On my way to college: Transport and access to higher education in a megacity

Fabiola Alba Vivar*

December, 2021

Click here for the most recent version

Abstract

In this paper, I study the effects of public transport in access to higher education. I focus on the opening of the first mass transportation system in a one of the largest cities in Latin America. Using detailed administrative data of college enrollment, I causally estimate the impact of the *Metropolitano* by implementing a difference-in-difference strategy that exploits variation of cohorts and universities who were exposed to new bus stations. I find a significant increase on college enrollment that is persistent up to 6 years after the opening. When looking at the heterogeneous effects, I find a reduction on public schools enrollment suggesting that the effects was driven by private colleges.

This is a preliminary draft. I have benefited from data provided by the Peruvian Ministry of Education and the Instituto Metropolitano PROTRANSPORTE de Lima.

* Teachers College at Columbia University (New York, New York), *fma2139@tc.columbia.edu*

1 Introduction

Transportation infrastructure is thought of an important component for gaining productivity in the economy and support economic growth. Although, much less has been studied regarding the potential effects it might have in education (Bryan et al., 2020). I study the effects of the opening of the very first public transportation system in a megacity and its effects on access to higher education. As Psacharopoulos and Patrinos (2018) document, individual education is correlated with upward mobility in cities as schools provide a setting that enables urban interactions. According to Glaeser and Resseger (2009), city density and education appear to be complements, suggesting that better education may enable poorer children to take advantage of urban opportunities. In this paper, I focus on tertiary education and document how less advantaged students benefit from reducing transportation costs.

I look at the case of a major transportation policy in Lima, a city of approximately 12 million people. In July 2010, after 4 years of construction and multiple delays, the *Metropolitano*, the first public transportation system in the city, was opened to service. This system connected the north side and the south side of the city (12 districts out of the 44 in the city) and reduced the transportation time from 2.5 hours to 1 hour on average, providing cleaner, faster and safer service. The *Metropolitano* included multiple bus stops near several colleges across the city. I study whether the introduction of this new public system changes college enrollment. I combine administrative data from college records and geocoded *Metropolitano* stations to document how college students were affected by this new policy.

I estimate a difference-in-differences model, exploiting exposure to the *Metropolitano* and cohort variation. My preliminary analysis of the effects shows positive effects on enrollment. This is consistent with the findings of Adukia et al. (2020), who study the effects of India's flagship road construction program and find that school enrollment increases and this is driven by heterogeneous effects: effects are larger when nearby labor markets offer high returns to education. In this paper's setting, heterogeneity might arise from different types of colleges.

I find that there is a significant decrease on public college enrollment. This suggests that the main effects have come from students choosing private universities over public ones. However, I do not find results when looking at heterogeneity by quality. To measure quality, I use the license status given by a reform in 2016, as documented by Alba-Vivar et al. (2021), that forced low quality universities to cease operations. However, in future work, I will study how selectivity might also affect the enrollment results.

This paper is linked to various branches of the economics literature. First, it relates to papers studying the effects of transportation costs on education attainment. I expect that the opening of the *Metropolitano* affected university enrollment as it directly decreased transportation costs in the city. In India, Muralidharan and Prakash (2017) study the impact of providing bicycles to female student. The authors use a triple difference-in-difference approach and find that being in a cohort that was exposed to the Cycle program increased girls' age-appropriate enrollment in secondary school by 32 percent and reduced the corresponding gender gap by 40 percent. I complement their work by showing how a reduction in transportation costs can increase higher education as well. Borker (2020) measures the extent to which perceived risk of street harassment can help explain women's college choices in Delhi. She finds that women are willing to choose a college in the bottom half of the quality distribution over a college in the top quintile to travel by a route that is perceived to be one standard deviation safer. This will also document gender-specific preferences. Our current event study shows an increment on the share of females enrolled in colleges exposed to the *Metropolitano*, suggesting that they were potentially benefited the most by this new system.

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 underserved areas, does indeed increased enrollment but the effects for minority students are only half the size of others, widening preexisting gaps. They also document that proximity is highly valued by less-advantaged students, and that meritocratic admission criteria hinder poor and minority students, who disproportionately attend lower-quality high schools. They also focus on all regions except Lima, the subject of this paper. I will complement their work by studying how an alternative policy that simply reduces transportation costs can also boost college enrollment.

This paper contributes to the current literature in education and urban economics. There little evidence on the impacts of major transportation reforms on higher education. I will be able to provide causal estimates of the opening of a city's major public transportation system and their effects on college choice as well consequent changes in college socio-economic composition. I plan to use a market access approach like Donaldson and Hornbeck (2016) and Tsivanidis (2019), which will allow me to get a general equilibrium effect in the higher education market as well as estimate welfare effects. Lastly, this paper relates to the literature studying of social dynamics within the school setting. By changing the social composition of cohorts exposed to the *Metropolitano*, one could expect that increased diversity changes subsequent labor market outcomes after graduation. These results will complement the work of Carrell, Hoekstra and West (2019) and Rao (2019) by showing how diversity generated because of a change in transportation costs can have positive effects on students.

The next section discusses the relevant institutional context. Section 3 describes the data used in my analysis. Section 4 presents the empirical strategy, while Section 5 presents preliminary results. Section 6 concludes with a discussion of my next steps.

2 Background

2.1 Higher Education in Peru

Peru, as many other developing countries, had a rapid expansion and increasing demand for higher education over the last two decades. This process was accompanied by a sustained economic growth, often referred as the *Peruvian miracle*. [1] The Peruvian Education system

[1]For more information, see: Peru: Economic Miracle or Just a Mirage?

is based on 3 levels: primary education (6 years), secondary education (5 years) and the higher education system which often lasts from 2 (technical school) to 10 years (School of Medicine). On average, university students take between 6 and 7 years to graduate. According to the 2017 Peruvian Census, approximately 4 out of 10 recent high school graduates (between 17 and 21) have access to some type of higher education. More specifically, 15% have access technical school, 22% access university, and the remaining 63% does not access any type of higher education. Following similar trends to other comparable countries, women access higher education in slightly greater proportion than men. Similarly, those who are Spanish native speakers have a greater access to higher education in comparison with other ethnic minorities (Quechua and Aymara native speakers). Non surprisingly, young people residing in urban areas as well as those whose parents has some type of higher education, have greater access to higher education compared to their peers.

Peruvian students face similar struggles to other college students in the developing world. Among those who access higher education, a significant amount of students work and study at the same time (28% of technical school students and 20% of university students). When we look at their household characteristics, a large proportion of students have one or both parents with a higher education degree (42% of students in universities and 25% of students in technical schools). Regarding the characteristics of their institutions, 35% of higher education students are studying in an institution located in Metropolitan Lima, [2] and a little more than 60% of students attend an higher education institution located in their province of birth and the number raises up to 90% when it comes to colleges. This sustains the fact that out-of state enrollment is quite uncommon.

In this paper, I focus on Lima, the capital of Peru, a city of approximately 12 million people and one of the biggest cities in Latin America. Lima concentrates around half of the college enrollment, making it a important market to study. Lima is quite diverse and inequal-

[2]Notably, in our study sample, the proportion raises up to 60 percent given that we are not capturing technical schools that are more popular across the country and whose enrollment increased by 2017.

ity withing the city is quite striking. Some aspects regarding this educational market are worth highlighting. First, there is no centralized admission system and students take admission exams at multiple colleges and re-apply multiple times. Public universities are virtually free, they only charge a small administrative fee, but they are highly competitive. On the other hand, private universities have a greater variance in price and quality. The market for loans is almost non existent and scholarships are quite limited. Most students live at home and commute to college. Dorms, if they exist, are reserved exclusively for out-of state students. Figure 1 shows the average travel time from home to university campus in minutes using the 2010 University Census data. Students living in the outskirts of the city travel on average 1.5 hours, whereas those living in Downtown Lima travel 40 minutes on average.

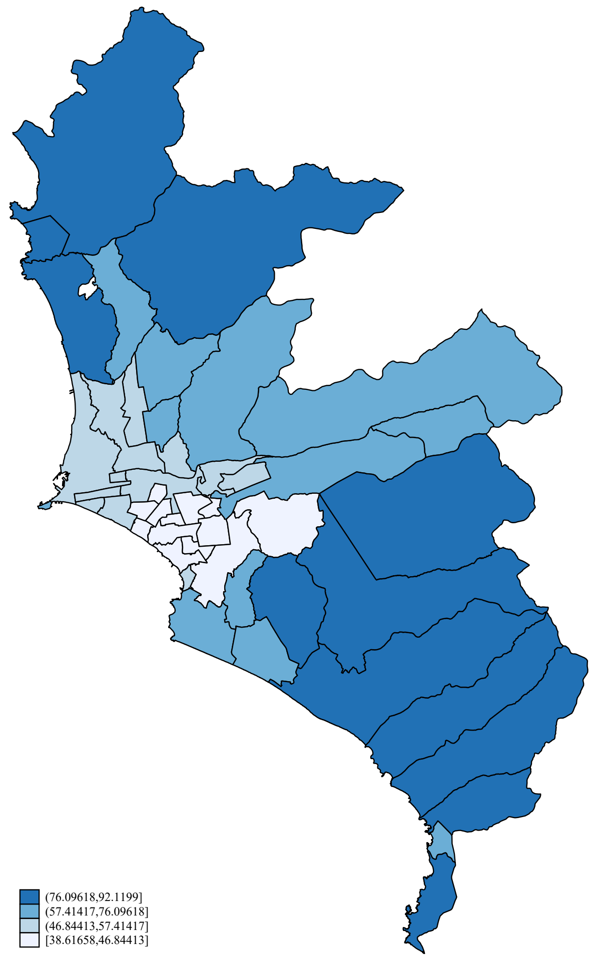
2.2 Transportation in Lima

Lima is considered a **megacity** with a population just as big as Tokyo or New York City. However, Lima is not as dense and most people travel long distances to cross the city. During the 90s, market liberation policies and a huge wave of migrants arriving to the city [3] spurred informal transportation which became one of the main problems in the city. By 2015, traffic accidents became the main cause of death for adults and adolescents. In July 2010, after 4 years of construction and multiple delays, the *Metropolitano* was opened to service. The *Metropolitano* was the very first mass transportation public system in the city. This system connected the north side 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 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 president inaugurated *Line No.1* of the Lima Metro which connected the north-east side of the city with the south-east side. Lima's Metro

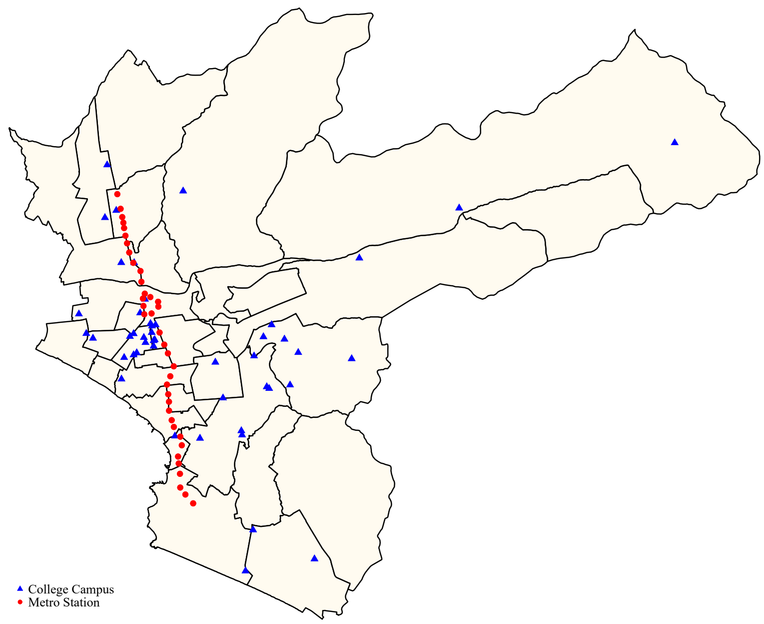
[3] The migration wave from rural and poor areas of Peru were consequence of the economic depression at the end of the 80s, mainly caused by the civil conflict and uncontrolled inflation. Peru's inflation by the end of the 80's was 7,481 percent.

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



project started in the 70's and it's construction began in 1986 but it was never finished. 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. The *Line No.1* and the *Metropolitano* reduced the transportation time from 2.5 hours to 1 hour on average, providing a cleaner, faster and safer service. Most importantly, the routes included multiple stops near several colleges across the city. In this paper, I focus on the *Metropolitano*. In Figure 2, I show the distribution of colleges and the *Metropolitano* stations across the city.

Figure 2: University Campuses and *Metropolitano* Stations across Lima



3 Data

In this draft I use multiple data sets, mainly administrative records compiled by the Peruvian Ministry of Education. First, I use multiple cross-sectional data spanning 2006-2016 of college enrollment. This data, at the student level, includes year of enrollment, college ID, student's address, major, age, sex. Second, I use the GPS location of each *Metropolitano* stations, provided by ProTransporte (the institution that runs the *Metropolitano*).

In future versions of this paper, I plan to integrate other datasets. More importantly, I will add a measure of income at the individual level. This is will be potentially linked with enrollment in Peru's CCT that classifies households as non poor, poor and extreme poor. Using, a more ambitious approach, given that I have students' addresses, I can use satellite images of their houses and neighborhoods a get a continuous measure of income (when proxy with granular census data at the neighborhood data).

Other sources of data that I could integrate:

- Graduation Records: looking at those who were already enrolled, I can estimate if being enrolled on a university exposed to the *Metropolitano* increases the graduation rates in expected time.

- Labor Market Outcomes: looking at cohorts about to graduate, I can see if they are more likely to get jobs or internships. This could be done using either administrative data and survey data.

- Metropolitano Usage: to document who are using the Metropolitano, there is data based on the student discounts given by Protransporte.

4 Empirical Strategy

My empirical strategy is based on a difference-in-differences (DiD) design comparing universities that had a differential exposure to *Metropolitano* both before and after its opening. As detailed above, I proxy the cross-sectional exposure to the *Metropolitano* by calculating the distance to the closest *Metropolitano* Station to the main campus. My main identifying assumption is that, absent the *Metropolitano*, the enrollment in universities with a higher exposure to the *Metropolitano* would have evolved in tandem with universities with a exposure to the *Metropolitano*. Second, since I do not explicitly observe students who are explicitly using the *Metropolitano* in a given university, I also assume that my proxy variable—distance to the closest *Metropolitano* station—is highly correlated with this ideal exposure measure and that any deviations from the latter are completely arbitrary. Given both of these assumptions, I can identify the effect of the opening of the *Metropolitano* in college enrollment.[4] Explicitly, I estimate the following equation:

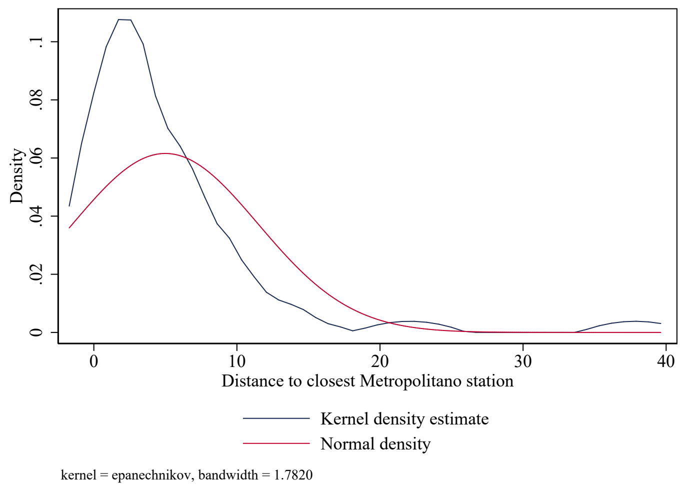
$$y_{mt} = \delta Exposure_m \times \mathbf{1}[\text{MetOpening}]_t + X'_{mt}\beta + \gamma_t + \gamma_m + \epsilon_{mt} \tag{1}$$

In this regression, δ measures the impact of the *Metropolitano* opening in university (m) and year (t)—represented by y_{mt}. Where $Exposure_m$ is at the university level and $\mathbf{1}[\text{MetOpening}]_t$ is a dummy variable that takes value 1 for all years after the opening of the *Metropolitano*. Figure 3 shows the density of the exposure measure. X_{mt} is a set of controls

[4] Given the classical nature of the measurement error in our exposure variable, I believe that I am identifying a lower bound for the effect of this policy on enrollment.

that vary at the university- and year-level. The year- and university-level fixed effects are represented by γ_t and γ_m, respectively. Lastly, all our regressions have heteroskedasticity-robust standard errors clustered at the treatment level in accordance with the difference-in-differences literature. [5]

Figure 3: Exposure to the *Metropolitano* as Distance to Closest Station



Additionally, I estimate a generalized DiD with variation in treatment timing ("event study") following indications in Borusyak and Jaravel (2018). The model takes the form:

$$y_{mt} = \sum_{j=T_1}^{T_2} \alpha_j \cdot I(ry_t = j) + \beta X_{mt} + \gamma_u + \delta_m + \epsilon_i \tag{2}$$

Where y_i is the dependent variable (enrollment, share of female students, age of enrollment) for university m in year t. T_1 (T_2) is the maximum number of years before (after) the *Metropolitano* opening and ry_t is the year relative to the date when the *Metropolitano* opened. γ_i is an university level fixed effect and δ_m is a year fixed effect.

[5]For simplicity, in this draft I am using a discrete version of the Exposure (continuous) measure. In particular, I take as treated universities those who have at least one *Metropolitano* station in less than 2.5 kilometers.

5 Preliminary Results

5.1 Main Results

In this draft, I am using a discrete version of the Exposure (continuous) measure. In particular, I take as as treated universities those who have at least one *Metropolitano* station in less than 2.5 kilometers and zero otherwise. Estimating Equation 1, I find that the opening of the *Metropolitano* had a positive sizable effects on college enrollment.[6] As seen in Table 1, the *Metropolitano* opening increased enrollment by 56% over the period studied. I do not find significant effects for the share of female enrolled and the age at which people enroll. This means that a university exposed to the Metropolitano withing 2.5 kilometers, experienced a significant increase in the number of students enrolled relative to the pre-ruling average for the who were not exposed. In this table and all other specifications, standard errors are clustered at the treatment level.

Notice that this coefficient is very high, which suggests that I should be more conservative in the future, I will discuss potential ways to fix this in Section 6. The surprisingly highly magnitude in the main enrollment effects was pretty similar to other similar specifications like using a continuous exposure (distance to closest metro station).

Table 1: Main Effects of the *Metropolitano* Opening

	Log(Enrollment)	Female Share	Age
	(1)	(2)	(3)
Treatment*Opening	0.459**	0.00512	0.208
	(0.182)	(0.0149)	(0.348)
Dep. Var. Mean	7.195	0.484	25.35
N	380	380	380
Year FE	Yes	Yes	Yes
Uni FE	Yes	Yes	Yes

Standard errors in parentheses

Errors clustered at the university-year level.

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

[6] In future work, I will show that these results are qualitatively similar if the continuous measure of exposure and other specifications.

5.2 Event Study

To make sure the parallel trends assumption hold, I estimated Equation 2 for each one of the main outcomes. I find no evidence of pre-trends for these outcomes. In Figure 4, I confirm positive effects found on Table 1 and they seem to be constant over time period I cover. Interestingly, as seen in Figure 5 and 6when looking at the share of female enrolled in college, it seems to go up and when looking at the average age of enrollment, it seems to go down. To put into context, unlike in the US or Europe, students do not automatically enroll in college when graduating. In most cases, they take time to either prepare for the admissions exams or work. What is more, several universities have an older set of students (above 30 years old). This results suggest that now students are enrolling at a younger age after graduation or that older students are enrolling less. It is not clear if universities opened more seats and increased their enrollment but we do know that no new universities were opened up since there was a moratorium in 2012 that prohibited the creation of new universities.

Figure 4: Results on log(Enrollment)

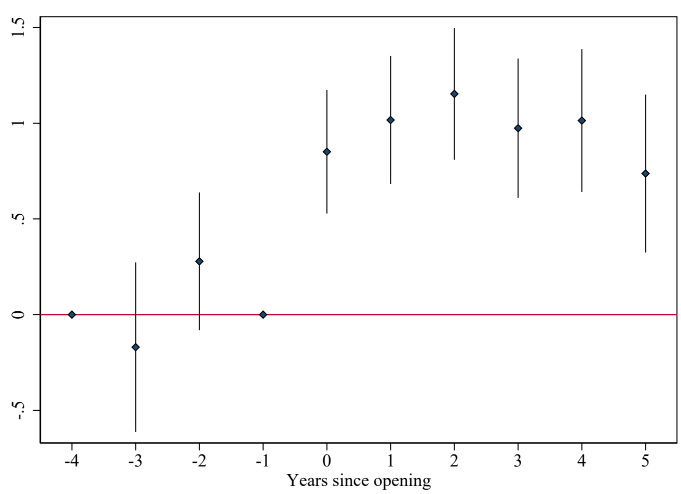


Figure 5: Results on the share of female enrollment

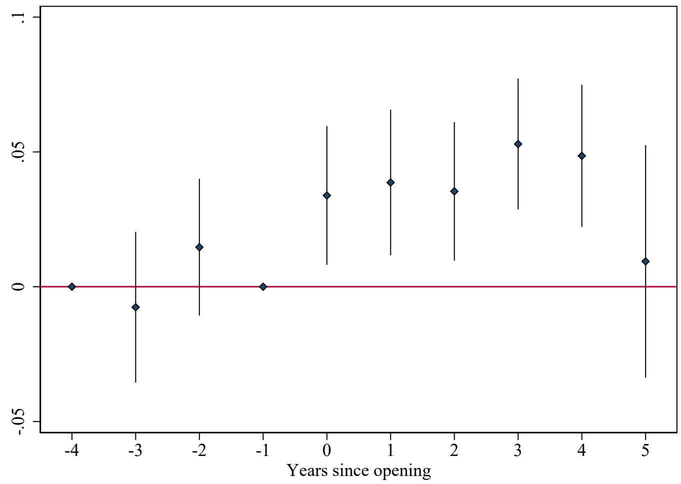
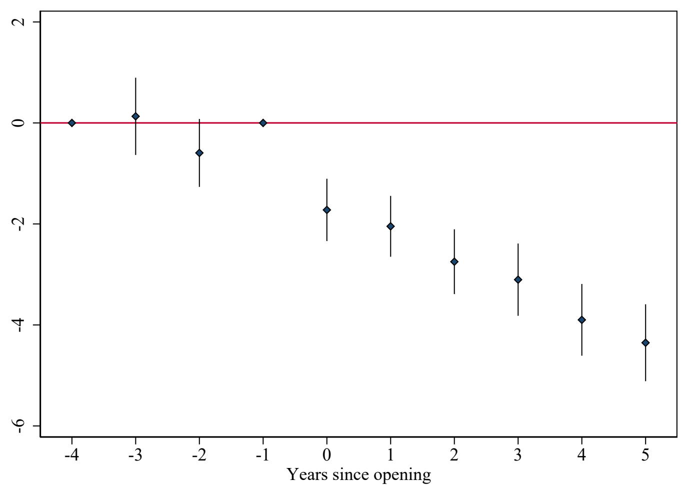


Figure 6: Results on the average age of enrollment



6 Mechanisms and Road-map to Robustness

In this section, I want to emphasize possible treats to the causal evidence I presented on the previous section as well as some mechanisms to be explored.

6.1 Mechanisms

One natural question one could ask if *Metropolitano* changed students school choice and if certain types of colleges capture that increase in demand. As seen on Table 2, there is a significant decrease on public college enrollment. This suggest that the increment in enrollment was driven by private universities it is consistent with the fact that public colleges are quite selective and often have a very strict enrollment cap whereas private universities have room to increase enrollment more easily. Nevertheless, in Table 3, I look at the effects on university quality. I find that there were no significant effects on high and low quality colleges. This was measure using a dummy variable for whether the university received a license to operate after the 2016 reform as documented by Alba-Vivar et al. (2021). Universities that did not received the license had to cease operations withing 2 years.

Table 2: Enrollment Effects of the *Metropolitano* Opening by University Type

	Log(Enrollment)	Female Share	Age
	(1)	(2)	(3)
Treatment*Opening	0.688***	0.00885	0.133
	(0.190)	(0.0163)	(0.386)
Treat*Public	-1.661***	-0.0270*	0.546
	(0.236)	(0.0150)	(0.374)
Dep. Var. Mean	7.195	0.484	25.35
N	380	380	380
Year FE	Yes	Yes	Yes
Uni FE	Yes	Yes	Yes

Standard errors in parentheses

Errors clustered at the university-year level.

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

6.2 Robustness Checks

6.2.1 New Literature on Differences and Differences

The frailties of two-way fixed effects estimators (TWFE) come from the inclusion of already-treated or partially treated groups within the comparison group: if treatment effects are heterogeneous, the TWFE estimator will be biased.[7] Several recent papers have introduced solutions to address these problems, by making sure that only never-treated or not-yet-treated

[7] Dynamic effects, e.g. when treatment effect is growing as time from the event passes, also lead to misspecification and inconsistency when a constant effect is assumed. This can be addressed by estimating treatment effects relative to event time (Borusyak et al. (2021)).

Table 3: Enrollment Effects of the *Metropolitano* Opening by University Quality

	Log(Enrollment)	Female Share	Age
	(1)	(2)	(3)
Treatment*Opening	0.503**	0.0162	0.709
	(0.254)	(0.0233)	(0.669)
Treat*Licensed	-0.0736	-0.0181	-0.820
	(0.303)	(0.0248)	(0.686)
Dep. Var. Mean	7.195	0.484	25.35
N	380	380	380
Year FE	Yes	Yes	Yes
Uni FE	Yes	Yes	Yes

Standard errors in parentheses

Errors clustered at the university-year level.

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

units are included in the comparison group. Similar to the basic difference in differences, these papers rely on parallel trends assumptions to build consistent estimators that do not suffer from the same problems.[8]

In the setting I am studying, it is possible to think that the effect of exposure to the Metropolino shock on outcome $y_{t,m}$ (e.g. enrollement) is biased. The term $\beta_{\tau,m}$ highlights the possibility of effects being heterogeneous for different colleges (m) and depending on exposure length ($t-j$).[9] Allowing for heterogeneity and dynamics requires us to avoid standard "two-way fixed effects" (TWFE) regressions, as they have been shown to be problematic in such setups (Goodman-Bacon (2021), Baker et al. (2021)).

I propose using two new estimators: Callaway and Sant'Anna (2020) provide estimates using only never-treated units.[10] This estimator is robust to dynamic effects (e.g. increasing with length of exposure to treatment) and effects being heterogeneous across colleges. I can also test using the de Chaisemartin and D'Haultfœuille (2020) estimator that can be use when

[8]See as examples of such new estimators Callaway and Sant'Anna (2020), Borusyak et al. (2021), and Sun and Abraham (2020).

[9]In our setting, heterogeneous treatment effects are likely to arise from heterogeneity of the *Metropolitano* users or colleges quality.

[10]By the end of the period studied, all colleges had received or were denied a license.

I have a continuous measure of exposure. Finally, I can also use a more ambitious estimator and have a synthetic DiD approach as in Arkhangelsky et al. (2021), which will help me to get a reasonable control group and a proper benchmark for the effects.

6.2.2 Additional Specifications and Randomize Inference

Besides the standard procedures for new difference-in-differences estimation discussed above. I also plan to show that my results are robust to a wide range of additional considerations. The three main problems that could arise when using these data. First, the effects could come from the oversized influence of outliers. In this sense, I can identify two potential types of outliers 1) college with an excessive enrollment and 2) college that are too far from the *Metropolitano* station to be relevant or that were affected by the Linea 1 of the Metro (opened in mid-2011). In this sense I can remove these universities and get a better control group, especially when using the new estimator as in Arkhangelsky et al. (2021). Second, we could have whether the measurement error or misreporting in the data. In this case, I could also use survey data reported in ENAHO that is representative at the district level and not at the university level. Third, I can be misreporting access to the *Metropolitano* since they have also smaller buses that extend the current route. In future work, I will map these additional routes.

Finally, in terms of inference, I will show that my results remain statistically significant when using a randomization test—as detailed in Bertrand et al. (2004)—instead of the usual asymptotically normal approximation to the distribution of the estimand. Following MacKinnon and Webb (2020) and Bertrand et al. (2004), I can test the statistical significance of my results results using a randomization test in place of the standard asymptotically normal approximation. This approach serves two main purposes: 1) it allows me to demonstrate that my results are representative of robust relationships in the data and 2) it provides an additional check on the statistical relevance of my results. This randomization inference exercise will estimate a modified version of Equation 1 where we resample the exposure to and timing of the *Metropolitano* from the empirical distribution of my data. One could repeat this process 300 times and use all these coefficients to generate an estimate of the finite sample distribution of the δ parameter in equation 1. The main drawback of this procedure is that it requires that $Exposure_m \times \mathbf{1}[\text{MetOpening}]_t \perp\!\!\!\perp \epsilon_{mt} \mid X_{mt},\ \gamma_t,\ \gamma_m$. While this is a stronger condition than the parallel trends assumption that I need to identify δ in Equation 1, I believe that it is palatable given the auxiliary nature of this exercise. The following equation could be implemented with this randomization exercise.

$$y_{mt} = \tilde{\delta}\ \overline{Exposure_m} \times \overline{\mathbf{1}[\text{MetOpening}]_t} + X'_{mt}\beta + \gamma_t + \gamma_m + \epsilon_{mt} \tag{3}$$

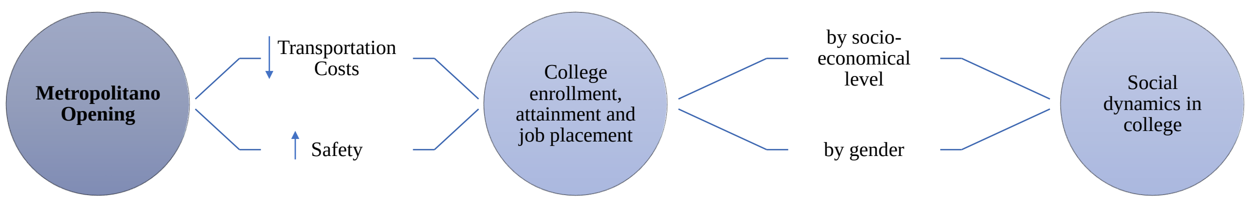
7 Discussion and Further Steps

There is little causal evidence of how transportation affects students college choice. This paper brings novel evidence and studies the case of the first transportation system in a megacity. The preliminary results suggests an increase in enrollment that is pottencially benefiting young recent graduates and females. What is more, there is some evidence that the increment in enrollment driven by private universities but there is no evidence that this affected low quality universities, measured by a policy that effectively closed low quality universities starting in 2016.

The results presented on this draft will be completed in the upcoming months as I gather and include more data that has already been obtained and that I could also potentially get as detailed in Section 3. I also plan to improve the selection of the control group and look at treatment effect heterogeneity depending college selectivity. I will also implement a DiD approach using individual level data once I obtained student's home address. Finally, I will complement the analysis with a formal structural model. This is based on the work of Donaldson and Hornbeck (2016) that introduced a "Market Access" approach and that included a new reduced-form expression derived from general equilibrium trade theory model. I will also take into account the work of Tsivanidis (2019), who studies the effects of improving transit infrastructure on city structure and welfare. He uses the commuter market access (CMAs) as a sufficient statistic, that delivers a log-linear reduced form. he find that high-skilled workers benefited slightly more, which is surprising given the reliance of the low-skilled on public transit.

In the future, I aim to build on these previous paper and estimate a estructural model, following a conceptual framework as seen in Figure 7.

Figure 7: Conceptual Framework



References

Anjali Adukia, Sam Asher, and Paul Novosad. Educational investment responses to economic opportunity: Evidence from indian road construction. *American Economic Journal: Applied Economics*, 12(1):348–76, January 2020. doi: 10.1257/app.20180036. URL https://www.aeaweb.org/articles?id=10.1257/app.20180036.

Fabiola Alba-Vivar, Jose Flor-Toro, and Matteo Magnaricotte. College licensing and reputation effects on the labor market. *Work in progress*, December 2021. URL https://falbav.github.io/research/research-licensing/.

Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118, December 2021. doi: 10.1257/aer.20190159. URL https://www.aeaweb.org/articles?id=10.1257/aer.20190159.

Andrew Baker, David F. Larcker, and Charles C. Y. Wang. How Much Should We Trust Staggered Difference-In-Differences Estimates? *SSRN Electronic Journal*, 3 2021. doi: 10.2139/SSRN.3794018. URL https://papers.ssrn.com/abstract=3794018.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275, 2004.

Girija Borker. Safety First: Perceived Risk of Street Harassment and Educational Choices of Women. 2020.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting Event Study Designs: Robust and Efficienct Estimation. *Work in Progress*, pages 1–48, 2021.

Gharad Bryan, Edward Glaeser, and Nick Tsivanidis. Cities in the developing world. *Annual Review of Economics*, 12(1):273–297, 2020.

Brantly Callaway and Pedro H.C. Sant'Anna. Difference-in-Differences with Multiple Time periods. *Journal of Econometrics*, 12 2020. ISSN 03044076. doi: 10.1016/j.jeconom.2020.12.001.

Clément de Chaisemartin and Xavier D'Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96, September 2020. doi: 10.1257/aer.20181169. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20181169`.

Dave Donaldson and Richard Hornbeck. Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858, 2016.

José Flor-Toro and Matteo Magnaricotte. College Expansion and Unequal Access to Education in Peru. *Job Market Paper*, 2021.

Edward L. Glaeser and Matthew G. Resseger. The Complementarity between Cities and Skills. NBER Working Papers 15103, National Bureau of Economic Research, Inc, June 2009. URL `https://ideas.repec.org/p/nbr/nberwo/15103.html`.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 6 2021. ISSN 0304-4076. doi: 10.1016/J.JECONOM.2021.03.014.

James G MacKinnon and Matthew D Webb. Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 2020.

Karthik Muralidharan and Nishith Prakash. Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics*, 9(3):321–50, July 2017.

George Psacharopoulos and Harry Antony Patrinos. Returns to investment in education: A decennial review of the global literature. *Policy Research working paper*, (WPS 8402), April 2018.

Gautam Rao. Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review*, 109(3):774–809, March 2019. doi: 10.1257/aer.20180044. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20180044`.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 12 2020. ISSN 0304-4076. doi: 10.1016/J.JECONOM.2020.09.006.

Nick Tsivanidis. Evaluating the impact of urban transit infrastructure: evidence from Bogotá's TransMilenio. 2019.