Opportunity Bound: Transport and Access to College in a Megacity

Fabiola M. Alba-Vivar ⋆

(Job Market Paper)

May, 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 and 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. Access to rapid transit increases college enrollment by 1%, mainly driven by private college enrollment. While all students are less likely to enroll in high-return and elite institutions, women are more likely to enroll in low-quality private colleges and men are more likely to enroll in public colleges, which are also low-cost but more selective. In the medium and long run, access to transport increases a person's likelihood of graduating from any college by 5% and this gain is largest for women and lower-income students. College graduates from areas exposed to these lines are more likely to secure white-collar employment. This illustrates that the effects I estimate come in part through increasing people's ability to connect to labor market opportunities that better reward human capital accumulation.

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

This project has benefited from data provided by the Peruvian Ministry of Education and the National Institute of Statistics (INEI). I thank Alex Eble, Kiki 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 Matteo Magnaricotte, Carolina Concha-Arriagada, Eric Verhoogen, Jack Willis, Bentley MacLeod, Daniel Bjorkegren, Kevin Donovan, Sam Asher, Nishith Prakash, Marianne Bitler, Michael Best, Daniel Velazquez, Sofia Amaral, and Stephen Keating. I am also thankful to all participants at AEFP, 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.

⋆ Teachers College 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

The speed, safety, and efficiency of public transportation shape the economic productivity of the billions of humans who live in cities. As megacities around the world arise, complexities involved in meeting the needs and aspirations of millions residing in a single geographic region present increasingly formidable challenges (Li et al., 2021).[1] Countries are actively reshaping and redesigning their systems with this in mind, and most others regularly consider it. Furthermore, inefficient, unsafe, or irregular public transportation represents a large cost for the economy. In the US, the estimated loss of not improving public transportation represents a loss of 180 USD billion in cumulative gross national product and a loss of 109 USD billion in household income (American Public Transportation Association, 2018).

In this paper, I estimate the impact of improved public transport on a key driver of economic productivity: human capital accumulation via tertiary education. I study this in Lima, the capital of Peru and a megacity of approximately 12 million people, where two new mass public transportation systems were provided in the 2010s: the Metropolitano, a rapid bus transit, and the first line of the Metro de Lima. Here onward, I refer to these as the $M\&M$. I focus on the services the $M\&M$ provide to *commuter* college students, a common type of student across much of the developing and developed world. In Lima, as in elsewhere, college campuses are spread across the city. 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]

I use detailed geo-coded administrative college enrollment records and data on the location and openings of new stations to generate causal estimates of access to the $M\&M$ on a variety of outcomes related to tertiary education, and post-tertiary labor market outcomes. To do so, I

[1] Currently, one in eight people lives in the 33 megacities around the world which are predominately in the Global South. A recent report by Euromonitor highlights how most of the future upcoming megacities in the world will be from developing countries.

[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.

employ a difference-in-differences (DiD) strategy that exploits variation by cohort and neighborhood location. I further refine the DiD strategy by comparing neighborhoods exposed to new $M\&M$ stations versus neighborhoods that were exposed to planned but non-opened stations, following the placebo strategy implemented in Donaldson (2018).

My first finding is that the expansion of these transit lines led to a significant increase in college enrollment. I estimate a one percent per year increase in enrollment in affected areas, as compared with baseline rates prior to the $M\&M$. Interestingly, these effects are mostly driven by private college enrollment rather than public, even when Peruvian public colleges are virtually free. Nevertheless, public institutions are quite competitive in comparison with private colleges that have relatively easier admission systems.[3] Women are slightly more likely to enroll in college than men, especially after 2 years since the $M\&M$ opening but the differences are not statistically significant. These results are consistent with previous literature that documents how increasing women's increasing mobility has positive effects on their human capital (Borker, 2020; Fiala et al., 2022; Muralidharan and Prakash, 2017)

When looking at the intensive margin, I find that students who get connected to the $M\&M$ are more likely to enroll in low-return private and public colleges and less like to enroll in high-return and elite colleges. At this point, gender differences start appearing. Women seem to be less likely to enroll in high-return or elite colleges but they are mostly driving the low-return enrollment increase. 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 a low-return private college 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.

Using the 2017 Census, I further explore if students living in neighborhoods exposed to the $M\&M$ are more likely to complete college compared with older age-cohorts students. The results suggest that college completion rates are 5 percent higher compared to the baseline rates. This implies that students who enrolled in college are more likely to finish on time when connected to the $M\&M$ and effects increase for those who were exposed to the longest. This is an important outcome since most Peruvian college students tend to complete their education later than expected

[3] Flor-Toro and Magnaricotte (2021) document the disparities among the admissions systems for both private and public in Peru.

(> 5 years). The effects are particularly higher for women, low-income students, and minorities.

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. Therefore, I also examine medium- and long-term outcomes, such as employment rates. My finding suggests that there is a positive effect on employment rates for those who enrolled in college. Especially, students who enrolled in college are more likely to be employed in a white-collar job.

This paper contributes to the current literature on education and urban economics. 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. This work provides causal estimates of the opening of a city's major public transportation system and its effects on college access. There are three 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).[4] Muralidharan and Prakash (2017) study the impact of providing bicycles to female students (a reduction on transportation costs) and find that being exposed to the program increased girls' age-appropriate enrollment in secondary school by 32 percent. In this paper, the *M&M* affects university enrollment as it directly decreases transportation costs in the city, therefore I complement previous work by showing how a reduction in transportation costs can affect college enrollment decisions (on an extensive margin).

Another important potential mechanism is travel safety which is important for women living in urban spaces that limit their mobility. Borker (2020) explores how the perceived risk of street harassment can help explain women's college choices in Delhi, especially those who commute through unsafe routes. 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, which is considered one of the worst large cities in the world for women's mobility and transportation, comparable to

[4]Tigre et al. (2017) document how the duration of commuting has a negative causal effect on academic achievement using data from Brazil.

New Delhi, Mexico City, and Jakarta.[5] My results are consistent with previous literature and the fact that the $M\&M$ increased safety and allowed for higher mobility for women. I find that female students are getting most of the benefits from the $M\&M$: they do not only access to 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 $M\%M$.

This project also contributes to the literature on the economics of transportation. Most of this literature regarding the economic impact of improving or building new transportation shows positive effects on economic activity, trade, and labor market opportunities. However, little is known about the effects on education. My conceptual framework builds up on the market access approach (Borusyak and Hull, 2022; Donaldson and Hornbeck, 2016; Tsivanidis, 2022). 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 $M\&M$ commute 17 percent less time than those who do not get connected, this is equivalent to almost 30 minutes per day. However, it is also worth highlighting that the $M\&M$ is not only reducing transportation costs but also providing better labor market opportunities as it connects people from the outskirts of the city to 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 a similar spirit, 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. I complement this literature by exploring how labor market opportunities changed for recent high school graduates and how this interacts with those who access college.

A final contribution of this paper is to the literature on the geography of inequality. Most of these papers have shown how moving to places with more opportunities can positively impact income mobility, especially for economically disadvantaged populations and minorities (Chetty et al., 2020; Chetty and Hendren, 2018). On the other hand, accessing better schools within multiple districts can

[5]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.

have positive effects on accessing college for minorities although it might also increase driving-related offenses Bergman (2018). In this project, I explore how $M\&M$ can create a way to commute to opportunities by increasing access to college for vulnerable populations and subsequently, accessing higher-quality jobs.[6] I also complement the work of Flor-Toro and Magnaricotte (2021) who study the effects of college openings in Peru. They find that the opening of new college campuses in low-SES areas increases enrollment, but the effects for minority students are only half the size of those for others, widening pre-existing gaps. They document how proximity is highly valued by less-advantaged students, who disproportionately attend lower-quality high schools. They focus on all regions except Lima, the subject of this paper. I show how an alternative policy that simply reduces transportation costs in a large metropolitan area affects college enrollment, with potential applications in a wide range of developing countries' metropolises considering major public transit investment.

The remainder of this paper is organized as follows: in section 2, I present the setting of the $M\&M$ and the data. In section 3, I present the empirical strategy. Section 4 includes the main results of the paper and Section 5 concludes.

2 Background and Data

2.1 College Education in Peru

The Peruvian Education system is based on 3 levels: primary education (6 years), secondary education (5 years), and the higher education level which often lasts from 2 (technical school) to 10 years (School of Medicine). On average, college students take between 5 and 7 years to graduate. According to the 2017 Peruvian Census, approximately 4 out of 10 recent high school graduates (between 17 and 21 years old) have access to some type of higher education. More specifically, 15% 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.

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

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.

2.2 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), 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 and the lack of regulation became the basis of the new transportation system for commuters. These smaller, privately-operated minibusses known as *combis* partially alleviated the demand for transportation across the city and became the main mode of transportation for students. However, as demand kept increasing and the lack of quality from these buses, this mode of transportation became a hazard and unsafe. The institutional efforts to regulate them were unsuccessful. In July 2010, after 4 years of construction, the *Metropolitano* was opened to service. The *Metropolitano* was the very first mass transportation public system in Peru. This system connected the north and the south side of the city (12 districts out of the 44 in the city). There is a flat fee of 2.50 PEN, approximately USD 0.83 for regular commuters but students have a 50 percent discount. 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 its current price.

A year after the *Metropolitano*'s opening, the Peruvian president inaugurated *Line No.1* of the *Metro de Lima* which connected the north-east side of the city with the south-east side. [7] This corridor connected two of the biggest districts in Lima and benefited over 2 million people. The *Line 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. As of today, there are two lines in operation, with several more under plans for construction. In this paper, I focus on the *Line No.1* and the *Metropolitano* ($M\&M$), which reduced the transportation time from 2.5 hours to 1 hour on average, providing a cleaner, faster, and safer service. This reduced transportation costs for thousands of students in Peru's capital. Notably, the $M\&M$ 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.

2.3 Data

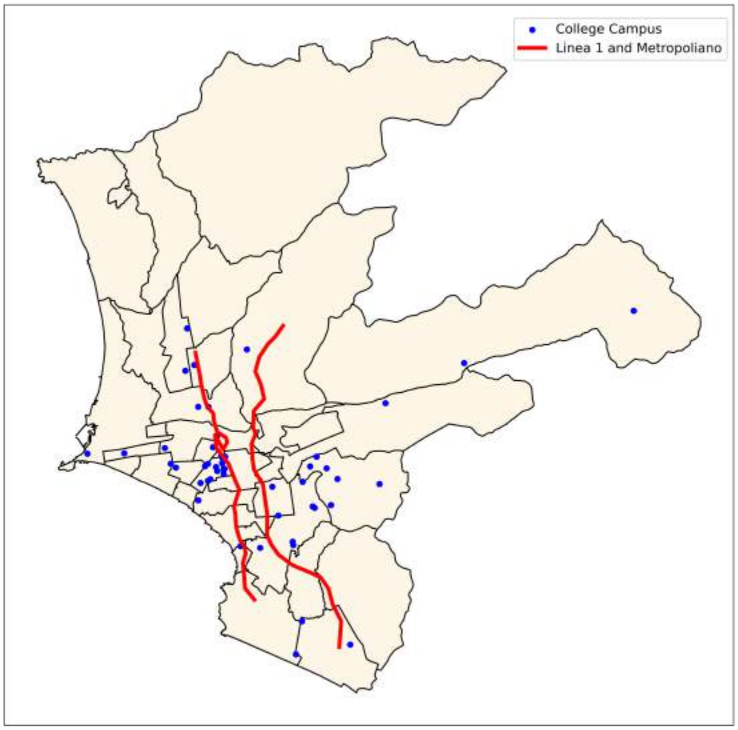
This paper relies on multiple sources of data: administrative data from college records, geocoded $M\&M$ stations, and census block-level data.

Educational Outcomes: 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 within the Lima and Callao Region boundaries and 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. [8] I further narrowed down the sample to recent high school graduates or

[7] Lima's Metro project started in the '70s and its construction began in 1986 but it was actually never finished as the economic and social crisis hit the country.

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

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



students under 19 years old for the primary analysis. The study's time frame ranges from 2006 to 2014, ending with 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 52 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 INEI. The resulting GPS coordinates were plotted in blue on Figure 1.

Peruvian Census. I obtained the Peruvian Census from 2007 and 2017 from INEI. Both datasets are geocoded at the block level (or *manzana* level). Importantly, these data sets include education achievement and employment status. I restrict my sample to individuals living in Lima and Callao Region. I use the data from 2007 to obtain block-level counts by age and use this as the denominator for college access rates. I use the 2017 data to explore long-term outcomes such as college completion and employment status.

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. I calculated the average commuting time to any college in the city before and after the $M\&M$. I use the road network data from OpenStreetMap API [9] which includes information on road type (highway, motorway, etc.). Then, I calculated the optimal route: the shortest possible route from a random sample of HH to each college in the city. I follow Velasquez's (2023) procedure and data to impute velocities for major highways and the $M\&M$, and I also complement it with the Google Maps API data to obtain primary and secondary highway speeds. With this information, I computed commute times before and after the $M\&M$.

Predicted College Access and College Choice. I use machine learning to measure whether a student was likely to go to college based on their neighborhood characteristics and to measure the probability to attend a certain type of college (public, private high returns, private low returns). First, I use the k-nearest neighbors algorithm in order to predict the probability of attending college for each neighborhood in the absence of the $M\&M$, meaning that I use the information before 2010. Second, I also predict whether students were likely to go to a *high return* private college, a *low return*, a public and an elite college at the individual level. *High return* college are those whose graduates in 2014 earned more than 2000 PEN or approximately twice the minimum monthly wage while *low return* are those whose graduates earned less than 1300 PEN. I further refined the algorithm using a k-fold validation strategy and choose an optimal k parameter using GridSearch.[10]

3 Empirical Strategy

This paper follows a Difference-in-Differences (DiD) strategy that exploits neighborhood exposure to the $M\&M$ 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:

[9] The information is public and the package *osmnx* is available on Python.
[10] Details are described in Appendix B.

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

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 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.[11] 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. According to different sources I collected, there were

[11] 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).

multiple plans for the city metro lines, but given budgetary restrictions, only one was implemented. The following lines suffered from several delays (up to 10 years) given some corruption scandals involving the government and the concession holder that built *Linea 1.* [12]

Figure 2 shows the neighborhood that belongs to the treatment and control groups. I define a neighborhood that is exposed to the $M\&M$ 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.

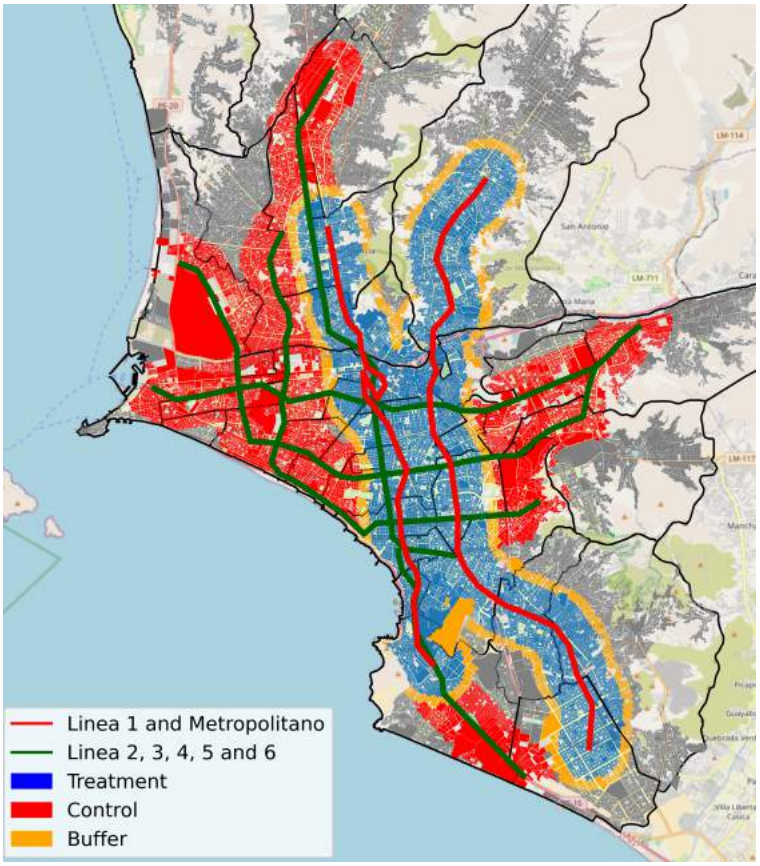
Finally, even when using the placebo lines to reduce selection bias might seem enough, an additional challenge is that most areas in the city center will be treated since most planned lines aimed to be connected in this area. This means Lima Downtown has higher levels of market access than those in the outskirts as this area typically gets connected by multiple ways of transportation. Recent work by Borusyak and Hull (2022) highlights this issue, which is particularly important for those using instrumental variables when, and propose to use a *recentered treatment* as an instrument that removes the bias from non-random shock exposure. Given that this paper follows a staggered DiD combined with an already improved placebo group, I further refine the sample by excluding neighborhoods in Lima Downtown to address these concerns. [13]

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 1. 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 2. Age cohorts will be considered treated if their residency block

[12] Appendix C compiles relevant information about the history of these projects.

[13] 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.

Figure 2: Neighborhoods exposed to the $M\&M$ and planned but not-executed lines



Notes. Blue-shaded areas are neighborhoods within 1.5 km distance from the nearest $M\&M$ executed station 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.

was exposed to the $M\&M$ 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.[14]

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

3.1 Descriptive Statistics

Table 1 shows the summary statistics for the main sample using information before the $M\&M$ openings. Panel A shows the college enrollment rates with different specifications. Enrollment

[14]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.

counts is 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.[15] 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 the $M\&M$ 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 $M\&M$ versus non $M\&M$ neighborhoods. However, the population over 25 years old seems to be slightly more educated in the treatment group.

4 Results

4.1 College Access

Using the enrollment administrative data, I find positive effects on college enrollment rates at the block level (1% increase) as seen in Table 2. Column 1 on in Table 2 shows the results. Despite this, the results including Lima Downtown and those using simple rates instead of logs still display similar and consistent results. Column 3 illustrates the impact on private college enrollment while Column 4 highlights the effects on public college enrollment, which are significantly lower. This aligns with the notion that public colleges in Peru are more competitive than private colleges, and therefore reducing transportation costs may have a limited impact on accessibility. There is no significant difference in college enrollment for both women and men, as shown in Columns 6 and 7.

When examining the dynamics of the effects, it can be observed that college enrollment has risen since the first year of the event, and the magnitude of the effects doubles up to a 2.5% increase

[15]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 before the $M\&M$

	(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 Pop. Share	0.197	0.215	0.201
	(0.143)	(0.141)	(0.146)
Secondary School Pop. Share	0.394	0.415	0.394
	(0.175)	(0.175)	(0.175)
Higher Ed. Pop. Share	0.409	0.370	0.405
	(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*).

after three years of the $M\&M$ openings (See Figure 7 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 $M\&M$. 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

Table 2: Effects of the $M\&M$ on College Enrollment Rates

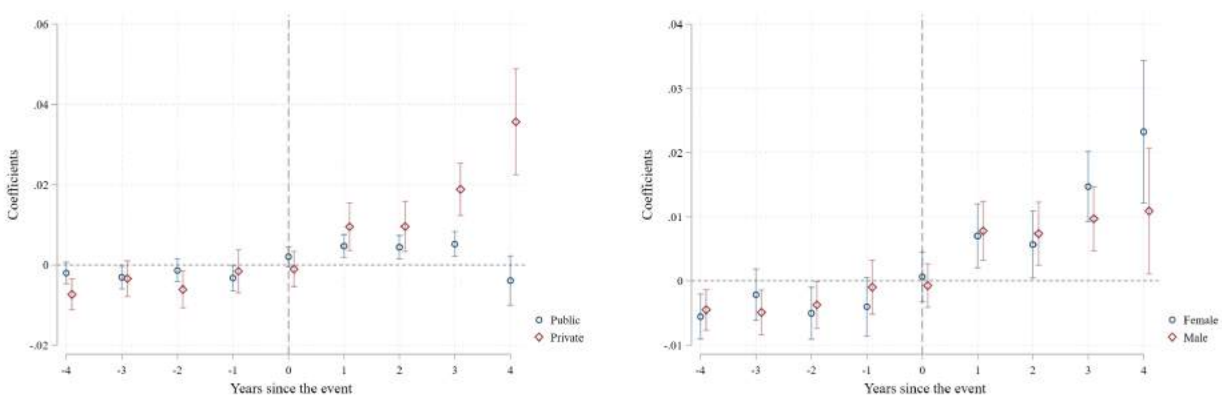
	All	Log(All)	Private	Public	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment*Opening	0.009***	0.012***	0.010***	0.003***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Dep. Var. Mean	0.142	0.080	0.059	0.027	0.047	0.039
N	411147	411147	411147	411147	411147	411147
Block FE	Yes	Yes	Yes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

shown in Figure 3b. This suggests that it took women a couple of years to take full advantage of the new system. This also highlights that the benefits of the $M\&M$ go beyond reducing transportation costs and extend to increasing travel safety for women, who are particularly vulnerable in this city. Figure 3a shows that most of the effects are driven by enrollment in private colleges. Interestingly, by the third year after the opening, there appears to be a trade-off between private and public colleges.

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.

4.2 College Choice

It is expected that the effects of the $M\&M$ 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 $M\&M$. Descriptive analysis suggests that colleges that get connected to the $M\&M$ enjoy an increase in their enrollment rates while the rest of the colleges suffer a decrease

In an ideal setting, students' college preferences and abilities would have been known to the researcher via a centralized admission system. However, in this context and like many others in the developing world, there is neither a centralized admission system nor standardized testing available. I develop 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 $m = [HighReturns, LowReturns, Public, Elite]$ using the parameters from all the data available from the Census and students' characteristics before the $M\&M$ takes place (details about who I calculated θ are in Appendix B). Then, I use the θ as a control variable in Equation 1 to generate my results.

Table 3 shows the results for different types of colleges. First, Column (1) shows that students connected to the $M\&M$ are 3% more likely to enroll in private colleges whose graduates enjoy the highest salaries in the market. Similarly, Column (4) shows similar results for elite colleges where the likelihood of enrolling in such colleges decreases by 2%.[16] There is also an increase in public colleges by 2,6%. On the other hand, the likelihood of enrolling in low-return colleges which are typically low-cost too, increases by 3,6%.

Notably, these effects are heterogeneous for different populations. Figure 4 shows the heterogeneous effects by groups for different college types. Figure 4a and 4d suggests that the negative effects on the likelihood to enroll in a high-return college affect all types of populations but mainly women and the poor. Similarly, the positive effects of enrolling in a low-return college affect all populations. However, women seem to be driving the effects in comparison with men. Women are twice as likely of enrolling in a low-return college than men. On the other hand, women are less likely to enroll in public colleges but men are 7% more likely to do it. These results suggest that gender

[16] I define elite colleges in Peru as the ones who are affiliated to the Consorcio de Universidades.

Table 3: Effects of the $M\&M$ on College Choice

	High Return	Low Returns	Public	Elite
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.030***	0.036***	0.026***	-0.020***
	(0.004)	(0.004)	(0.006)	(0.004)
Dep. Var. Mean	0.101	0.077	0.383	0.052
N	185114	185114	185114	185114
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 2000 PEN while Low Return private colleges are the ones whose graduates earn less than 1000 PEN using administrative data from wage records in 2014.

differences in a reduction in transportation costs do not show up in terms of access but in terms of choice. Women are more attracted to the low-return colleges because they are located in Lima Downtown and they are easily connected to the $M\%M$ while men are attracted to public colleges that are more dispersed across the city and take the advantage of traveling further distances.

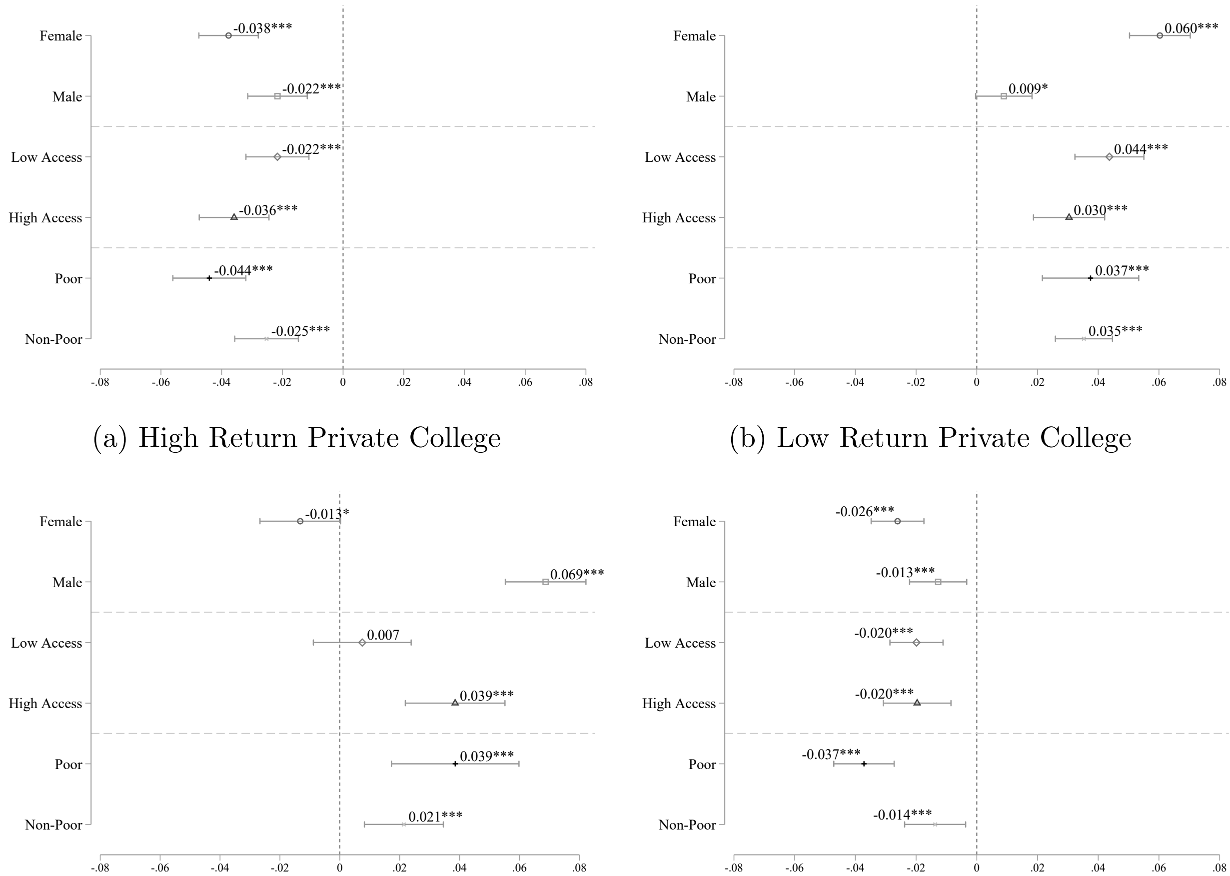
4.3 Medium and Long-Term Effects

4.3.1 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. 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 $M\&M$ can increase access to internship opportunities, which are a crucial requirement for graduation. Given the lack of data on each channel, I estimated the overall effect using the 2017 Census (7 years after the opening) and a DiD model that employs a cohort-exposure variation. The results show a positive impact on college completion rates (5%) compared to baseline rates, as shown in Table 4. The estimated coefficients are similar to whether Lima Downtown is included or not.

The dynamic effects on the event study are shown in Figure 5 which shows that the more exposed to the $M\&M$ a student, 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

Figure 4: Heterogeneous Effects by College Type



(a) High Return Private College

(b) Low Return Private College

(c) Public College

(d) Elite College

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 2000 PEN while Low Return private colleges are the ones whose graduates earn less than 1000 PEN using administrative data from wage records in 2014.

displays the estimated coefficients by groups which shows that the most vulnerable populations enjoy the benefits of the $M\&M$. Most of the effects are coming from women and not men and students living in neighborhoods where the average income is below the national median. Also, students who self-declare being part of a minority group (indigenous or Afro-peruvian) have a higher likelihood to complete college thanks to the $M\&M$.

Table 4: Effects of the $M\&M$ on College Completion Rates

	All	Excl. DT
	(1)	(2)
Treat*Opening	0.008***	0.008***
	(0.002)	(0.002)
Mean	0.156	0.147
N	941754	828292
Block FE	Yes	Yes
Cohort FE	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 manzana and cohort-fixed effects. ATTs are calculated using the Borusyak et al. (2021) estimator for the first 4 years after a station opening.

Figure 5: Dynamic Effects of the $M\&M$ on College Completion Rates

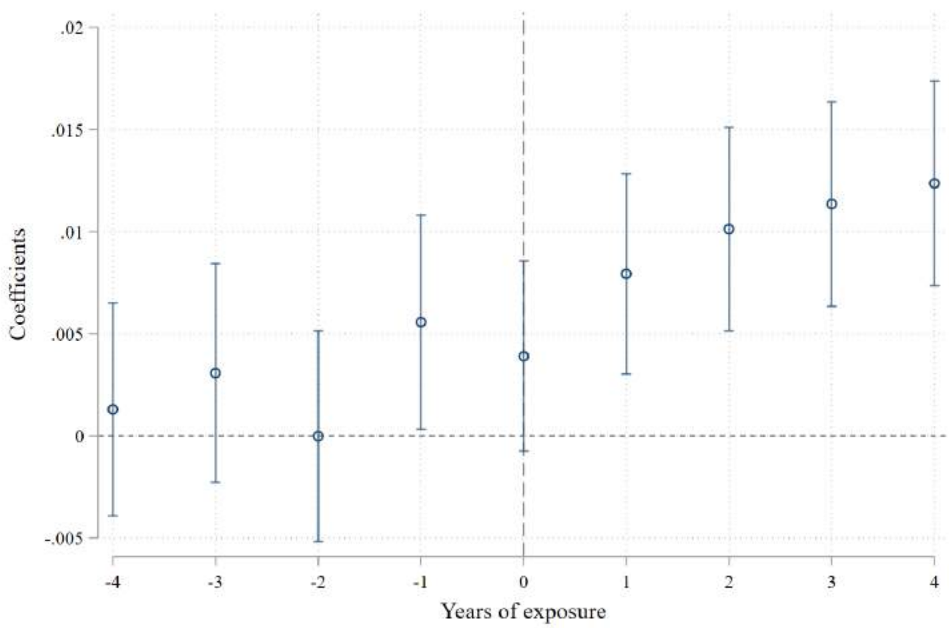
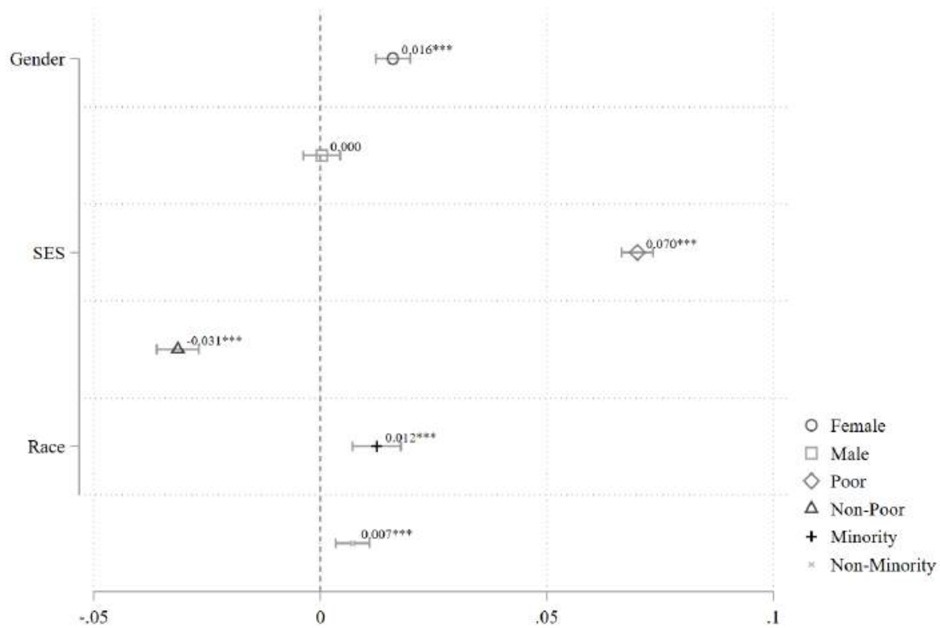


Figure 6: Heterogeneous Treatment Effects of College Completion



5 Mechanisms

5.1 Commuting Time

A natural concern is that $M\&M$ might have not actually affected commuting patterns for students if the colleges are located far away from stations. To address this, I investigate whether the $M\&M$ routes reduced commuting time on average to any college when a student's neighborhood is connected. To do this, I use a simple 2 by 2 difference-in-difference model that leverages the opening of the $M\&M$ system and the treatment status of being connected to either the $M\&M$ system or planned but not executed lines. [17]. 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 $M\&M$ system was approximately 1 hour, which is consistent with the self-reported data from the College Census of 2010. The results in Table 2 indicate that the introduction of the $M\&M$ system reduced the average commuting time to any college in the city by 17%, which

[17] The before and after variation can be visualized in Figure A.3a. Notably, people living in the northeast of the city seem to have the most benefits of the new transportation system

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 more significant.

Table 5: Effects of the $M\&M$ on Commuting Time (mins) from HH to *any* College

	All Colleges	Excl. Downtown	Private	Public	Elite
	(1)	(2)	(3)	(4)	(5)
Treat*Opening	-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

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

5.2 Distance to College

An additional channel I explore is whether the students who enroll 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 greater freedom to select colleges that are more distant from their homes. Table 6 validates this hypothesis and presents the impact of $M\&M$ on the distance between home and college. The results suggest that following the establishment of $M\&M$, students travel 6% farther to attend college. Compared to baseline

Table 6: Effect of the $M\&M$ on Distance (kms.) from Households to College

	Log Distance	Distance	Distance (incl. DT)
	(1)	(2)	(3)
Treatment*Opening	0.0659***	0.148***	0.130***
	(0.00654)	(0.0286)	(0.0262)
Mean	10.53	10.53	9.656
N	411147	411147	461151
Block FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

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

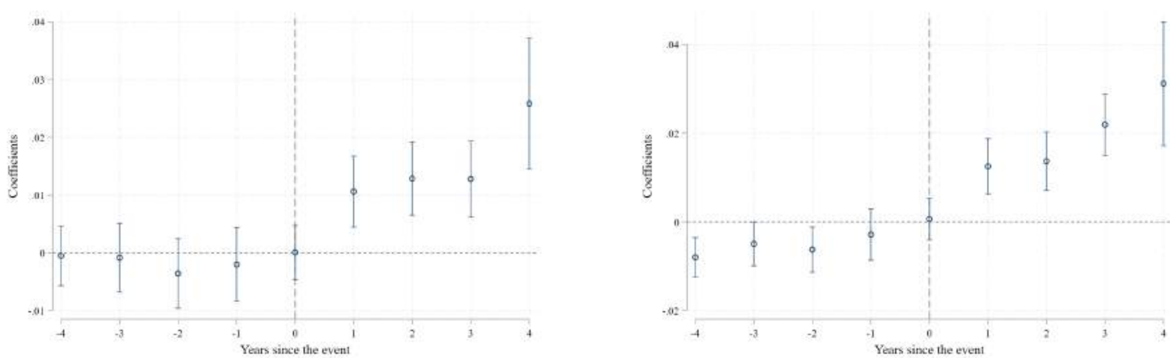
6 Robustness Checks

6.1 Dynamic Effects

Figure 7 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 7: 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.

6.2 Full Sample Including Lima Downtown

In Section 3, 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 2, when adding the districts in this central area of the city does no change the results significantly as seen on Table 7. 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 2 but in a slightly lower magnitude suggesting that the neighborhoods in Downtown Lima experienced no impact.

Table 7: Effects of the $M\&M$ on College Enrollment Rates including Lima Downtown

	All	Log(All)	Private	Public	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment*Opening	0.009***	0.011***	0.008***	0.003***	0.006***	0.005***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
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.

6.3 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 8 show a similar pattern to the my main results.

Table 8: Effects of the $M\&M$ on College Choice (relative to Predicted)

	High Return	Low Returns	Public	Elite
	(1)	(2)	(3)	(4)
Treatment*Opening	-0.033***	0.016**	0.018**	-0.020***
	(0.005)	(0.007)	(0.007)	(0.004)
Dep. Var. Mean	0.086	0.077	0.135	0.052
N	185114	185114	185114	185114
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.

7 Conclusions

This paper studies the link between urban features and educational attainment. In particular, aims to narrow the gap between education and urban economics and study the effects of a public transportation system in a large city in South America, similar to many other megacities around the world. 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 younger students attending basic education.

The results of this paper are important in the context of big cities where inequality is more striking. 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. On the other hand, the literature on the impacts of transportation improvements has mostly looked at the effects on macroeconomic indicators such as growth, trade, and labor, but education has often been overlooked.

This paper brings novel evidence and studies the case of the first transportation system in a megacity. The results suggest an increase in enrollment rates but this is mostly driven by private colleges, that is this context, includes low-cost and low-quality institutions. These results suggest that the benefits of enrolling to college are limited when students are choosing to enroll in colleges that will not provide the best possible labor market returns after graduation. The results also suggest gender differences in college choice. Women are trading off high-return, elite, and especially public colleges to attend low-return colleges given that they are getting relatively closer thanks to the new system. Even when women are more likely to graduate and secure employment, choosing to attend low-return college seats them back on 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.

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.

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.

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

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

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

Li, Z., Moore, A., and Staley, S. (2021). Megacity mobility: Integrated urban transportation development and management. *CRC Press, First Edition.*

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.

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*.

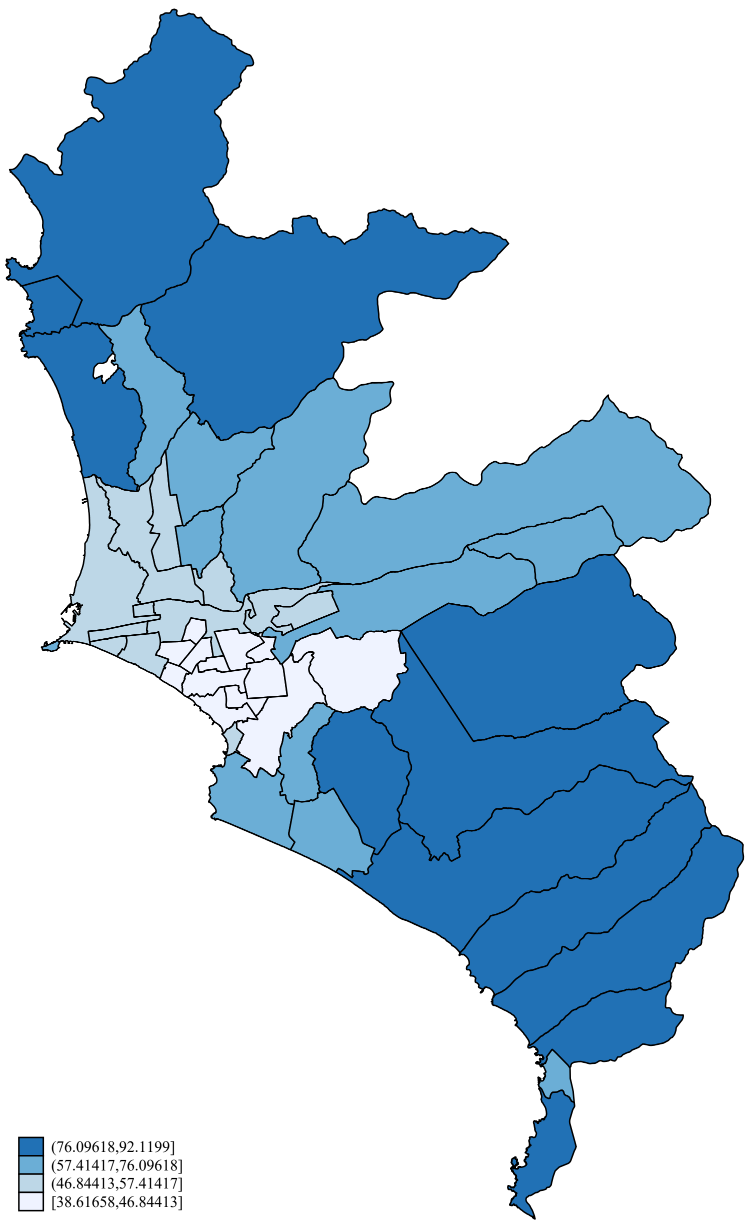
Sviatschi, M. and Trako, I. (2021). Gender violence, enforcement, and human capital: Evidence from women's justice centers in peru. Technical report.

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.

A Additional Figures and Tables

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



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

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

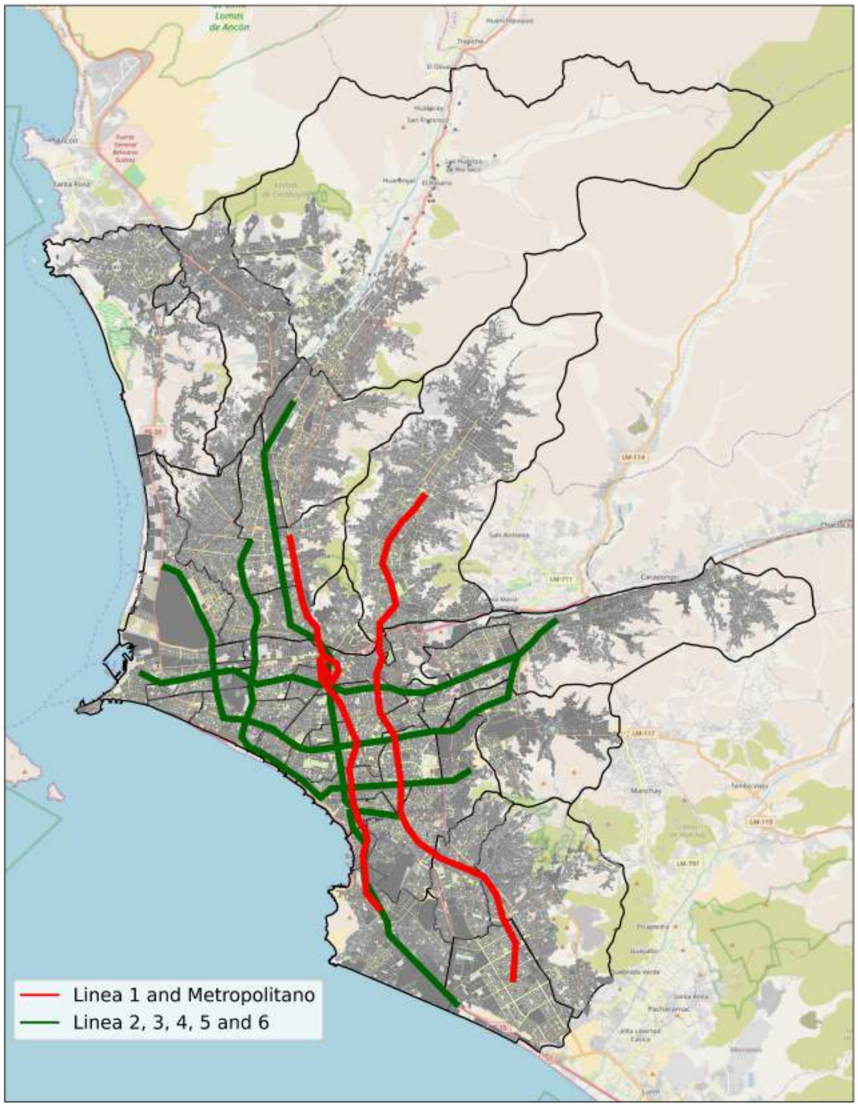
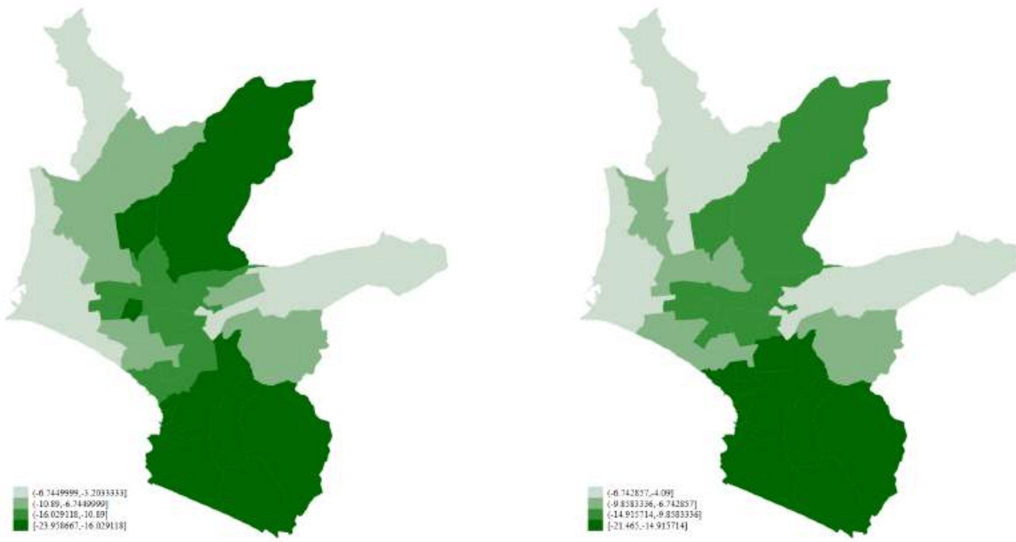


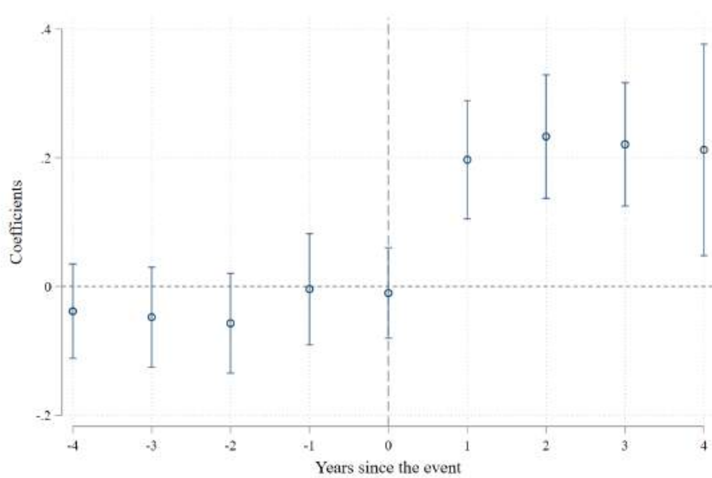
Figure A.3: College Market Access (before/after M&M)



(a) All colleges

(b) Private colleges

Figure A.4: Dynamic Effects of the $M\&M$ on Distance (kms.) to College



B 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 B.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 B.1: kNN Scores

	Access	Private			Public
	to College	Low Returns	High Returns	Elite	
Best mean cross-validation score	0.995	0.896	0.887	0.936	0.672
Best k parameters	19	19	19	19	19
Test-set score	0.995	0.895	0.888	0.936	0.673

C Brief History of Public Transportation in Lima

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

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

The construction of the project began in 1986, but by 1990 only 1.5 km was built. The economic and political crisis stopped the project from continuing. In the 1990s, AATE was part of the Council of Ministers and proposed a complementary study of the Lima Metro Network in 1998 (MTC, 2005). Then, at the beginning of 2000's, the Municipality of Metropolitan Lima developed the COSAC study, a preliminary study of a BRT (a Metropolitano antecedent). As a consequence, the AATE changed the route of line 1 (tramo II) from Av. Grau to San Juan de Lurigancho instead of Comas. [18] 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 construction 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 [19] 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.

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