Social Mobility and Higher Education in Brazil*

Javier Feinmann   Roberto Hsu Rocha

UC Berkeley   UC Berkeley

May 29, 2025

[Preliminary and Incomplete]

Abstract

We follow high school graduates through college and the labor market to study income segregation and intergenerational mobility across colleges in Brazil, a unique context where admissions are mostly determined by exam scores, public institutions are free and of high quality. We show that public college admissions are income neutral once controlling for grades, but elite public colleges are composed mostly of higher-income students, as they have higher exam scores. Intergenerational mobility rates in elite public colleges are low, but higher than in comparable private institutions. We integrate different margins of response in a framework to evaluate how policies aimed at reducing income segregation across colleges impact the future earnings of different groups. We use this general framework to evaluate affirmative action in public colleges and subsidized loans for private institutions. Both policies increased the mobility of low-income students, but subsidized loans have a larger effect. While AA increases the representation of disadvantaged students in elite schools and subsidized loans do not, the latter policy reallocates a larger number of students to better college tiers overall. All the results in the paper are based on data we make publicly available for the first time. It was constructed based on confidential records and aggregated so other colleagues could use it for their research.

*We would like to thank Sebastian Otero, Emmanuel Saez, Gabriel Zucman, Ricardo Perez-Truglia, and Sydnee Caldwell for insightful comments. We also thank Marco Cesar Araujo Pereira for his permanent support regarding the use of data in the secure room. We gratefully acknowledge financial support from the Stone Center at UC Berkeley and the Institute for Research on Labor and Employment. All data at the college level is available upon request. Email jfeinmann@berkeley.edu

1 Introduction

College is a formative experience for individuals' lives. Higher education shapes, among other things, the future earnings perspectives of individuals. However, if good colleges can only be attended by a selective part of the population, it can end up being a mechanism for diminishing intergenerational mobility. Furthermore, segregation across colleges can be intensified if admission policies favor richer students. For example, in the US, Chetty et al. (2020) find that allocating students in an income-neutral way could substantially increase intergenerational mobility.

In this paper, we study income segregation across colleges and its consequences for intergenerational mobility in Brazil. We leverage individual-level longitudinal data covering students graduating high school with information on standardized test scores and economic background, college attendance, and the universe of the formal labor market. This allows us to paint a full picture of how college attendance and future labor market outcomes interact with students' economic backgrounds. We then explore the consequences of two large policies from the Brazilian federal government, affirmative action in public colleges and subsidized loans to low-income students in private colleges. We outline a generalized conceptual framework and apply empirical estimates to calculate the effect of these policies on intergenerational mobility patterns.

Brazil provides a unique context given its institutional setting. In Brazil, public universities are free to enroll and are, on average, of higher quality than private ones. Furthermore, admissions are determined by entrance exam scores and do not include subjective analysis of students' characteristics or extra-curricular activities, which usually favors students from more advantaged backgrounds. At the same time, Brazil is one of the most unequal countries in the world, and recent estimates show it has one of the lowest rates of intergenerational mobility (GC Britto et al., 2022).

We begin by showing that entrance exam scores and college attendance are strongly correlated with students' economic background. Our sample comprises all high school graduates between 2010 and 2012 who took the national standardized exam (ENEM) that serves both as an entrance exam for colleges and to evaluate high schools across the country. To measure students' economic backgrounds, we developed a granular metric based on the high school they attended by linking zip-code level income per capita from the 2010 census to students' schools[1]. We find that 85% of the students in the top 10% of the economic background distribution enroll in college at some point in their life, compared to only 43% for students in the bottom 10%. Once controlling for ENEM scores, the relationship remains strong, especially between the extremes (77% vs. 54%).

We find important differences when comparing college attendance between private and public institutions. When looking at private institutions, controlling by grades does not change the relationship between economic background and college attendance. On the other hand, while top 10% students are almost three times more likely to attend a public university than students in the bottom 10%, this association is entirely explained by differences in grades. This indicates that attendance in public colleges is income neutral, as individuals from different economic backgrounds, but with the same exam scores, have the same probability of attending public colleges.

Next, we examine how segregated Brazilian higher education institutions are. When comparing the composition of public and private colleges, we see that both have similar shares of students from the bottom 40% and bottom 20% of the economic background distribution (28% and 10% respectively), but public colleges have a higher share of students from the top quintile (39%) than private ones (28%). Following a data-driven approach to group colleges into quality tiers, we find that elite colleges are highly segregated in terms of income. Only 2.5% of their students come from the bottom 20% of the economic background distribution and 9% from the bottom 40%. At the same time, 60% of students are from the top quintile and almost 30% come from the top 5%. These levels of segregation are equivalent between elite public and

[1]We provide different evidence on the validity of our school-level measure by showing its' correlation with student-level measures of household income and parents' education.

private colleges. This changes when looking at colleges lower in our rank. Non-elite public colleges are less segregated than comparable private ones.

Our findings show that even in a setting with income-neutral admissions, colleges remain highly segregated, especially at elite colleges, which have a large overrepresentation of students in elite colleges even when they have income-neutral admissions and attendance policies.

We then turn our focus to intergenerational mobility by examining how students' wages when they are adults correlate with their economic background and college attendance. We find a strong correlation between economic background and future wages. Students from the top 10% of the economic background distribution rank, on average, in the percentile 62 relative to their cohort (with the 0.1% ranking 80), while the bottom 10% ranks, on average, in the percentile 36. When we control for college fixed effects, the correlation decreases substantially (54 vs. 45). This demonstrates that most of the difference between students from low and high economic backgrounds arises between colleges and not within them. After controlling by college fixed effects interacted with ENEM scores, the correlation decreased slightly more, suggesting that there still is some remaining segregation within colleges.

We follow Chetty et al. (2020) and create mobility rate measures across colleges, that consist of the product of the share of students from disadvantaged backgrounds in each college with their future wage rank. We find that mobility rates decrease monotonically as college rank increases since the share of students from disadvantaged backgrounds decreases at a higher rate than the increase of future outcomes. Elite colleges have the worst mobility rate, but public elite colleges still have a higher mobility rate than comparable private ones.

In our last section, we outline a simple conceptual framework to guide us in understanding the effects of policies that shift college attendance on intergenerational mobility. With small assumptions, the framework illustrates that we can divide the effects of policies into composition effects and college effects. Composition effects are cross-elasticities of how students' higher education decisions change in response to the policy. College effects are the value added in income corresponding to changing from college A to college B. We show that average estimates of these measures are sufficient to understand the aggregate impacts of these policies on intergenerational mobility.

We then apply our framework to understand two different policies: affirmative action in public institutions and subsidized loans in private ones. Affirmative action policies have existed since 2004 in Brazil but were vastly expanded in 2012 with a federal law establishing that 50% of all seats in public colleges run by the federal government should be reserved for students from public high schools or underrepresented minorities. On the other hand, federal loans to higher education have long existed in the country. However, in 2010, the government expanded the loans program by increasing its budget and reducing by half interest rates on loans.

We specify a difference in differences model to understand how both policies changed the composition of student bodies across colleges. We find that both policies increase the share of students from the most disadvantaged backgrounds, but an additional seat with subsidized loans is more likely to be filled by disadvantaged students than an additional seat reserved for affirmative action. Effects are distinct between both policies across colleges. Affirmative action has the highest effect in elite colleges, whereas subsidized loans do not change the composition of the student body in elite private colleges.

Our findings suggest that both policies increased intergenerational mobility, but the effects are small. In our preferred specification, subsidized loans increase the income of disadvantaged groups four times more than affirmative action. Despite being more effective in increasing disadvantaged students in elite colleges, affirmative action has smaller aggregate effects than subsidized loans because they targeted a much smaller number of students.

This paper contributes directly to the literature on the role of higher education in shaping intergenerational mobility. Due to data limitations, very few papers provide a comprehensive picture of income segregation across colleges, linking students' backgrounds to future earnings and standardized test scores for an entire country. Chetty et al. (2020) is the big exception, making this characterization for the United States. To the best of our knowledge, our paper is the first one extending this type of evidence to a developing country.

A much larger body of research explains the role of college on future earnings, with a focus on who benefits the most from it(Neelsen, 1975; Iannelli and Paterson, 2005; Dale and Krueger, 2002, 2014; Kearney and Levine, 2014; Zimmerman, 2019). There has also been a growing body of work on how specific policies that work through colleges affect social mobility(Bucarey et al., 2020; Bleemer, 2022, 2021; Black et al., 2023). Mountjoy (2022) shows that the increase of two-year colleges in the United States has an ex-ante ambiguous effect on social mobility because it does not only move people from *no college* to a *two-year college*, but it can also move students from a *four-year college* to a *two-year college*. This highlights the importance of thinking through counterfactual options to understand the aggregate impacts of policies that target college attendance. Londoño-Vélez et al. (2023) shows how financial aid in Colombia improves college enrollment, quality, and attainment, particularly in STEM-related fields. The earnings gains are substantial, growing, and driven partly by high-quality universities improving students' skills. We expand this literature by shedding light on what is behind the unequal access to elite colleges, even in a context where it is publicly provided and tuition-free. Identifying the limitations in admission to higher education is crucial to guide efficient policy interventions.

Our paper also relates to a growing literature studying the Brazilian higher education system. Duryea et al. (2019) investigates the topic of income segregation and intergenerational mobility in one prominent public college, comparing it to U.S. measures. Dobbin et al. (2021) studies subsidized loans in Brazil and their general equilibrium effects on enrollment and prices. Several papers have studied the effects of affirmative action and the Brazilian centralized admission system on income segregation (Senkevics and Mello, 2019; Mello, 2023; Machado and Szerman, 2021; Estevan et al., 2019). Results are consistent and show that affirmative action policies significantly increase the share of disadvantaged students in public universities. On the other hand, the centralized admission system increased competition in application, leading to a crowd-out of high-achieving students over low-achieving students. Otero et al. (2021) expanded this previous work, providing a theoretical framework to account for the general equilibrium effects of affirmative action policies. These general equilibrium effects take displaced students into consideration in order to provide guidance on welfare analysis. In our paper, we leverage both affirmative action and subsidized loans policies to understand their effects on social mobility. We provide a unifying conceptual framework that takes elements from the previous work. It extends Mello (2023) by separating its estimates by college tier to recover compositional changes in enrollment of disadvantaged students. It also incorporates the notion of displaced students developed by Otero et al. (2021), to consider students who cannot enroll because someone else took their seat.

Lastly, our paper contributes to the general debate about intergenerational mobility. In developed countries, researchers have linked parents and children's income, from adminstrative records to measure intergenerational mobility and its determinants(Chetty et al., 2014, 2016; Kenedi and Sirugue, 2023) In developing countries, this task is much harder given the lack of administrative data that gives family links, thus much of the research relies on survey data(Asher et al., 2018; Alesina et al., 2021; Mahlmeister et al., 2019; Bautista et al., 2023). An exception is GC Britto et al. (2022) who estimate aggregate patterns of mobility and the role of geography on children's future outcomes in Brazil. They find a higher correlation than us between economic background and future outcomes, but our aggregate measures are not comparable to theirs, as we focus on a sample of high school graduates whereas their analysis is done with the whole population.

2 Institutional Background

Brazil is the largest country in Latin America with over 200 million inhabitants. Despite recent reductions in income inequality, it remains one of the most unequal countries in the world where the top 10% represent 58.6% of the country's income and the top1% retain 26.6% of the country's income. It's GINI index of 52.9 in 2021 is one of the highest in the world and much higher than other Latin American countries such as Argentina (42.0), Chile (44.9), or Mexico (45.4). Inequality also manifests in the education system. Although rates of illiteracy reduced, and high school completion rates increased, education attainment in both years of education and quality of education remains highly dispersed around the population.

The educational system is divided into three major stages. First, Elementary education (Ensino Fundamental) comprises grades 1-9, for children aged 6-14. It is followed by High School (Ensino Médio) with three grades for individuals between 15 and 17. Higher education (Ensino Superior) encompasses universities and other institutions that offer undergraduate programs. There are three types of higher education degrees, *Bacharelado* which is the equivalent of a bachelor's degree, *Licenciatura* which are degrees that allow you to teach a given topic in high schools and elementary schools, and *Técnico* which are degrees that are equivalent to vocational higher education degrees.

All public education institutions, including public universities, have no financial costs. Up to the high school level, most of the education provision is made by public institutions. Public schools are responsible for 87.4% of high school enrollment. This changes in higher education, where private institutions are responsible for over 70% of enrollment. Private institutions can be for-profit or non-profit, and can charge for education provision at all levels. Public provision is made by the three levels of the Brazilian administrative structure. Municipalities are usually responsible for elementary education, States for High School, ands public universities are divided between the Federal and State government.

Enrollment in High School: Each state has discretion on how to organize admissions to public High Schools. Usually, students graduating from elementary levels, are forwarded to the high school closest to where they live. Sometimes they can claim to change high schools but are left under the discretion of principals or personnel of the state administration. Despite some initial initiatives, centralized admission systems where families rank their preferences are not common in Brazil.

Private schools have discretion over who to accept, as long as it does not violate discrimination forbidden by law. The fees charged by private schools end up being the biggest mechanism of selection. In 2023, the average fee for a private high school was around 1000 Brazilian Reais per month, which is almost the value of the minimum wage.

Enrollment in Higher Education: In Brazil, students apply directly to a major in a university. Selection systems may vary across private institutions, but not for public institutions. All public institutions select students based on large exams called vestibulares. Federal universities use a national exam called ENEM (detailed below), whereas state universities can opt between ENEM and their own entrance exam. There are no subjective criteria in admissions for public universities, such as valuing extracurricular activities or letters of recommendation.

Despite being allowed to use other selection criteria, most private universities also follow admission strategies based on exams, as subjective analysis of students is not common. It is common for private universities to use the ENEM score as their criteria, but many institutions also organize their own entrance exam. We show that private universities are on average less selective than public universities.

2.1 Affirmative Action

Due to this discrepancy in access to public free education, social movements in the beginning of the 2000s started to push for policies that increased the opportunities for students from public high schools and non-white individuals. Starting in 2004, with the State University of Rio de Janeiro, 113 higher education institutions adopted some kind of affirmative action policy between 2004 and 2012. These policies included quotas for non-white individuals or students from public high schools, or bonus in the entrance exam for the targeted groups.

In April 2012, the national congress approved a law that mandated that all federal institutions that provided higher education adopt affirmative action policies reserving 50% of the spots for non-white students or students from public high schools. Colleges had some room for discretion in how to allocate these spots between race quotas or public high school education quotas, but all Universities had to maintain a level of racial quotas. These institutions had 4 years to adequate to the policy. For state institutions, affirmative action policies are still at the discretion of the state administration.

2.2 Financial Aid Policies

In addition to the provision of public colleges, the federal government plays a significant role in extending financial assistance to students enrolled in private institutions. Two primary programs drive higher education financing in Brazil: PROUNI (Programa Universidade Para Todos) and FIES (Fundo de Financiamento ao Estudante do Ensino Superior). Notably, while PROUNI offers scholarships to students, FIES provides subsidized loans.

PROUNI, established in 2005, grants partial or full scholarships to students for private university attendance, with participating institutions receiving corresponding tax incentives based on the allocated seats. Eligibility hinges on factors such as household income, public high school attendance, and performance in the ENEM exams.

Conceived in 1999 as a replacement for previous financial aid programs, FIES initially provided loans with interest rates aligned with Brazil's benchmark rate (SELIC) in the subsequent decade. In 2010, the government augmented the program's budget and substantially reduced the interest rate to less than half of the Central Bank's benchmark rate. This adjustment led to a significant surge in the number of students benefiting from FIES loans. To qualify for FIES loans, students must have a household income below 3 minimum wages. In 2015, eligibility criteria underwent a transformation, limiting loans exclusively to students scoring above 450 points across all ENEM exams.

3 Data

We use three main data sources in our analysis. ENEM and CESUP are our sources for individual-level education data, and RAIS is our source for labor market outcomes.[2]

ENEM (Exame Nacional do Ensino Médio): Enem is a national standardized exam created in 1998 that serves both for college admissions and for evaluating the quality of high school education. Since 2009, the exam is comprised of four different parts with 45 multiple choice questions that cover math, reading comprehension, social sciences, and natural sciences, and one essay[3]. It is conducted yearly in November and takes place over the course of two consecutive Sundays. Scores are standardized using Item Response Theory making results comparable across years. Anyone can register to take the exam, for a fee of around 80 BRL (18USD), but public school students are registered for free.

We access individual-level data on ENEM. Besides exam scores for all 5 different parts, the data also contains answers to a detailed survey including questions on socioeconomic background and student perceptions. Furthermore, for individuals who are graduating from high school when taking the exam, it also provides the identifier of the school at which the individual is concluding their studies.

[2] All analysis that requires combinations of these data was done in a secured room at the Instituto Nacional de Estudos e Pesquisas educacionais Anísio Teixeira (INEP). The data contains masked social security numbers that allow us to follow individuals in different data sets, but no individual could be identified during the analysis.

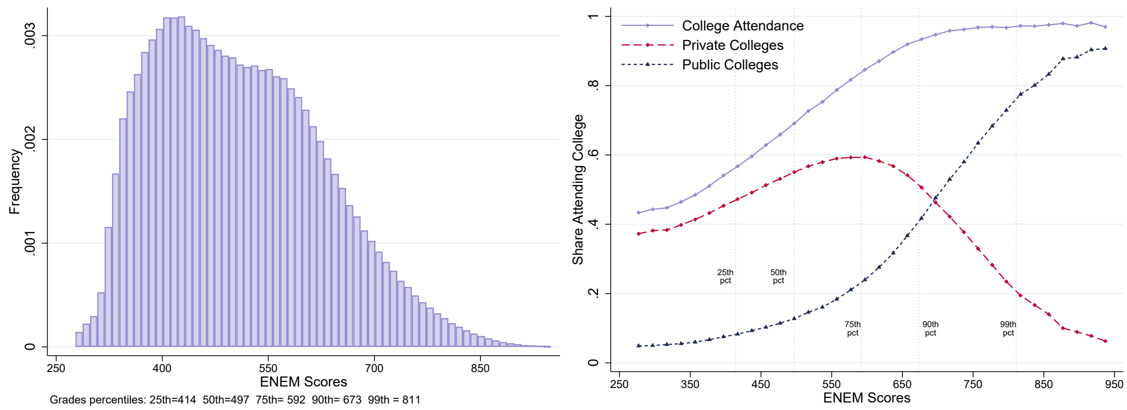
[3] Prior to 2009, ENEM was comprised of 63 questions covering all areas and scores were not comparable across years

CESUP (Censo da Educação Superior) : The higher education census contains information on every student enrolled at any higher education institution in Brazil, allowing us to observe the educational path of every student in any degree program between 2009 and 2021. In a given year, the observation unit is at the degree-student level and includes a variable indicating if the individual graduated, dropped out, or is successfully enrolled at the end of the academic year. These data are of very high quality, as most institutions have their systems integrated with the census in real time. In addition to the student data, this database also includes administrative information at the level of degrees and institutions, allowing us to characterize both over time.

We show the distribution of ENEM math scores[4] and how it correlates with college attendance, by private and public institutions in Figure 1. In Figure 1a, we see that although scores can theoretically vary between 0 and 1000, virtually the whole distribution concentrates between 250 and 900 points. Figure 1b shows how math grade correlates with college attendance, by private and public institutions. We observe that the probability of attending any college is increasing in ENEM grades and over 95% for students in the top 10% of ENEM scores. Attendance in private and public colleges varies substantially according to ENEM grades. Students below the median grade are highly unlikely to go to public colleges. On the other hand, almost 90% of students in the top 1% of ENEM go to public colleges.

[4] We choose to use math scores because of two main reasons. First, students take ENEM exams on two different weekends. Math is on the first weekend, reducing the selection of who attends the second weekend. Second, averaging grades shields a smoother *ability* distribution, while math provides a wider range of values. In Figure A1, we show that math scores are highly correlated with other exam scores. If we run a linear regression of math scores on the average of the other five tests, we find a coefficient of 0.91.

Figure 1: ENEM scores Distribution and Attendance at Public and Private Colleges



(a) Histogram of ENEM Scores **(b)** College Attendance by ENEM Score

Panel (a) shows the histogram of ENEM grades among high school graduates between 2010 and 2012. Panel (b) shows college attendance of these high school graduates according to ENEM scores in math. College attendance is defined as 1 if the individual appeared at least once in CESUP in the 7 years after they graduated High School and 0 otherwise. If an individual appears both in private and public universities, we select the first observation after they graduate high school. Vertical lines in Panel (b) show percentiles of the ENEM score distribution.

The last main data set used in the analysis is **RAIS (Relação Anual de Informações sociais)**. RAIS is a matched employer-employee data set organized by the Brazilian Ministry of Labor and contains information on formal labor market outcomes from both the private and public sectors. Firms have high incentives to report the correct information as it determines a series of government policies for both employers and employees, besides the possibility of being fined for misreporting. It provides information on wages, duration of job relation, number of hours worked, age, gender, race and education of each employee, and firm-level characteristics such as industry and location.

3.1 Main Analysis Sample and Economic Background Measure

Main Analysis Sample: Our main sample restricts the population of ENEM takers between 2010 to 2012 to those who just graduated from high school. This guarantees certain homogeneity in age, college records, and work experience. We follow them over time to observe their college and labor market activities. In particular, we match our main sample to CESUP until 2020, and to RAIS 9 years after taking ENEM. As students usually graduate when they are around 18 or 19 years old, This implies that we can observe labor market variables when they are between 27-28 years old.

Economic Background Measure: Linking students to parents, and parents to their labor market variables in the past to directly measure social mobility is not feasible in Brazil. While ENEM data contains a questionnaire on household income, responses are grouped in large bins not comparable across years. Some progress can be made with this data, but a deeper analysis requires higher granularity on students' distribution. Therefore, we develop a measure of students' economic conditions in high school based on the school from which they graduated. Every individual in our main sample contains the high school identifier from where they graduated. We match these identifiers to the Censo da Educação Básica[5], finding information on over 99% of graduating high schools. We then match the zip code where each high school is located to a zip code level income per capita from the 2010 population census. In short, we proxy students' household income with a very granular income measure of their high school neighborhood.

There are several reasons to take this measure for good. First, there is quite a lot of income segregation across schools in Brazil. Therefore, allocating the same value to all students coming from the same high school is not unreasonable. Second, families typically live very close to where they send their children to high school. This supports the idea of assigning students the economic conditions of the zip code where their school is located. In Appendix Figure A3, we combine the measure we developed with other socio-economic variables self-reported by individuals in ENEM. We see that our economic background measure is highly correlated with measures such as living with less than one Minimum wage per month and parents' education. At the same time, there are no differences by gender, but a strong negative correlation

[5] This data has a similar structure than CESUP but for primary and secondary education.

with the probability of declaring as nonwhite.

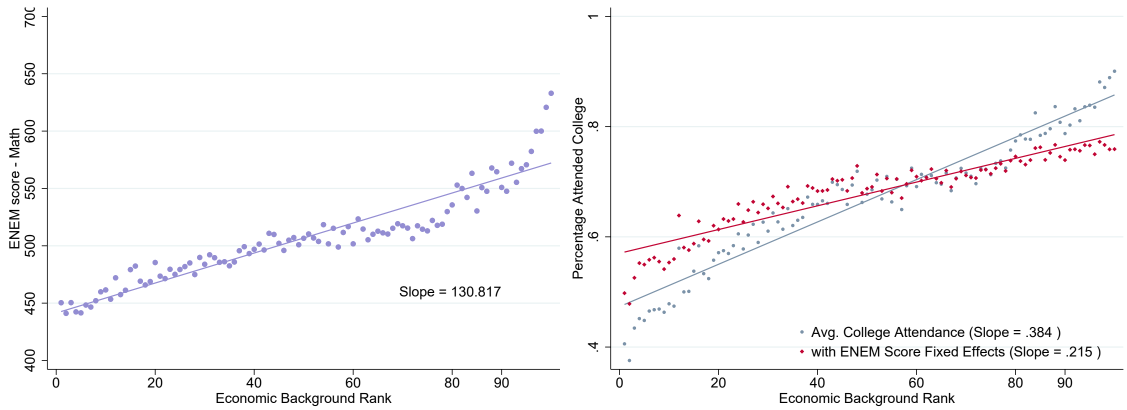
4 Attendance and Income Segregation Across Colleges

In this section, we show how college attendance varies across income levels and present statistics on income segregation in different types of higher education institutions.

4.1 College Access Across the Income Distribution

We begin showing that ENEM scores and college attendance increase substantially across the high school income distribution. We plot the average ENEM score in the math exam in Figure 2a, where we can see a monotonic increase throughout the income distribution, with a particular increase in the steepness of the slope starting from percentile 90. In Figure 2b we show how college attendance varies across the income distribution. College attendance is measured as being found enrolled in CESUP after high school graduation. For individuals in the top 10 percent of the high school income distribution, we see that over 90% of them attend college, whereas the probability is around 45% for those in the bottom 20 percent of the distribution. When controlling for ENEM scores, the slope of college attendance over the income distribution decreases substantially. We see that the slope reduces to 0.215 and it is mostly driven by the beginning of the income distribution, as the steepness decreases from percentile 50.

Figure 2: ENEM scores and College Attendance by High School Income Rank



(a) ENEM Score **(b)** College Attendance

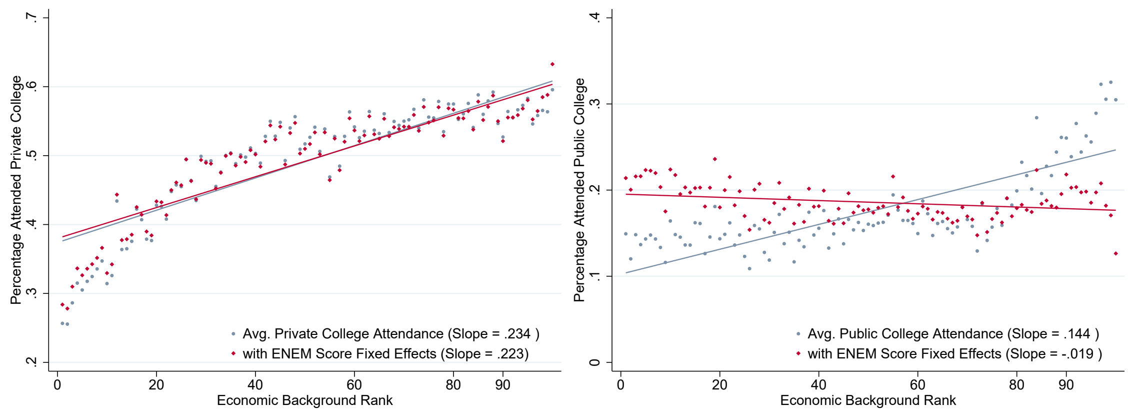
This Figure shows how exam scores and college attendance vary across the High School Income Rank distribution. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012. Each bin represents individuals at the respective percentile of the distribution. ENEM scores are measured as the standardized grades in mathematics. College attendance is defined as 1 if the individual appeared at least once in CESUP in the 7 years after they graduated High School and 0 otherwise. Red dots in Panel (b) are the average within bins of residuals from a linear regression of College Attendance on Fixed Effects of 5 points of math score in ENEM, summed with the average college attendance in the sample.

The patterns of college attendance across income distribution are strikingly different between public and private colleges. In Figure 3, we show college attendance rates with and without adjusting for ENEM scores divided by Public and Private institutions. As we can see in Figure 3a, attendance at private colleges is increasing across the whole income distribution and particularly in the bottom half of the distribution. When we adjust for exam scores, we do not observe any large change in the pattern of attendance. We interpret this as evidence that costs are an important barrier to private college attendance in Brazil, as students with the same ENEM grades, but with different economic backgrounds have substantially different attendance rates.

In Figure 3b we see that public college attendance is also increasing across the income distribution, but with a different pattern. The probability of attending a public college is flat until percentile 80 and then increases at the top of the distribution from around 15% to over 30%. But when we look at the probability adjusted for

ENEM scores, we see that the slope becomes flat indicating that, controlled by grades, the economic background of students is not determining public college attendance.

Figure 3: Attendance at Public and Private Institutions by High School Income Rank



(a) Private Colleges **(b)** Public Colleges

This Figure shows private and public college attendance vary across the High School Income Rank distribution. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012. Each bin represents individuals at the respective percentile of the distribution. College attendance is defined as 1 if the individual appeared at least once in CESUP in the 7 years after they graduated High School and 0 otherwise. If an individual appears both in private and public universities, we select the first observation after they graduate high school. Red dots are the average within bins of residuals from a linear regression of College Attendance on Fixed Effects of 5 points of math score in ENEM, summed with the average college attendance by administration type in the sample.

We summarize college attendance across the income distribution in Table A1, which shows the actual attendance rate and the attendance rate residualized by ENEM scores, summed by the sample average for different types of administration, degrees, and majors. On top of the already discussed differences in attendance between private and public colleges, we see that the type of degree also varies significantly across the income distribution. *Licenciatura* and *Tecnico* degrees are more common at the bottom of the distribution, whereas *Bacharelado* has the opposite pattern.

Attendance by major also varies substantially among groups. In the bottom part of Table A1 we show the probability of attending some selected majors. In Computer Science or Psychology, we observe that attendance is flat across the income distribution. In turn, majors such as Economics, Law, Medicine, and Engineering show a steep increase in attendance as income increases. A student in the top 5 percent of income is 10 times more likely to attend a Medicine major than one in the bottom 25 percent.[6] When controlling for ENEM scores, we observe that the differences across the distribution are attenuated significantly for engineering and medicine, two degrees with a large share of students in public universities. On the other hand, the probability of studying Law, a degree with a high share of private college students, does not change when controlling for ENEM grades.[7]

4.2 Income Segregation across Colleges

We next provide evidence of income segregation across different types of colleges. In Brazil, besides the public and private distinction, there are no clear groups of colleges such as the Ivy League in the United States or the Grandes Écoles in France, thus, to group universities into different tiers, we implement a data-driven ranking. We find all individuals in CESUP graduating between 2010 and 2012 who 9 years after graduating appear as formally employed in RAIS. We then calculate the average wages of graduates by college, and rank colleges by this measure. Henceforth, we refer to Elite Colleges as those in the top 5% of the rank.[8]

We show the income composition across different colleges in Figure 4. When comparing public and private colleges in Figure 4a, we observe that both have similar shares (25%) of students from the bottom 40% of the income distribution, but public colleges have a higher share of students from the top 20% of the income distribution (39% vs 31%). However, this masks important differences within both groups. Dividing the college sample across the graduates' earnings rank, we observe in Figure 4b that Elite Colleges are highly segregated, with more than 60% of its students coming

[6] Most majors that decrease and explain the remaining share are *Licenciatura* degrees.

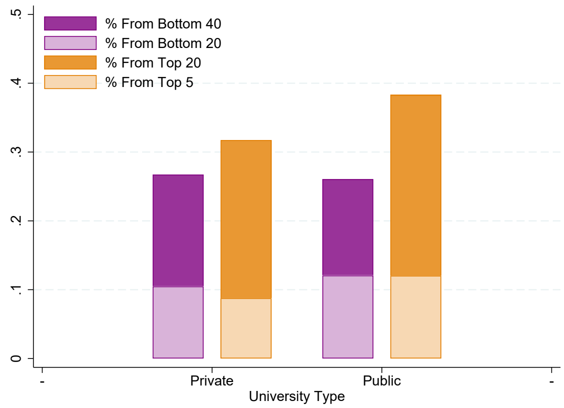
[7] In Brazil, Law and Medical school are not post-graduate degrees. Students enter directly from high school to those degrees.

[8] We restrict ourselves to this time range such that none of the students in our main analysis sample are included in the sample for the ranking. Institutions that had less than 10 graduates in each year are excluded from the rank. The rank is done without weighting colleges by their number of students.

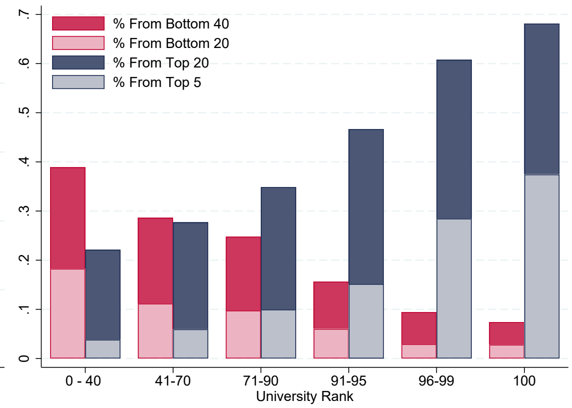
from the top 20% of the income distribution and less than 10% from the bottom 40%. In Figure A7b we show that the pattern across the graduates' earnings rank is consistent for both public and private universities, albeit elite public universities are slightly less segregated than elite private universities.

Looking at specific college examples helps us understand these patterns. In Figure 4c we see that University of Sao Paulo (USP) and University of Brasilia (UNB), two elite public universities, have less than 8% of their students coming from the bottom 40% of the income distribution, and around 23% and 35% of its students respectively coming from the top 5%. On the other hand, a still really good public, but not an elite university such as the Federal University of Bahia (UFBA) has around 25% of its students from the bottom 40% and less than 20% of its students from the top 5%. When looking at private colleges examples, we see that PUC-Rio, an elite private college with high tuition fees, has almost 70% of its students from the top 5% of the income distribution and less than 5% from the bottom 40. In turn, Universidade Paulista (UNIP), a large private university with accessible tuition fees, has an equal share of students (25%) from the bottom 40% and top 20% of the income distribution.

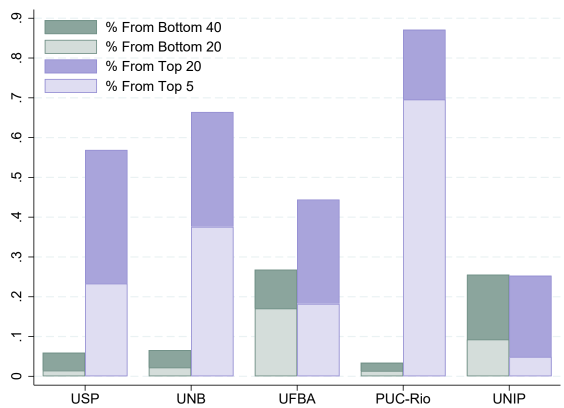
Figure 4: Income Composition in Different Colleges



(a) Public and Private



(b) Rank by Future Income

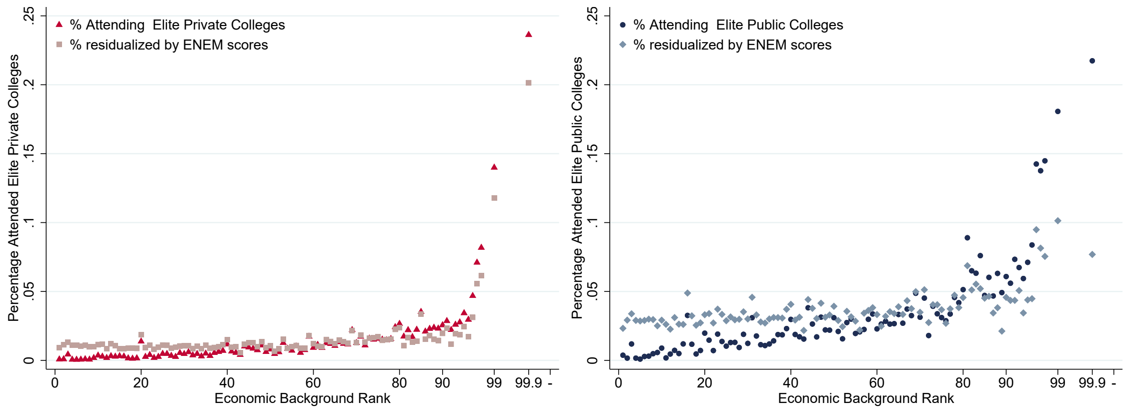


(c) Selected Colleges

This Figure shows the income composition of entrants across different types of colleges. The sample comprises of all entrants between 2010 and 2015 for whom we have information in ENEM of where they graduated from in high school. The remaining observations of entrants are treated as missing. In Panel (b) colleges are ranked by average earnings of graduates between 2010 and 2013.

There are still substantial differences between elite public and private colleges in how ENEM grades affect enrollment. In Figure 3 we plot elite college attendance across the high school income rank across private and public colleges with and without controlling for ENEM scores. We see in both panels that the probability of attending elite colleges increases substantially from the 90th percentile. When controlling for ENEM grades, we see small changes in the probability of attending elite private colleges, on the other hand, the steep increase in the probability of enrollment in elite public colleges decreases substantially. [9]

Figure 5: Attendance at Elite Public and Private Colleges by Income Rank



(a) Private Colleges **(b)** Public Colleges

This Figure shows elite private and public college attendance vary across the High School Income Rank distribution. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012. Each bin represents individuals at the respective percentile of the distribution. College attendance is defined as 1 if the individual appeared at least once in CESUP in the 7 years after they graduated High School and 0 otherwise. If an individual appears both in private and public universities, we select the first observation after they graduate high school. Red dots are the average within bins of residuals from a linear regression of College Attendance on Fixed Effects of 5 points of math score in ENEM, summed with the average college attendance by administration type in the sample.

In Table 1 we summarize income segregation levels across different types of colleges as well as characterize them in terms of other demographic and institutional characteristics. [10]. First, we observe that the share of students living in households that had less than one minimum wage of per capita earnings is highly decreasing across the college earnings rank, and the share of individuals whose parents had college degrees increases substantially across the college's rank. This serves also to validate our high school income distribution as these measures are self-reported at the individual level.

[9]In Figure A8 we show the same graph for the different types of colleges ranked by the graduates' earnings rank and divided by public and private colleges.

[10]Demographic characteristics such as the share of households with less than one minimum wage per capita and parents' education are calculated from the self-reported ENEM questionnaire. In the case of gender and race, if we do not observe these students in ENEM, we input the values from CESUP. The number of students and majors as well as the majors' composition are collected from CESUP

We see that the share of female and nonwhite students also decreases in higher-ranked colleges. This reflects the correlation of income and race in Brazil, as the nonwhite population is significantly less favored economically than the white population.[11]

We also observe that the number of students is increasing in the college rank for public universities, but that is not true for private ones, which on average have fewer majors and a smaller number of students than mid-level private colleges. There are also differences in the major composition across college types. First, we see that almost half of the students in private institutions are in Business, Social Sciences, and Law majors. In Public schools, education majors who are mostly targeted to eventual elementary and high school teachers, are overrepresented at the bottom of the college rank, but at the top, the students are more equally distributed across the remaining areas. Engineering is the area with the biggest increase across the college rank distribution for both public and private universities.

[11] In the case of female students, the decrease can be explained by higher share of women attending college on average, which is equalized in elite institutions.

Table 1: Income Segregation Across Colleges

	Private Colleges					Public Colleges				
College Rank	0-40	41-70	71-90	91-95	96-100	0-40	41-70	71-90	91-95	96-100
Income Characteristics										
% from bottom 40%	39.36	27.35	23.03	16.28	7.63	43.20	35.86	32.59	16.03	10.25
% from bottom 20%	18.60	9.94	7.72	6.08	2.53	22.50	17.61	16.22	6.46	3.15
% from top 20%	21.72	27.05	34.02	46.65	63.54	20.35	25.85	32.98	47.24	60.22
% from top 10%	10.22	13.18	17.87	29.02	45.48	8.17	11.63	18.29	26.35	39.43
% from top 5%	3.87	6.05	9.71	17.39	33.70	1.53	4.48	8.58	14.93	27.26
Avg. Income Rank	51.93	60.42	63.93	70.73	81.10	47.38	53.02	57.07	69.13	76.87
% in H.H. with < M.W.	21.82	14.71	12.85	8.59	3.59	26.46	22.51	18.83	9.63	6.06
Demographics										
% Female	63.05	59.17	57.91	54.24	46.27	59.49	57.21	53.63	50.62	49.70
% Nonwhite	58.91	49.84	48.31	44.82	27.09	53.25	56.22	56.97	42.36	36.10
% Father had Coll. Degree	11.63	14.28	17.92	29.17	55.30	12.50	14.81	22.17	31.31	44.59
% Mother had Coll. Degree	15.72	17.17	19.90	31.13	56.66	17.92	20.32	27.12	35.57	46.48
Univ. Characteristics										
Avg. Number of Majors	43.66	47.45	281.79	70.61	29.08	29.28	94.59	95.87	117.68	158.80
Avg. Number of Students	4,148.15	7,651.64	16,016.64	10,782.69	4,475.58	2,958.16	9,218.37	16,822.71	18,841.36	20,765.15
Avg. ENEM score	513.24	523.05	537.01	568.38	636.08	529.96	557.94	588.55	644.75	682.43
Share of Students in Each type of Major										
Education	20.64	15.68	12.62	6.43	2.48	40.57	39.97	34.73	22.90	17.85
Humanities and Arts	1.19	1.51	2.55	4.62	4.93	0.56	2.22	2.06	5.11	5.49
Business Social Sci. and Law	41.10	44.48	45.74	45.49	50.95	27.51	21.84	19.35	16.90	18.81
Natural Sciences and Math	4.22	5.20	6.19	6.17	7.63	6.40	7.18	10.06	12.04	12.95
Engineering	7.39	12.50	15.02	18.74	26.15	5.59	9.84	13.20	19.51	18.26
Agric. and Vet	1.25	1.81	1.06	1.11	0.00	5.59	7.64	5.69	5.65	2.69
Health and Services	21.38	16.44	15.07	12.91	6.31	10.66	9.28	11.99	11.72	12.82
Others	2.83	2.37	1.73	4.53	1.54	3.13	2.03	2.91	6.17	11.14
Observations	589	434	256	45	53	61	52	68	36	27

Notes: This table shows college attendance across the high school income rank distribution using individuals graduating high school and taking ENEM between 2010 and 2012. Avg share by degrees and majors are calculated conditional on attending college. In case an individual enrolls in multiple institutions, we use the first observation after graduating high school. All results are in percentages and are not adjusted by ENEM scores.

5 Students' Earnings Outcomes

In this section, we show how students' wages vary across colleges. We describe how students' wage ranks are persistent across the distribution of high school income. We then show how wages vary across different types of colleges.

We begin defining our wage rank measure. We start by matching our main analysis sample which comprises of high school graduates taking ENEM between 2010 and 2012, to labor market outcomes in RAIS 9 years after high school graduation. Then, within those who we find in the RAIS sample, we rank each student within their cohort of high school graduation according to their December monthly wage. In Appendix XX we show labor market participation rate and a series of robustness checks.

We find that high school income rank strongly correlates with students' wage rank 9 years after graduating. We show our findings in Figure 6, where we plot the average wage rank in each percentile of the H.S. income rank. We see in the National line that on average an increase in one percentile in the H.S. income rank correlates with a 0.23 increase in the wage rank[12]. We also observe that the steepness of the slope increases starting in percentile 90, suggesting a higher persistence in earnings at the top of the high school income distribution.

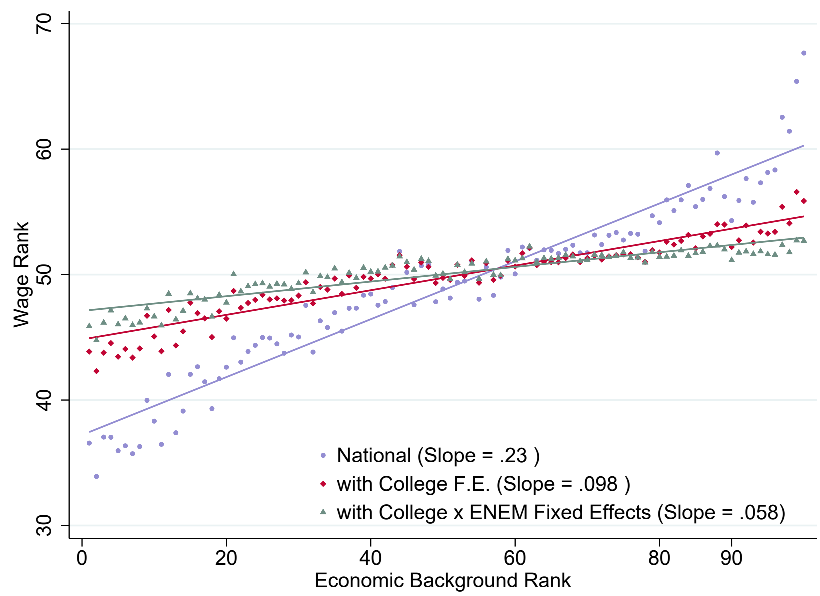
There are two main reasons for that difference. First, The other two variables plotted in Figure 6 are wage ranks residualized by College Fixed Effects and College interacted with ENEM scores fixed effects. To construct those variables we execute a linear regression of wage rank on the respective fixed effects, recover the residuals, and sum the average rank in the sample. We then take the average of the corresponding measure within each percentile of the high school income distribution.

The correlation between high school income rank and students' future wage rank decreases substantially when controlling for ENEM scores and college fixed effects, but still persists across all areas of the high school income distribution. Our results suggest that increasing one percentile in the high school income distribution increases the students' wage rank by 0.098 when controlling for College Fixed Effects, and 0.058 when controlling for fixed effects the interaction of College and bins of the ENEM

[12] Our estimates for the slope are smaller than intergenerational mobility estimates in Brazil(GC Britto et al., 2022). Different from other estimates, our sample comprises only individuals graduating from high school and taking ENEM. This initial selection to our estimates that makes results not comparable.

grade.[13]

Figure 6: Relationship Between Students' Wage Rank and Economic Background Rank



This Figure shows how Wage rank 9 years after high school graduation varies across the High School Income Rank distribution. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012 who are found employed in RAIS 9 years after graduating. Each dot represents the averages of individuals at the respective percentile of the high school income rank. The red and green markers show the average wage rank in the sample summed with residuals from a linear regression of wage rank on College F.E. and College interacted with bins of 5 points in ENEM math score respectively. Individuals who do not attend college are grouped into one single category. For individuals with multiple colleges in CESUP, we select the first one they appear after graduating high school.

Next, we show how wage rank varies across the high school income rank within the different groups of colleges. We provide four different measures in Figure 7, grouped in bins of 5 percentiles of the high school income rank[14]. In Figure 7a we show the average wage rank. In Figures 7b and 7c we show the residuals of a linear regression of wage rank on the respective set of fixed effects summed with the average wage rank of each college type.

[13]In Figure A10 we show how wage rank behaves when controlling only for ENEM scores. The slope is 0.127, slightly higher than when we control for College Fixed Effects.

[14]We group the percentiles because of the smaller sample in each percentile with each college group. We also restrict the Figure to 5 types of colleges, grouping the top 5 percent because the last group by itself has very small sample, particularly in the bottom part of the H.S. income rank. We show the result with the top percentile in Figure A11

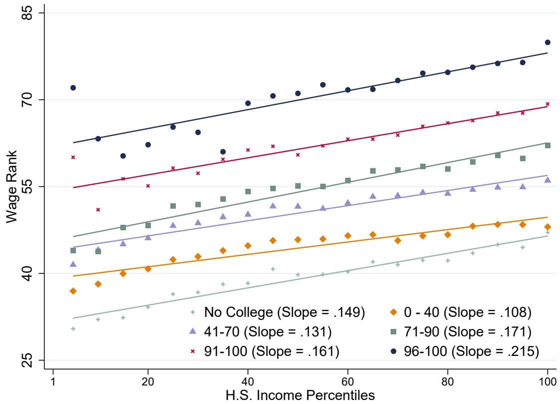
The relationship between future wages and attending higher-ranked colleges is consistent within all economic backgrounds. We can see that by comparing lines vertically, as the average wage rank is increasing across college types for all bins of the high school income distribution. Albeit smaller, these differences are persistent when controlled by College interacted with Major fixed effects, and by 5 point bins of ENEM grades.

However, within higher-ranked colleges, the relationship between previous economic background and future wage ranks is stronger than within lower-ranked colleges. In Figure 7a, the slope for the top 5% of colleges is 0.215, almost as big as the national slope in Figure 6. Within colleges in the bottom 40% of our rank, the slope is 0.108, half of the measure in Elite Colleges. This is the opposite of what Chetty et al. (2020) document in the United States. They show that within elite colleges, the slope is smaller than in other four-year colleges or two-year colleges.[15]

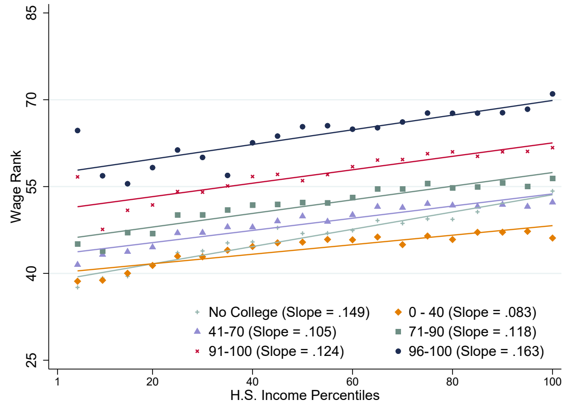
The strong correlation between economic background and future wage rank within elite colleges cannot be explained by differences in the composition of colleges or majors within elite institutions. We see that in Figure 7b, where we show wage rank residualized for college interacted with major fixed effects. We still observe a correlation of 0.16 between High School income Percentiles and wage rank. However, the correlation decreases substantially when controlling for ENEM scores as we show in Figure 7c. This illustrates that even within majors there is significant variation in previous grades, which are strong predictors of future wage rank.

[15] Slope magnitudes are not exactly comparable to their measure, as Chetty et al. (2020) run the slope with college fixed effects whereas we do not.

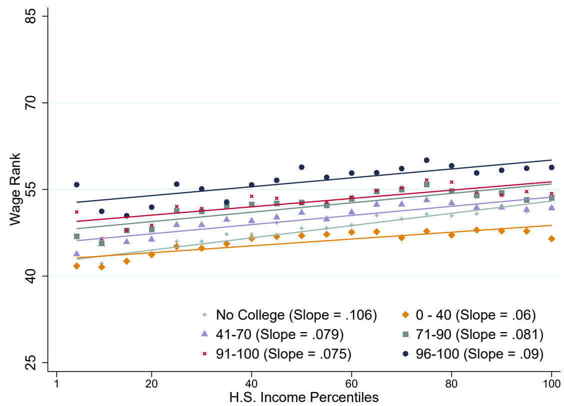
Figure 7: Relationship Between Wage Rank and H.S. Rank Within Different Colleges



(a) Average Wage Rank



(b) Controlling by College x Major F.E..



(c) Contr. College x Major x ENEM F.E.

This Figure shows how Wage rank 9 years after high school graduation varies across the High School Income Rank distribution for different colleges ranked by their graduates' between 2010 and 2012 future earnings. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012 who are found employed in RAIS 9 years after graduating. Each dot represents the averages of individuals at the respective percentile of the high school income rank. Panels (b)-(d) show the average wage rank in the sample summed with residuals from a linear regression of wage rank on College X Major F.E., ENEM Grades, and College X Major X bins of 5 points in ENEM math scores respectively. Individuals who do not attend college are grouped into one single category. For individuals with multiple colleges in CESUP, we select the first one they appear after graduating high school.

5.1 Robustness with Other Earnings Measure

Having a comprehensive measure of income is challenging in the Brazilian context. A large fraction of business owners attended college, implying that their source of income is not registered in RAIS. Therefore, it is not possible to directly observe their income in the absence of tax records. Even with them, informality and self-employment impose additional challenges. We do several things to address this issue, providing robustness to our findings.

Our main results rely on wages reported in *RAIS*, meaning that we compute missing income for those individuals who are not formally employed. We show in Figure A4 that employment status has a soft inverted U-shape, especially at the bottom of the income distribution. This means that we are likely overestimating the wage rank of low-income students, translating into an overestimation of social mobility. Therefore, we consider our estimates lower bonds of how little social mobility we find in Brazil.

Moreover, in Figure A5, we document how entrepreneurship and *Pejotas* are distributed across our measure of household income. *Pejotas* are individuals who opened a firm to receive wages, a widespread phenomenon in Brazil. It also includes real contractors with no employment relationship with a firm. The key characteristic is that they are firms with no employees. In both cases, we observe that students from high-ranked schools are overrepresented in these groups. This means that by using only wages, we are also overestimating social mobility, because students from rich households who are business owners are likely to earn more than employees, leading to an underestimation of their income.

We partially correct the last issue by producing an alternative measure of earnings that complements wages. For business owners with employees, we input the wage of the highest-paid employee to the owner. For business owners with no employees, we input the average wage of students in their same class that does appear in *RAIS*. If no classmate appears in *RAIS*, we move to a higher level of aggregation, such as students with the same type of degree in the same university (we exploit the fact that degrees can be grouped at different levels of detail). The reason to do so is to capture the outside option (wage) these individuals could get. In Figure A6, we show that the main results are unchanged, except for the very top, where social mobility is slightly lower (larger slope) when we take into consideration other labor activities that rich students are more likely to perform.

5.2 College Level Intergenerational Mobility

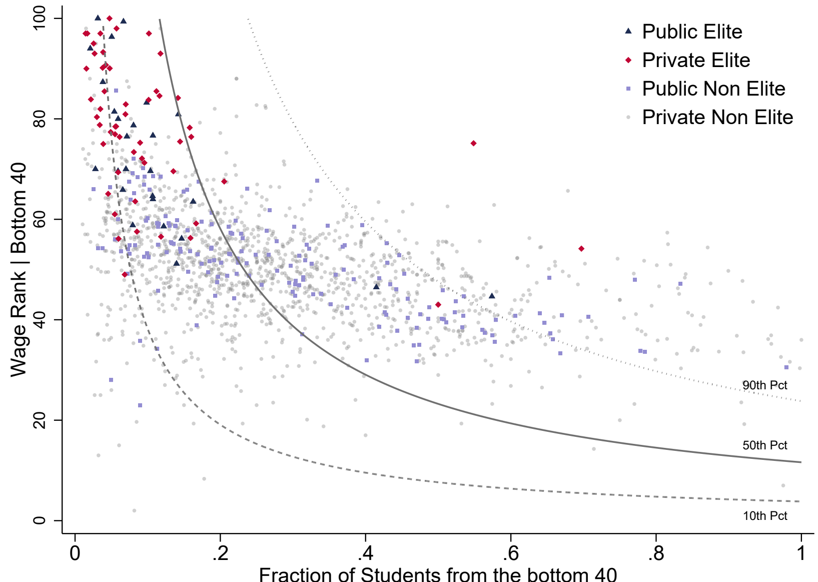
Next, we estimate college-level intergenerational mobility rates. To understand intergenerational mobility, we need to take into account both income segregation in colleges and students' future earnings. We create a measure similar to Chetty et al. (2020), by examining what is the average wage rank of students in a given who were from the bottom part of the income distribution, multiplied by the fraction of students in a given college that originate from a low economic background. The expression below defines our mobility rate in a given college j. We use the bottom 40% of the income distribution as our measure of being from a disadvantaged economic background, multiplied by the average of these students' future wage rank.[16].

$$\text{Mobility Rate}_j = Prob(\text{Bottom } 40\%)_j \cdot (\text{Avg. Wage Rank}|\text{Bottom } 40\%)_j$$

In Figure 8, we plot the share of students from disadvantaged backgrounds and their future wage ranks, highlighting elite private and public colleges, and the indifference curves of mobility rates for the 10th percentile, median and 90th percentile. We observe that, despite having high-achieving students, elite colleges are on average at the bottom of the college mobility rate distribution, as the share of students from disadvantaged backgrounds is small. Elite public colleges have on average a slightly higher representation of disadvantaged students, but most of them are still between the 10th and the 50th percentile of mobility rates.

[16]Chetty et al. (2020) use the share of students from the bottom quintile instead of the bottom 40%. In our setting, the share of students from the bottom 40% is very low in many universities, this generates noisy estimates of students' future earnings. Figure A12 shows the same exercise as in Figure 8 but using the bottom quintile of economic background.

Figure 8: Mobility Rates Across Colleges



This Figure show in the x-axis the fraction of students in each university who comes from the bottom 40 of the H.S. income rank. In the y-axis we show the average wage rank at the age of 30 that students from the bottom 40 of the H.S. income rank reach. It shows a very clear pattern highlighting that universities with lower fraction of students from the bottom 40 are the same that place these disadvantaged students at the top of the income distribution. Among those, elite schools (both private and public) are clearly highly segregated and with high returns for low income students.

We show average mobility rates across different types of colleges in Table 2. Columns (1) - (3) show results using the bottom 40% as the economically disadvantaged group, whereas columns (4)-(6) show the same exercise but with the bottom 20% of the economic background measure distribution. We see that public and private colleges have similar mobility rates, with comparable composition and outcomes. Students from the Bottom 40% have slightly better outcomes in public colleges but are also slightly underrepresented. When we look at college rank, we observe that mobility rates decrease monotonically as college rank increases. This happens because lower-ranked colleges have a much higher share of students from disadvantaged economic backgrounds. When focusing on elite colleges, those at the top 5 percentiles of college rank, we observe that private colleges have higher outcomes, but a lower share of disadvantaged students. This makes mobility rates higher in public elite colleges than in private ones.

Table 2: Mobility Rate for Different Colleges

	(1) Avg. Wage Rank for bottom 40%	(2) Share of Students From bottom 40%	(3) Mobility Rate	(4) Avg. Wage Rank for bottom 20%	(5) Share of Students From bottom 20%	(6) Mobility Rate
Public and Private Colleges						
Public Colleges	53.41	0.26	12.43	52.53	0.12	5.38
Private Colleges	51.40	0.27	12.72	50.07	0.10	4.71
College Rank						
0-40	43.31	0.39	16.20	41.90	0.18	7.34
41-70	49.37	0.29	13.53	48.04	0.11	5.05
71-90	52.62	0.25	12.39	51.01	0.10	4.54
91-95	60.05	0.16	9.01	60.19	0.06	3.35
96-100	70.04	0.09	6.08	69.99	0.03	1.91
Elite Colleges Public and Private						
Elite Public	66.88	0.10	6.39	67.16	0.03	1.95
Elite Private	77.42	0.07	5.36	76.78	0.02	1.81
Selected Colleges						
USP	79.99	0.06	4.75	83.36	0.01	1.21
UNB	65.86	0.07	4.33	64.73	0.02	1.43
UFBA	61.81	0.27	16.57	61.81	0.17	10.54
PUC-Rio	61.81	0.03	2.68	61.81	0.01	1.23
UNIP	78.79	0.26	12.18	89.86	0.09	3.90

Notes: This table shows intergenerational mobility measures across different groups of colleges. The sample comprises all entrants in CESUP between 2010 and 2015. Moments are weighted by the number of students in each college. Mobility rate corresponds to the multiplication of the average wage rank of students from the bottom 40% and the share of students from the bottom 40%. Mobility Rate in column (3) is the multiplication of columns (1) and (2) and mobility rate in column (6) is the multiplication of columns (4) and (5).

6 From Income Segregation to Social Mobility

In this section, we develop a sufficient statistic framework to evaluate how policies targeting college attendance can affect the income of the most disadvantaged students. We then apply it to understand how two different policies conducted by the Brazilian government affected income segregation and mobility.

6.1 *Conceptual Framework*

The framework allows for different responses from individuals to policies and different returns to each higher education option. Our goal is to provide sufficient statistics at an aggregated level that allow us to understand how policies shape mobility rates in each part of the economic background distribution.[17]

We start by defining our outcome of interest for individuals from economic background b, i.e., how the average income of this group responds to a given policy"

$$\Delta \bar{I}_b = \bar{I}_{bp} - \bar{I}_{b0} \tag{1}$$

The first thing to note is that average income in a given group b can be written as a weighted average of the different subgroups multiplied by their income:

$$\bar{I}_b = \frac{1}{N_b} \sum_{j=0}^{J} N_{bj} I_{bj} \tag{2}$$

In our application, j refers to a particular college tier, including not attending college. Second, a policy that aims to increase college attendance of a particular group b will change the share of students in group b attending each tier j. Therefore, let's define for each group b the transition vector ϵ_j for individuals in each college tier j:

$$\epsilon_j^b = \begin{pmatrix} \epsilon_{j0}^b & \epsilon_{j1}^b & \cdots & \epsilon_{jj}^b & \epsilon_{jJ}^b \end{pmatrix}$$

Where $\epsilon_{jj'}^b$ is the number of individuals in group b that moved from college tier j to j'. Note that $\sum_{j'=0}^{J} \epsilon_{jj'}^b = N_{bj}^{pre}$, meaning all the individuals from college tier j that went to others or stay in the same must some up to all the individuals that were in that college tier beforehand. Therefore, we can write income of college tier j in

[17] In our exposition, we focus on analysis aggregated at a given group level. In Appendix A we show how to derive our aggregate conceptual framework from a potential outcomes model at the individual level.

group b after the policy as:

$$I_{bp}^{j} = \begin{pmatrix} \epsilon_{j0}^{b} & \epsilon_{j1}^{b} & \cdots & \epsilon_{jj}^{b} & \epsilon_{jJ}^{b} \end{pmatrix} \cdot \begin{pmatrix} I_{j0}^{b} \\ I_{j1}^{b} \\ \vdots \\ I_{jj}^{b} \\ I_{jJ}^{b} \end{pmatrix}$$

Where $I_{jj'}^{b}$ means the income that an individual in group b moving from tier j to j' due to the policy would earn. This can be decomposed between how much this individual would earn in the absence of the policy (defined by the tier and group he would belong to), and the value added that the new tier brings to someone like her.

$$I_{jj'}^{b} = I_{j}^{b} + \mathrm{VA}_{jj'}^{b} \tag{3}$$

Therefore we can rewrite I_{bp}^{j}:

$$I_{bp}^{j} = I_{j}^{b} \underbrace{\sum_{j'=0}^{J} \epsilon_{jj'}^{b}}_{=N_{bj}^{pre}} + \sum_{j'=0}^{J} \epsilon_{jj'}^{b} \mathrm{VA}_{jj'}^{b} \tag{4}$$

$$\bar{I}_{bp} = \underbrace{\frac{1}{N_b} \sum_{j=0}^{J} I_{j}^{b} N_{bj}^{pre}}_{=\bar{I}_{b0}} + \frac{1}{N_b} \sum_{j=0}^{J} \sum_{j'=0}^{J} \epsilon_{jj'}^{b} \mathrm{VA}_{jj'}^{b} \tag{5}$$

Concluding

$$\Delta \bar{I}_{b} = \frac{1}{N_b} \sum_{j=0}^{J} \sum_{j'=0}^{J} \underbrace{\epsilon_{jj'}^{b}}_{\text{Composition Effect}} \underbrace{\mathrm{VA}_{jj'}^{b}}_{\text{College Effect}} \tag{6}$$

Equation (6) provides a sufficient statistic for the change in the average income of group b, which can be decomposed in two parts. The first one is the *composition effect*, which captures which college tiers individuals in group b attend to, relative to where they would attend in the absence of policy. Note that $\epsilon_{jj'}^{b}$ are semi-elasticities, meaning that it not only matters increasing the *fraction* of students in tier q from group b, but also *how many* students two alternative policies are able to reallocate. On the other hand, the *college effect* accounts for the role of higher education in shaping income distribution. Note that if all returns to college are purely driven by selection (there is no value added to going to a given college), then reallocating students across college tiers has no impact on income. In the same way, if the policy is not effective at increasing the number of disadvantaged students in high value-added college tiers, we should expect no change in disadvantaged students' income.

An important advantage of this framework is that it makes all the assumptions explicit. As an example, take an affirmative action policy that mandates that elite colleges must increase the number of disadvantaged students. In this case, elite colleges refer to one particular college tier j and disadvantaged students to a particular group b. Empirical research designs need to assume the absence of *defiers* to recover the AA policy's causal effect on enrollment in elite schools. This assumption translates into setting $\epsilon_{ej}^{d} = 0$ with $j \neq e$, meaning that no disadvantaged students (d) that would go to an elite school in the absence of the policy will now attend a non-elite college. Moreover, when evaluating changes in total income $\sum_b \Delta \bar{I}_b$, it makes explicit the general equilibrium effects of displaced students coming from groups $b \neq d$. Finally, note that improvements in measuring outside options of targeted groups and college value-added can be used within this framework. Therefore, we hope this framework provides guidance on the key elasticities we need to improve measurement if we want to evaluate particular programs that promote college enrollment in order to improve certain groups' outcomes.

In what follows, we use Equation (6) for two policies the Brazilian government has implemented in the past to learn about their effectiveness through the lens of our framework. **Disclaimer: In the secure room, we can access the data to run the specification that matches Equation (6). However, on our last visit, we did not extract the right variables we needed for the exercise. We will**

use other values in order to illustrate how the framework works. In the next version of this draft, we will provide the right estimates.[18]

6.2 Estimating the Effects of Affirmative Action and Subsidized Loans

We follow our conceptual framework to understand the effects of affirmative action and subsidized loans on aggregate patterns of social mobility. We first investigate empirically how the composition of the student body in affected colleges change because of the affirmative action program. We then describe how we compute measures of value added per college. Lastly, we use our empirical estimates to calculate how income changes across students from different economic backgrounds.

Due to sample limitations, we restrict our focus here to two different backgrounds, which we will refer to as disadvantaged students, and non-disadvantaged students. Furthermore, we will group our higher education options into 6 different groups according to our college rank. Those groups are the same as the ones describe in section 4 and 5, with the addition of no college as a higher education option. Thus individuals are dividided between no college, colleges ranked 0-40, 41-70, 71-90, 91-95, 96-100.

Changes in Student-Body Composition caused by Affirmative Action

As detailed in section 2, since 2004, public universities in Brazil have adopted affirmative action policies that reserved seats to students from underrepresented background. In August 2012, the Brazilian federal government passed the *Law of Social Quotas*, requiring all federal institutions to reserve half of their admission spots in each degree program for students coming from public high schools. Using CESUP, we observe that during this period the share of seats reserved for affirmative action from around 12%

[18] First, we did not extract the variables to match the semi-elasticity for the number of new admits of group b matching with the same group b for with we analyzed social mobility. Second, we did not extract the average income at the group level b, but the average income rank of individuals in group b. The second point is more complicated because there is no 1-to-1 function from income to rank. In the Appendix we provide an illustration on how results will look like when extracting the right variables.

in 2010 to over 30% in 2015[19].

To estimate the effects of affirmative action on student-body composition by type of college, we extend the analysis by Mello (2023) to the different types of colleges from our rank. This consists of estimating the following equation:

$$Y_{pt} = \alpha_p + \delta_t + \beta \cdot \text{Sh. of Affirmative Action}_{u(p)t} + \Gamma' X_{pt} + \varepsilon_{pt} \tag{7}$$

where $\{\alpha_p, \delta_t\}$ are program and year fixed effects, respectively. Sh. of Affirmative Action$_{u(p)t}$ corresponds to the share of vacancies filled through affirmative action in university u where program p is offered. X_{pt} is a vector of time-varying program characteristics. We include the number of students in each program as a covariate, to control for potential changes in the supply of seats that could be correlated with affirmative action. We also include the share of seats reserved to centralized admissions (SISU)[20] Y_{pt} is our main dependent variable which is the share of students entering in program p at year t who come from a disadvantaged background.

We interpret β as the increase in the share of students from disadvantaged backgrounds in programs going from 0% of seats reserved for affirmative action to 100% of seats reserved. Our identifying assumption is that in the absence of changes in the share of seats reserved for affirmative action, different programs would have parallel trends in the share of disadvantaged students.

We restrict our sample to public colleges and estimate Equation 7 using program-level data from 2010 and 2015[21]. We show our findings in Table 3. Column (1) shows estimates using the full sample of programs in public colleges. Columns (2)-(6) are estimated restricting the sample to programs from the respective college rank.

[19] We show the share of seats in public colleges reserved for affirmative action by year in Panel (a) of Figure A17

[20] Machado and Szerman (2021) shows that SISU, the centralized admission system which was expanded during this period, increases selectivity in colleges and reduces the share of disadvantaged students.

[21] Mello (2023) main estimates are done at the student-level, which changes the interpretation of the coefficient. However, in robustness checks, she estimates a similar model at the program level.

We observe that on average, the increase in the share of students from disadvantaged backgrounds is relatively small. Using all colleges, we estimate β as 0.0195. When we look at heterogeneity according to college types, we observe that colleges in the bottom part of the distribution have no change due to affirmative action. This is increasing in college rank, up to elite colleges who observe a substantial increase in disadvantaged students due to affirmative action.

Table 3: Affirmative Action and Income Segregation in Public Colleges

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep Var: Share of Disadvantaged Students					
	All Colleges	College Rank				
		0-40	41-70	71-90	91-95	96-100
Sh. of Affirmative Action	0.0195***	-0.000441	0.0170	0.0221***	0.0160**	0.0454***
	(0.00323)	(0.0198)	(0.00896)	(0.00556)	(0.00493)	(0.0104)
Dep. Var Mean	.14	.27	.23	.17	.08	.05
Sh. A.A. in 2010	.11	.06	.12	.12	.1	.09
Sh. A.A. in 2015	.3	.2	.22	.34	.32	.26
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31707	833	5041	12148	7385	5555

Notes: This table shows estimates of coefficient β from equation 7. The sample comprises all programs in Public colleges that have at least one entrant in all years between 2010 and 2015. A program is defined as a major x college interaction. Share of disadvantaged students is measured as the share of students that report less than one minimum wage of household income. Standard errors are clustered at the program level. In Columns (2)-(6), sample is restricted to programs within the respective college rank.

There are different explanations for the small coefficient. While affirmative action guarantees 50% seats reserved in each university, the quotas do not typically target a unique group. Figure (A16) shows that the law targeted students from public high schools, which represent 81% of total high school graduates, and included race and income criteria. Moreover, the income criteria was less than 1.5 minimum wages.[22] Therefore, it is far from obvious to what extent the policy is able to increase total enrollment of disadvantaged students. Mello (2023) estimates suggest that affirmative action increases enrollment of students from households with less than one minimum wage by 2.4 percentage points (representing a 34% increase) for the cohorts 2010-

[22] It has been recently changed to 1 minimum wage because it was too broad.

2015. Furthermore, we can see that especially in lower-ranked colleges, the share of disadvantaged students was already high (0.27%). If disadvantaged students who were previously enrolling through open contest seats are now enrolling in the affirmative action seats, it is expected that a change in affirmative action seats has small effects on the composition of the student body.

Changes in Student-Body Composition caused by Subsidized Loans

On top of the public provision of colleges, the federal government is also heavily involved in financial aid to students in private colleges. Two main programs finance higher education in Brazil: PROUNI (*Programa Universidade Para Todos*) and FIES (*Fundo de Financiamento ao Estudante do Ensino Superior).* The main difference between both is that PROUNI gives scholarships for students, whereas FIES gives out subsidized loans. In our empirical analysis, we will focus on FIES, as it has the most variation during our analysis period.[23]

FIES was structured in 1999 by the federal government replacing previous programs of financial aid. In the following decade, the program gave financial loans charging the benchmark interest rate in Brazil (SELIC), set by the Central Bank of Brazil. In 2010, the government increased the program's budget and reduced the interest rate to less than half of the Central Bank's benchmark rate. That increased substantially the number of students that obtained FIES loans as we can see in Panel (b) of Figure A17.

Students are eligible for FIES loans if they have less than 3 minimum wages of household income. In 2015, there was a change in eligibility restricting loans only to students with more than 450 points in all ENEM exams.

To understand how FIES affects income segregation across colleges, we use the

[23] ***PROUNI*** was created in 2005 by the federal government, and provides partial or full scholarships to students to attend private universities. Private universities that offer seats in PROUNI are given tax incentives that match the amount of seats made available by PROUNI. Eligibility is determined by household income, having studied in a public high school, and ENEM grades.

same strategy as before, estimating the following equation but with a sample of private college programs.

$$Y_{pt} = \alpha_p + \delta_t + \beta \cdot \text{Sh. of Students with FIES}_{u(p)t} + \Gamma' X_{pt} + \varepsilon_{pt} \tag{8}$$

where $\{\alpha_p, \delta_t\}$ are program and year fixed effects, respectively. Sh. of Students with $\text{FIES}_{u(p)t}$ corresponds to the share of students with FIES loans in university u where program p is offered. We also include the number of students in each program as a covariate, to control for potential changes in the supply of seats that could be correlated with affirmative action.

We observe that FIES has bigger effects on increasing the share of disadvantaged students than affirmative action. On average, one additional student with FIES loan, implies 0.084 students from disadvantaged backgrounds. However, effects are stronger at the bottom of the college rank distribution, and despite an increase in the share of students with FIES in elite private colleges from 0.02 to 0.07, we see no change in the share of disadvantaged students.

Table 4: Financial Aid and Income Segregation in Private Colleges

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep Var: Share of Disadvantaged Students					
	All Colleges	College Rank				
		0-40	41-70	71-90	91-95	96-100
Sh. of Students with FIES	0.0845***	0.0750***	0.0768***	0.0760***	0.0615***	0.00474
	(0.00342)	(0.00659)	(0.00573)	(0.00639)	(0.0136)	(0.0174)
Dep. Var Mean	.12	.19	.13	.1	.07	.03
Sh. FIES in 2010	.02	.02	.02	.02	.02	.02
Sh. FIES in 2015	.15	.2	.16	.14	.14	.07
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	67087	15172	21409	22719	2947	1505

Notes: This table shows college attendance across the high school income rank distribution using individuals graduating high school and taking ENEM between 2010 and 2012. Types of degrees and majors are calculated conditional on attending college. In case an individual enrolls in multiple institutions, we use the first observation after graduating high school. All results are in percentages and are not adjusted by ENEM scores.

Effects on Income by Economic Background

Next, we show how we approximate value-added changes for each higher education and background pair change ($VA^b_{jj'}$), and how we approximate the elasticities that build up the composition effect.

In section 5, we detailed how wage measures change across groups of colleges and economic backgrounds. We show that on average, students from all economic backgrounds who go to elite colleges have much better future outcomes than those who go to lower-ranked institutions or no higher education. However, once we control for ENEM grades, we find that those differences reduce substantially. Across all economic backgrounds, the difference between elite college and the lower ranked college students goes from 45 percentiles in the wage rank to less than 10 percentiles. This indicates that the vast majority of differences in future outcomes across college ranks come from selection into better colleges than the actual value added of the institutions.

To approximate changes in value added by college tier for each economic background, we first run a linear regression of income measures on college fixed effects[24] and granular ENEM score fixed effects[25]. Controlling by ENEM fixed effects allows us to adjust for the selection of good students into good colleges.

We recover the fixed effects by college option and average them across the six higher education groups (5 college tiers and no college) for students from advantaged and disadvantaged backgrounds. We then compute the value-added change of going from tier j to tier j' as the difference between the average of college fixed effects in tier j and college fixed effects in tier j'.[26]

To estimate elasticities of composition responses to policies we need to make assumptions on the counterfactual options of students who are affected by the policies.

[24] The sample includes students who do not attend college, who are grouped as one option in the college fixed effects.

[25] We divide the sample in bins of 20 points of ENEM score

[26] In our current exercise, we are using individuals' wage rank as the income measure. This is imprecise as we show in Appendix A that changes in income rank and changes in income do not necessarily match. We are currently working on an adequate measure of income changes using actual wages, and then recomputing changes in average rank by economic background.

We outline two different scenarios for composition elasticities, based on our estimates of changes in student composition caused by affirmative action and subsidized loans. In both scenarios, we assume a constant number of seats in each college tier.

In our first scenario, we assume that students who reach higher-ranked colleges come from their second-best option, and *move up the ladder* of college rank. In our framework, this assumption implies that $\epsilon^b_{jj'} = 0$ if $j' - j \neq 1$. Thus, students who because of policies are going to elite colleges (96-100 pct of college rank) would have gone to the previous tier (91-95) in the absence of the policy.

An example is useful to illustrate our method. We estimate that Affirmative action increased by 0.7% the share of disadvantaged students in elite colleges and 0.35% in colleges from tier 4[27]. We then multiply 0.7% by the number of students in elite colleges, to arrive at ϵ^{dis}_{45}, the number of students going from tier 4 to tier 5 from disadvantaged backgrounds. To arrive at ϵ^{dis}_{34} we first observe that ϵ^{dis}_{45} students left tier 4 to tier 5, but tier 4 still observed an increase of 0.35% in the share of disadvantaged students. Thus, ϵ^{dis}_{34} must account for the students who moved up from 4 to 5. We first multiply 0.35% by the number of students in college tier 4, and then add ϵ^{dis}_{45} to arrive at ϵ^{dis}_{34}. [28]

We execute the same simulation for the effects of subsidized loans, replacing the estimated coefficients and the share of AA, with their corresponding values for subsidized loans.

We show our estimates of changes in income rank from this simulation in Panel (a) of Figure 9. We observe that both policies have relatively small effects on average outcomes of disadvantaged students. The average future rank changes by 0.09 due

[27] We arrive at 0.7% by multiplying our coefficient 0.0454 with the actual change from 9% of seats reserved in 2010 to 0.26% of seats reserved in 2015. Equivalently for college tier 4.

[28] We can formalize this. Define β_j the coefficient estimated in regression 7 for a given tier j, N_j the number of seats in a given college tier j, and $ShAA_{j,t}$ the share of reserved seats in tier j at year t. We can write :

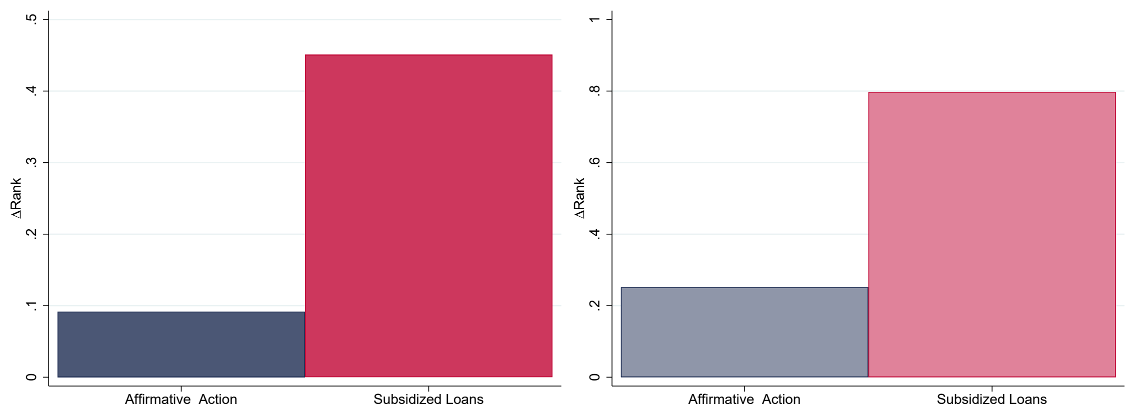
$$\epsilon^b_{jj'} = \beta_{j'} * (ShAA_{j',2015} - ShAA_{j',2010}) * N_{j'}) + \epsilon_{j'j''} \quad if \quad j'' = j' + 1, j' = j + 1, j \in \{0, 1, 2, 3, 4\}$$

and $\epsilon^b_{45} = \beta_5 * (ShAA_{5,2015} - ShAA_{5,2010}) * N_5)$.

to Affirmative Action and 0.44 with subsidized loans. Subsidized loans have a larger overall effect because the number of students benefited is larger, despite having a smaller effect in moving students to elite colleges.

In a second simulation, we assume that instead of moving up the ladder, all students come from the option *no college*. This implies that $\epsilon^b_{jj'} = 0$ if $j \neq 0$. We show our estimates in Panel (b) of Figure 9. We observe that increases in income are two times bigger than in our previous simulation. Subsidized loans still have a larger effect than affirmative action. The difference between both simulations highlights the role of counterfactual options in estimating the effects of such policies.

Figure 9: Affirmative Action vs. Financial Aid (FIES)



(a) *Moving up the Ladder*　　**(b)** No college

This Figure shows Equation (6) calibrated for the affirmative action policy implemented in Brazil and FIES. Following Chetty et al. (2020), we use the difference between a residualized version (by ENEM grades) of our outcome variable to recover the value added of college tier j relative to j' ($VA_{jj'}$) for group b. We allocate all additional (due to the policy) disadvantaged students in a given year, to the previous cohort of high school graduates. Therefore, we allocate all additional disadvantaged students in 2011-2013 to our main sample (high school graduates 2010-2012). Panel (a) compares the effect of AA and FIES, taking into consideration the real elasticities estimated in Tables (3) and (4). The assumption is that all compliers moving to a better college tier come from the best outside option available (one lower college tier). Panel (b) does the same, but it sets the worst outside option, meaning they all come from *no college*, maximizing the college effects.

7 Conclusion

Our research uncovers a pronounced pattern of income segregation within Brazilian higher education, where students from affluent backgrounds are significantly more represented in elite colleges, despite a system that ostensibly does not consider economic status in admissions. This reflects deep-rooted social divides that merit-based entrance exams alone cannot address. Our findings further indicate that these wealthy students gain access to top institutions and earn substantially more than their peers from lower-income backgrounds. A considerable portion of this earning disparity can be traced back to the segregation evident in the higher educational system, highlighting the role of institutional inequality in perpetuating social and economic divides.

This income-specific segregation and the resultant earnings gap raise questions about the efficacy of existing policies in promoting genuine intergenerational mobility. By developing a generalized conceptual framework, we've provided a tool to assess how different educational interventions might pave the way for increased social mobility. We use this general framework to evaluate affirmative action in public colleges and subsidized loans for private institutions. Both policies increased the mobility of low-income students, but subsidized loans have a larger effect. While AA is able to increase the representation of disadvantaged students in elite schools and subsidized loans are not, the latter policy reallocates a larger number of students to better college tiers overall.

For policymakers, these findings underscore the need to carefully consider pre-existing inequalities when assessing the role of higher education in social mobility. Policies need to account for the fact that the higher education system does not operate in isolation, but rather selects from a pool of candidates already shaped by significant socio-economic disparities. Our work suggests that without a thorough understanding and addressing of these pre-existing differences, any educational reform aimed at promoting equality may be less effective.

Bibliography

Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou, “Intergenerational mobility in Africa,” *Econometrica*, 2021, *89* (1), 1–35.

Asher, Sam, Paul Novosad, and Charlie Rafkin, “Intergenerational mobility in India: Estimates from new methods and administrative data,” *World Bank Working Paper*, 2018.

Bautista, Maria Angelica, Felipe González, Luis R Martinez, Pablo Munoz, and Mounu Prem, “Dictatorship, higher education and social mobility,” *Higher Education and Social Mobility (September 12, 2023). University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2023, (2023-121).

Black, Sandra E, Jeffrey T Denning, and Jesse Rothstein, “Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes,” *American Economic Journal: Applied Economics*, 2023, *15* (1), 26–67.

Bleemer, Zachary, “Top Percent Policies and the Return to Postsecondary Selectivity, by Zachary Bleemer, CSHE 1.21,” 2021.

_ , “Affirmative action, mismatch, and economic mobility after California’s Proposition 209,” *The Quarterly Journal of Economics*, 2022, *137* (1), 115–160.

Britto, Diogo GC, Alexandre de Andrade Fonseca, Paolo Pinotti, Breno Sampaio, and Lucas Warwar, “Intergenerational mobility in the land of inequality,” 2022.

Bucarey, Alonso, Dante Contreras, and Pablo Muñoz, “Labor market returns to student loans for university: Evidence from Chile,” *Journal of Labor Economics*, 2020, *38* (4), 959–1007.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, “Income segregation and intergenerational mobility across colleges in the United States,” *The Quarterly Journal of Economics*, 2020, *135* (3), 1567–1633.

_ , Nathaniel Hendren, and Lawrence F Katz, “The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment,” *American Economic Review*, 2016, *106* (4), 855–902.

_ , _ , Patrick Kline, and Emmanuel Saez, “Where is the land of opportunity? The geography of intergenerational mobility in the United States,” *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Dale, Stacy B and Alan B Krueger, "Estimating the effects of college characteristics over the career using administrative earnings data," *Journal of human resources*, 2014, *49* (2), 323–358.

Dale, Stacy Berg and Alan B Krueger, "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables," *The Quarterly Journal of Economics*, 2002, *117* (4), 1491–1527.

Dobbin, Cauê, Nano Barahona, and Sebastián Otero, "The Equilibrium Effects of Subsidized Student Loans," Technical Report, Working paper, Stanford University 2021.

Duryea, Suzanne, Luísa Baptista de Freitas, Luana Marques-Garcia Ozemela, Breno Sampaio, Gustavo R Sampaio, and Giuseppe Trevisan, "Universities and intergenerational social mobility in Brazil: examining patterns by race and gender," *Journal of Economics, Race, and Policy*, 2019, *2*, 240–256.

Estevan, Fernanda, Thomas Gall, and Louis-Philippe Morin, "Redistribution without distortion: Evidence from an affirmative action programme at a large brazilian university," *The Economic Journal*, 2019, *129* (619), 1182–1220.

Iannelli, Cristina and Lindsay Paterson, *Does education promote social mobility?*, Vol. 35, Centre for Educational Sociology, 2005.

Kearney, Melissa S and Phillip B Levine, "Income inequality, social mobility, and the decision to drop out of high school," Technical Report, National Bureau of Economic Research 2014.

Kenedi, Gustave and Louis Sirugue, "Intergenerational income mobility in France: A comparative and geographic analysis," *Journal of Public Economics*, 2023, *226*, 104974.

Londoño-Vélez, Juliana, Catherine Rodriguez, Fabio Sánchez, and Luis E Álvarez-Arango, "Financial Aid and Social Mobility: Evidence from Colombia's Ser Pilo Paga," Technical Report, National Bureau of Economic Research 2023.

Machado, Cecilia and Christiane Szerman, "Centralized college admissions and student composition," *Economics of Education Review*, 2021, *85*, 102184.

Mahlmeister, Rodrigo, Sergio Guimarães Ferreira, Fernando Veloso, Naercio Menezes-Filho, and Bruno Kawaoka Komatsu, "Revisitando a mobilidade intergeracional de educação no Brasil," *Revista Brasileira de Economia*, 2019, *73*, 159–180.

Mello, Ursula, "Affirmative action and the choice of schools," *Journal of Public Economics*, 2023, *219*, 104824.

Mountjoy, Jack, "Community colleges and upward mobility," *American Economic Review*, 2022, *112* (8), 2580–2630.

Neelsen, John P, "Education and social mobility," *Comparative Education Review*, 1975, *19* (1), 129–143.

Otero, Sebastián, Nano Barahona, and Cauê Dobbin, "Affirmative action in centralized college admission systems: Evidence from Brazil," *Unpublished manuscript*, 2021.

Senkevics, Adriano Souza and Ursula Mattioli Mello, "O perfil discente das universidades federais mudou pós-Lei de Cotas?," *Cadernos de Pesquisa*, 2019, *49* (172), 184–208.

Zimmerman, Seth D, "Elite colleges and upward mobility to top jobs and top incomes," *American Economic Review*, 2019, *109* (1), 1–47.

A Deriving Aggregate Statistics from Individual Level Model

We start by defining our outcome of interest for individual i, i.e., how her income responds to a given policy:

$$\Delta \text{Income}_i = \text{I}_{ip} - \text{I}_{i0} \tag{A.9}$$

where I_{ip} and I_{i0} are the potential outcomes of individuals with and without the policy p.

To understand the role of college on individuals' future outcomes, it is useful to decompose potential outcomes to account for college attendance.

We define $d_{j(i)p}$ as an indicator variable that individual i's higher education option in the presence of policy p was j, where $j \in \mathcal{J} = \{\text{No College}, 1, ...N\}$ where 1 to N indexes all the college options available. Furthermore, we define $\gamma_{ij(i)}$ as the potential income obtained by individual i in case they opt for higher education option j. We can then, rewrite the potential outcomes as:

$$\text{I}_{ip} = \sum_{j \in \mathcal{J}} d_{j(i)p} \cdot \gamma_{ij(i)} \qquad \text{I}_{i0} = \sum_{j \in \mathcal{J}} d_{j(i)0} \cdot \gamma_{ij(i)} \tag{A.10}$$

Therefore, to understand the effects of policies, we need to understand both the changes in the higher education decision of individuals' ($\{d_{j(i)p}$ and $d_{j(i)0}\}$), and also recover the value added of the colleges for individual i, $\gamma_{ij(i)}$[29]

To understand the general patterns of social mobility, we do not need to calculate ΔIncome_i for each individual i. We aggregate from the individual level to a given population from economic background b. We can then define the average change in income for those from background b caused by the policy as:

$$\Delta \text{Income}_b = \underbrace{\frac{1}{N_b} \sum_{i \in b} \sum_{j \in \mathcal{J}} d_{j(i)p} \cdot \gamma_{ij(i)}}_{\text{Avg. Income with Policy}} - \underbrace{\frac{1}{N_b} \sum_{i \in b} \sum_{j \in \mathcal{J}} d_{j(i)0} \cdot \gamma_{ij(i)}}_{\text{Avg. Income w/o Policy}} \tag{A.11}$$

where N_b is equal to the sum of individuals i from background b. Equation A.11 simply takes the average income across individuals from the same background. We can rewrite the average income change for economic background b as:

$$\Delta \text{Income}_b = \frac{1}{N_b} \sum_{j \in \mathcal{J}} \sum_{i \in b} \gamma_{ij(i)} \cdot (d_{j(i)p} - d_{j(i)0})$$

The expression above shows that under this framework, all variation in average income for a given background comes from those who change their college decision

[29] We make the simplifying assumption that returns to a given college is the same for each individual independently of the policy. A more sophisticated framework could let $\gamma_{ij(i)}$ be also a function of p, which could then incorporate mechanisms such as peer effects.

because of the policy.

In order to aggregate from the individual level to the economic background level, it is useful to slightly modify our notation. Define $j(i)$ as the higher education option in the absence of the policy, and $j'(i)$ individual i's choice in its presence.

At each economic background b, we define the changes in the number of students going to a given higher education option j from option j' as:

$$\epsilon_{bj}^{j'} = \sum_{i \in b} d_{j(i)} \cdot d_{j'(i)}$$

Furthermore, we can define the average value-added change for an individual in economic background b that has higher education option j in the absence of the policy and higher education option j' in its presence:

$$VA_{bj}^{j'} = \frac{1}{\sum_{i \in b} d_{j(i)} \cdot d_{j'(i)}} \sum_{i \in b} d_{j(i)} \cdot d_{j'(i)} \left(\gamma_{ij'(i)} - \gamma_{ij(i)} \right)$$

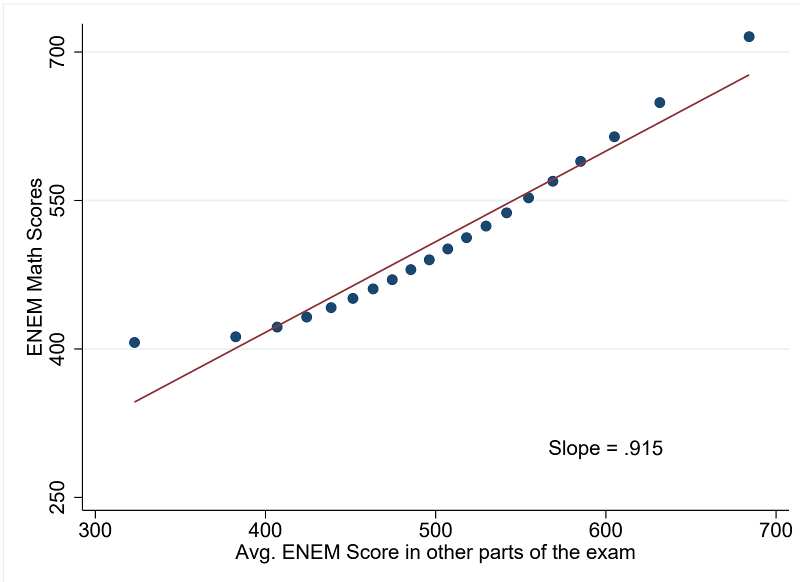
We can then rewrite the variation in income caused by the policy as:

$$\Delta \text{Income}_b = \sum_{j \in \mathcal{J}} \sum_{j' \in \mathcal{J}} \underbrace{\frac{1}{N_b} \epsilon_{bj}^{j'}}_{\text{Composition Effect}} \cdot \underbrace{VA_{bj}^{j'}}_{\text{College Effect}}$$

It is noteworthy that with no other assumption to our framework, there is no sufficient statistic to analyze the average rank change for a group of background b. To aggregate average rank, it is necessary to calculate rank changes for all individuals in the sample, and then calculate the average changes for each group b. This happens because the rank of individuals who do not change their college decision due to the policy could still be potentially affected by the policy because of changes in the *compliers'* income.

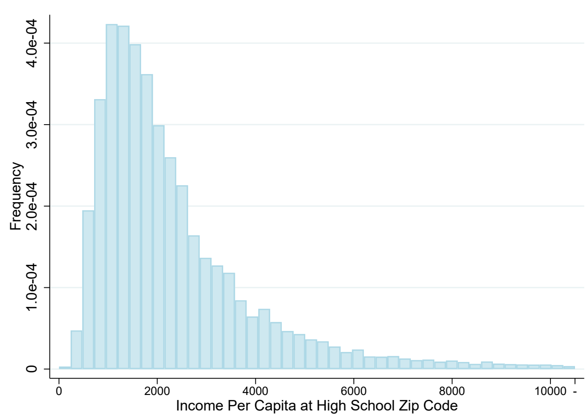
B Appendix Figures

Figure A1: Correlation Between ENEM Math Scores and Scores in other exams.



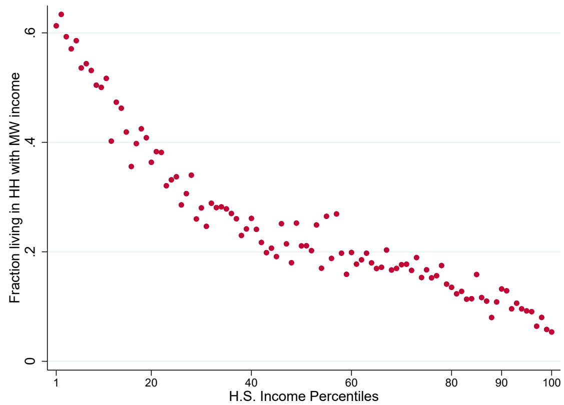
This Figure shows the correlation between math grades and average grades in ENEM exam. Sample comprises of all students graduating high school and taking ENEM between 2010 and 2012. Bins have the same equal numbers of individuals. Slope is the coefficient β estimated in a regression Math Scores$_i = \alpha + \beta$Avg. Other Scores$_i + \varepsilon_i$

Figure A2: Histogram of High School Income Distribution

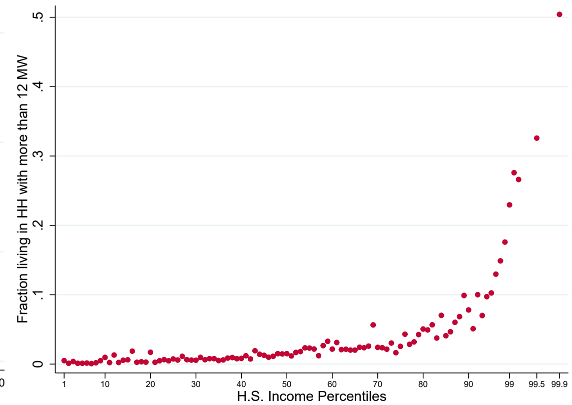


This Figure shows a histogram of the income per capita levels across colleges. The sample comprises of all high schools in *Censo Escolar* (2009-2013).

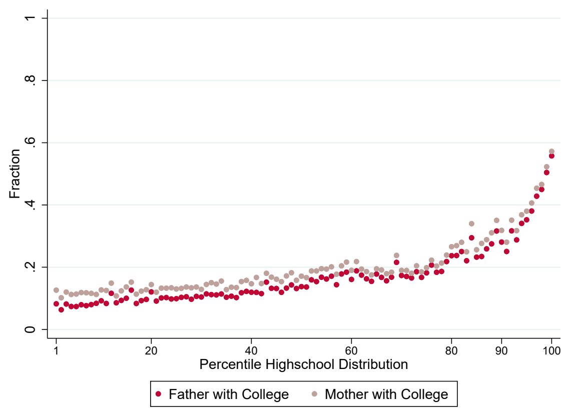
Figure A3: Validation of Students' Economic Background Measure



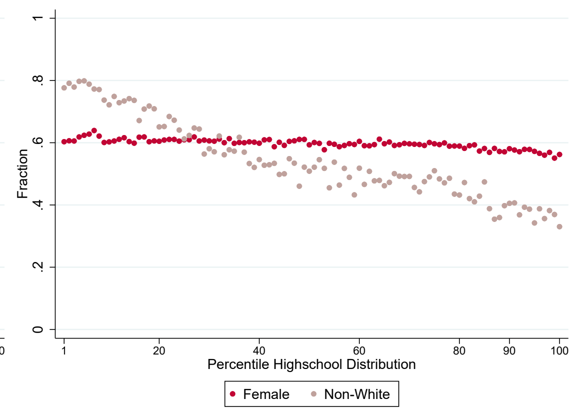
(a) Household < M.W.



(b) Household > 12 M.W.



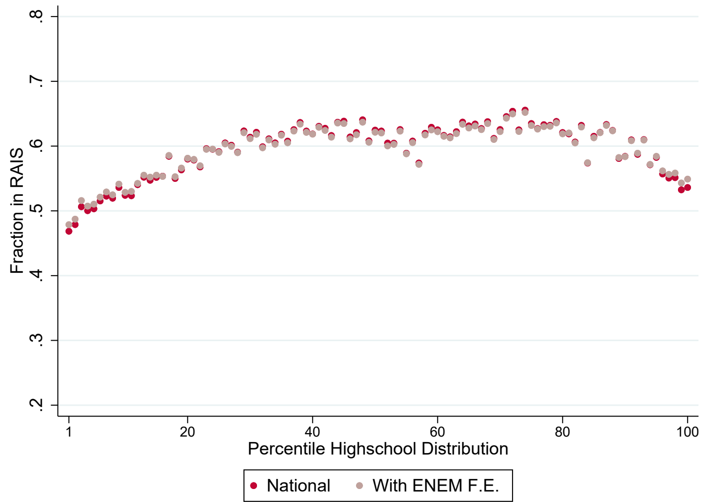
(c) Parents with College



(d) Gender and Race

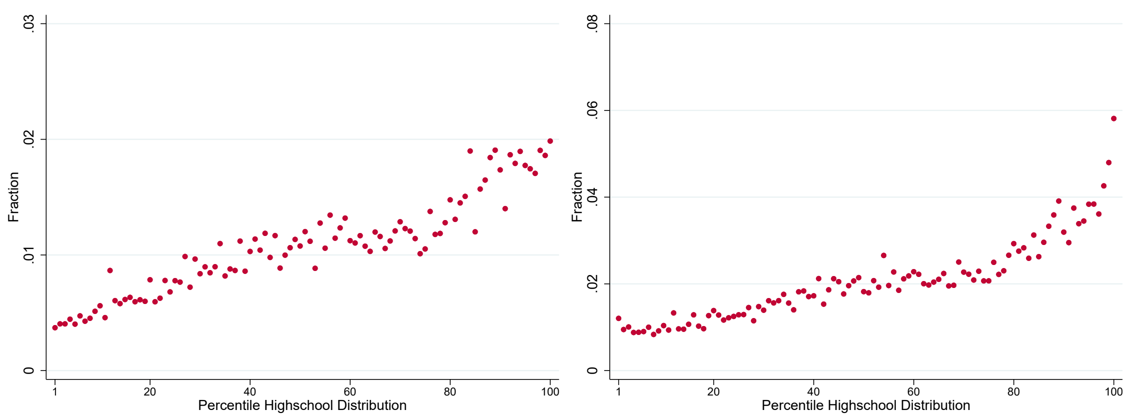
This Figure shows how the economic background measure correlates with self-reported individual-level characteristics. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012. Panel (a) shows the share of individuals who report living in a household with less than 1 minimum wage of household income. Panel (b) shows the share of individuals living in households with more than 12 minimum wages of household income. Panel (c) shows the share of parents with college. Panel (d) shows the share of women and share of individuals who self-report as nonwhite.

Figure A4: Probability of Being Formally Employed



This Figure shows the fraction of students from each percentile of the economic background distribution who appear in the matched employer-employee nine years after high school graduation.

Figure A5: Alternative Labor Market Activities



(a) Firm owner (with employees) **(b)** Firm owner (no employees)

This Figure shows the fraction of students from each percentile of the economic background distribution who perform activities in the formal labor market differently from being employees. Panel (a) shows the fraction that becomes owners of firms with employees, nine years after high school graduation. Panel (b) shows the fraction of students owning active firms with no employees, nine years after high school graduation. We separate both types of ownership activity because, in Brazil, the second one is associated to a tax evasion practice known as *pejotizacao*. This practice consists of employees who, rather than being registered in the payroll as employees, they open a firm to receive their wages as revenues, taking advantage of more generous tax rates.

Figure A6: Distribution of Wages and Earnings Measure across Economic Background

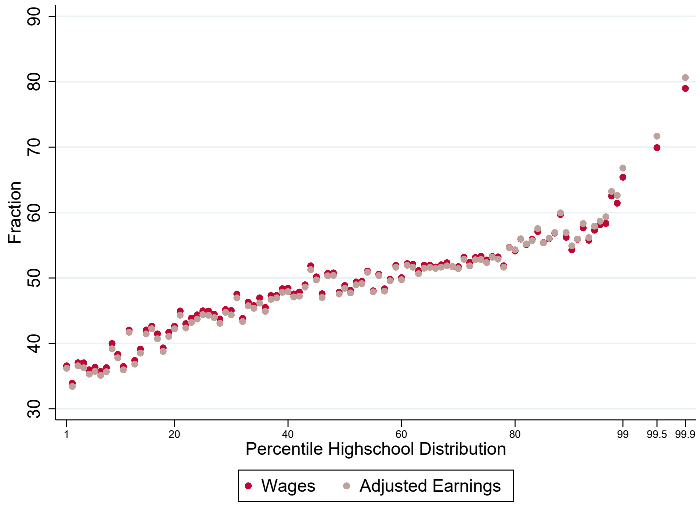
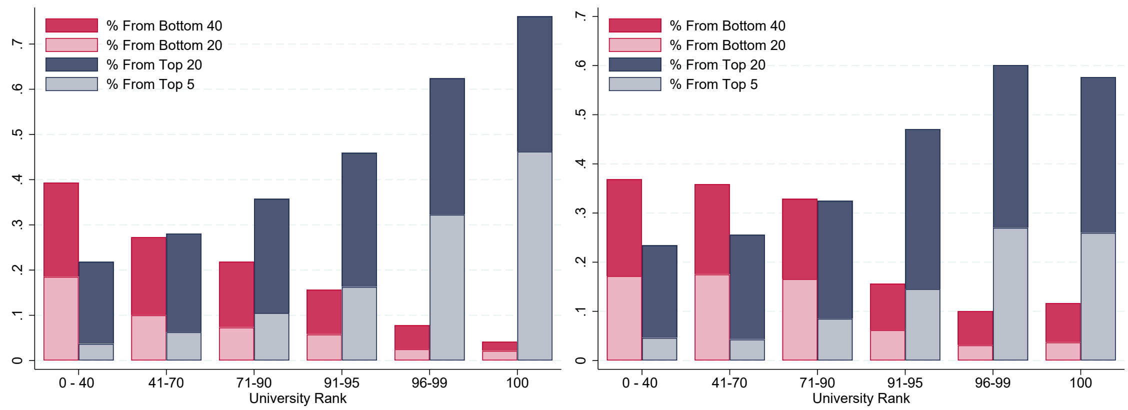


Table A1: Attendance Across Economic Background Distribution by Types of College and Majors

	Percentiles of High School Income Distribution						
	1-25	26-50	51-75	75-90	91-94	95-98	99th
Administration Type							
Attended College	51.34	65.16	70.31	77.92	82.08	86.42	90.05
	58.43	67.88	70.66	74.25	76.00	76.23	75.91
Private College	36.89	50.36	54.11	56.30	56.49	56.26	59.57
	38.11	49.96	53.42	55.74	55.97	57.43	63.27
Public College	14.45	14.80	16.20	21.62	25.59	30.16	30.49
	20.31	17.92	17.24	18.50	20.03	18.80	12.64
Type of Degree							
Licenciatura	23.14	16.73	13.71	11.48	10.13	8.18	5.35
	21.00	15.51	13.28	12.26	11.61	11.17	9.91
Bacharelado	64.61	69.52	71.32	75.69	78.99	82.55	86.83
	67.94	71.48	72.07	74.47	76.63	77.72	79.47
Técnico	11.84	13.15	14.07	11.74	9.77	7.86	6.05
	10.39	12.24	13.67	12.28	10.83	10.16	9.68
Majors							
Architecture	1.31	1.93	2.34	2.72	3.06	3.35	3.63
	1.57	2.05	2.37	2.62	2.89	3.06	3.27
Accounting	3.46	3.91	3.61	3.14	3.03	2.39	1.73
	3.40	3.80	3.54	3.17	3.11	2.69	2.31
Computer Sciences	3.52	4.16	4.54	4.44	4.09	3.64	3.02
	4.00	4.37	4.57	4.26	3.79	3.18	2.49
Economics	0.53	0.50	0.61	0.78	0.87	1.25	2.26
	0.74	0.64	0.67	0.70	0.71	0.91	1.74
Engineering	9.30	12.49	13.76	15.70	16.39	17.67	19.06
	12.74	14.68	14.69	14.45	13.96	12.18	10.16
Law	6.73	8.25	8.46	9.67	11.00	12.56	13.49
	6.96	8.33	8.46	9.58	10.86	12.39	13.38
Medicine	0.53	0.78	1.16	2.32	2.99	4.67	6.79
	1.49	1.50	1.52	1.97	2.26	2.81	3.57
Psychology	2.49	2.62	2.66	2.65	2.57	2.64	2.64
	2.22	2.45	2.60	2.75	2.76	3.04	3.26
Num. of Individuals	620,372	723,202	726,211	652,833	187,248	243,876	45,188

Notes: This table shows average college attendance by groups of economic background. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012. If an individual appears more than one institution, we select the first observation after they graduate high school.

Figure A7: Attendance at Public and Private Institutions by High School Income Rank

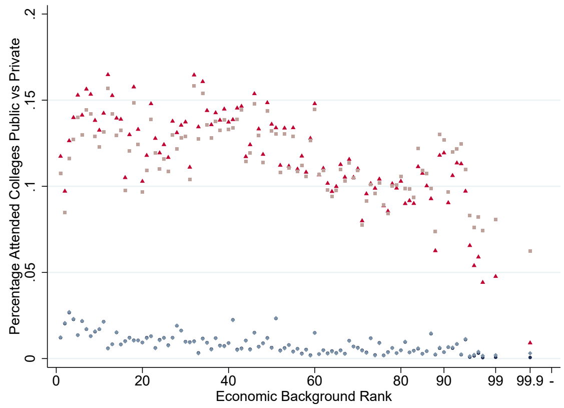


(a) Private Colleges

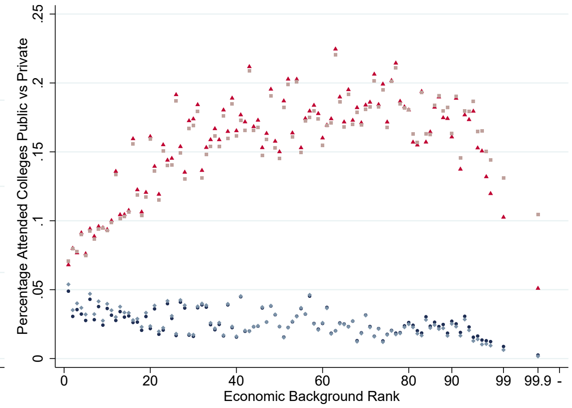
(b) Public Colleges

This Figure shows the income composition of entrants across different types of colleges. The sample comprises of all entrants between 2010 and 2015 for whom we have information in ENEM of where they graduated from in high school. The remaining observations of entrants are treated as missing. Colleges are ranked by average earnings of graduates between 2010 and 2013.

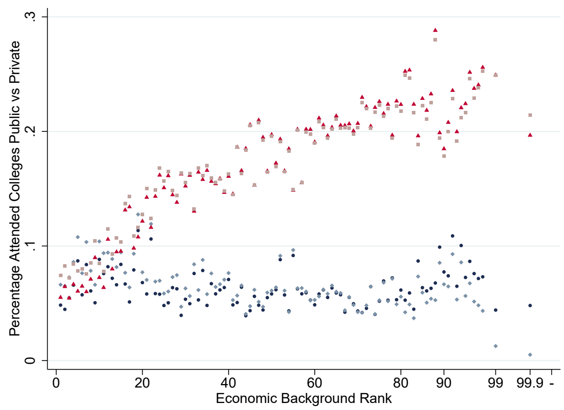
Figure A8: Public and Private College Attendance Across Income Distribution by College Rank



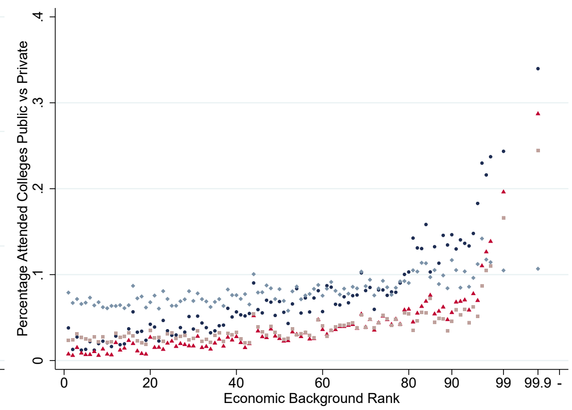
(a) Bottom 40 Percentiles of Colleges



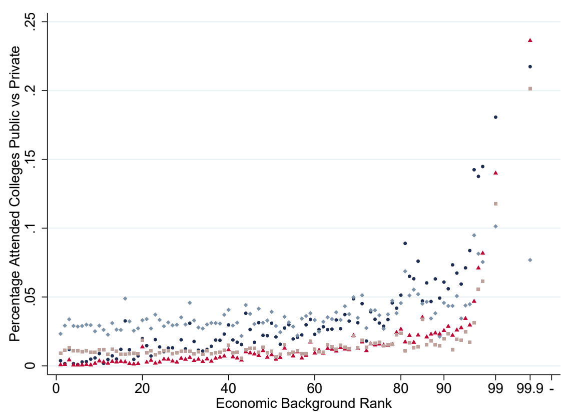
(b) 41st-70th Percentiles of Colleges



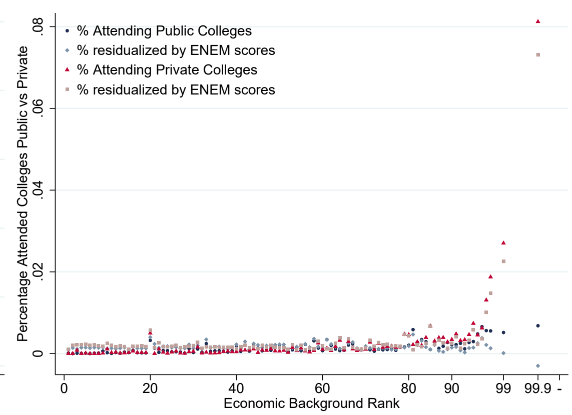
(c) 71th-90th Percentiles of Colleges



(d) 91st-100th Percentiles of Colleges



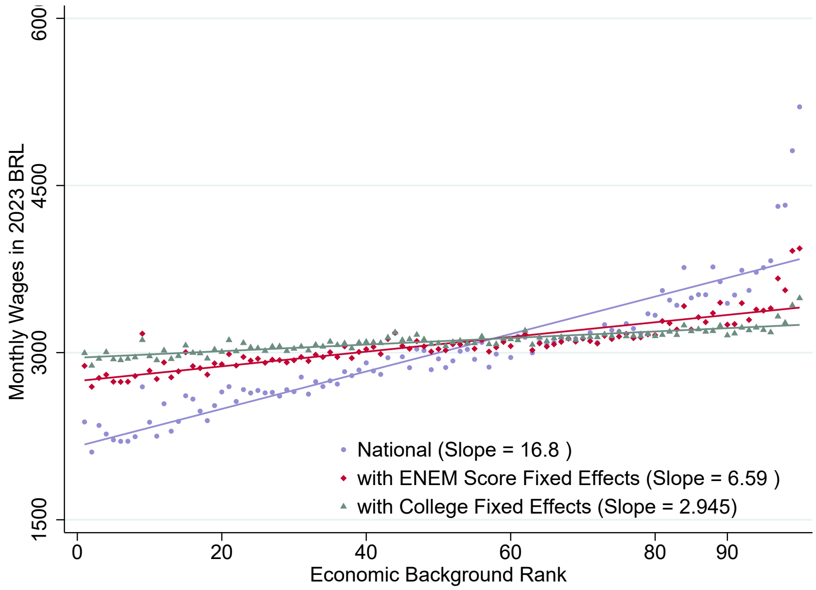
(e) 96th-100th Percentiles of Colleges



(f) Top 1% of Colleges

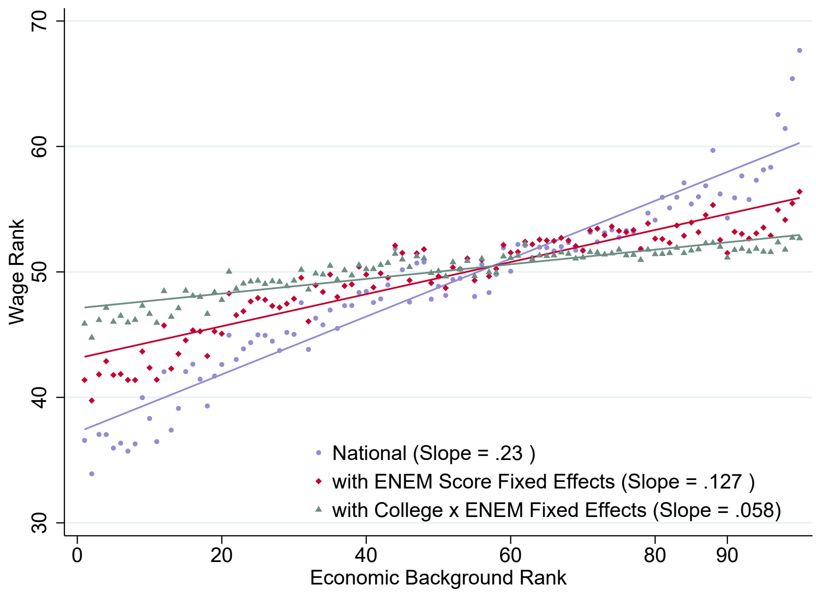
This Figure shows elite private and public college attendance vary across the High School Income Rank distribution. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012. Each bin represents individuals at the respective percentile of the distribution. College attendance is defined as 1 if the individual appeared at least once in CESUP in the 7 years after they graduated High School and 0 otherwise. If an individual appears both in private and public universities, we select the first observation after they graduate high school. Red dots are the average within bins of residuals from a linear regression of College Attendance on Fixed Effects of 5 points of math score in ENEM, summed with the average college attendance by administration type in the sample.

Figure A9: Relationship Between Students' Monthly Wages and High School Income Rank



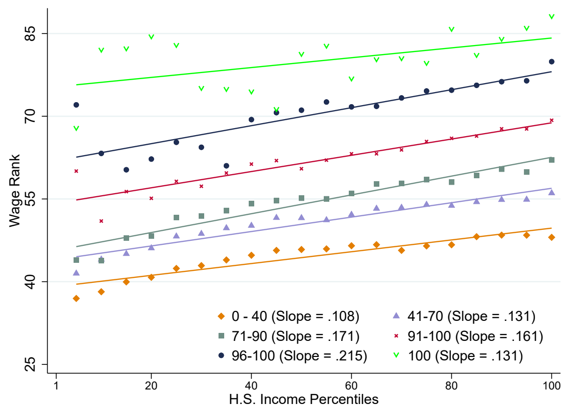
This Figure shows how wages 9 years after high school graduation varies across the High School Income Rank distribution. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012 who are found employed in RAIS 9 years after graduating. Each dot represents the averages of individuals at the respective percentile of the high school income rank. The red and green markers show the average wage rank in the sample summed with residuals from a linear regression of wage rank on College F.E. and College interacted with bins of 5 points in ENEM math score respectively. Individuals who do not attend college are grouped into one single category. For individuals with multiple colleges in CESUP, we select the first one they appear after graduating high school.

Figure A10: Relationship Between Students' Wage Rank and Economic Background with ENEM scores fixed Effects not Interacted

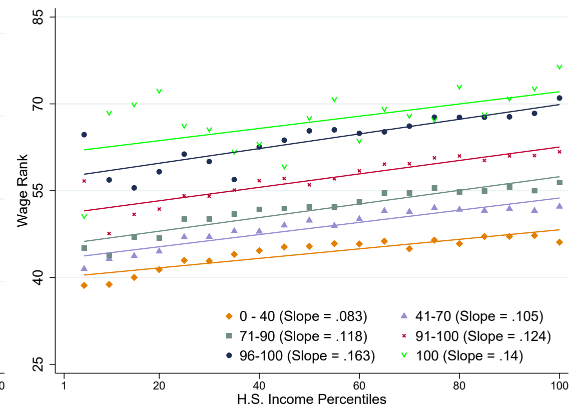


This Figure shows how Wage rank 9 years after high school graduation varies across the High School Income Rank distribution. The sample comprises of all individuals graduating high school and taking ENEM between 2010 and 2012 who are found employed in RAIS 9 years after graduating. Each dot represents the averages of individuals at the respective percentile of the high school income rank. The red and green markers show the average wage rank in the sample summed with residuals from a linear regression of wage rank on fixed effects of 20 points bins in ENEM math score and College interacted with bins of ENEM score respectively. Individuals who do not attend college are grouped into one single category. For individuals with multiple colleges in CESUP, we select the first one they appear after graduating high school.

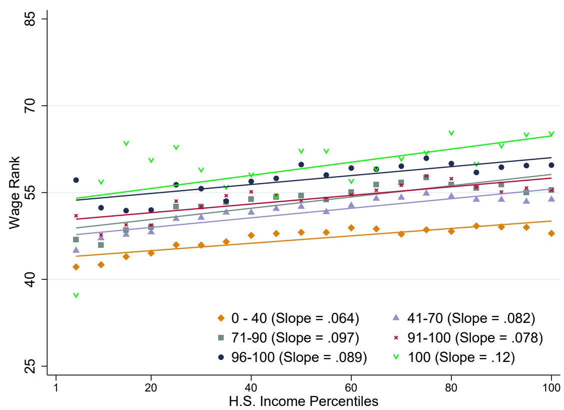
Figure A11: Relationship Between Wage Rank and H.S. Rank Across Different Colleges with Top Percentile



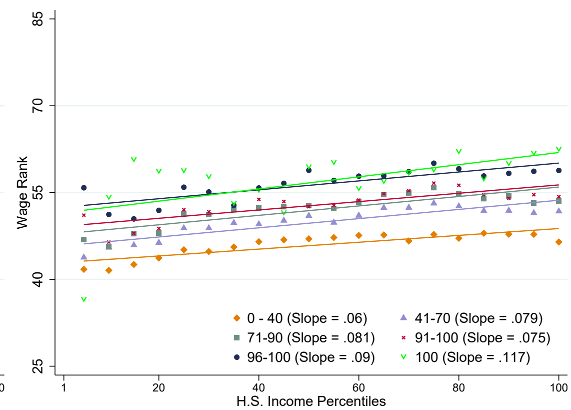
(a) National Averages



(b) Controlling by College x Major F.E..



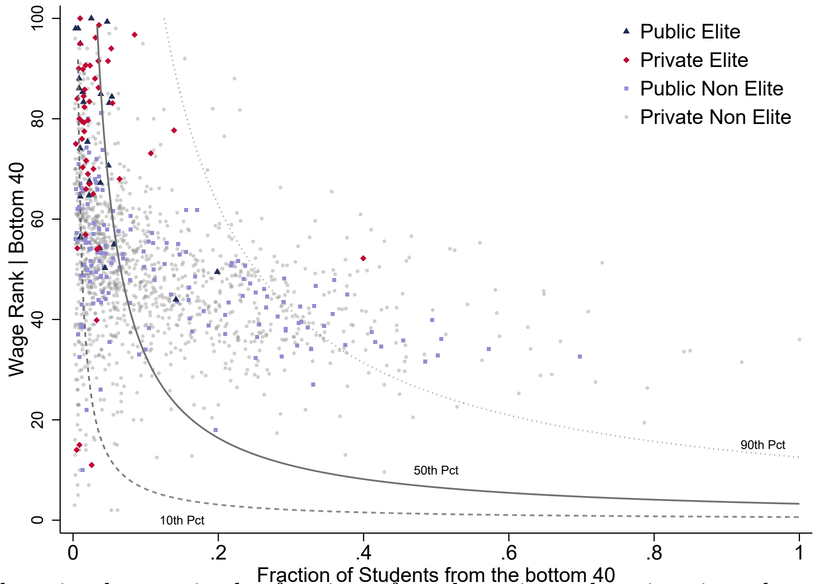
(c) Controlling by ENEM F.E.



(d) Contr. College x Major x ENEM F.E.

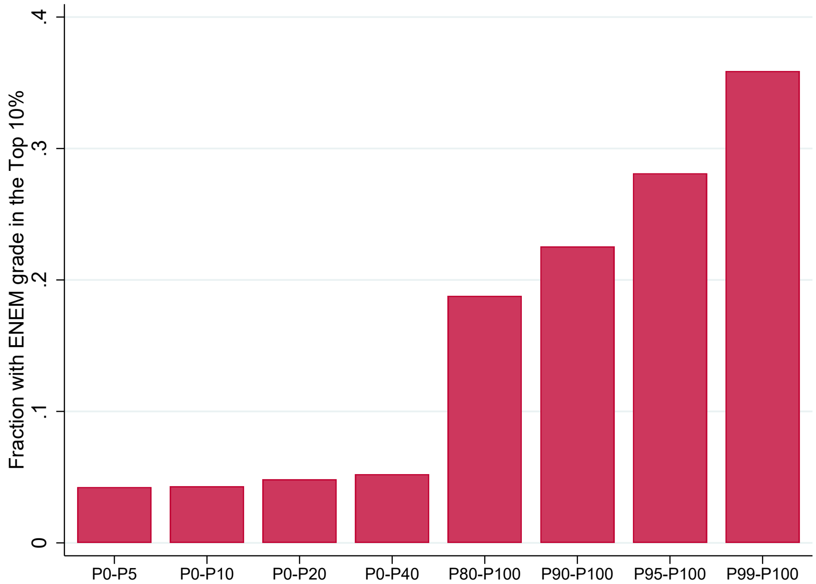
This Figure shows the income composition of entrants across different types of colleges. The sample comprises of all entrants between 2010 and 2015 for whom we have information in ENEM of where they graduated from in high school. The remaining observations of entrants are treated as missing. In Panel (b) colleges are ranked by average earnings of graduates between 2010 and 2013.

Figure A12: Social Mobility versus Income Segregation - Bottom 20%



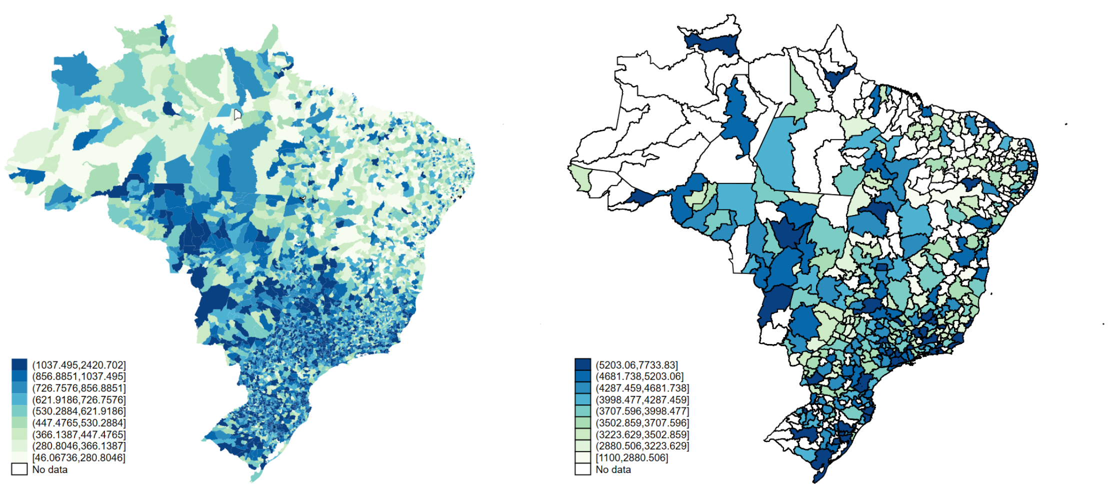
This Figure show in the x-axis the fraction of students in each university who comes from the bottom 40 of the H.S. income rank. In the y-axis we show the average wage rank at the age of 30 that students from the bottom 40 of the H.S. income rank reach. It shows a very clear pattern highlighting that universities with lower fraction of students from the bottom 40 are the same that place these disadvantaged students at the top of the income distribution. Among those, elite schools (both private and public) are clearly highly segregated and with high returns for low income students.

Figure A13: Relationship between income and ENEM performance



This Figure shows the fraction of students performing in the top 10% of the math grade distribution by HS rank.

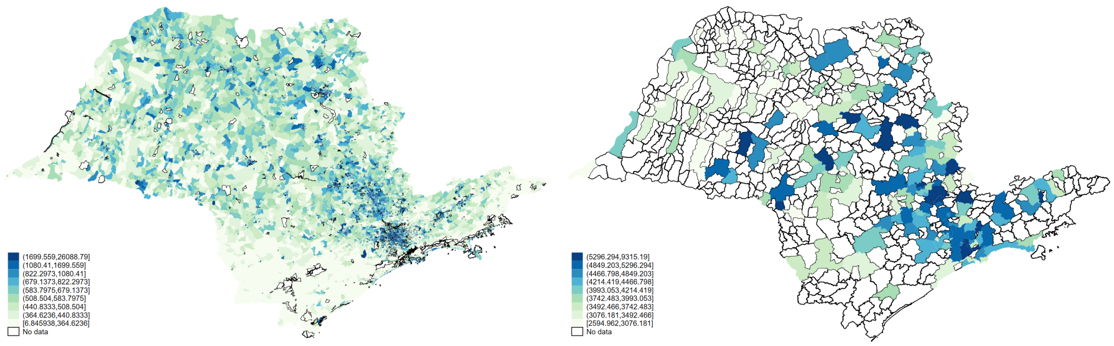
Figure A14: Geographic Distribution of Economic Background and College Graduates' Wages



(a) Economic Bacgkround

(b) Graduates' Wages

Figure A15: Geographic Distribution of Economic Background and College Graduates' Wages - S ao Paulo

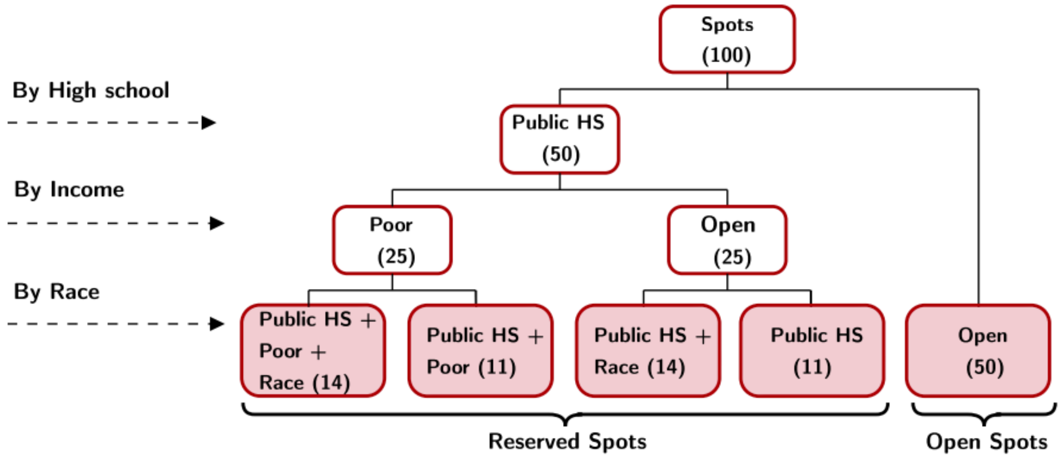


(a) Economic Bacgkround

(b) Graduates' Wages

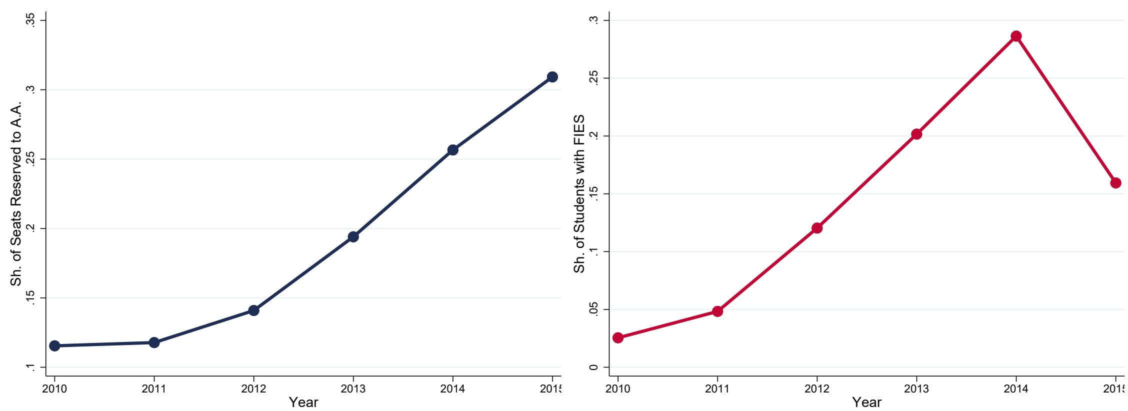
Location and Colleges: Location has been documented as an important determinant in intergenerational mobility (Chetty et al., 2014; Kenedi and Sirugue, 2023). It is no different in Brazil, as documented by GC Britto et al. (2022). This interacts with our analysis, as places with lower economic conditions also have lower labor market perspectives. In Figure A14, we show the geographical distribution of the economic background measure and the average wage of college graduates. The high schools and colleges are aggregated at the microregion level, often argued as equivalent to commuting zones in the United States. We see a high correlation between places with a better economic condition and higher wages for college graduates. Since a large share of students stay in the same geographical area when going to college, (Machado and Szerman (2021) shows that 85% of students in public schools stay in the same state and 45% in the same municipality for college), we interpret this as suggestive evidence that location plays an important role in explaining the relationship between colleges and intergenerational mobility.

Figure A16: Affirmative Action Regulation



Source: Otero et al. (2021)

Figure A17: AA and Disadvantaged Students' Outcomes



(a) Sh. of Students admitted through A.A.

(b) Sh. of Students with FIES

This figure shows the share of students admitted through affirmative action in Panel (a) and the share of students with FIES in Panel (b). The sample comprises all programs in Public colleges that have at least one entrant in all years between 2010 and 2015.