Opportunity Bound: Transport and Access to College in a Megacity

Fabiola M. Alba-Vivar ⋆

(Job Market Paper)

September, 2023

Click here for the most recent version

Abstract

This paper examines the causal effects of new transportation infrastructure on college enrollment, choice, completion, and on early labor market outcomes. I use novel geo-located administrative data to estimate a difference-in-difference model that exploits the rollout of two new public transportation lines in Lima, a city of 12 million people. Neighborhoods connected to new lines increase college enrollment rates by 1 *p.p.* relative to the 14% baseline, mainly driven by private college enrollment. Interestingly, students migrate toward low-return colleges, which are the ones connected to the new stations, especially women. In contrast, men are more likely to enroll in public colleges which are more dispersed over the city. A model of college choice shows that male students are willing to travel twice as much as females to enroll in a college which gives them one standard deviation higher salaries. In the medium and long run, access to transport increases a person's likelihood of graduating from college by 12% and access to white-collar jobs by 4%.

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

I thank Alex Eble, Cristian Pop-Eleches, Micaela Sviatschi, Sandra Black, and Maria Rosales-Rueda for their constant guidance and support. I also gratefully acknowledge all the comments and suggestions from Sofia Amaral, Mark Anderson, Sam Asher, Dan Bjorkegren, Girija Borker, Kirill Borusyak, Carolina Concha-Arriagada, Kevin Donovan, Paul Goldsmith-Pinkham, Stephen Keating, Bentley MacLeod, Matteo Magnaricotte, Nishith Prakash, Sebastian Tello-Trillo, Anja Tolonen, Daniel Velasquez, Eric Verhoogen, Jack Willis. I am also thankful to all participants at the WEAI Graduate Student Workshop, AEFP, Seminario MAP, the Development Colloquium at CU, and the Economics of Education Colloquium at TC. This project was possible thanks to the generous funding of the Center for Development Economics and Policy at CU and the Department of Education Policy & Social Analysis at TC. I am deeply thankful to Roberto Mendoza, who provided excellent research assistance. I have benefited from data provided by the Peruvian Ministry of Education (MINEDU) and the National Institute of Statistics (INEI). I also thank members of the *Dirección General de Educación Superior Universitaria* at MINEDU for their support in this project.
⋆ Teachers College (TC) at Columbia University, *fma2139@tc.columbia.edu*,, New York, NY.

"Few factors express better the disdain towards the marginalized urban sectors than the state of abandonment of public transport in Lima." - Danilo Martucelli in "*Lima y sus arenas*"

1 Introduction

One in eight people in the world lives in a "megacity", defined as a city with over 10 million people.[1] As these megacities continue emerging, new opportunities and challenges accompany their growth. A key aspect that can either limit or further boost economic productivity is the state of public transportation. Recent work in the literature has shown that an upgrade in public transportation can improve welfare in aggregated terms and has the potential to reduce inequalities within cities (Balboni et al., 2021; Tsivanidis, 2022; Zarate, 2022). This paper estimates how improved public transportation directly affects a key driver of economic productivity: human capital accumulation via college education.

The setting of this paper is Lima, the capital of Peru and a megacity of 12 million people. Identification comes from the creation of two new mass public transportation systems, created in the early 2010s: a rapid bus transit line and, separately, a new train line which, I show, dramatically reduced transit times to college for 50,000 college students on average each year. Prior to the rollout of these new transit lines, students in Lima spent two hours per day, on average, commuting to and from school.[2]

First, I create a novel dataset of college location, student location, and transit options that captures variance in commuting time and educational choice over time. I begin with administrative college enrollment records for 450,000 students over 10 years. I geocode the locations of both colleges and students' households in each year and then link these to data on the location and date of opening of the public transit stations from the two lines. I use a difference-in-differences approach to generate causal estimates of how access to improved public transportation affects college

[1] As of today, there are 33 megacities around the world which are predominately in the Global South and this trend is set to continue. A recent report by Euromonitor highlights how most of the future upcoming megacities in the world will be from developing countries. These upcoming megacities include Luanda, Dar es Salaam, Baghdad, Chennai, Bogota and Chicago, which are currently experiencing higher economic and population growth.

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

access, choice and match quality. My approach exploits variation by cohort and neighborhood location as well as the staggered nature of the station openings. My strategy compares educational choices and outcomes of cohorts in neighborhoods exposed to new public transit stations to the same cohorts in neighborhoods that were exposed to planned but non-opened stations, inspired by the placebo strategy implemented in Donaldson (2018), which further improves the difference-in-difference estimates.

I show that access to improved public transit significantly increases the likelihood of college enrollment among college-aged youth living in connected neighborhoods. I estimate a 1 percentage point increase in enrollment rates in affected areas. At this point, I do not observe significant differential effects by gender, and most of the effects are driven by middle-income households and neighborhoods that typically had low or zero college enrollment rates. More importantly, the results show that the increase in access is mostly driven by private college enrollment, whereas the impact on public college enrollment remains positive but economically negligible. In this setting, Peruvian public colleges are more competitive and tuition-free in comparison with private colleges that have relatively easier admission systems and a large proportion of them are low-cost.[3]

Second, I find that a decrease in commuting time to college also corresponds with a change in college choice. Students who live in neighborhoods connected to the new stations are more likely to enroll in low-return private and public colleges and less likely to enroll in high-return private colleges. Crucially, this change in college choice is concentrated among female students. After increased access to public transportation, women are driven to colleges whose graduates' wages are at the bottom of the distribution. On the other hand, men are also less likely to enroll in high-return or elite colleges but more likely to enroll in public colleges. The implications of women choosing low-return private colleges compared with men taking advantage of attending free public colleges are of interest since they can help explain the relatively large wage gender gaps in the Peruvian labor market.

Linking these data to the 2017 Peruvian national census, I also find that access to improved public transportation increases the likelihood of college completion. I estimate a 12 percent increase

[3]See the work of Flor-Toro and Magnaricotte (2021) who document the disparities among the admissions systems for both private and public in Peru, where the interaction of initial advantage from high school quality and meritocratic criteria in public colleges increases educational inequality.

in college completion among students living in affected areas. Using a cohort-exposure analysis relative to the year of high-school graduation, I show that these effects increase in magnitude among those who were exposed to improved transit options for the longest. This implies a major reallocation of time and resources, given that most Peruvian college students tend to complete their education over more than five years, longer than typical in other countries. These effects are particularly higher for women and low-income students.

The focus of this paper is on tertiary education, which has the power to switch labor market trajectories dramatically given the higher labor market returns to college education and more importantly, reducing inequalities in the long run. The results suggest that students' college choice switches towards low-quality colleges, which yields a decrease in projected earnings by 3 percent. However, this effect is heterogeneous by gender, with women experiencing significantly more negative effects compared to men, who, in contrast, experience positive effects. I also examine medium-term outcomes such as employment rates. My finding suggests that there is a positive effect on employment rates for those who enrolled in college measured in the 2017 Census, 7 years later after the first opening. More importantly, these effects are coming from students who enrolled in college employed in white-collar jobs.

This paper contributes to the literature on the economics of transportation. Most of this literature regarding the economic impact of improving or building new transportation in large cities shows positive effects on economic activity and labor market opportunities (Balboni et al., 2021; Tsivanidis, 2022; Zarate, 2022) [4]. However, less is known about the direct causal effects on human capital investment. First, I document whether neighborhoods see a decrease in commuting time to *any* college in the city: students living in neighborhoods who get connected to the new stations see a 17 percent reduction in the average commuting time to any college in the city compared to those who do not get connected, this is equivalent to almost 30 minutes per day. It is also worth highlighting that there is not only a reduction in transportation costs but the new system is also providing better labor market opportunities as it connects people from the outskirts of the city to

[4] Currently, there is evidence on different types of transportation such as railroads and roads (Brooks and Donovan, 2020; Donaldson, 2018; Donaldson and Hornbeck, 2016) which shows positive effects in trade and economic growth. At the city level, there is evidence on buses or BRTs (Balboni et al., 2021; Tsivanidis, 2022) and metros (Zarate, 2022)

Lima Downtown (where most white-collar jobs are located). In this sense, potential college students might also benefit from better job prospects after graduation. In this sense, Adukia et al. (2020) find that children stay in school longer and perform better on standardized exams in rural areas that get connected to urban areas that offer higher returns to education. The results of this paper show reduced-form results that combine both channels: increased labor market opportunities (both for current jobs and after graduation) and reduced commuting times.

As mentioned above, there is little evidence of the impacts of city transportation policies on higher education given the lack of data or empirical strategy challenges. In that sense, this paper also contributes to the current literature on education and transportation. This work provides the first causal estimates of a new public transportation system and its effects on college access. There are two main channels that can explain these effects that have already been explored in the literature: reducing transportation costs and increasing travel safety. Currently, evidence from developing countries has shown that reducing transportation costs has positive effects on women's access to basic education as well as improving schooling outcomes and aspirations (Fiala et al., 2022; Muralidharan and Prakash, 2017). Other related work has shown how transportation affects human capital, especially in primary and secondary education. Tigre et al. (2017) document how the duration of commuting has a negative causal effect on academic achievement using data from Brazil. Asahi and Pinto (2022) show how the extension of the subway in Santiago de Chile reduces the gap between low-skill and high-skill students. Using the same expansion, Herskovic (2020) shows that families connected to new subway stations travel further and even get to enroll in higher-quality schools. In Mexico City, Dustan and Ngo (2018) shows that new trains increased the demand for elite and more distant schools. However, given the test-based assignment mechanisms in the city, only high-achieving students with highly-educated parents take advantage of the new lines. Muralidharan and Prakash (2017) study the impact of providing bicycles to female students, a reduction in transportation costs, and find that being exposed to the program increased girls' enrollment in secondary school by 32 percent. In this paper, I focus on a higher level of education: college, which is becoming relevant in the developing world as access to basic education gets covered. What is more, the results of this paper also apply to big cities in both the developing and developed worlds where access to college is still limited, especially access to *high quality* colleges for poor students. My results show that living in a neighborhood that gets connected to new lines affects not only college enrollment rates, as it directly decreases transportation costs to college, but also can change college decisions and change incentives to invest in higher quality education.

Another important potential mechanism related to transportation costs is travel safety. This is particularly important in large cities in the developing world, where women face harassment and crime at significantly higher rates than men, and this can yield differential responses by gender when travel costs change. There is a small but growing literature on this matter, where most results show that women have a higher demand for safe transportation and this can directly affect their labor supply Field and Vyborny (2022); Kondylis et al. (2020). [5]. The closest paper to my work is the one from Borker (2020), which explores how the perceived risk of street harassment can help explain women's college choices in Delhi. She finds that women are willing to choose a low-quality college over a top college that is perceived to be one standard deviation safer. In this paper, I focus on Lima, a city considered one of the worst large cities in the world for women's mobility and transportation, comparable to Delhi, Mexico City, and Jakarta.[6] My results are consistent with previous literature given that the new transportation systems provided a safer ride compared to the informal buses that circulate the city. I find that female students are getting most of the benefits: they do not only access college at a higher rate than men but they are also entirely driving the effects for college completion. However, similar to Borker (2020), women are more likely to enroll in lesser-quality private colleges since these colleges are the ones getting connected to the new stations, minimizing their exposure to harassment and facilitating their commute.

A final contribution of this paper is to the literature on the geography of inequality and place-based policies. Prominent work has shown how growing up in places with more opportunities can positively impact income mobility, especially for economically disadvantaged populations and minorities (Chetty et al., 2020; Chetty and Hendren, 2018). What is more, accessing better schools can have positive effects on attending college Bergman (2018). In this project, I explore how

[5]In Lima, the introduction of the new stations also had a differential impact on women's labor supply Martinez et al. (2020); Velásquez (2023)

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

transportation can create a way to commute to opportunities by increasing access to college for new students and subsequently, accessing higher-quality jobs. However, this gain in access gets limited by the supply of colleges in the city. Most students prefer to commute less and access lower-return colleges connected to the new stations.[7]

The remainder of this paper is organized as follows: in section 2, I introduce a simple conceptual framework. In section 3, I present the setting of the new transportation lines and in section 4 I describe the data. In section 5, I present the empirical strategy. Section 5 includes the main results of the paper and Section 6 concludes.

2 Conceptual Framework

In this section, I present a simple conceptual framework that outlines the decision-making process for students. In this stylized model, students are utility maximizers. A student will choose college c from the choice set $C = [c_1, c_2, ...c_N]$ based on the utility coming from Equation 1. Other important factors like college reputation, non-monetary preferences, etc. are omitted for simplification.

$$U_i^C = f(W_i^C, D_i^C, X_i^C) \tag{1}$$

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

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

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

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

$$
\begin{aligned}
U^A - TC^A - P^A > U^B - TC^B - P^B, \quad C^* = A \\
U^B - TC^B - P^B > U^A - TC^A - P^A, \quad C^* = B
\end{aligned} \tag{3}
$$

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

3 Background

3.1 College Education in Peru

The Peruvian Education system is based on 3 levels: primary education (6 years), secondary education (5 years), and the higher education level which often lasts from 2 (technical school) to 10 years (School of Medicine). On average, college students take between 5 and 7 years to graduate. According to the 2017 Peruvian Census, approximately 4 out of 10 recent high school graduates (between 17 and 21 years old) have access to some type of higher education. More specifically, 15% have access to a technical school or community college, 22% access to university while the remaining 63% do not access any type of higher education (Alba-Vivar et al., 2020). Following similar trends to the rest of the world, Peruvian women access college in a slightly greater proportion than men. Similarly, those who are Spanish native speakers access college at higher rates in comparison with other ethnic minorities (Quechua and Aymara native speakers). Some aspects of the college education system are worth highlighting. First, there is no centralized admission system and students can take admission exams for multiple colleges, similar to the US. However, there are no standardized exams like the SAT. What is more, public universities are virtually free as they only charge a small administrative fee. On the other hand, private universities have a greater variance in price and quality. Typically, public and elite private colleges have competitive admission exams while the rest of the private colleges enroll students on demand.

Notably, 60% of students attend a higher education institution located in their province of birth, and the number rises to 90% when it comes to colleges suggesting that out-of-state college enrollment is quite uncommon. College housing is virtually nonexistent; if they exist, they are reserved exclusively for out-of-state students, but most live in off-campus housing. The focus of this paper is Lima, the capital of Peru, which concentrates around half of the Peruvian college enrollment. Most students live at home with their parents and commute to college. Figure A.1 shows the average travel time from home to the university campus in minutes using the 2010 University Census data. Students living on the outskirts of the city travel from home to college an average of 1.5 hours, whereas those living in Downtown Lima (city center) travel 40 minutes.

3.2 Transportation in Lima

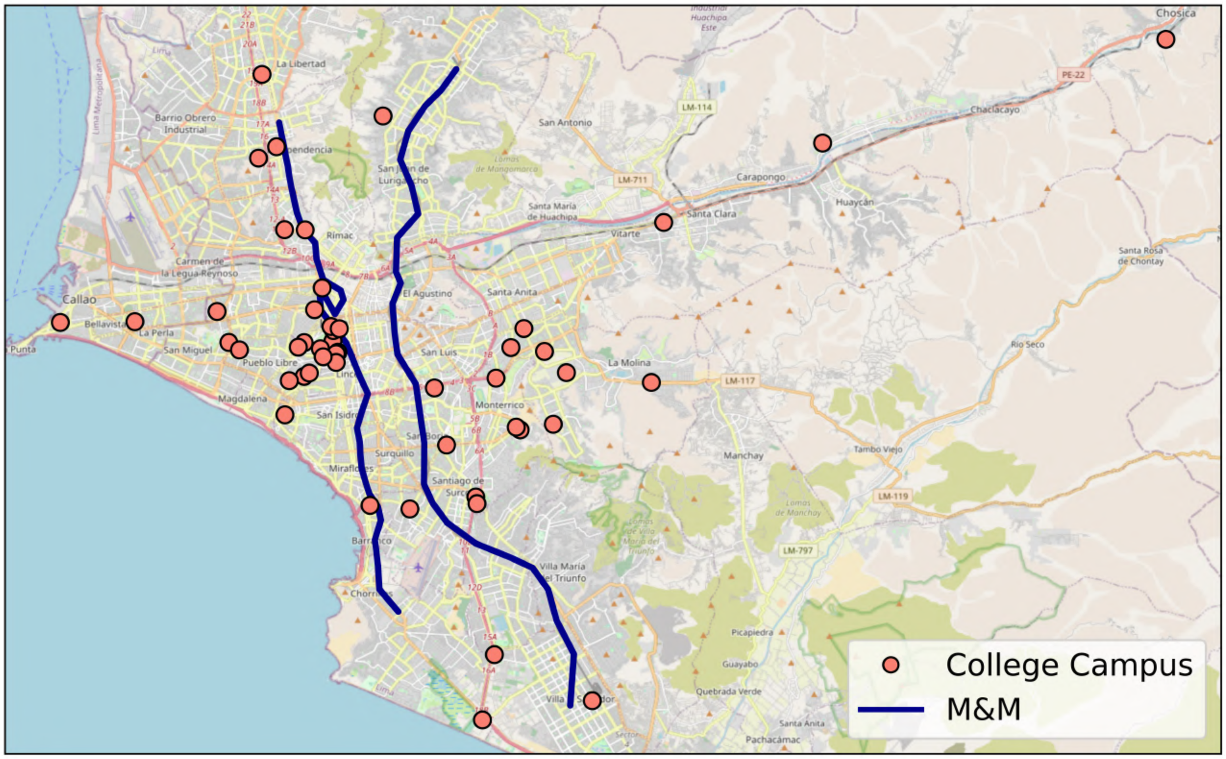
Lima's population is comparable to other large cities around the world such as New York City, Paris, Xi'an, Chennai, Jakarta, Bogota or Los Angeles Metropolitan Area. However, Lima is not nearly as dense (4,000 hab/km^2, compared to NYC's 10,194 hab/km^2), and commuting across the city can take up to 3 hours during rush hour. During the 90s, market liberation policies facilitated the import of used cars and mini-buses which became the basis of the new transportation system for commuters. These smaller, privately-operated minibuses known as *combis* partially alleviated the demand for transportation across the city and became the primary mode of transportation. However, as demand kept increasing and given the poor quality of these *combis* and lack of regulations, this mode of transportation became unsafe for commuters. Subsequently, most institutional efforts to regulate them were unsuccessful. A regional initiative to implement a rapid transit line that connected the north and south of the city was opened to service in July 2010, after 4 years of construction. The *Metropolitano* was the very first mass transportation public system in Peru, connecting 12 districts out of the 44 in the city. There was a flat fee of 1.50 PEN, approximately 0.50 USD for regular commuters but college students had a 50 percent discount.[8]

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

A year after the *Metropolitano*'s opening, the Peruvian president inaugurated the *Linea No.1* of the *Metro de Lima.* According to different sources I collected, in the 70's, the Peruvian Ministry of Transport had designed a complete metro system for the city, which aimed to connect multiple districts, specially emergings neighborhoods in the outskirts. The original plans for the city included 8 lines but, as years passed, only 6 of them were properly studied and evaluated. Appendix Figure A.2 shows these lines. A very small part of the project was initiated in the 1980s during the first government of President Alan Garcia, but it remained incomplete for 20 years and never opened to the public. Peru's major economic crash avoided future developments of this project until it revived in 2006 during Garcia's second presidential term Campos et al. (2021). Due to budgetary restrictions and President Garcia's wish to inaugurate the project he had promised in the 80s, *Linea No.1* was rushed to open before it was completed and before the end of Garcia's term in 2011. The remaining half was subsequently inaugurated in 2014. The *Linea No.1* was built on an elevated viaduct and was the longest metro-type train viaduct in the world for 6 years until it was overtaken by the Wuhan Metro in 2017. This train connected the northeast side of the city with the southeast side and benefited over 2 million people. Until today, no other lines have opened to the public and *Linea Nro. 2* is still under construction since 2014. These delays are mostly explained due to several corruption scandals involving Garcia's government and Odebrecht, the consortium in charge of the construction, which allegedly paid more than 20 USD millions in bribes for this project. Appendix B compiles relevant information about the history of these projects.

In this paper, I focus on the *Linea No.1* and the *Metropolitano* (*M&M*), which reduced the transportation time from 2.5 hours to 1 hour on average. The new bus line and the metro provided a cleaner, faster, and safer service compared with traditional *combis.* This reduced transportation costs for thousands of students in Peru's capital. Notably, both systems crossed the city from north to south connecting several neighborhoods to downtown Lima which is the hub of several university campuses as seen in Figure 1.

Figure 1: University Campuses and $M\&M$ Stations across Lima



4 Data

4.1 Data Sources

This paper relies on multiple sources of data which include administrative data from college records, geocoded stations, and census block-level data.

Educational Outcomes: This information was provided by the Peruvian Ministry of Education, which annually compiles enrollment data for every university. These records contain information about students' year of enrollment, college ID, addresses, declared major, age, and gender. I focused on students whose home addresses are located within the Lima Metropolitana and Callao Region boundaries. I used the Google Maps API to collect GPS coordinates for their homes. For a small portion of cases where the algorithm failed, I imputed GPS coordinates at the block or neighborhood level. This sample accounts for less than 5% of the total cases.[9] I further narrowed

[9]The home address is self-declared by students at age 18 when they obtain their national ID, and this is typically validated with utility bills by the National Identification Agency in Peru (RENIEC).

down the sample to recent high school graduates or students under 19 years old for the analysis. The study's time frame ranges from 2006 to 2014, this year is particularly important since it marks the beginning of a significant higher education reform in Peru that denied operational licenses to one-third of colleges in the country for failing to meet basic quality standards (Alba-Vivar et al., 2023).

Geocoded College Campuses. The locations of 44 college campuses in the Lima and Callao Regions were manually collected and geocoded. These addresses were obtained from the 2010 College Census compiled by the Ministry of Education and the National Institute of Statistics and Informatics (INEI). The resulting GPS coordinates are plotted in red dots in Figure 1.

Peruvian Census. I obtained the Peruvian Census from 2007 and 2017 from INEI. Both datasets are geocoded at the block level, or to be specific, I use the *manzana* level, which is a unit bigger than a block but smaller than a ZIP code. 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. 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 and Callao Region and between 17 and 28 years old.

Transportation Data. I obtained all the information on stations from the *M&M* systems from the *Autoridad de Transporte Urbano para Lima y Callao* (ATU). This included the GPS location and address of all stations. I also collected information on planned but non-executed lines from the *Metro de Lima*. This information comes from multiple technical records from the national government (Ministry of Transport). I geocoded all planned stations from 6 routes as seen in Figure A.2.

Commuting Time. A key variable in this paper is how much time students travel when commuting to college. First, I calculate the average commuting time from a student's household to *any* college in the city, with and without the new systems in place. I use the road network data from OpenStreetMap API [10] which includes information on road type (highway, motorway, etc.). Then, I calculate the optimal route, defined as the shortest possible route from households to each college in the city. I follow (Velásquez, 2023) procedure and data to impute velocities for major highways

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

and the new lines, I also complement it with the Google Maps API data to obtain primary and secondary highway speeds. With this information, I computed commute times with and without the new lines. Given computational restrictions, I computed this travel time for a subset of households across the city that are representative at the district level.

Predicted College Access and College Choice. In an ideal setting, having information about students' preferences for college. This information is typically available in systems with centralized admissions systems. However, in the case of Peru, and similarly to most countries, there is a decentralized admission system, where each college has its own admission procedures. A contribution of this paper is that I use *machine learning* as a second best to recover students' choices. First, I measure whether a student is likely to go to college based on their neighborhood characteristics and then I measure the probability to attend a certain type of college. I use the k-nearest neighbors algorithm in order to predict the probability of attending college for each neighborhood before the new lines open to the public, meaning that I use the information before 2010. I also predict whether students were likely to go to a *high-return* private college, a *low-return*, or a public at the individual level. [11] *High-return* college are those whose graduates in 2014 earned more than 2250 PEN or approximately twice the minimum monthly wage while *low-return* are those whose graduates earned less than 1450 PEN. Appendix Figure A.3 shows the distribution of college wage returns and the thresholds used. I further refined the algorithm using a k-fold validation strategy and choose an optimal k parameter using GridSearch.[12]

Labor Market Returns. I use labor market outcomes compiled by the Ministry of Labor, *Planilla Electrica*. This dataset is a panel spanning from January 2014 to November 2019, and it includes monthly labor market outcomes such as wages, and hours work, etc.[13] It also includes some college information as major, gender, and college. I restrict my sample to students who graduated in 2014 and 2015 as these cohorts are not affected by the policy. I collapse the data at the $college \times major \times gender$ cell.

[11] 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.

[12] Details are described in Appendix D.

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

4.2 Descriptive Statistics

Table 1 shows the summary statistics for the main sample using information before the station's openings. Panel A shows the college enrollment rates with different specifications. Enrollment counts are the average number of students who enroll in any college in the city at the block level. As seen in the table, less than 1 student per block enrolls in college in my sample. When dividing by the denominator of population counts of the same age, we can see that on average 17 percent of students under 19 years old enroll in college.[14] I then transform these variables in logs (to be specific, I use the inverse hyperbolic sine function to correct for the substantial amount of zeros). It is also clear that, on average, women enroll in college at higher rates than men and that private college enrollment is higher than public college enrollment. Importantly, we can also observe that the distance from students HH to college is on average quite similar for both neighborhoods connected to new lines and those connected to the planned but not-executed lines. Panel B in Table 1 shows average statistics using the 2007 Census, which includes the total population and education levels achieved for people over 25 years old. There is no significant difference in terms of the population size for affected versus non-affected neighborhoods. However, the population over 25 years old seems to be slightly more educated in the treatment group.

5 Empirical Strategy

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

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

Let Y_{it} represent the outcome of interest, such as the college access rate, at the neighborhood level i during year t. The binary treatment variable, D, equals one if the neighborhood is connected

[14] To be specific, the college enrollment rates are defined as the following: $Access_{it}^{College} = \frac{TotalEnroll_{it}^{16-19}}{TotalPop_{it}^{16-19}}$

Table 1: Summary statistics at the neighborhood level (Pre-treatment)

	(1) Total	(2) M&M	(3) No M&M
Panel A :	College Admin. Records Data		
Enrollment (rates)	0.172	0.163	0.176
	(0.363)	(0.357)	(0.368)
Enrollment (log)	0.0911	0.0701	0.0842
	(0.331)	(0.285)	(0.325)
Female Enroll. (logs)	0.0543	0.0416	0.0503
	(0.251)	(0.216)	(0.247)
Male Enroll. (logs)	0.0446	0.0335	0.0422
	(0.225)	(0.192)	(0.223)
Public Enroll. (logs)	0.0281	0.0223	0.0285
	(0.179)	(0.157)	(0.184)
Private Enroll. (logs)	0.0690	0.0517	0.0622
	(0.285)	(0.242)	(0.277)
Distance HH to College	9.695	11.24	10.10
	(5.557)	(5.156)	(6.047)
Panel B:	2007 Census		
Total Population	128.2	123.0	123.6
	(123.8)	(122.2)	(111.2)
Primary School	0.197	0.215	0.201
Pop. Share	(0.143)	(0.141)	(0.146)
Secondary School	0.394	0.415	0.394
Pop. Share	(0.175)	(0.175)	(0.175)
Higher Ed.	0.409	0.370	0.405
Pop. Share	(0.240)	(0.230)	(0.238)
Observations	211,824	78,788	110,284

Notes. This table shows the means at the block (*manzana*) level before 2010. Panel A shows the college enrollment rates using the administrative data from MINEDU. Logarithmic transformations are adjusted for 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*).

to the $M\&M$ and zero if the neighborhood is connected to the planned but not executed line. $\mathbb{1}(\tau = t - T^*)$ consists of event-year dummies that represent the four years before and after the new service was opened. The coefficients of interest, ϕ_τ, demonstrate how the outcomes evolve over time following the opening, allowing for the possibility of heterogeneous effects on different routes. α_τ

indicates the pre-treatment effects in eventually treated neighborhoods relative to untreated ones, enabling us to test the parallel pre-trends assumption. Additionally, μ_i are the neighborhood fixed effects and psi_t are the year fixed effects.

However, there are a few empirical challenges when using this strategy. First, the staggered nature of the treatment might arise 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. What is more, in this setting, heterogeneous treatment effects are likely to arise from heterogeneity in how the *Metropolitano* and *Metro de Lima* connect to different colleges in the city. To address these potential issues, this study relies on the recent advances of the DiD literature.[15] In particular, I follow Borusyak et al. (2021) and implement their imputation estimator which allows for treatment-effect heterogeneity and dynamic effects.

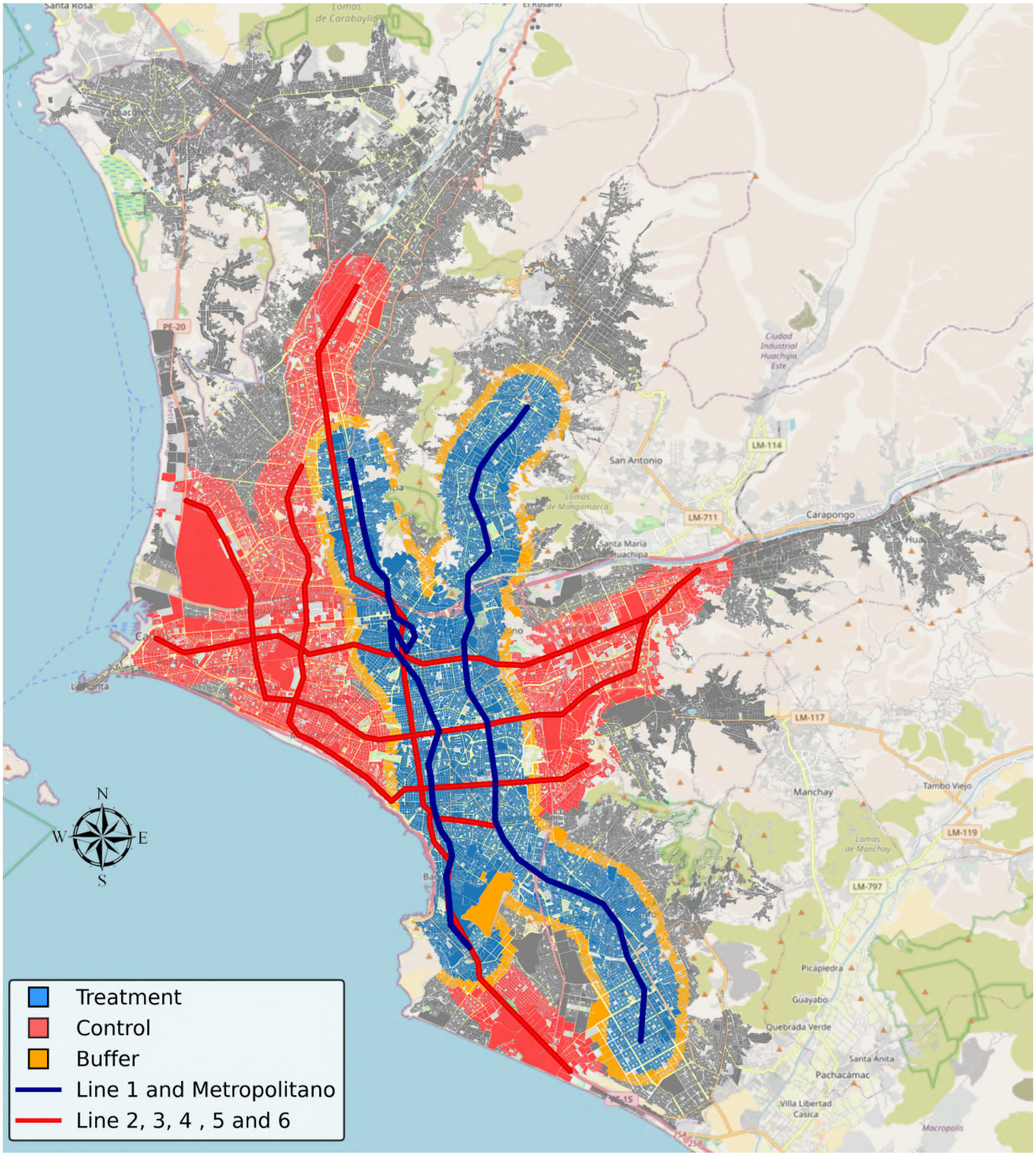
A second empirical challenge is to establish a proper control group. In this sense, simply comparing connected neighborhoods to non-connected neighborhoods within the city might overestimate our results since the allocation of the new routes is not completely random. One way to address this concern is using a placebo group as in Donaldson (2018). In this paper, the control group comes from those neighborhoods that could have been affected by the new transport system because there were planned lines that have not yet happened.

Figure 2 shows the neighborhood that belongs to the treatment and control groups. I define a neighborhood that is exposed to the executed lines as one that is within 1.5 kilometers of the nearest station as seen in Figure 2. The control group comes from those neighborhoods that are within 1.5 kilometers of the planned but not executed station. I also excluded neighborhoods that are simultaneously exposed to opened and planned but not executed lines as seen on the yellow shaded areas in 2. Restricting the control group to neighborhoods connected to planned but not-executed lines reduces the selection bias due to a potential correlation between the *M&M* placement and unobserved changes in access to college.

Finally, even when using the placebo lines to reduce selection bias, another potential issue can

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

Figure 2: Neighborhoods exposed to the executed and planned but not-executed lines



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

arise in this setting. When estimating the effects of these new transportation lines, using the timing of the opening of the station as an exogenous shock, other determinants of the outcome of interest (college enrollment rates) are still not random, recent work by Borusyak and Hull (2022) highlights this issue. For example, neighborhoods located in the city center (which is both, economically and geographically central) might experience higher enrollment growth since they were more likely to have a new line than those in the outskirts. What is more, families living in these areas are also more educated (see Appendix Figure **??**). Borusyak and Hull (2022) propose to use a *recentered treatment* as an instrument that removes the bias from the non-random shock exposure. Given that this paper follows a staggered DiD combined with an already improved control group, I further refine the sample by excluding neighborhoods in Lima Downtown to address these concerns. [16]

With the data available, I look at neighborhoods that were exposed to the *M&M* and use the administrative data and the latest Census available (2017) to obtain estimates at the block or *manzana* level. First, using the administrative enrollment data, I calculate yearly block-level college enrollment rates. The denominator comes from the total population counts from the 2007 Census. For this, I estimate Equation 4. Second, using individual-level data from the 2017 Census, I calculate age cohort completion rates and labor market outcomes at the block level. I estimate an exposure DiD as in Equation 5. Age cohorts will be considered treated if their residency block was exposed to the excuted lines by the time they were 17 years old, the age at which most high school students graduate. I also restrict the analysis to those individuals born in the period 1991 to 2000.[17]

$$y_{c,i} = \sum_{\tau=-4}^{0} \alpha_\tau D_i^{pre} \mathbb{1}(\tau = c - T^*) + \sum_{\tau=1}^{4} \phi_\tau D_i^{post} \mathbb{1}(\tau = c - T^*) + X\beta_{c,i} + \psi_c + \mu_i + e_{c,i} \tag{5}$$

[16] I define Lima Downtown as all neighborhoods in the following districts: Lima (Historical Center), Lince, Jesus Maria, San Isidro, Miraflores, Breña, La Victoria, and Rímac.

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

6 The impact of improved transportation on college education

6.1 Effects on commuting time

As a first-stage analysis, I investigate whether the new stations reduced commuting time to *any* college when a student's neighborhood is connected. To do this, I use a simple 2 by 2 different-in-difference model that leverages the opening of new stations and the treatment status of being connected to either the executed or planned but not executed lines. [18]. Note that this estimate is the most conservative one as informal routes, such as *combis*, are not included in the data, and I assume that students are commuting by car, which is an overestimate of their actual transportation. On average, the commuting time to any college before the new systems was approximately 1 hour, which is consistent with the self-reported data from the College Census of 2010. The results in Table 3 indicate that the introduction of the new system reduced the average commuting time to *any* college in the city by 17%, which translates to almost 30 minutes per day saved on commuting. It is important to note that these findings are based on the most conservative estimate, and the actual impact on commuting time could be even higher.

Table 2: Effects of new stations on commuting time (mins) to college

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

6.2 Effects on college enrollment

Using the enrollment administrative data, I find positive effects on college enrollment rates at the block level as seen in Table 3 Column (1). The coefficient shows a 1% p.p. increase relative to a

[18]The before and after variation can be visualized in Figure A.4a. Notably, people living in the northeast of the city seem to have the most benefits of the new transportation system

baseline 14% enrollment rate, measured as the number of students enrolled in a block divided by the counts of people of the same age living in the same block. Column (3) shows the impact on private enrollment while Column (4) shows the impact on public enrollment, which is significantly lower. This aligns with the fact that public colleges in Peru are more competitive than private colleges, and therefore reducing transportation costs may have a limited impact on enrollment since there is a heavier weight on skills. Columns (5) and (6) show that there is no significant difference in college enrollment for both women and men, with both showing positive and significant results.

Table 3: Effects of new stations on college enrollment rates

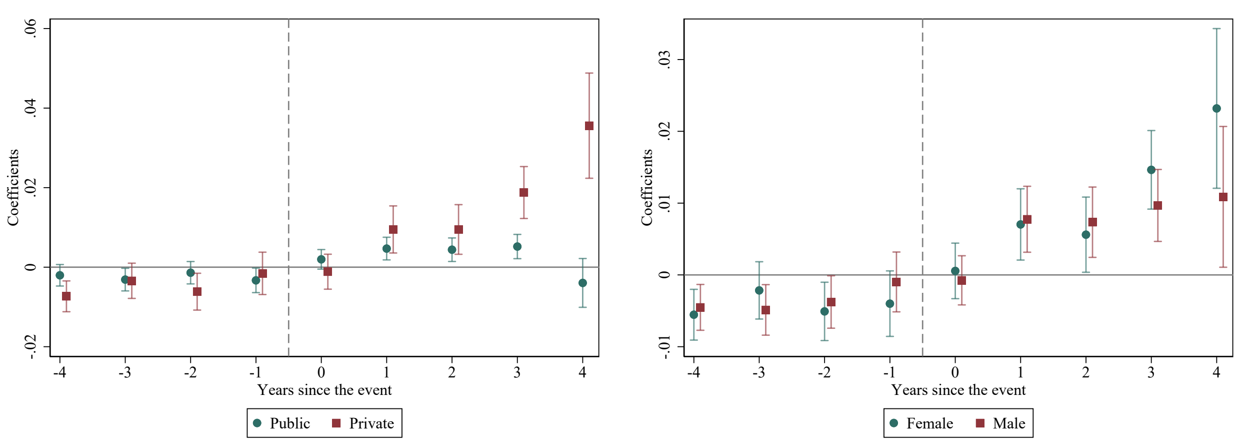
	Rates	Log(All)	Log(Private)	Log(Public)	Log(Female)	Log(Male)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment*Open	0.009***	0.012***	0.010***	0.003***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Dep. Var. Mean	0.142	0.080	0.059	0.027	0.047	0.039
N	411147	411147	411147	411147	411147	411147
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$.Regressions include block (*manzana*) and year-fixed effects. ATTs are calculated using the Borusyak et al. (2021) 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.

Figure 3: Dynamic Effects of the *M&M* on College Enrollment Rates by Groups

(a) By College Type

(b) By Gender



Notes. Regressions include block and cohort-fixed effects. The event study is calculated using the Borusyak et al. (2021) estimator. Panel (a) and Panel (b) shows the logarithm transformation of enrollment counts by each group adjusted by the hyperbolic sine.

When examining the dynamics of the effects, I observe that college enrollment rates increase since the first year of the event, and the magnitude of the effects doubles up to a 2.5% increase after three years of the new station openings (See Figure D.1 and further details in the robustness section). Additionally, the pre-treatment coefficients validate our findings as they demonstrate no prior trends before the implementation of the new lines. When analyzing the dynamic effects among subgroups, it can be seen that the positive effects for females grow faster over time compared to those for men, as shown in Figure 3b. This suggests that it couple of years for women to take full advantage of the new system. This also highlights that the benefits of the new systems go beyond reducing transportation costs and extend to increasing travel safety for women, who are particularly vulnerable in this city. Very importantly, there are significant differences by college type. Figure 3a shows that most of the effects are driven by enrollment in private colleges, whose enrollment keeps increasing over time. The effects for public college are also positive but significantly smaller in magnitudes with no economic relevance.

6.3 Effects on college choice

It is expected that the effects of the new lines are not only on the extensive margin (access to college) but also on the intensive (college choice). This is mainly motivated by the fact that there is an additional variation based on *which* college gets connected to the new lines (Appendix Figure A.5 shows these patterns over time.) Descriptive analysis suggests that colleges that get connected to the new stations enjoy an increase in their enrollment rates while the rest of the colleges suffer a decrease, suggesting that students are trading off between them.

In an ideal setting, students' college preferences and abilities would have been known to the researcher. This can typically be recovered using their revealed preferences in centralized admission systems. However, in this context and like many others in the developing world, there is neither a centralized admission system nor standardized testing available. I developed a new approach to deal with this limitation using machine learning techniques. I first obtained a predicted probability θ that a student i might have to go to a certain college type m, where $m = [HighReturns, LowReturns, Public]$ y also calculate the unconditional probability to enroll in a STEM major. Using the parameters from all the data available from the Census and students' characteristics before the policy takes place (details about who I calculated θ are in Appendix D). Then, I use the θ as a control variable in Equation 4 to generate my results. Table 4 shows the results for different types of colleges. First, Column (1) shows that students connected to the new stations are less likely to be enrolled in a high-return college by a statistically significant 1.3 percentage point. In contrast, students become more likely to be enrolled in a low-return college by 2 percentage points and by 3 percentage points in public colleges. Alternatively, I also test for other definitions of college quality, such as being an elite institution and being part of a national college college consortium or receiving an operational license after the 2014 Higher Education Reform, and the results are similar showing that students are less likely to enroll in such institutions.[19] I do not find statistically significant results on the likelihood of enrolling in a STEM major.

Table 4: Effects of the *M&M* on College Choice

	High Return	Low Return	Public	STEM
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.013**	0.020***	0.029***	-0.004
	(0.006)	(0.007)	(0.008)	(0.008)
Dep. Var. Mean	0.112	0.165	0.377	0.301
N	101187	101187	101187	101187
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

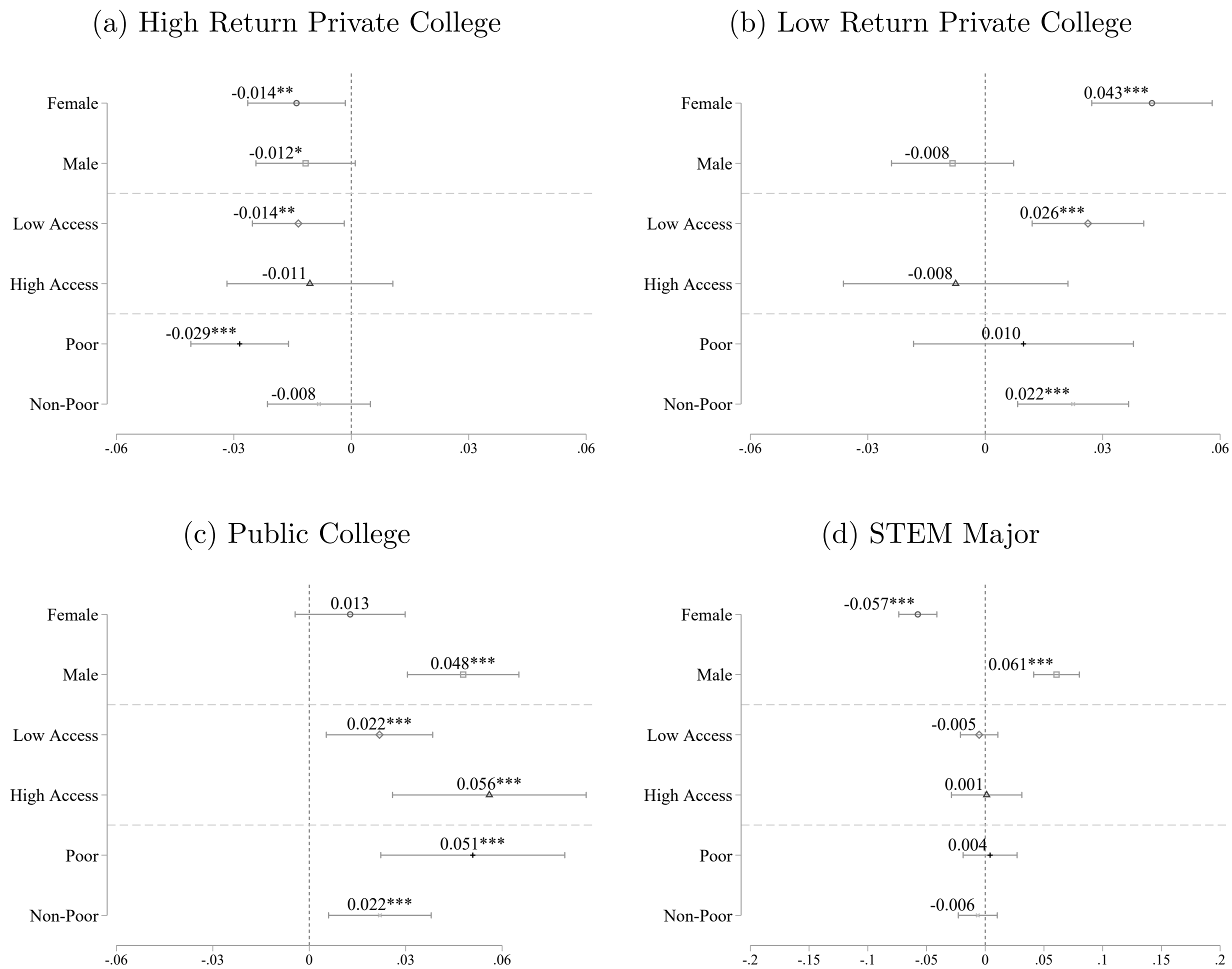
Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$.Regressions include block (*manzana*) and year-fixed effects. ATTs are calculated using the Borusyak et al. (2021) estimator for the first 4 years after a station opening. High Return Colleges are all private colleges whose graduates earn more than 2250 PEN while Low Return private colleges are the ones whose graduates earn less than 1300 PEN using administrative data of formal wages from recent graduates in 2014.

Notably, these effects are heterogeneous for different populations. Figure 4 shows the heterogeneous effects by groups for different college types. Figure 4a suggests that the negative effects on the likelihood of enrolling in a high-return college affect all types of populations but mainly women and the poor. However, the positive effects of enrolling in a low-return college suggest the effects are only significant for female students and non-poor populations. These effects are also driven by those new students who come from neighborhoods with typically low access to college. On the other hand, all populations are more likely to enroll in public colleges except female students. These results suggest that gender differences in a reduction in transportation costs do not show up in terms

[19]See Appendix Table ?? shows the results for these alternative definitions. I define elite colleges in Peru as the ones that are affiliated with the Consorcio de Universidades.

of access but they do in terms of choice.

Figure 4: Heterogeneous Effects by College Type



Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and year-fixed effects. High return Colleges are all private colleges whose graduates earn more than 2250 PEN while Low Return private colleges are the ones whose graduates earn less than 1300 PEN using administrative data from wage records in 2014.

6.4 Mechanisms

6.4.1 Distance to College

I explore whether the students who enroll in college are now opting for colleges that are located farther away. The rationale behind this is that since transportation becomes less of an issue, students have higher incentives to travel further and select colleges that are more distant from their homes.

Table 5 column 1 evaluates this hypothesis and shows the impact of the new routes on the distance between home and college. Contrary to what is expected, I find that students connected to new lines are enrolling in colleges that are 9% closer to them. What is more, they are 7% more likely to enroll in a college that is connected to the new lines as well. However, when breaking down these by whether the college is public or private, I find that students are more likely to enroll in private colleges that get connected.

Table 5: Effects on distance to college and college connectivity

	Distance (km.)	Connected	Priv. Connected	Pub. Connected
	(1)	(2)	(3)	(4)
Treatment*Open	-0.998***	0.0261***	0.0252***	0.001
	(0.089)	(0.007)	(0.007)	(0.005)
Dep. Var. Mean	10.894	0.382	0.223	0.159
N	120212	120212	120212	120212
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Distance is measured in kilometers using Euclidean distance from block centroid to college location. Columns 2-4 measure the probability of enrolling in a college connected to the new lines, with the closest station less than 3km away. Column 3 and Column 4 measure the probability of enrolling in a private and a public college connected to new lines, respectively. Regressions include block and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2021) estimator for the first 4 years after a station opening.

Additionally, I explore the heterogeneous effects by sub-groups. Appendix Figure A.6 panel shows that all types of students are less likely to enroll in a college that is further away from their homes, contrary to what was expected. The effects are significantly driven by students living in poor neighborhoods. Panel B shows that all types of students except the ones living in poor neighborhoods are more likely to enroll in colleges connected to new lines. However, when looking at the results by whether the college is public or private, females are more likely to enroll in private colleges that get connected while men are more likely to enroll in public colleges get connected.

7 Trade-off between college returns and travel time to college

The reduced-form results suggest that there are differential responses by gender, women are more likely to enroll in low-return private colleges while men are more likely to enroll in public ones. Therefore, to understand such differences, a key parameter to estimate is the student's willingness to pay for commuting time to school and college returns (or the wage premium of attending school c). I outline a simple model of college choice to recover students' preferences [20] I rely on a random utility model, where each student i maximizes an indirect utility function denoted as:

$$\begin{aligned} U_{ic} = & \quad \beta_i V_{ic} + \epsilon_{ic} \\ = & \quad \beta_i^w W_{ic} + \beta_i^t T_{ic} + \delta_c + \epsilon_{ic} \end{aligned} \tag{6}$$

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 effect of observed characteristics of students and ϵ_{ic} captures the effect of unobserved variable. V_{ic} can take a linear combination of W_{ic} wage premiums of graduating from college c, T_{ic} that represents travel time to college c and δ_c college fixed effects. A student i will choose a college c that maximizes their utility over the set of colleges available. The probability that a student i chooses college c, as in

$$\begin{aligned} P_{ic} = & \quad Pr(U_{ij} > U_{ik}) \quad \forall j \neq k \\ & \quad Pr(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}) \quad \forall j \neq k \\ & \quad Pr(\epsilon_{ik} - \epsilon_{ij} > V_{ij} - V_{ik}) \quad \forall j \neq k \end{aligned} \tag{7}$$

To summarize the effects I find in the reduced form, I am interested in measuring the trade-off between college returns and travel time to college. Similarly to Borker (2020), I use the marginal rate of substitution to measure how much travel time to college a student is willing to give up for an additional unit of college wage returns. This relationship is represented with the following equation:

[20] I follow Borker (2020) who implemented a similar model, although she incorporates more sophistication by using college routes.

$$MRS_i^{WT} = -\frac{\Delta T_{ic}}{\Delta W_{ic}} = -\frac{\beta_i^w}{\beta_i^t} \tag{8}$$

To obtain estimates of β_i^w and β_i^t, I estimate this model using a mixed logit framework with random coefficients. This mixed logit model can approximate any random utility model and it relaxes the Independence of Irrelevant Alternatives (IIA) property. This model overcomes the limitations of standard logit models by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time (Train, 2002). In particular, having random taste variation is useful since we expect heterogeneity in students' observables. To estimate this model, I assume that ϵ_{ic} is i.i.d extreme value and assume normally distributed coefficients for both T_{ic} and W_{ic}. The mixed logit probabilities are integrals of standard logit probabilities over a density of parameters, as seen in the following equation:

$$P_{ic}(\delta) = \int \frac{exp(V_{ic}\beta_i)}{\sum_{c=1}^{N} exp(V_{ic}\beta_i)} f(\beta_i|\delta)d\beta \tag{9}$$

where f is the mixing distribution. Since I am interested in evaluating female and male students' different responses suggested by the reduced form results, I estimate this model separately by gender. I also estimate the effects separately by college private and public colleges. As mentioned before, public colleges in Peru have higher requirements for admissions but they To make interpretation easier, I standardize college returns (wage premium), and travel time to college is measured in minutes.

7.1 Results

Table 6 shows the estimated of the mixed logit model. Column (1) shows the results for female students enrolling in private colleges. As expected, they are less likely to choose a college that demands more commute, in contrast, they are more likely to choose a college that provides higher wage returns. Column (2) shows the results for females enrolling in public colleges and the results show an interesting result: women are less likely to choose public colleges that provide higher returns. In this context, the public college that provides the highest wages in the market is the National University of Engineering (19% female enrollment in 2019) while among the lowest wages come from the National University of Education, a majority-female college. Column (3) shows that results for men enrolling in private colleges and their results have the same direction as their female peers. However, Column (4) shows that male students are more likely to enroll in public colleges that provide higher returns.

Table 6: Mixed Logit Results

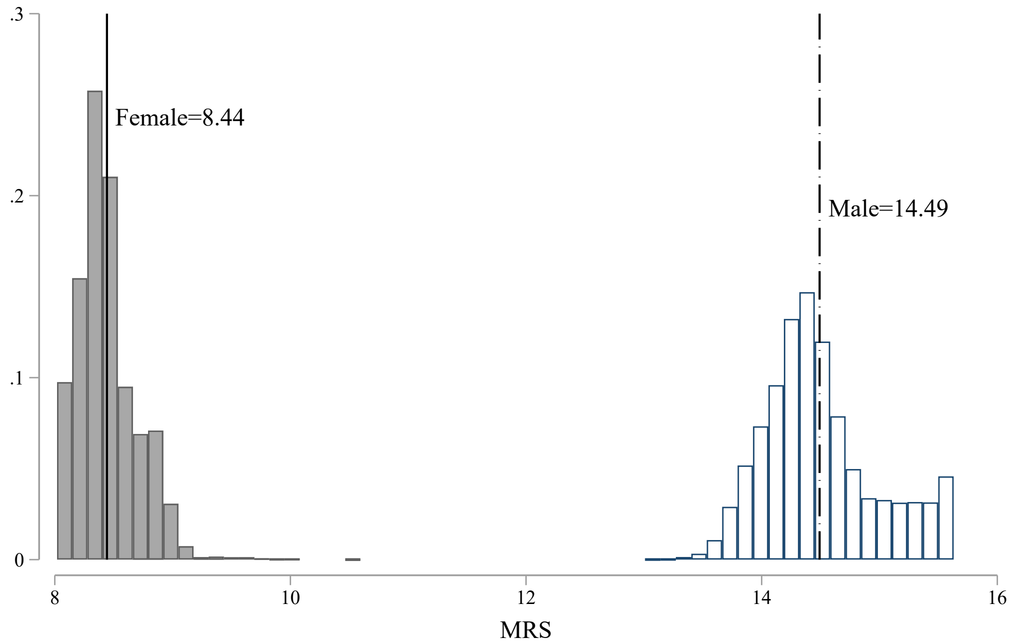
	Female		Male	
	Private	Public	Private	Public
	(1)	(2)	(3)	(4)
Mean				
Commuting Time	-0.016***	-0.019***	-0.018***	-0.024***
	(0.000)	(0.000)	(0.000)	(0.000)
Wage Premium	0.133***	-0.154***	0.254***	0.130***
	(0.003)	(0.004)	(0.003)	(0.004)
SD				
Commuting Time	0.004***	0.016***	0.005***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)
Wage Premium	-0.000	0.000	-0.0001	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Students	423,08	13,180	31,055	15,902
Observations	1,480,780	461,300	1,086,925	556,570
Log-Likelihood	-146174.1	-45448.1	-105800.4	-54003.7

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

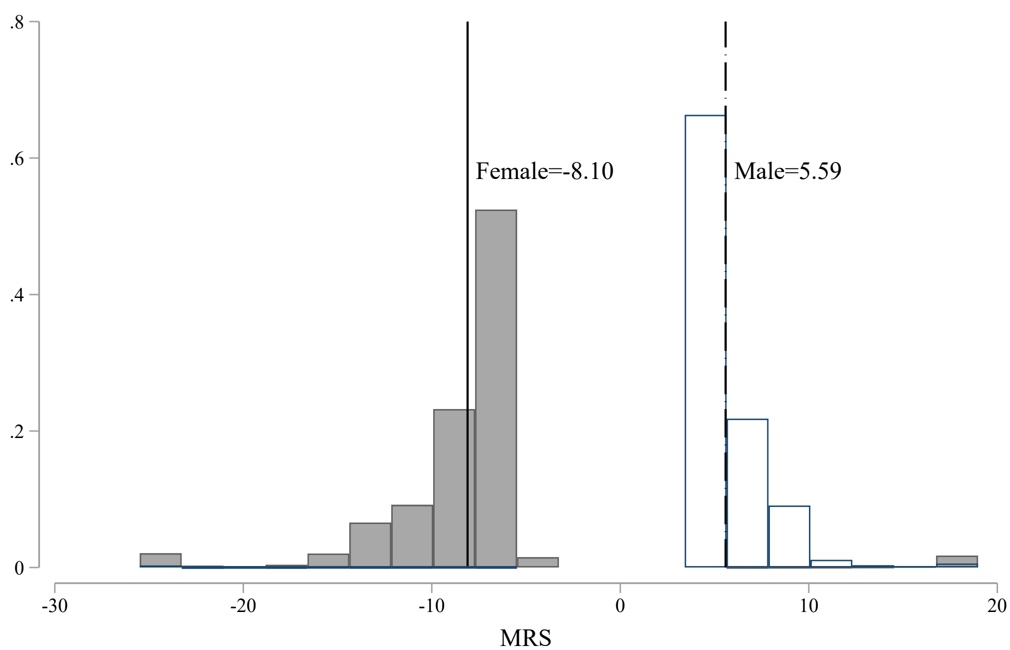
To evaluate the trade-off between college returns and commuting time to college, I estimate Equation 8 after calculating individual-level parameters corresponding to T and W, using the method proposed by Revelt and Train. (2000). Figure 5a shows the results for students enrolling in private colleges by gender. As expected, both men and female are willing to travel for one SD higher in returns. However, men are willing to travel almost twice as much as women. Note the distribution of both MRS do not overlap, showing that men and female are statistically different. When looking at the results for public colleges in Figure 5b, given that female students are less likely to enroll in high-return public colleges, they are also less willing to travel for higher returns. In contrast, men are willing to travel 5 more minutes for higher returns but this is almost 3 times less than what they are willing to travel for private colleges.

Figure 5: Marginal Rate of Substitution by Gender

(a) Private College



(b) Public College



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

8 Medium and Long-Term Effects

8.1 College Completion

When evaluating the medium-term effects, it is expected that the $M\&M$ will not only improve access to college but also enhance students' college experience and increase their chances of graduating on time. This can occur through two channels: (i) reduced commuting time can positively impact academic performance (as documented in Tigre et al. (2017)) and (ii) the new lines can increase access to internship opportunities, which are a crucial requirement for graduation in several programs. Given the lack of data on each channel, I estimated the overall reduced-form effect using the 2017 Census (7 years after the opening) and a DiD model that uses a cohort-exposure variation. The results show a positive impact on college completion rates (12%) compared to baseline rates, as shown in Table 7 column 1. The estimated coefficients are similar to whether Lima Downtown is included or not as seen in column 2. The dynamic effects on the event study are shown in Figure A.7. As a student is more exposed to the new lines, the more likely she is to complete college by 2017. Thanks to the rich individual-level information in the Census, I am able to analyze heterogeneous effects.

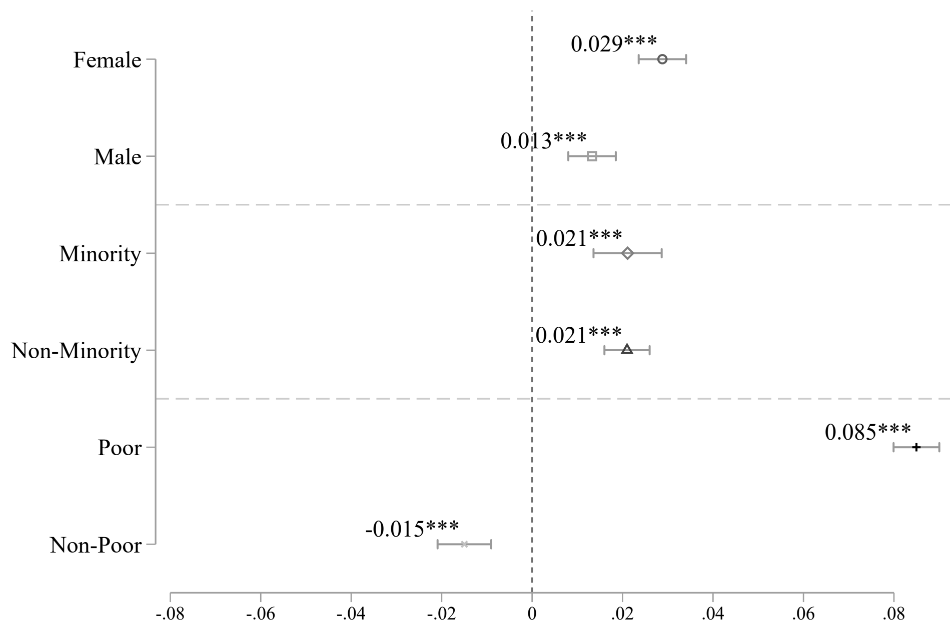
Figure 6 displays the heterogenous effects by sub-groups. Women and low-income students enjoy the benefits of the new lines in terms of college completion. The effects are driven by women and students living in neighborhoods where the average income is below the national median. I do not find significant differences between students who self-declare being part of a minority group (indigenous or Afro-Peruvian) and the majority ethnic group (mestizos). Surprisingly, non-poor students are getting negative effects. These types of students are the same who are enrolling in both public and low-return colleges. Two effects could be happening at the same time: (i) enrolling in a low-return quality might have been disappointing and induced students to drop out or take longer and (ii) an increase in labor market opportunities might have also encouraged students to work instead of study.

Table 7: Effects on College Completion

	No DT	Inc. DT
	(1)	(2)
Treatment*Open	0.021***	0.021***
	(0.002)	(0.002)
Dep. Var. Mean	0.173	0.182
N	497962	607928
Block FE	Yes	Yes
Year FE	Yes	Yes

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2021) estimator for the first 4 years of exposure.

Figure 6: Heterogeneous Treatment Effects of College Completion



8.2 Employment Rates

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

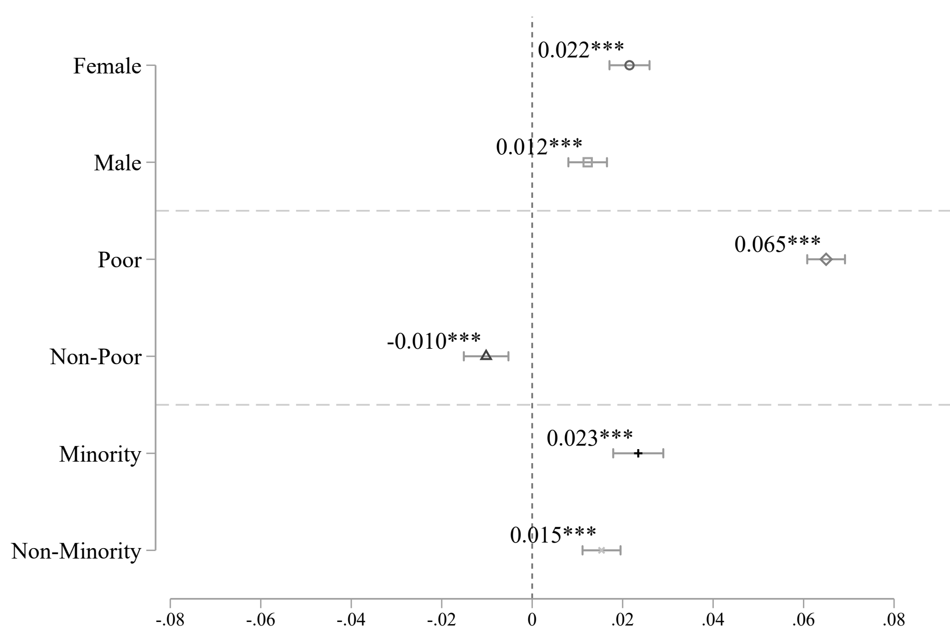
Figure 7 shows the heterogeneous effects of the impact on white-collar jobs by sub-groups. All populations, except, the non-poor enjoy these benefits. Women are twice as much likely to be working in a white-collar job than men. I do not find significant differences between students who self-declare being part of a minority group and the majority ethnic group. These results suggest that even when women and the poor are attending low-quality institutions, they are still more likely to graduate and obtain a higher-quality job in comparison with peers who are not connected to the new lines.

Table 8: Effects on Employment Rates

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

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$. Regressions include block (*manzana*) and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2021) estimator for the first 4 years of exposure.

Figure 7: Heterogeneous Treatment Effects of White Collar Employment



9 Conclusions

Countries are actively reshaping and redesigning their transportation systems since inefficient, unsafe, or irregular public transportation represents a large cost to the economy. What is more, recent advances in the literature are focusing on understanding the optimal transportation routes within cities. For example, Kreindler et al. (2023) show that having a less concentrated network would increase ridership and commuter welfare in Jakarta.

This paper studies the link between urban features and educational attainment. In particular, I study of two new public transportation systems in a large city in South America, and their effects on human capital investment. Unlike most of the related literature, this paper focuses on college education since these types of students are the ones who commute the most across big cities and greatly benefit from reducing transportation costs compared with students attending basic education, who typically commute shorter distances.

The results of this paper are important in the context of big cities where inequality is more striking. I find that a reduction of 17% in commuting time to any college available in the city can increase college enrollment rates by 1 percentage point. This is mostly driven by private colleges, that is this context, includes low-cost and low-quality institutions. When it comes to college access, most of the literature on education has focused on how monetary restrictions and other institutional factors limit access to higher education, but less is known about how reducing transportation costs for students can affect not only the decision to enroll in college but also in *which* college. In this sense, students connected to the new systems are choosing to travel less and are opting for colleges connected to the new system, even when these colleges are of lesser quality.

Women are particularly attracted to these colleges suggesting that they are more likely to trade off higher-quality colleges for safety reasons. Using a simple model, I estimate these trade-offs and show that men are willing to travel twice as much as what females will travel to a college which gives one standard deviation higher salaries. These results suggest that the benefits of enrolling in college are limited when students are choosing to enroll in colleges that will not provide the best possible labor market returns after graduation. Even when women are more likely to graduate and secure employment, choosing to attend low-return college seats them back in the race to break the existing glass ceiling.

References

Adukia, A., Asher, S., and Novosad, P. (2020). Educational investment responses to economic opportunity: Evidence from indian road construction. *American Economic Journal: Applied Economics*, 12(1):348–76.

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

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

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

Balboni, C., Bryan, G., Morten, M., and Siddiqi, B. (2021). Could gentrification stop the poor from benefiting from urban improvements? *AEA Papers and Proceedings*, 111:532–37.

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

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

Borusyak, K. and Hull, P. (2022). Non-Random Exposure to Exogenous Shocks. *(Revise and Resubmit, Econometrica)*.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. *Work in Progress*, pages 1–48.

Brooks, W. and Donovan, K. (2020). Eliminating uncertainty in market access: The impact of new bridges in rural nicaragua. *Econometrica*, 88(5):1965–1997.

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

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

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

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

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

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

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

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

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

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

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

Herskovic, L. (2020). The effect of subway access on school choice. *Economics of Education Review*, 78:102021.

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

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

Kreindler, G., Gaduh, A., Graff, T., Hanna, R., and Olken, B. A. (2023). Optimal public transportation networks: Evidence from the world's largest bus rapid transit system in jakarta. Working Paper 31369, National Bureau of Economic Research.

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

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

MTC (2005). Plan maestro de transporte urbano para el área metropolitana de Lima y Callao en la república del perú. *Agencia de Cooperación Internacional de Japón–JICA*.

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

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

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

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

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

Sviatschi, M. and Trako, I. (2021). Gender violence, enforcement, and human capital: Evidence from women's justice centers in peru. Technical report.

Tigre, R., Sampaio, B., and Menezes, T. (2017). The impact of commuting time on youth's school performance. *Journal of Regional Science*, 57(1):28–47.

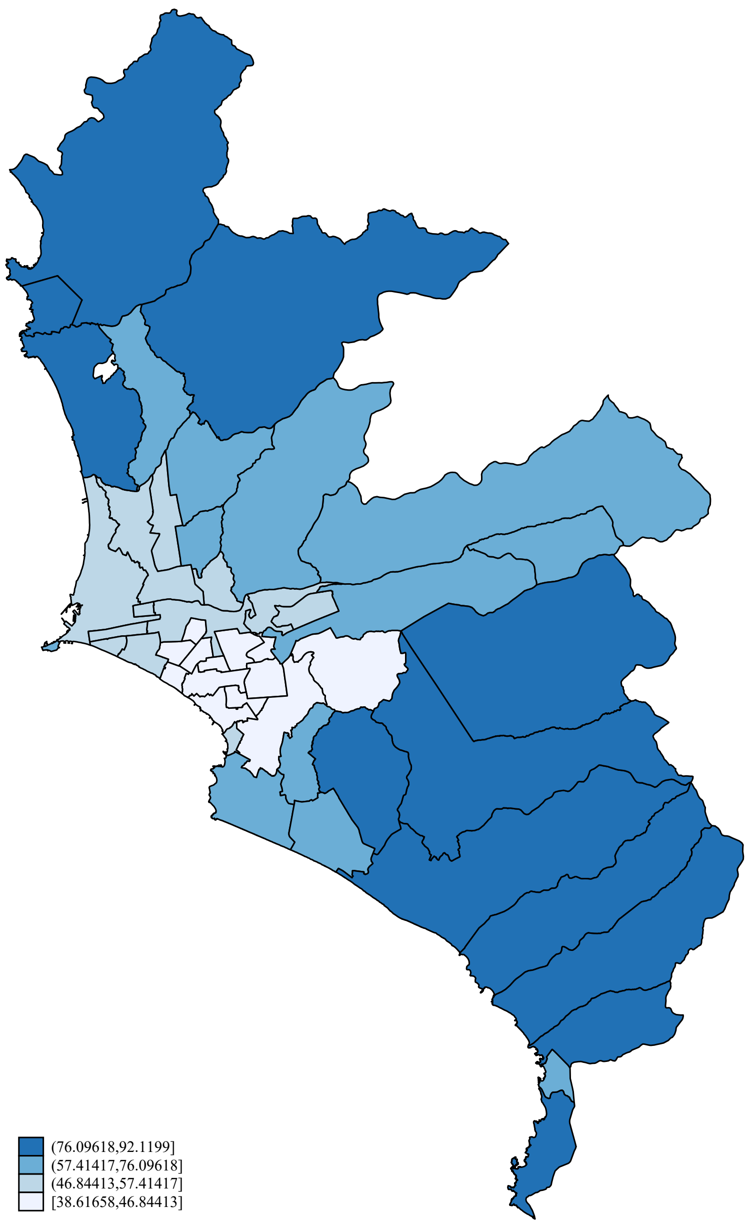
Tsivanidis, N. (2022). Evaluating the impact of urban transit infrastructure: evidence from Bogota's TransMilenio.

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

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

A Additional Figures and Tables

Figure A.1: Average Travel Time from Home to University (in min)



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

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

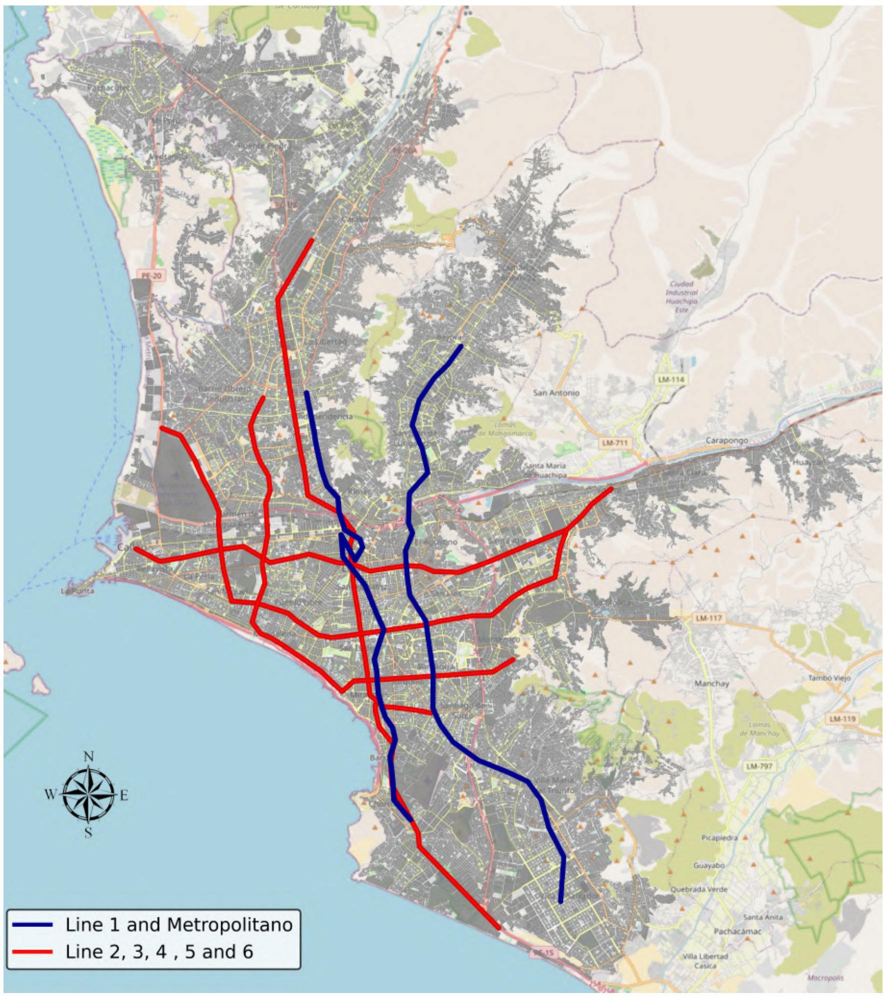
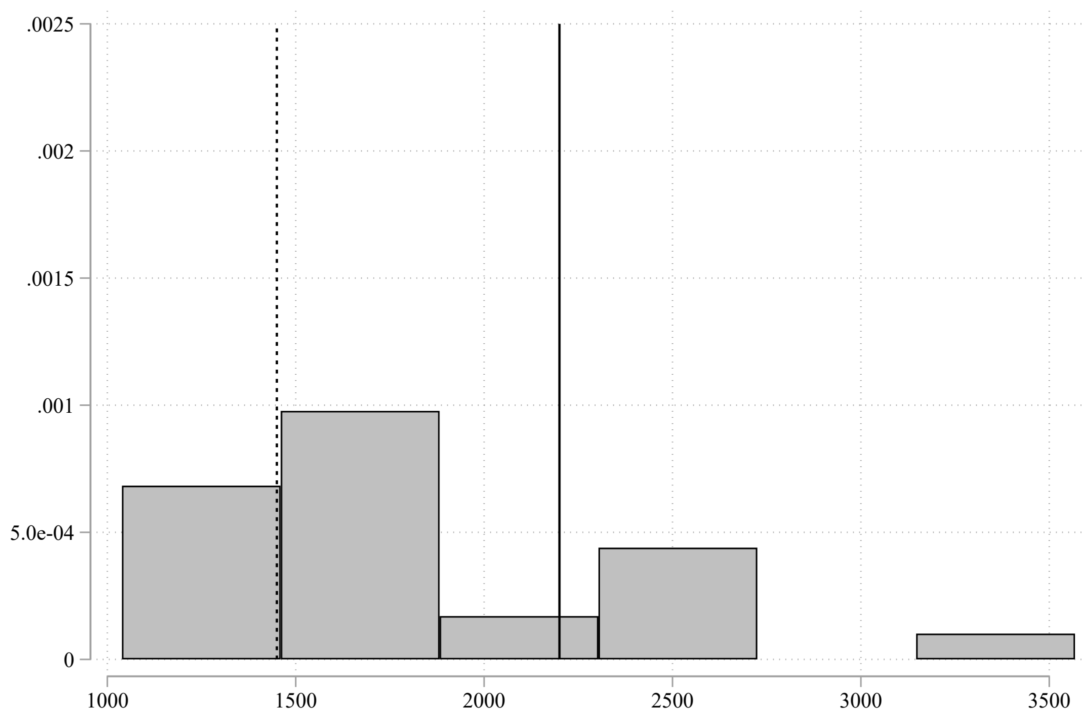
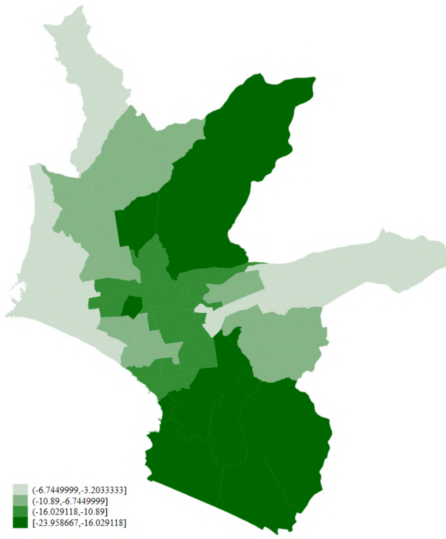


Figure A.3: Wage Distribution of Recent Graduates (Average by College)

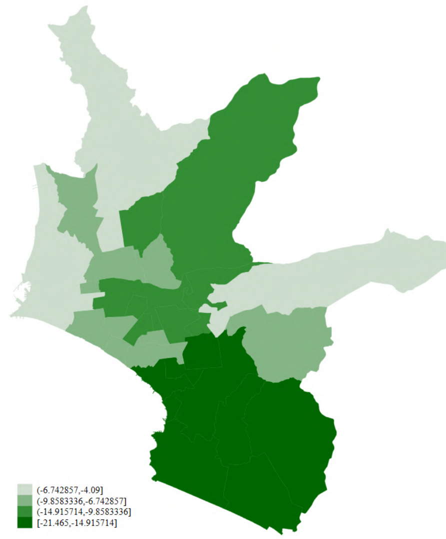


Source: Planilla Electronica. The dashed line shows the limit for the 25% bottom of the distribution below 1450 PEN while the plain line marks the 25% top of distribution above 2250 PEN.

Figure A.4: College Commuting Time (before/after the $M\&M$)



(a) All colleges



(b) Private colleges

Figure A.5: Log(enrollment) for connected and non-connected colleges

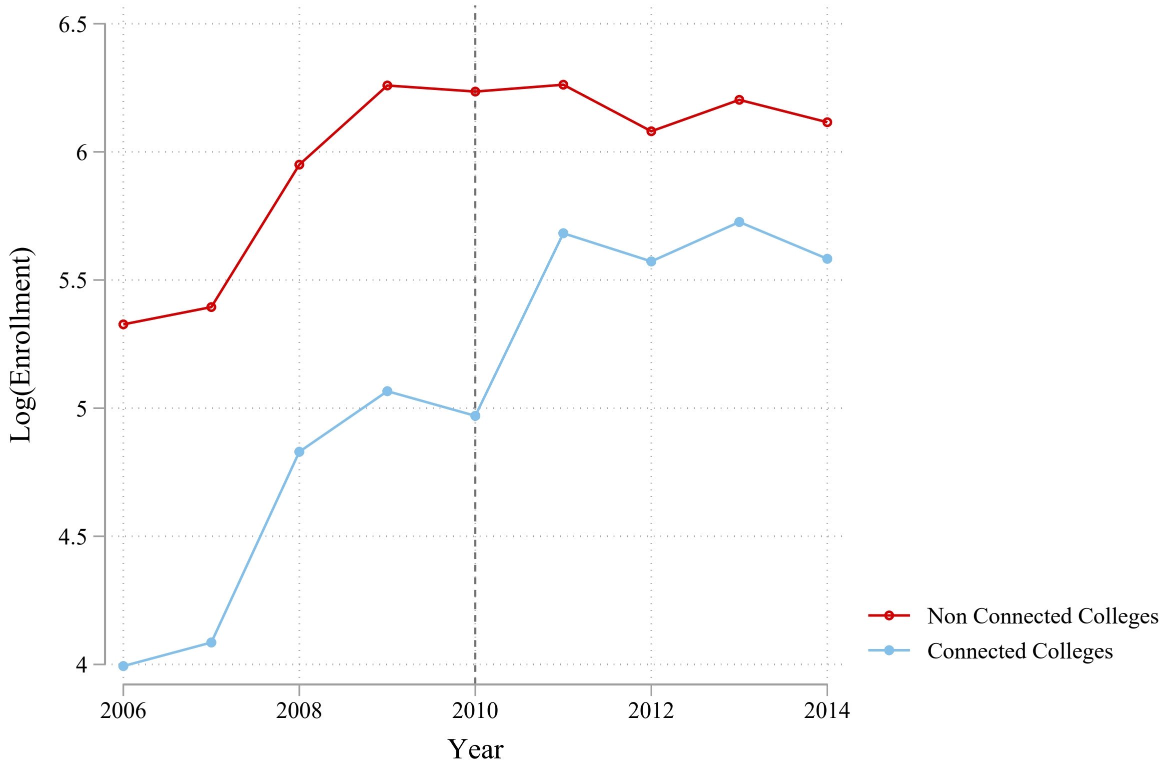


Table A.1: Effects on the likelihood of enrolling in a high-quality college

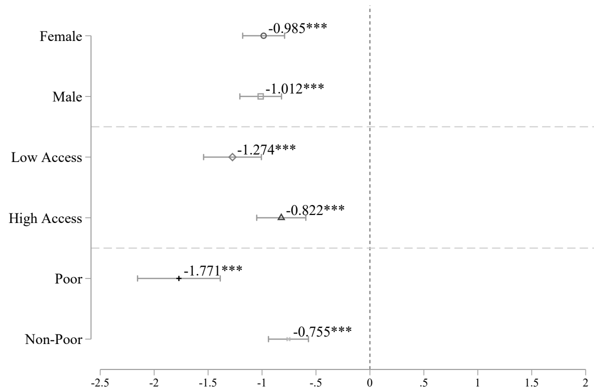
	Diff. Elite	Elite	Diff. Licensed	Licensed
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.035***	-0.032***	0.025***	-0.016***
	(0.007)	(0.007)	(0.007)	(0.006)
Mean	0.075	0.077	-0.835	0.882
N	101187	101187	101187	101187
Block FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Standard errors in parentheses

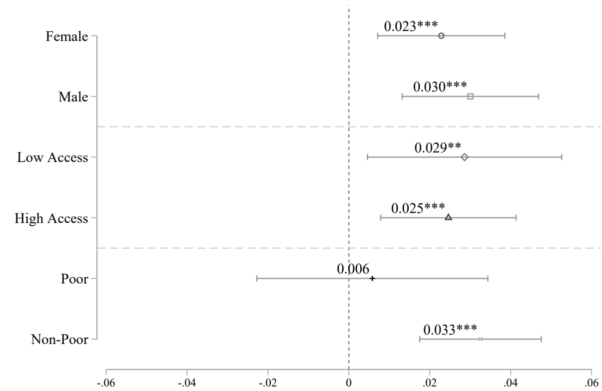
Errors clustered at the block level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

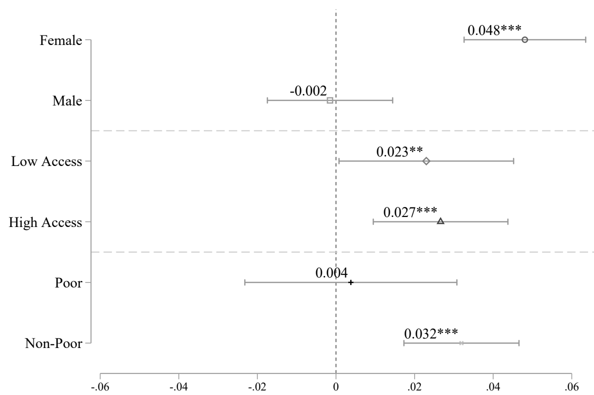
Figure A.6: Heterogeneous Effects of distance to college and college connectivity



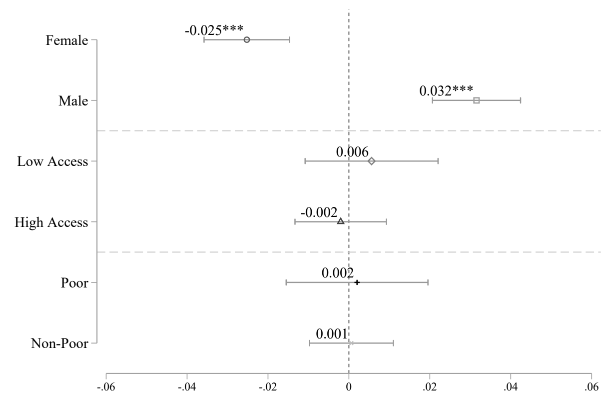
(a) Distance (km)



(b) College connected to the $M\&M$



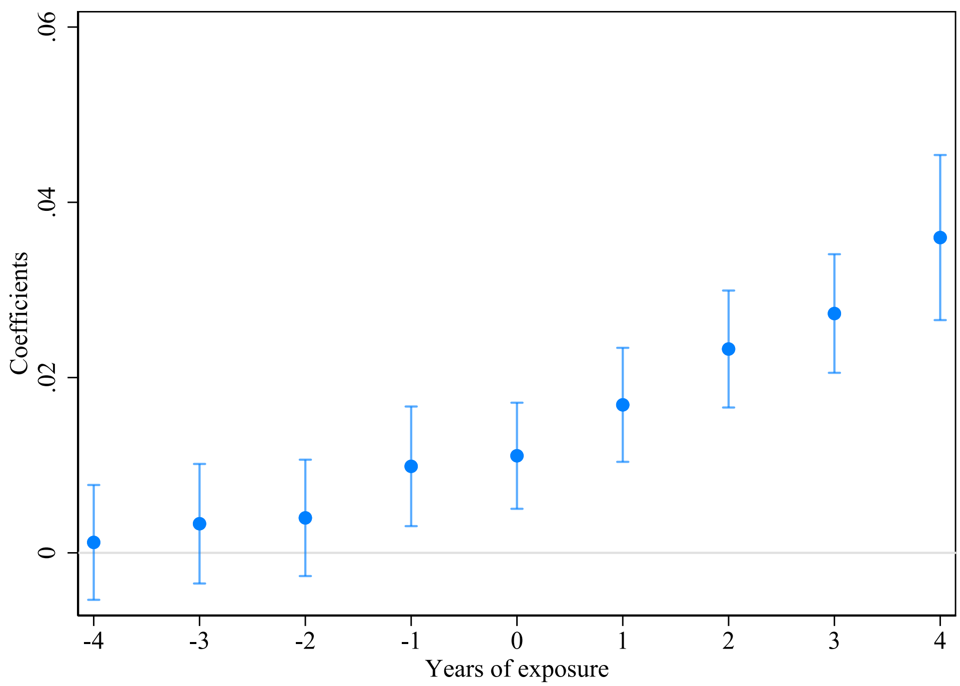
(c) Private connected to the $M\&M$



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

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

Figure A.7: Dynamic Effects of the $M\&M$ on College Completion Rates



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 economic and political crisis stopped the project from continuing. In the 1990s, AATE was part of the Council of Ministers and proposed a complementary study of the Lima Metro Network in 1998 (MTC, 2005). Then, at the beginning of 2000's, the Municipality of Metropolitan Lima developed the COSAC study, a preliminary study of a BRT (a Metropolitano antecedent). As a consequence, the AATE changed the route of line 1 (tramo II) from Av. Grau to San Juan de Lurigancho instead of Comas. [21] 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.

To sum up, the delay in the implementation of the Metro is explained by several factors. On the institutional side, the AATE [22] 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.

[21] The AATE also left the Council of Ministers and became part of the Municipality of Metropolitan Lima.
[22] In 2019, ATTE became the Urban Transport Authority (ATU, in Spanish)

C Predicted College Access and Predicted College Choice

In this section, I describe the procedures to calculate both the predicted access to college and the predicted college choice for certain types of colleges. For both analyses, I use the k-nearest neighborhood (kNN) algorithm, which is particularly good for a setting like this. The kNN is a non-parametric algorithm that uses proximity to make a prediction. One key advantage is that it is able to capture non-linearity. To avoid its sensitivity to the choice of a distance metric, I use the grid search technique to find the optimal value of k, the number of nearest neighbors to consider when making predictions. This optimal k maximizes the test dataset score using cross-validation. Table C.1 shows the results of this procedure.

The algorithm proceeds as follows:

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

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

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

Table C.1: kNN Scores

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

D Robustness Checks

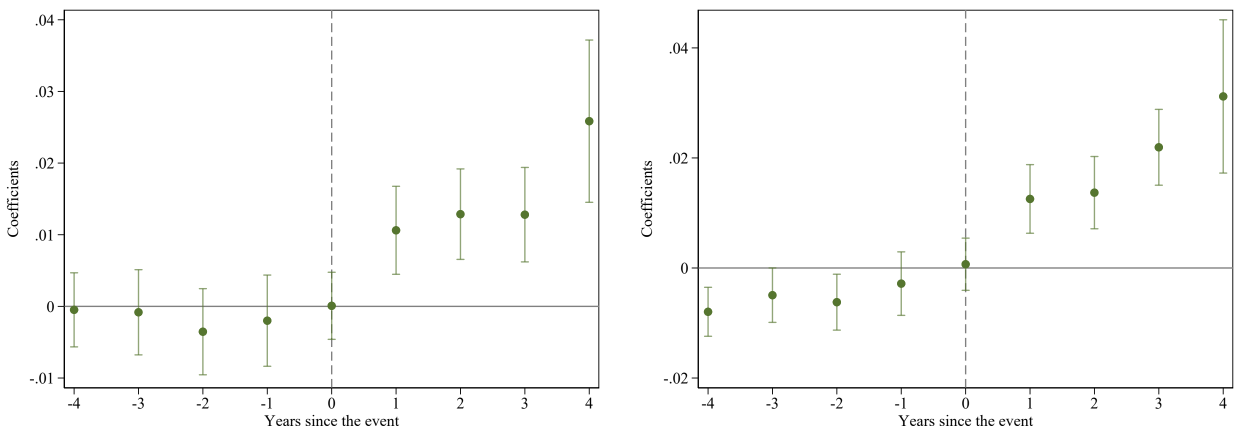
D.1 Dynamic Effects

Figure D.1 shows the dynamic effects of the *M&M* opening. First, this Figure validates the parallel trends assumption from the differences-in-differences strategy since there are no significant trends in the pre-treatment period. This is for two measures as seen in Panel (a) and Panel (b) which show the effects on enrollment rates and the logarithm transformation of the enrollment counts per block.

Figure D.1: Dynamic Effects of the *M&M* on College Enrollment Rates

(a) Enrollment Rates

(b) Log(Enrollment Counts)



Notes. Regressions include block and cohort-fixed effects. The event study is calculated using the Borusyak et al. (2021) estimator. Panel (a) shows the enrollment rates where the denominator is the total count of students enrolled divided by the total population of (potential) students of the same cohort. Panel (b) shows the logarithm transformation of enrollment counts adjusted by the hyperbolic sine.

D.2 Full Sample Including Lima Downtown

In Section 5, I highlighted that the main specification excludes districts in Lima Downtown since these areas of the city were very likely to get treated by any *potential* transportation line. What is more, this part of the city has a higher level of market access in comparison with the outskirts. Similar to the results on Table 3, when adding the districts in this central area of the city does no change the results significantly as seen on Table D.1. I still find positive effects on college enrollment rates at the block level (1% increase). The results from Columns 2-6 reveal a similar pattern to the main results in Table 3 but in a slightly lower magnitude suggesting that the neighborhoods in Downtown Lima experienced no impact.

Table D.1: Effects of the $M\&M$ on College Enrollment Rates including Lima Downtown

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

Notes. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.10$.Regressions include *manzana* and year-fixed effects. ATTs are calculated using the Borusyak et al. (2021) 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.

D.3 College Choice

I use an alternative specification to estimate the effects on college choice. Instead of controlling for the baseline predicted probability θ of enrolling in a type m college, I use as an outcome the relative choice on whether a student enrolls in a type m college relative to what is predicted on the baseline. The effects on Table D.2 show a similar pattern to the my main results.

Table D.2: Effects of the $M\&M$ on College Choice (relative to Predicted)

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

Standard errors in parentheses
Errors clustered at the block level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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. (2021) estimator for the first 4 years after a station opening.