In Lima, the introduction of the new stations had a positive impact on women's labor supply (Martinez et al., 2020).

[9]See the ranking on women's safety in transportation in a Reuters study here. As documented by Sviatschi and Trako (2021), Peru is a country that has experienced a large increase in gender violence, where the number of domestic violence cases registered in local police departments has increased substantially: from 29,759 in 2002 to more than 60,000 in 2016.

that circulate the city. I find that female students are getting most of the benefits: they do not only access college at a higher rate than men but they are also entirely driving the effects for college completion. However, similar to Borker (2020), women are more likely to enroll in lesser-quality private colleges since these colleges are the ones getting connected to the new stations, minimizing their exposure to harassment and facilitating their commute.

A final contribution of this paper is to the literature on the geography of inequality and place-based policies. Prominent work has shown how growing up in places with more opportunities can positively impact income mobility, especially for economically disadvantaged populations and minorities (Chetty et al., 2020; Chetty and Hendren, 2018). What is more, accessing better schools can have positive effects on attending college (Bergman, 2018). In this project, I explore how transportation can create a way to commute to opportunities by increasing access to college for new students and subsequently, accessing higher-quality jobs. My results also suggest that these gains are mostly for women and low-income families.[10] However, I also show that this can be limited by the supply of colleges in the city. Most students prefer to commute less and access lower-quality colleges connected to the new stations, especially women. This has direct implications for future labor market outcomes as they are likely to receive less returns than their peers.

The remainder of this paper is organized as follows. In section 2, I describe the background and in section 3, I describe the data. In section 4, I present the empirical strategy. Section 5 discusses the reduced-form 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. //

[10] In a related work, Meneses (2022) studies how new subway lines in Santiago de Chile yield positive effects on intergenerational mobility given that families are able to attend better schools and subsequently access higher quality college-majors.

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 3 levels: primary education (6 years), secondary education (5 years), and the higher education level which often lasts from 2 (technical school) to 10 years (School of Medicine). On average, college students take between 5 and 7 years to graduate. According to the 2017 Peruvian Census, approximately 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 percent enroll in a technical school or community college, 22 percent enroll in college while the remaining 63 percent do not access any type of higher education (Alba-Vivar et al., 2020). Following similar trends to middle and high-income countries, Peruvian women access college slightly higher than men. However, Peruvian students face similar struggles to other college students in the developing world. Among those who access higher education, a significant amount of students work and study at the same time (28 percent of technical school students and 20 percent of college students).

Some additional aspects of the college education system are worth highlighting and are relevant to understanding students' college choices. Typically, a student faces a decision tree as depicted in Figure 1. First, a student might choose to attend either college, community college or choose not to get any higher education and keep their high school diploma. Second, if a student chooses to attend college, they might choose to attend either private or public colleges. There are several differences between private and public colleges and in their admission process. On one hand, public institutions offer free tuition, have a decentralized admission system (i.e. each university organizes its own admission exam), and 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 this 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. In general, their admission process is very straightforward and students face less or no uncertainty about their admission process: admission exams, if they exist, are typically not binding. Only a few elite private colleges tend to have more selective admission exams. For simplification, I am assuming two types of private institutions: high-quality and low-quality colleges. High-quality colleges tend to charge higher tuition fees in comparison with low-quality colleges (the price range goes from 150 USD monthly up to 1300 USD, while the minimum monthly wage was 180 USD in 2010). For private universities, prestige is correlated with prices given the selection of its students. Graduates from low-cost and low-quality colleges enjoy fewer returns in the labor market in comparison with their peers in high-cost and high-quality institutions.

Unlike the US, there are no general standardized exams like the SAT which makes it difficult to assess students' abilities. Also, there is no centralized admission system to enroll in college and students face a more complicated process as each college maintains its own distinct procedure to admit students. This system incentivizes students to apply only to their most preferred and feasible choices. Another relevant fact is that 90 percent of college students attend a college located in their province of birth, suggesting that most students do not move to go to college. Hence, college housing is almost nonexistent and if it does, they are reserved exclusively for out-of-state students. In Lima, 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 the 2010 University Census data. 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 a reduction in transportation costs can switch not only their decisions to attend college but also *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 or Los Angeles Metropolitan Area. However, Lima is not nearly as dense (8,000 hab/m^2, compared to NYC's 29,302 hab/m^2), and commuting across the city can take up to 3 hours during rush hour. During the 90s, market liberation policies facilitated the import of used cars and mini-buses which became the basis of the new transportation system for commuters. These privately operated minibuses, known as *combis*, partially alleviated the demand for transportation across the city. However, *combis'* 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 percent of people who suffered a traffic accident were under 25. A regional initiative to implement a bus rapid transit line that connected the north and south of the city was opened to service in July 2010. The *Metropolitano* was the very first mass transportation public system in Peru, connecting 12 districts out of the 44 in the city. There was a flat fee of 1.50 PEN, approximately 0.50 USD for regular commuters but college students had a 50 percent discount.[11] A year after the *Metropolitano*'s opening, 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 connected over 2 million people.

The *Metro de Lima* was a project that took almost 40 years. During the 70s, the Peruvian Ministry of Transport had 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 8 lines but, as years passed, only 6 of them were properly studied and

[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 6 years until it was overtaken by the Wuhan Metro in 2017.

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 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 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 80s, the first line was rushed to open before it was completed, and 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 is still under construction since 2014. These delays are mostly explained due to several corruption scandals involving Garcia's government and Odebrecht, the consortium in charge of the construction of the first line, which allegedly paid more than 20 USD millions in bribes for this project. Appendix C compiles relevant information about the history of these projects.

In this paper, I focus on the first line of the *Metro de Lima* and the *Metropolitano* new stations. The new bus line and the metro line provided a faster and safer service compared with traditional informal minibuses *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 relies on multiple sources of data that include administrative data from college records, geocoded stations, and census data at the block level.

College Enrollment: This information was provided by the Peruvian Ministry of Education, which annually compiles enrollment data for every college, both public and private ones. These records contain information about students' year of enrollment, college, home addresses, declared 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 used the Google Maps API to collect GPS coordinates for their homes. For less than 5 percent of the total cases, where the algorithm failed, I imputed GPS coordinates at the block or neighborhood level.[13] Additionally, I refined 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. The locations of 44 college campuses in Lima's metro area were manually collected and geocoded. The addresses were obtained from the 2010 College Census compiled by the Ministry of Education and the National Institute of Statistics and Informatics (INEI), verified using Google Maps API. The resulting GPS coordinates are plotted in red dots in Figure 2.

Peruvian Census. I obtained the Peruvian Census from 2007 and 2017 from 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 Data. I obtained all the information on stations from the new transportation systems from the *Autoridad de Transporte Urbano para Lima y Callao* (ATU). This included the GPS location and address of all stations. I also collected information on planned but non-executed lines from the *Metro de Lima*. This information comes from multiple technical records from the national government (Ministry of Transport), details about

[13]The home address is self-declared by students at age 18 when they obtain their national ID, and 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. Also, in 2016, the Ministry of Education started a licensing process 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 potentially could have change the way students make their college decisions.

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

this information are explained in Appendix C. I geocoded all planned stations from 6 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. First, I 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 use the road 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, then I complement it with the Google Maps API data to obtain primary and secondary highway speeds, with this information, I computed 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 dataset is a panel spanning from January 2014 to November 2019, and it includes monthly labor market outcomes such as wages, and hours work, etc.[18] It also includes college information such as major, gender, and college. I restrict my sample to students who graduated in 2014 and 2015. I 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 station's openings. Panel A shows the college enrollment rates using 2 measures: (1) Enrollment counts are the average number of students who enroll in any college in the city at the block level. As seen in the table, less than 1 student per block enrolls in college in my sample. When dividing by the denominator of population counts of the same age, we can see that on average 17 percent of students under 19 years old enroll in college. To be specific, the

[16] This data is publicly available, I use the package *osmnx* available on Python.

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

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

college enrollment rates are defined as the following:$Access_{it}^{College} = \frac{TotalEnroll_{it}^{16-19}}{TotalPop_{it}^{16-19}}$. I also use the counting of students enrolled in logs (or the inverse hyperbolic sine function to correct for the presence of zeros). It is also clear that, on average, women enroll in college at higher rates than men and that private college enrollment is higher than public college enrollment. Importantly, we can also observe that the distance from students home to college is on average quite similar for both neighborhoods connected to new lines and those connected to the planned but not-executed lines. Panel B in Table 1 shows average statistics using the 2007 Census, which includes the total population and education levels achieved for people 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 years old seems to be slightly more educated in the treatment group.

4 Empirical Strategy

4.0.1 Main Outcomes

This paper uses a Difference-in-Differences (DiD) strategy that exploits neighborhood exposure to new lines as well as variation across student cohorts. Additionally, I use a flexible event study framework to account for dynamic and heterogeneous treatment effects. The specification is the following:

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

Y_{it} represents the outcome of interest, such as the college enrollment rate, 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 following the opening of new stations, 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. Additionally, μ_i are the block fixed effects and ψ_t are the year fixed effects.

However, there are a few additional empirical challenges 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, this study relies on the recent advances of the DiD literature.[19] In particular, I follow Borusyak et al. (2023) and implement their imputation estimator which allows for treatment-effect heterogeneity and dynamic effects.

A second empirical challenge is to establish a proper control group of never-treated neighborhoods. In this sense, simply comparing connected neighborhoods to non-connected neighborhoods within the city might overestimate our results since the allocation of the new routes is not completely random. For example, poor people in the city live in remote areas where it is unlikely they could implement a new line. 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 because there were planned lines that had not been constructed yet.

Figure 3 shows the neighborhood that belongs to the treatment and never-treated groups. I define a neighborhood that is exposed to the executed lines as one that is within 1.5 kilometers (about 20 minutes walking) of the nearest station as seen in Figure 3. The

[19] Several papers have shown 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 the work of Borusyak et al. (2023); Callaway and Sant'Anna (2021); de Chaisemartin and D'Haultfœuille (2020); Sun and Abraham (2020).

never-treated group comes from those neighborhoods that are within 1.5 kilometers of the planned but not executed station. I also excluded neighborhoods that are simultaneously exposed to opened and planned but not executed lines as seen on the yellow shaded areas in 3. Restricting the never-treated neighborhoods connected to planned but not-executed lines reduces the selection bias due to a potential correlation between the new lines' placement and unobserved changes in access to college.

Third, even when using the placebo lines to reduce selection bias, another potential issue can arise in this setting. When estimating the effects of these new transportation lines, using the timing of the opening of the station as an exogenous shock, other determinants of the outcome of interest (college enrollment rates) are still not random, recent work by Borusyak and Hull (2023) highlights this issue. For example, neighborhoods located in the city center (which is both economically and geographically central) might experience higher enrollment growth since they were more likely to have a new line than those in the outskirts. Families living in these areas are substantially more educated, which is a strong predictor of college enrollment (see Appendix Figure A.6).[20] To avoid issues with neighborhoods that were going to receive the treatment regarding which line opens, I improve my main sample by excluding such neighborhoods located in Downtown Lima.[21]

4.0.2 Medium and long-term outcomes

I look at neighborhoods that were exposed to the new lines and use the administrative data and the latest Census available (2017) to obtain estimates at the block level. First, using the administrative enrollment data and census counts, I calculate yearly block-level college enrollment rates. The denominator comes from the total population counts from the 2007 Census. For this, I estimate Equation 1. Second, using individual-level data from the 2017 Census, I calculate age cohort completion rates and labor market outcomes at the block level.

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

[21] 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.

I estimate an exposure DiD as in Equation 2. Age cohorts will be 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. I also restrict the analysis to those individuals born in the period 1991 to 2000.[22]

$$y_{c,i} = \sum_{\tau=-4}^{0} \alpha_\tau D_i^{pre} \mathbb{1}(\tau = c - T^*) + \sum_{\tau=1}^{4} \phi_\tau D_i^{post} \mathbb{1}(\tau = c - T^*) + X\beta_{c,i} + \psi_c + \mu_i + e_{c,i} \quad (2)$$

4.0.3 Identifying access to college and college choice

?? I expect that a reduction in transportation costs will not only change access to college but given that *some* colleges get connected and become more attractive, students might change their college choice. In an ideal setting, the researcher could have information about students' abilities and preferences for college, both crucial information that could help identify students who are likely to go to college and *which* college they are attending.

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, in the case of Peru, and similarly to most countries around the globe, there is a decentralized admission system, where each college has its own admission procedures. Furthermore, there is no standardized testing like the SATs, meaning that there is no measure of a student's ability when applying to college.

To tackle this limitation, I use *machine learning* as a second-best tool to recover students' choices. First, I 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. I use the k-nearest neighbors algorithm in order to predict the probability of attending college for each neighborhood before the new lines open to the public, meaning that I use pre-treatment information. I further refined the algorithm using a k-fold

[22]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 by 2017.

validation strategy and chose an optimal k parameter using *GridSearch*.[23] 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.[24] *High-quality* private colleges are those whose graduates in 2014 earned more than 2250 PEN (25 percent top of the wage distribution) or approximately twice the minimum monthly wage while *low-quality* are those whose graduates earned less than 1450 PEN (25 percent bottom of the wage distribution). Appendix Figure A.4 shows the distribution of college wage returns and the thresholds used.

5 The impact of transportation on college education

In this section, I explore the impact of the new lines on access to college and college choice. As a first step, I also show the results of a reduction in commuting time to college in the next subsection.

5.1 Effects on Commuting Time to College

I investigate whether the new stations reduced 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 different-in-difference 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.[25]. This estimate can be considered a lower bound and very conservative since informal routes, such as *combis*, are not included in the data, and I assume that students are commuting by car, which is an overestimate of their actual

[23] Details are described in Appendix D.

[24] 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 the ones who 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). Appendix Figure A.3 shows the distribution of the colleges that were closed in the city, and most of them are located in Downtown Lima and connected to the new stations.

[25] The before and after variation can be visualized in Figure A.5. Notably, people living in the northeast of the city seem to have the most benefits of the new transportation system.

transportation time. The average commuting time to any college in these neighborhoods, prior to the new systems being implemented, is estimated to be around 1 hour. This aligns with the self-reported data from the College Census of 2010, as illustrated in Figure A.1 which shows a similar number. The results in Table 2 suggest that the introduction of the new system reduced the average commuting time to *any* college in the city by 17 percent (13 minutes reduction relative to 66 minutes baseline per trip), which translates to almost half an hour per day saved on commuting. The results for private colleges are even higher, a 20 percent reduction in commuting time in contrast with a 14 percent reduction in commuting time to public colleges relative to baseline. It is important to 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, I estimate the post-treatment results from Equation 1 until after 4 years of the opening of new lines. I find positive effects on college enrollment rates at the block level as seen in Table 3, Column (1). The coefficient shows a 1 p.p. increase relative to a baseline 14 percent enrollment (pre-treatment)[26], measured as the number of students enrolled in a block divided by the counts of people of the same age living in the same block. Column (2) shows similar results when enrollment is measured in logarithms or to be specific, the hyperbolic sine function. 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 and closer to zero. This is in accordance with the fact that public colleges in Peru tend to have higher competitiveness compared to private colleges. Consequently, decreasing transportation expenses may have a limited influence on enrollment, as students' abilities carry

[26]The sample for this estimate is restricted to recent high school graduates, up to 19 years old.

greater significance in their admission and student's decision-making process. Conversely, private universities not only offer more streamlined admission processes but can also swiftly adapt to accommodate an influx of new students. Columns (5) and (6) show that there is 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 the opening, and the magnitude of the effects doubles up to a 2.5*p.p.* increase after three years of the new station openings (See Figure 4a). Additionally, the pre-treatment coefficients validate our findings as they demonstrate no prior trends before the implementation of the new lines.

Examining the dynamic effects within subgroups 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 females intensify more rapidly over time in contrast to those on males, as illustrated in Figure 5b. This suggests that it took a couple of years for women to fully harness the benefits of the new system. It also underscores that the advantages of the new systems extend beyond merely reducing transportation costs to enhancing travel safety, which is especially crucial for women who are particularly vulnerable in this city.

5.2.1 Robustness Checks

Dynamic Effects. Figure 4a shows the event study results using the Borusyak and Hull (2023) estimator. First, this figure shows that there are no significant trends in the pre-treatment period. I additionally test another functional form, using the hyperbolic sine transformation. Figure 4b shows that the results are similar. I also test for a different estimator than Borusyak and Hull (2023), following the recent advances in the DiD literature. The results using the Callaway and Sant'Anna (2021) estimator, as seen in Appendix Figure

A.7, show similar results.

Sample Including Downtown Lima. In Section 4, I highlighted that the main specification excludes districts in Downtown Lima since these areas of the city were very likely to get treated by any *potential* transportation line. What is more, this part of the city has a higher level of market access in comparison with the outskirts. Similar to the results in Table 3, adding the districts in this central area of the city does not change the results significantly as seen in Table A.2. I still find positive effects on college enrollment rates at the block level (1*p.p.* increase). The results from Columns 2-6 reveal a similar pattern to the main results in Table 3 but in a slightly lower magnitude, suggesting that the neighborhoods in Downtown Lima experienced almost no impact.

5.3 Effects on College Choice

It is expected that the effects of the new lines are not only on the extensive margin (access to college) but also on the intensive margin (college choice). This is mainly motivated by the previous results suggesting that students are more likely to enroll in private institutions and by the fact that there is an additional variation based on *which* college gets connected to the new lines. Appendix Figure A.8 shows these patterns over time which suggests that colleges connected to the 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. In an ideal setting, students' ability and college preferences are known to the researcher[27]. However, in this context and like many others in education systems in the world, this information is unknown.

To deal with this limitation and to take advantage of the big data I collected and use machine learning techniques (described in Section **??** and Appendix D with more detail) to predict students' college choices using a non-parametric algorithm, k-nearest neighborhood. An advantage of using this tool is that allows me to exploit the Census's big dataset and

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

allows for non-linear relationships, unlike other similar methods. I feed this algorithm with all the data available before the policy takes place, and then I obtain the predicted probability θ_i^m that a student i will go to college type m, where $m = [HighQuality, LowQuality, Public]$, 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. I also estimated the unconditional probability of enrolling in a STEM major and whether a student is likely to enroll in college or not. The latter will help me identify if the results are coming from new students induced by the policy (*new* students sample) or by students who were going to college regardless of the policy (*typical* students sample) but are now switching their choices. 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 Ten policy given data limitations that prevented them to directly identify eligibility. In my case, I use it to explore whether students who were 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 by a statistically significant 1.3 percentage point. This represents a 12 percent reduction in the probability of enrolling in high-quality colleges relative to the baseline. In contrast, students become more likely to be enrolled in a low-quality college by 2.6 percentage points (16 percent increase relative to baseline). When examining the likelihood of enrolling in a public college, I also find positive results. Column (3) shows an increase of 3 percentage points in the likelihood of enrolling in public colleges (7 percent increase relative to baseline). Additionally, I estimate the impact on the likelihood of enrolling in a STEM major, but no statistically significant effects are observed as seen in Column (4).

Second, I investigate the extent to which the previous results are driven by new students who are induced by the policy (*new* students sample) or students who had high chances of attending college (*typical* students) regardless of the policy. Figure 6 shows that the negative effects on the likelihood of enrolling in a high-quality college and the increase in the likelihood of enrolling in a low-quality college come from the *new* students sample. However, the increase in the likelihood of enrolling in a public college comes from both types of students but mostly those in the *typical* students sample.

Third, I also explore the effects on the likelihood that a student enrolls in a college based on how high or low probability θ they have of enrolling in college type m. I assess the heterogeneous impacts by quintiles of θ, where θ^{q1} is the lowest quintile and θ^{q5} is the highest quintile of the probability of enrolling in school m. Figure 7a shows the results for high-quality private colleges by quintile and also broken down by either the *new* or *typical* students sample. First, students who were very likely to go to college experienced no effects, even the ones who were very likely to go to a high-quality college. On the other hand, students induced by the policy seem to be driving all the effects, where students with the lowest probability (up to the third quintile) of enrolling in these institutions are responsible for the observed negative effects. Interestingly, students in the highest quintile experience positive effects suggesting that a reduction in transportation costs positively those who were likely to attend.

Conversely, Figure 7b shows that students with the least probability of enrolling in these institutions are the primary contributors to the observed positive effects, regardless of whether they come from the *typical* or *new* students sample. Both results suggest that marginal students (with the least probability of attending college type m) are the ones driving the effects. However, it is worth highlighting that students with a high probability of attending a low-quality college from the *typical* sample are less likely to enroll, suggesting that these students are trading off their choices towards higher-quality options.

Turning to public institutions, I observe that there are no significant differences between the *typical* and *new* students sample. The increase in the likelihood of enrolling comes from both, suggesting that not only students who were already likely to attend a public college are incentivized but also some of the one who were not likely to attend a public college as seen in Figure 7c. Regarding the probability of pursuing a STEM major, my findings indicate that students who initially had a low likelihood of enrolling in a STEM major are now even less inclined to do so, whereas those with a preexisting inclination towards STEM exhibit a positive effect (Figure 7d) and these effects are similar for the *typical* and *new* students sample. The effects on STEM are relevant since these majors are the ones that yield the highest wages after graduation, and the lack of effects on this outcome is explained by students re-enforcing their predicted choices.

5.3.1 Heterogeneous Effects by Gender and SES

Moreover, I explore whether these effects are heterogeneous for different populations. Figure 8 and Figure 9 illustrate these diverse impacts for different types of colleges by gender and socioeconomic status (SES). Figure 8a suggests that male and female students respond differently to the same shock of reduced transportation costs. However, the results are also different for the *typical* and *new* students sample. Figure 8a shows that females in the *typical* students sample show no effects in terms of their college of choice. However, females in the *new* students 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-quality college. On the other hand, males in the *typical* students sample are more likely to attend public college and less likely to attend low-quality colleges. Males in the *new* students sample also show a higher likelihood of enrolling in public college and are less likely to enroll in high-quality colleges. These divergent reactions carry significant implications for the gender gap, as male students appear to be gravitating towards more lucrative educational pathways than women.

The results by SES, where **poor** students are those students living in neighborhoods with households income below the median in the income distribution and **non-poor** are those above the median, show that income doesn't seem to explain college choice for students in the *typical* students sample as seen in Figure 9a. In contrast, the results for those in the *new* students sample suggest that poor students are the ones 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.

5.3.2 Robustness Checks

First, 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 that is 1 is a student is likely to enroll in a college type m relative to what is predicted at baseline (probability θ_m ¿ 0.5). The effects on Appendix Table A.3 show a similar pattern to my main results.

Second, I also test for an alternative definition of college quality, such as being an elite institution and being part of an elite college consortium[28] or if the college receives an operational license after the 2014 Higher Education Reform, and the results are similar showing that students are less likely to enroll in such institutions. Appendix Table A.1 shows the results for these alternative definitions.

5.4 Mechanisms

First, I explore whether the students who enroll in college are now opting for colleges that are located farther away. 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 like to enroll in colleges that are 9% closer to them. This suggests that students are not taking advantage of traveling longer distances within the city.

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

Second, 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 7 percent more likely to enroll in a college that is connected to the new lines as well. However, when breaking down these by whether the college is public or private, I find that students are 11 percent more likely to enroll in private colleges that get connected while I find no significant effects on the probability of enrolling in public college that gets connected to the new lines. These results are consistent with previous results that suggest that most effects on college enrollment rates are coming from enrollment in private institutions.

Additionally, I explore the heterogeneous effects by sub-groups for these estimates. Appendix Figure A.9 shows that all types of students are less likely to enroll in a college that is further away from their homes, contrary to what was expected. The effects are significantly driven by students living in poor neighborhoods. Panel B shows that all types of students except the ones living in poor neighborhoods are more likely to enroll in colleges connected to new lines. However, when looking at the results by whether the college is public or private, females are more likely to enroll in private colleges that get connected while men are more likely to enroll in public colleges get connected. These results confirm that effects on college choice are substantially different by gender. On one hand, women are choosing private colleges whose commute becomes easier (not only does their neighborhood get connected but also the college they choose is connected to new lines, reducing commuting times significantly) while men are taking advantage of enrolling in public colleges. Note that, among the few public colleges that get connected to the new lines is the National University of Engineering, where 80 percent of students enrolled are male.

6 Trade-off between College Quality and Travel Time to College

The reduced-form results for college choice suggest discernible gender-specific responses. Specifically, a reduction in travel time to college leads to a higher probability of women opting for low-quality private institutions, while men tend to lean towards enrollment in 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, key parameters to estimate are the student's willingness to pay for commuting time to school 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 [29] 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}

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

that 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

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

I am interested in measuring the trade-off between college quality and travel time to college. Similarly to Borker (2020), 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, as a proxy for college quality. 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 I also assume normally distributed coefficients for both T_{ic} and W_{ic}. This mixed logit model is a highly flexible model that can approximate any random utility model (McFadden and Train, 2000), as it relaxes the Independence of Irrelevant Alternatives (IIA) 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 we 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. On the other hand, selection into private universities is mostly explained by the household's ability to cover the tuition fees rather than ability. To facilitate interpretation, I standardize college returns (wage premium) to have a mean zero and a standard deviation of one. I measure travel time from home to college in minutes.

6.1 Results

Table 6 shows the estimated of 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 more commute and, in contrast, they are more likely to choose a college that provides higher wage returns. Column (2) shows the results for females enrolling in public colleges and similarly to Column (1), they dislike commuting to college. However, in this case, women are less likely to choose public colleges that provide higher returns. This not surprising for this context, where the public college that provides the highest wages in the market is the National University of Engineering (19 percent female enrollment in 2019) while among the lowest wages come from the National University of Education, a majority-female college. The results for male students are similar for both public and private colleges: they dislike commuting to college but even more to public ones and they are more likely to enroll in colleges that provide higher returns, especially private colleges.

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, it can be observed that students are less willing to travel to attend public colleges than private ones, and the gender differences are even more striking. On 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 one SD higher in returns. To be specific, men express a 55 percent higher willingness to travel compared to women measured in minutes per SD in returns. It will be expected then, that a reduction in commuting time will yield differential results by gender, as seen in Section 5.3, which suggests that with the opening of the new lines, men are less likely to attend low-quality colleges while women are induced to them. In this case, since most low-quality colleges get connected, they become relatively more attractive for women.

On the other hand, Figure 10b shows that the 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 express a 7 percent higher willingness to travel compared to women measured in minutes per SD in returns. With this, it is expected that a reduction in transportation costs have similar impacts for men and women. However, results in Section 5.3 show that only men take advantage of the reduction. Is it important to highlight that among the public colleges that get connected to the new lines is the National University of Engineering, which attracts mostly male students, while others like the National University of Education do not get 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, it is anticipated that the introduction of these new transportation lines will not only enhance college access but also 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 estimated the overall reduced-form effect using the 2017 Census (7 years after the opening) and a DiD model that exploits cohort-exposure variation. The results show a positive impact on college completion rates (12 percent) compared to baseline rates, as shown in Table 7 Column 1. The estimated coefficients are similar to whether Downtown Lima is included or not as seen in column 2. The dynamic effects on the event study are shown in Appendix Figure A.10, which shows that the more time a student was exposed to the new transportation lines, the higher the likelihood of her completing college by 2017.

Leveraging the comprehensive individual-level data available in the Census, I can delve into the varied impacts across different groups. Appendix 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 are getting negative effects. These types of students are the same who are enrolling in both public and low-quality colleges. Two effects could be happening at the same time: (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 ?? shows that there is a 10 percentage point increase in employment rates relative to the baseline. However, when breaking down these effects by job quality measured as blue-collar or white-collar jobs, I find that there is a 17 percentage point increase or a 6 percent increase that is statistically significant.

Figure 12 shows the heterogeneous effects of the impact on white-collar jobs by sub-groups. 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 women and the poor are attending low-quality institutions, they are still more likely to graduate and obtain a higher-quality job in comparison with peers who are not connected to the new lines.

8 Conclusion

This paper studies the link between urban features and higher education, and it is the first to show the causal effects of improved public transportation on both college access and college choice. While the prevailing body of educational literature has traditionally emphasized the influence of financial constraints and institutional factors on higher education access, limited attention has been devoted to understanding how the ramifications of reduced transportation costs for students can affect not only the decision to enroll in college but also in *which* college, and subsequently, how this affects their potential labor market outcomes.

In this paper, I focus on a higher level of education: college, which is becoming relevant in the developing world as access to basic education gets covered. The results of this paper also apply to big cities in both the developed world where access to college is still limited, especially access to *high quality* colleges for poor students. In a context where countries are actively revamping and reimagining their transportation systems, it is important to recognize that inefficient, unsafe, or unreliable public transportation can take a significant toll on the economy. Recent developments in the literature highlight the importance of having the most efficient transportation routes within urban areas. For instance, Kreindler et al. (2023), shows that adopting a less concentrated network can improve commuter well-being in cities like Jakarta. However, less is known about how 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 shed light 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. I find that a reduction of 17 percent in commuting time to any college available in the city can increase college enrollment rates by 1 p.p. in neighborhoods that get connected. 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 are choosing to travel less distances and are opting for colleges connected to the new system, even when these colleges are of lesser quality and this will be reflected in their future wages.

The findings of this study 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 schools of choice. Women appear to be more inclined towards choosing colleges of lower quality 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 find that men are willing to travel 50 percent more than women in order to attend a college that offers one standard deviation higher salaries upon graduation. 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.

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.

Bergman, P. (2018). The Risks and Benefits of School Integration for Participating Students: Evidence from a Randomized Desegregation Program.

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.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income Segregation and Intergenerational Mobility Across Colleges in the United States. *The Quarterly Journal of Economics*, 135(3):1567–1633.

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

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.

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.

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.

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.

Meneses, F. (2022). "Intergenerational Mobility After Expanding Educational Opportunities: A Quasi Experiment".

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.

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. (2021). Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification. *Accepted,The Review of Economics and Statistics*.

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

Sviatschi, M. and Trako, I. (2021). Gender Violence, Enforcement, and Human Capital: Evidence from Women's Justice Centers in Peru. Technical report.

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.

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

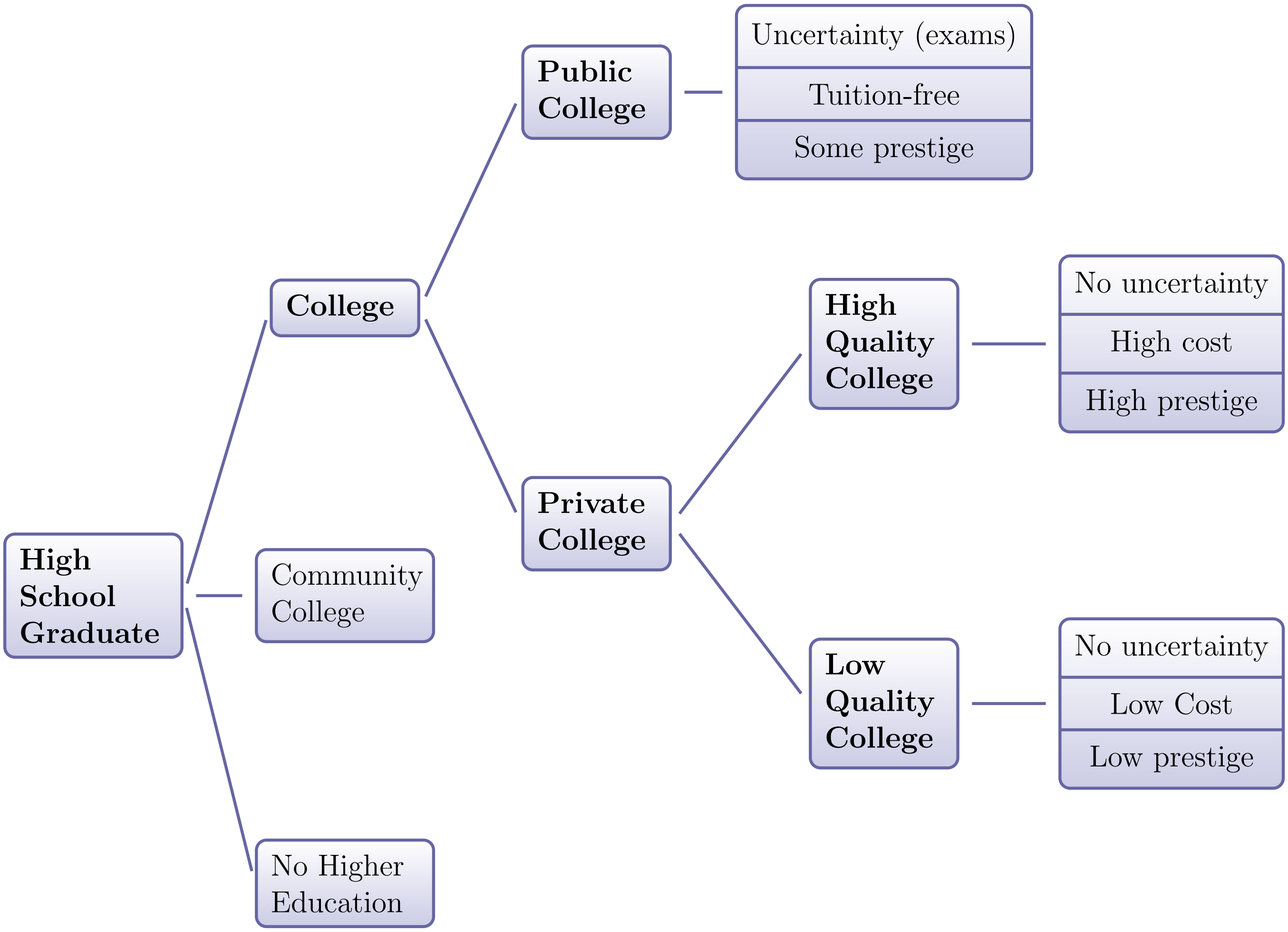
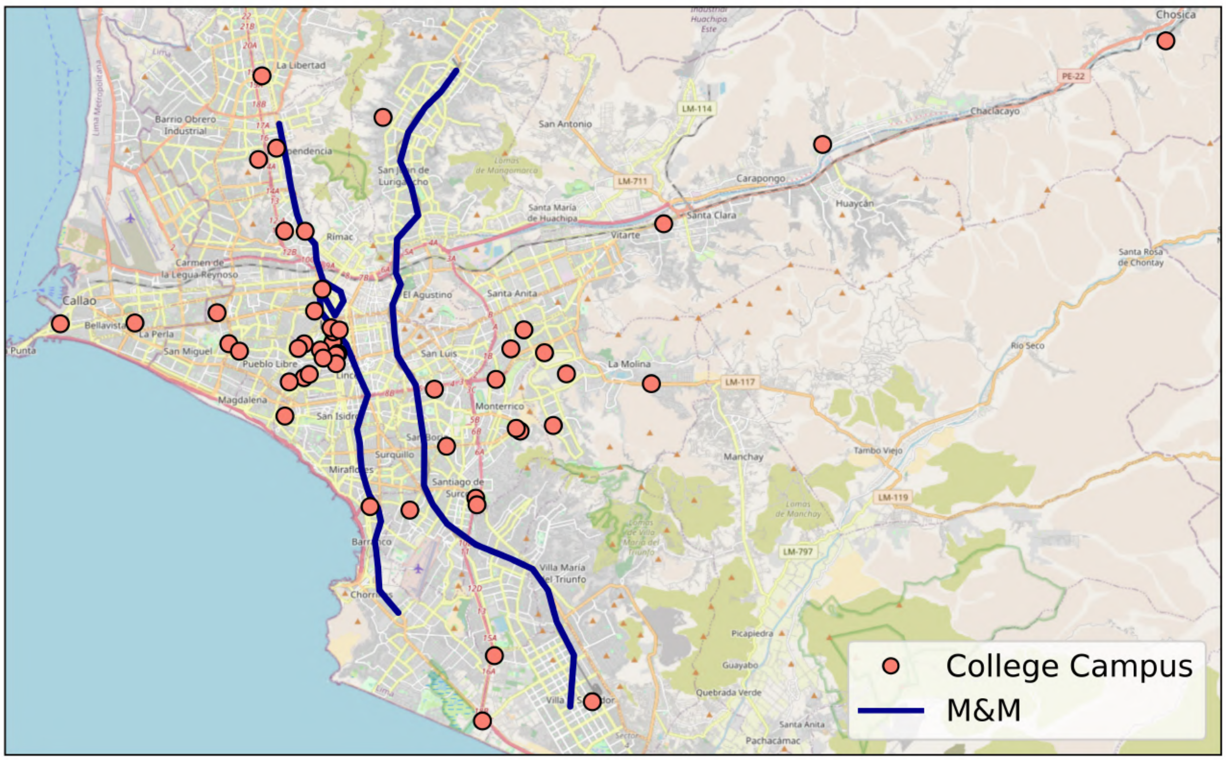
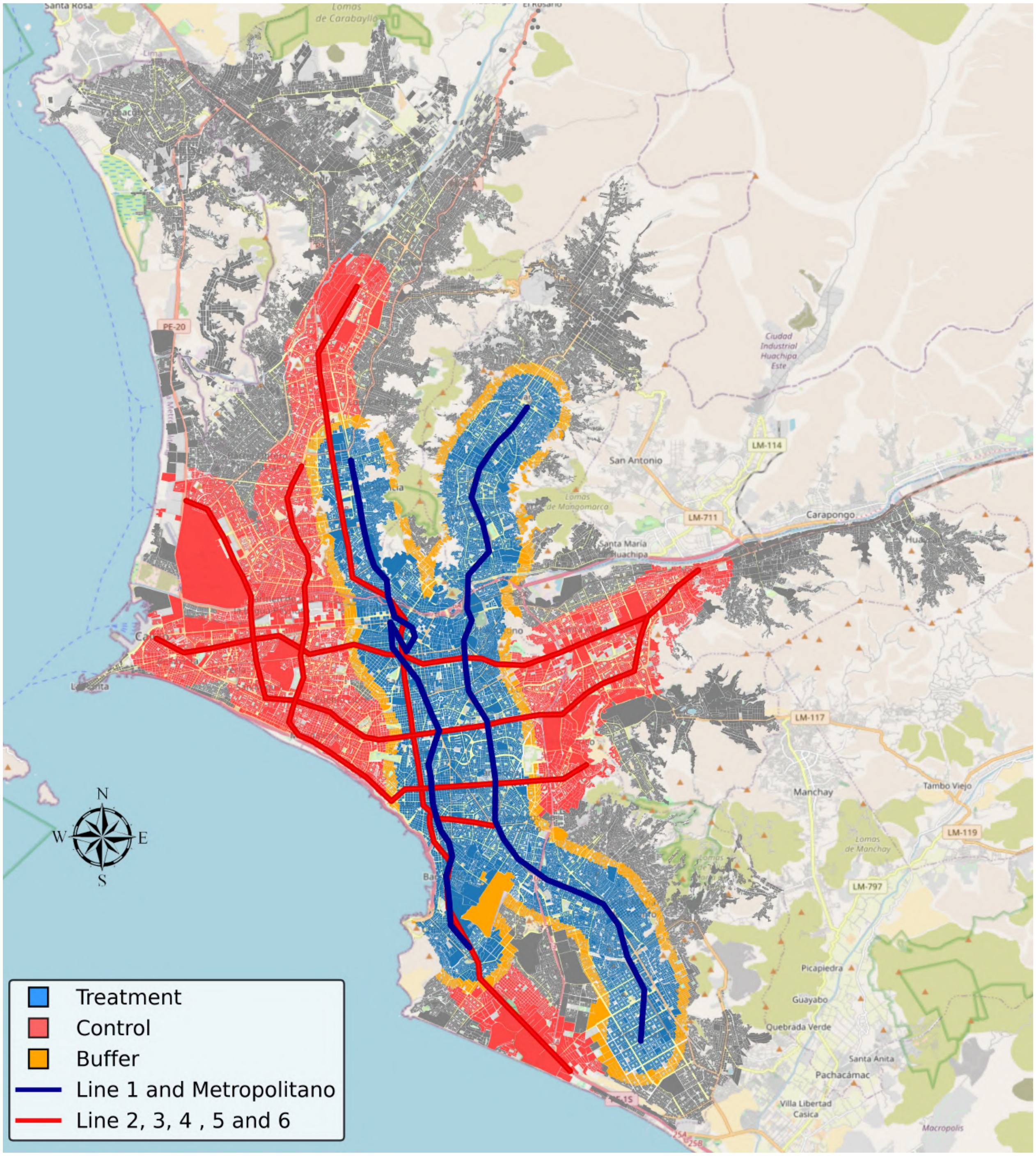


Figure 2: College Location and New Stations Across Lima



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.

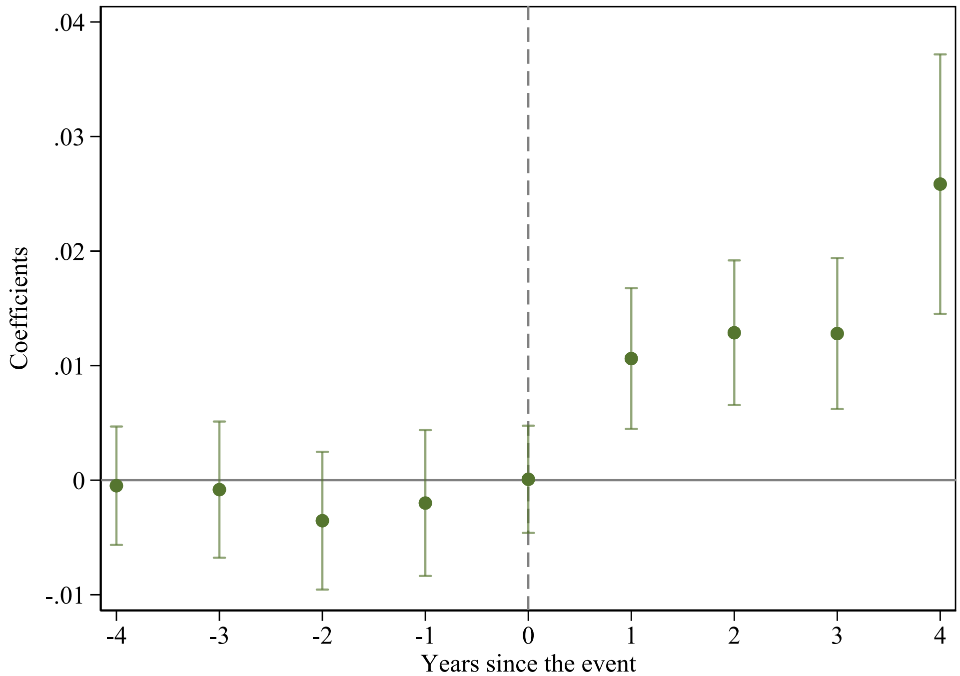
Figure 3: Neighborhoods Exposed to the Executed and Planned but Not-Executed Lines



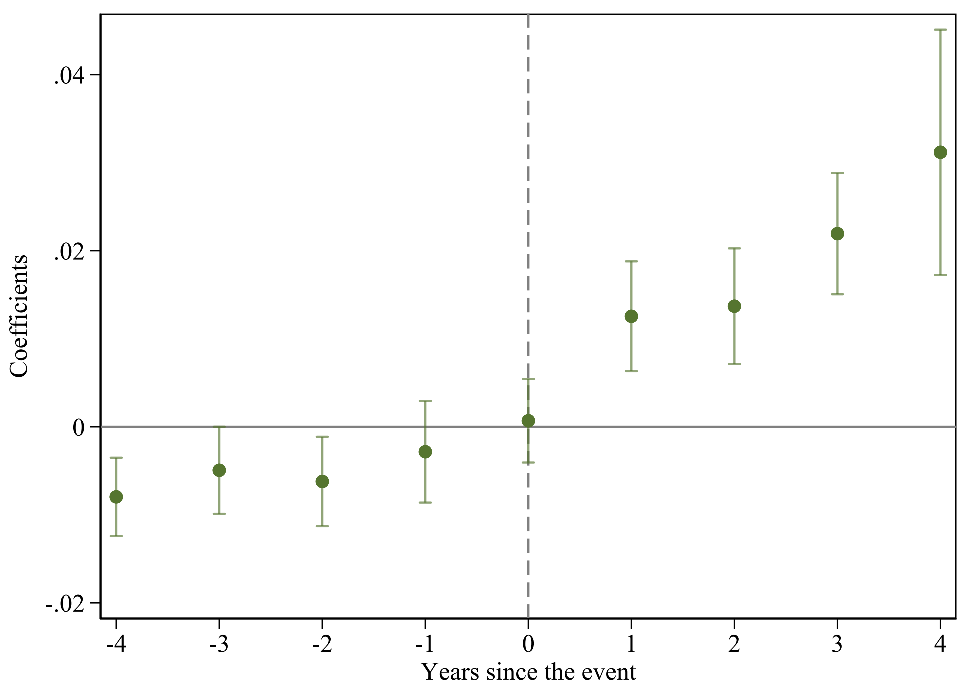
Notes. Blue-shaded areas are neighborhoods within 1.5 km distance from the nearest executed stations while red-shaded areas are neighborhoods within 1.5 km distance from the planned but not executed stations. Yellow-shaded areas are buffer zones that are excluded from the main sample.

Figure 4: Dynamic Effects on College Enrollment Rates

(a) Enrollment Rates



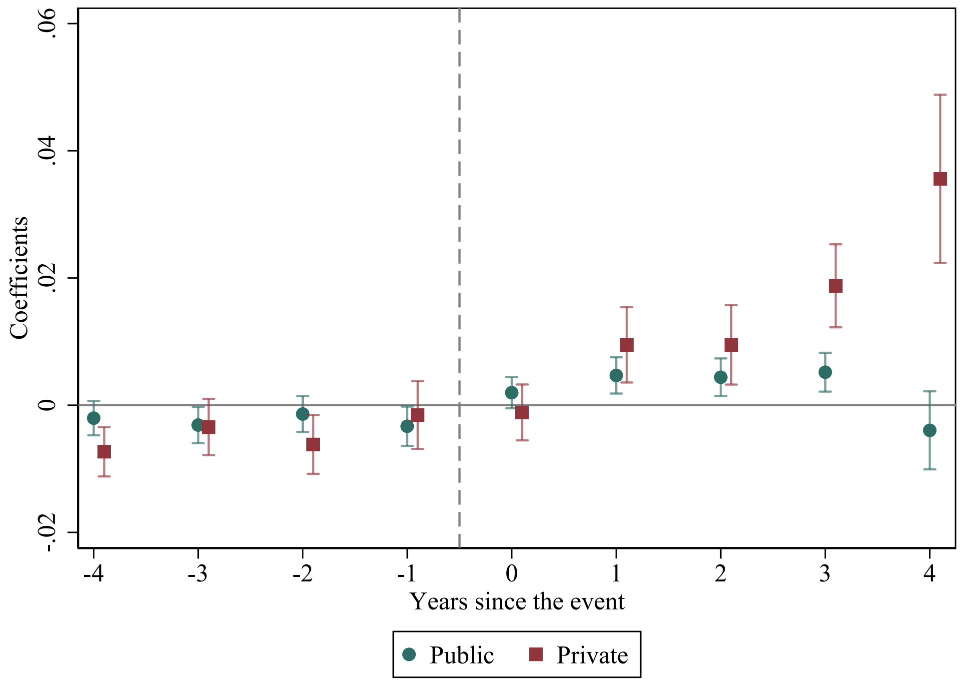
(b) Log(Enrollment Counts)



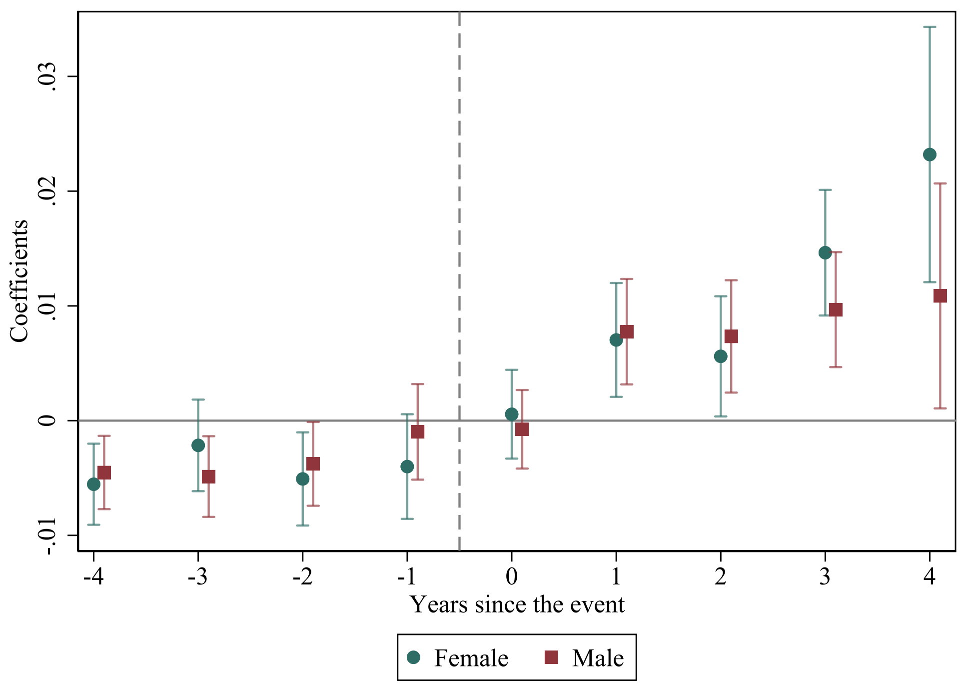
Notes. Regressions include block and cohort-fixed effects. The event study is calculated using the Borusyak et al. (2023) estimator. Panel (a) shows the enrollment rates where the denominator is the total count of students enrolled divided by the total population of (potential) students of the same cohort.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

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

(a) By College Type

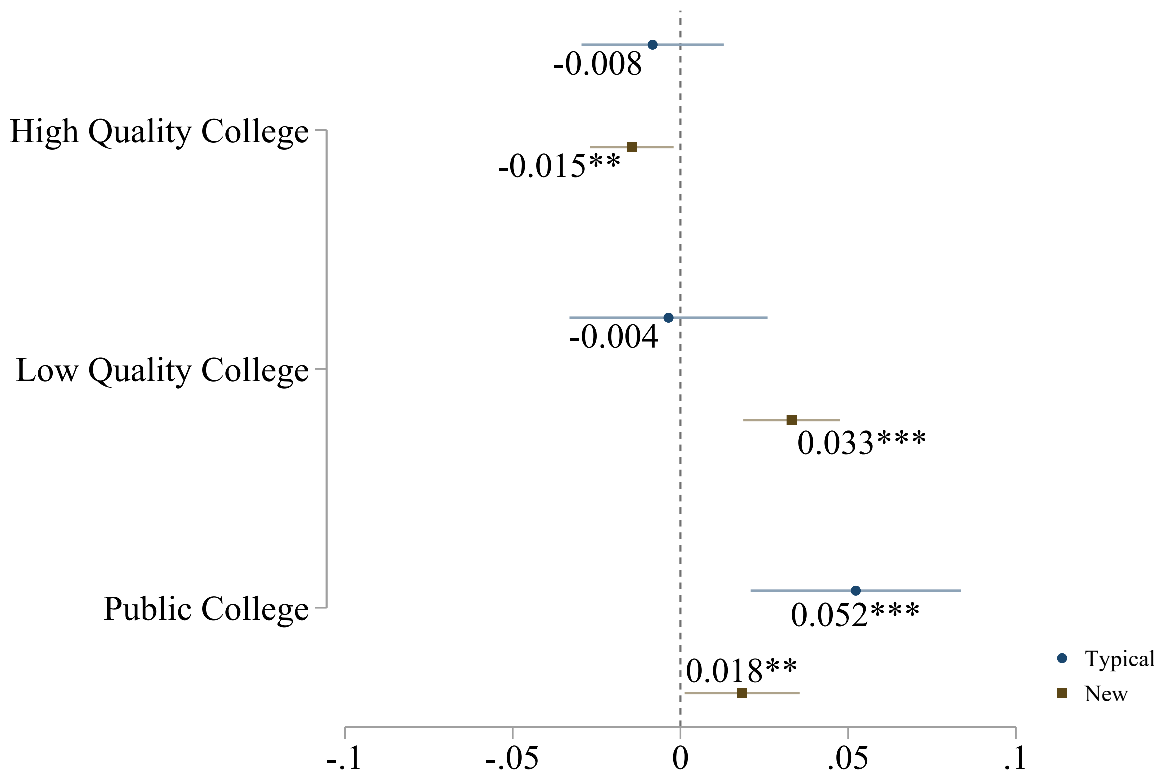


(b) By Gender



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 (*manzana*) and year fixed effects. High-quality Colleges are all private colleges whose graduates earn more than 2250 PEN while Low-quality private colleges are the ones whose graduates earn less than 1450 PEN using administrative data from wage records in 2014. *High Access* is the sample of students who come from neighborhoods with high enrollment rates while *Low Access* is the sample of students who come from neighborhoods with low enrollment rates.

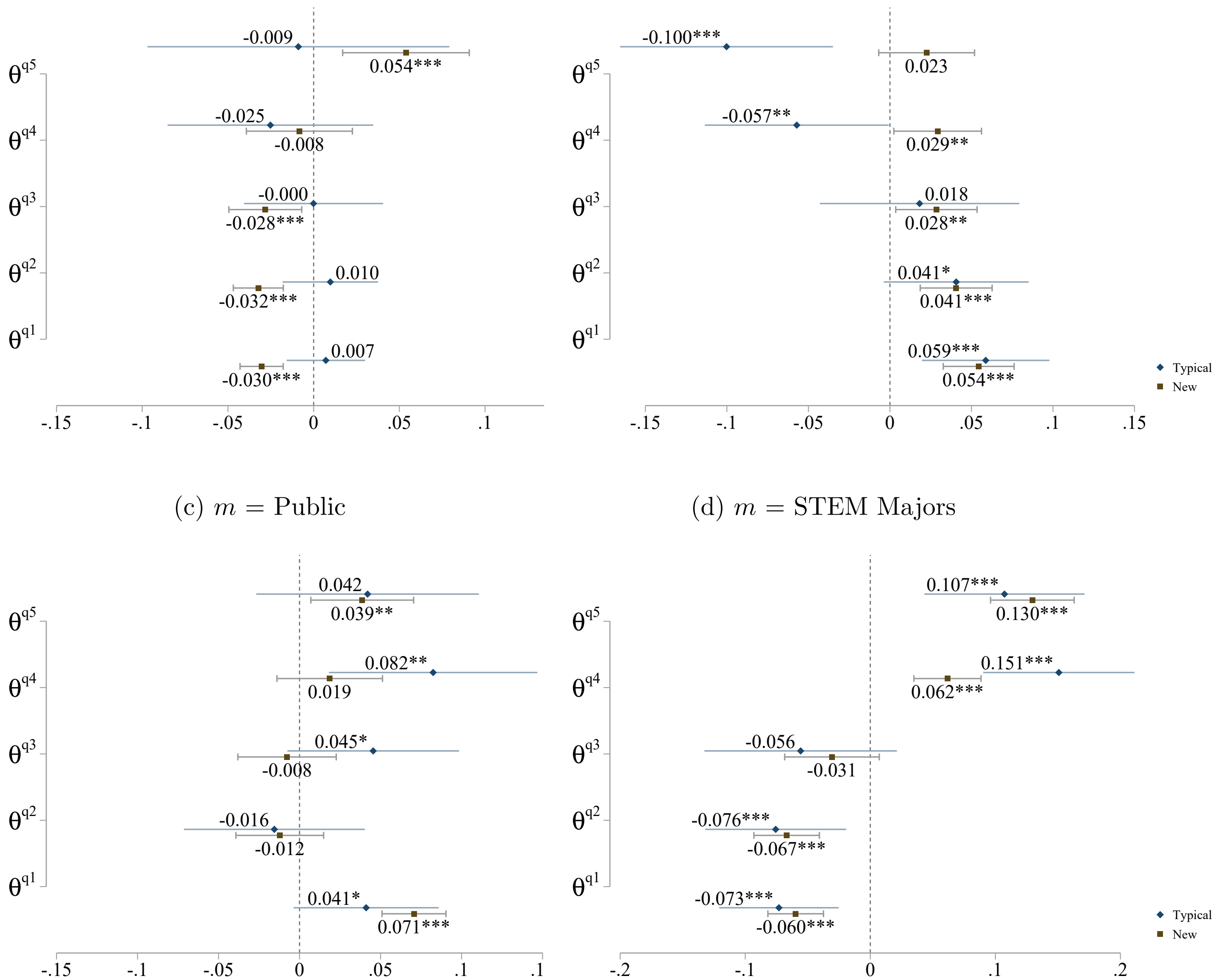
Figure 7: Heterogeneous Effects by Predicted Probability θ (quintiles) of Enrolling in College Type m by Access to College

(a) m = High-quality Private

(b) m = Low-quality Private

(c) m = Public

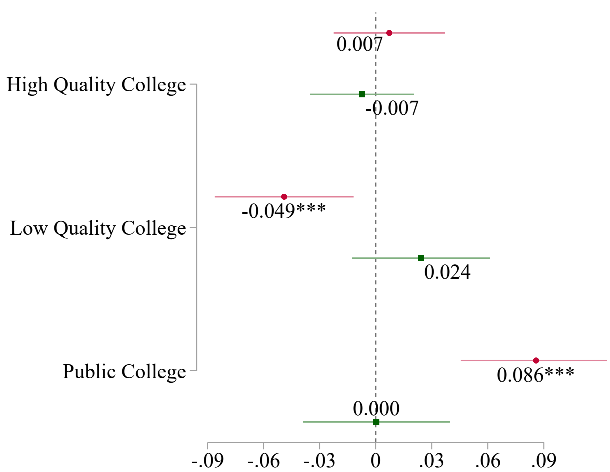
(d) m = STEM Majors

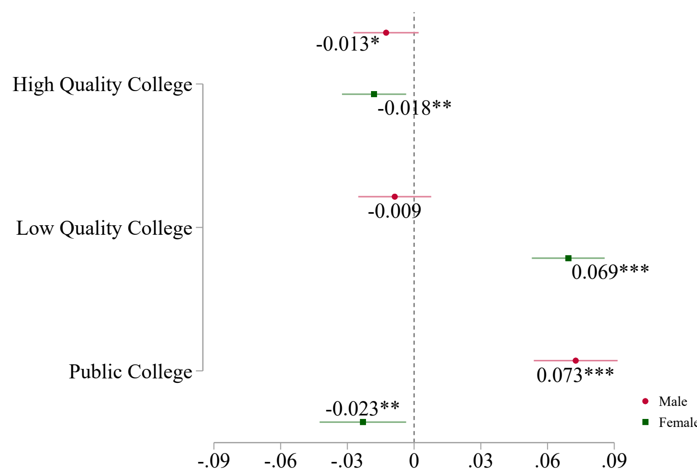


Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. High-quality colleges are all private colleges whose graduates earn more than 2250 PEN while low-quality private colleges are the ones whose graduates earn less than 1450 PEN using administrative data from wage records in 2014. θ^{qN} represents the quintile N of the predicted probability of enrolling in college type m.

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

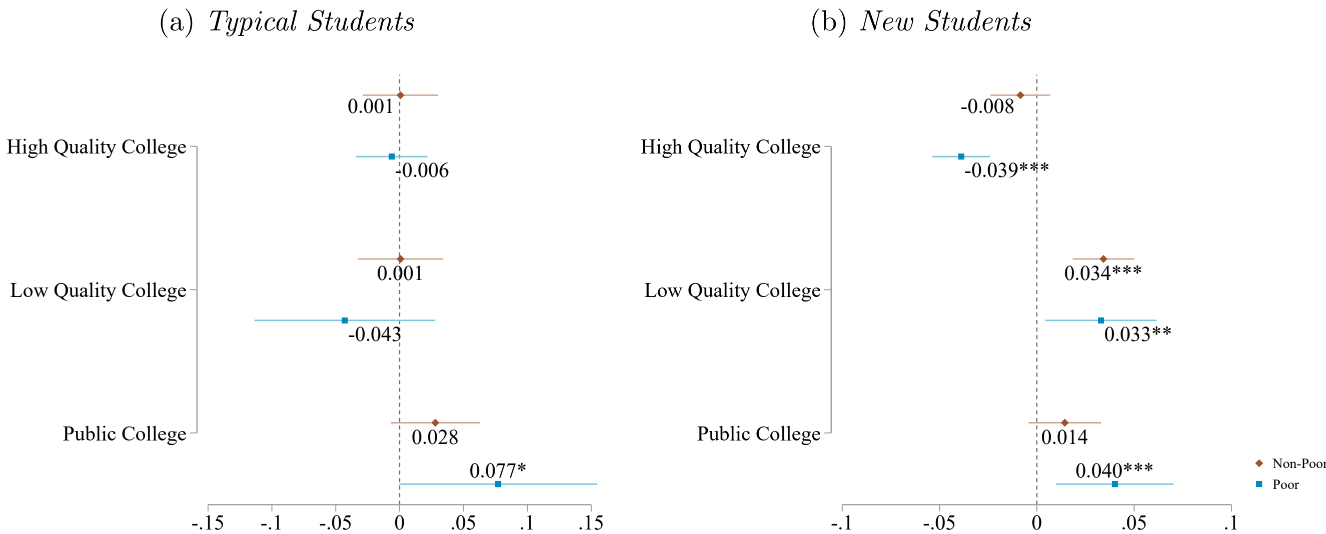
(a) *Typical Students* (b) *New Students*





Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and year fixed effects. High-quality colleges are all private colleges whose graduates earn more than 2250 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 Students* is the sample of students who were very likely to attend college while *New Students* are the ones induced by the policy.

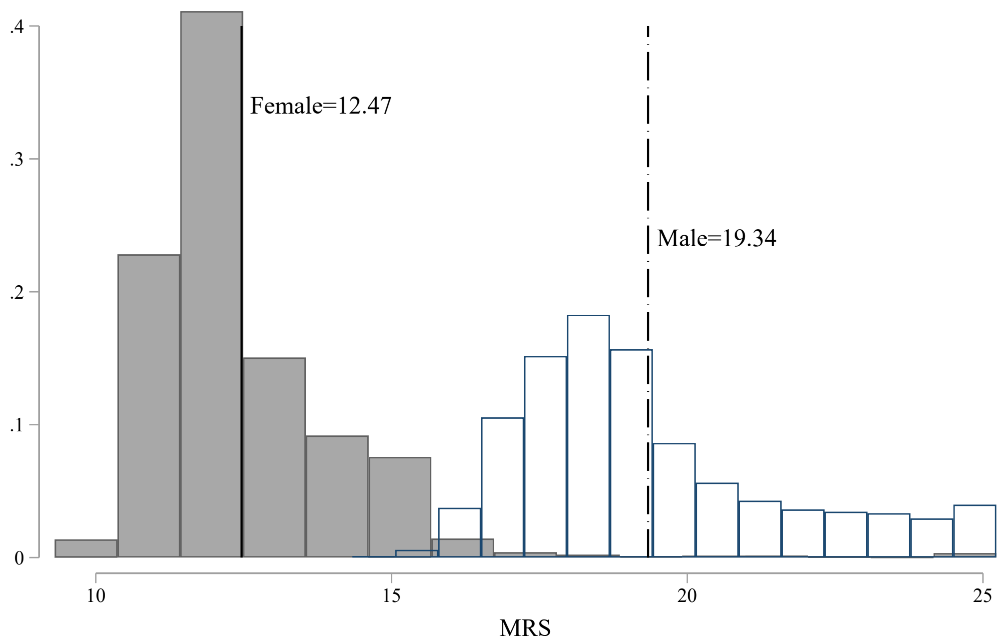
Figure 9: Effects on the Likelihood of Enrolling in College Type m by SES



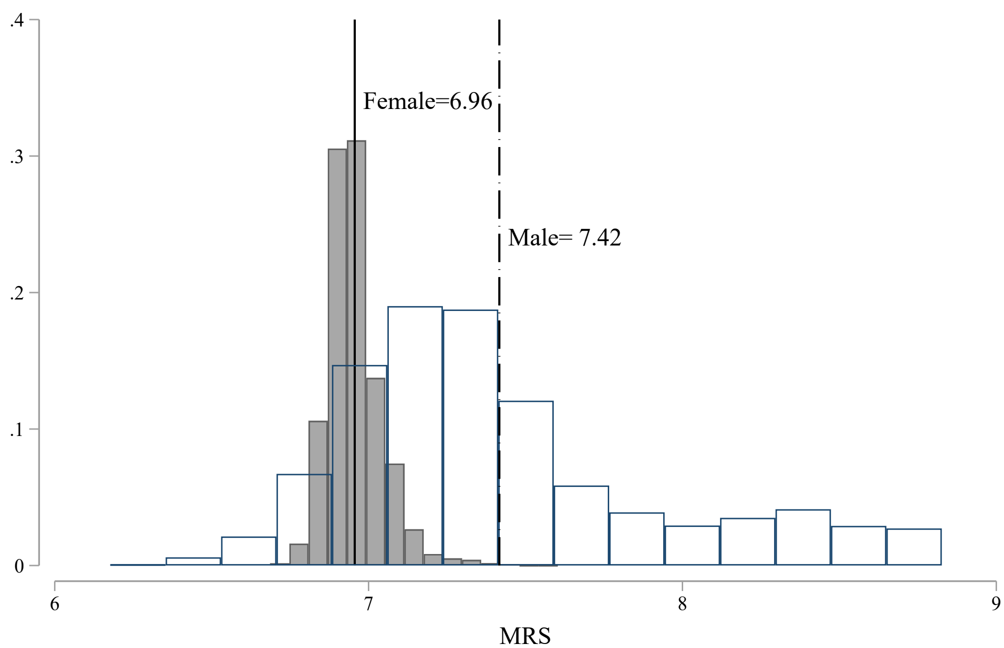
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. High-quality Colleges are all private colleges whose graduates earn more than 2250 PEN while Low-quality private colleges are the ones whose graduates earn less than 1450 PEN using administrative data from wage records in 2014. *Typical Students* is the sample of students who were very likely to attend college while *New Students* are the ones induced by the policy.

Figure 10: Marginal Rate of Substitution by Gender

(a) Private College

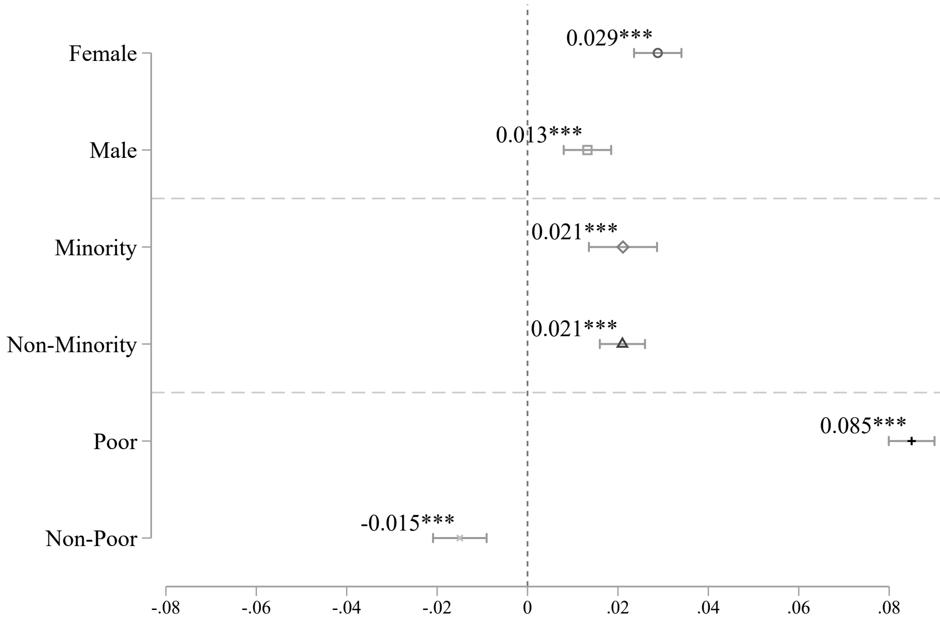


(b) Public College



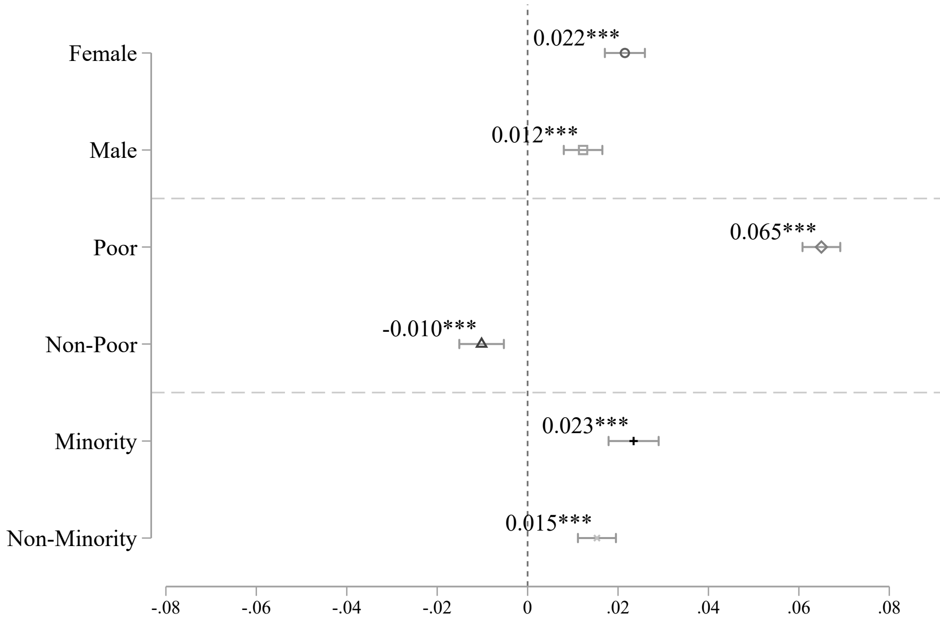
Notes. MRS distributions are winsorized at the 1% and 99% percent.

Figure 11: Heterogeneous Treatment Effects of College Completion



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*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.

Figure 12: Heterogeneous Treatment Effects of White Collar Employment



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 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	Excl. Downtown	Private	Public	Elite
	(1)	(2)	(3)	(4)	(5)
Treatment*Open	-11.60***	-13.33***	-11.81***	-11.66***	-9.492***
	(1.127)	(1.297)	(1.136)	(1.081)	(1.589)
Mean Control	62.01	66.54	59.91	79.07	52.14
N	582	472	582	582	582
District FE	Yes	Yes	Yes	Yes	Yes

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. Column (5) shows the results on commuting time to elite colleges.

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	Low Quality	Public	STEM
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.013**	0.026***	0.025***	-0.001
	(0.006)	(0.007)	(0.008)	(0.007)
Dep. Var. Mean	0.112	0.165	0.377	0.301
N	101187	101187	101187	101187
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All 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.

Table 5: Effects on Distance to College and College Connectivity

	Distance (km.)	Connected	Priv. Connected	Pub. Connected
	(1)	(2)	(3)	(4)
Treatment*Open	-0.998***	0.026***	0.025***	0.001
	(0.089)	(0.007)	(0.007)	(0.005)
Dep. Var. Mean	10.89	0.382	0.223	0.159
N	120212	120212	120212	120212
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Distance is measured in kilometers using Euclidean distance from block centroid to college location. Columns 2-4 measure the probability of enrolling in a college connected to the new lines, with the closest station less than 3km away. Column 3 and Column 4 measure the probability of enrolling in a private and a public college connected to new lines, respectively. 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

	Female		Male	
	Private	Public	Private	Public
	(1)	(2)	(3)	(4)
Mean				
Commuting Time	-0.015***	-0.019***	-0.017***	-0.025***
	(0.000)	(0.000)	(0.000)	(0.001
Wage Premium	0.223***	-0.032***	0.334***	0.269***
	(0.003)	(0.004)	(0.004)	(0.004)
SD				
Commuting Time	0.004***	0.017***	-0.003***	0.018***
	(0.001)	(0.001)	(0.001)	(0.001)
Wage Premium	0.000	0.002	-0.002	0.000
	(0.000)	(0.001)	(0.001)	(0.000)
Students	42308	13180	31055	15902
Observations	1480780	461300	1086925	556570
Log-Likelihood	-145313.8	-45570.5	-104759.8	-53481.4

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

	No DT	Inc. DT
	(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.

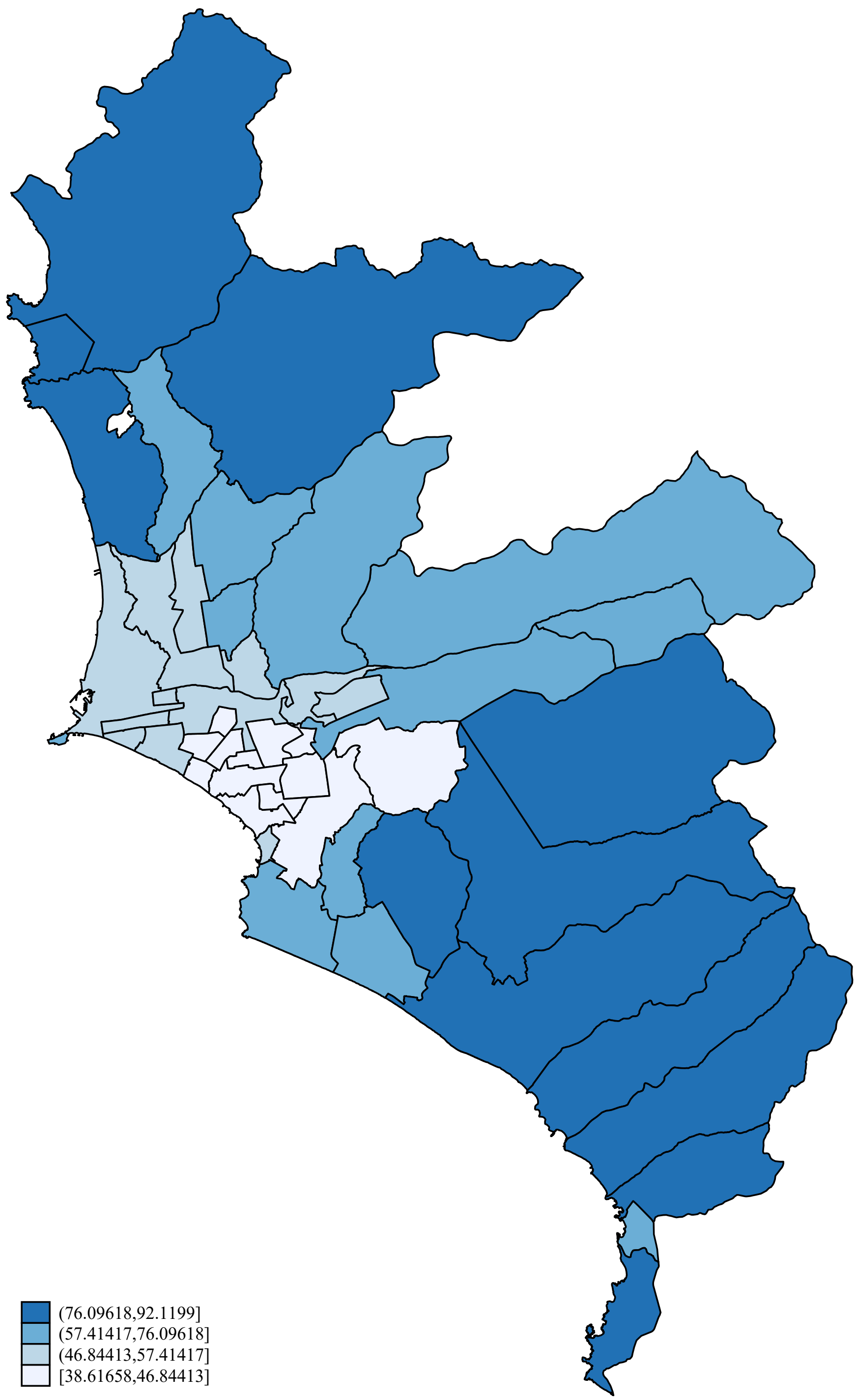
Table 8: Effects on Employment Rates

	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 Travel Time from Home to University (in min)



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

Figure A.2: Planned but non-executed Metro lines in Lima

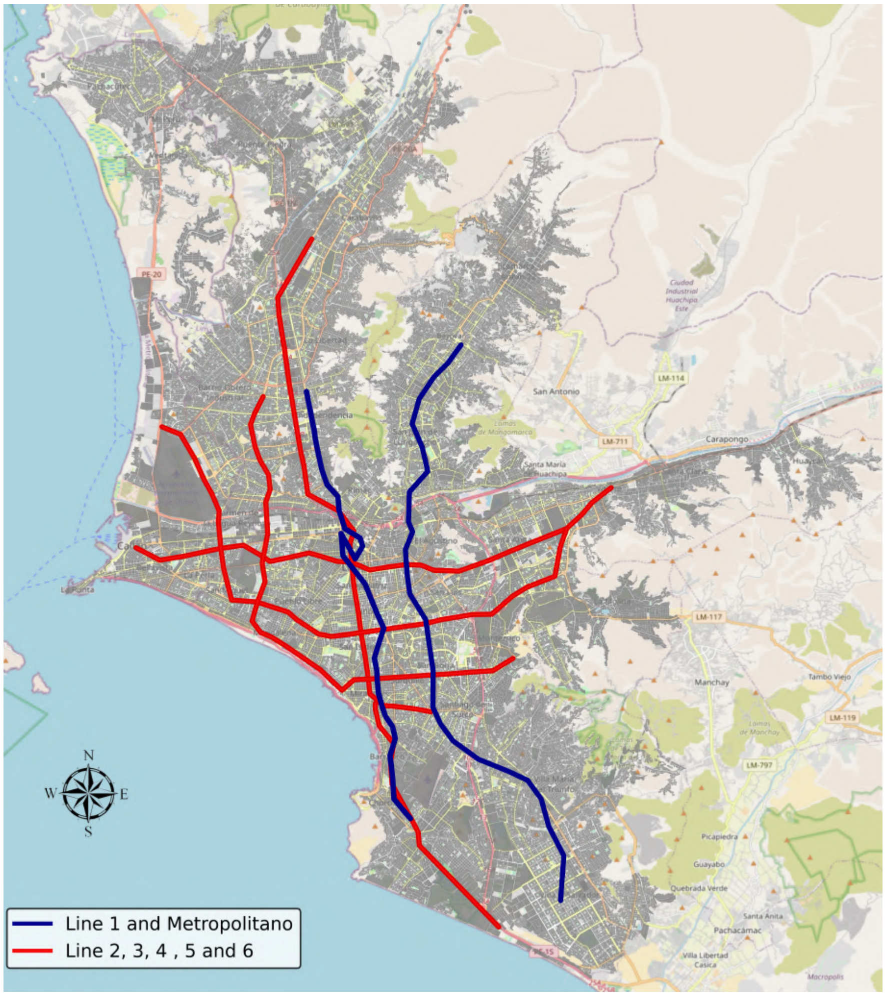


Figure A.3: College location by license status and new stations

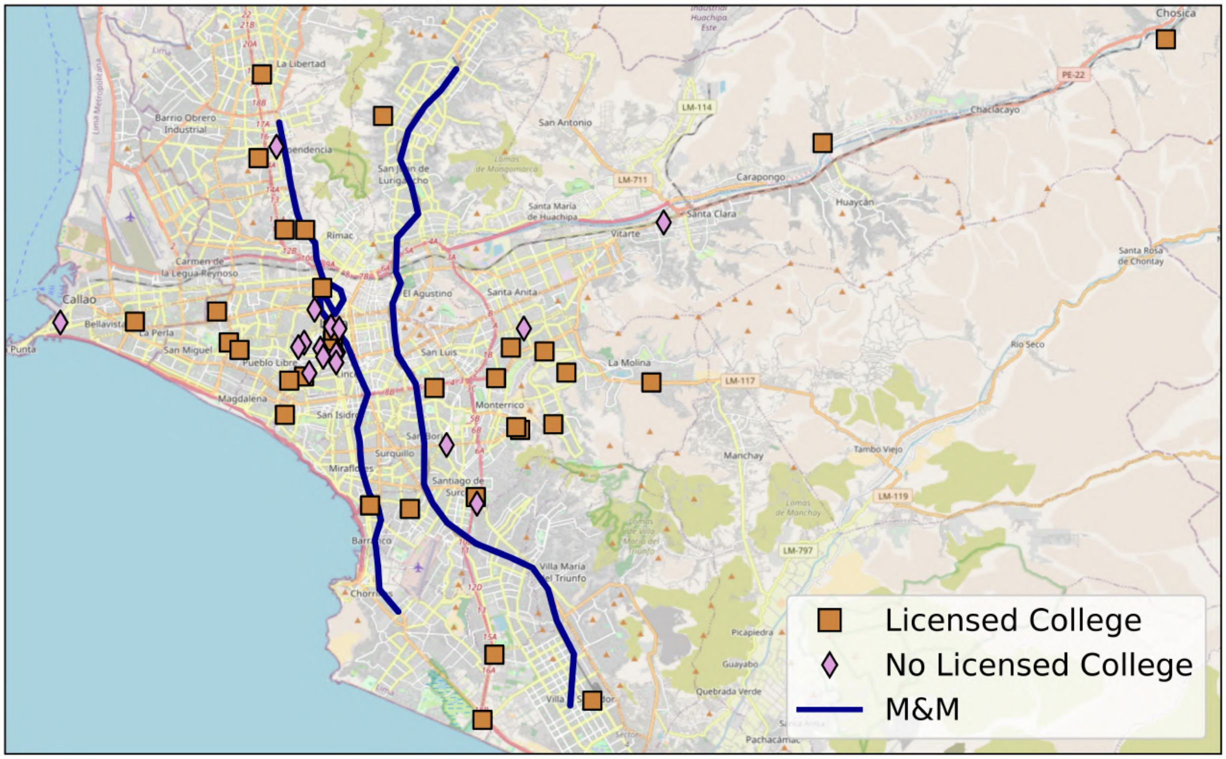
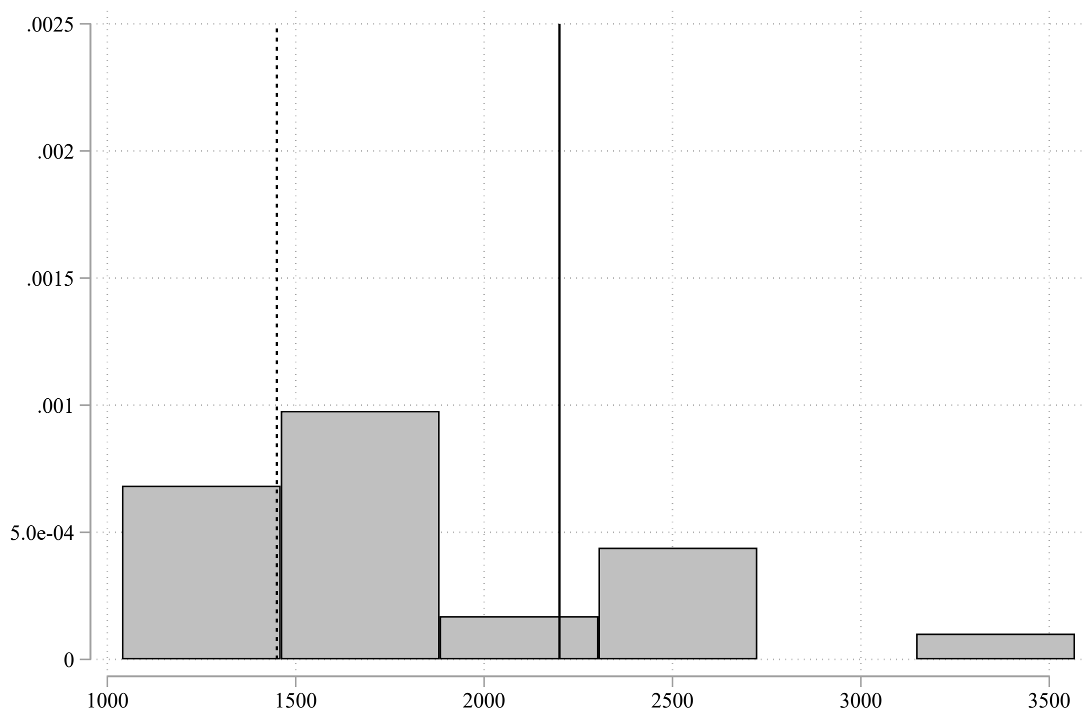
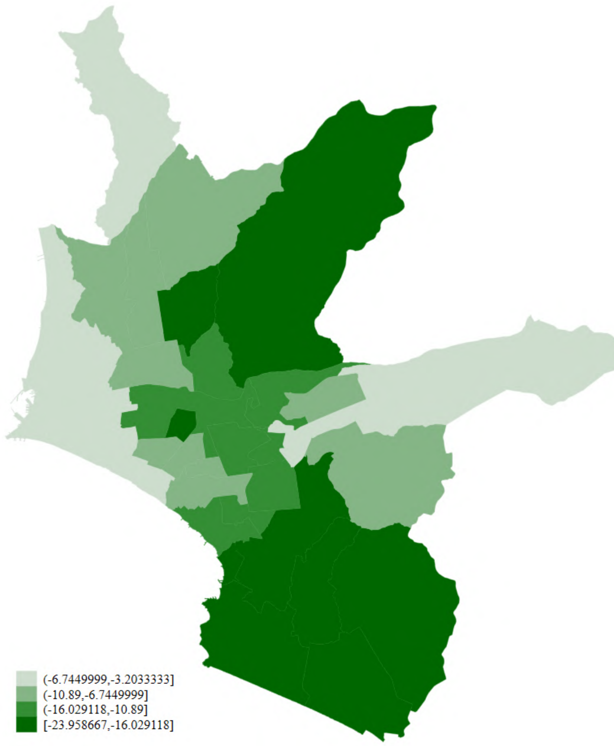


Figure A.4: Wage Distribution of Recent Graduates (Average by College)

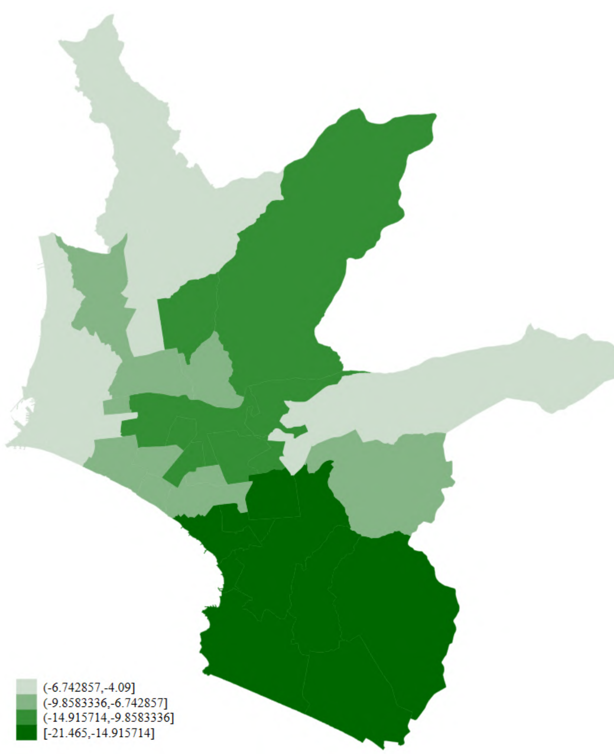


Source: Planilla Electronica. The dashed line shows the limit for the 25% bottom of the distribution below 1450 PEN while the plain line marks the 25% top of distribution above 2250 PEN.

Figure A.5: Commuting Time to College (before/after new stations)



(a) All colleges



(b) Private colleges

Figure A.6: Share of population who access higher education, Census 2007

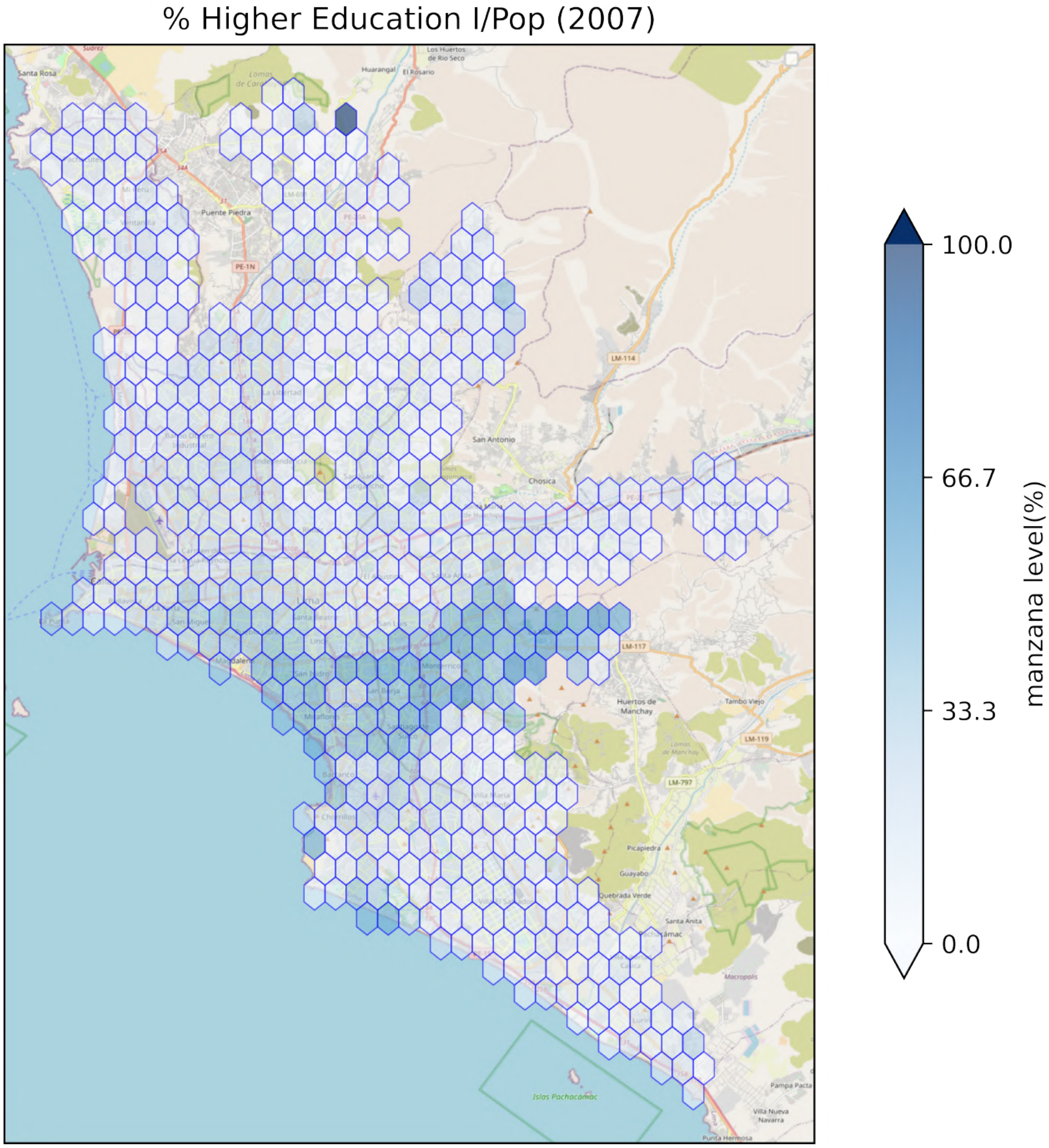


Figure A.7: Effects on College Enrollment Rates using Callaway and Sant'Anna (2021) estimator

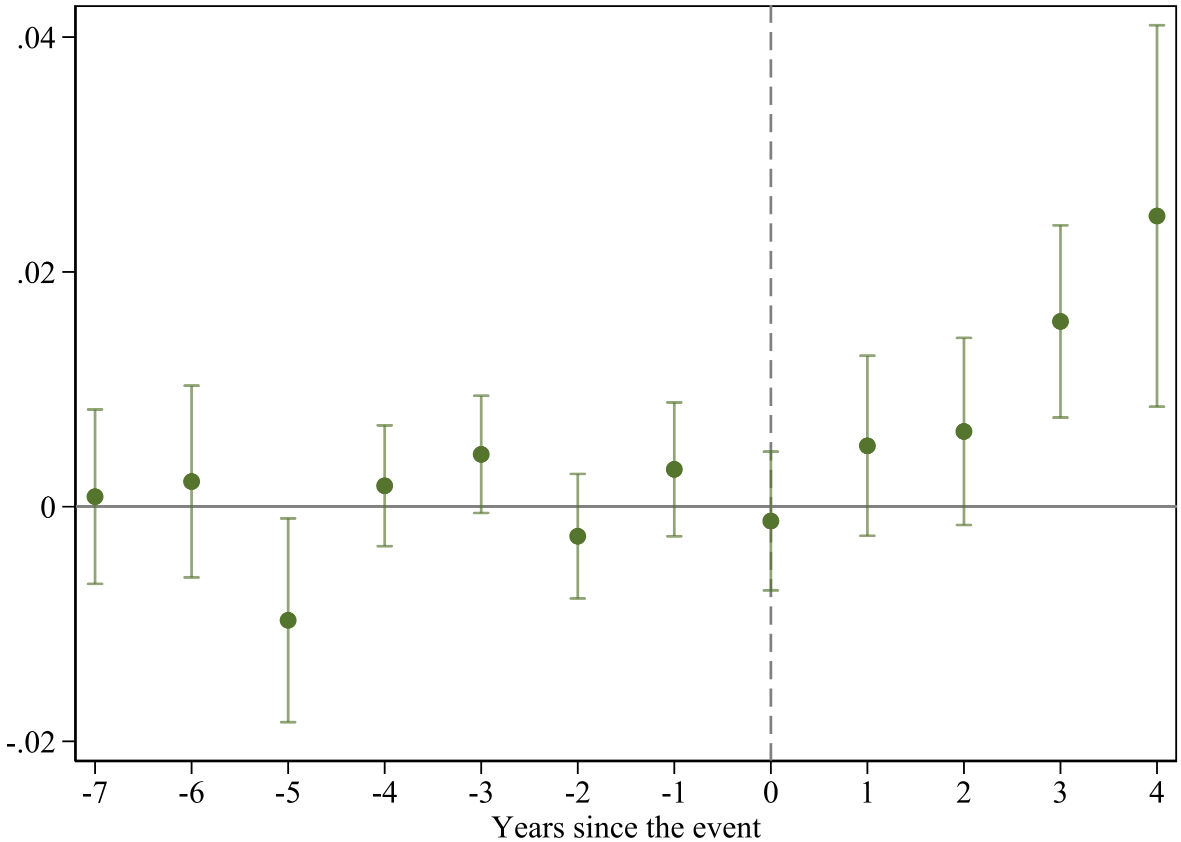


Figure A.8: Log(enrollment) for connected and non-connected colleges

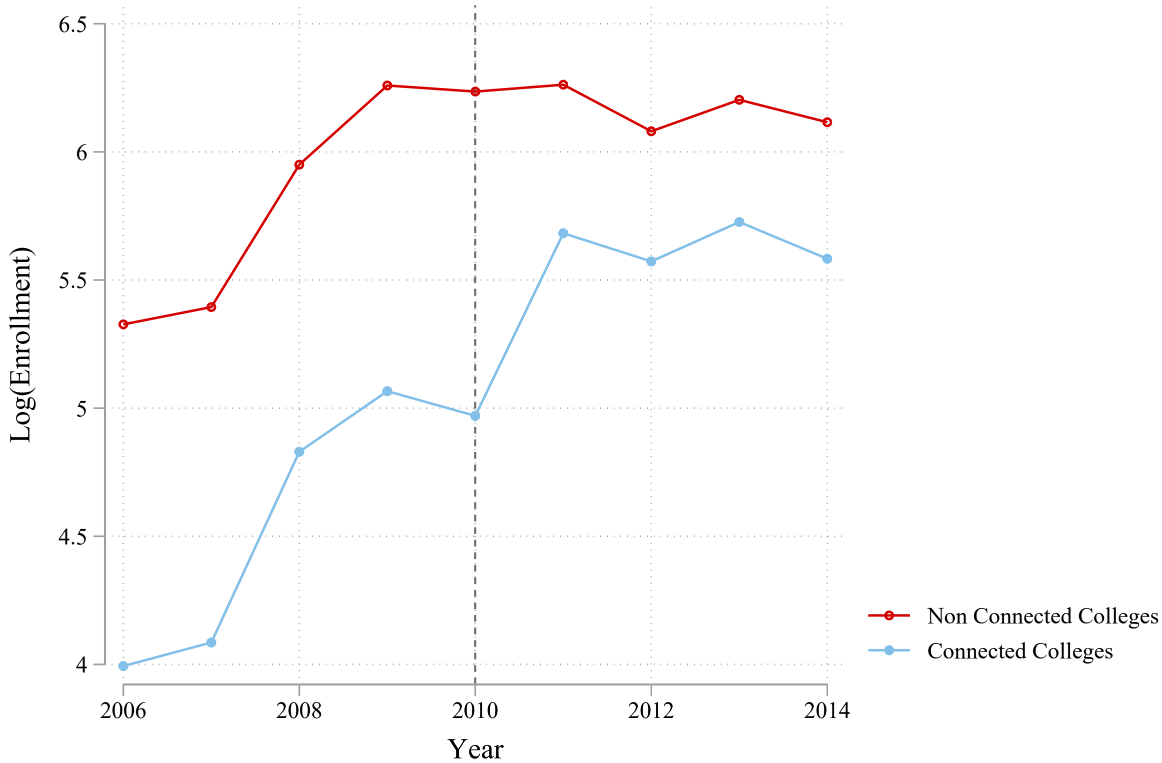
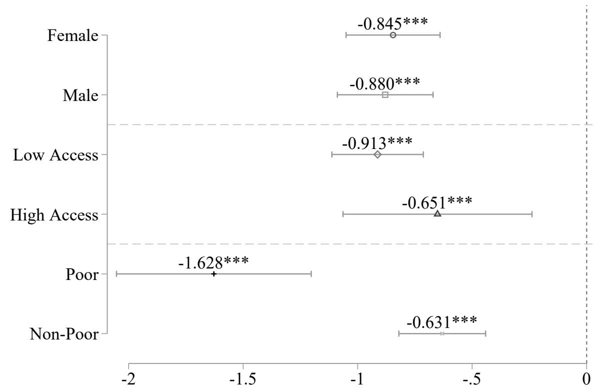
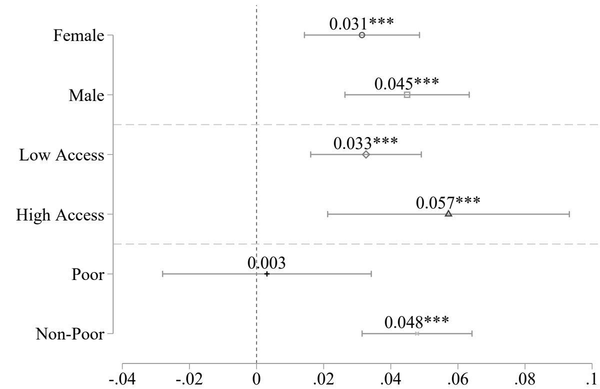


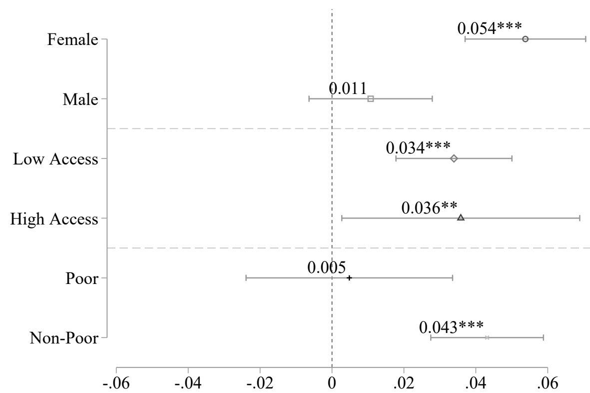
Figure A.9: Heterogeneous Effects of Distance to College and College Connectivity



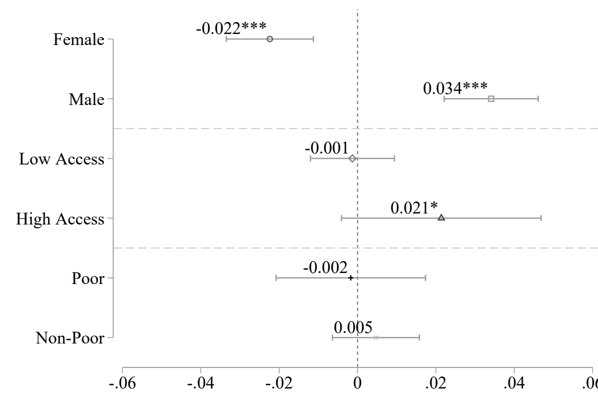
(a) Distance (km)



(b) College Connected



(c) Private College Connected



(d) Public connected to the $M\&M$

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 the $M\&M$, 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 $M\&M$, respectively. Regressions include block and cohort-fixed effects.

Figure A.10: Dynamic Effects of the $M\&M$ on College Completion Rates

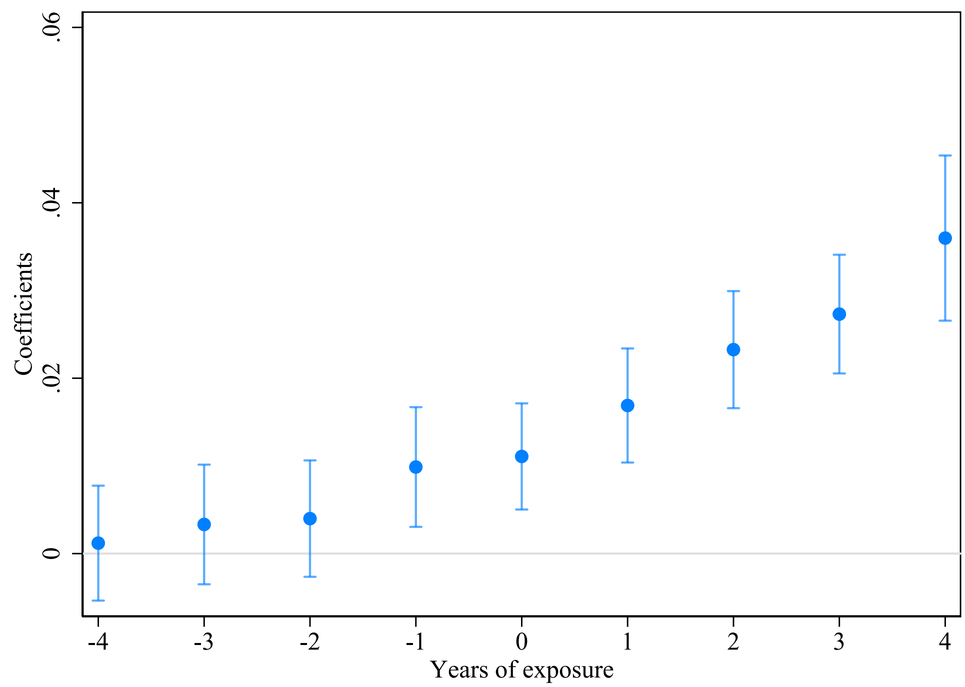


Table A.1: Effects on the likelihood of enrolling in a high-quality college

	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

Table A.2: 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 *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 A.3: Effects of the New Lines on College Choice (relative to Predicted)

	High Quality	Low Quality	Public
	(1)	(2)	(3)
Treatment*Opening	-0.019***	0.008	-0.012
	(0.006)	(0.009)	(0.011)
Dep. Var. Mean	0.112	0.165	0.301
N	101187	101187	101187
Block FE	Yes	Yes	Yes
Year FE	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.

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. [30] 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.

[30] 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 [31] 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.

[31] 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 $M\&M$).

Table D.1: kNN Scores

	Access to College	Private			Public
		Low Returns	High Returns	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