Opportunity Bound: Transport and Access to College in a Megacity

Fabiola Alba Vivar*

April, 2023

Click here for the most recent version

Abstract

A common problem urban areas face is the distance between key education services, such as universities which are often located in high-rent central areas, and the location of many of the urban poor, who live on the periphery. This paper examines the effects of new transportation infrastructure on college access. By leveraging detailed administrative records of college enrollment, as well as information on the timing, location, and routes of two newly established systems, I employ a difference-in-differences strategy to exploit variations in exposure to new stations and neighborhood locations in Lima, a city of 12 million people. The findings reveal that improved transportation increases college enrollment by 1%, with private colleges driving the effect. Furthermore, those who gain access to college are 5% more likely to graduate, and these benefits are most pronounced for vulnerable populations such as women and low-SES students. Over the long term, students who enroll in college are more likely to secure white-collar employment suggesting that they also connect to better labor market opportunities.

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 Eric Verhoogen, Matteo Magnaricotte, Carolina Concha-Arriagada, Daniel Velazquez, Michael Best, Jack Willis, Kevin Donovan, Sam Asher, Daniel Bjorkegren, Nishith Prakash, Marianne Bitler, 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

Transportation infrastructure is thought of as an important component for gaining productivity in the economy and supporting economic growth but much less has been studied regarding the potential effects it might have in education (Bryan et al., 2020). In this sense, a common problem urban areas face is the distance between key education services, such as universities, which are often located in high-rent central areas, and the location of many of the urban poor, who live on the periphery. This problem can exacerbate economic and social inequality and can have long-lasting consequences. In this paper, I explore the causal effects of providing efficient, clean, and safe transportation infrastructure on access to higher education in a large city in South America. This work studies the case of two new mass public transportation systems the Metropolitano, a rapid bus transit, and the first line of the Metro de Lima, for now on the $M\&M$. These systems provided Lima, a city of approximately 12 million people, with an upgrade in transportation quality. More importantly, the $M\&M$ serve *commuter college students* who, given the location of college campuses across the city, travel on average 2 hours daily to attend classes.[1]

Using detailed administrative college enrollment records and data on the location and openings of new stations, I causally estimate the impact of the $M\&M$ with a difference-in-differences (DiD) strategy that exploits variation by cohort and neighborhood location. I further refine the DiD strategy by comparing neighborhoods that were exposed to new $M\&M$ stations versus neighborhoods that were exposed to planned but non-opened stations, following the placebo strategy implemented by Donaldson (2018). I collected information on Lima's strategic plans for the city for the last decades as well as several transportation studies

[1]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.

that were used to build the Metro de Lima plan, which included eight lines connecting several areas of the city to Lima downtown. Only six of them were properly studied and examined, and eventually, only one of the lines was built.[2] I use the neighborhoods surrounding the 5 planned but not executed lines as a control group. However, even using the placebo strategy, the urban features of this design might also suffer from bias. Neighborhoods in Lima Downtown have systemically higher market access given that these areas were meant to be connected anyway, regardless of which line gets executed. I exclude such neighborhoods from the main sample to address these concerns.

The results suggest positive effects on access to higher education: there is a significant increase in college enrollment that is persistent up to 4 years after the opening of the first $M\&M$ station. The increase is on average 1 percent per year compared with the baseline rates. 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]. What is more, private universities have a higher likelihood of increasing spots and rapidly responding to the increasing demand. The effects also differ by gender even when both female and male enrollment increase. Women are more likely to enroll in college than men, especially after 2 years since the $M\&M$ opening. These results are consistent with previous literature that documents how increasing women's mobility has positive effects on their human capital (Borker, 2020; Fiala et al., 2022; Muralidharan and Prakash, 2017).

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.

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

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

This is an important outcome since most Peruvian college students tend to complete their education later than expected (> 5 years). The effects are particularly higher for vulnerable populations. Women, low socioeconomic status, and minorities are the ones who take most of the benefits.

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 preliminary finding suggests that there is a positive effect on employment rates but the effects are higher 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).

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

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 cities in the world for women, comparable to New Delhi and Kampala.[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.

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); **?**. 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

[5] As documented by Sviatschi and Trako (2021), Peru is a country that has experienced a huge increment 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.

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 have a positive impact on 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 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.

[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.

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. 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

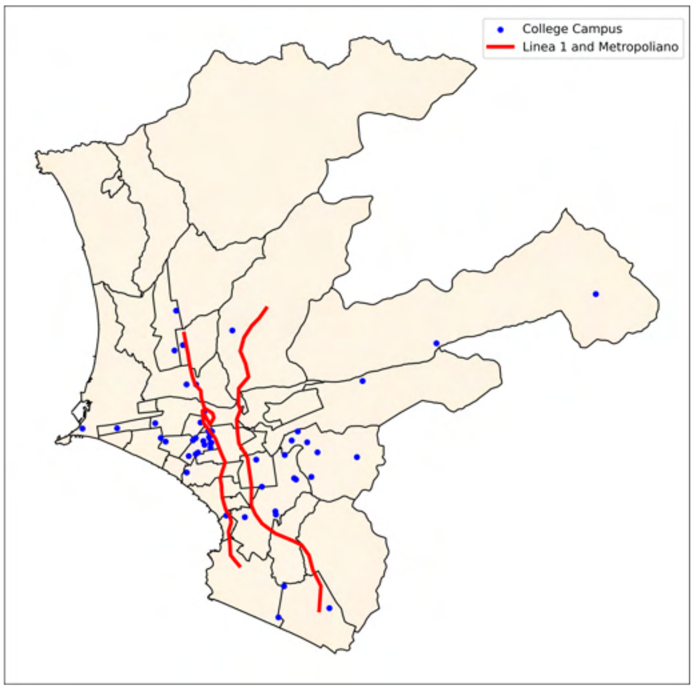
Lima's population is comparable to other large cities around the world such as New York City, Paris, Xi'an, Chennai, Jakarta, Bogota or Los Angeles Metropolitan Area. However, Lima is not nearly as dense (4,000 hab/km^2), and commuting across the city can take up to 3 hours during rush hour. During the 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.

[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.

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.

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



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 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 (2023) procedure and data to impute velocities for major highways and

[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).

[9] The information is public and the package *osmnx* is available on Python

the $M\&M$, and I also complement it with Google Maps API data to obtain primary and secondary highway speeds. With this information, I computed commute times before and after the $M\&M$.

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:

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

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.[10] In particular, I follow Borusyak et al. (2021) and implement their imputation estimator which allows for treatment-effect heterogeneity and dynamic effects.

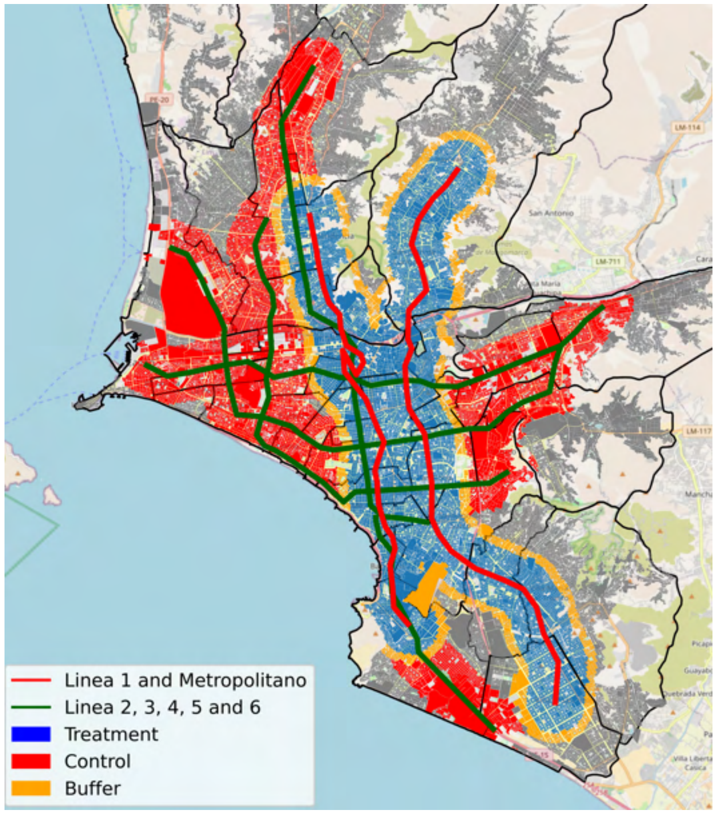
A second empirical challenge is to establish a proper control group. In this sense, simply comparing connected neighborhoods to non-connected neighborhoods within the city might overestimate our results since the allocation of the new routes is not completely random. One way to address this concern is using a placebo group as in Donaldson (2018). 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 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 company that built *Linea 1*.

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, recent work by Borusyak and Hull (2022) shows that we could still have issues. Most areas in the center of the city will tend to have higher levels of market access than those in the

[10] 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 $M\&M$ and planned but not-executed lines



outskirts since they typically get connected by multiple ways of transportation. I exclude such neighborhoods in Lima Downtown from the main sample to address these concerns.

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 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.[11]

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

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 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.[12] 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 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.

[11] 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.

[12] 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 Statistic at the Neighborhood Level before the *M&M*

	(1) Total	(2) M&M	(3) No M&M
Panel A :	College Admin. Records Data		
Enroll. (counts)	0.367	0.339	0.388
	(3.487)	(4.154)	(2.916)
Enroll. (rates)	0.172	0.165	0.177
	(0.363)	(0.357)	(0.367)
Enroll. (log)	0.0910	0.0819	0.0976
	(0.331)	(0.310)	(0.345)
Female Enroll. (logs)	0.0543	0.0490	0.0581
	(0.251)	(0.236)	(0.262)
Male Enroll. (logs)	0.0446	0.0392	0.0485
	(0.225)	(0.209)	(0.236)
Public Enroll. (logs)	0.0281	0.0242	0.0309
	(0.179)	(0.166)	(0.188)
Private Enroll. (logs)	0.0690	0.0624	0.0737
	(0.285)	(0.268)	(0.297)
Distance HH to College	1.621	1.641	1.607
	(4.267)	(4.280)	(4.257)
Panel B:	2007 Census		
Total Population	128.2	127.8	128.6
	(123.8)	(130.9)	(118.3)
Primary School	0.197	0.206	0.191
Pop. Share	(0.143)	(0.140)	(0.144)
Secondary School	0.394	0.409	0.384
Pop. Share	(0.175)	(0.175)	(0.175)
Higher Ed.	0.409	0.386	0.426
Pop. Share	(0.240)	(0.235)	(0.243)
Observations	211,824	88,436	123,388

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 total population counts by block. The population shares by education levels consider people above 25 years old. Higher education level includes college and community college.

4 Results

4.1 College Access

Using the enrollment administrative data, I find positive effects on college enrollment at the block level (1% increase) compared to the baseline as seen in Table 2. The preferred estimation, as shown in Table 2, is in Column 3, which excludes Lima Downtown. 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.

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

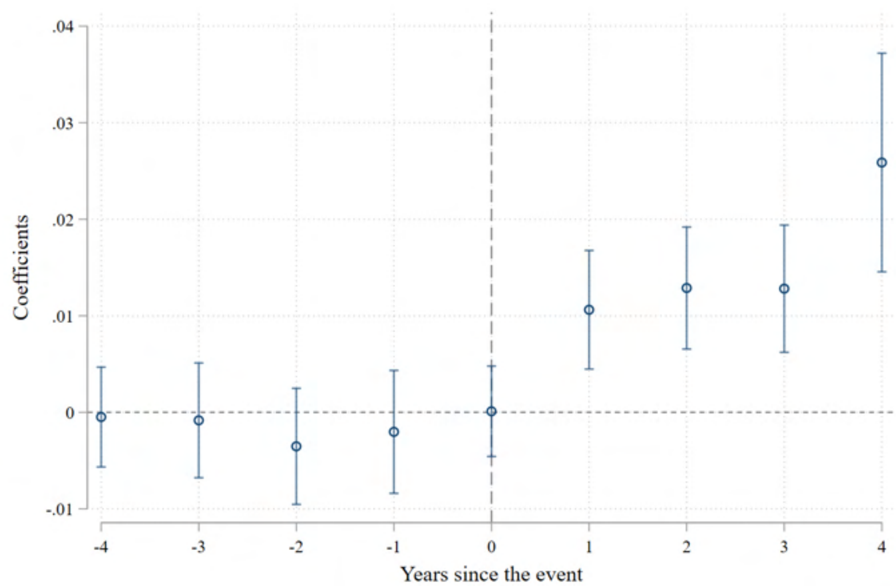
	All	Log(All)	Excl. DT	Private	Public	Female	Men
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat*Open	0.009***	0.011***	0.012***	0.010***	0.003***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean Control	0.144	0.093	0.080	0.059	0.026	0.047	0.039
N	461151	461151	411147	411147	411147	411147	411147
Block FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

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 has doubled to a 2.5% increase after three years of the $M\&M$ openings. 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 shown in Figure 4b.

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 4a 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.

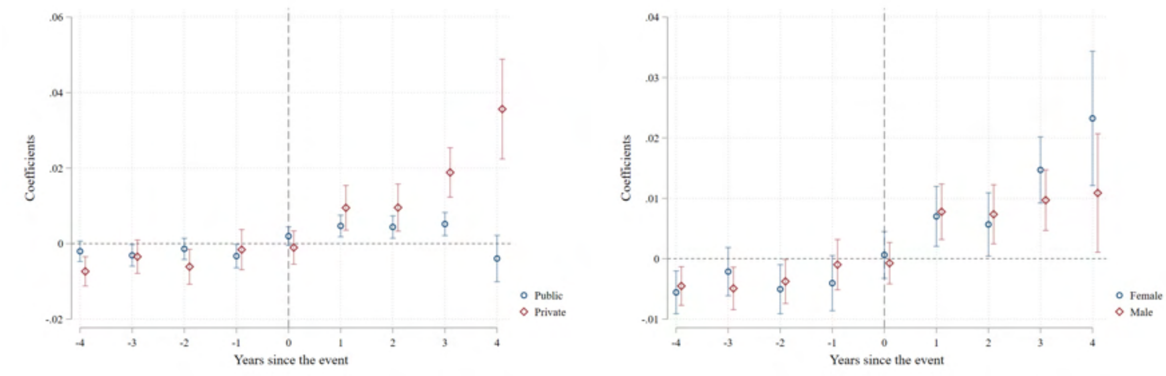
Figure 3: Dynamic Effects of the $M\&M$ on College Enrollment Rates



4.2 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

Figure 4: Dynamic Effects of the $M\&M$ on College Enrollment Rates by Groups



(a) By College Type

(b) By Gender

results show a positive impact on college completion rates (5%) compared to baseline rates, as shown in Table 3. The estimated coefficients are similar to whether Lima Downtown is included or not.

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

	All	Excl. DT
	(1)	(2)
Treat*Open	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.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 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 $M\&M$.

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

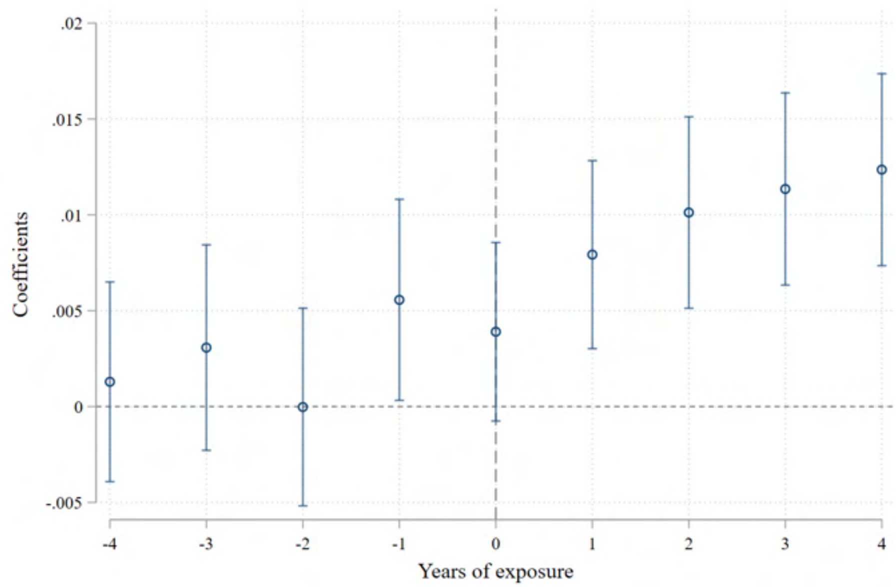
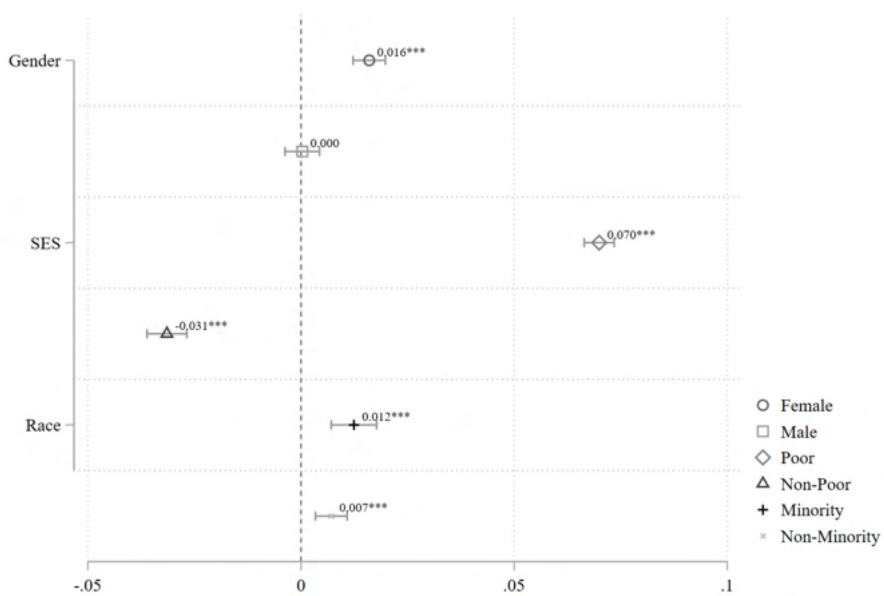


Figure 6: Heterogeneous Treatment Effects of College Completion



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. [13]. 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 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 4: 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*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

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

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

5.2 Distance to College

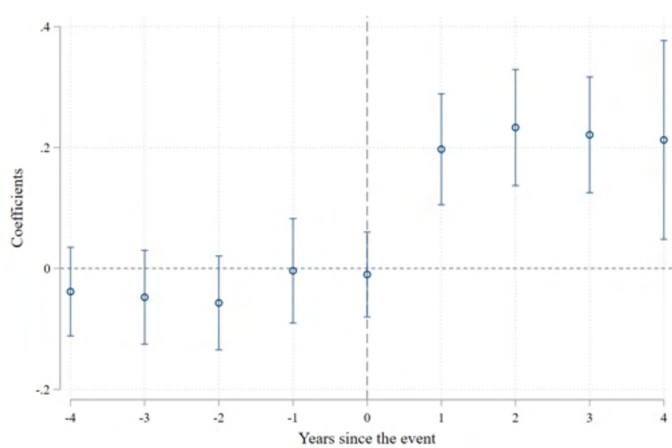
An additional channel I explore is whether the students who enroll are now opting for colleges that are located farther away. The rationale behind this is that since transportation becomes less of an issue, students have greater freedom to select colleges that are more distant from their homes. Table 5 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 9% farther to attend college as compared to the baseline.

Table 5: Effect of the $M\&M$ on Distance (kms.) to College

	(1) Distance	(2) Distance Excl. DT
Treat*Open	0.130***	0.148***
	(0.0262)	(0.0286)
Mean Control	1.654	1.586
N	461151	411147
Block FE	Yes	Yes
Year FE	Yes	Yes
Controls	Yes	Yes

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

Figure 7: Dynamic Effects of the $M\&M$ on Distance (kms.) to College



6 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, similar to many other metropolises around the world. But 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 from basic education. Furthermore, access to college is often studied through the lens of monetary constraints, but less is known about reducing transportation costs for students. This is particularly important in the context of big cities where inequality is more striking. When it comes to access to college, while most of the literature on education has focused on how monetary restrictions and other institutional factors limit access to higher education, little is known about transportation costs. 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 that is potentially upper-middle income class and females. Additionally, there is some evidence that the increment in enrollment was driven by private universities, which are typically less competitive than public ones. What is more, accessing college education has long-term benefits such as higher wages and this could change the geography of inequality. The results of this paper show how less advantaged students (low social-economic status (SES) and women) benefit from a large reduction in transportation costs and better access to college education and labor market opportunities.

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.

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.

Bryan, G., Glaeser, E., and Tsivanidis, N. (2020). Cities in the Developing World. *Annual Review of Economics*, 12(1):273–297.

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*.

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

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.

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

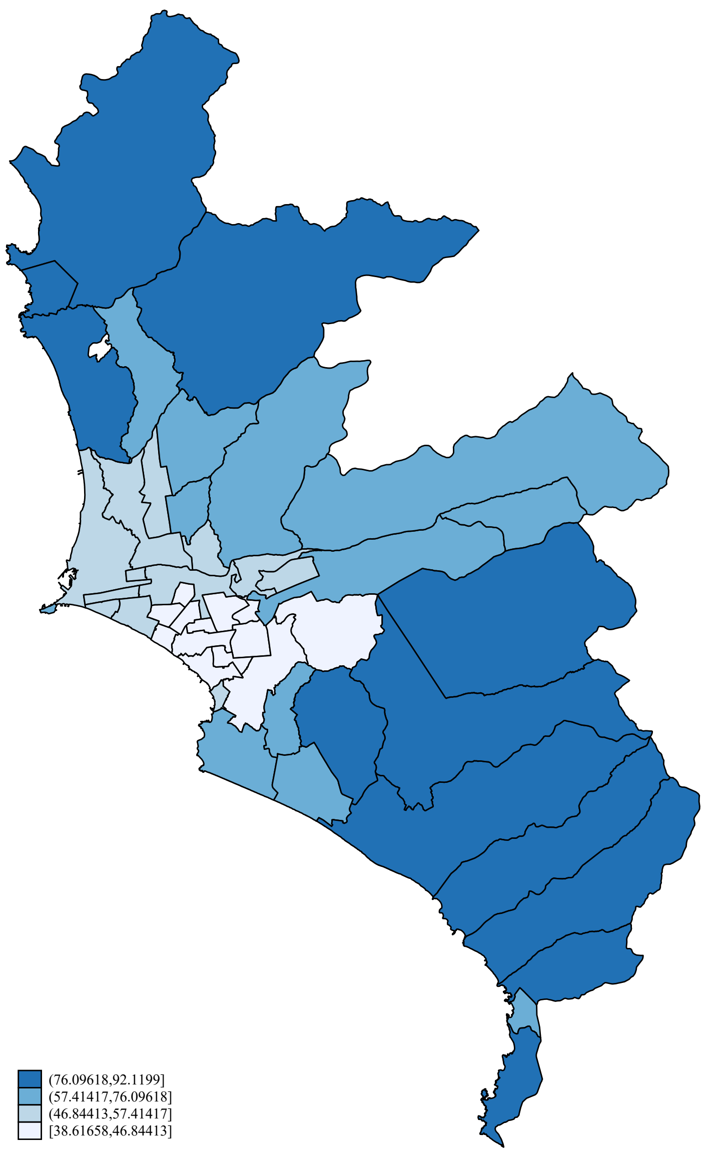
Sviatschi, M. and Trako, I. (2021). Gender violence, enforcement, and human capital: Evidence from women's justice centers in peru. Technical report.

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.

A Appendix

A.1 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: Potential but unexecuted Metro lines in Lima

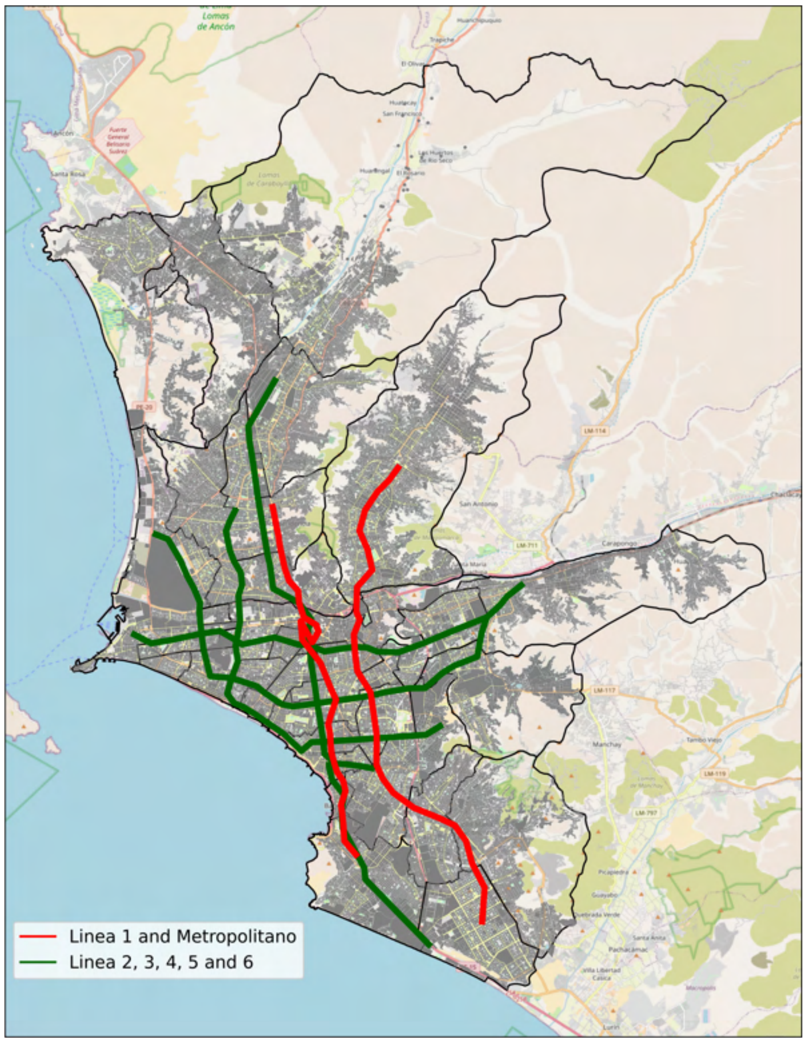
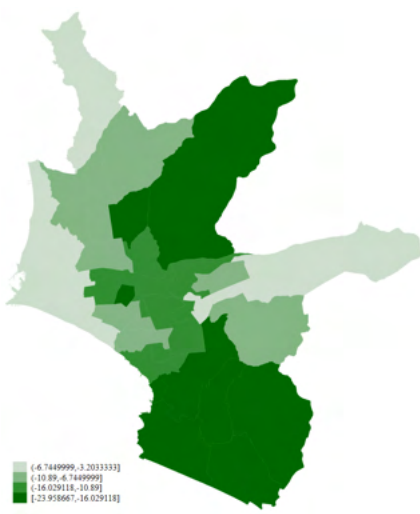
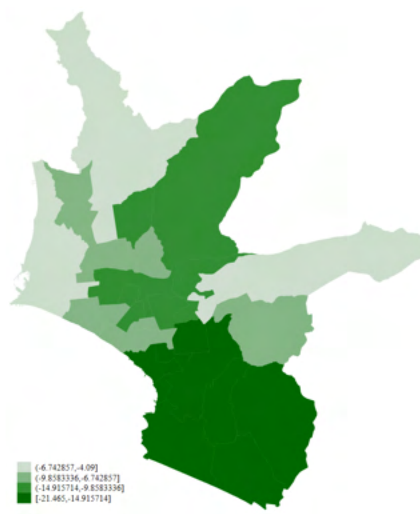


Figure A.3: College Market Access (before/after M&M)



(a) All colleges



(b) Private colleges