Opportunity Bound: Transport and Access to College in a Megacity

Fabiola M. Alba-Vivar *

September, 2025

First Version: November, 2023

Click here for the most recent version

Abstract

Access to safe and reliable transportation is a critical yet understudied determinant of college enrollment, particularly in low- and middle-income countries where commuting is the norm for students and infrastructure often falls short. This paper examines the effects of major public transit expansions in Lima, Peru—specifically a Bus Rapid Transit (BRT) system and an elevated metro line—on college enrollment and institutional choices. Leveraging a novel geocoded dataset and a staggered difference-in-differences strategy that compares neighborhoods that gained access to those exposed to planned-but-not-executed lines, I find that improved connectivity increases college enrollment. However, this expansion disproportionately boosts enrollment in lower-quality institutions, especially among women. In contrast, men become more likely to enroll in public colleges, which typically yield higher wages. A college choice model further shows that women are less responsive to potential earnings gains and more sensitive to commuting burdens. These findings underscore that while transit improvements expand access to higher education, gender-specific mobility constraints shape who benefits and how.

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

I thank Alex Eble, Cristian Pop-Eleches, Sandra Black, and Micaela Sviatschi for their constant guidance and support. I also gratefully acknowledge all the comments and suggestions from Sofia Amaral, Mark Anderson, Liz Ananat, Andy de Barros, Michael Best, Daniel Bjorkegren, Girija Borker, Kirill Borusyak, Carolina-Concha Arriagada, Sarah Cohodes, Rob Garlick, Paul Goldsmith-Pinkham, Jose Flor-Toro, Carolina Lopez, Stephen Keating, Leah Lakdawala, Bentley MacLeod, Matteo Magnaricotte, Stephen Ross, Sebastian Otero, Sarah Pearlman, Nishith Prakash, Randall Reback, Maria Rosales-Rueda, Kartini Shastry, Judy Scott-Clayton, Sebastian Tello-Trillo, Anja Tolonen, Daniel Velasquez, Eric Verhoogen, and Jack Willis. I am also thankful to all conference participants at the WEAI Graduate Student Workshop, Seminario MAP, AEFP, UEA, NEUDC, SEA, LACEA LAUrban, SOLE, SEEDS, RIDGE. I also thank seminar participants at Columbia, PUC, TecMonterrey, Oberlin, Bucknell, Wake Forest, Rutgers, Barnard, MSU, Delaware and Duke. 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 Teachers College. I am deeply grateful to Roberto Mendoza, who provided excellent research assistance. Mia Annunziata also provided valuable 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.
* Department of Economics - Wake Forest University, *albavif@wfu.edu*, Winston-Salem, NC.

"I have spent so many hours sitting on a combi, and it is a waste of time. On average, I spend 3 hours per day going back and forward to school, and I am not even counting the rush hour. Living so far away is difficult; it means waking up very early and I cannot stay up late at night. It means not being able to eat well, that I skip meals and I cannot choose early morning classes because I know I will not make it on time." - A female college student in Lima (2023)

1 Introduction

The majority of the world's college students—over 500 million—live in urban areas and depend on transportation infrastructure to commute to school every day. Even in well-developed contexts like the United States, 85% of college students live off campus and face daily commuting challenges (Inside Higher Ed, 2020)[1]. This dependence becomes particularly acute in low- and middle-income countries, where transport systems are often slow, unsafe, and unreliable, lagging far behind high-income nations in both coverage and quality (Gonzalez-Navarro and Zarate, 2023). Yet despite this widespread reliance on transportation for educational access, research has focused primarily on how geographic distance affects student outcomes (Acton et al., 2024; Agarwal and Somaini, 2019; Card, 2001; Kane and Rouse, 1995), while far less is known about the direct effects of transportation policies that actively reduce commuting time on college access—a crucial determinant of upward mobility (Chetty et al., 2017)[2]. Understanding these policy effects is especially critical because mobility barriers disproportionately affect women, who typically face higher travel costs (Borker, 2024). Importantly, gender differences in travel behavior, shaped by social norms and safety concerns, may unintentionally influence college decisions and potentially make transportation improvements regressive for women.

This paper examines the causal impact of major public transportation improvements on college enrollment decisions, and their gendered effects. The analysis focuses on Lima,

[1] See this article about the role of commuter students in the US.

[2] Most existing studies on college access focus on financial barriers or affirmative action policies to increase representation among marginalized groups (Page and Scott-Clayton, 2016).

Peru—a megacity of 12 million residents that ranks among the most challenging environments for women's mobility, comparable to cities such as Mexico City, New Delhi, and Jakarta. To identify these effects, I exploit the phased rollout of two large-scale mass transit systems introduced in the early 2010s: a Bus Rapid Transit (BRT) line and an elevated metro line. These new systems offered not only faster travel times but also a safer alternative to the widely used but informal and often unsafe transportation network. The new lines substantially reduced commuting times for thousands of students across the city. Crucially, commuting is the norm rather than the exception: nearly 90% of college students in Lima live off campus and rely on daily travel to attend school. Prior to the transit upgrades, these students spent an average of two hours per day commuting.[3]

I construct a novel dataset that integrates detailed information on student and college locations with the rollout of the two new lines. By geocoding both college campuses and students' home addresses, and linking them to data on the location and opening dates of new transit stations, I am able to trace the impact of improved transportation access on college enrollment behavior. The primary identification challenge in estimating causal effects of transportation improvements is that line placement is not random—areas selected for new transit may differ systematically from those that are not, potentially confounding any observed changes in educational outcomes. To address this challenge, I implement a staggered difference-in-differences (DiD) strategy that exploits a unique institutional feature: while six metro lines were formally planned for the city, only one was ultimately built due to budget constraints and corruption scandals.[4] This setting enables a placebo-style identification strategy, following the approach of Donaldson (2018), where I compare outcomes for students in neighborhoods that gained access to new stations with those in neighborhoods where stations were planned but not executed. To identify these planned-but-not-executed lines, I collected

[3]According to the 2010 National University Census, students living on the urban periphery reported average one-way commutes of 1.5 hours, while those in central Lima traveled approximately 40 minutes, similar to commute times for college students in Delhi, as documented by Borker (2020).

[4]The remaining 5 lines were postponed due to budget constraints as well as several corruption scandals associated with the Lava Jato and Odebrecht case in Latin America. See Campos et al. (2021) for further details regarding this case.

and digitized data from Lima's official strategic urban plans and multiple feasibility studies used in the design of the Metro de Lima, providing a more credible comparison group for isolating the impact of transit expansion.

First, I document how improved access to public transportation significantly reduces commuting time to college and modestly increases enrollment rates. Students living in neighborhoods connected to the new transit lines experienced a 17% reduction in average commuting time to colleges citywide, equivalent to a 13-minute decrease per trip (or nearly 30 minutes saved per day).[5] These commuting gains translate into higher college enrollment rates. Neighborhoods connected to newly constructed lines saw a 0.8 percentage point increase in enrollment from a baseline of 11%, representing a 7% relative gain. While this effect size is modest compared to policies explicitly designed to increase college access—such as providing information (8 pp increase in FAFSA applications, Bettinger et al. (2012)), early commitment to free tuition (15 pp increase for high-achieving low-income students, Dynarski et al. (2021)), or offering scholarships (10 pp increase in Colombia's Ser Pilo Paga program, Londoño-Vélez et al. (2023))—it is nonetheless economically meaningful given that these transportation investments were not primarily intended to boost educational access. The fact that infrastructure improvements designed for general mobility generate substantial educational spillovers suggests that transportation barriers represent a significant but underappreciated constraint on college enrollment, particularly in developing urban contexts. The effects are concentrated among private enrollment, which experienced a 1.3% increase. In contrast, the estimated effect on public college enrollment was just 0.2% and not statistically significant, likely due to the limited capacity and highly selective admissions processes at public universities in Peru.[6] While both female and male students responded positively to improved transit access, dynamic analyses suggest that the gains for women increase more steadily over time. This may reflect not only reduced travel time but also improved safety,

[5] These estimates likely represent lower bounds, as they exclude informal transport routes and assume car-based travel times.

[6] See Flor-Toro and Magnaricotte (2021) for evidence on admissions disparities between public and private colleges.

an especially salient factor for women in Lima.[7]

Second, the introduction of new public transit lines significantly altered students' college choices, with effects that differ across both the extensive and intensive margins. I use machine learning techniques to classify students based on their predicted likelihood of enrolling in college in the absence of the policy. This allows me to distinguish between *typical* students—those who would have enrolled regardless—and *new* students, whose enrollment was induced by improved connectivity.[8] These college choice patterns reveal striking gender differences with important implications for educational equity. Among *typical* students, men increasingly switch from low-quality private to public colleges, while women show no significant changes. Among *new* students, men are disproportionately more likely to enroll in public institutions, whereas women are more likely to enroll in low-quality private colleges. This gendered response is particularly concerning given that public colleges provide superior labor market returns in Peru, yet women—despite higher overall enrollment rates—are systematically channeled away from these higher-return institutions, potentially reinforcing existing gender wage gaps.

Gender-specific mobility constraints and institutional barriers combine to explain these patterns. Male students exhibit stronger preferences for connected public colleges, which are typically located on the city's outskirts, while female students increasingly choose connected low-quality private institutions that are centrally located and provide an easier commute. Social norms and safety concerns also appear central to these choices: using district-level femicide rates as a proxy for an adverse environment for women, I find that women in lower-

[7]Robustness checks confirm the consistency of these results across alternative DiD estimators (Borusyak and Hull, 2023; Callaway and Sant'Anna, 2021), functional forms, and sample definitions (e.g., including or excluding downtown Lima). Moreover, heterogeneity by distance to transit and job centers suggests that the observed effects are driven by improved access to colleges rather than proximity to labor markets.

[8]Across both groups, I find that students increasingly enroll in colleges located near the new transit stations, with this spatial reallocation pointing to a strong behavioral response to geographic accessibility. Among *typical* students, the shift is primarily toward public colleges, which are tuition-free but highly selective and limited in capacity. Among *new* students, the increase is concentrated in low-quality private institutions—those with weaker academic profiles and lower labor market returns. Robustness checks using alternative definitions of college quality and outcome specifications confirm that these patterns hold consistently, reinforcing the validity of the findings.

risk areas are more likely to enroll in lower-quality institutions, suggesting that mobility and safety considerations override quality concerns. These patterns echo findings from Borker (2020) in Delhi, where women similarly sacrifice institutional quality to minimize safety risks. Institutional characteristics—including gender composition, STEM enrollment, and low female enrollment—further constrain women's choices, creating a reinforcing cycle where these barriers limit women's access to high-return educational pathways.

To assess the medium- to long-term effects of improved transportation access, I implement an exposure DiD strategy using the 2017 National Census, seven years after the first line opening. The results show that among *new* students, women are more likely to complete college, while men see a decline in completion rates. This pattern aligns with earlier findings that women disproportionately enroll in lower-quality private colleges, which may facilitate earlier graduation. Male students, more likely to attend selective public institutions, may face longer time-to-degree due to stricter requirements or more flexible academic pacing. Employment outcomes mirror these educational trends: access to new lines increases overall employment rates among students, particularly through gains in white-collar jobs. Particularly, female students shift into white-collar occupations, likely facilitated by their higher college completion rates. These results highlight a nuanced tradeoff: although improved transit access draws women into potentially lower-quality institutions, it also increases their graduation and entry into higher-quality employment.

To understand these behavioral patterns, I use a random utility model with a mixed logit framework, estimated separately by gender, to capture heterogeneous preferences for college quality (proxied by post-graduation wages) and commuting time college. The model estimates show that all students prefer colleges with higher expected earnings and shorter commutes, but men are generally more willing to trade off commuting time for higher wages than women, particularly those with a higher baseline probability of attending a public college. Female students likely to attend public universities exhibit a notably lower responsiveness to wage differences, possibly reflecting non-financial constraints such as safety concerns

or social norms. These findings imply that transportation improvements may differentially benefit students by gender and that targeted policies addressing mobility barriers and safety concerns could enhance college access and quality choices for female students.

This paper contributes to several strands of the literature. Despite the lack of granular data and identification challenges, evidence indicates that reduced transportation costs affect human capital investment, specially in the K-12 setting in developing countries. Bicycles, for instance, bolster women's access to education and improve schooling outcomes and aspirations (Fiala et al., 2025; Muralidharan and Prakash, 2017).[9] Moreover, reduced transportation costs can reduce gaps between low- and high-skill students (Asahi and Pinto, 2022), as families with reduced transportation costs often travel further (Herskovic, 2020). However, in places with meritocratic systems, only high-achieving students with highly educated parents take advantage of reduced transportation costs (Dustan and Ngo, 2018). My results show that living in a neighborhood connected to new lines not only increases college enrollment rates but also influences college selection among students. I also find differential responses by gender: male students become more likely to enroll in public colleges, while female students enroll in low-quality private colleges.

What is more, travel safety and social norms are closely linked to mobility. In Lima, informal transportation is widely considered unsafe due to frequent accidents, sexual harassment, and muggings (Dominguez Gonzalez et al., 2020). This situation mirrors that of other large cities in the developing world, where women face significantly higher rates of harassment and crime than men, leading to gender-specific responses when travel costs change. The new transportation systems I study offer a safer alternative to the informal buses that circulate the city. Thus, I expect that the opening of these new lines also represents an expansion of safe(r) travel options for women. There is a small but growing literature on this issue, with most findings indicating that women have a higher demand for safe transporta-

[9]Both papers study a similar policy: providing bicycles to female students so they can mobilize easily and therefore reduce transportation costs. This particularly relevant in rural setting where distance to schools plays a significant role.

tion (Borker, 2024; Kondylis et al., 2020), which in turn can directly influence their labor supply (Field et al., 2025; Martinez et al., 2020).[10] Beyond employment, travel safety can also shape human capital investment. In Delhi, Borker (2020) documents how women often choose lower-quality colleges over more prestigious but less safe institutions.[11] My findings align with Borker (2020)'s results in India but also provide causal evidence on the effects of a policy that reduces commuting time to college and, plausibly, exposure to harassment and crime. This reduction in risk led female students to enroll in lower-quality private colleges located near the new stations, prioritizing safety and convenience over institutional prestige and expected returns.

Finally, this paper also contributes to the literature studying the impact of improving or building new transportation in large cities in developing countries. Some evidence shows positive effects on labor market opportunities and welfare in the medium and long run (Balboni et al., 2021; Tsivanidis, 2022; Zarate, 2022).[12] However, Severen (2023) finds no impact on local productivity or amenities for the Los Angeles Metro Rail. Warnes (2021) shows that the BRT line in Buenos Aires increased segregation between high- and low-skill workers, and a similar result is found by Lee and Tan (2024) on the Downtown Line in Singapore. In the case of Peru, Velásquez (2023) finds no reduced-form effects on rents, income, poverty, and household expenditures for the opening of the new bus rapid transit line in Lima. Nevertheless, less is known about the direct causal effects on human capital investment. More specifically, in terms of higher education, and to the best of my knowledge, this paper is the first to study the causal effects of a city transportation upgrade on college education. I document how neighborhoods experience reduced commuting time to any college in the city

[10] In Lima, the new transit stations have been shown to positively impact women's labor supply (Martinez et al., 2020; Velásquez, 2023).

[11] Lima, the setting of this study, is considered one of the most dangerous large cities in the world for women's mobility and transportation, comparable to Delhi, Mexico City, and Jakarta. See the ranking on women's safety in transportation in a Reuters study here.

[12] Other similar research on different types of transportation such as railroads and roads (Brooks and Donovan, 2020; Donaldson, 2018; Donaldson and Hornbeck, 2016) shows long-run, positive effects on trade and economic growth.

and how this boosts college enrollment rates.[13]

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

2 Background

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

2.1 College Education in Peru

The Peruvian education system is based on three levels: primary education (6 years), secondary education (5 years), and the higher education level, which often lasts from 2 (technical school) to 10 years (medical school). On average, college students graduate in 5 to 7 years. According to the 2017 Peruvian Census, 4 out of 10 recent high school graduates (between 17 and 21 years old) enroll in some type of higher education. More specifically, 15% enroll in a technical school or community college, 22% enroll in college, and the remaining 63% do not enroll in any type of higher education (Alba-Vivar et al., 2020). Following trends similar to those in middle- and high-income countries, Peruvian women enroll in college at slightly higher rates than men.

Several additional aspects of the college education system are worth highlighting, as they

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

are relevant to understanding students' college choices. Unlike in the United States or other countries in the Latin American region, Peru does not have general standardized exams at the end of high school, making it difficult to assess students' abilities. Also, unlike Chile or Brazil, there is no centralized admission system to enroll in college, as each institution maintains its own distinct procedure for admitting students. Students typically face a college choice decision tree as depicted in Figure 1. A student may choose to attend college, community college, or not attend any college at all (instead staying home or working, which represents approximately one third of recent high school graduates). If a student chooses to attend college, they might choose between a private or public institution.

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

On the other hand, private colleges exhibit greater variance in price and quality. Their admission process is very straightforward, and students face little or no uncertainty about their probability of enrolling, as only a few elite private colleges tend to have more selective admission exams. For simplification, I assume that there are three types of private institutions: high-, medium-, and low-quality private colleges. High-quality colleges tend to charge higher tuition fees compared to low-quality colleges (college fees range from 150 USD monthly to 1,300 USD; for reference, the minimum monthly wage in Peru was 180 USD in 2010). For private universities, prestige is correlated with school fees, given the selection of their students. Graduates from low-cost and low-quality colleges have lower wages in the labor market compared to their peers from high-cost and high-quality institutions. Appendix Figure A.5 shows the average wages after graduation for all available choices, providing evidence that

higher-quality colleges yield significant premiums in the labor market.

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

2.2 Public Transportation in Lima

Lima's population is comparable to other major global cities, such as New York City, Paris, Xi'an, Chennai, Jakarta, Bogotá, and Los Angeles. However, Lima is significantly less dense (8,000 inhabitants per square mile compared to New York City's 29,302 inhabitants per square mile), and commuting across the city can take up to three hours during rush hour. In the 1990s, market liberalization policies allowed the importation of used cars and mini-buses, which became the foundation of a new transportation system for commuters. These privately operated mini-buses, known as *combis*, partially alleviated the city's growing demand for transportation. However, their poor quality and the lack of transit regulations made this mode of transportation unsafe for commuters, particularly for young commuters. In 2010, 57% of people involved in traffic accidents were under 25 years of age.

In July 2010, the city opened the Metropolitano, a bus rapid transit (BRT) line that connected the northern and southern parts of Lima. The Metropolitano was the first mass public transportation system in Peru, linking 12 of the city's 44 districts. Regular commuters

paid a flat fee of 1.50 PEN (approximately 0.50 USD), while college students received a 50% discount.[14] A year after the Metropolitano opened, the Peruvian president inaugurated the first line of the Metro de Lima (*Linea 1*, which was built on an elevated viaduct.[15] This line connected the northeast and southeast areas of the city, serving over two million people.

The Metro de Lima took almost 40 years to complete. In the 1970s, the Peruvian Ministry of Transport designed a comprehensive metro system for the city, aiming to connect multiple districts, especially emerging neighborhoods on the outskirts. The original plans included eight lines; however, as years passed, only six were properly studied and evaluated (Appendix Figure A.4 shows these lines). A small part of the project was initiated in the 1980s during the first government of President Alan García, but it remained incomplete for more than 20 years and never opened to the public. Peru's major economic crash prevented further development of this project until it was revived in 2006, during García's second presidential term (Campos et al., 2021).

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

In this paper, I focus on the first line of the Metro de Lima (*Linea 1*) and the new stations of the Metropolitano. Both systems provided a faster and safer service compared to *combis*, reducing transportation costs and travel time for thousands of students in Peru's capital. Notably, both systems crossed the city from north to south, connecting several

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

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

neighborhoods to downtown Lima, which is the hub of numerous college campuses, as shown in Figure A.2.

3 Data

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

College Enrollment. Data on college enrollment come from the Peruvian Ministry of Education, which annually compiles enrollment data for every college in the country. These records include information on students' year of enrollment, college, home address, major, age, and gender. I restrict my sample to students whose home addresses are located within the boundaries of Lima and Callao (the Lima metropolitan area). I use the Google Maps API to collect GPS coordinates for these home addresses. For fewer than 5% of cases where the algorithm failed, I impute GPS coordinates at the block or neighborhood level.[16] Additionally, I refine the sample by including only recent high school graduates or students under 20 years old in the analysis. The study period spans from 2006 to 2014.[17]

College Data Information. I use the 2010 university census *"II Censo Nacional Universitario"*(CENAUN) compiled by the Ministry of Education and the National Institute of Statistics and Informatics (INEI), which contains information on students, faculty, and infrastructure for each college. In addition, I obtained the GPS location for each university and verified it using the Google Maps. Figure A.2 plots the resulting GPS coordinates as red dots.

Peruvian Census. The 2007 and 2017 Peruvian Census data come from the INEI. Both

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

[17]I do not include information after 2014, as it marks the beginning of a significant higher education reform in Peru. In subsequent years, the Ministry of Education changed the format in which it collected enrollment data and initiated a licensing process in 2016 that denied operational licenses to one-third of colleges in the country for failing to meet basic quality standards (Alba-Vivar et al., 2023). This may have affected students' college choices.

datasets are geocoded at the block level.[18] I use the data from 2007 to obtain block-level counts by age and use this as the denominator for college enrollment rates at the block level. Then, I use the 2017 data to explore long-term outcomes such as college completion and employment status. Here, I restrict my sample to individuals living in Lima's metro area and between 17 and 28 years old.

Transportation. All data on stations from the new transportation systems come from the Autoridad de Transporte Urbano para Lima y Callao (ATU). This includes the GPS location and address of all stations. Data on Metro de Lima's planned-but-not-executed lines come from the national government's (Ministry of Transport) multiple technical records (for details, see Appendix B). I geocoded all planned stations from six routes as seen in Figure A.4.

Commuting Time. An important variable in this paper is the amount of time students spend commuting to college. I first calculate the average commuting time from each student's household to any college in the city for a random sample of students. I use road network data from the OpenStreetMap API,[19] which include information on road types (highways, motorways, etc.). I then calculate the optimal route, defined as the shortest possible route from each household to each college in the city. I follow the procedure and data from Velásquez (2023) to impute velocities for major highways and the new lines, and I complement this with data from the Google Maps API to obtain speed information for primary and secondary roads. With this information, I compute commuting times with and without the new lines. Additional details are included in Appendix C.[20]

Adverse environment for women. I use data on femicides collected by the Peruvian Ministry of Women and Vulnerable Populations (formerly the Aurora Program). Femicides, defined as the intentional killing of women motivated by gender-based reasons, represent the most severe and extreme form of violence against women and girls. My analysis relies on

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

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

[20] Given computational constraints, I calculate travel times for a random subset of households across the city that are representative at the district level, and then use these values for the entire sample.

district-level femicide counts for 2009 and 2010, prior to the introduction of the new lines. Due to the lack of alternative data on women's safety or social norms, femicides serve as a proxy; however, this likely underestimates other safety issues such as street harassment or restrictive social norms that limit women's mobility. Appendix Figure A.3 illustrates the distribution of femicide rates across Lima Metropolitana and Callao. In this paper, districts highlighted in yellow are identified as areas of low risk for women.

Recent Graduates Wages. I use data compiled by the Peruvian Ministry of Labor (*Planilla Electronica*) which includes monthly wages for recent graduates.[21] It also includes college information such as major, gender, and college. I restrict my sample to students who graduated in 2014 and 2015, and collapse the data at the college × major × gender cell. This information is used in Section 7.

3.1 Descriptive Statistics

Table 8 presents the summary statistics for the main sample, using information before the stations opened. Panel A reports college enrollment rates using two measures: (1) the logarithm of enrollment counts, and (2) the rate of students in a given year-cohort who enroll in any college at the block level. In my sample, fewer than one student per block enrolls in college. However, by accounting for the population in the same age group, it becomes evident that, on average, 14% of students under 19 years old enroll in college. The college enrollment rate is defined as follows: $Rate_{it} = \frac{TotalEnroll_{it}^{17-18}}{TotalPop_{it}^{17-18}}$. Furthermore, the table indicates that, on average, women enroll in college at higher rates than men, and that enrollment in private colleges is higher than in public colleges. Importantly, the average distance from students' homes to college is quite similar across neighborhoods, whether they are connected to the new lines or to the planned-but-not-executed lines.

Panel B reports average statistics from the 2007 census, which include total population and educational attainment for individuals over 25 years old. There is a small significant

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

difference in population size between affected and non-affected neighborhoods. Additionally, the population over 25 appears to be slightly more educated in the treatment group, with higher shares of people with some type of higher education. For this reason, I control for the share of the population with higher education in the main estimates.

4 Empirical Strategy

In this section, I present the main empirical strategy, which leverages a difference-in-differences methodology that exploits variation in neighborhood exposure to new transit lines and across student cohorts. I describe in detail the estimation approach for short-term outcomes, specifically college enrollment, and discuss the empirical challenges involved in quantifying the impact of new transit infrastructure on students' college access and choices. Additionally, I outline the methodology used to identify students along both the extensive margin (i.e., those who attend college because of the policy) and the intensive margin (i.e., those who would have attended college regardless of the new lines). Finally, I explain the analysis of medium-term outcomes, including college completion and employment.

4.1 Short-Term Analysis

I use a difference-in-differences approach and a flexible event study framework to exploit variation in neighborhood exposure to new lines and across student cohorts. One important concern is the potential for selection bias when establishing a proper control group of never-treated neighborhoods. Simply comparing connected neighborhoods to non-connected ones within the city might overestimate the results, as the allocation of new routes is not completely random. For example, poorer individuals often live in remote areas where the implementation of a new line is unlikely. To address this concern, I follow a placebo strategy similar to Donaldson (2018).

In this paper, the control group consists of neighborhoods that could have been affected

by the new transportation system through planned-but-not-executed lines, as explained in Section 2.2. Figure 2 shows the neighborhoods included in the treatment and never-treated groups. I define a neighborhood as exposed to the executed lines if it is within 1.5 kilometers (approximately a 20-minute walk) from the nearest station. The never-treated group comprises neighborhoods located within 4 kilometers of planned-but-not-executed lines. I also exclude neighborhoods that are simultaneously exposed to both executed and planned-but-not-executed lines, as indicated by the yellow shaded areas in Figure 2.

By restricting the never-treated group to neighborhoods connected to planned-but-not-executed lines, I reduce potential selection bias arising from correlations between the placement of new lines and unobserved shifts in access to college. The event study specification is as follows:

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

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

A few additional empirical challenges arise when using this strategy. 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 *Linea 1* connect to different colleges in the city. To address these potential issues, I draw upon recent advancements in the DiD literature,[22] following Borusyak et al. (2024). I implement their imputation estimator, which allows for treatment-effect heterogeneity and dynamic effects. The estimation process begins by employing Ordinary Least Squares (OLS) regression exclusively on the untreated observations, accounting for any fixed effects at both the unit and time period levels. This step helps establish a baseline of the data before introducing any treatment variables. Next, in the second step, the estimator extrapolates the data to impute the potential outcome in the absence of the treatment. This allows for a comparison between the observed outcome and the imputed counterfactual, providing a treatment effect estimate for each treated unit. Finally, in the third step, the outcome of interest is estimated by combining these treatment effect estimates. This is achieved through a weighted summation of each estimated effect. I estimate the Average Treatment Effect on the Treated (ATT) using the administrative data and Census data at the block level. All standard errors are clustered at the block level.

Additionally, in an ideal experimental scenario, one would randomly allocate a reduction in commuting time for students. Yet, determining the impacts of policies that universally reduce commuting times for all individuals can be notably challenging. First, even when using the placebo lines to reduce selection bias, another potential issue can arise in this setting. For example, neighborhoods located in the city center (which is both economically and geographically central in Lima) might experience higher enrollment rates since they are more likely to have a new line than those in the outskirts. In this sense, other determinants of the

[22]Several papers show that using the two-way fixed effects estimator in a staggered design might yield biased estimates given the presence of both heterogeneous and dynamic effects. See Borusyak et al. (2024); Callaway and Sant'Anna (2021); de Chaisemartin and D'Haultfœuille (2020); Sun and Abraham (2020).

outcome of interest (college enrollment rates) are still not random (recent work by Borusyak and Hull, 2023 highlights this issue). This could happen since families living in these areas are also much more educated, which is a strong predictor of college enrollment (see Appendix Figure A.8).In this paper, I use variation not only in the location of new lines but also in the timing of the opening. Nevertheless, to avoid issues related to neighborhoods located in the geographical center of the city, that were going to receive the treatment regardless of the allocation of newly opened lines, I exclude those located in downtown Lima.[23]

Another concern involves potential general equilibrium (GE) effects, which refer to broader economic impacts resulting from families gaining improved access to better job opportunities, potentially increasing their ability to afford college education. As previously discussed, the recent literature on the GE impacts of new intracity transportation is mixed (Balboni et al., 2021; Lee and Tan, 2024; Severen, 2023; Tsivanidis, 2022; Warnes, 2021; Zarate, 2022). Moreover, such effects typically materialize over a longer time horizon. In this analysis, I focus on short-term impacts, specifically within a four-year window. Notably, Velásquez (2023) find no significant effects on outcomes such as rent or household income following a similar transportation shock in Lima. This absence of significant effects may be attributed to the lack of major changes in population and the still underdeveloped.

Finally, another concern arises from the fact that the new lines might also increase labor market opportunities for students, as the trade-off between the long-term advantages of education and the more immediate gains from engaging in the labor market might arise Adukia et al. (2020). A student who gets connected to new lines might experience both of these effects simultaneously and this can bias my estimates. Even though I am not able to fully disentangle these two effects, the context suggests that this bias is relatively small in my setting. First, in Peru, students are only legally authorized to work at 18 years old, partially ruling out the option of working immediately in the formal sector after high school graduation (16 or 17 years old in this context). Second, most formal jobs (high-skill) are located in the

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

geographical center of the city while most informal jobs (low-skill) are typically located in the outskirts. However, it could be the case that the long-term advantages of education become more salient as students can now access better formal jobs that can potentially hire high-skill workers. In Section 5.2.1, I show how the effects vary on how close students are to where most high-skill jobs are located, suggesting most of the effects are actually coming from students who had higher access to these formal jobs and not those who might have experienced an higher increase in future labor market opportunities.

4.2 Identifying the intensive and extensive margins

I expect that the connection to new lines will affect both college enrollment rates and students' college choices. This occurs not only because commuting times decrease for students living in connected neighborhoods, but also because colleges connected to the new lines become more accessible. Disentangling these two effects is particularly challenging. Ideally, I would have access to detailed information on students' abilities and preferences, which is critical for predicting both who is likely to attend college and which institutions they might choose. For example, in countries with centralized admissions systems, this information is often available through standardized admission exams (as a proxy for ability) and reported college preferences. However, as noted earlier, Peru's decentralized admissions system lacks standardized testing, and there is no uniform measure of student ability when applying to college. This context is similar to that of many other developing countries.

To overcome this limitation, I use machine learning as a second-best tool to recover students' choices. I first estimate whether a student is likely to attend college using rich data on neighborhood characteristics at the block level. This allows me to identify neighborhoods that typically send students to college, as well as neighborhoods that are now becoming new feeders of college students. I also estimate the probability of attending a certain type of college. Using the k-nearest neighbors algorithm, I predict the probability of attending college for each neighborhood based on the pre-treatment data, before the opening of the

new transit lines. I further refine the algorithm using a k-fold validation strategy and choose an optimal k parameter using GridSearch.[24] Using this tool enables me to exploit non-linear relationships, unlike other similar methods. I feed this algorithm with all the data available before the policy takes place (both neighborhood and student characteristics). I then obtain the predicted probability θ_i that a student i will attend college

I also predict whether students are likely to enroll in a high-quality private college, a low-quality private college, or a public college at the individual level.[25] High-quality private colleges are defined as those whose graduates in 2014 earned over 2,200 PEN (placing them in the top 25% of the wage distribution), roughly double the minimum monthly wage. In contrast, low-quality private colleges are those where graduates earned less than 1,450 PEN (the bottom 25% of the wage distribution). All other private universities are considered medium-quality private colleges. Appendix Figure A.6 shows the distribution of post-graduation wages and the thresholds used. I estimate θ_i^m, the predicted probably that a student will enroll in college type m, where m = [High-Quality Private, Medium Quality Private, Low-Quality Private, Public].

The results on college choice will be presented separately for two groups of students: those living in neighborhoods that typically sent students to college (i.e., students with a high baseline probability of attending college) and those living in neighborhoods likely affected by the new lines and now sending students who otherwise would not have attended college (i.e., the marginal students). From now on, I refer to the first group as the "typical students" sample and to the second group as the "new students" sample. A close example of the use of machine learning to identify groups can be found in (Black et al., 2023), where a random forest method was employed to distinguish between students who met and did not meet the eligibility criteria for the Texas Top 10 policy given data limitations that prevented them

[24] See Appendix D for details.

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

to directly identify eligibility. In my case, I use θ_i, the predicted probability of attending college, to differentiate results across two samples: students who are unlikely to enroll in college (*new student sample*) and those who are very likely to attend college regardless of the policy (*typical student sample*), but may now be changing their choices.

4.3 Medium-Term Analysis

For medium-term outcomes, I use individual-level data from the 2017 Census (7 years after the first opening) and estimate an exposure difference-in-differences strategy as in Equation 2. The notation is similar to Section 4.1, where I use as a placebo group the set of neighborhoods that are connected to planned-but-not-executed lines. As in the previous section, I also rely on the imputation estimator proposed by Borusyak et al. (2024). Age cohorts are considered treated if their residency block was exposed to the executed lines by the time they were 17 years old, which is the typical age at which most students graduate from high school in Peru. I further restrict the analysis to individuals born between 1991 and 2000 who reported living in these neighborhoods for at least five years. [26]

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

5 Impact of Access to New Lines on College Education

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

[26] For this analysis, I also exclude the opening of the second half of Line 1 in 2014, since students affected by this event would not have had enough time to graduate from college by 2017.

5.1 Effects on Commuting Time to College

In this section, I investigate whether the new stations reduce the average commuting time from students' households to any college in the city. To do this, I use a simple 2-by-2 difference-in-differences strategy that leverages the opening of new stations and the treatment status of living in a district connected to either the executed or planned-but-not-executed lines.[27]

This analysis should be interpreted as a simple exercise aimed at providing an estimated magnitude of the reduction in commuting time to college and understanding its potential impact on the relevant population: commuter students. The estimates should be considered lower bounds or conservative for two main reasons: (i) informal routes (*combis*) are not included due to lack of data, and (ii) I assume students are commuting by car, which likely overestimates their actual travel time.

The average commuting time to any college in these neighborhoods, before the introduction of the new system, is estimated to be around one hour. This aligns with self-reported data from the 2010 National University Census, as illustrated in Appendix Figure A.1.

The results in Table 2 suggest that the new system reduced the average commuting time to any city college by approximately 17% (a 13-minute reduction from a baseline commute of 66 minutes per trip), which translates to almost 30 minutes saved per day. The reductions are even larger for private colleges, at 20%, compared to a 14% reduction for public colleges relative to their respective baselines. It is important to note that these findings are based on the most conservative estimates, and the actual impact on commuting time could be even greater.

[27] Appendix Figure A.7 visualizes the before-and-after variation. Notably, residents in the northeastern part of the city appear to benefit the most from the new transportation system.

5.2 Effects on College Enrollment

I estimate the impact of access to newly constructed lines on college enrollment rates at the neighborhood level, using neighborhoods connected to planned-but-not-executed lines as the comparison group. Using block-level administrative enrollment data, I estimate the post-treatment effects from Equation 1 for up to four years after the new lines opened.

I find positive effects on college enrollment rates at the block level (Table 3, Column (1)). The coefficient indicates that neighborhoods connected to new lines exhibit a 0.8 percentage point (pp) increase in enrollment rates from a baseline enrollment of 11% (pre-treatment),[28] measured as the number of students enrolled in a block divided by the number of people of the same age living in that block. This represents an average increase of 7% relative to the baseline rates. Column (2) shows similar effects when enrollment is measured in logarithms, showing an increase of 1.4%. Nevertheless, the magnitudes of the estimated coefficients are significantly smaller compared to policies that aimed to directly increase college enrollment rates. For instance, initiatives like providing information (which resulted in an 8 pp increase in FAFSA applications, as demonstrated by Bettinger et al. (2012)), early commitment to free tuition (leading to a 15 pp increase for high-achieving low-income students in a flagship university, as studied by Dynarski et al. (2021)), or offering scholarships (resulting in a 10 pp increase in Colombia's Ser Pilo Paga program, as outlined by Londoño-Vélez et al. (2023)) have shown more substantial effects. When examining the dynamics of the effects using the event study specification, I observe that college enrollment rates increase starting from the first year after the station openings, and the magnitude of the effects remains relatively stable after three years (see Figure 3a). Additionally, the pre-treatment coefficients support the validity of my findings, as they show no evidence of pre-existing trends before the implementation of the new lines.

I also investigate the effects by type of institution, as public and private colleges in this context have different admission systems. Column (3) presents the impact on private

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

college enrollment rates, indicating a 1.3% increase relative to the baseline. In contrast, Column (4) displays the positive impact on public college enrollment, which shows a 0.2% increase; however, this effect is not statistically significant. This aligns with the fact that public colleges in Peru are generally more competitive than private colleges. As a result, reduced commuting time may have a limited impact on public college enrollment due to the strong role of student ability in the admissions process. In contrast, private colleges, which have more flexible admissions processes and the capacity to quickly adapt, can more readily accommodate an influx of new students. Analyzing the dynamic effects within sub-groups reveals that the majority of the impacts stem from enrollment in private colleges, which continue to experience growth over time, as seen in Figure 4a.

Additionally, I explore the effects on enrollment rates by gender. Columns (5) and (6) indicate a positive response for both female and male students, respectively. Nevertheless, the average impact on female enrollment appears to be larger compared to that for male enrollment. When examining the dynamic effects by gender, I observe that the positive impacts on women increase more rapidly over time compared to those on men. This pattern aligns with the fact that the advantages of the new transportation system extend beyond merely reducing commuting time; they also improve travel safety, which is particularly important for women in this context. However, at no point do I find a statistically significant difference between them, as illustrated in Figure 4b.

5.2.1 Robustness Checks

Alternative Specifications and Estimators. Figure 3a presents the event study results using the Borusyak and Hull (2023) estimator and shows no significant trends in the pre-treatment period. I test another functional form, using the hyperbolic sine transformation (Figure 3b), and the results are similar. Furthermore, in line with recent advances in the DiD literature, I use an estimator distinct from that of Borusyak and Hull (2023). Using the Callaway and Sant'Anna (2021) estimator, as seen in Appendix Figure A.11, I find similar

results. The average post-treatment effects show a 10 percent increase.

Sample Including Downtown Lima. In Section 4, I noted that the main specification excludes districts in downtown Lima, as they were highly likely to benefit from any potential transportation line. This area of the city also has greater market access compared to the outskirts. Similar to the results in Table3, adding these districts does not significantly change the results, as seen in Appendix Table A.1. I still observe positive effects on college enrollment rates at the block level, with a 1.4 pp increase. The results from Columns (2)–(6) reveal a similar pattern to the main results in Table 3, albeit with slightly higher magnitudes.

Distance to Stations and Distance to High-Skill Jobs. It is expected that the effects will be higher for students living closer to the new stations. Appendix Figure A.9 confirms that this is the case. Additionally, in Section 4, I noted a potential confounder: students may be motivated to pursue college due to a reduction in commuting time to high-skill job opportunities. Therefore, I would expect the effects to be greater for students living farther from these job locations. However, Appendix Figure A.10 demonstrates that this is not the case; in fact, most of the observed results are driven by students who already live closer to such jobs, showing that the results are driven by reductions in commuting time to college and not commuting time to potential jobs.

5.3 Effects on College Choice

I anticipate that the introduction of new transit lines may influence students along both the extensive margin (whether to enroll in college) and the intensive margin (which college to attend). This expectation is informed by prior evidence showing that students in this context are generally more inclined to enroll in private rather than public institutions (as seen on Section 5.2), as well as by the exogenous variation in geographic accessibility introduced by the connection of certain colleges to the new lines. Importantly, given the wide heterogeneity in the quality of private institutions, it is not evident ex ante whether students will respond by selecting higher-quality institutions or simply opt for colleges that are newly accessible.

Appendix Figure A.12a documents these trends, showing that colleges located near the new stations experienced substantial increases in enrollment following the openings, while non-connected institutions exhibited relatively stable enrollment levels. This pattern suggests that students are responding to changes in accessibility by reevaluating and adjusting their college choices. [29]

First, I estimate the impact of the new transit lines on the likelihood of enrolling in college type m, using individual-level data conditional on enrollment. The empirical strategy mirrors the approach used in the previous enrollment analysis. Specifically, I compare students residing in neighborhoods connected to the newly opened transit lines with those living in neighborhoods along transit lines that were planned-but-not-executed lines, serving as a plausible counterfactual group. Table 4 presents the estimated effects by college type. Column (1) shows that students from neighborhoods connected to the new stations are 9% less likely to enroll in high-quality private colleges, a statistically significant reduction. Column (2) reveals a 19% decline in enrollment in medium-quality private colleges, while Column (3) shows a corresponding 19% increase in enrollment in low-quality private institutions. In contrast, Column (4) indicates a 10% increase in the likelihood of attending a public college, although this estimate is not statistically significant.

Importantly, these results reflect the combined responses of students along both the extensive and intensive margins—those who newly enroll in college due to improved access, and those who would have enrolled regardless but now adjust their choices. To disentangle these mechanisms, the subsequent analysis exploits a machine learning-based prediction model that recovers students' counterfactual enrollment probabilities in the absence of the policy, as detailed in Section 4.2 and Appendix D. Given the lack of standardized admissions data in Peru's decentralized system, I estimate baseline enrollment likelihoods using neighborhood- and student-level pre-treatment characteristics through a k-nearest neighbors

[29] However, Appendix Figure A.12b shows that the number of majors offered by connected colleges did not expand significantly, suggesting that the increase in enrollment was not driven by program diversification but rather by improved access alone.

algorithm with k-fold cross-validation and parameter tuning via *GridSearch*. This procedure yields block-level predicted probabilities θ_i of college enrollment and θ_i^m of enrollment by institution type m (high-, medium-, or low-quality private; public). I use these predictions to stratify the analysis into two subsamples: (i) students from neighborhoods with a high ex ante probability of college enrollment (*typical students*), and (ii) students from neighborhoods with a low ex ante enrollment probability who are plausibly induced to enroll due to the policy (*new students*). This allows me to disentangle students who are at the extensive margin and students who are on the intensive margin.

To distinguish between effects along the extensive and intensive margins, I re-estimate the college choice results separately for the *new student* sample (extensive margin) and the *typical student* sample (intensive margin). Figure 5 presents these heterogeneous effects. Among *typical* students—those with a high ex-ante probability of enrolling in college—I find an increase in the likelihood of attending public colleges, indicating that improved accessibility may lead students who would have enrolled regardless to reoptimize their institutional choices. In contrast, *new* students—those induced to enroll by the new transit connections—exhibit a shift away from high- and medium-quality private institutions toward low-quality private colleges. This pattern raises concerns about the educational and labor market outcomes for marginal students, given the lower expected returns associated with these institutions. Appendix Table A.2 reports the corresponding regression estimates.

5.3.1 Robustness Checks

I adopt an alternative specification to analyze changes in college choice. Instead of modeling the probability of enrolling in college type m, I define the outcome as a binary variable equal to 1 if a student make a different choice relative to predicted probability to enroll in college type m ($\theta_m > 0.5$). Appendix Table A.3 presents the results from this approach, which closely mirror the main findings: students shift enrollment toward lower-quality public and private colleges.

I also examine alternative definitions of college quality. Specifically, I define a college as *elite* if it is part of an elite consortium[30], and as *high quality* if it obtained an operational license following the 2014 Higher Education Reform. The results, shown in Appendix Table A.4, consistently indicate that students are less likely to enroll in either elite or licensed institutions. Columns 1 and 3 test the probability of enrolling in an elite or licensed college, while Columns 2 and 4 assess whether students who were likely to attend such institutions (based on their predicted probabilities) ultimately did so. The findings suggest that students who were initially likely to enroll in elite colleges became less likely to do so. However, this pattern does not hold for students predicted to attend licensed colleges. These analyses serve as robustness checks that reinforce and complement the main results, confirming that the observed shifts in college enrollment are not sensitive to the specific definition of the outcome or college quality.

5.3.2 Heterogeneous Effects by Gender

A particularly salient feature of the results on enrollment is the substantial heterogeneity in responses by gender. As shown in Appendix Figure A.13, female students exhibit a significantly stronger enrollment response to improved access to transportation via newly connected colleges compared to their male counterparts. This differential response underscores the gendered impacts of transportation infrastructure and suggests that reductions in commuting time and gains in perceived or actual mobility disproportionately benefit women. Motivated by this evidence, the subsequent analysis disaggregates the effects by gender to explore underlying patterns in terms of college choice.

Figure 6 presents the results on college choice by gender. The findings indicate meaningful differences in how male and female students—both in the *new* and *typical* student samples—respond to improved access to transportation. Among *typical* female students, there are limited changes in college selection, except for an increase in enrollment at low-

[30]Elite colleges in Peru are defined as those affiliated with the Consorcio de Universidades.

quality private colleges. In the *new* student sample, women exhibit a significant shift away from high- and medium-quality private institutions toward low-quality colleges, with a corresponding decline in enrollment at public institutions. By contrast, *typical* male students respond by reallocating toward public colleges and away from low-quality private institutions. Similarly, *new* male students are more likely to enroll in public colleges and less likely to attend low- or medium-quality private institutions.

These contrasting patterns have important implications for gender disparities in educational outcomes and future earnings. Male students appear to adjust toward educational pathways associated with higher expected wages, while female students—particularly those newly induced to enroll—are more likely to select into institutions with lower labor market payoffs. This may exacerbate existing gender wage gaps, even as overall access expands.[31]

5.4 Mechanisms

Distance to college and connectivity. I first explore whether students who enroll in college are now opting for colleges that are located farther away in terms of distance. The rationale behind this is that since transportation becomes less of an issue, students have higher incentives to travel further and select better colleges that are more distant from their homes. Table A.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 are more likely to enroll in schools that are closer to them by half a kilometer or 5% closer, suggesting they are not taking advantage of traveling longer distances within the city.

I also explore whether students enroll in a college that is connected to the new lines.

[31] I also explore heterogeneity by socioeconomic status (SES) using a coarse proxy based on the 2017 Census. While this measure is available at the block level rather than the household level, it provides suggestive evidence. Figure A.14 shows results by SES, classifying students from neighborhoods below the median income as poor and those above as non-poor. Among *typical* students, SES appears to play a limited role in college choice results (Figure A.14a). However, for *new* students, those from poorer areas are significantly less likely to enroll in high-quality institutions, whereas the increase in enrollment at low-quality colleges is more pronounced among non-poor students.

Table A.5, Column (2) shows that students connected to new lines are 11% more likely to enroll in a college that is connected to the new lines as well. However, when breaking these results down by whether the college is public or private, I find that students are 14% more likely to enroll in private colleges that become connected to new lines, and more likely to enroll in a low-quality private college that is connected, as seen in Column (4). Note that the magnitude increase of 0.4 p.p. increase seems small in absolute terms but represents a large increase for this population relative to the baseline. I also find that students become more likely to enroll in public connected colleges by 10%. These results suggest that college connectivity does play a role among students choices.

Additionally, I examine heterogeneous effects of improved college connectivity by gender for both the new and typical student samples. Figure 7 presents these results. Panels (a) and (b) show that male students are more responsive to improvements in college connectivity, particularly for public institutions. In contrast, female students, on average, do not significantly adjust their enrollment in response to improved connectivity. However, they are more likely to enroll in better-connected private institutions—especially low-quality private colleges, which tend to be located near the geographic center of the city. These patterns highlight how transportation improvements interact with institutional characteristics in shaping college choices and may unintentionally reinforce existing gender disparities in access to higher education.

Adverse environment for women. Given the strong gender differences observed in the results, a potential explanatory factor is the presence of constraints that limit women's mobility across the city—such as safety concerns or restrictive social norms. To the best of my knowledge, granular data on public transportation safety or household-level gender norms is not available for this period (prior to 2010). As a proxy for an adverse environment for women, I use the femicide rate, as described in Section 3. Femicides, the most extreme manifestation of gender-based violence, serve as an indicator of both pervasive safety concerns and deeply rooted discriminatory norms. However, it is important to note that these data

are only available at the district level, which limits the granularity of the analysis.

Figure 8 presents the heterogeneity in college choice by gender and local femicide rates, comparing students living in districts with relatively high versus low exposure. Panel (a) shows the results for *typical* students (those with a high baseline probability of college enrollment), where no significant differences emerge based on femicide exposure. In contrast, Panel (b) focuses on the *new* student sample—those marginal students induced into college by the policy—and reveals that female students residing in low-femicide districts (i.e., low-risk areas) are significantly more likely to enroll in low-quality private colleges. As mentioned before, these institutions tend to be more centrally located and offer a more accessible alternative for women facing mobility and safety constraints. Among women in high-risk areas, the estimated effects are smaller but still statistically significant suggesting that these females students suffer from limited mobility. While these patterns are consistent with the hypothesis that safety concerns shape educational decisions, it is important to note that the differences in effects between high- and low-risk areas are not statistically significant. Therefore, the results should be interpreted as suggestive rather than conclusive. Nevertheless, this also echoes the findings of Borker (2020), who documents that women in Delhi tend to enroll in colleges in the lower half of the quality distribution due to safety concerns, while men prioritize institutional quality. This highlights the broader implications of safety and social norms for women's educational and career outcomes, and underscores the need for policy interventions aimed not only at improving safety but also at expanding inclusion in selective institutions.

Interestingly, male students in high-risk districts are more likely to enroll in public universities. This suggests that men living in these areas may be more willing to travel and choose a riskier path in order to access public higher education, which often involves more uncertainty. This gendered response to women-safety conditions underscores the different ways in which mobility constraints shape educational trajectories, for both male and female students.

College Characteristics. Potentially certain college characteristics may help explain the gender differences observed in college choice decisions. One mechanism is the share of female students enrolled. Figure 9a Panel (a) shows that female students are less likely to enroll in institutions with a low proportion of female students, whereas the opposite pattern is observed for male students. Notably, one of the few public colleges connected to the new transit lines—the National University of Engineering—has an enrollment that is 80% male, further emphasizing this trend. In addition to enrollment composition, I examine major choice mechanisms. Since students make simultaneous decisions about both college and major, the availability of specific programs can significantly influence their academic trajectories. To investigate this further, I focus on the choice of STEM majors, given their association with higher post-graduation earnings. Figure 9b Panel (b) illustrates that, among both typical and new students, women are less likely to enroll in STEM programs, while men are more likely to do so. Note that public schools tend to be more selective in admitting students into STEM programs, making them competitive and potentially riskier choices. In contrast, private institutions often have the flexibility to expand non-STEM enrollment and are generally less selective, which may make them more attractive to female students. Together, these results show that college characteristics—ranging from the gender composition of student bodies to the availability of specific academic programs— play a critical role.

Overall, the patterns suggested by these mechanisms indicate that female students navigate an educational landscape shaped both by the adverse environments they face and by accessibility, both in terms of commuting and institutional factors.

6 Medium-Term Effects

In this section, I examine the medium to long-term effects of residing in neighborhoods that become connected to newly opened lines. To do so, I implement a cohort-exposure difference-in-difference analysis using data from the 2017 National Census, conducted seven years after

the initial line opening, as described in Section 4.3.

6.1 College Completion

When assessing medium-term impacts, the new transportation lines are expected to improve college access, enhance students' overall academic experience, and increase the likelihood of timely graduation. These effects can operate through two main channels: (i) reduced commuting time, which can improve academic performance and persistence among students, as documented by Tigre et al. (2017); and (ii) expanded access to internship opportunities, which are often mandatory for graduation in several degree programs.

Due to data limitations on the specific mechanisms, I estimate the overall reduced-form effect using the 2017 National Census—seven years after the first line openings—leveraging a difference-in-differences strategy based on variation in cohort exposure. The results, presented in Table 7, Column (1), show a positive but statistically insignificant effect on college completion rates.

Disaggregating the results by student type (new versus typical) and gender, as shown in Figure 11, reveals important heterogeneity. There is no effect among *typical* students. However, among *new* students—those coming from neighborhoods that typically do not sent students to college—female students are more likely to complete college relative to their male peers, who actually exhibit a decrease in completion likelihood by 2017.

This gendered pattern may reflect shifts in college choice induced by improved connectivity. As shown earlier, women are increasingly enrolling in lower-quality private institutions rather than public colleges. While this may lead to earlier graduation, it raises concerns about educational quality and labor market outcomes. In contrast, men—who are more likely to attend public institutions—may take longer to graduate due to more rigorous academic requirements or the tuition-free nature of public education, which often encourages extended enrollment. Overall, these findings underscore that infrastructure improvements can differentially affect students' educational trajectories.

6.2 Employment Rates

Building on the previous cohort-exposure strategy, I next examine the effects of the new transportation stations on employment outcomes, using data from the 2017 Peruvian Census. Table 7 columns (2-4) shows that access to new lines is associated with a 1 percentage point increase in employment rates among students who enrolled in college, relative to baseline levels. When disaggregating by job quality (distinguishing between blue-collar and white-collar occupations) the results reveal that this employment gain is primarily driven by an increase in white-collar employment.

Further heterogeneity analysis by student type (new vs. typical students) and gender, presented in Figure 11, reinforces the earlier narrative. First, among *typical* students (Panel A), I see a significant reduction on employment among female students and specially blue-collar jobs. Among *new* students (Panel B), males exhibit higher employment rates compared to their female peers. However, the composition of employment differs notably: men are more likely to be employed in blue-collar jobs, while women show a shift from blue-collar toward white-collar employment, consistent with their higher rates of college completion documented in the previous section.

Taken together, these results suggest that, even though female students are increasingly selecting lower-quality private institutions, they are more likely to complete college and transition into higher-quality employment. This pattern highlights a meaningful tradeoff: improved access may lead some marginal female students to enroll in institutions of lower academic quality, yet the resulting degrees still enable entry into white-collar occupations.

That said, these findings should be interpreted with caution. The 2017 Census captures outcomes for only the earliest cohorts exposed to the new infrastructure. As male students—who are more likely to attend tuition-free public institutions with longer degree completion times—begin to graduate in larger numbers, the observed gender differences in employment and job quality may evolve over time. Continued tracking of later cohorts will be essential to understanding the full labor market impact of access to transportation.

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

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

To understand such differences, I estimate key parameters such as a student's willingness to pay for commuting time in terms of average wages (when attending school c). I outline a simple model of college choice to recover students' preferences,[32] 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

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

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}, the average wages of graduating from college c;

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

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

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

I am interested in measuring the trade-off between college quality and travel time to college. Similar to Borker (2020), 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 average wages. This relationship is represented with the following equation:

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

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

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

Given my interest in assessing distinct responses from both female and male students as suggested by the reduced-form results, I proceed to estimate this model separately by gender. Additionally, I conduct separate estimations based on the baseline likelihood that students go to either private or public colleges since these two types of colleges attract different types of students. Public colleges in Peru typically impose more stringent admission criteria, attracting high-ability students who cannot afford private universities. In contrast, selection into private universities is mostly explained by the household's ability to cover the tuition fees rather than student ability. To help with interpretation, I standardize the average wages to have a mean of zero and a standard deviation of one. I measure travel time from home to college in minutes as described in Section 3.

7.1 Results

Table 6 presents the results from the mixed logit model. Column (1) reports estimates for the full sample, showing that students generally prefer colleges that are closer to home and associated with higher average earnings. Columns (2) and (3) show estimates for female students with a high baseline probability of enrolling in a private or public college, respectively, while Columns (4) and (5) present analogous results for male students.

Across all groups, students exhibit a significant disutility from commuting time, with male students being slightly more sensitive, particularly those likely to attend public institutions. In terms of average earnings, male students place greater weight on wages than their female counterparts. Among female students, those likely to attend private colleges respond more strongly to wage expectations, whereas responsiveness is notably lower among those likely to attend public institutions. A similar but less pronounced pattern emerges for male students.

Estimates of the standard deviations capture heterogeneity in preferences. There is statistically significant heterogeneity in how students value commuting time, especially among students likely to attend public colleges (both female and male). In contrast, preferences for average earnings display minimal heterogeneity across individuals, suggesting that most students value higher wages similarly, even if the strength of their response differs.

These patterns suggest important policy implications. First, reducing commuting burdens through improved transportation infrastructure could improve access, especially for students sensitive to distance. Second, the muted responsiveness to average earnings among women likely to attend public colleges may reflect non-financial constraints such as safety concerns, limited mobility, or social norms. Addressing these barriers through targeted scholarships, safe and reliable transit, or outreach that improves the salience and feasibility of higher-return college options—could help students make more informed and economically beneficial decisions. More broadly, policies that expand perceived and actual agency in college choice may enhance the responsiveness to quality signals like earnings, particularly among female students.

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

willing to commute less.

On the other hand, Figure 10b Panel (b) shows that the Marginal Rate of Substitution (MRS) for the sample of student that are likely to enroll in public schools. First, they are in general less willing to commute for higher wages. The difference by gender is statistically significant but in a higher magnitude compare to students that have a high probability of attending private college at baseline. Male students show a 126% higher willingness to travel compared to women (7 more minutes), measured in minutes per one standard deviation in average wages. The results in Section 5.3 show that only men take advantage of the reduced commuting time to attend public schools. Is it important to highlight that among the public colleges that become connected is the National University of Engineering, which mostly attracts male students, while others like the National University of Education do not become connected and attract more female students. In this sense, since public colleges are more specialized, they are also only attracting students who are likely to enroll in such majors. The STEM differences emerge quite strongly as women tend to stay away from such majors.

8 Conclusion

This paper studies the relationship between intracity transportation and higher education, making it the first exploration of the causal effects of improved public transportation on both college access and college choice. While the literature on education 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 their decision to attend college but also *which* college.

The focus on college education is particularly relevant in the developing world, where access to K-12 education is becoming universally available and there is an increasing demand for higher education. The results also resonate with large cities in the developing world, where

access to college, especially access to high-quality colleges for poor students, is still limited. As nations actively re-design and re-imagine their transportation systems, it is important to understand the economic consequences of inefficient, unsafe, or unreliable transportation. Recent studies in the literature highlight the importance of having the most efficient transportation routes within urban areas. For instance, Kreindler et al. (2023) show that adopting a less concentrated network can improve commuter well-being in cities like Jakarta. However, incorporating access to college education, a crucial driver of an individual's productivity, can be integrated into the planning of optimal transportation networks within cities.

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

The results also hold particular significance in the context of cities where disparities in mobility by gender are especially pronounced. I show that men and women have different responses in terms of their college choices. Women appear to be more inclined toward choosing lower-quality colleges that are conveniently located along new transportation routes, suggesting a willingness to compromise on the quality of education for the sake of a shorter and easier commute. Using a random utility model, I quantify these trade-offs and show that compared to women, men are willing to travel more minutes to attend a college where graduates earn higher salaries. 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. Choosing lower-return colleges can impede their progress in breaking through the already existing glass ceiling.

References

Acton, R., Cortes, K. E., Miller, L., and Morales, C. (2024). Distance to degrees: How college proximity shapes students enrollment choices and attainment across race-ethnicity and socioeconomic status. (1055).

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

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

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

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

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

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

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

Black, S. E., Denning, J. T., and Rothstein, J. (2023). Winners and Losers? The Effect of

Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes. *American Economic Journal: Applied Economics*, 15(1):26–67.

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

Borker, G. (2024). Understanding the constraints to women's use of urban public transport in developing countries. *World Development*, 180:106589.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting Event-Study Designs: Robust and Efficient Estimation. *The Review of Economic Studies*, page rdae007.

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

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

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

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

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

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

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

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

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

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

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

Fiala, N., Hernandez, A. G., Narula, K., and Prakash, N. (2025). Wheels of Change: Transforming Girls' Lives with Bicycles. (Forthcoming).

Field, E., Garlick, R., and Vyborny, K. (2025). Women's Mobility and Labor Supply: Experimental Evidence from Pakistan . *Working paper*.

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

Gonzalez-Navarro, M. and Zarate, R. D. (2023). Land transport infrastructure. *VoxDevLit*, 9.

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

JICA (2013). Data Collection Survey on Urban Transport for Lima and Callao Metropolitan Area, Final Report. Technical report, Japan International Cooperation Agency (JICA).

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

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

Lee, K. H. and Tan, B. J. (2024). Urban Transit Infrastructure and Inequality. *The Review of Economics and Statistics*, pages 1–46.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Main Figures and Tables

Figure 1: Decision Tree of College Choice

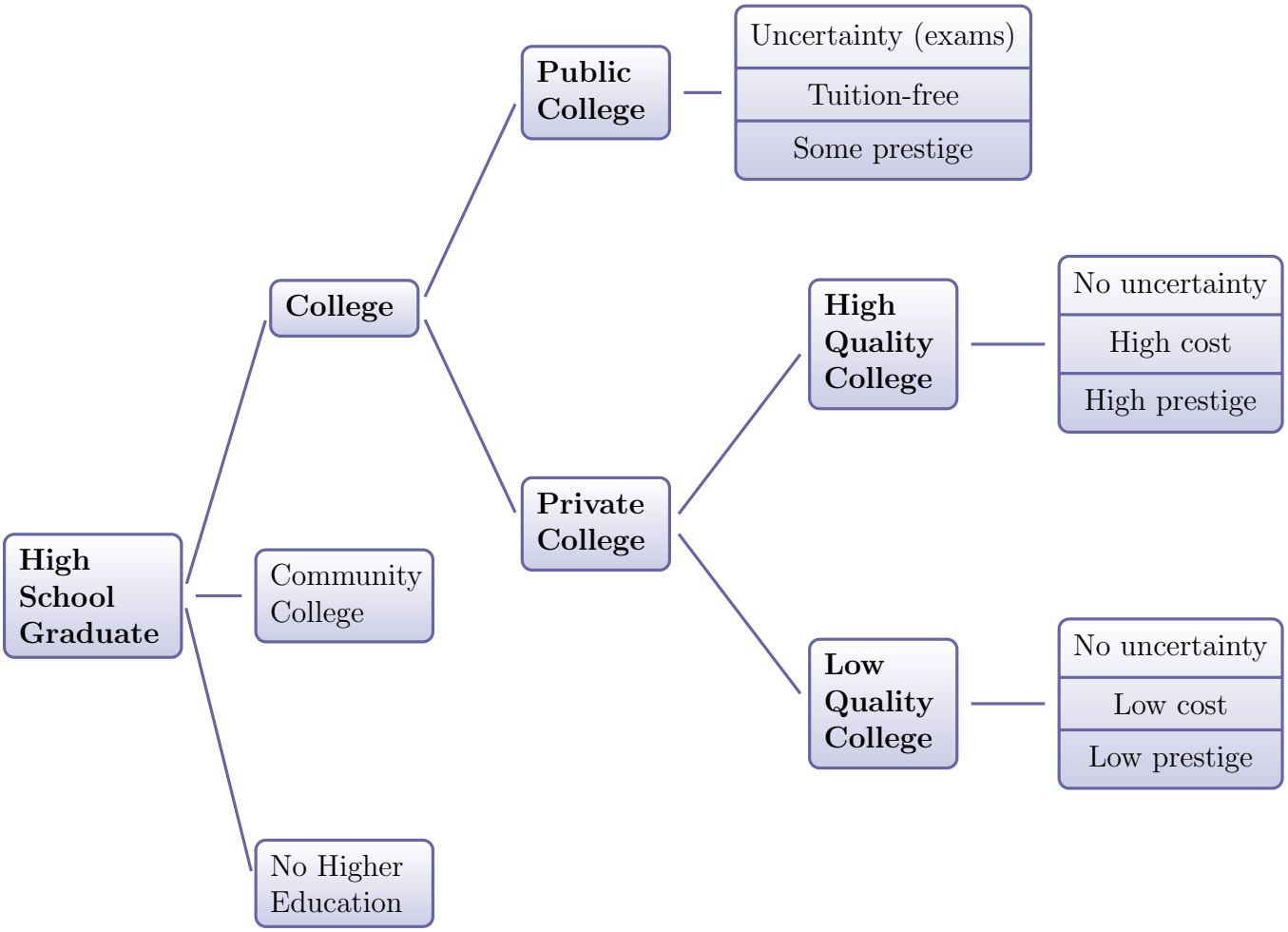
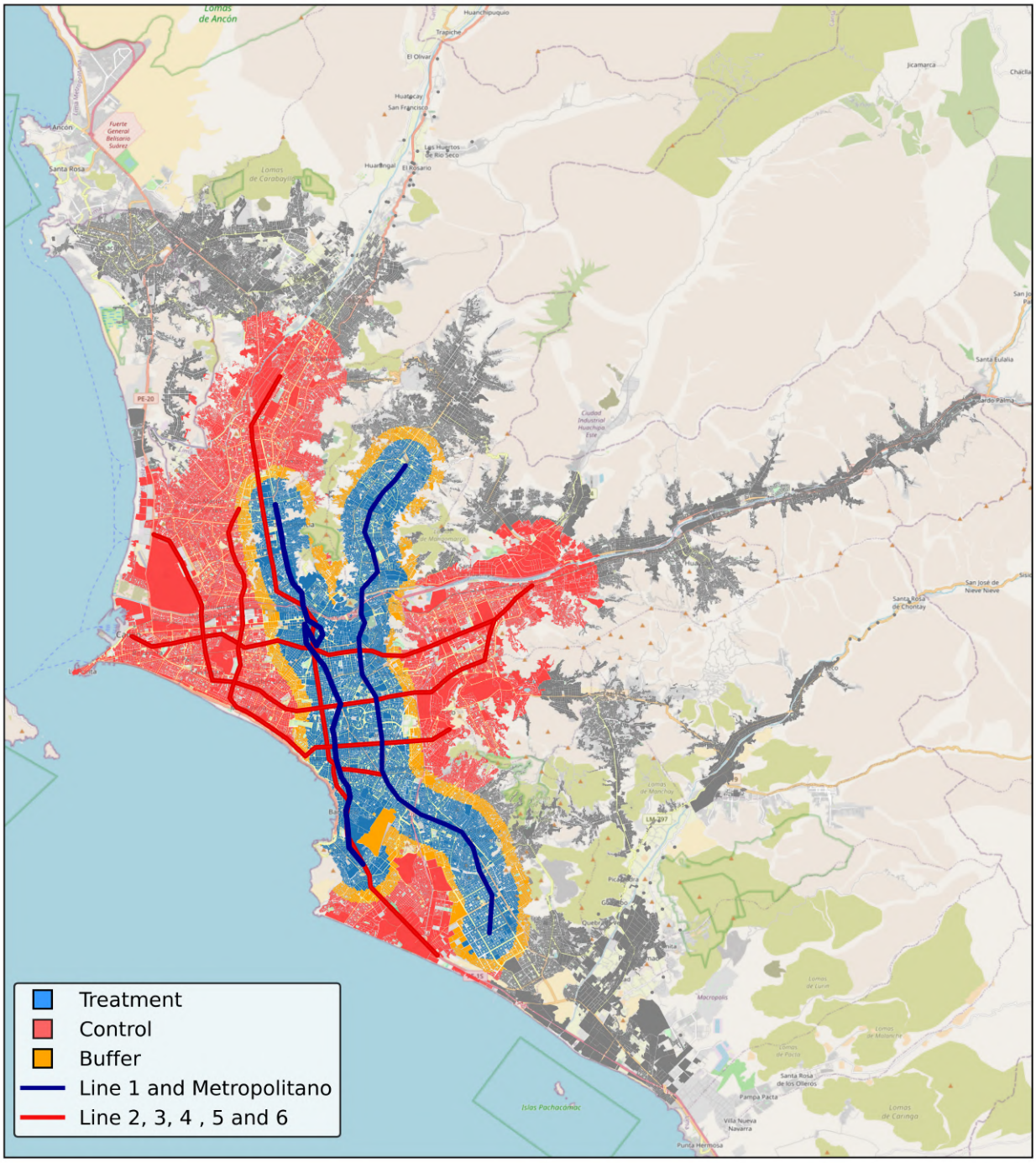


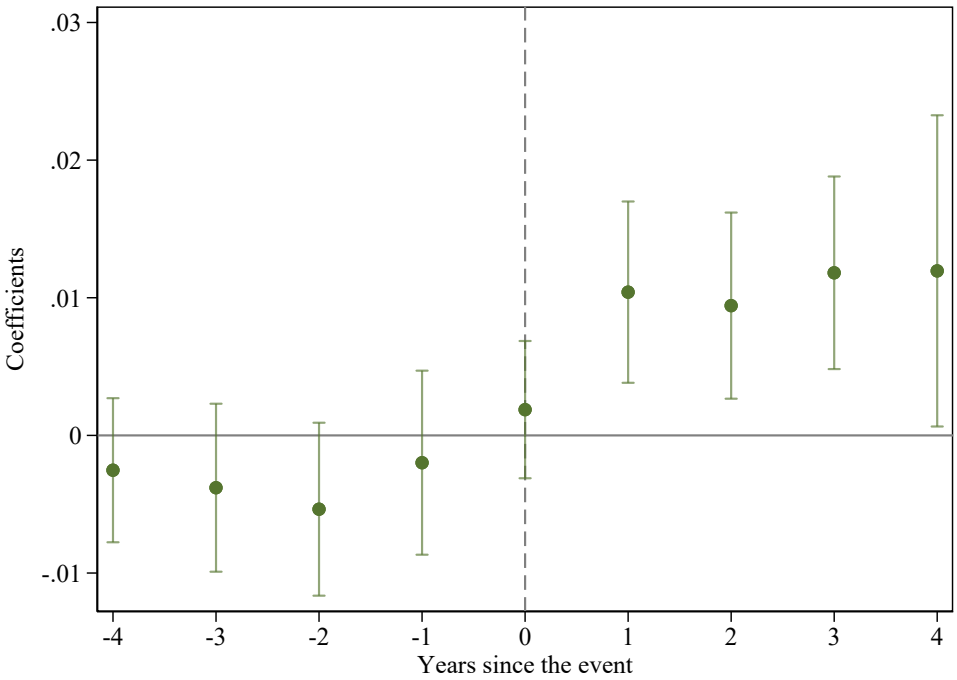
Figure 2: Neighborhoods Exposed to Executed and Planned-but-not-executed Lines



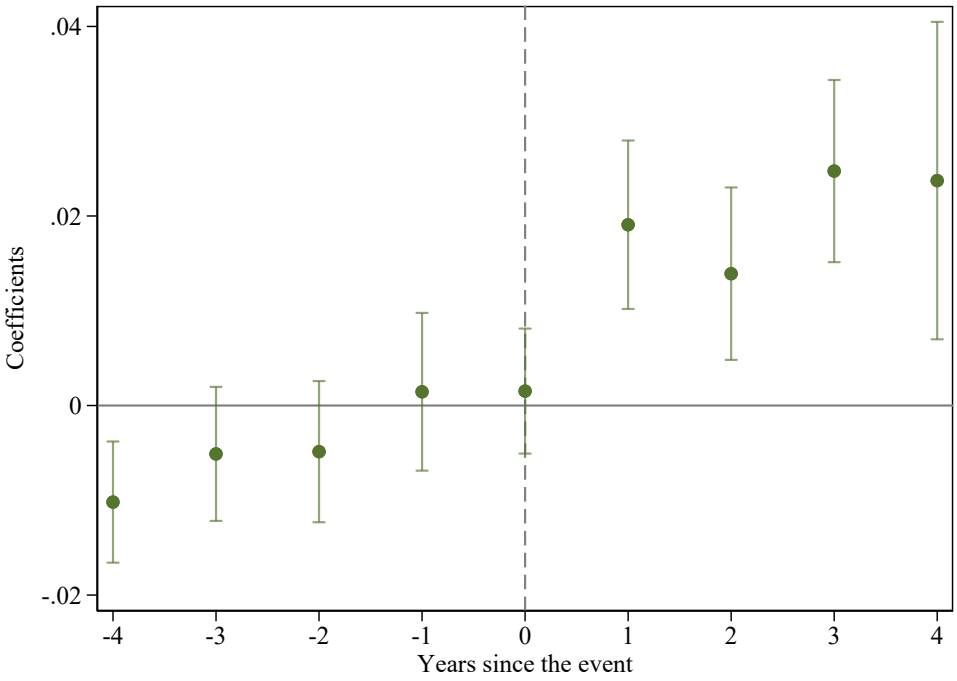
Notes. Blue-shaded areas are neighborhoods within *1.6 km* or 1 mile distance from the nearest executed stations while red-shaded areas are neighborhoods within 4 km distance from the planned-but-not-executed stations. Yellow-shaded areas are 0.5 km buffer zones that are excluded from the main sample.

Figure 3: Dynamic Effects of College Enrollment Rates

(a) Enrollment Rates



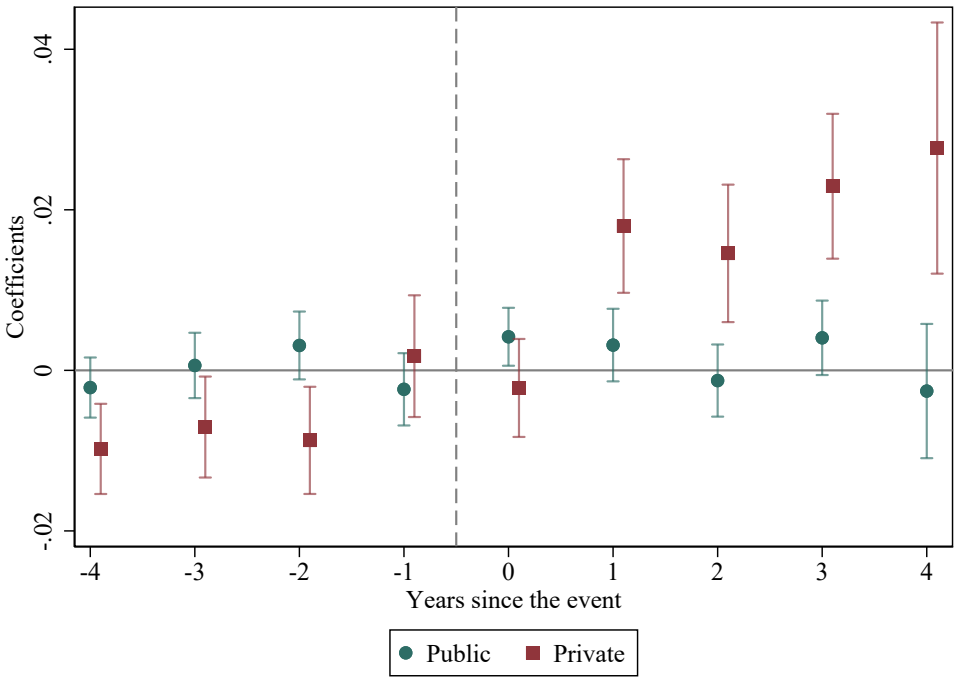
(b) Log(Enrollment Counts)



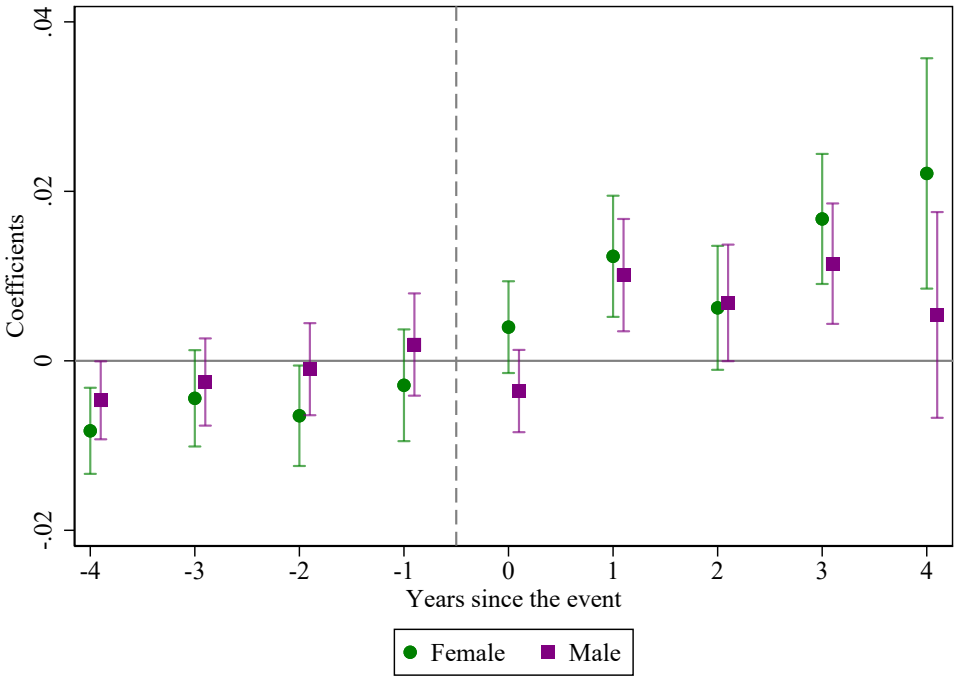
Notes. Regressions include block and year-fixed effects. The event study estimates are calculated using the Borusyak et al. (2024) estimator.Errors are clustered at the block level. Panel (a) shows the enrollment rates where the denominator is the total count of students enrolled divided by the total population of (potential) students of the same cohort. Panel (b) shows the logarithm transformation of enrollment counts using the hyperbolic sine.

Figure 4: Dynamic Effects of New Lines on College Enrollment by Groups

(a) By college type

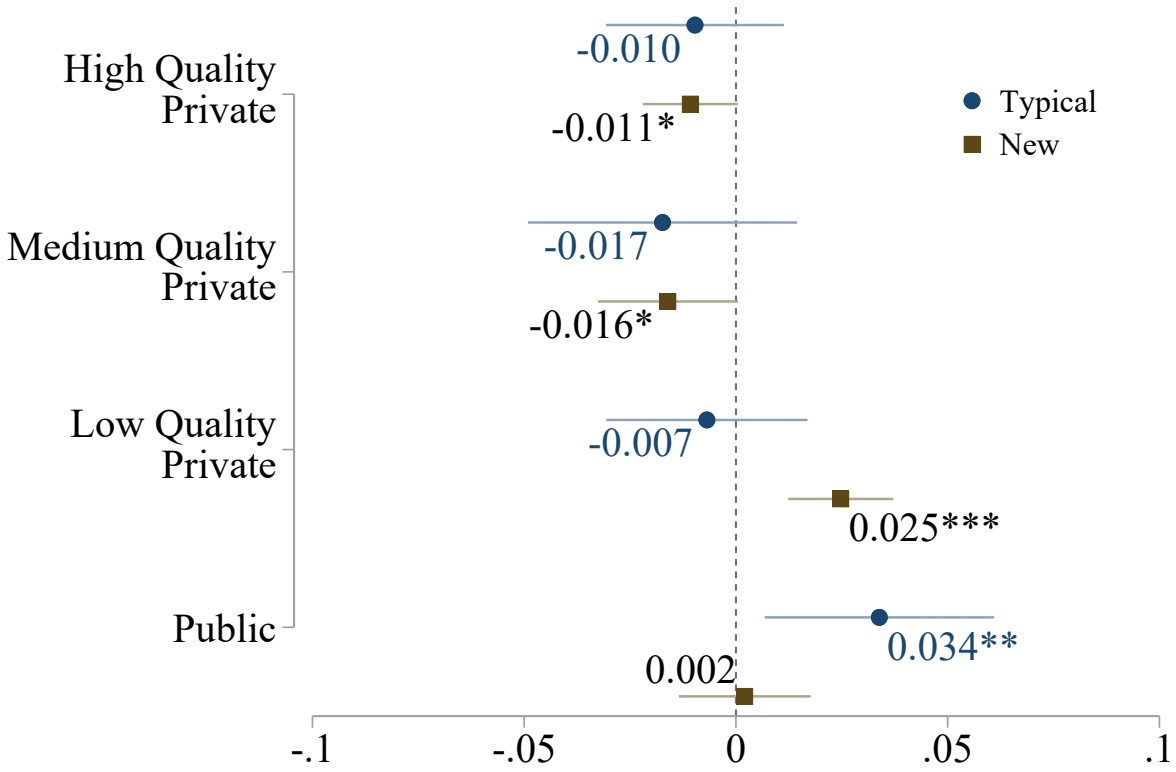


(b) By gender



Notes. Regressions include block and year fixed effects. Standard errors are clustered at the block level. The event study is calculated using the Borusyak et al. (2024) estimator. Panel (a) and Panel (b) shows the logarithm transformation of enrollment counts.

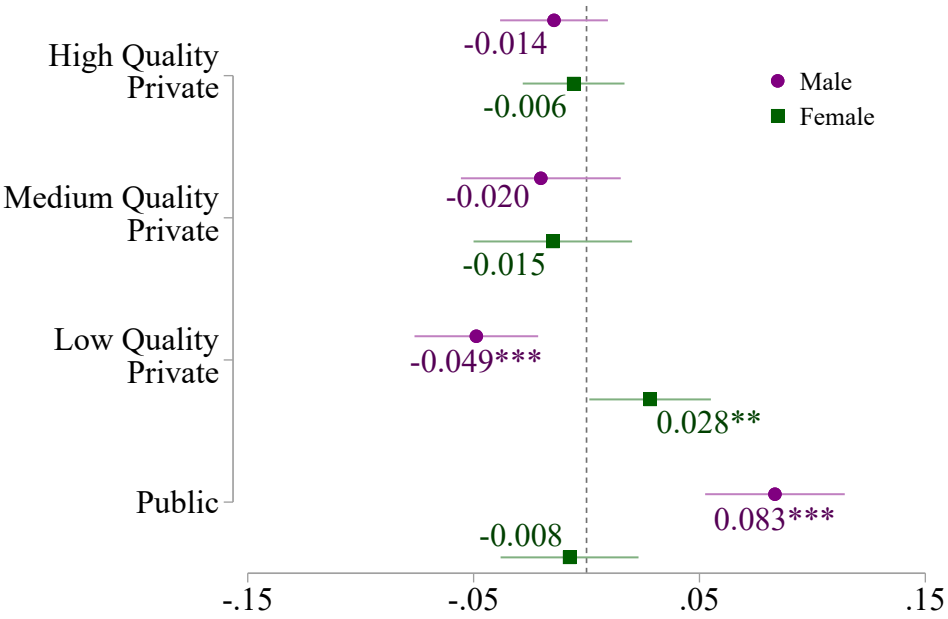
Figure 5: Effects of New Lines on College Enrollment Type by Access to College



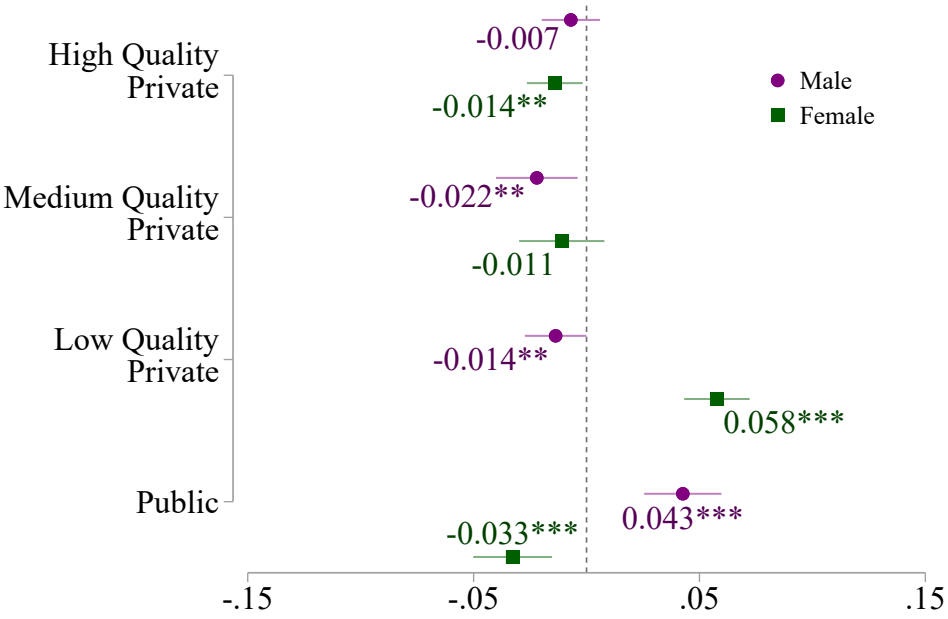
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. Standard errors are clustered at the block level. High-quality private colleges are all private colleges whose graduates earn more than 2200 PEN while Low-quality private colleges are the ones whose graduates earn less than 1450 PEN using administrative data from wage records in 2014. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low.

Figure 6: Effects of New Lines on Prob. of Enrolling in College Type m by Gender

(a) *Typical Students*



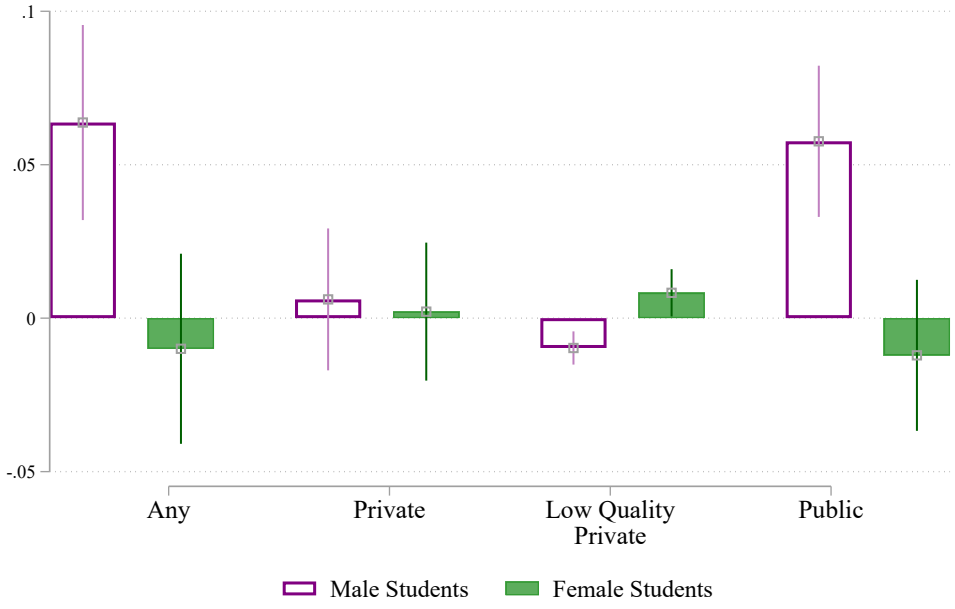
(b) *New Students*



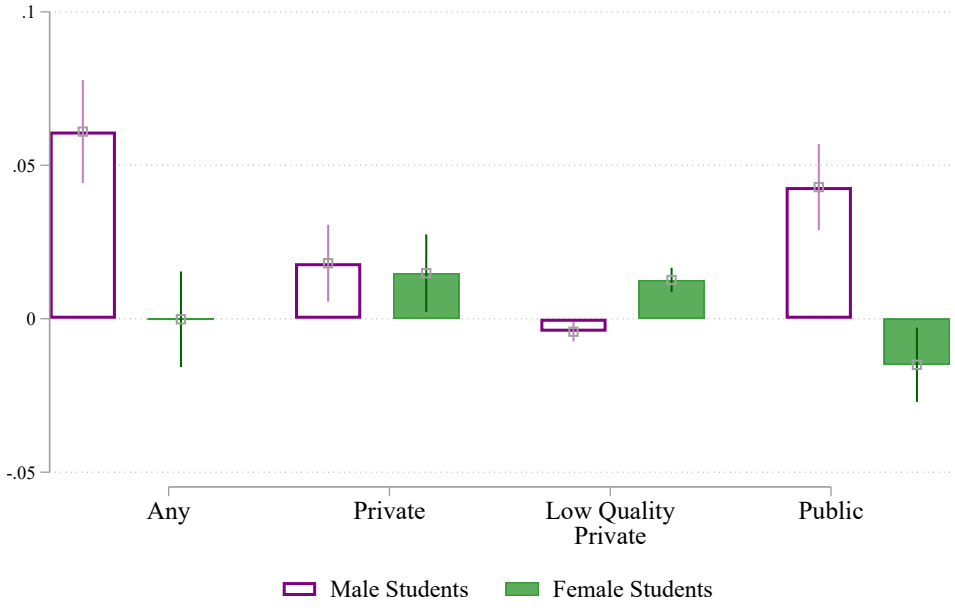
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. Standard errors are clustered at the block level. High-quality colleges are all private colleges whose graduates earn more than 2200 PEN, low-quality private colleges are the ones whose graduates earn less than 1450 PEN, and medium-quality private colleges are those whose graduates earn in between. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low.

Figure 7: Effects of New Lines on College Connectivity by Access and Gender

(a) *Typical students*



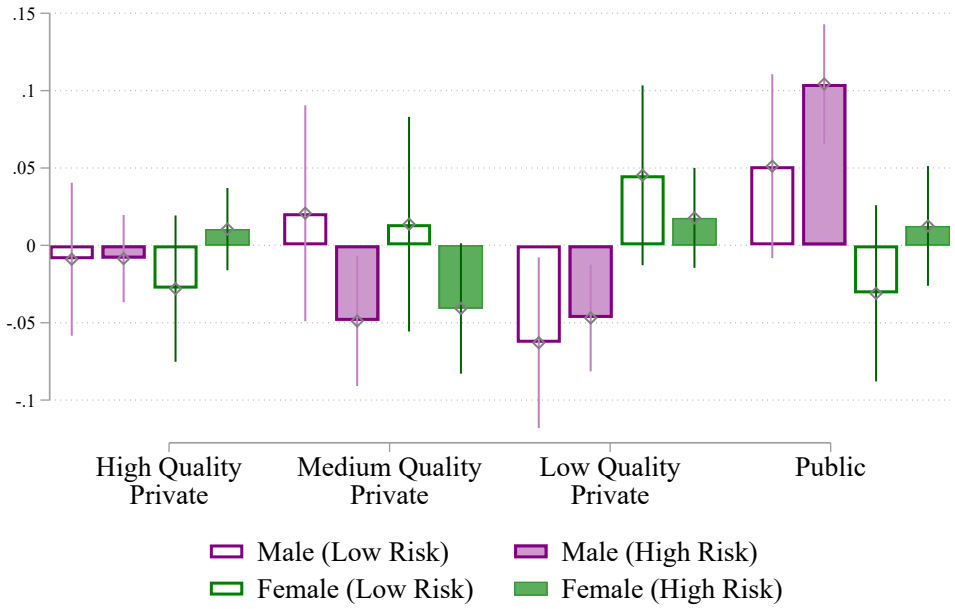
(b) *New students*



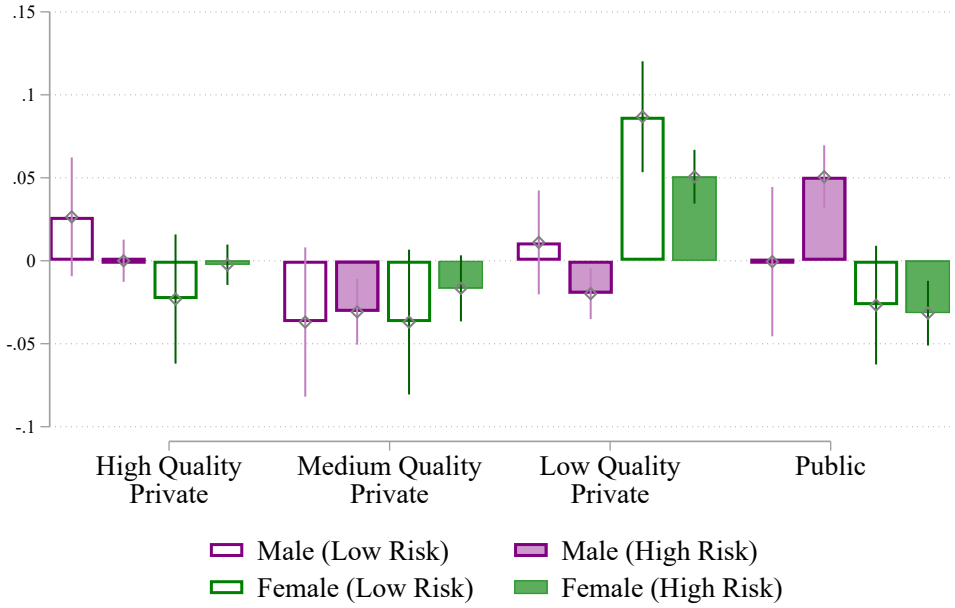
Notes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions include block and year fixed effects. Standard errors are clustered at the block level. High-quality private colleges are those whose graduates earn more than 2200 PEN; low-quality private colleges are those whose graduates earn less than 1450 PEN; and medium-quality private colleges are those in between. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low.

Figure 8: Effects of New Lines on Enrollment by College Type m by Gender and Risk

(a) *Typical students*



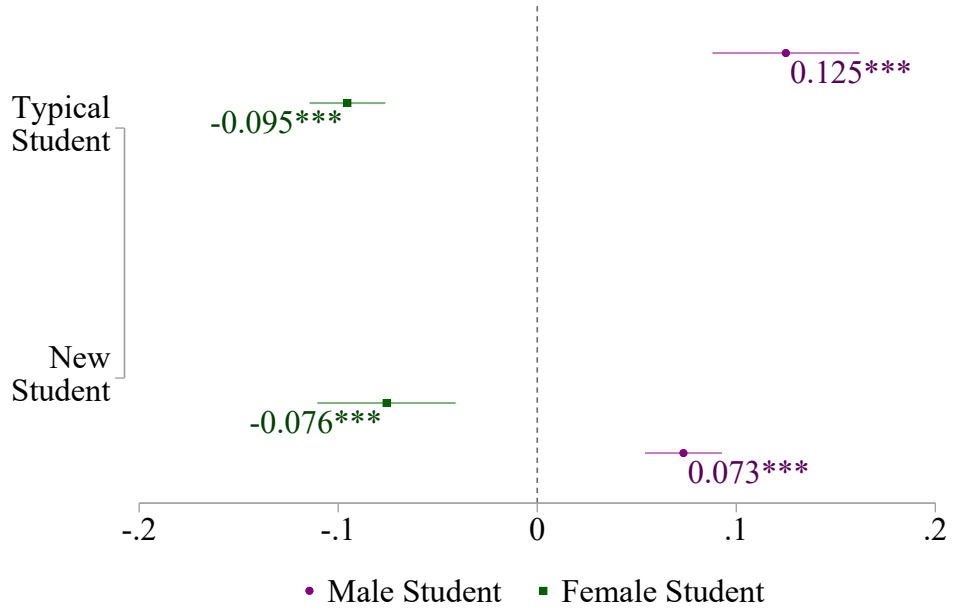
(b) *New students*



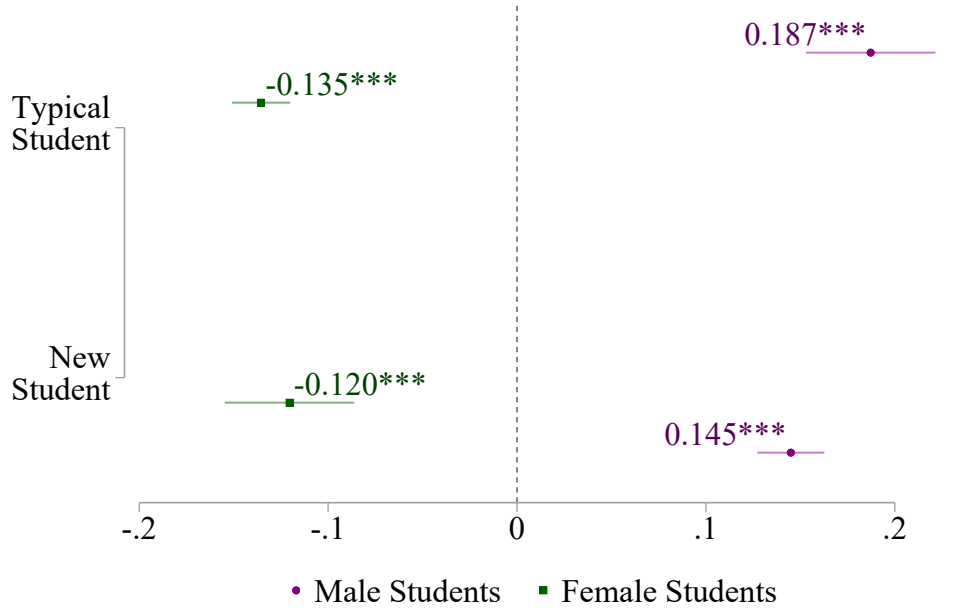
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. Standard errors are clustered at the block level. High-quality colleges are all private colleges whose graduates earn more than 2200 PEN, low-quality private colleges are the ones whose graduates earn less than 1450 PEN, and medium-quality private colleges are those whose graduates earn in between. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low. Low-risk districts are defined as those with the lowest rates of femicide, whereas high-risk districts are characterized by high femicide rates.

Figure 9: Effects of New Lines on College Enrollment for Schools with Low Share of Female Students and STEM Major by Gender

(a) *College with a Low Share of Female Students*



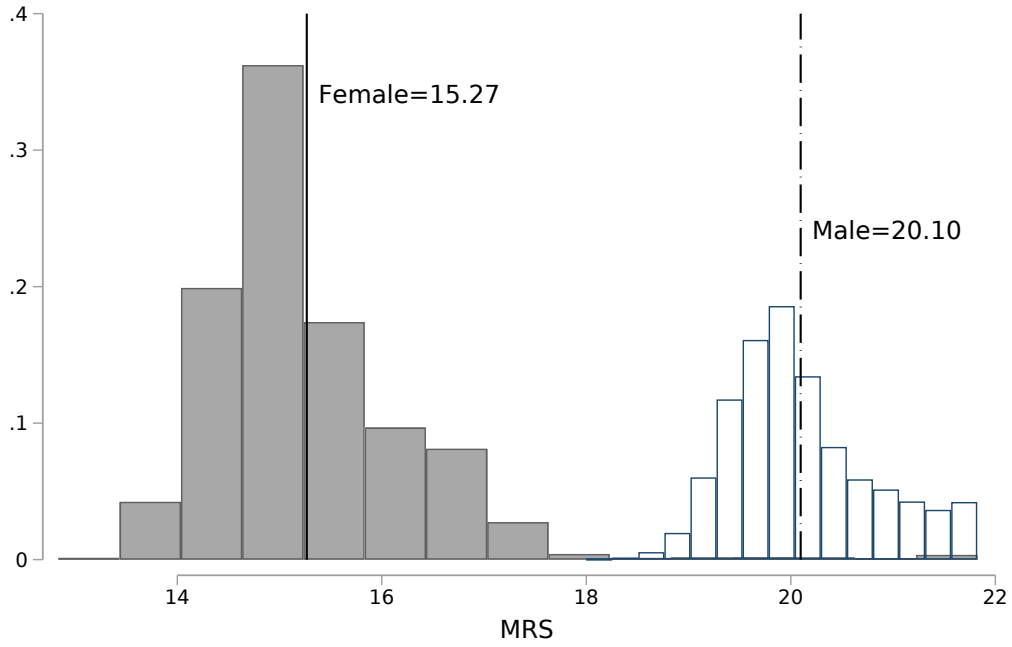
(b) *STEM Major*



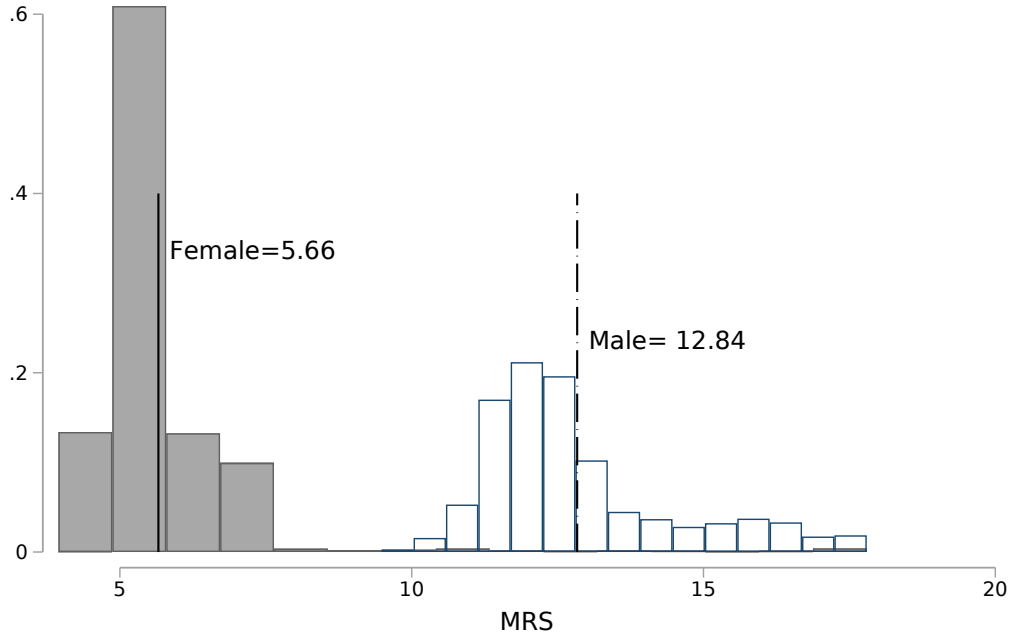
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. Standard errors are clustered at the block level. High-quality colleges are all private colleges whose graduates earn more than 2200 PEN, low-quality private colleges are the ones whose graduates earn less than 1450 PEN, and medium-quality private colleges are those whose graduates earn in between. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low.

Figure 10: Marginal Rate of Substitution (MRS) by Gender

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



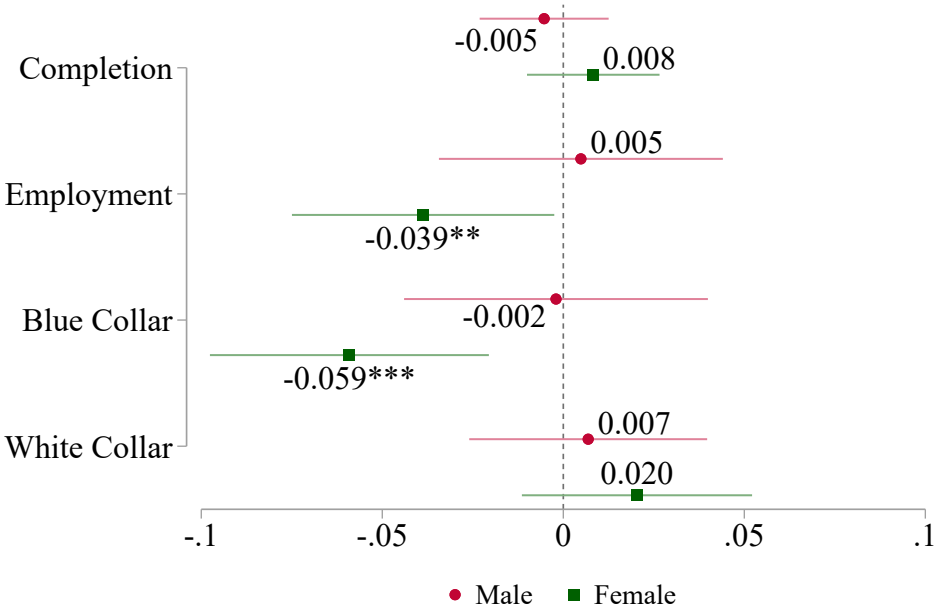
(b) Students with High Prob. of Attending Public Colleges



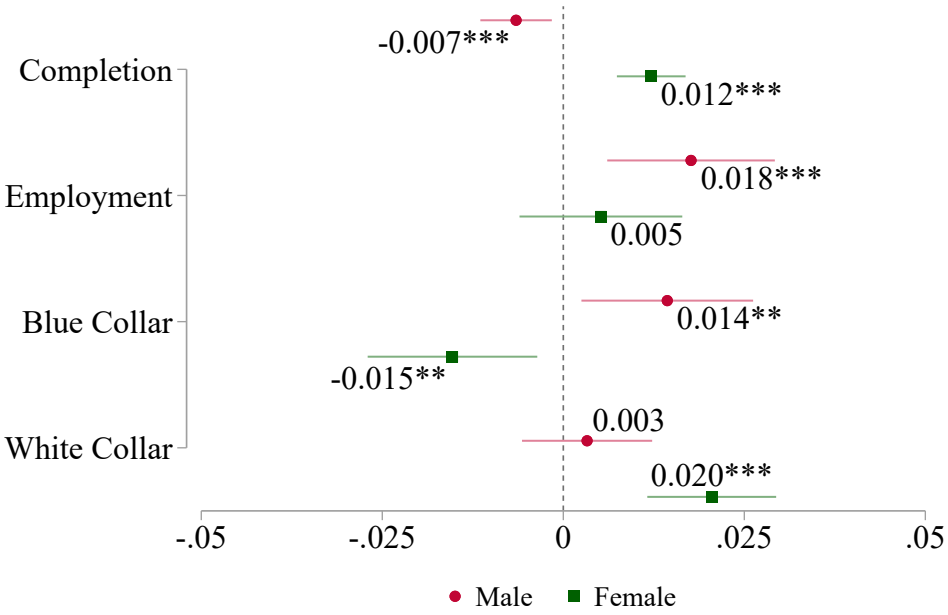
Notes. MRS distributions are winsorized at the 1% and 99% percent.

Figure 11: Effects of New Lines on College Completion and Employment by Access and Gender

(a) *Typical Students*



(b) *New Students*



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and cohort-fixed effects. Standard errors are clustered at the block level. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low.

Table 1: Baseline Summary Statistics

	Total	Treatment	Control
	(1)	(2)	(3)
Panel A :	College Admin. Records Data		
Enrollment (rates)	0.137	0.133	0.141
	(0.323)	(0.320)	(0.327)
Enrollment (log)	0.119	0.113	0.125
	(0.376)	(0.357)	(0.393)
Female Enroll. (logs)	0.0699	0.0657	0.0738
	(0.284)	(0.268)	(0.298)
Male Enroll. (logs)	0.0600	0.0558	0.0639
	(0.261)	(0.245)	(0.275)
Public Enroll. (logs)	0.0348	0.0333	0.0362
	(0.198)	(0.189)	(0.207)
Private Enroll. (logs)	0.0920	0.0858	0.0978
	(0.330)	(0.310)	(0.347)
Distance HH to College	1.978	2.030	1.930
	(4.552)	(4.546)	(4.557)
Panel B:		2007 Census	
Total Population	149.8	153.1	146.7
	(131.7)	(148.1)	(114.5)
Primary School	0.182	0.191	0.175
Pop. Share	(0.128)	(0.128)	(0.128)
Secondary School	0.374	0.384	0.365
Pop. Share	(0.162)	(0.158)	(0.164)
Higher Ed.	0.443	0.425	0.460
Pop. Share	(0.226)	(0.222)	(0.229)
Observations	127,104	60,780	66,324

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 and Census counts. Logarithmic transformations use 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 Lines on Average Commuting Time (mins) to College

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

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include district and year-fixed effects. Standard errors are clustered at the district level. Column (2) excludes districts in downtown Lima. Column (3) shows the effects on commuting time to private colleges only while Column (4) shows the effects on public institutions.

Table 3: Effects of New Lines on College Enrollment

	Rate	Log(All)	Log(Private)	Log(Public)	Log(Female)	Log(Men)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment*Opening	0.008***	0.014***	0.013***	0.002	0.010***	0.005**
	(0.002)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Dep. Var. Mean	0.115	0.121	0.093	0.035	0.071	0.061
N	278874	278874	278874	278874	278874	278874
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. (2024) estimator for the first 4 years after a station opening. Standard errors are clustered at the block level. 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 of New Lines on College Choice

	High Quality	Medium Quality	Low Quality	Public
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.009*	-0.019**	0.019***	0.010
	(0.005)	(0.008)	(0.006)	(0.007)
Dep. Var. Mean	0.116	0.400	0.156	0.329
N	100687	100687	100687	100687
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and year-fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening. Columns 1-3 show the results for private colleges. High-quality private colleges are those whose graduates earn more than 2200 PEN; low-quality private colleges are those whose graduates earn less than 1450 PEN; and medium-quality private colleges are those in between.

Table 5: Effects of New Lines on Distance to College and College Connectivity

	Distance (km.)	Connected College	Private Connected	Low Quality Priv. Connected	Public Connected
	(1)	(2)	(3)	(4)	(5)
Treatment*Opening	-0.465***	0.027***	0.014***	0.004***	0.013**
	(0.080)	(0.006)	(0.005)	(0.001)	(0.005)
Dep. Var. Mean	9.744	0.236	0.103	0.002	0.133
N	100687	100687	100687	100687	100687
Block FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Distance is measured in kilometers using the Euclidean distance from a block centroid to a college location. Columns 2-5 measure the probability of enrolling in a college connected to the new lines, with the closest station less than 3 km. away. Columns 3 and 4 measure the probability of enrolling in a private and low-quality private college connected to new lines, respectively. Column 5 measures the probability of enrolling in a public college connected to a new line. All regressions include block and year-fixed effects. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening. Standard errors are clustered at the block level.

Table 6: Mixed Logit Results

	All	Female		Male	
		Prob(Public)	Prob(Private)	Prob(Public)	Prob(Private)
	(1)	(2)	(3)	(4)	(5)
Mean					
Commuting Time	-0.017***	-0.015***	-0.015***	-0.018***	-0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Expected Wages	0.270***	0.229***	0.080***	0.369***	0.243***
	(0.002)	(0.003)	(0.006)	(0.003)	(0.005)
SD					
Commuting Time	0.008***	0.006***	0.009***	0.005***	0.010***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Expected Wages	-0.000*	-0.000	-0.001	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Students	99,936	45,520	7,774	33,674	12,968
Observations	3497760	1593200	272090	1178590	453880
Log-Likelihood	-341257.7	-156713.2	-27017.8	-112778.9	-44159.4

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Standard errors in parentheses. Columns (2) and (4) shows results for those who have a high probability of enrolling in a private college and Columns (3) and (5) show the results for the sample of students with a high probability of enrolling in a public college. Expected wages are standardized to mean zero. Robust standard errors.

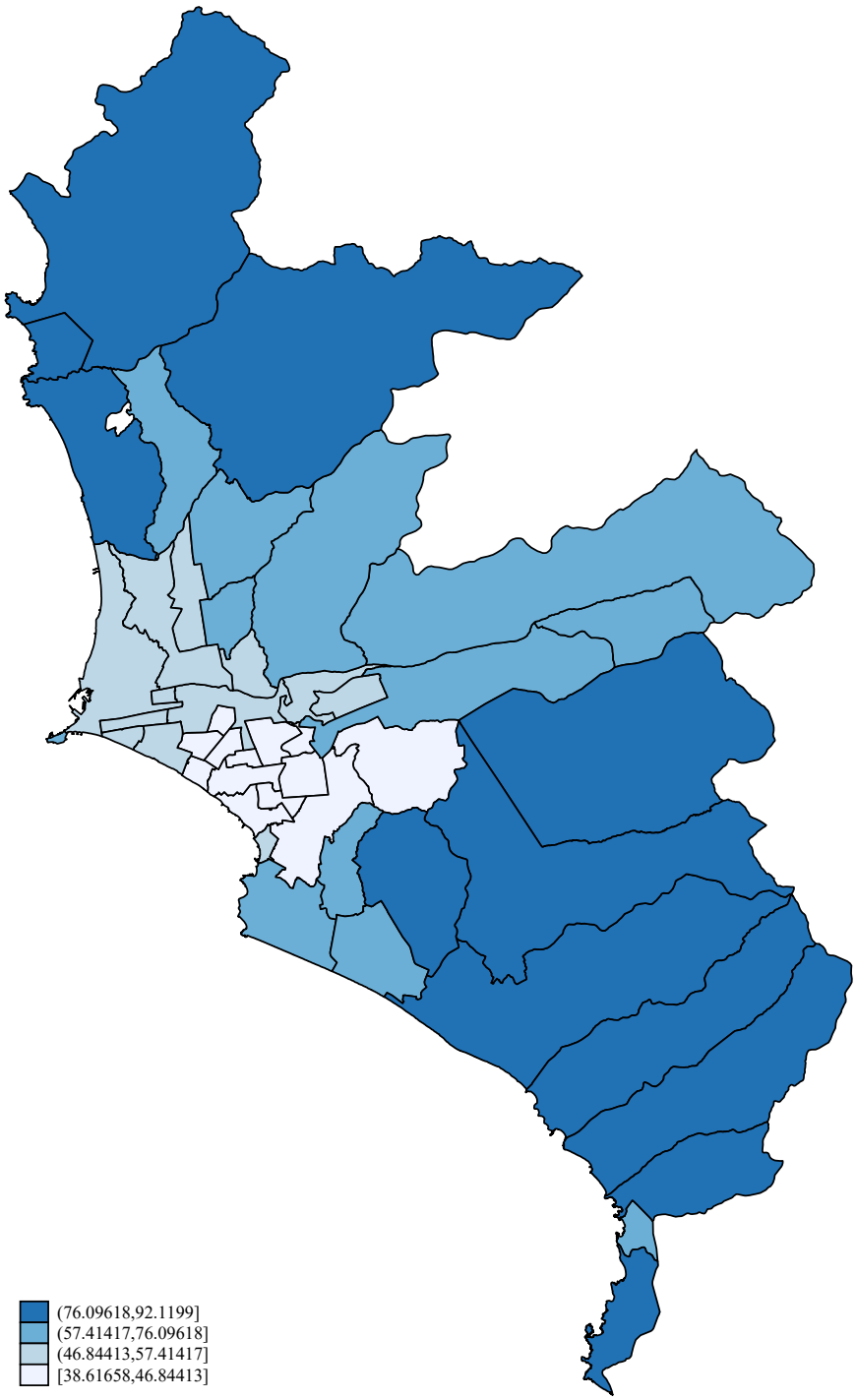
Table 7: Effects of New Lines on College Completion and Employment

	Completion	Employment	Blue Collar	White Collar
	(1)	(2)	(3)	(4)
Treatment*Open	0.002	0.009*	-0.004	0.013***
	(0.002)	(0.005)	(0.005)	(0.004)
Mean	0.159	0.664	0.408	0.256
N	644560	234482	234482	234482
Block FE	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and cohort-fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years of exposure. *Data Source.* National Census 2017.

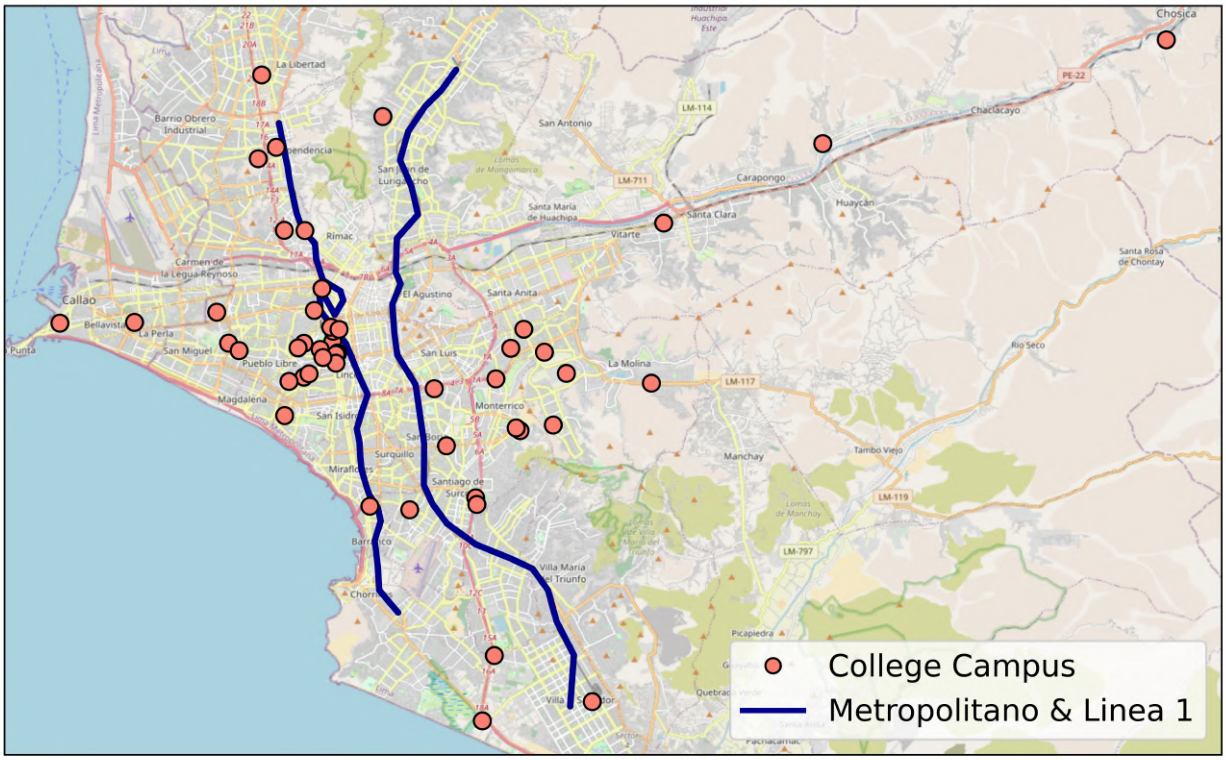
A Additional Figures and Tables

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



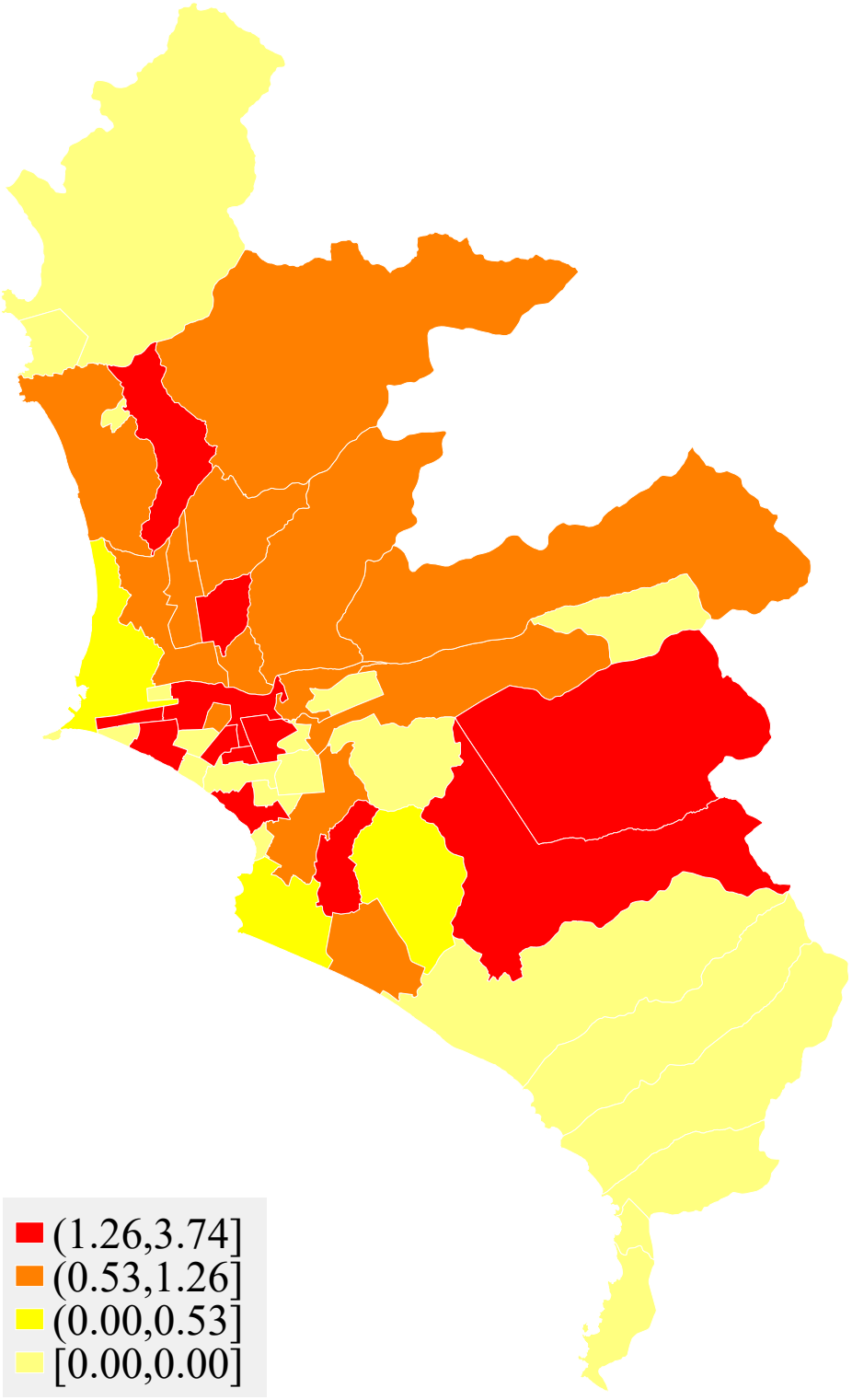
Source. CENAUN 2010. Travel time is self-reported in minutes.

Figure A.2: College location and new lines across lima



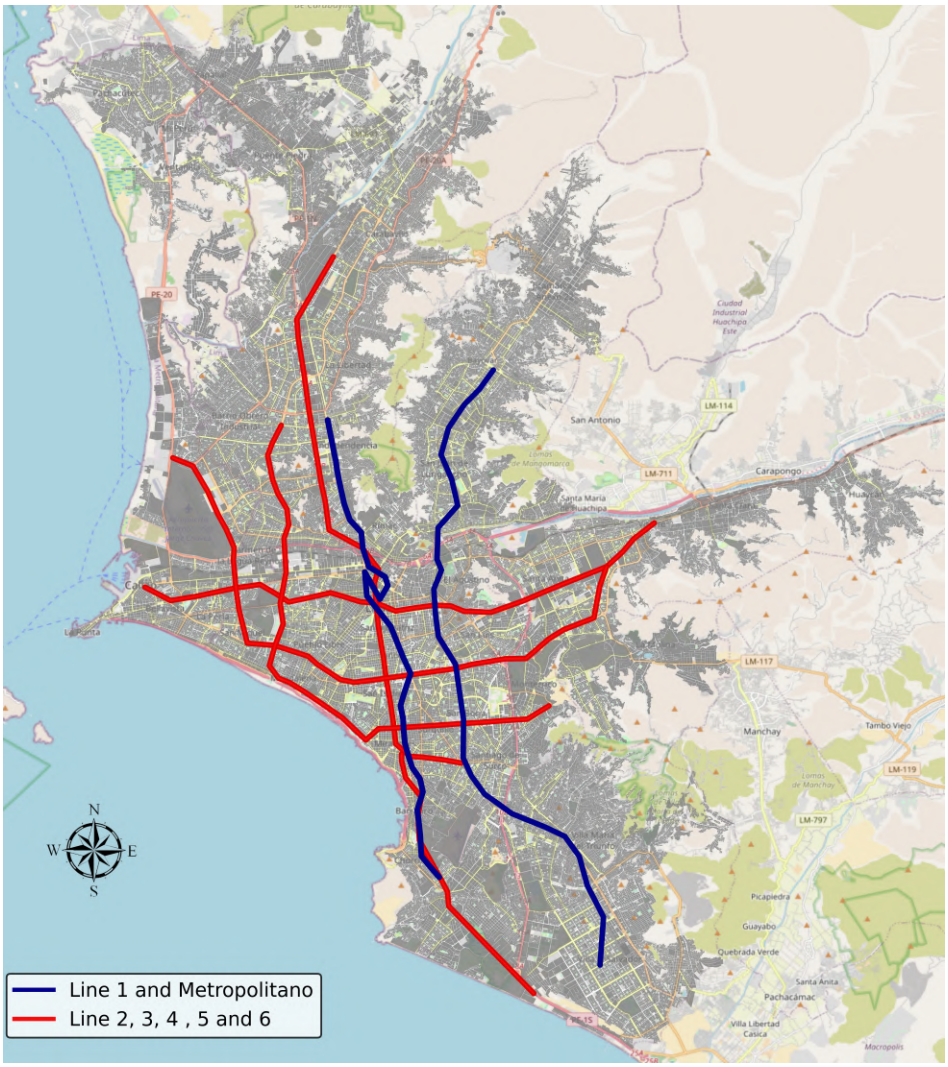
Notes. The blue lines show the new routes, to the left is the *Metropolitano*, and to the right is the *Linea No.1*. The pink dots show the location of college campuses across the city.

Figure A.3: Femicide rate (accumulated 2009-2010)



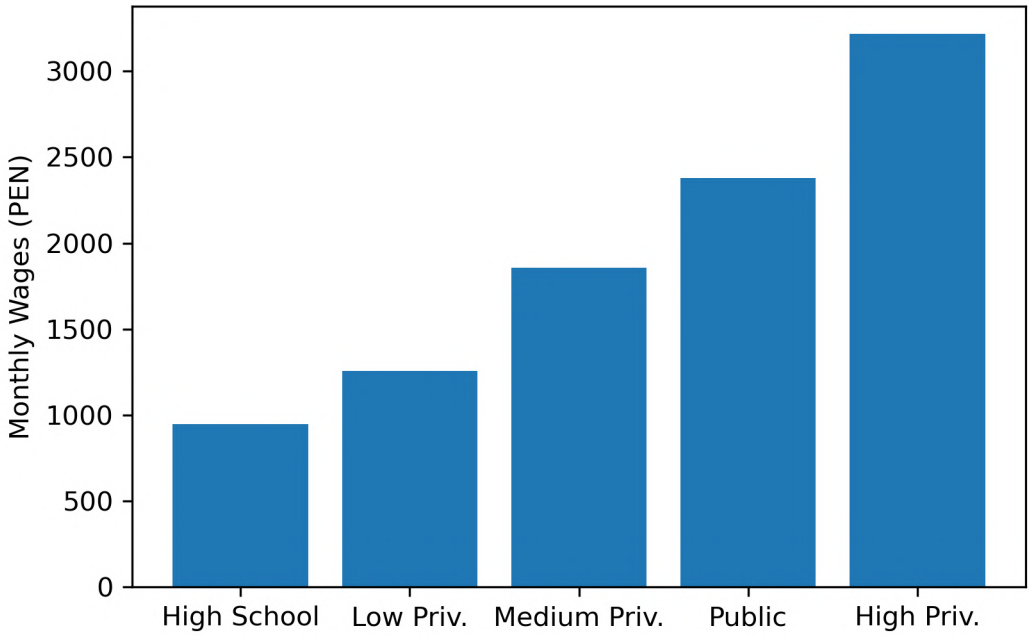
Notes. Femicide rates are the number of femicides (in the years 2009 and 2010) per 100,000 women by district. Yellow districts are defined as *low risk* districts for women.

Figure A.4: Planned-but-not-executed Metro lines in Lima



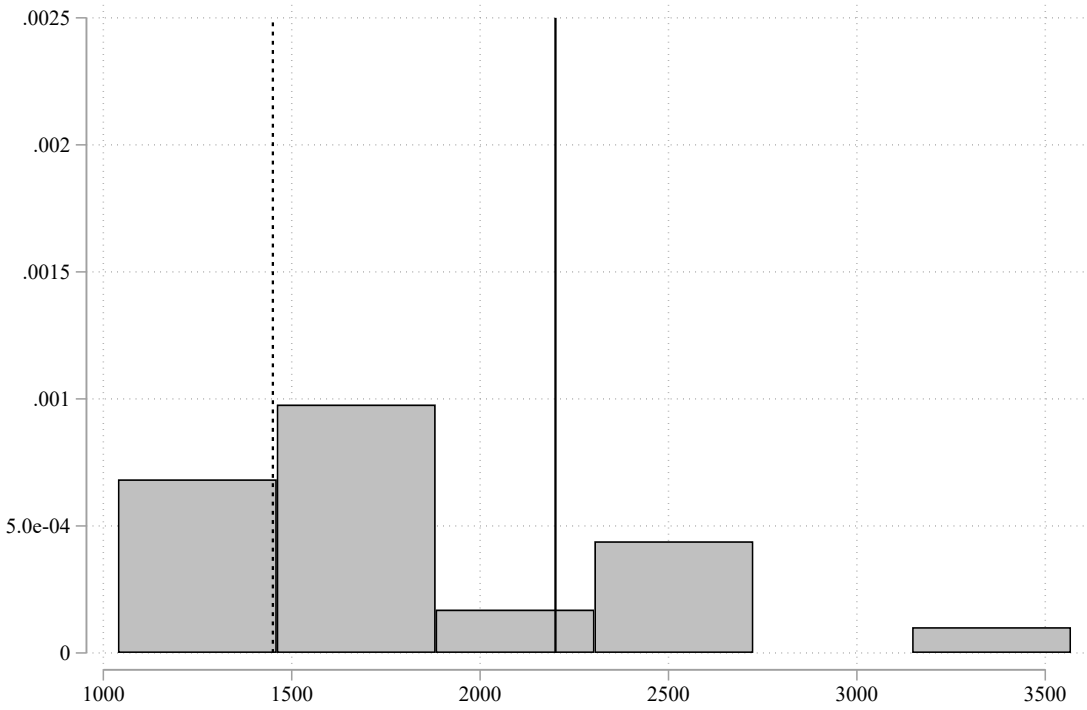
Notes. Red lines are the planned lines according to the Peruvian Ministry of Transport. Blue lines represent Linea 1 (to the left) and the Metropolitano (to the right).

Figure A.5: Wage returns by educational level



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

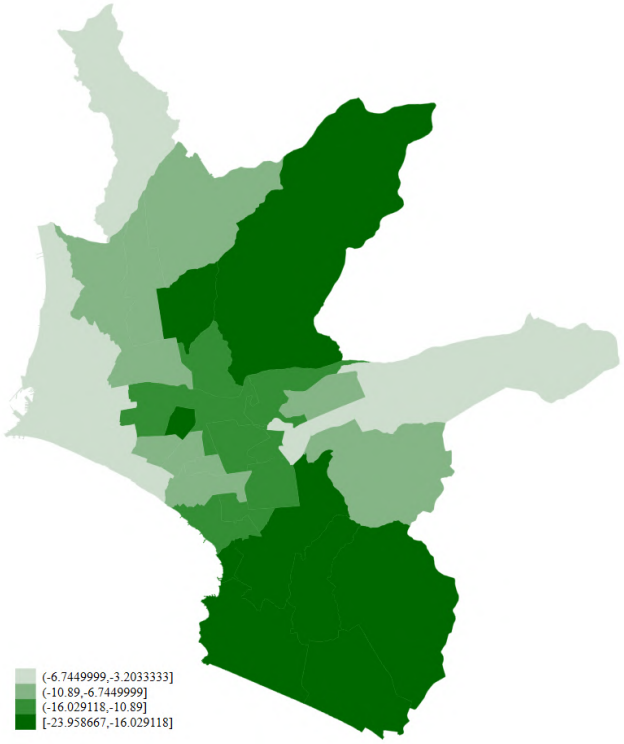
Figure A.6: College wage distribution (average by college)



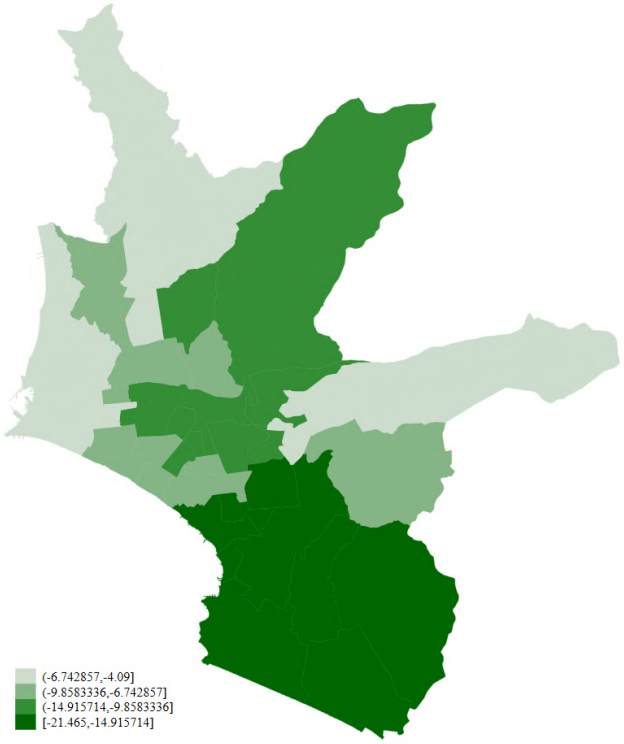
Source. Planilla Electronica, details on data are described in Alba-Vivar et al. (2023). The dashed line shows the limit for the 25% bottom of the distribution below 1450 PEN while the plain line marks the 25% top of distribution above 2200 PEN.

Figure A.7: Commuting time to college (with/without new lines)

(a) All colleges



(b) Private colleges



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

Figure A.8: Share of population who access higher education, Census 2007

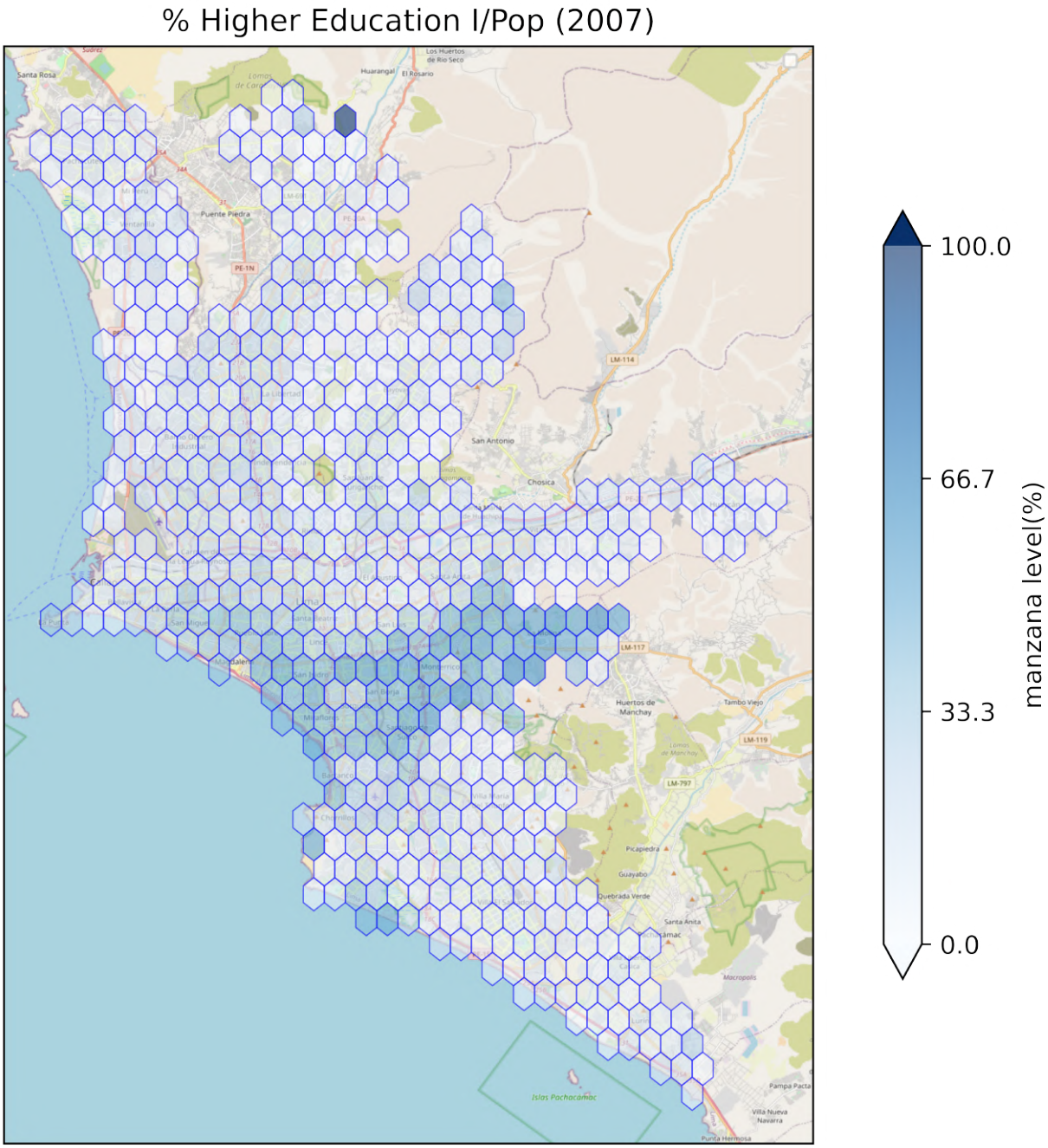
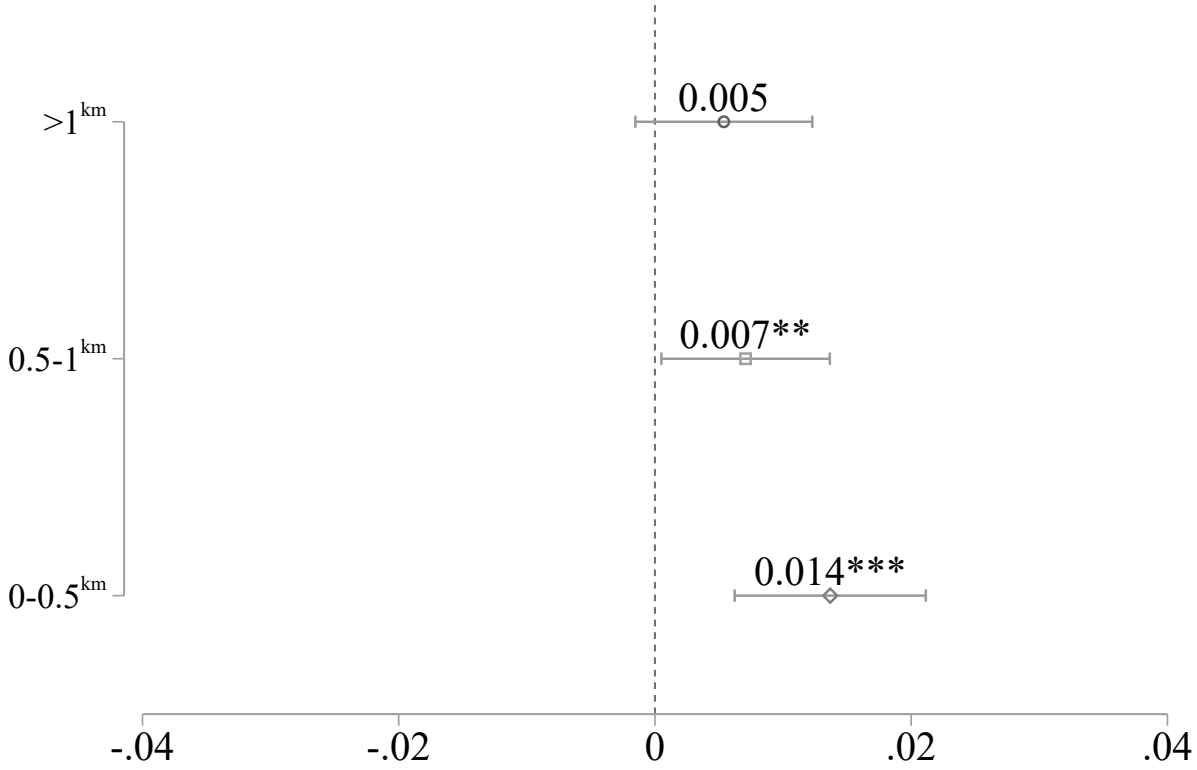
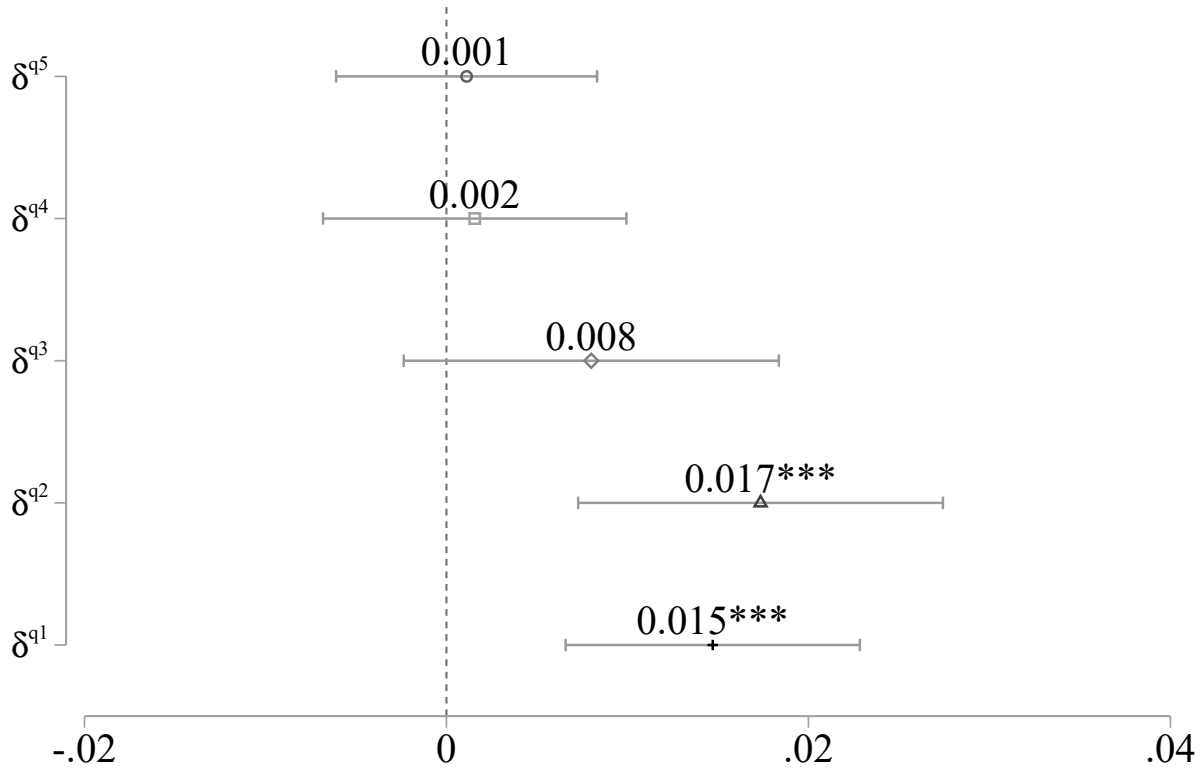


Figure A.9: Impact on college enrollment rates by distance to stations



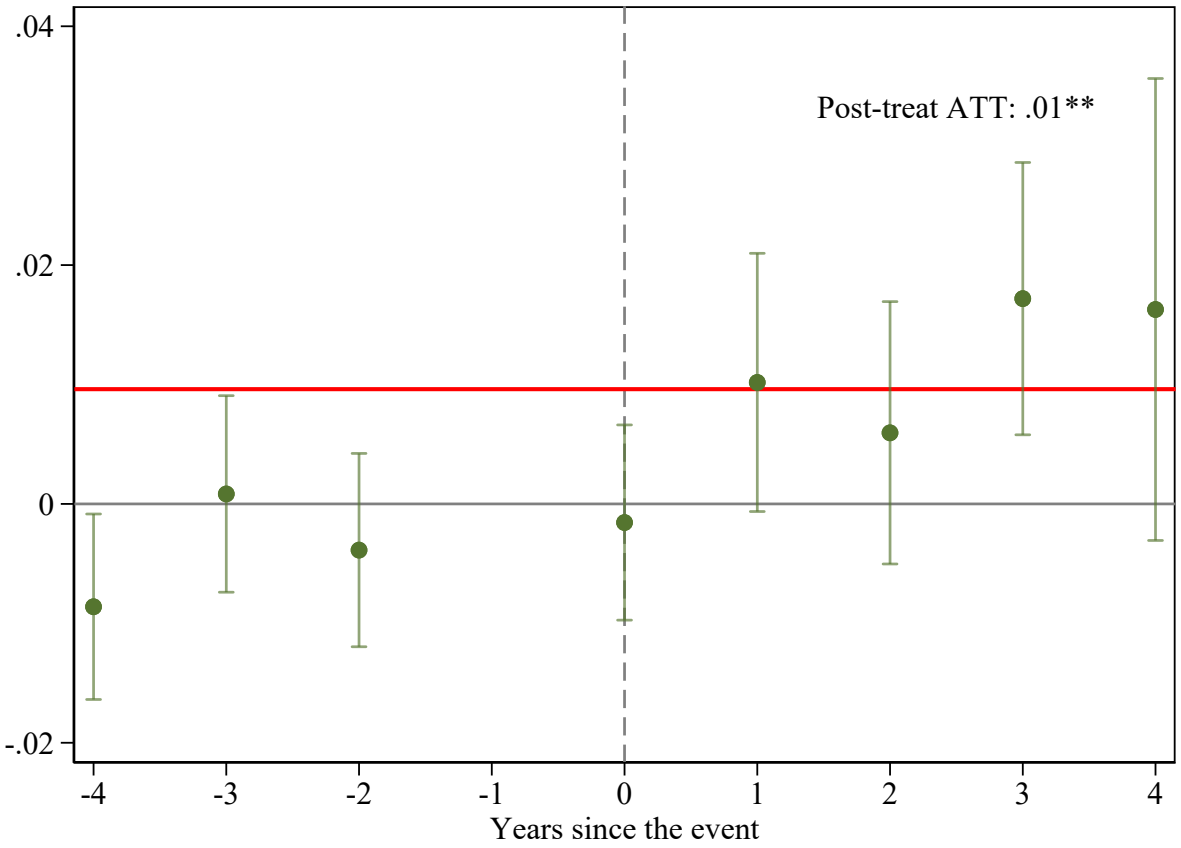
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and year fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening.

Figure A.10: Impact on college enrollment rates by distance to downtown



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. δ^{qN} is the quintile N of the distance to downtown distribution measured in kilometers. All regressions include block and year fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening.

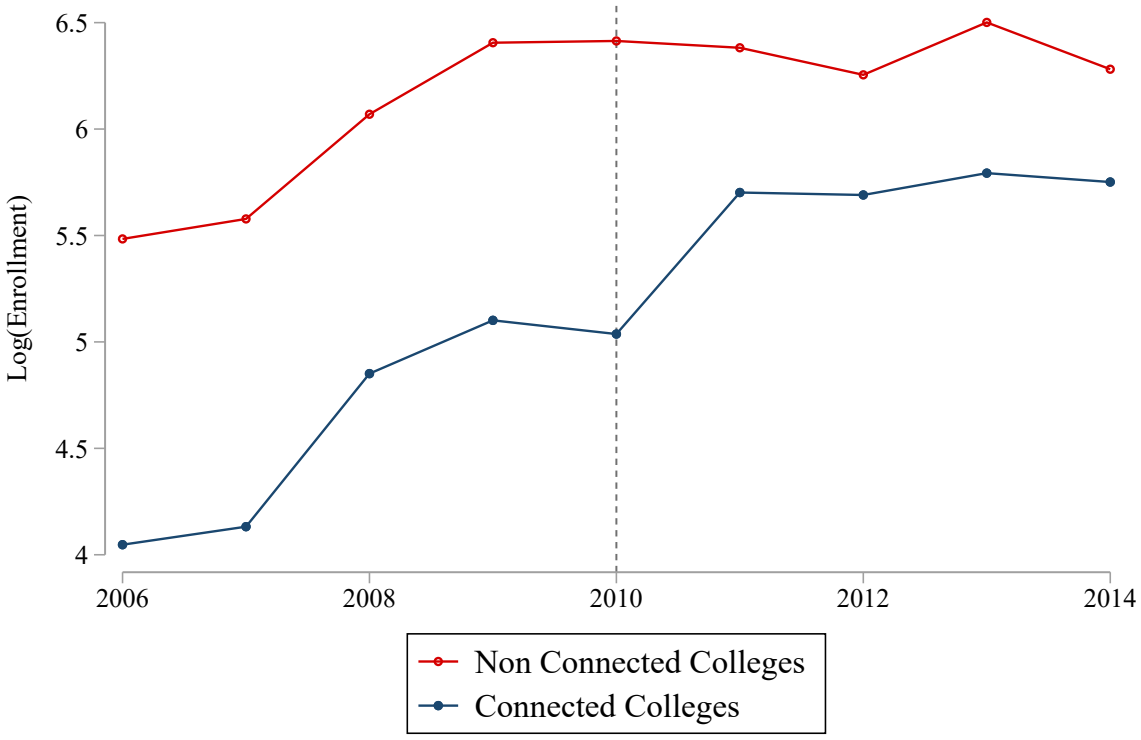
Figure A.11: Effects on college enrollment rates using Callaway and Sant'Anna (2021) estimator



Notes. Average treatment effects are calculated by averaging the group-time average treatment effects across different lengths of exposure to the treatment, with the x-axis representing exposure duration. The outcome of this regression is the hyperbolic sine transformation of college enrollment at the neighborhood level.

Figure A.12: Enrollment and major trends over time

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



(b) Major counts for connected and non-connected colleges

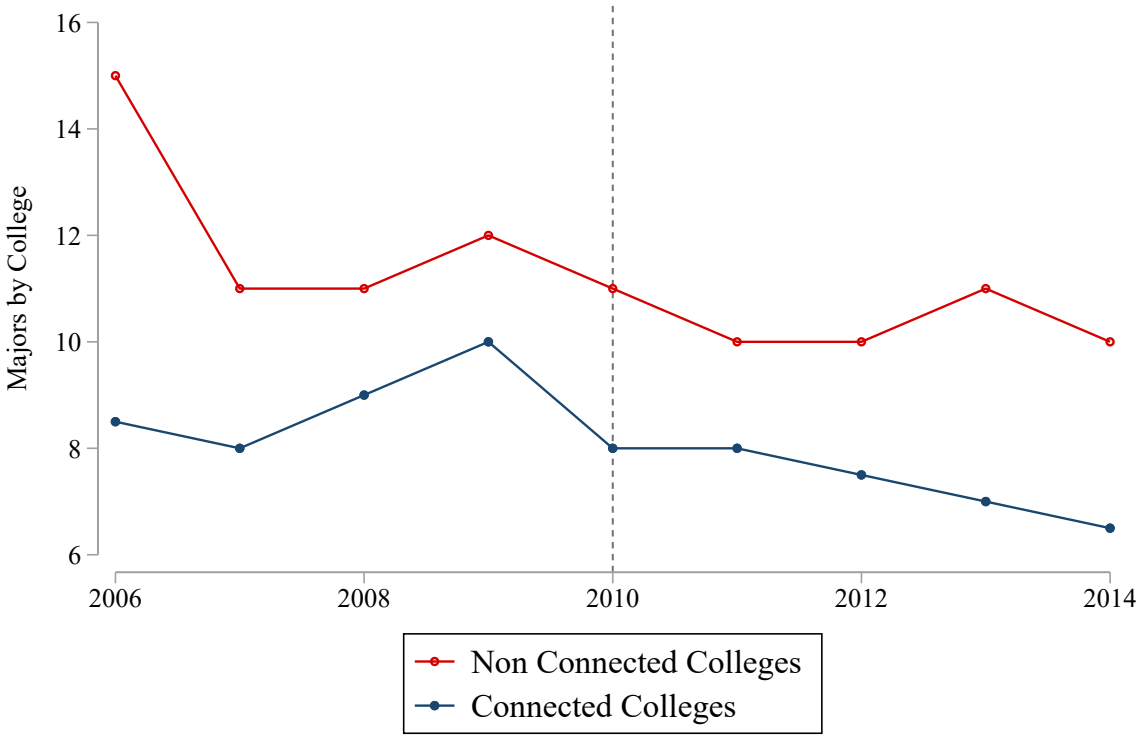


Figure A.13: Log(enrollment) by gender for connected and non-connected colleges

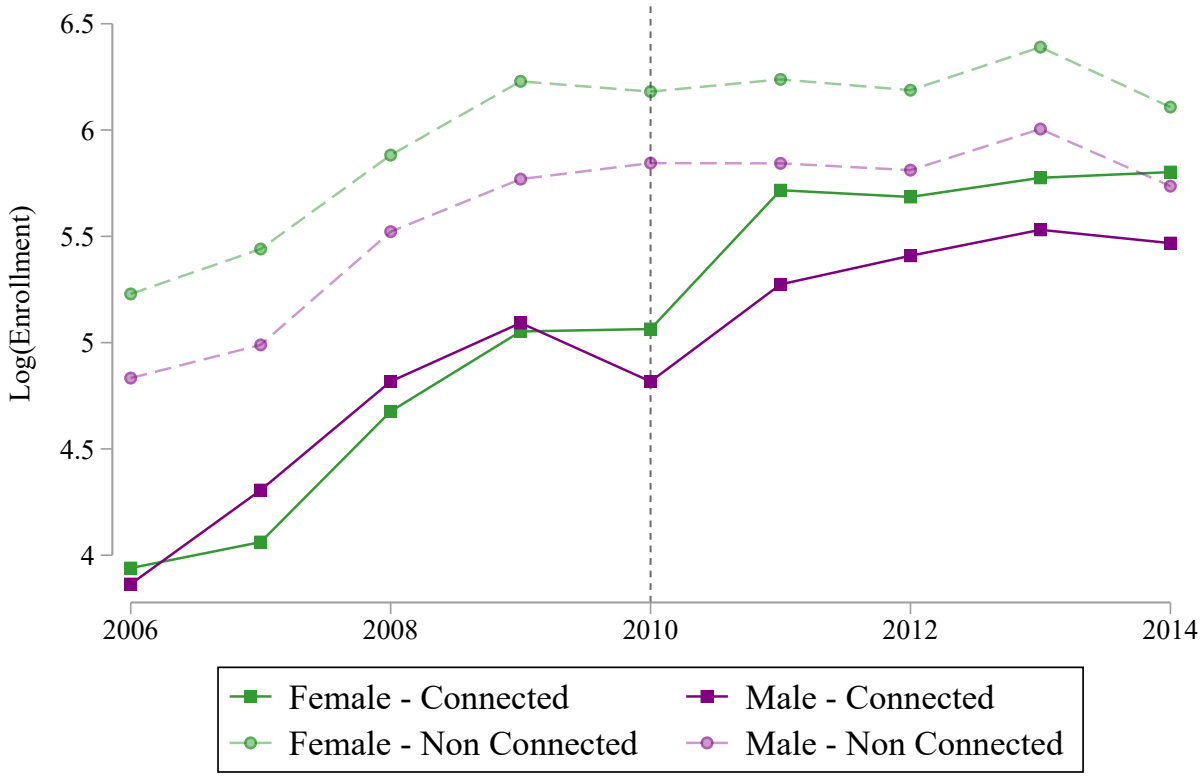
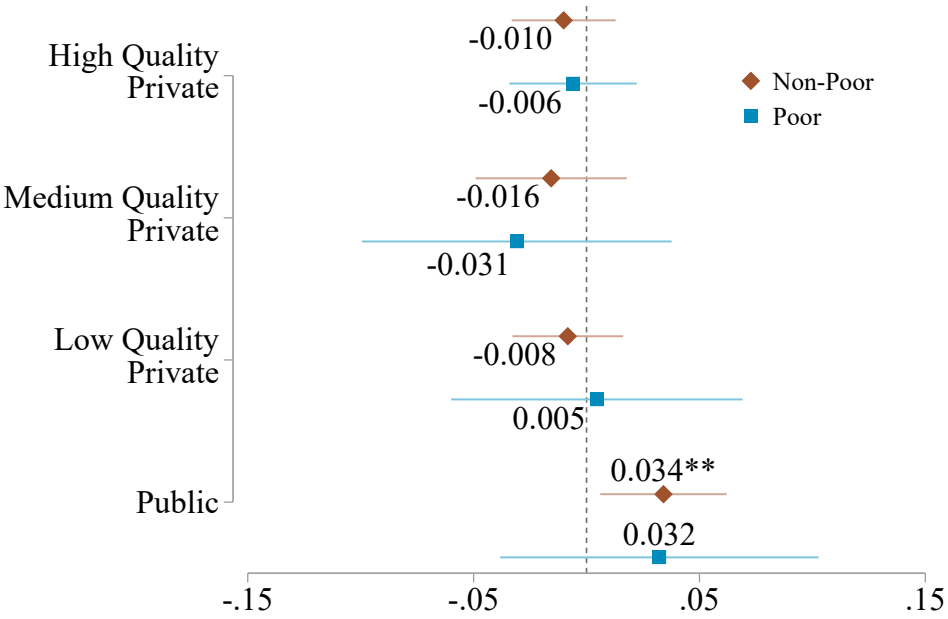
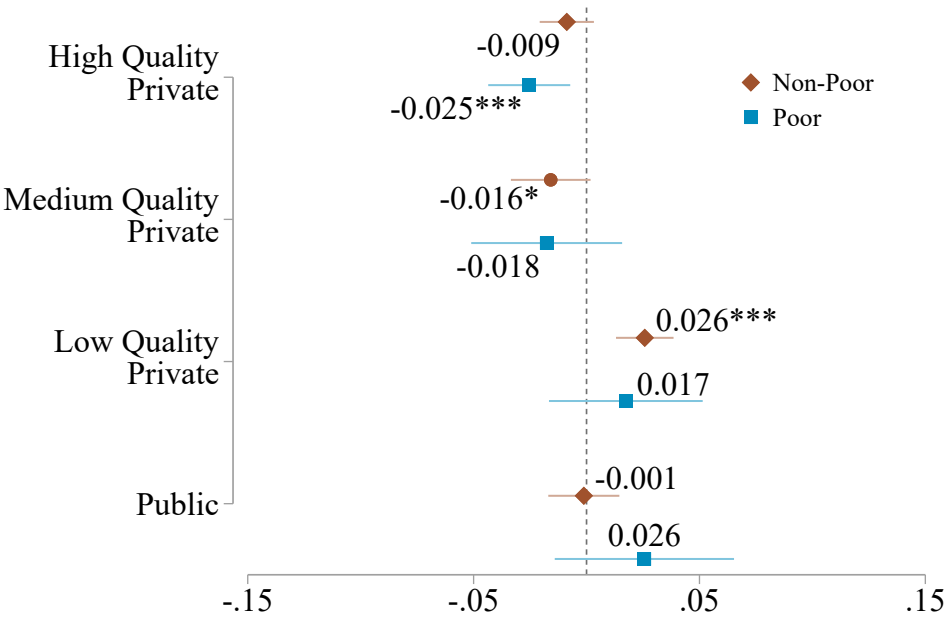


Figure A.14: Effects on the likelihood of enrolling in college type m by SES

(a) *Typical Students*

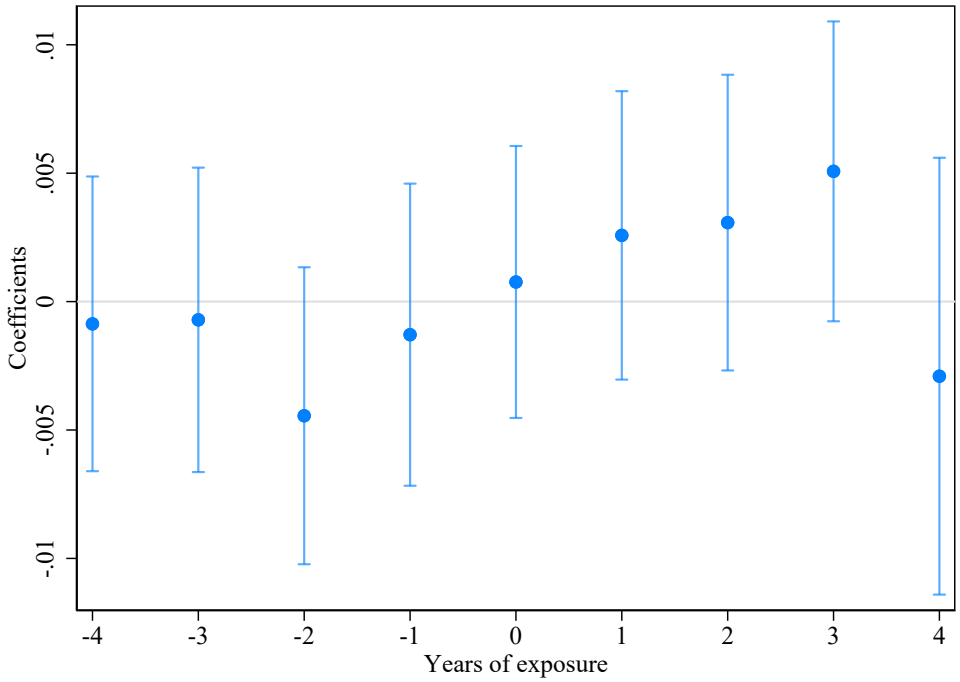


(b) *New Students*



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. Standard errors are clustered at the block level. High-quality colleges are all private colleges whose graduates earn more than 2200 PEN, low-quality private colleges are the ones whose graduates earn less than 1450 PEN and medium-quality private colleges are those whose graduates earn in between. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low.

Figure A.15: Dynamic Effects of College Completion



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and cohort fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening.

Table A.1: Effects of new lines on college enrollment rates including downtown Lima

	Rate	Log(All)	Log(Private)	Log(Public)	Log(Female)	Log(Men)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment*Opening	0.014***	0.016***	0.015***	0.002*	0.012***	0.007***
	(0.002)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Dep. Var. Mean	0.124	0.132	0.104	0.036	0.077	0.066
N	328959	328959	328959	328959	328959	328959
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$.Regressions include block and year-fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) 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.2: Effects of new lines on college choice by access

	High Quality		Medium Quality		Low Quality		Public	
	New	Typical	New	Typical	New	Typical	New	Typical
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment*Opening	-0.011*	-0.010	-0.016*	-0.017	0.025***	-0.007	0.002	0.034**
	(0.006)	(0.011)	(0.008)	(0.016)	(0.006)	(0.012)	(0.008)	(0.014)
Dep. Var. Mean	0.117	0.110	0.400	0.397	0.154	0.164	0.328	0.329
N	78166	22521	78166	22521	78166	22521	78166	22521
Block FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. All regressions include block and year fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening. *Typical students* refers to students residing in neighborhoods that typically send students to college, while *new students* refers to students residing in neighborhoods where the probability of attending college is low. High-quality private colleges are those whose graduates earn more than 2200 PEN; low-quality private colleges are those whose graduates earn less than 1450 PEN; and medium-quality private colleges are those in between.

Table A.3: Effects of the new lines on college choice (relative to predicted)

	High Quality	Medium Quality	Low Quality	Public
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.006	-0.019**	0.015**	0.016*
	(0.006)	(0.009)	(0.007)	(0.009)
Dep. Var. Mean	0.116	0.400	0.156	0.329
N	100687	100687	100687	100687
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block and year fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening. High-quality private colleges are those whose graduates earn more than 2200 PEN; low-quality private colleges are those whose graduates earn less than 1450 PEN; and medium-quality private colleges are those in between.

Table A.4: Effects of New Lines on High-quality Private College Enrollment (Alternative Definitions)

	Elite	Diff. Elite	Licensed	Diff. Licensed
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.022***	-0.020***	-0.015***	0.023***
	(0.006)	(0.006)	(0.005)	(0.007)
Mean	0.135	0.122	0.886	-0.764
N	100687	100687	100687	100687
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$.Regressions include block and year fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening.

Table A.5: Effects on Distance to College and College connectivity

	Distance (km.)	Connected College	Private Connected	Low Quality Priv. Connected	Public Connected
	(1)	(2)	(3)	(4)	(5)
Treatment*Opening	-0.465***	0.027***	0.014***	0.004***	0.013**
	(0.080)	(0.006)	(0.005)	(0.001)	(0.005)
Dep. Var. Mean	9.744	0.236	0.103	0.002	0.133
N	100687	100687	100687	100687	100687
Block FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Distance is measured in kilometers using the Euclidean distance from a block centroid to a college location. Columns 2-5 measure the probability of enrolling in a college connected to the new lines, with the closest station less than 3 km. away. Columns 3 and 4 measure the probability of enrolling in a private and low-quality private college connected to new lines, respectively. Column 5 measures the probability of enrolling in a public college connected to a new line. All regressions include block and year fixed effects. Standard errors are clustered at the block level. ATTs are calculated using the Borusyak et al. (2024) estimator for the first 4 years after a station opening.

B 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 eco-

nomic 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. [33] 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.

[33] 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 [34] 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.

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

C Commuting Time Calculation

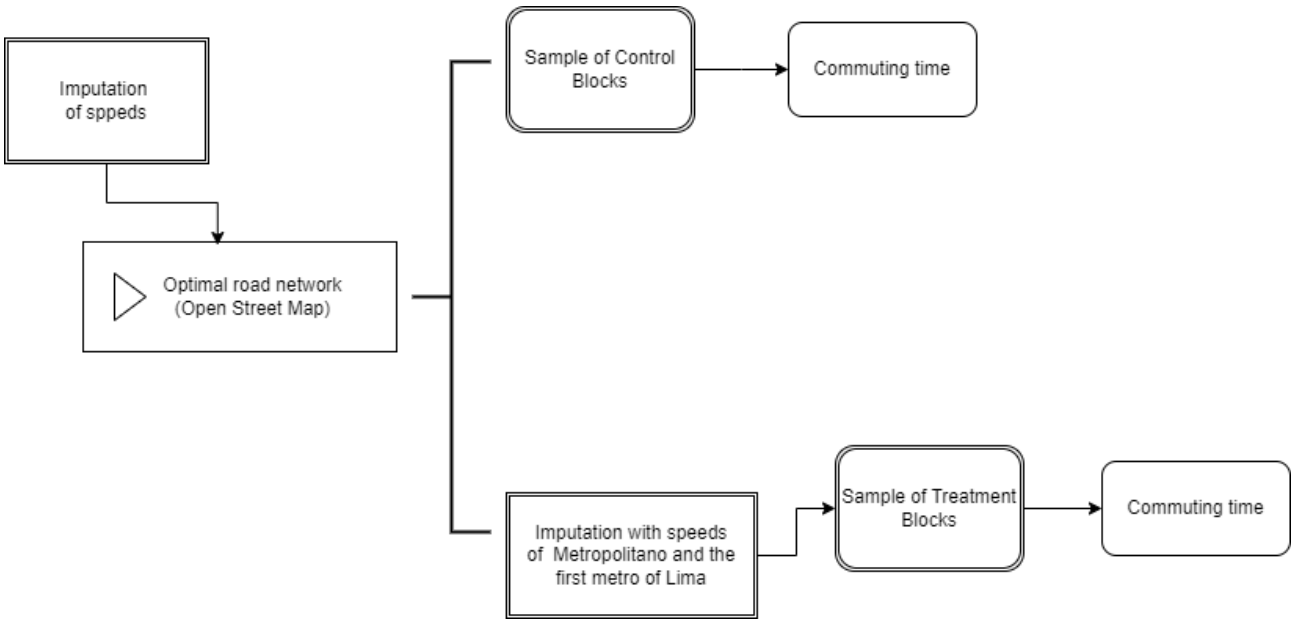
In this section, I explain the procedure to obtain the commuting times used in the paper. I estimate the travel time from each block to all colleges in Lima, creating a commuting time matrix for a random sample of treatment and control blocks. This analysis was facilitated by tools such as the optimal road network and imputed speeds.

Optimal Road Network: OpenStreetMap (OSM) is used to determine the shortest travel time between two points by calculating the optimal route. This procedure relies on three key steps. First, OSM requires a reference polygon, so we use the shapefiles of Lima and Callao. Second, it generates alternative road networks connecting the two locations. Finally, the travel time for each possible route is computed, with the optimal route being the one with the shortest duration. However, these estimates are based on the maximum allowed speed for each road segment, assuming ideal driving conditions. As a result, they do not account for real-world traffic congestion and cannot be directly used for commuting time calculations. This process was implemented in Python using the 'osmnx' library.

Imputed Speed: Since OSM estimates commuting time without accounting for traffic, I independently computed travel times by adjusting for congestion. To do this, I used the optimal road network, which Python exports as a table containing detailed attributes for each road segment, including highway type (primary, secondary, tertiary, motorway, residential), section length, name, and maximum allowed speed. I then imputed travel speeds for each segment based on rush hour conditions specific to its highway type. These values were sourced from a 2013 report by the Ministry of Transport and Communications, conducted in collaboration with the Japan International Cooperation Agency (JICA, 2013). The imputed speeds for each highway type are as follows: motorway (15 km/h), primary (15 km/h), residential (15 km/h), secondary (20 km/h), and tertiary (10 km/h). This is a similar procedure used by (Velásquez, 2023).

Commuting Time: On the final step, I estimate travel times for a sample of control and treatment blocks (with and without connectivity to the new lines). For control blocks, commuting time was calculated by applying the imputed speeds to the corresponding road segment lengths in the optimal road network. For treatment blocks, I incorporated the impact of public transportation by assigning travel speeds of 35 km/h for the Metropolitano and 45 km/h for Lima's first metro line in segments where the optimal road network overlaps with these transit routes.

The following figure provides a flowchart of the commuting time calculation process, offering a concise summary of the steps described above.



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 can capture non-linearity. To avoid its sensitivity to the choice of a distance metric, I use the grid search technique to find the optimal value of k, the number of nearest neighbors to consider when making predictions. This optimal k maximizes the test dataset score using cross-validation. Table D.1 shows the results of this procedure.

The algorithm proceeds as follows:

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

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

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

Table D.1: kNN Scores

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