College Quality and Tuition Subsidies in Equilibrium*

Fernando Cordeiro† Alvaro Cox‡

December 5, 2023

Click here for the latest version.

Abstract

In this paper, we study how demand-side subsidies interact with the equilibrium level of price and quality in the higher education sector in Brazil. More precisely, we consider two policies that assist low-income students in attending private institutions: scholarships and subsidized student loans. First, we develop a quality measure for undergraduate programs using value-added models, where a student's post-graduation outcome is determined by their pre-enrollment characteristics. To do so, we link multiple administrative datasets to track individual students before enrollment, during college, and after college. We consider two post-graduation outcomes: a standardized "exit" exam, which tests students' major-specific knowledge, and income from a matched employer-employee database. We document key patterns and correlations of our quality measure and extensively validate it. Next, we develop a static equilibrium model of demand, pricing, and quality provision. We consider two counterfactuals: decreasing the supply of loans by 10% and decreasing the supply of scholarships by 10%. We find similar patterns under both scenarios but much stronger effects for scholarships. Specifically, by lowering the supply of scholarships by 10% in each program across the country, there would be 13 thousand fewer students in college, corresponding to 80% of the total reduction in scholarships. Most programs would have an incentive to decrease quality and price, with a median change in value-added of -5% and of -0.7% in price. Crucially, our approach uncovers significant heterogeneity across programs, particularly in terms of student composition and the effects of these policy changes, offering valuable insights into the nuanced impacts of educational subsidies.

*We are very grateful to Katja Seim, Steve Berry, and Christopher Neilson for their invaluable comments and support on this project. We also thank Claudia Allende, Phil Haile, Charles Hodgson, Jonas Leiber, and seminar participants at Yale University for their comments and suggestions. We gratefully acknowledge financial support from the Yale Economic Growth Center. We are grateful for the technical assistance from the SEDAP team at INEP.

†Department of Economics, Yale University. Email: fernando.cordeiro@yale.edu

‡Department of Economics, Yale University. Email: alvaro.cox@yale.edu

1 Introduction

Several countries subsidize higher education. Two common approaches used to facilitate student enrollment are reducing out-of-pocket expenditures by offering scholarships or alleviating credit constraints by offering subsidized student loans. This is typically motivated by two types of arguments: (1) higher education generates positive externalities to society by fostering higher levels of innovation and productivity[1]; (2) as a way of providing equal opportunities to all well-qualified students. Even so, simply increasing the quantity of college-educated workers may not be enough for a society to reap the full benefits of higher education. Recent research has shown that education quality, as opposed to just quantity, represents an important factor in explaining variation in income within and across countries [Hendricks, Herrington, and Schoellman (2021); Martellini, Schoellman, and Sockin (2022).] However, in a market populated by private for-profit institutions, there is a concern that institutions with market power will imperfectly pass through subsidies by adjusting both price and quality.

In this paper, we address precisely this question and study the interplay between quality and subsidies in Brazil's higher education market. Brazil provides an interesting test case for examining the role of subsidies. Following a series of liberalization policies in the late 1990s, undertaken under tight fiscal constraints on the government, private provision has been the leading way to increase access to higher education. Such an expansion has led private institutions—nonprofits and for-profits—to dominate the market, serving more than 70% of enrolled students. In the early 2000s, the Brazilian government introduced targeted subsidies that facilitated the enrollment of low-income students in private colleges. Specifically, the ProUni program was designed to grant scholarships to low-income students and minorities, while the FIES program offered subsidized student loans. Parallel to the surge in private providers, the government created mechanisms to monitor the market: detailed administrative data and a mandatory standardized "exit" exam provide fundamental tools for assessing the quality of education. In this context, the objective of this paper is twofold: first, to create a quality measure for all undergraduate programs in Brazil and extensively validate it; second, to use this measure in a model of supply and demand to gauge how students value quality and how colleges choose their levels of quality and price.

We measure quality with value-added models, where a student's post-graduation outcome is predicted by their pre-enrollment characteristics and a program fixed effect, which we interpret as value added. To do so, we leverage rich individual-level data and link multiple administrative datasets to track students from high school graduation through college and into the workforce. We use a nationwide standardized entry exam, ENEM, to collect data on the students' characteristics before enrollment and then link these students to undergraduate programs using the Higher Education Census. We focus on

[1] Arrow (1973); Kremer (1993) argue that subsidies to education might be optimal even in the absence of productivity gains because under imperfect matching of workers due to imperfect information, there will be underinvestment in education.

two post-graduation outcomes: a standardized "exit" exam that tests graduating cohorts on major-specific knowledge, ENADE, and students' labor income sourced from a matched employer-employee database, RAIS. Therefore, we derive two value-added measures: one based on an exam and another on income.

The first descriptive result we document is that exam value-added is positively correlated with income value-added. In other words, our estimates indicate that colleges' contributions to students' academic achievement and to their post-college salaries are correlated. We then show that a sizeable part of the variation in value added is explained by what could reasonably be called inputs to value-added production. Namely, the composition of instructors—in terms of educational attainment and type of work contract—the student-instructor ratio, and the average ability of students are relevant predictors of value added. In this setting, we also find that, conditional on these inputs, private institutions have higher value added than publics. This finding suggests that private institutions may be more efficient in producing quality than public institutions, conditional on their level of resources. Finally, we find that value added increases in the share of students receiving subsidy support.

Moreover, we investigate whether a program's quality varies systematically across subpopulations. In other words, we examine the possibility that students of a particular group might benefit more from their education than others. To do so, we replicate our value-added model using data only from students of a particular subpopulation to obtain a value-added measure specific to that subpopulation. We find small but significant heterogeneity in terms of income and subsidy status. However, we do document a substantial gap for students who are ranked in the bottom quartile of the ability distribution of their programs. More precisely, the value-added for students attending for-profit institutions and in the bottom quartile of the ability distribution is, on average, 0.3 standard deviations below the value-added for the entire population. We also find a 0.25 gap between students in the top quartile versus students in the bottom quartile of ability in for-profit institutions. This disparity is not only significant in its own right but also informs our approach to incorporating student heterogeneity into the model, allowing for different value-added valuations between students of varying abilities.

Next, we develop a static equilibrium model of demand, pricing, and quality provision. Students follow a canonical discrete choice model to select an academic program—or not to enroll in college at all. They consider price, distance, value-added quality, probability of receiving either form of subsidy, and the student's entry exam performance relative to the program's average. This last factor is particularly crucial as it determines admission probabilities to competitive programs, effectively making some programs inaccessible to students with lower scores. Scholarships affect students by changing their expected out-of-pocket tuition, while loans affect students by changing their sensitivity to price. On the supply side, there is a fixed set of private colleges, which compete by simultaneously and strategically choosing price and quality.

We estimate the demand side as a Logit model using student-level data from students who took the national entry exam. This allows us to incorporate a rich amount of observed student heterogeneity in the model while keeping it tractable in a setting with thousands of programs and millions of students. One robust feature we learn from our estimates is that high-ability students value quality more than less qualified students and that men value quality significantly less than women. Moreover, we also estimate a strong distaste for distance among students. We then estimate the supply side separately via GMM to recover cost parameters, drawing on moments derived from the program's first-order conditions.

With the model at hand, we consider two counterfactuals: decreasing the supply of loans by 10% and decreasing the supply of scholarships by 10%. We find similar patterns under both scenarios but much stronger effects for scholarships. Specifically, by lowering the supply of scholarships by 10% in each program across the country, there would be 13 thousand fewer students in college, corresponding to approximately 80% of the total reduction in scholarships. For-profit institutions would lose about 9% of their market share, while nonprofits would lose only about 0.9%. Most programs would have an incentive to decrease quality and price, with a median change of -5% in value-added and of -0.7% in price. Furthermore, these reductions would be more pronounced in for-profit programs.

A notable feature of our model is that it allows us to understand the student composition of each program and to uncover the vast heterogeneity underlying the aggregate effects. Under the scholarship counterfactual outlined above, the change in market share ranges from -30% to +30%, with the bulk of the variation happening in the -8% to +0.5% range. In terms of student composition, programs that would lose market share have, on average, students of higher ability but of lower income than programs that would gain market share. We interpret our results as suggesting that the counterfactual reduction in scholarships would set into place a reshuffling of students in the market. Students targeted in the counterfactual are of slightly higher ability and lower income than a typical student enrolled in their program. As the likelihood of obtaining a subsidy decreases, these students decide not to enroll in college or to move to a cheaper program with students of lower ability, which may trigger further movements down the ability ladder. Overall, the counterfactual simulation indicates that most programs would experience an increase in the average ability of students and a decrease in the average income. An analogous pattern would happen to students who do not enroll in college because students of relatively high ability and low income would leave higher education altogether.

Taken together, the evidence presented in this paper paints a nuanced picture of the higher education sector. We conclude that a targeted and merit-based subsidy scheme could potentially encourage marginal growth in quality by selecting high-ability students who are invested in the quality of their education. Furthermore, our estimates also suggest that, conditional on their resources, private institutions might be more efficient than public institutions in producing quality. Thus, channeling students who could not access public institutions to private programs might be a cost-effective approach to increase access while keeping quality under control.

Related Literature. Our paper is connected to a few branches of literatures. First of all, we contribute to the growing body of work that studies how imperfect competition interacts with quality in education markets; chief among these are Neilson (2021); Allende (2021); Bau (2022); Armona and Cao (2022); Bodéré (2022). Our work is also closely related to the work of Dobbin, Barahona, and Otero (2022), who also study the Brazilian higher education market but focus on the effect of subsidized student loans on price. We see our work as a complement to this analysis, as we incorporate a quality measure and also focus on scholarships. Our study of how subsidized student loans interact with quality is also close to Barahona, Dobbin, Ho, Otero, and Yannelis (2021), which investigates how private institutions responded to new regulations that imposed a financial cost to institutions where subsidized students dropped out or defaulted. Similar to our setting, they find that institutions do respond to these incentives. Furthermore, Otero, Barahona, and Dobbin (2023) also study the Brazilian context and investigate how affirmative action quotas in the public institutions affected several student outcomes.

Our model of college choice is also connected with recent work of Kapor, Karnani, and Neilson (2023), who model college choice in Chile combining data from a centralized mechanism as well as with off-platform options. Even though we do not have application data, our setting also combines options on a platform and off-platform options. Another relatively close study of subsidies in the higher education sector is Lau (2020), who investigates what would be the effect of tuition-free communit colleges in the US. More broadly, we contribute to the literature that investigates subsidies' pass-through to tuition in higher education: Cellini and Goldin (2014); Turner (2017); Lucca, Nadauld, and Shen (2018); Kelchen (2019).

Our paper also speaks to the broader literature on value-added models in education, surveyed by Koedel, Mihaly, and Rockoff (2015), which has traditionally focused on teacher value-added, notably Chetty, Friedman, and Rockoff (2014). A notable exception is the work of Mountjoy and Hickman (2021) who study college value-added in the US. Our rich data environment with linked individual-level data allows us to estimate value-added measures for almost all undergraduate programs in Brazil.

2 Institutional Details

2.1 Overview

The Brazilian higher education sector is fairly large. In 2019[2], there were about 5.5 million students enrolled in undergraduate in-person programs, out of which about 1.1 million were newly enrolled students. In 2018, there were about 7.7 million students enrolled in high school, with about 2.1 million in the last year of their studies. Most students attend a

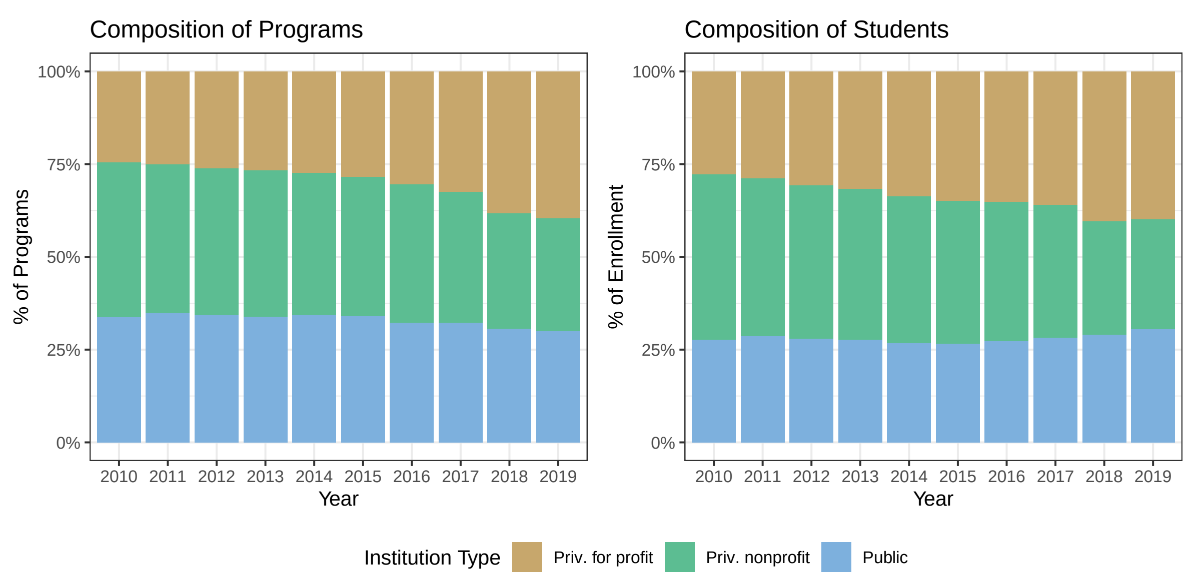
[2]In this paper, we will most of the time focus on the group of students who took an entry exam in 2018 and enrolled in 2019

public high school: only 13% of enrolled students attended a private institution, a pattern that is reversed in higher education, as we will see.

To attend a higher education institution, students apply to a specific institution-major-location, which we will call a "program" henceforth. Liberal Arts colleges are extremely rare, as colleges tend to be more technical or geared towards specific professional tracks. For instance, popular majors include Law, Medicine, Business Management, and Accounting. Typically, each program announces a fixed number of slots before opening applications, and students are admitted based solely on their grades in an entry exam, generically called the *Vestibular*. Thus, a program with n slots will accept eligible students with the n highest grades. Public institutions and subsidies also have affirmative action quotas, where a fraction of slots is set apart for eligible students who compete separately for these slots.

Public institutions are free of charge for every student. Typically, the oldest and most prestigious universities are public. Public programs admit students based on their grades in a nationwide standardized exam, the ENEM. Application for these public institutions is carried out in a centralized platform called SISU. Moreover, as we will show later, public institutions are, on average, of better quality. On the other hand, application to private institutions is decentralized—for each option the student is considering, he must submit a separate application and pay a fee.

Figure 1: Composition of Programs and Students Across Time.

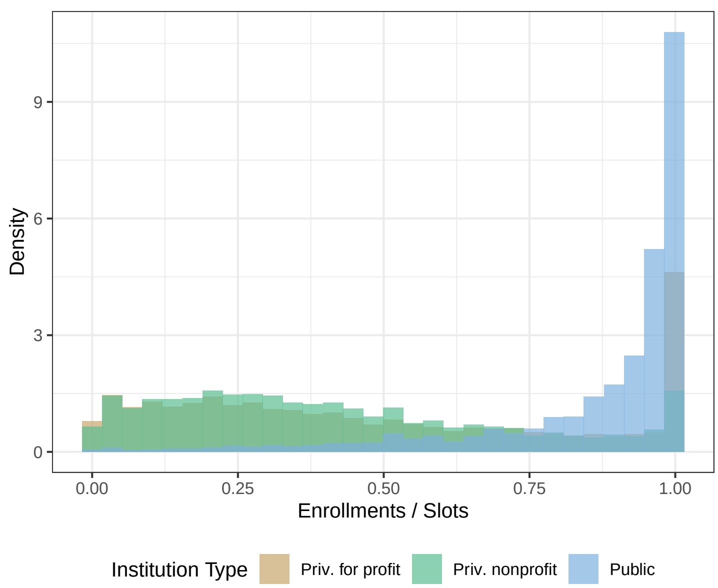


Note: This figure shows the evolution across time of institution type market shares. The left panel presents the composition of in-person undergraduate programs and the right panel presents the composition of enrolled students. Most students attend private institutions and most programs are run by private institutions. Source: Higher Education Census.

Despite the qualities of public institutions, most college students in Brazil attend a private institution. Among these institutions, for-profit colleges constitute a large fraction. For instance, figure 1 shows that in 2019, approximately 70% of undergraduate students attended private institutions, and 40% attended for-profit colleges. The composition of programs follows a very similar pattern. Moreover, we can also see that for-profit programs have been steadily increasing their market share.

One reason why private institutions dominate the market is that public institutions are capacity-constrained. Following a series of liberalization policies in the late 1990s, undertaken under tight fiscal constraints on the government, private provision has been the leading way to increase access to higher education. Figure 2 shows the distribution of applications per slot across programs in our sample. One can clearly see that public institutions are oversubscribed, with most programs operating at capacity. Private institutions, on the other hand, operate below capacity for the most part. In fact, 89% of private programs report a new enrollment to slot ratio below 1. This is a feature that will inform our model later on, as we will assume that private institutions do not face capacity constraints.

Figure 2: Distribution of Capacity Constraints



Note: The figure presents the distribution of the ratio of new enrollments to slots across undergraduate programs in 2019. A program operating at capacity shows a ratio of 1. Source: Higher Education Census.

2.2 ENEM

The ENEM (National Secondary Education Examination) was established in 1998 and reached its current format in 2009. Strictly speaking, the exam is designed to assess students' ability in the secondary education curriculum and not necessarily as an entry exam.

We will, however, refer to it as an entry exam to higher education, as it is, in practice, the main criterion for admission in most colleges. The exam is used by all public institutions and subsidy allocation mechanisms. The vast majority of private institutions also accept ENEM scores in their applications. Anyone who has graduated from or is enrolled in high school can take the ENEM to access higher education programs. The ENEM exam is formulated by INEP[3] and is administered in two days in early November. Participants take multiple choice exams in four subjects: Languages[4], Sciences, Humanities, and Mathematics, summing up to 180 questions. Participants are also evaluated in essay writing, which requires the development of an argumentative text. Typically, students will use their grades to enroll in a program or apply for subsidies the following year.

Some sources have claimed that ENEM is the second-largest standardized entry exam in the world, only behind its Chinese equivalent. In 2018, for example, almost 4 million students took the exam, out of which about 0.9 million students enrolled in college in the following year. We use ENEM to learn about students' pre-enrollment characteristics, especially because students must fill out an extensive socioeconomic questionnaire when registering for the exam.

2.3 Tuition Subsidies

The federal government uses two main types of subsidies to facilitate access to higher education: scholarships and subsidized student loans. Two key features common to them are their targeted and merit-based nature. In particular, they all require students to have a household income below a certain threshold and assign subsidies based on entry exam grades. Application to subsidies is independent of application to admission in a program, so students can apply before or after being admitted to a particular program. For the scholarships, however, acceptance implies being admitted for enrollment in the program.

Figure 3 depicts the evolution of subsidy use across time. The vast majority of programs received subsidized students—more than 80% for most years. From 2011 to 2014, subsidized student loans skyrocketed. After facing a dire fiscal crisis in 2015, the government decided to roll back the expansion of subsidized loans, cutting the program to levels similar to 2013.

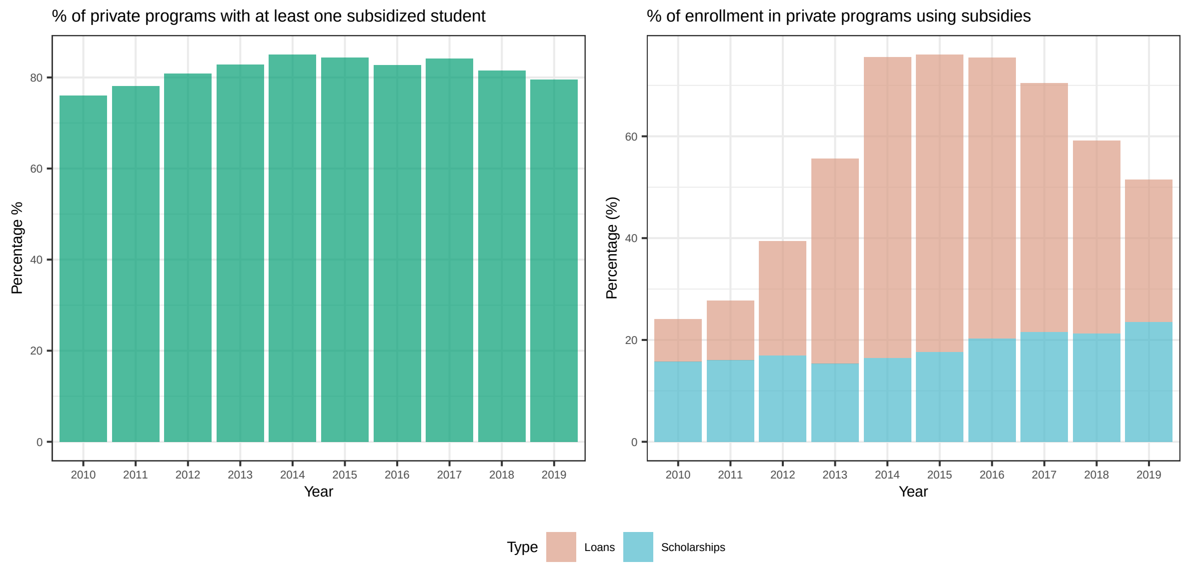
To apply for a subsidy, students use an online platform to register and list their preferences. During the period we studied, each form of subsidy operated on a different platform. The actual assignment mechanism follows a variation of the iterative deferred acceptance mechanisms, whose theoretical properties have recently been analyzed by Bó and Hakimov (2022). The details vary from year to year, but a simplified timeline can be sketched as:

1. A fixed number of scholarships and subsidized loans is allocated to eligible programs

[3] INEP (National Institute of Education Research) is a federal government agency whose mission is to produce and foster education research and to elaborate education assessment exams.

[4] All students are tested in Portuguese and a foreign language, either English or Spanish.

Figure 3: Subsidy Use Across Time



Note: The left panel shows the share of private programs with at least one subsidized student, while the right panel shows the composition of enrolled students in terms of subsidy use in private programs. Source: Higher Education Census.

2. Students take ENEM
3. Eligible students use a platform to list up to 2 options of scholarships/loans[5]
4. During a period of a few days, students can update their list every day.
5. At the end of each day, a cutoff is estimated and made public
6. On the last day, students are matched by selecting students with the highest grades
7. After the first round, unmatched students can form a second list and repeat the process.
8. If a student wasn't matched after the second round, he could choose to join a waitlist.

ProUni. Created in 2005[6], this program offers partial and full scholarships to low-income students attending private institutions. The program also has an affirmative action component in the form of quotas based on race. Eligibility criteria varied slightly across time, but in 2019—the year we focus our analysis—to qualify, a student must:

1. not have a higher education degree;

[5] An option is a particular form of subsidy for a particular program

[6] Formally established by Federal Law 11.096 of January 13, 2005

2. have graduated from a public high school or private high school with a full scholarship, or be a teacher in a public secondary or primary institution, or have some disability;
3. have a per capita household income of at most 1.5 minimum wages per month (full scholarship) or at most 3 minimum wages (partial scholarship)[7];
4. have scored an ENEM average grade of at least 450 points, with an essay grade above zero, in the latest exam edition.

FIES. Created in 2001[8], this program offers subsidized student loans to low-income students attending private colleges. The terms of the loan are generous but varied over time. In 2019—the year we focus our analysis—it offered a zero interest rate, the student was only required to start paying off their debt 18 months after graduation and could spread payments in a timeline three times the duration of his studies, which for most programs corresponds to 12 years. To be eligible, a student must:

1. have a per capita household income of at most 3 minimum wages per month;
2. have scored an ENEM average grade above 450 points, with essay grade above zero, in any edition of the exam since 2010.

2.4 ENADE

As part of an effort to monitor and regulate the higher education sector, the Brazilian government created the ENADE (National Student Performance Exam), a standardized exam designed to assess the performance of students from undergraduate programs in Brazil; it has been formulated by INEP since 2004. The exam is taken by students of a given major every three years and tests students' knowledge of major-specific content. Registration is mandatory for qualified incoming and graduating students, but we only use students who are graduating in our sample. Therefore, we will refer to it as an "exit" exam, as opposed to the entry exam we mentioned earlier.

One issue we would like to address head-on is that ENADE is not necessarily a high-stakes exam—students do not suffer any grave penalty for doing poorly in the exam. We would like to offer some counter-points to this view. Even though there are no direct mechanisms to incentivize students to exert high effort, there are indirect ways to affect them. One clear mechanism through which students might care about their ENADE grades is if they care about the reputation of their institution and how it might affect their post-graduation outcomes. Moreover, the ENADE exam is well-known among participants in the higher education industry; the exam's result is publicized by the government and private institutions to advertise their programs. Thus, private institutions have an incentive

[7] Teachers are not subject to the income requirement

[8] Formally established by the Federal Law No. 10.260 of July 12, 2001

to encourage students to a put high effort into the exam in order to advertise the quality of their program. Public institutions, on the other hand, do not face the same market incentives. Thus, these incentives should bias our results towards a small gap between public and private institutions. However, we do find a large gap between public and private programs, as shown in section 4.3. Furthermore, we conduct a series of validation exercises where we show that (1) entry exam grades are reliable predictors of the ENADE grade, (2) inputs to the production of value-added are strong predictors of value-added, and(3) income-based value-added is correlated with ENADE value-added. Thus, the collection of evidence suggests that the ENADE value-added does capture fundamental features that are commonly viewed as "quality" by the public.

3 Data

Higher Education Census. The cornerstone of this paper is the Higher Education Census. This database tracks the universe of students, instructors, programs, and institutions in the higher education sector in Brazil. We use this database to link students to programs and to construct program-level characteristics such as the composition of instructors. In our sample, we only used in-person undergraduate programs, which were matched with at least three students taking ENADE. Table 1 presents summary statistics taken from the Higher Education Census for the sample of programs we use in our analysis.

ENEM. The next crucial dataset we use is the ENEM administrative records. As discussed in Section 2.2, the ENEM is a nationwide standardized entry exam. From this dataset, we collect students' pre-enrollment characteristics: grades in all five subjects tested, age, race, income bracket, parents' education, and municipality of residency. To create a single continuous measure of income, we simulate an income draw for each student. We leverage the fine intervals used to report students' income brackets and fit a log-normal distribution to students' income. Then, we draw an income for each student from the fitted distribution conditional on having income within his reported income bracket. In appendix B, we show more details of this procedure. Table 2 presents summary statistics taken from the ENEM microdata for the sample of students we use in our analysis.

Subsidies. We recover the student's subsidy status from the Higher Education Census. However, we use the public registry of FIES loans to construct a measure of the average full tuition price per program. Since some programs extend discounts to some students, the full tuition financed by the loans might change from one student to another. More specifically, for each program, we compute the average full tuition price implied by the FIES loan registry. The tuition price is measured in thousands of Reais per semester. This procedure does not yield perfect coverage in our sample, so a few programs are left without tuition prices in our data. In Section 6, we describe in more detail how we deal with this issue when estimating the model.

Table 1: Summary Statistics of Program Sample

Variable	Mean	SD	P10	P25	P50	P75	P90
Public $n = 4,890$							
Enrolled Students	249.59	190.01	93.00	132.0	196.0	310.00	456.10
Incoming Students	61.90	39.92	27.90	38.0	51.0	79.00	103.00
Loans	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Scholarships	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Average ENEM Grade	0.63	0.55	-0.06	0.2	0.6	1.02	1.35
Tuition ($1,000 BRL / Semester)	0.00	0.00	0.00	0.0	0.0	0.00	0.00
Private nonprofit $n = 4,311$							
Enrolled Students	296.64	400.79	70.00	110.00	180.00	337.0	625.00
Incoming Students	81.47	125.54	18.00	29.00	48.00	88.0	169.00
Loans	6.49	22.74	0.00	0.00	1.00	5.0	14.00
Scholarships	8.40	14.52	0.00	0.00	3.00	11.0	22.00
Average ENEM Grade	0.19	0.40	-0.25	-0.07	0.15	0.4	0.68
Tuition ($1,000 BRL / Semester)	6.57	5.35	2.58	3.98	5.62	7.8	10.57
Private for-profit $n = 4,599$							
Enrolled Students	327.40	328.25	82.00	133.00	225.00	406.00	680.20
Incoming Students	106.03	112.16	23.00	38.00	70.00	136.00	224.00
Loans	11.05	33.03	0.00	0.00	3.00	9.00	25.00
Scholarships	15.74	25.56	0.00	3.00	9.00	19.00	36.00
Average ENEM Grade	0.04	0.29	-0.30	-0.15	0.02	0.20	0.39
Tuition ($1,000 BRL / Semester)	6.37	4.13	3.24	4.30	5.73	7.57	9.56

Note: This table shows summary statistics for the sample of we used. Figures correspond to 2019 values of all programs used in the model estimation sample. We only consider in-person undergraduate programs. Tuition is measured in thousands of Reais per semester. Columns named Px refer to the xth percentile of the distribution. Source: Higher Education Census and FIES Registry.

Post-Graduation. To measure students' performance upon graduation, we use the administrative records of the ENADE exam, from where we collect students' scores. As a second post-graduation outcome, we use RAIS. This matched employer-employee database tracks the universe of Brazilian workers hired in the formal labor market under the country's Labor Protection Law (CLT).

4 Measuring the Quality of Undergraduate Programs

We measure quality using a standard value-added model. Given a measure of post-graduation outcome Y_{ij}, and student-specific pre-enrollment characteristics, Z_i, the pro-

Table 2: Summary Statistics of Student Sample

Variable	Mean	SD	P10	P25	P50	P75	P90
Age	21.41	7.09	17.00	17.00	19.000	22.00	30.0
Income (Min. Wage)	2.76	4.06	0.44	0.94	1.455	2.99	6.0
Low Income	0.52	0.50	0.00	0.00	1.000	1.00	1.0
Parent Graduate HS	0.60	0.49	0.00	0.00	1.000	1.00	1.0
Male	0.41	0.49	0.00	0.00	0.000	1.00	1.0
Non-White	0.58	0.49	0.00	0.00	1.000	1.00	1.0
Graduated HS	0.87	0.34	0.00	1.00	1.000	1.00	1.0
Never Married	0.89	0.31	0.00	1.00	1.000	1.00	1.0
ENEM Sciences	0.00	1.00	-1.19	-0.78	-0.138	0.66	1.4
ENEM Humanities	0.00	1.00	-1.46	-0.73	0.178	0.74	1.2
ENEM Languages	0.00	1.00	-1.35	-0.70	0.068	0.73	1.2
ENEM Mathematics	0.00	1.00	-1.16	-0.78	-0.182	0.63	1.5
ENEM Essay	0.00	1.00	-1.17	-0.85	0.021	0.67	1.3
ENEM Average	0.00	1.00	-1.19	-0.74	-0.106	0.63	1.4
Distance (Full)	80.3	268.3	0	0	0	43.7	169.5
Distance (for-profits)	60.8	235.9	0	0	0	30	99.4
Distance (nonprofits)	62	230.6	0	0	0	33.2	104.2
Scholarship 100% (for-profits)	0.12	0.33	0	0	0	0	1
Scholarship 50% (for-profits)	0.05	0.21	0	0	0	0	0
Scholarship 100% (nonprofits)	0.11	0.31	0	0	0	0	1
Scholarship 50% (nonprofits)	0.05	0.22	0	0	0	0	0
Subsidized Loan (for-profits)	0.08	0.27	0	0	0	0	0
Subsidized Loan (nonprofits)	0.07	0.25	0	0	0	0	0

Note: This table shows summary statistics for the sample of 2018 ENEM takers used to estimate the model and simulate counterfactuals. We only use students who took all the ENEM 2018 exams in our sample. Columns named Px refer to the xth percentile of the distribution. Distance represents the geodesic distance measured in kilometers between the centroid of the student's municipality and the centroid of the program municipality, conditional on enrolling in college. Indicator for scholarship or Loan is conditional on enrolling in college. Source: ENEM Microdata.

gram value-added VA_j is implicitly defined by the following equation

$$Y_{ij} = Z_i\gamma + VA_j + \varepsilon_{ij} \tag{1}$$

We consider two types of post-graduation outcomes: grades from an "exit" exam and income from the formal labor market. We use a rich set of student-specific controls: ENEM score across five subjects, age, distance from the student's municipality to his program's municipality, race, sex, type of high school he graduated from, family income bracket, and mother and father education level. We estimate value-added as a program fixed-effect.

4.1 First Outcome: Exit Exam

Our first main measure of quality is based on the ENADE exit exam. This exam is taken every three years by the graduating cohort of a given major, and it tests students' knowledge of major-specific topics[9]. In our main implementation, we pool all students in the graduating cohorts from 2013 to 2017. We chose this approach because we do not have enough data to cover two full three-year cycles of tests across majors, which prevents us from obtaining time variation in value-added. The outcome variable, Y_{ij}^E, is defined as the student's grade standardized relative to that year's average grade among students from his major. More precisely,

$$Y_{ijmt}^E = \frac{G_{it} - \overline{G}_{mt}}{SD(G_{mt})} \tag{2}$$

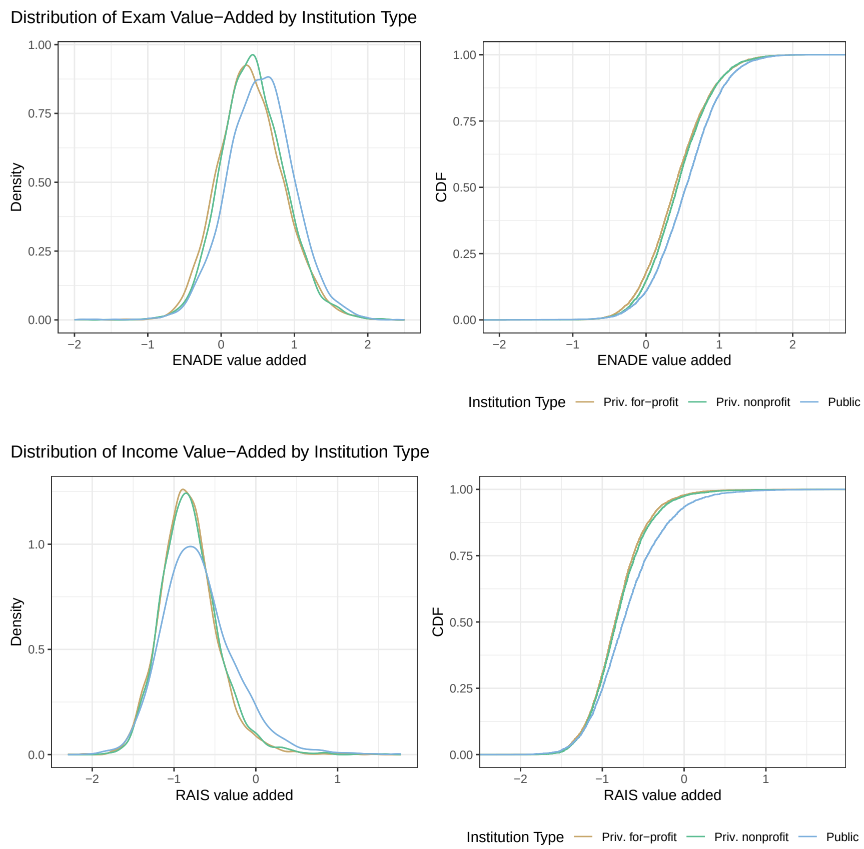
where Y_{ijmt}^E is the outcome measure for student i who graduated from program j, major m, at time t; G_{it} is student i's grade at time t; $\overline{G}_{mt}$ is the average grade of all students who graduated from major m at time t, including students from other programs; and $SD(G_{mt})$ is the standard deviation of the grades of all students who graduated from major m at time t. To make the notation lighter, we will omit subscripts m and t from the outcome variable.

To form our sample of students, we proceed in the following way. First, we select the graduating cohort of a given year on the higher education census and then link those students with ENADE takers of the same year. Then, for each student, we search for their ENEM data in the nearest year before enrollment. Our final sample is the collection of all students that were fully linked across ENEM, Higher Education Census, and ENADE. In our sample, we only used in-person undergraduate programs, which were matched with at least three students taking ENADE.

Figure 4 shows the distribution of recovered value-added estimates. One can clearly see a normal-like shape with a symmetrical and single-peaked distribution. In Section 4.3, we will dive into more detail regarding the properties of this distribution. Moreover, we

[9] The ENADE exam also has "general knowledge" questions. We only use the major-specific component of the exam when estimating value-added.

Figure 4: Distribution of Value-Added by Institution Type



Note: Here we show the distributions of value-added based on the exit exam in the two top panels and based on income in the two bottom panels. Panels on the left show the estimated densities, while panels on the right show the estimated cumulative distribution. Both measures are based on standardized outcomes.

capture significant spatial variation, as shown in Figure 5, where the regions with higher quality programs are agglomerated in the south and southeast of the country. The regions with the poorest performance are in the north and northwest regions. Finally, table 3 shows selected coefficients of the value-added model. In particular, we show the coefficients for the entry exam grades to emphasize the point made in section 2.4, that students' academic ability is a relevant predictor of the "exit" exam, so that our quality measure is, to some extent, able to control for observed sources of selection.

4.2 Second Outcome: Labor Income

We also construct value-added measures based on labor income sourced from the Brazilian matched employer-employee database, RAIS. The student's labor income is defined as

Figure 5: Spatial Distribution of Exam Value-Added

Note: This figure shows the spatial distribution of the exit exam value-added across Brazil. Here we show the average value-added for each micro-region.

the total annual income[10], measured two years after graduation. We measure income in minimum wages, which is a common measurement unit for workers with formal jobs in Brazil; we also adjust it for inflation to make numbers comparable across years. We have two versions of the outcome variable, Y^R_{ijt}. The first version is simply the total income defined previously, while in the second, we use the standardization applied to the ENADE grade in equation (2). In our main estimation, we pool all students from three graduating cohorts: 2014, 2015, and 2016.

To form our sample of students, we proceed in the following way. First, we select the graduating cohort of a given year on the Higher Education Census, and for each student, we search for their ENEM data in the nearest year before enrollment. Repeating this procedure for the 2014 to 2016 cohorts results in a sample with about 1.1 million students linked through ENEM and Higher Education Census. Next, we link those students to RAIS income data from two years ahead of their graduation year. We managed to match about 30% of the original 1.1 million students. The main reason why we cannot link most students to income data is that our database only includes formal jobs regulated under Brazil's labor protection laws. Thus, we cannot observe informal jobs or professionals working as contractors. Traditionally, some professionals, such as architects and lawyers, have mostly worked as contractors. Our sample is the collection of all students that were

[10] We only consider income from jobs started during or after the graduation year.

Table 3: Value-Added Model: Student's Ability

Dependent Var.:	ENADE Grade
ENEM Science	0.073***
	(0.002)
ENEM Humanities	0.112***
	(0.002)
ENEM Portuguese	0.118***
	(0.002)
ENEM Mathematics	0.070***
	(0.002)
ENEM Essay	0.012***
	(0.002)
Fixed-effects	
Program	Yes
Controls	
Student's Demographics and SES	Yes
S.E.: Clustered	by: Program
Observations	563,170
R2	0.297
Within R2	0.096

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table shows selected coefficients of the value-added model based on the ENADE exam. Both ENADE and ENEM grades are standardized.

fully linked across ENEM, Higher Education Census, and RAIS[11].

4.3 Descriptive Analysis

Now we will briefly discuss a series of descriptive exercises with two objectives in mind: first, to validate our value-added measure as a reasonable measurement of quality; second, to inform our analysis and model construction later.

Ordering of Institution Type. As a first exercise to validate our quality measures, we test for the apparent stochastic dominance between institution types. By inspecting figure 4, one may wonder whether public institutions first-order stochastically dominate private nonprofits and whether private nonprofits first-order stochastically dominate private for-profits. This would mean that, at every percentile, public institutions have a higher value-added than private nonprofits and, similarly, at every percentile, private nonprofits are better than for-profits. We use the statistical test formulated by Barrett and Donald (2003) to formally test the null hypothesis that such ordering is the case. We find strong evidence for the ordering as indicated in Table 4. This result is interesting because we

[11] In the actual estimation of value-added, we dropped students from programs with less than four fully linked students.

believe such ordering in terms of *absolute* levels of quality has long been held by the general public. Thus, our quality measure is aligned with what we believe is a common view in the country. Nonetheless, the result is still striking because it is in terms of *value added*; not only public institutions are more competitive and select better candidates, but they also seem to provide better value added. Furthermore, this result establishes that there is a clear divide in the market in terms of institution type, something we will take into account when modeling the market.

Table 4: Testing First-Order Stochastic Dominance of Value-Added Distribution

	Exam		Income	
Null Hypothesis	Test Statistic	P-Value	Test Statistic	P-Value
Public $\succeq_{FSD}$ Private Non-Profit	0.046	1.00	0.372	0.71
Private Non-Profit $\succeq_{FSD}$ Private for-profit	0.175	0.91	0.011	1.00
Observations				
Public	5,645		3,034	
Private nonprofit	6,100		4,220	
Private for-profit	5,904		4,317	

Note: This table shows the results of first-order stochastic dominance hypothesis tests for the distributions of value-added of different types of institutions. We use the methodology of Barrett and Donald (2003). To compute test statistics, we used a grid with 500 points and 300 bootstrap samples. Hypothesis tests implemented using the PySDTest package.

Income vs. Exam value-added. Our next question is whether exam value-added is correlated with income value-added. We test this by regressing our "exit" exam value-added on our income value-added. Table 5 shows that income value-added is positively correlated with our preferred exam-based measure. One downside of our income measure is that we only match students who are formally employed under Brazil's labor protection laws. To mitigate this issue, we run the same regression using only programs with high coverage (more than 50%) in the matched employer-employee database. In this specification, results are even stronger. Here, we present results with both outcomes standardized. In Appendix A, we show that we get qualitatively similar results when the outcomes are not standardized.

Inputs to Production of Value Added. To further validate our quality measure, we investigate whether reasonable inputs are good predictors of value-added. In particular, in Table 6, we regress our exam-based value-added measure on a group of supply-side and demand-side inputs. For the supply-side inputs, we consider the share of instructors with graduate degrees, the share of instructors with part-time contracts, and the student-instructor ratio; the demand-side input is the average ENEM score of students enrolled in the program. In column (2), one can see that all inputs are strong predictors of value-added and explain a sizeable fraction of the variation. Moreover, the sign of inputs corresponds to our prior on how these inputs would affect value-added: the share of instructors with

Table 5: Income vs. Exam Value-Added

Dependent Variable:	Exam Value-Added (ENADE)			
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	0.6569***		0.8237***	
	(0.0102)		(0.0213)	
Income Value-Added (RAIS)	0.2255***	0.1952***	0.4105***	0.3136***
	(0.0116)	(0.0231)	(0.0258)	(0.0494)
Fixed-effects				
Major	No	Yes	No	Yes
Sample	Full	Full	High coverage	High coverage
Fit statistics				
Observations	9,902	9,902	2,738	2,738
R^2	0.03665	0.25127	0.08483	0.31678
Within R^2		0.03243		0.06032

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table shows the result of regressing ENADE value-added (exit exam) on income value-added. Both value-added measures were estimated with outcomes standardized by major. We consider two samples: full sample and a sub-sample containing only programs with high coverage (greater than 50%) in the RAIS database.

a graduate degree is positively correlated, while the share of instructors with part-time contracts and the student-instructor ratio are negatively correlated. We also note that the composition of student quality, measured by the average ENEM score, is also positively correlated, suggesting that some form of peer effects play a role in determining the contribution of a program on individual student outcomes. One interesting feature is that without controlling for inputs, the coefficients on the type of private institution are negative, as indicated in column (1). Once we add inputs to the regression, the sign of these variables flips to positive. We hypothesize that this can be understood as suggesting that private institutions are, in some sense, more efficient than public institutions. That is, conditional on their resources and student composition, private institutions are, on average, better than public institutions.

Correlation Between Subsidies and Value Added. In this context, a natural question to ask is how exposure to subsidies is correlated with quality from a program's perspective. To get some insight on this question, we maintain the set of controls used so far and add two variables: the fraction of enrolled students with scholarships and fraction of enrolled students with subsidized loans, both measured in 2011. We use these measures of exposure to subsidies in the past to avoid a mechanical relationship induced by the possibility that subsidies were allocated based on value-added itself. Column 3 in Table 6 shows positive coefficients for these variables, and column 4 shows a similar correlation without conditioning on the inputs. There are multiple explanations for the sign of these coefficients, including that the merit-based nature of subsidy allocation brings high-ability students who offer a positive peer effect to the program, and that there is some persistence

Table 6: Value-Added Production Inputs and Past Subsidies

Dependent Variable:	ENADE Exam Value-Added			
Model:	(1)	(2)	(3)	(4)
Variables				
Private for-profit	-0.2363***	0.2198***	0.2006***	-0.2875***
	(0.0327)	(0.0174)	(0.0190)	(0.0332)
Private nonprofit	-0.1892***	0.1787***	0.1658***	-0.2302***
	(0.0258)	(0.0140)	(0.0172)	(0.0269)
Share Instructors Graduate Degree		0.2379***	0.2256***	
		(0.0148)	(0.0140)	
Share Instructors part-time		-0.2121***	-0.2217***	
		(0.0158)	(0.0169)	
Students / Instructors ratio		-0.0039***	-0.0034***	
		(0.0005)	(0.0005)	
Average ENEM score		0.3766***	0.3771***	
		(0.0233)	(0.0229)	
Share Students Scholarship 2011			0.1805***	0.4496***
			(0.0402)	(0.0464)
Share Students Loans 2011			0.1499***	0.1171**
			(0.0459)	(0.0573)
Fixed-effects				
Major	Yes	Yes	Yes	Yes
Fit statistics				
Observations	17,649	16,018	14,577	16,005
R^2	0.231	0.393	0.397	0.236
Within R^2	0.048	0.241	0.250	0.057

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table shows regressions of value-added on a set of inputs to quality production. In columns 3 and 4, we include a program's share of subsidized students in the past, 2011. Our value-added sample only considers students who graduated from 2013 to 2017. Standard-errors are clustered at the major level.

of value-added across time such that high quality programs received more subsidies and persisted to be of high quality in the future.

Heterogeneity. Now we study whether a program's value-added varies across different subgroups. We replicate our value-added model estimation using only data from students who belong to various populations and denote it VA_j^{Subgroup}. To investigate systematic differences between the whole population and subgroups, we define the following variable:

$$\Delta VA_j^{\text{Subgroup}} := VA_j^{\text{Subgroup}} - VA_j,$$

where VA_j is value-added computed using the full sample of students. We consider the following subgroups: students from low-income households, students who were subsidized, students who ranked in the top quartile of the ability distribution of their program, and students who ranked in the bottom quartile. Then, we regress $\Delta VA_j^{\text{Subgroup}}$ on a constant and indicators for the type of private institution. Table 7 shows the results in columns 1–4. We do find statistically significant differences for all subgroups but relatively small coefficients for low-income and subsidized students. There is a relatively large difference for the subgroup of students in the bottom quartile of private programs, reaching a mean difference of about -0.3 standard deviations.

A related question we explore is the direct comparison between students in the top and bottom quartile of ability. To be precise, we run the same regression but with the following dependent variable $\Delta VA_j^{\text{Top-Bottom}} := VA_j^{\text{Top Quartile}} - VA_j^{\text{Bottom Quartile}}$. In column 5 of Table 7, we can see that there is a relevant gap for students attending private institutions, where the difference is about 0.2 standard deviations. Interestingly, the gap between value-added for top and bottom ability is much smaller in public institutions, possibly because student ability in public institutions is substantially higher and less heterogeneous than in private institutions. The presence of some degree of heterogeneity across groups will inform our model decisions later on. In the college choice utility, we will interact value added with an indicator for low-income and with the students' ENEM grade. Thus, we will allow low-income and high-ability students to value quality differently.

Table 7: Heterogeneity in Exam Value-Added

Dependent Variables:	$\Delta VA_j^{\text{Subgroup}}$				$\Delta VA_j^{\text{Top-Bottom}}$
Subgroup:	Low-income	Subsidized	Top Quartile	Bottom Quartile	
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Constant	0.0262***	0.0319***	-0.1272***	-0.1531***	0.0296***
	(0.0059)	(0.0035)	(0.0042)	(0.0049)	(0.0085)
Priv. for-profit	-0.0400***	-0.0165***	0.0574***	-0.1545***	0.2212***
	(0.0084)	(0.0046)	(0.0059)	(0.0071)	(0.0121)
Priv. nonprofit	-0.0373***		0.0517***	-0.1372***	0.1931***
	(0.0090)		(0.0060)	(0.0070)	(0.0120)
Fit statistics					
Observations	7,304	8,538	13,098	11,420	10,410
R^2	0.00380	0.00153	0.00853	0.04839	0.03772

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table shows the results of regressing value-added computed on the full sample of students on value-added measures computed on subsamples. Heteroskedasticity-robust standard-errors in parentheses.

Lastly, not only can we precisely estimate differences in the mean value added across populations, but in Appendix A, we also show that there is a strong positive correlation between value-added for subgroups and full sample.

5 Model

In this section, we build a model of demand and supply of the higher education sector. Our aim is to capture (i) how students choose their college or not to go to college, (ii) how subsidies affect their decisions, and (iii) how private colleges set their levels of quality and price. We proceed in three steps: first, we describe how a student forms expectations about his prospect of receiving either form of subsidy; second, using the probability of winning subsidies, he chooses an undergraduate program; third, colleges choose price and quality to maximize their profits.

5.1 Subsidy Allocation

As described in section 2.3, application to subsidies is independent of the student's application to a college. Unfortunately, our data only allows us to observe the final outcome for each student, i.e., which option he chose and whether he received any form of subsidy. Therefore, we cannot directly model the subsidy assignment mechanism. To circumvent this issue, we model how a student forms his expectations about the prospect of being selected for a type of subsidy in a given program. For each program j, government fixes a number of subsidies of type d: N_j^d Student i receives subsidy of type d for program j according to

$$S_{ij}^d = 1\left\{Z_i\theta_Z + DE_{ij}^{d+}\theta_E^+ + DE_{ij}^{d-}\theta_E^- + M_j\theta_M + N_j^d\theta_N + \zeta_{ij} > 0\right\} \tag{3}$$

where Z_i is a vector of student-specific variables such as household income bracket, type of high school, race, age, and sex; M_j is a vector of institution- and market-specific variables, and ζ_{ij} are iid standard normal. Finally, we include the positive and negative parts of DE_{ij}^d, which is the deviation of the student's score from the program's mean. That is to say, if E_i is the student's entry exam score, and if $\boldsymbol{E}_j^d$ is the collection of scores of students attending j with subsidy d, then

$$DE_{ij}^{d+} := \max\{0, E_i - mean(\boldsymbol{E}_j^d)\} \qquad DE_{ij}^{d-} := \min\{0, E_i - mean(\boldsymbol{E}_j^d)\}$$

These variables capture the competitive nature of the subsidy allocation, such that a student with poor grades is very unlikely to receive subsidies in a competition program with high average grades. Thus, we impose that students' expectations about subsidy allocation are consistent with observed data. The next step is to incorporate the student's expectations into his program choice. In particular, we will use the estimated probabilities conditional on the student's characteristics and program characteristics:

$$\hat{S}_{ij}^d = \Phi\left(Z_i\hat{\theta}_Z + DE_{ij}^d\hat{\theta}_E + M_j\hat{\theta}_M + N_j^d\hat{\theta}_N\right) \tag{4}$$

5.2 College Choice

Students follow a discrete choice model to select which program they will enroll in or whether they will not go to college at all. This model captures in a tractable way the key trade-offs students face when deciding where to go: price, quality, distance, and likelihood of subsidies. In particular, student i's indirect utility associated with program j is given by

$$U_{ij} = \delta_j + \beta^{E+} DE^{+}_{ij} + \beta^{E-} DE^{-}_{ij} + \beta^{D} \text{Dist}_{ij} + \beta^{V}_i VA_j + \beta^{P}_i P_{ij} + \beta^{S} \hat{S}^{L}_{ij} \cdot P_{ij} + \varepsilon_{ij} \quad (5)$$

where Dist_{ij} is the distance from student i's municipality to the j's municipality; VA_j is the value-added of program j; P_{ij} is the expected tuition of program j given student i's scholarship probability. For instance, if a program only offers full scholarships, then

$$P_{ij} = P_j(1 - \hat{S}^{FS}_{ij}),$$

where $\hat{S}^{FS}_{ij}$ is the estimated probability of student i receiving a full scholarship at program j. For a program with both partial (50% funding) and full scholarships, we define

$$P_{ij} = \min \left\{ P_j(1 - \hat{S}^{FS}_{ij}), \quad 0.5 \cdot P_j \cdot \hat{S}^{PS}_{ij} + P_j(1 - \hat{S}^{PS}_{ij}) \right\}.$$

Similarly, the probability of getting a subsidized student loan, S^{L}_{ij}, changes the student's sensitivity to price. Finally, a key component of the model is the deviation of a student's entry exam grade relative to the program's mean:

$$DE^{+}_{ij} = \max\{0, E_i - mean(\boldsymbol{E}_j)\} \qquad DE^{-}_{ij} = \min\{0, E_i - mean(\boldsymbol{E}_j)\}.$$

These two variables capture two relevant components of a student's choice. First, given the competitive admissions for some programs, these variables ensure that a student who did poorly in the entry exam is very unlikely to enroll in a competitive program where most students got high grades. Second, these variables also capture taste for being close or distant to the program's mean in terms of academic ability. Using the positive and negative part of the grade deviation allows asymmetric effects for being below or above the program's mean. The outside option represents the choice of not enrolling in college at all, and we normalize its utility to $u_{i0} = \varepsilon_{i0}$.

We include a program-specific intercept, also referred to as mean utility, which also captures unobservable features of the program that might lead to endogeneity in prices and quality:

$$\delta_j = \overline{\beta}_V VA_j + \xi_j.$$

Since our measure of price is student-specific, it is not included in the mean utility. Lastly, the unobservable ε_{ij} is an iid Type-1 extreme value distribution. Thus, in our model, scholarships affect the student's decision by decreasing his expected price, while loans make him less price-sensitive. Thus, conditional on student i's characteristics, the probability of him choosing program j is

$$\pi_{ij}(Z_i, E_i, \boldsymbol{S}_i, \boldsymbol{P}, \boldsymbol{VA}) = \frac{e^{V_{ij}}}{1 + \sum_l e^{V_{il}}}.$$

where $V_{ij} := U_{ij} - \varepsilon_{ij}$ is the non-random component of his utility. Since we observe the universe of all students taking the ENEM entry exam, the aggregate demand faced by program j is defined by the aggregation of choice probabilities across all students in the market:

$$D_j(\boldsymbol{P}, \boldsymbol{VA}) = \frac{1}{N} \sum_{i=1}^{N} \pi_{ij}(Z_i, E_i, \boldsymbol{S}_i, \boldsymbol{P}, \boldsymbol{VA}).$$

5.3 Post-Graduation Outcome

Even though we already described in section 4 how we estimate value added, we now take a stance on how this relates to our model. Given that student i chose program j, his post-graduation outcome, Y_{ij}, is determined according to our value-added model:

$$Y_{ij} = Z_i \lambda_Z + VA_j + \eta_{ij} \tag{6}$$

In the manner of Allende (2021) and Otero, Barahona, and Dobbin (2023), this can be viewed from a potential outcomes perspective. Our model is saying that student i's potential outcome of going to program j is determined by equation (6). Moreover, VA_j is an equilibrium object in the model, determined jointly by demand and supply forces.

5.4 Supply

We follow Neilson (2021) and let private colleges compete à la Nash-Bertrand by choosing price and VA levels simultaneously. More precisely, they solve the following problem

$$\max_{P_j, VA_j} \; (P_j - MC_j(W_j, VA_j))\, D_j(\boldsymbol{P}, \boldsymbol{VA})$$

Where we parameterize marginal cost as $MC_j = c_0 + c_W W_j + c_V VA_j + \omega_j$. In our implementation, we allow cost coefficients to vary by the institution type, i.e., nonprofit or for-profit. Our formulation yields the typical first-order conditions:

$$P_j = MC(W_j, VA_j) - D_j \cdot \left(\frac{\partial}{\partial P_j} D_j \right)^{-1} \tag{7}$$

$$VA_j = \frac{1}{c_V} (P_j - c_0 - c_w W_j - \omega_j) - D_j \cdot \left(\frac{\partial}{\partial VA_j} D_j \right)^{-1} \tag{8}$$

Equation (7) gives rise to the traditional pattern in oligopolistic competition models, where the equilibrium price is equal to the competitive price—here, this is simply the marginal cost—plus a markup. The markup is determined by how price-sensitive the aggregate demand for a particular program is, in other words, how much market power the college has. Notice that such market power depends on the full market structure, i.e., on the particular collection of competitors and their respective levels of price and quality. As Neilson (2021) points out, equation (8) represents an analogous phenomenon for the program's quality. The equilibrium level of quality is equal to a competitive component plus a markdown, which subtracts from the competitive level. Just as with price markups, the quality markdown is determined by the college's market power: the higher the market power, the further away it sets its quality from the competitive level.

We will not model how public institutions operate and will treat them as an exogenous component of the model. Public programs do not respond to market incentives as tuition is always set to zero for all students, and hiring follows different patterns. Moreover, we do not impose any capacity constraints for private institutions; Figure 2 indicates that the vast majority of private programs operate below capacity.

6 Identification and Estimation

Preliminaries. Before we estimate the college choice model itself, we separately estimate value-added via equation (6) and subsidy probabilities by equation (3). This approach relies on an important assumption we will impose to estimate the model. Namely, the unobservables from each of the three components we have seen so far are independent across i and j:

$$\eta_{ij} \perp\!\!\!\perp \varepsilon_{ij} \perp\!\!\!\perp \zeta_{ij}.$$

Even though one could relax this assumption in multiple ways, we adopted this approach to allow a tractable estimation, where each component is estimated independently.

Market Definition. To estimate the model, we partition the country into 24 geographical areas, each corresponding to a different market. The demand side corresponds to all students residing in the region who took the entry exam, and the supply side corresponds to all in-person undergraduate programs that operate in the region. Our main constraint was running our estimation procedure with limited time and computational power available in a secure data room. Thus, a market is the largest possible contiguous region where we could run our estimation routine. To define the market regions for less densely populated states, we start by agglomerating states up to the point where the next neighboring state would lead to more than 1,000 programs. For more densely populated regions, we aggregate meso-regions within a state up to the same threshold used for states. The only two exceptions are the markets for the two most populated regions in the country: the capitals of Rio de Janeiro and São Paulo. These two are separate markets corresponding to the micro-region containing the two capitals. The definitions of micro- and meso-regions were formulated by the Brazilian Institute of Geography and Statistics (IBGE) and are common ways to partition the country's space. To deal with students who choose a program outside their market, we add an option in this choice set that aggregates all programs outside the market. In appendix C, we describe the markets used in the estimation and counterfactuals in more detail.

College Choice. The choice model is a multinomial Logit estimated in each market separately by maximum likelihood. Thus, we exclusively use within market variation of student characteristics and product characteristics to identify parameters. This empirical strategy is informed by the fact that we use rich microdata with the universe of every student who took the ENEM exam in the country, which allows us to incorporate observed heterogeneity among students directly.

After estimating the Logit coefficients, we use the mean-utility estimates, δ_j, to recover the parameter $\overline{\beta}_V$ while dealing with possible endogeneity generated by program-specific unobservable, ξ_j. To be more precise, our specification is the following.

$$\delta_{jm} = VA_{jm}\overline{\beta}_V + \xi_m + \xi_{jm},$$

where δ_{jm} is the mean-utility associated with program j of major m and ξ_m represents a major fixed-effect. Adding such a fixed effect to the model allows us to capture the fact that quality has major-specific levels of competition and attention given to quality. Because our expected price measure varies by student-program, we can recover $\overline{\beta}_P$ directly from our Logit estimates.

We estimate $\overline{\beta}_V$ by running a 2SLS where we regress δ_j on VA_j using a series of instruments. In particular, we instrument value-added with cost shifters: average instructor wage in a micro-region, average instructor wage of the program's major in a market, and average wage of workers with graduate degrees. Moreover, we also include the number of competitors within the same major and a "Waldfogel-Fan" Instrument. As Berry and Haile (2021) explain, this type of instrument essentially uses average demographics of nearby markets as exogenous shifters of equilibrium markups. Here, we use the average ENEM score of neighboring micro-regions as an instrument. We construct wage averages using the matched employer-employee database, RAIS. To run our estimation, we pool the mean utilities recovered across all markets, but we drop programs that are capacity-constrained from our sample. Our reasoning is that the mean utility of capacity-constrained programs does not reflect students' preferences as if they were freely choosing among options, which would bias our estimates for the utility parameters. We provide more details about the estimation of the college choice model in Appendix C.

Supply Side. After recovering the demand-side parameters, we estimate the supply side separately. First, we recover marginal costs, markups, and markdowns and then use GMM to estimate cost parameters. The moment conditions are based on the two first-order conditions described in equations (7) and (8). Our cost shifters are average wages taken from the matched employer-employee database. In particular, we use the average wage of instructors in a micro-region, the average wage of instructors of a particular major in a market, and the average wage of employees with graduate degrees in a micro-region.

Moreover, we instrument value added because it is a choice made by the program under full information of demand and supply. Thus, value added is possibly correlated with the marginal cost unobservable, ω_j. Our instruments are demand shifters: average ENEM score and average income in the micro-region where the program is located.

Results. Table 8 shows selected coefficients from the discrete choice model. One can see that the ENEM grade deviations, DE^-, DE^+, play an important role in disciplining the model regarding how the student's exam score affects his choice probabilities. Moreover, students have a strong distaste for distance: we find that students are willing to pay up to 36 BRL more per term to compensate for a reduction of 1 km in distance. Another noteworthy coefficient is the interaction $VA_j\times$ ENEM grade, which indicates that high-ability students value quality more than low-ability students. We believe this is an important finding for our study; because of the merit-based nature of the subsidies, targeted students are likely to be of higher ability. Thus, subsidies, on average, are shifting the demand of high-ability students who generally have a higher willingness to pay for quality. Table 9 shows the baseline coefficient for quality. In the baseline, students are willing to pay up to 1,478 BRL to compensate for a one standard deviation increase in value-added. Finally, Table 10 shows that for-profits have a smaller marginal cost associated with value-added than nonprofits.

Table 8: Estimation Result of Student College Choice Model

Variable	Estimate	Standard Error
DE_{ij}^{-}	1.526	0.028
DE_{ij}^{+}	-0.931	0.027
Distance$_{ij}$	-7.440	0.082
P_{ij}	-2.049	0.131
$P_{ij}\times$ Low Income	-0.015	0.004
$P_{ij}\times$ Income	0.004	0.0001
$P_{ij}\times$ Parents HS Grads	0.020	0.004
$P_{ij}\times$ Age	-0.002	0.0002
$P_{ij}\times$ Man	-0.003	0.003
$P_{ij}\times$ Non-White	-0.021	0.003
$P_{ij}\times$ FIES Probability	0.075	0.014
$VA_j\times$ ENEM grade	0.375	0.026
$VA_j\times$ Low Income	0.088	0.033
$VA_j\times$ Income	-0.0003	0.003
$VA_j\times$ Parents HS Grads	-0.094	0.035
$VA_j\times$ Age	0.016	0.002
$VA_j\times$ Man	-0.163	0.030
$VA_j\times$ Non-White	0.137	0.028

Note: This table shows selected coefficients from a multinomial logit model estimated by maximum likelihood. Sample corresponds to 29 thousand students who took the ENEM exam in 2018 linked to their college choices in 2019. Here we show the estimates for the Rio de Janeiro market, which corresponds to the Rio de Janeiro microregion.

7 Counterfactuals

We consider two counterfactuals: decreasing the supply of scholarships by 10% or decreasing the supply of subsidized student loans by 10%. In our model, these exogenous shifts translate to a reduction in the probability of a student being subsidized in a given program via equation (4), which leads to a higher expected price (scholarships) or an increase in price sensitivity (loans) for targeted individuals. To be more precise, let N_j^d be the number of subsidies of type d allocated for program j in the status quo. The counterfactual allocation is defined as $\tilde{N}_j^d = \text{floor}(0.9 \cdot N_j^d)$, where the floor$(x)$ function returns the greatest integer less than or equal to x. In aggregate terms, this represents removing 16 thousand scholarships and 11.5 thousand loans.

Table 9: Mean Utility IV Regression

Dependent Variable:	VA_j	δ_j
IV stage	First	Second
Model:	(1)	(2)
Variables		
Instructor Wage	0.0028***	
	(0.0008)	
Major-specific Instructor Wage	0.0048**	
	(0.0019)	
Graduate Degree Wage	0.0164***	
	(0.0016)	
# of Competitiors	-0.0020***	
	(0.0002)	
WF-Instrument Average ENEM	0.1533***	
	(0.0153)	
VA_j		3.029***
		(0.5461)
Fixed-effects		
Major	Yes	Yes
Fit statistics		
Observations	10,350	10,350
R^2	0.24848	0.27807
F-test, stat.	25.592	53.447

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table shows 2SLS estimates of the baseline coefficient VA_j in the model. We instrument value-added with: average instructor wage in a micro-region, average instructor wage of a major in a market, average wage of workers with graduate degree, number of competitors within the same major, and a Waldfogel-Fan instrument based on average ENEM score in neighboring micro-regions. Heteroskedasticity-robust standard-errors in parentheses.

Table 10: Cost Parameters Estimates

Variable	Estimate	Standard Error
Constant	3.650	0.441
Graduate Degree Wage	0.186	0.014
Instructor Wage	0.002	0.008
Major-specific Instructor	0.002	0.016
$VA\times$ for-profit	1.684	0.834
$VA\times$ nonprofits	2.423	0.813

Note: This table shows GMM estimates of cost parameters. We construct moment conditions based on the program's first-order conditions: equations (7) and (8). We instrument value added with the average income and average ENEM score in the program's micro-region. Our sample, we only consider private programs charging tuition.

7.1 Supply Side Fixed

To gain some intuition on our model estimates, we first conduct the counterfactuals described above but hold the price and quality of all programs fixed. This exercise has two main benefits: it offers some clear insights into how demand is shifted by subsidies without contamination from the supply response, and it is computationally cheap, so we are able to cover all the markets across the country.

Counterfactual Computation. Even though the supply side is fixed, our model still requires some adjustment in the demand side. In particular, a student's utility for a specific program depends on the deviation of his entry exam grade relative to the mean grade of the program. However, the model-implied mean grade of the program depends on the students' utility via their choice probabilities. Therefore, to compute this counterfactual, we search for a mean grade fixed point for each program, similarly to how one finds a price fixed point when simulating a counterfactual with the supply side. We proceed in the following iterative way.

1. Given program-specific average grade, $\overline{E}_j^t$, compute students' deviations DE_{ij}^+, DE_{ij}^-, and then choice probabilities, π_{ij}^t.
2. Given choice probabilities, update model-implied average grade:
$$\overline{E}_j^{t+1} = \frac{\sum_i \pi_{ij}^t \cdot E_i}{\sum_i \pi_{ij}^t} \tag{9}$$
3. Repeat steps 1–2 until convergence of mean grade vector.

Change in Market Shares. Table 11 shows how nationwide aggregate market shares would change, relative to the status quo, in each counterfactual. One can readily see that the effect of the subsidy counterfactual is much stronger than the loan—more than an order of magnitude. On the one hand, this result is not surprising, as subsidized student loans still require students to pay for their education, only providing better terms than the ones they may find in the private sector, while scholarships are actual transfers. On the other hand, the magnitudes suggest that if the government's ultimate goal is to increase access to higher education, it might even be more efficient on a per-dollar basis to use scholarships rather than loans. A final say on this question requires a more detailed analysis of the costs faced by the government associated with each form of subsidy. Figure 6 shows the heterogeneity masked in the aggregate numbers. We can see that there is a big mass of programs, especially nonprofits, whose share is not affected. Most private programs experience a decline in market share. However, there is substantial heterogeneity, as can be inferred from the long left tail of the distribution.

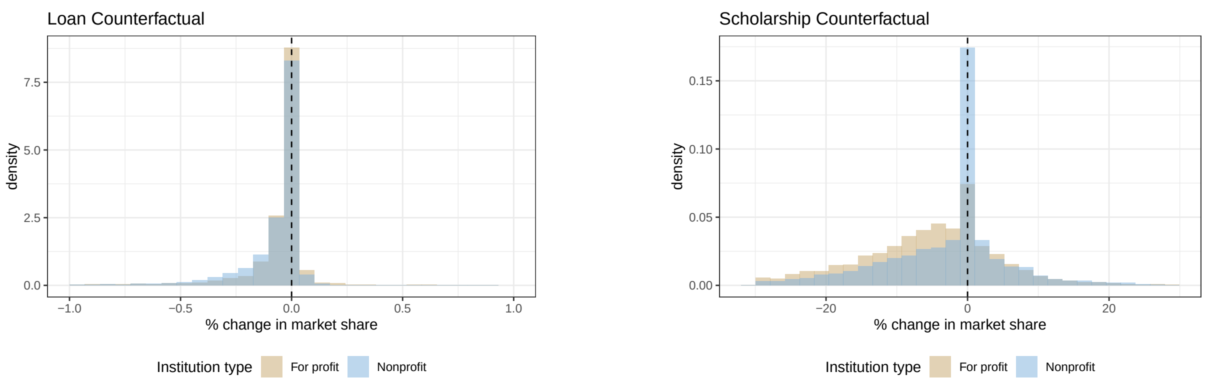
Composition of Students. Figure 7 shows the distribution of programs' characteristics and composition of students in the status quo scenario across the entire country. We split programs into two groups based on the *scholarship* counterfactual: winners and losers, i.e.,

Table 11: Change in Market Shares for Aggregate Options

Option	Loan		Scholarship	
	Δ	%Δ	Δ	%Δ
Outside Option	288	0.01	18,409	0.71
Private for-profit	-327	-0.21	-18,656	-12.08
Private nonprofit	-36	-0.06	-4,303	-7.28
Public	39	0.02	1,587	0.65

Note: This table shows how nationwide aggregate market shares would change, relative to the status quo, in each type of counterfactual: decreasing supply of loans by 10% or decreasing supply of scholarships by 10%. Here, Δ denotes our estimate for the absolute change in the number of students, while %Δ denotes the percentage change in market share.

Figure 6: Distribution of Market Share Change Across Private Programs



Note: This figure shows the distributions of the percentage change in market shares for private institutions, with the supply side fixed. The left panel depicts the loan counterfactual, whereas the right panel depicts the scholarships counterfactual.

those whose market share increased and decreased, respectively. One can readily observe a few clear patterns. Winners, on average, have students of lower ability, higher income, and a higher proportion of women than losers. On the other hand, there is no discernible difference in terms of value-added. Next, Figure 8 shows how the composition of programs changes with the scholarship counterfactual. In particular, we can see that the average ability of students would increase in most programs, while the average income would increase mostly for winners. Finally, table 12 shows that, in most markets, average ability would increase and average income would decrease for students in the outside option.

We interpret this collection of findings as suggesting that the decline in scholarships sets into place a reshuffling of students in the higher education sector. Students who are targeted in the counterfactual are of slightly higher ability and lower income than the typical student in the program they initially enrolled in. As we decrease the offer of scholarships, these targeted students then mostly move out of the market or move to cheaper options, which initially had students with low ability. These lower-ability students in cheaper options then either leave the higher education sector for the outside option if they are of low income or then move to other programs that are even cheaper and with lower-ability students and continue the chain. Thus, overall, we observe an increase in average ability across most programs and the outside options.

Table 12: Distribution of Percentage Change in Income and Ability Under Scholarships Counterfactual for Aggregate Options

Option	P10	P25	P50	P75	P90
		Percentage change in Average Income			
Outside Option	-0.21	-0.164	-0.10	-0.01	0.045
Priv. for-profit	-1.52	-0.887	-0.31	0.16	0.698
Priv. nonprofit	-1.47	-0.865	-0.16	0.14	0.369
Public	0.00	0.019	0.12	0.25	0.410
		Percentage change in Average ENEM			
Outside Option	1.284	2.54	5.2	13.0	20.9
Priv. for-profit	0.368	0.95	2.0	3.1	4.8
Priv. nonprofit	0.290	0.87	1.8	3.1	4.9
Public	0.023	1.22	2.9	6.3	16.2

Note: This table shows percentiles of the percentage change in average income and ability across aggregated options.

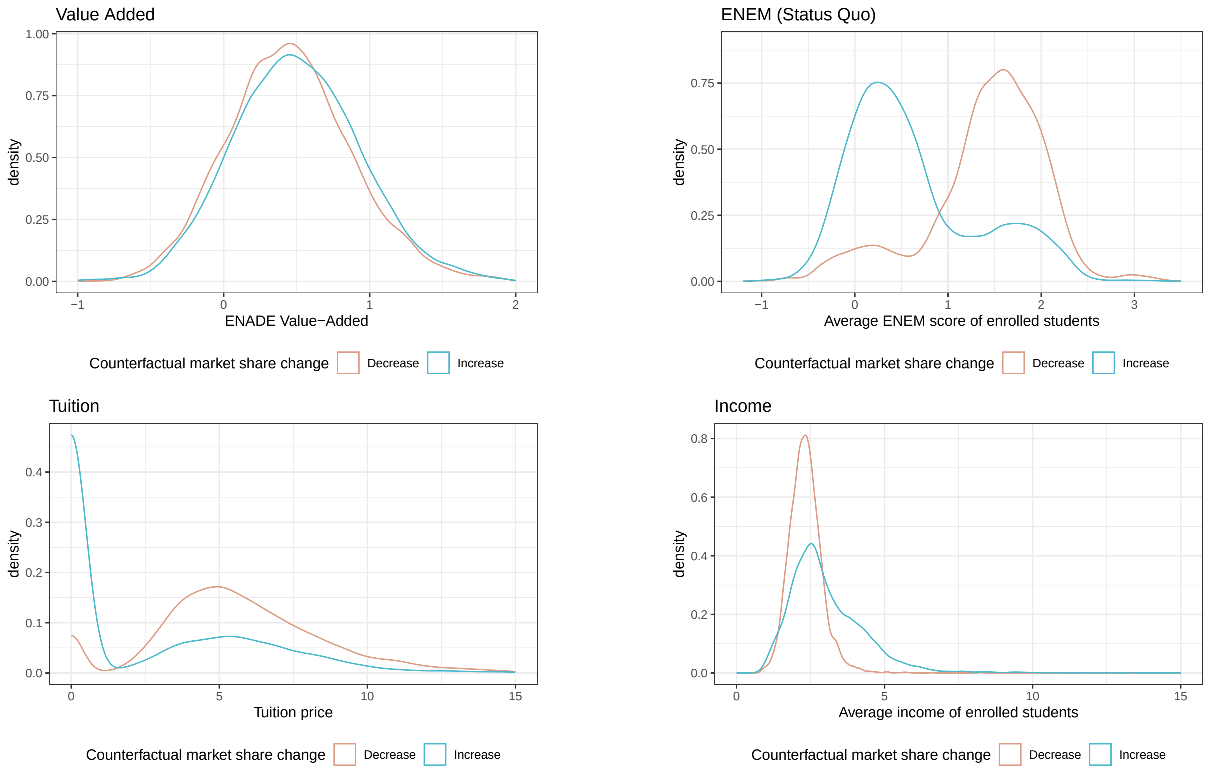
7.2 Supply Side Response

Now, we move to the counterfactual simulations with supply side response. That is, we now search for a fixed point in terms of price and quality for all private institutions. We keep public institutions fixed, as we assume they do not respond to market forces when determining their level of quality. Furthermore, public institutions never charge tuition, so their price is always zero.

Counterfactual Computation. With programs choosing price and quality, our simulation is more challenging as there are three fixed points to search: prices, quality, and mean grade. To do so, we solve the system of equations defined by equations (7) and (8) using a standard nonlinear solver. Each time the objective function is evaluated, we run an inner loop to find the program-specific average grade fixed point as described in the previous section. Due to computation constraints, we will only report the results for the Rio de Janeiro market, which corresponds to the Rio de Janeiro micro-region. This is the second-largest micro-region in our setting and a good representative example of markets in the southeast and south of the country.

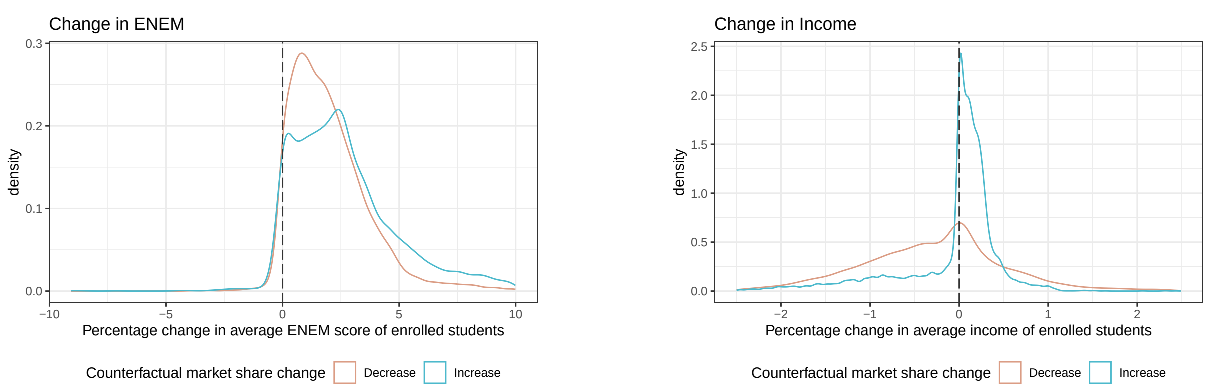
Change in Market Shares. The supply-side response induces some interesting substi-

Figure 7: Winners vs. Losers: Distribution of Program Characteristics



Note: This figure shows the distributions of programs' characteristics and composition of their students. Distributions correspond to values in the status quo. Programs are divided into two groups based on the *scholarship* counterfactual: winners, i.e., those who gained market share, or losers, i.e., those who lost market share. On average, winners have students of lower ability, higher income, higher tuition, than losers.

Figure 8: Distribution of Change in Average Income and Ability



Note: This figure shows the distributions of the change in programs' average income and average student ability, for the *scholarship* counterfactual. Programs are divided into two groups: winners, i.e., those who gained market share, or losers, i.e., those who lost market share.

tution patterns, especially when compared to the counterfactual without price and quality adjustment. Table 13 shows the percentage changes in market shares for aggregate options, where one can see three patterns that remain similar to the counterfactual without price and quality adjustment: (1) magnitudes are much larger for the scholarship counterfactual than for the loan counterfactual; (2) large number of students leaving for-profit institutions; and (3) a vast number of students leaving higher education for the outside option. Nonetheless, there is one key difference: in the scholarships counterfactual, nonprofit institutions would lose significantly less market share. As we will discuss in more detail later, under this counterfactual, a sizeable share of nonprofit institutions would cut prices more aggressively than for-profits. Figure 9 shows the distribution of change in market shares and elucidates the divide between pro-profit and nonprofits we mentioned earlier. One can see that a fairly large share of nonprofits would even gain market share under the scholarships counterfactual.

Table 13: Change in Market Shares of Aggregate Options

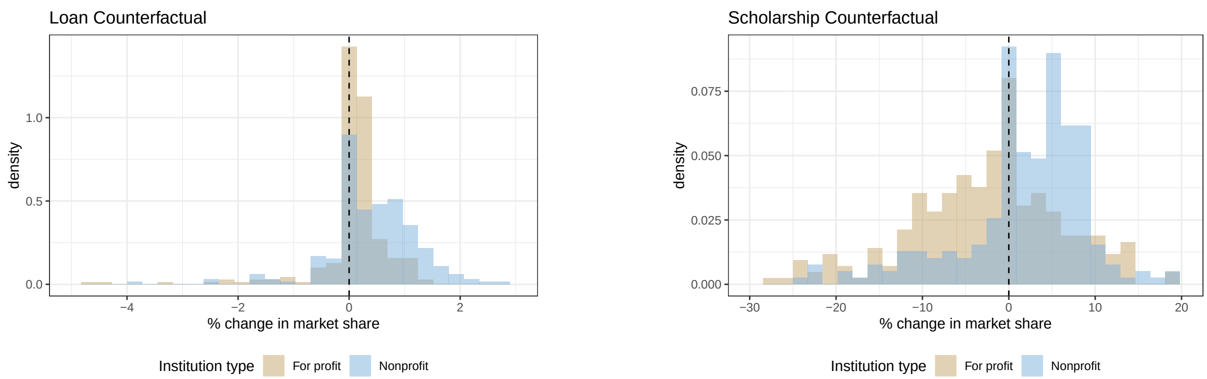
	Supply Side Fixed				Supply Side Response			
	Loan		Scholarship		Loan		Scholarship	
Option	Δ^*	$\%\Delta$	Δ^*	$\%\Delta$	Δ^*	$\%\Delta$	Δ^*	$\%\Delta$
Outside Option	31	0.001	15,232	0.58	-1,129	-0.043	12,646	0.48
Private for-profit	-48	-0.031	-15,071	-9.63	-3	-0.002	-14,245	-9.10
Private nonprofit	7	0.006	-8,666	-7.40	993	0.848	-928	-0.79
Public	2	0.001	735	0.39	-121	-0.065	459	0.25

Note: This table shows how aggregate market shares for the *Rio de Janeiro Market* would change, relative to the status quo, in each type of counterfactual: decreasing supply of loans by 10% or decreasing supply of scholarships by 10%. Here, $\%\Delta$ denotes the percentage change in market share, while Δ^* denotes our estimate for the absolute change in the number of students if this change was applied to the *entire country*.

Change in Price and Quality. Now, we explore how programs adjust price and quality in response to the counterfactual reduction in scholarships. In the left panel of Figure 10, we show the distribution of the percentage change in value-added[12]. About 82% of programs would reduce value-added, with the percentage change ranging from -50% to +10%; the median percentage change is -5%. Even though the percentage change is fairly large, in absolute terms, the change is not substantial. To be more precise, the median change in value-added would be -0.02 standard deviations. The right panel of Figure 12 shows the distribution of price changes. Similarly to the change in quality, most programs would decrease price. About 83% of programs would reduce price, with a median change of -0.7%. In absolute terms, the median reduction is equal to -0.035 (thousands of Brazilian Reais). Table 14 shows more details of selected percentiles of the distributions discussed here.

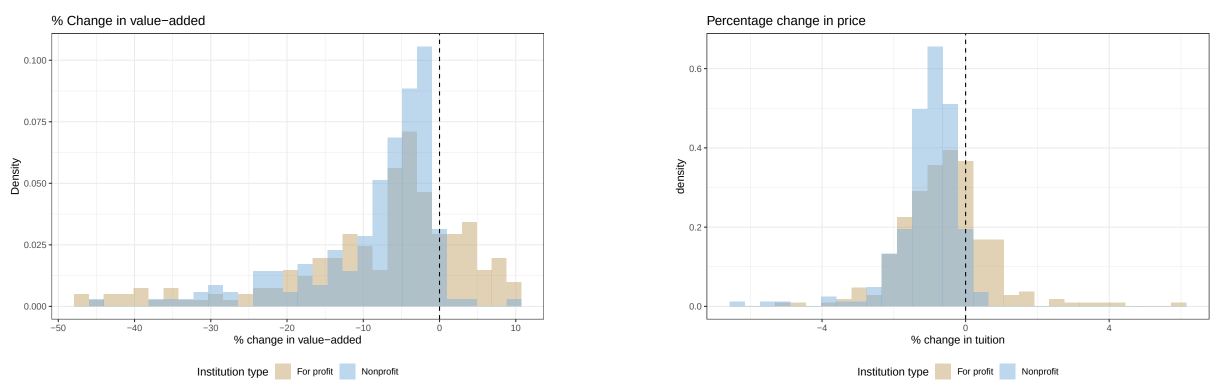
[12] Because VA takes negative and positive values, we define the percentage change in VA as $(VA_{Counterfacutal} - VA_{StatusQuo})/|VA_{StatusQuo}|$.

Figure 9: Distribution of Market Share Change Across Private Programs



Note: This figure shows the distributions of the percentage change in market shares for private institutions, with the supply side fixed. The left panel depicts the loan counterfactual, whereas the right panel depicts the scholarships counterfactual. Numbers based on the *Rio de Janeiro* market.

Figure 10: Distribution of Price and Quality Change under Scholarship Counterfactual



Note: This figure shows the distribution across private programs of the percentage change in value-added and price under the *scholarship* counterfactual. The percentage change in value-added is defined as $(VA_{Counterfacutal} - VA_{StatusQuo})/|VA_{StatusQuo}|$

In absolute terms, the magnitudes of the effects are not large, especially for price. Nonetheless, these figures are reasonable, as the counterfactual change in scholarships we induce is small by design. A 10% decline in the supply of scholarships for most programs does not translate to large decreases in the probability of obtaining a subsidy. Furthermore, only a small fraction of the population of students is really targeted by the subsidies. Thus, in this counterfactual, we are moving a rather small fraction of the market. We designed such a counterfactual because we believe the model's predictions are more credible in a neighborhood of the status quo equilibrium.

To understand how the counterfactual change in value-added correlates with other variables, we regress it on program-level outcomes and characteristics. Table 15 shows the regression results. Change in value-added is strongly positively correlated with change in price, suggesting that firms that decrease quality do so in conjunction with price re-

Table 14: Summary of Change in Price and Quality

Variable	Mean	SD	Percentile 10th	25th	50th	75th	90th
VA Status Quo	0.138	0.521	-0.444	-0.208	0.053	0.4720	0.776
VA Counterfactual	0.123	0.535	-0.467	-0.232	0.024	0.4525	0.750
Price Status Quo	7.391	5.966	2.438	3.505	5.528	9.2354	15.034
Price Counterfactual	7.358	5.972	2.388	3.451	5.465	9.1910	14.984
%Δ Price	-0.746	1.115	-1.898	-1.284	-0.745	-0.2114	0.303
Δ Price	-0.029	0.085	-0.083	-0.055	-0.035	-0.0159	0.033
%Δ VA	-9.451	16.764	-29.356	-13.612	-5.072	-1.5087	4.290
Δ VA	-0.013	0.047	-0.045	-0.030	-0.016	-0.0075	0.020

Note: This table shows summaries of distributions related to change in price and quality in the scholarship counterfactual. The top panel shows the distribution of the levels of price and quality, before and after the counterfactual. The bottom panels show the distributions of the changes in price and quality, both in percentage and absolute terms.

ductions. Such behavior is connected with our supply model, where value added shifts marginal cost. Thus, when a program faces pressure to reduce price in response to a demand shift, it may make sense to reduce quality to some extent to lessen the impact on margins. Moreover, nonprofits are expected to have a higher level of change in quality, as well as programs that initially had a high-quality level.

Composition of Students. Figure 11 and Table 23 suggest that students respond to the counterfactual policy by substituting across options in a similar pattern as in the supply-side fixed counterfactual such that the effects in terms of composition of students are in line with the previous section. That is to say, programs that lose market share have students who are on average of higher ability and lower income. The average ability of students who do not enroll in college—the outside options—increases substantially, whereas their average income decreases, suggesting that students of relatively high ability and low income exit the higher education sector. Similarly, we also observe the average ability of programs increasing across the board. Thus, we interpret that a rearrangement of students takes place here in a similar manner to the previous counterfactual without supply-side adjustment. Students affected by the reduced likelihood of receiving a subsidy—typically those with slightly higher abilities and lower incomes than average—might opt out of college enrollment or switch to less expensive programs with a generally lower ability cohort, potentially triggering a cascade of similar decisions.

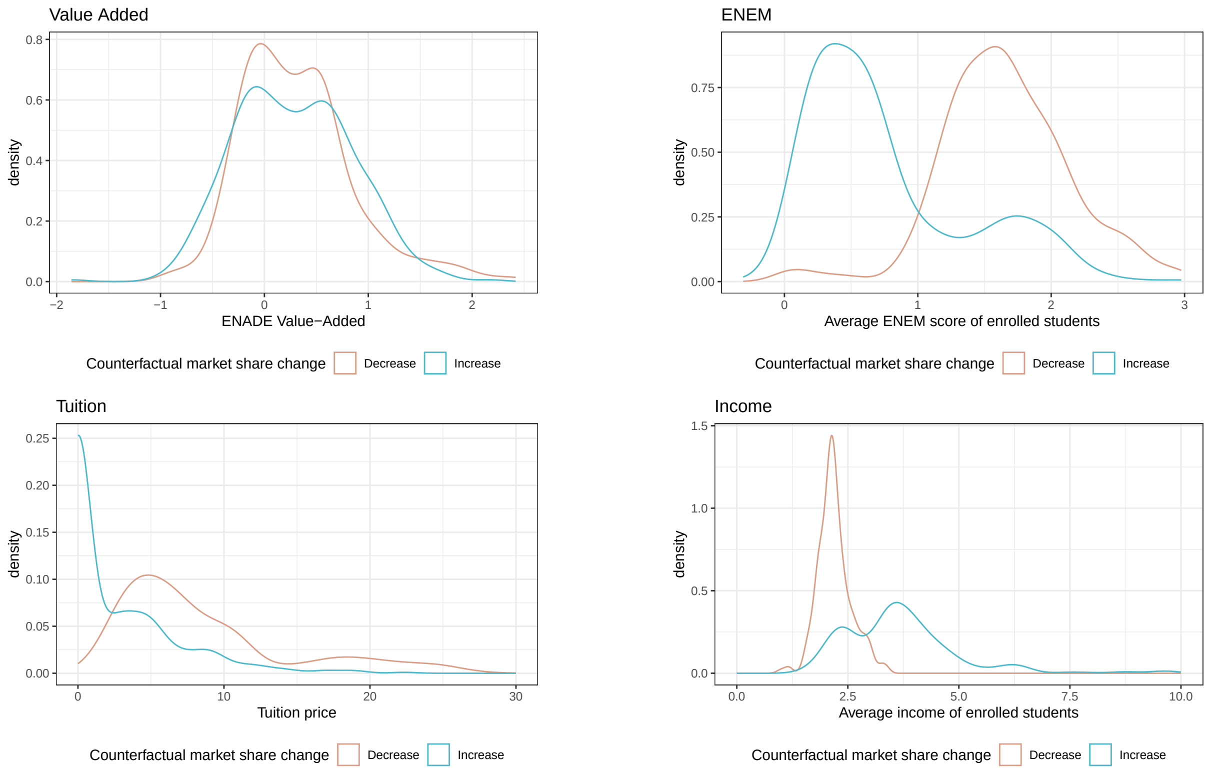
Table 15: Counterfactual Change in Value-Added: Regression on Program Characteristics

Dependent Variable:	$\%\Delta VA_j$
Model:	(1)
Variables	
Constant	-4.220***
	(0.9644)
Priv. Nonprofit	1.940
	(1.189)
VA_j Status Quo	2.747***
	(0.9277)
$\%\Delta P_j$	5.728***
	(0.5871)
Fit statistics	
Observations	405
R^2	0.31265
Adjusted R^2	0.30751

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

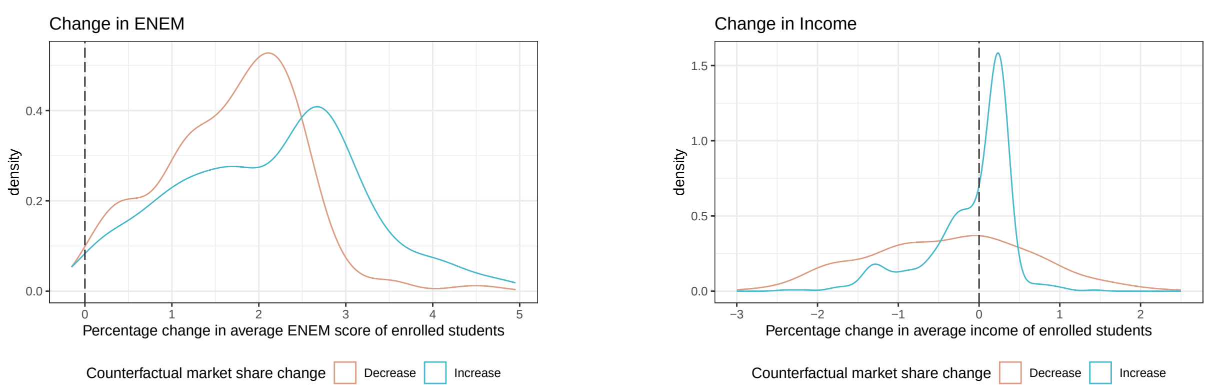
Note: This table shows the results of regressing the percentage change in value-added under the *scholarship* counterfactual on several controls. Heteroskedasticity-robust standard-errors in parentheses.

Figure 11: Winners vs. Losers: Distribution of Program Characteristics



Note: This figure shows the distributions of programs' characteristics and composition of their students. Distributions correspond to values in the status quo. Programs are divided into two groups based on the *scholarship* counterfactual: winners, i.e., those who gained market share, or losers, i.e., those who lost market share. On average, winners have students of lower ability, higher income, higher tuition, than losers.

Figure 12: Distribution of Change in Average Income and Ability



Note: This figure shows the distributions of the change in programs' average income and average student ability, for the *scholarship* counterfactual. Programs are divided into two groups: winners, i.e., those who gained market share, or losers, i.e., those who lost market share.

Student Welfare. To understand how students are affected by the counterfactual, we compute its effect on consumer surplus. More precisely, we follow Small and Rosen (1981) and compute the change in expected consumer surplus for each student according to

$$\Delta E(CS_i) = \frac{1}{|\beta_i^P|}\left[\log\left(\sum_j e^{V_{ij}^{\text{Counterfactual}}}\right) - \log\left(\sum_j e^{V_{ij}^{\text{StatusQuo}}}\right)\right]. \tag{10}$$

We assume throughout that the marginal utility of income is constant, so this welfare measure can be interpreted as a form of compensating variation, i.e., the monetary amount a consumer would need to be just indifferent between the equilibrium in the status quo and the counterfactual equilibrium.

The first row in Table 16 shows the distribution of change in consumer surplus across students for the *scholarship* counterfactual. Here we represent the effects as percentages of the student's average price, which is computed as the weighted average full price faced by each student with choice probabilities as weights. the While most students experience a marginal increase in consumer surplus—approximately 78% of students—we find significant dispersion across students. To characterize what type of students benefit or not from the reduction in scholarships, we define students targeted by the policy in the following way. First, we define the student's 90th percentile scholarship probability across all programs that accept scholarships.

$$S_i^{P90} := \inf\{x : \widehat{F}_i(x) \geq 0.9\},$$

where $\widehat{F}_i$ is an empirical CDF of student i's scholarship probabilities among institutions that offer scholarships. Then, we define a targeted student as someone who has such a probability above 50%.

$$\text{Target}_i := 1\{S_i^{P90} > 0.5\}.$$

Intuitively, this student has a fair chance of getting a scholarship in at least 10% of the programs offering such subsidy. This definition implies that about 14% of students in the Rio de Janeiro market are targeted. Figure 17 in the Appendix shows the distribution of S_i^{P90} among students in our sample. Going back to Table 16, one can see that targeted students experience welfare changes in higher magnitudes than the non-targeted population. We still find that about 80% of students experience an increase in welfare; however, the reasons for such an increase differ between the targeted and non-targeted populations.

We hypothesize that targeted students who benefit from the reduction in scholarships are top-ranked students among their targeted peers and were not the marginal students who switched courses. The targeted students whose welfare decreases, on the other hand, are the marginal students affected by the counterfactual. As an illustrative example, consider a program offering 10 scholarships. Effectively, the 10 students with the highest probability of getting a scholarship are the ones who would enroll in the program. In the counterfactual, this program will lose 1 scholarship, affecting the student at the bottom of the ranking. The other 9 students, however, will still enjoy their scholarship and probably lower prices in other options as well.

To substantiate our hypothesis, we move to Table 17, which shows the average characteristics of students belonging to different groups. For this comparison, we divide students into four groups. The first division is between targeted and non-targeted students. Within each of these categories, we further split students between those who lost welfare and those who gained. One can clearly see a couple of stark differences. First, among targeted students, those whose welfare increased had higher ENEM scores than those whose welfare decreased; the trend is reversed among non-targeted students, as those who gained had significantly smaller ENEM scores. Second, among targeted students, those who gained welfare had a much greater S_i^{P90} than those who lost welfare, while the opposite is true for non-targeted students. Third, we find a similar pattern for low-income indicator, as well.

Table 16: Change in Consumer Surplus: Distribution Moments and Percentiles

			Percentile					
Sample	Mean	SD	10%	25%	50%	75%	90%	CDF(0)
Full	3.36	13.66	-0.30	0.02	0.16	0.33	0.71	0.22
Targeted	23.79	29.43	-0.83	0.37	8.87	43.09	71.04	0.21
Non-Targeted	0.09	0.31	-0.25	0.02	0.13	0.27	0.41	0.22

Note: This table shows the distribution of change in consumer surplus associated with the scholarship counterfactual for different groups of students. Here we represent the effects as percentages of the student's average price, which is computed as the weighted average full price faced by each student with choice probabilities as weights.

Table 17: Change in Consumer Surplus: Heterogeneity

	Targeted			Non-targeted		
	$\Delta CS_i > 0$	$\Delta CS_i < 0$	Diff.	$\Delta CS_i > 0$	$\Delta CS_i < 0$	Diff.
Low Income	0.388	0.569	-0.182	0.488	0.370	0.119
			(0.009)			(0.004)
Avg. Distance	0.126	0.148	-0.022	0.147	0.124	0.023
			(0.002)			(0.001)
ENEM Score	1.591	0.870	0.722	-0.241	0.894	-1.135
			(0.007)			(0.005)
Parents HS Grads	0.775	0.671	0.104	0.627	0.720	-0.093
			(0.009)			(0.003)
Non-White	0.556	0.698	-0.142	0.570	0.542	0.028
			(0.009)			(0.004)
Prouni Prob. P90	68.187	54.563	13.624	7.469	35.226	-27.757
			(0.113)			(0.065)
Share (%)	11%	3%		67%	19%	

Note: This table shows the average characteristic of students from four mutually exclusive subgroups, with rows indicating characteristics and columns subgroups. First we divide students between targeted and non-targeted, and then split them between those who experienced an increase in consumer surplus and those who experienced a decrease. For each characteristic, we report the average value for each subgroup. We report the difference between the mean among those who gained CS and those who lost CS in the columns named "Diff". We also report the standard error associated with this difference in means, which we estimate as the robust standard error of regressing the characteristic variable on an indicator of positive change in CS.

8 Conclusion

In this paper, we have conducted a thorough investigation of quality in the higher education sector in Brazil. Using rich individual-level data, we devised a quality measure for undergraduate programs in Brazil. After extensive validation, we argue that our measure is a reasonable way to capture essential features typically attributed to quality. Next, we developed a model with demand and supply forces to understand how the equilibrium levels of price and quality are determined in the market. We found that a marginal decrease in scholarships would lead to fewer students enrolled in college and incentivize programs to reduce price and quality at the margin.

Taken together, the evidence presented in this paper paints a nuanced picture of the higher education sector. We conclude that a targeted and merit-based subsidy scheme might be used to foster marginal growth in quality by selecting high-ability students who are invested in the quality of their education. Furthermore, our estimates also suggest that, conditional on their resources, private institutions might be more efficient than public institutions in producing quality. Thus, channeling students who could not access public institutions to private programs might be a cost-effective approach to increase access while keeping quality under control.

We hope the framework developed in this paper can be used as a springboard for technical analysis of policy proposals in this market. In particular, future research avenues include investigating how policies that assign subsidies to programs according to rules based on quality can affect market equilibrium. Another important matter is to study how the pricing of public institutions would interact with the vigorous private market, possibly considering income-based price discrimination. Finally, an important next step in this framework is to allow value added to be determined by inputs chosen by the program and by the student composition of the program.

References

Allende, C. (2021): "Competition Under Social Interactions and the Design of Education Policies," Working Paper.

Armona, L., and S. Cao (2022): "Redesigning Federal Student Aid in Sub-Baccalaureate Education," Working Paper.

Arrow, K. J. (1973): "Higher education as a filter," *Journal of Public Economics*, 2(3), 193–216.

Barahona, N., C. Dobbin, H. Ho, S. Otero, and C. Yannelis (2021): "Skin in the Game: Colleges' Financial Incentives and Student Outcomes," Working Paper.

Barrett, G. F., and S. G. Donald (2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, 71(1), 71–104.

Bau, N. (2022): "Estimating an Equilibrium Model of Horizontal Competition in Education," *Journal of Political Economy*, 130(7), 1717–1764.

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25(2), 242–262.

Berry, S. T., and P. A. Haile (2021): "Chapter 1 - Foundations of demand estimation," in *Handbook of Industrial Organization, Volume 4*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, vol. 4 of *Handbook of Industrial Organization*, pp. 1–62. Elsevier.

Bodéré, P. (2022): "Dynamic Spatial Competition in Early Education: an Equilibrium Analysis of the Preschool Market in Pennsylvania," Working Paper.

Bó, I., and R. Hakimov (2022): "The iterative deferred acceptance mechanism," *Games and Economic Behavior*, 135, 411–433.

Cellini, S. R., and C. Goldin (2014): "Does Federal Student Aid Raise Tuition? New Evidence on For-Profit Colleges," *American Economic Journal: Economic Policy*, 6(4), 174–206.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104(9), 2633–79.

Dobbin, C., N. Barahona, and S. Otero (2022): "The Equilibrium Effects of Subsidized Student Loans," Working Paper.

Hendricks, L., C. Herrington, and T. Schoellman (2021): "College quality and attendance patterns: A long-run view," *American Economic Journal: Macroeconomics*, 13(1), 184–215.

Kapor, A., M. Karnani, and C. Neilson (2023): "Aftermarket Frictions and the Cost of Off-Platform Options in Centralized Assignment Mechanisms," *Journal of Political Economy, (Forthcoming)*.

Kelchen, R. (2019): "An empirical examination of the Bennett hypothesis in law school prices," *Economics of Education Review*, 73, 101915.

Koedel, C., K. Mihaly, and J. E. Rockoff (2015): "Value-added modeling: A review," *Economics of Education Review*, 47, 180–195.

Kremer, M. (1993): "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, 108(3), 551–575.

Lau, C. (2020): "The Effect of Tuition-Free Community College," FTC Working Paper No. 347.

Lucca, D. O., T. Nadauld, and K. Shen (2018): "Credit Supply and the Rise in College Tuition: Evidence from the Expansion in Federal Student Aid Programs," *The Review of Financial Studies*, 32(2), 423–466.

Manski, C. F., and S. R. Lerman (1977): "The Estimation of Choice Probabilities from Choice Based Samples," *Econometrica*, 45(8), 1977–1988.

Martellini, P., T. Schoellman, and J. Sockin (2022): "The Global Distribution of College Graduate Quality," *Federal Reserve Bank of Minneapolis Working Paper*, 791.

Mountjoy, J., and B. R. Hickman (2021): "The Returns to College(s): Relative Value-Added and Match Effects in Higher Education," Working Paper 29276, National Bureau of Economic Research.

Neilson, C. (2021): "Targeted Vouchers, Competition Among Schools and the Academic Achievement of Poor Students," Working Paper.

Otero, S., N. Barahona, and C. Dobbin (2023): "Affirmative Action in Centralized College Admission Systems," Working Paper.

Small, K. A., and H. S. Rosen (1981): "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 49(1), 105–130.

Turner, L. J. (2017): "The Economic Incidence of Federal Student Grant Aid," Working Paper.

Appendix

A Value-Added

Table 18: Income vs. Exam value-added

Dependent Variable:	Exam Value-Added (ENADE)			
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	48.45***		49.05***	
	(0.0962)		(0.2201)	
Income Value-Added (RAIS)	0.1187***	0.1285***	0.0152	0.2383***
	(0.0074)	(0.0263)	(0.0195)	(0.0306)
Fixed-effects				
Major	No	Yes	No	Yes
Fit statistics				
Observations	9,902	9,902	2,738	2,738
R^2	0.02503	0.51421	0.00022	0.52580
Within R^2		0.02804		0.06017

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

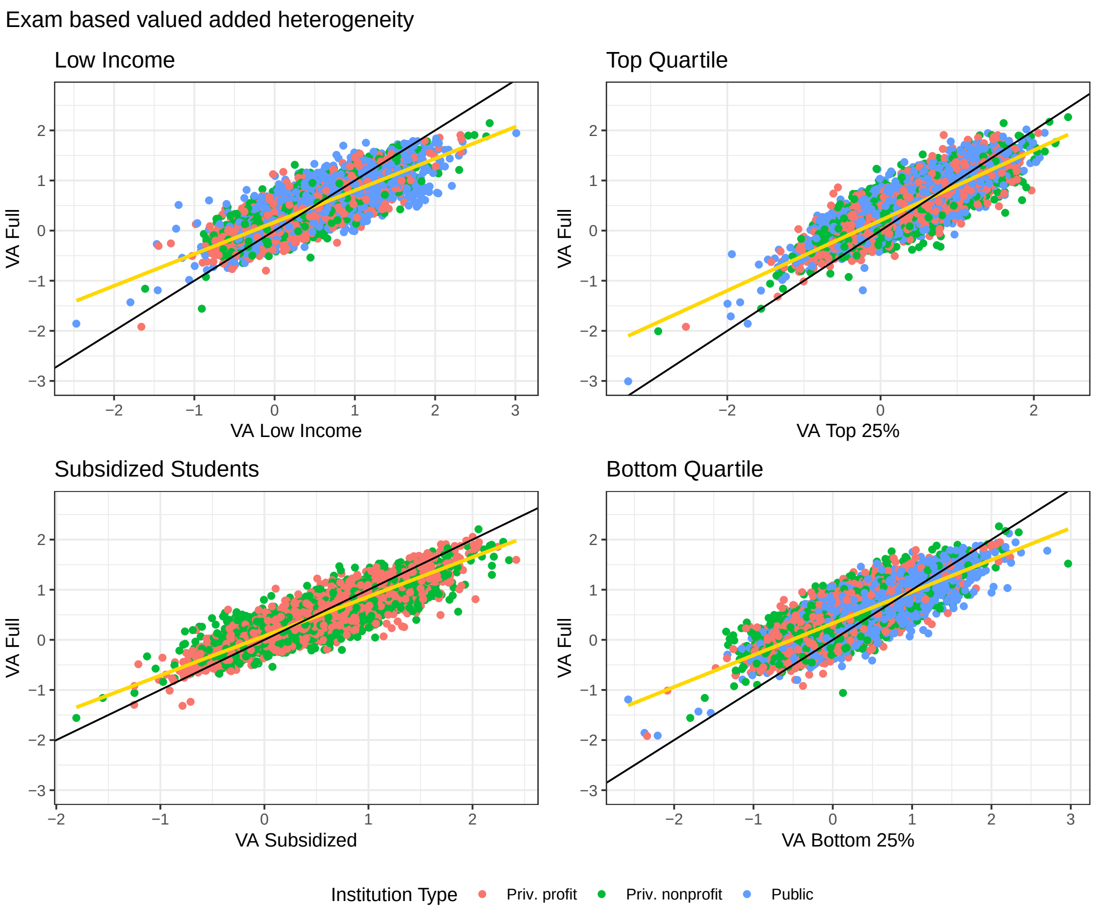
Note: This table shows the result of regressing ENADE value-added (exit exam) on income value-added. Both value-added measures were estimated with outcomes *not* standardized by major. This table is the analogue of table 19 with non-standardized outcomes. We consider two samples: full sample and a subsample containing only programs with high coverage in the RAIS database.

Table 19: Heterogeneity in Exam Value-Added

Dependent Variables:	VA Full Sample				VA Top Quartile
Subgroup:	Low-income	Subsidized	Top Quartile	Bottom Quartile	Bottom Quartile
	(1)	(2)	(3)	(4)	(5)
Constant	0.1675***	0.0751***	0.2038***	0.3282***	0.2773***
	(0.0040)	(0.0027)	(0.0026)	(0.0026)	(0.0046)
Slope	0.6346***	0.7868***	0.7001***	0.6342***	0.4690***
	(0.0063)	(0.0049)	(0.0044)	(0.0048)	(0.0085)
Statistics					
Observations	7,304	8,538	13,098	11,420	10,410
R^2	0.655	0.814	0.700	0.637	0.245
Pearson's correlation	0.810	0.902	0.837	0.798	0.495

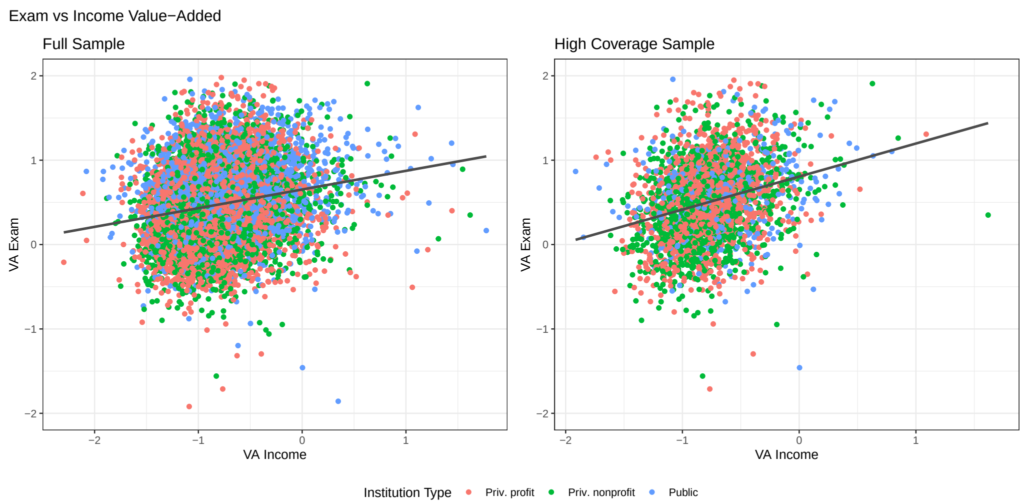
Note: This table shows the results of regressing value-added computed on the full sample of students on value-added measures computed on subsamples: $VA_j = \text{Constant} + \text{Slop} \cdot VA_j^{\text{Subgroup}}$. Each columns represents a different subgroup. Heteroskedasticity-robust standard-errors in parentheses.

Figure 13: Heterogeneity in Exit Exam Value-Added: Scatter Plots



Note: Here we show scatter plots representing regressions shown in table 19. The solid black lines represents the 45 degree line, while the yellow line represents the linear regression fit.

Figure 14: Income vs. Exam Value Added



Note: Here we show scatter plots representing regressions shown in table 5.

B Fitting Income Distribution

We use the granular structure of income brackets reported by students in ENEM to estimate a continuous distribution of income. Let X_i denote the (continuous) income of student i. Instead of observing X_i, we only observe the student's income bracket report, which we will assume is truthfully reported. Suppose there are M income brackets defined as $(0, B_1], (B_1, B_2], \dots, (B_{k-1}, B_k], \dots, (B_{M-1}, +\infty)$. We will assume that income is distributed according to a log-normal: $X_i \sim e^{\mu+\sigma Z_i}$, where $Z_i \sim N(0,1)$. Thus, the probability of observing income bracket $(B_{k-1}, B_k]$ is given by

$$\begin{aligned} P\left(X_i \in (B_{k-1}, B_k]\right) &= P\left(B_{k-1} < e^{\mu+\sigma Z_i} \leq B_k\right) \\ &= \Phi\left(\frac{\log B_{k(i)} - \mu}{\sigma}\right) - \Phi\left(\frac{\log B_{k(i)-1} - \mu}{\sigma}\right) \end{aligned}$$

Where $B_{k(i)}$ denotes the cutoff of the income bracket reported by student i. We employ a slight abuse of notation to use the convention that endpoints are covered in the notation above. Thus, $B_0 = 0$ and $B_M = +\infty$, so that

$$\Phi\left(\frac{\log B_0 - \mu}{\sigma}\right) = 0 \qquad \text{and} \qquad \Phi\left(\frac{\log B_M - \mu}{\sigma}\right) = 1.$$

We assume students' income are iid draws, which yields the following log-likelihood function:

$$\ell(\mu, \sigma) = \sum_{i=1}^{N} \log\left(\Phi\left(\frac{\log B_{k(i)} - \mu}{\sigma}\right) - \Phi\left(\frac{\log B_{k(i)-1} - \mu}{\sigma}\right)\right)$$

We use the sample of ENEM takers who reported non-zero income from 2018 with 3.5 million students to recover the following estimates: $\hat{\mu} = 0.532$ and $\hat{\sigma} = 0.966$. For each student in our sample, we draw a continuous measure of income conditional on his reported income bracket. In particular, if student i reports bracket $(B_{k-1}, B_k]$, then we draw income X_i from a Log-Normal distribution with parameters $\hat{\mu}$ and $\hat{\sigma}$, conditional on $X_i \in (B_{k-1}, B_k]$.

C College Choice Estimation

To estimate the model described in section 5.2, we augment it with two variables:

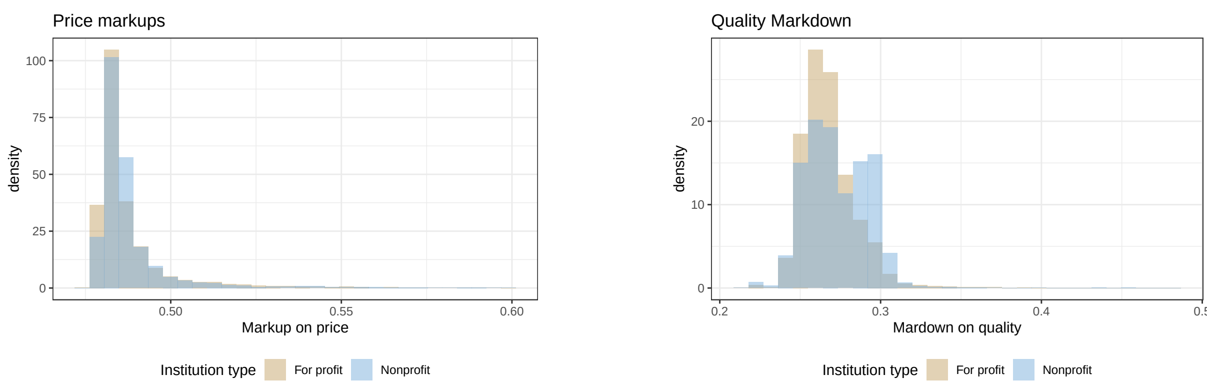
$$\begin{aligned} U_{ij} = \delta_j + \beta^{E+} DE^{+}_{ij} + \beta^{E-} DE^{-}_{ij} + \beta^{D}_{i}\mathrm{Dist}_{ij} + \beta^{V}_{i} VA_j + \beta^{P}_{i} P_{ij} + \beta^{S}_{i} \hat{S}^{L}_{ij} \cdot P_{ij} \\ + \mathrm{OutMarket}_j \beta^{OM}_{i} + \mathrm{NoTuition}_j \beta^{NT}_{i} + \varepsilon_{ij} \end{aligned}$$

where $\mathrm{OutMarket}_j$ is an indicator for the option of attending a program outside of the market, and $\mathrm{NoTuition}_j$ is an indicator for a program without tuition data.

Sample and Estimation. In our sample, we only consider in-person programs for which we managed to compute a value-added measure. Moreover, we drop programs wich enrolled 2 or less students in 2019. For most markets, we cannot use the full sample of ENEM takers to estimate the model, as it would dramatically increase the computational cost. Thus, in our estimation, we use a random choice-based sample of students. Manski and Lerman (1977) show that, in a multinomial Logit model with alternative-specific constants (δ_j in our case), if one estimates it assuming exogenous sampling, then all parameters except the alternative-specific coefficients are consistently estimated. Thus, our estimation proceeds in two steps:

1. Estimate the multinomial Logit model assuming exogenous sampling
2. Keep the individual-specific coefficients from the first step fixed and recover alternative-specific constants via Berry (1994) contraction mapping, using the true observed market shares.

Figure 15: Price Markups and Quality Markdowns Recovered from Estimation.



Note: The left panel shows the distribution of price markups by institution type. The panel on the right shows the distribution of quality markdowns by institution type.

Table 20: Value-added models

	ENADE Exam		Income	
Variable	Estimate	SE	Estimate	SE
ENEM Science	0.073	0.002	-0.007	0.066
ENEM Humanities	0.112	0.002	0.528	0.070
ENEM Languages	0.118	0.002	0.161	0.070
ENEM Mathematics	0.070	0.002	0.543	0.058
ENEM Essay	0.012	0.002	0.110	0.050
23 <Age <= 30	-0.097	0.003	0.559	0.112
Age >30	-0.020	0.008	1.199	0.327
Age	-0.015	0.000	0.037	0.019
Distance	0.007	0.000	0.031	0.006
Race = White	0.024	0.005	0.075	0.143
Race = Black	-0.053	0.007	-0.644	0.203
Race = *Pardo*	-0.003	0.006	-0.549	0.151
Race = Asian	-0.015	0.011	-0.431	0.397
Race = Indigenous	-0.104	0.025	0.123	0.756
Race = NA	0.018	0.009	0.733	0.211
Male	0.136	0.003	2.937	0.105
High School Type = Private	-0.002	0.004	-0.036	0.127
High School Type = NA	0.001	0.007	0.704	0.164
Household Income 0-1 MWs	-0.002	0.012	-0.478	0.431
Household Income 1-2 MWs	-0.021	0.012	-0.440	0.424
Household Income 2-5 MWs	-0.030	0.012	0.143	0.431
Household Income 5-10 MWs	-0.076	0.013	0.130	0.455
Household Income 10+ MWs	-0.116	0.013	0.269	0.471
Father Education = HS	-0.082	0.007	-0.228	0.229
Father Education = Primary	-0.039	0.007	-0.124	0.215
Father Education = Unknown	-0.035	0.008	-0.239	0.259
Father Education = College+	-0.105	0.008	-0.780	0.271
Mother Education = HS	-0.069	0.008	0.156	0.240
Mother Education = Primary	-0.026	0.008	0.137	0.228
Mother Education = Unknown	-0.034	0.012	-0.016	0.382
Mother Education = College+	-0.097	0.008	0.005	0.261
Time employed			0.712	0.009
Fixed-effects				
Program	Yes		Yes	
Nobs	563170		236371	
R2	0.297		0.361	
WithinR2	0.096		0.135	
MeanY	0.173		22.069	

Note: This table shows the full specification of value-added models discussed in this paper. The dependent variable for the ENADE exam is the standardized grade of the major-specific component of the exam. The dependent variable for the income specification is the total annual income two years after graduation, measured in real minimum wages.

Market	No. Students	No. Programs	States	Meso-regions
1	147,018	507	Rondônia + Acre + Amazonas + Roraima	All meso-regions in the state(s) assigned to market
2	398,089	747	Pará + Amapá + Tocantins + Maranhão	All meso-regions in the state(s) assigned to market
3	283,892	674	Piauí + Ceará	All meso-regions in the state(s) assigned to market
4	379,099	952	Rio Grande do Norte + Paraíba + Pernambuco	All meso-regions in the state(s) assigned to market
5	110,227	325	Alagoas + Sergipe	All meso-regions in the state(s) assigned to market
6	74,366	340	Espírito Santo	All meso-regions in the state(s) assigned to market
7	77,379	651	Santa Catarina	All meso-regions in the state(s) assigned to market
8	99,116	595	Mato Grosso do Sul + Mato Grosso	All meso-regions in the state(s) assigned to market
9	182,471	817	Goiás + Distrito Federal	All meso-regions in the state(s) assigned to market
10	187,297	500	Bahia	2902 + 2904 + 2903 + 2905 + 2901
11	76,631	209	Bahia	2906 + 2907
12	96,287	441	Minas Gerais	3105 + 3104 + 3102 + 3103 + 3101
13	192,978	706	Minas Gerais	3106 + 3108 + 3109 + 3107
14	90,362	534	Minas Gerais	3112 + 3110 + 3111
15	61,622	359	Rio de Janeiro	3305 + 3301 + 3304 + 3303 + 3302 + 3306
16	70,827	498	São Paulo	3508 + 3501 + 3502 + 3503
17	97,951	791	São Paulo	3507 + 3504 + 3506 + 3505
18	213,609	879	São Paulo	3511 + 3512 + 3509 + 3513 + 3515 + 3510 + 3514
19	53,513	442	Paraná	4104 + 4102 + 4101 + 4103 + 4105
20	93,969	650	Paraná	4110 + 4107 + 4106 + 4109 + 4108
21	54,073	464	Rio Grande do Sul	4301 + 4303 + 4302 + 4304
22	96,637	496	Rio Grande do Sul	4306 + 4305 + 4307
23	179,789	726	Rio de Janeiro	*Micro-region*: 33018
24	202,898	892	São Paulo	*Micro-region*: 35061

Table 21: Summary of markets used in the model. Each market is defined as a contiguous geographical area. For regions with high population and program density, we split states into multiple markets by their meso-reions. The two most densely populated capitals in the country correspond to their micro-regions: Rio de Janeiro and São Paulo.

Variable	Loans Estimate	Loans SE	Scholarships (Partial) Estimate	Scholarships (Partial) SE	Scholarships (Full) Estimate	Scholarships (Full) SE
Constant	-2.090	0.059	-3.815	0.061	-3.924	0.042
Household Income = (0, 1] MWs	0.024	0.022	0.017	0.032	-0.040	0.022
Household Income = (1, 1.5] MWs	-0.049	0.021	0.038	0.031	-0.079	0.021
Household Income = (1.5, 2] MWs	-0.138	0.022	0.017	0.033	-0.178	0.022
Household Income = (2, 2.5] MWs	-0.166	0.022	0.030	0.032	-0.208	0.022
Household Income = (2.5, 3] MWs	-0.212	0.024	0.012	0.034	-0.332	0.024
Household Income = (3, 4] MWs	-0.227	0.023	0.042	0.033	-0.383	0.022
Household Income = (4, 5] MWs	-0.313	0.025	-0.007	0.035	-0.704	0.025
Household Income = (5, 6] MWs	-0.356	0.026	-0.058	0.036	-0.911	0.026
Household Income = (6, 7] MWs	-0.417	0.029	-0.119	0.040	-1.278	0.034
Household Income = (7, 8] MWs	-0.430	0.033	-0.138	0.045	-1.492	0.043
Household Income = (8, 9] MWs	-0.514	0.037	-0.277	0.053	-1.746	0.053
Household Income = (9, 10] MWs	-0.539	0.040	-0.423	0.063	-1.928	0.063
Household Income = (10, 12] MWs	-0.661	0.037	-0.619	0.063	-2.207	0.065
Household Income = (12, 15] MWs	-0.794	0.045	-1.005	0.105	-2.410	0.087
Household Income = (15, 20] MWs	-0.944	0.051	-1.333	0.152	-2.560	0.103
Household Income = 20+ MWs	-1.141	0.053	-1.647	0.237	-2.977	0.123
High School Type = Public	-0.232	0.007	-0.007	0.009	-0.127	0.007
High School Type = Private	-0.336	0.058	-0.100	0.061	-0.116	0.042
High School Type = Abroad	-0.248	0.011	-0.828	0.022	-0.896	0.012
Race = White	0.053	0.025	0.062	0.031	0.005	0.023
Race = Black	0.100	0.026	0.061	0.033	0.273	0.024
Race = *Pardo*	0.100	0.025	0.068	0.031	0.217	0.023
Race = Asian	0.119	0.032	-0.066	0.044	-0.027	0.031
Race = Indigenous	-0.034	0.057	-0.147	0.091	0.093	0.055
Male	-0.028	0.006	-0.030	0.008	0.050	0.006
Age	0.009	0.001	0.008	0.001	0.017	0.001
DE^d_+ Mathematics	-0.127	0.008	-0.198	0.014	0.112	0.009
DE^d_+ Sciences	-0.094	0.008	-0.144	0.015	-0.102	0.010
DE^d_+ Humanities	-0.210	0.012	-0.167	0.024	-0.142	0.016
DE^d_+ Languages	-0.144	0.011	-0.192	0.020	0.122	0.013
DE^d_+ Essay	-0.062	0.007	-0.112	0.013	0.573	0.008
DE^d_- Mathematics	0.153	0.008	0.228	0.009	0.443	0.007
DE^d_- Sciences	0.104	0.008	0.144	0.009	0.363	0.007
DE^d_- Humanities	0.149	0.007	0.146	0.009	0.317	0.008
DE^d_- Languages	0.168	0.008	0.163	0.010	0.318	0.009
DE^d_- Essay	0.258	0.008	0.365	0.010	0.716	0.010
# Loans	0.002	0.000				
# Scholarships (Partial)			0.006	0.001		
# Scholarships (Full)					0.008	0.000
# Loans / Enrollment	4.109	0.086				
# Scholarships (Partial) / Enrollment			5.278	0.132		
# Scholarships (Full) / Enrollment					3.840	0.181
# Loans in Microregion	0.005	0.000				
# Scholarships (Partial) in Microregion			0.022	0.001		
# Scholarships (Full) in Microregion					0.007	0.001
Fit statistics						
No. Observations	895264		895264		895264	
Log Likelihood	-103007		-59215		-116182	
BIC	206712		119129		233063.9	
Squared Correlation	0.232		0.343		0.405	
Pseudo R2	0.358		0.482		0.491	

Table 22: This table shows selected coefficients for the subsidy probability models described in section 5.1.

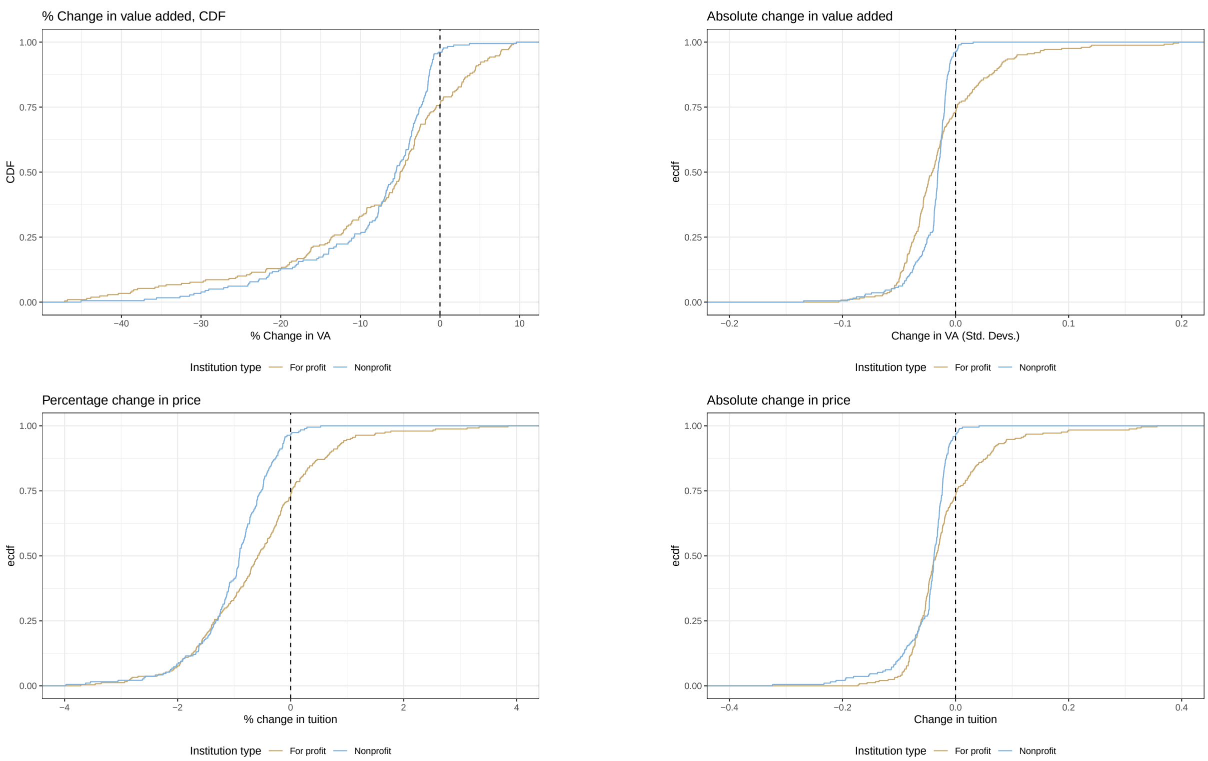
D Counterfactuals

Table 23: Distribution of Percentage Change in Income and Ability Under Scholarships Counterfactual for Aggregate Options

Option	P10	P25	P50	P75	P90
			Percentage change in Average Income		
Outside Option	—	—	-0.22	—	—
Priv. for-profit	-1.68	-1.08	-0.38	0.228	0.83
Priv. nonprofit	-1.25	-0.58	-0.24	0.034	0.16
Public	0.16	0.21	0.26	0.291	0.30
			Percentage change in Average ENEM		
Outside Option	—	—	20.5	—	—
Priv. for-profit	0.73	1.16	1.7	2.1	2.4
Priv. nonprofit	0.30	0.73	1.5	2.3	4.1
Public	2.35	2.54	2.8	3.2	4.0

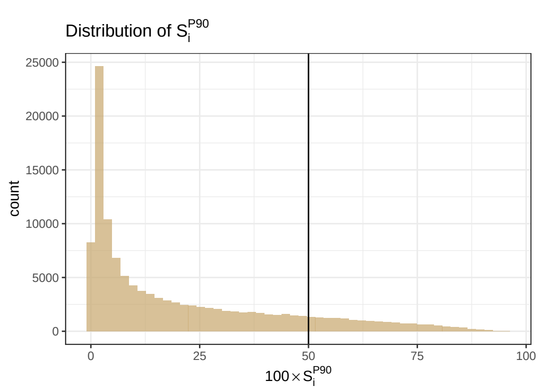
Note: This table shows the percentage change in average income and ability across aggregated options. Here we consider only the scholarship counterfactual with supply-side response for the Rio de Janeiro market. Because we only consider one market, there is no variation for the outside option.

Figure 16: CDFs of Price and Quality Change. Scholarship Counterfactual



Note: This figure shows the distributions of change in price and quality under the scholarships counterfactual with supply-side adjustment. Top panels show the distributions for price change, whereas bottom panels show the distributions for price change. Left panels show the percentage change distribution and right panels show the absolute change distribution. Simulations are based on the Rio de Janeiro Market.

Figure 17: Distribution of Scholarship P90 Probabilities



Note: The figure depicts the distribution of S_i^{P90}, i.e., the 90-th percentile of a student's scholarship probability among programs granting scholarships. We define a targeted student as someone with $S_i^{P90} > 0.5$, which correspond to the mass of students to the right of the vertical line on the graph. Numbers based on the Rio de Janeiro Market.