The Effects of Widespread Online Education on Market Structure and Enrollment*

Nano Barahona   Cauê Dobbin   Joaquín Fuenzalida   Sebastián Otero

March 9, 2025

Abstract: We examine the rapid growth of Brazil's private online higher education sector and its impact on market structure and college enrollment. Exploiting regional and field-specific variation in online education penetration, we find that online programs increase enrollment for older students but divert younger students from higher-quality in-person programs. Increased competition lowers the prices of in-person programs but leads to a decline in their provision. Using an equilibrium model of college education, we quantify that in the absence of online education, the average student would experience 3.4% higher value added. While young students benefit from fewer online options, older students are disadvantaged. Targeted policies limiting online education to older cohorts have the potential to improve value added across all groups.

Keywords: Online education, higher education, market structure.
JEL Codes: I23, I24, I26, J24, L11, L13.

*We thank Claudia Allende, José Ignacio Cuesta, Liran Einav, Matthew Gentzkow, Gautam Gowrisankaran, Caroline Hoxby, Gastón Illanes, Pat Kline, Cristóbal Otero, Kiki Pop-Eleches, Anna Popova, Paulo Somaini, and Pietro Tebaldi for helpful conversations and suggestions, and to seminar participants at Chicago Booth, Northwestern University, Stanford University, University of Tokyo, Peder Sather Conference on IO, O-Lab Conference, and the Bay Area IO Fest for useful comments. We also thank Romina Quagliotti, José Tomás Feliú, Enrico Ruggeri, and Arthur Weiss for excellent research assistance. Barahona: UC Berkeley and NBER, nanobk@berkeley.edu. Dobbin: Georgetown University, caue.dobbin@georgetown.edu. Fuenzalida: UC Berkeley, jfuenzalida@berkeley.edu. Otero: Columbia University, seb.otero@columbia.edu.

1. Introduction

Over the past two decades, rapid advances in digital technology have transformed numerous industries, expanding their offerings to include the online delivery of goods and services. Sectors traditionally reliant on physical, face-to-face interactions—such as healthcare and education—are increasingly adopting hybrid or fully virtual models of operation. This shift has been particularly pronounced in higher education, where colleges have increased their online offerings to meet the growing demand for more affordable and flexible learning (Deming et al., 2012; Aucejo et al., 2024). In 2019, around 15% of all U.S. undergraduate students were enrolled exclusively in distance education, a figure that rose to 24% by 2022, accelerated by the COVID-19 pandemic (NCES, 2022).

Expanding education—or any service—to online formats presents both opportunities and risks. On the one hand, online education can increase access for previously excluded students, democratizing access to higher education (Barrow et al., 2024). On the other hand, the shift to remote delivery can fundamentally change the nature of these services, potentially compromising their quality (Bettinger et al., 2017; Garrett et al., 2022). As a result, online education may boost overall enrollment in higher education but also risks diverting students from high-quality, in-person instruction to potentially inferior online alternatives. This is particularly problematic in higher education, where quality is difficult to assess.

These concerns are amplified when considering the equilibrium effects of online education. The entry of online programs can increase the competitive pressure on traditional in-person programs (Deming et al., 2016), initially benefiting students through lower tuition costs (Deming et al., 2015). However, because in-person programs operate with substantial fixed costs, sustained declines in enrollment and revenue can lead to program closures, ultimately reducing the availability of high-quality educational options. This risk is especially problematic in markets with limited in-person institutions, where students may be left with lower-quality online alternatives as their only choice.

This paper examines the impact of online education expansion in Brazil, the world's largest market for online higher education. We leverage plausibly exogenous variation in the differential entry of online programs across regions and fields of study to assess how the introduction of online undergraduate degrees impacts market expansion, student diversion, and in-person program availability. We integrate this analysis with an equilibrium model of college education to quantify the overall effects of online education and evaluate alternative policy designs.

Brazil provides an ideal setting for studying the expansion of online education for two reasons. First, analyzing equilibrium effects requires a setting where online education constitutes a significant share of the education market. In Brazil, fully remote programs have grown rapidly, comprising 17% of all new undergraduate enrollments in 2010, 44% by 2019, and 65% post-COVID. Second, to isolate the causal impact of online education on local market outcomes from nationwide time trends, we need variation in online penetration across local markets. Brazilian regulation provides this variation by requiring online programs to establish local physical hubs, which students need to attend periodically, generating geographic differences in online market entry. Additionally, by prohibiting online education in certain majors, the regulation introduces further variation across fields of study.

For our analysis, we use several administrative datasets. First, we draw on detailed data from the Brazilian Higher Education Census to assess market shares and track the entry and exit of degree programs. We also collect tuition fee data from various sources to examine colleges' pricing strategies. Additionally, we combine university entrance exam data with the Higher Education Census and matched employer-employee records to estimate the labor market returns of specific degrees, allowing us to compute the value added of online and in-person programs. We thus have a rich window into student demand for in-person and online degrees, colleges' behavior, and overall degree quality.

Our analysis focuses on the private sector, which accounts for 82% of incoming students between 2010 and 2019, and nearly all online programs. During this period, private sector enrollment grew from 1.7 million to 3 million students, with 89% of this growth driven by online programs, particularly at for-profit institutions. The growth of online education has been characterized by two main factors. First, it has improved access for older, lower-income adults seeking affordable and flexible learning options. While the average incoming in-person student is 24 years old, the average online student is 29 years old. Second, there has been a shift away from traditional in-person programs. Since 2010, in-person enrollment has stagnated, with significant declines in business and education programs—fields that have seen the largest increases in online enrollment.

We begin our analysis by comparing online degree programs to traditional in-person programs in terms of duration, tuition fees, dropout rates, and value added. To do this, we examine equivalent programs offered by the same institutions, differing only in delivery mode. Our findings show no significant differences in program duration, consistent with regulations mandating that both formats follow the same curriculum. We also document that tuition fees for online degrees are 57% lower than those for in-person programs. In terms of quality, online programs show lower dropout rates but are associated with a 38% lower value added, measured by students' gains in labor market outcomes after accounting for their entrance exam scores.

To examine the causal effects of online program expansion in Brazil, we use a linear model that regresses changes in outcomes—such as student enrollment, market structure, and tuition—on the change in the number of online degrees from 2010 to 2019. We define our unit of analysis as the interaction of a commuting zone and a field of study. We begin by estimating the model using OLS. The validity of this approach relies on a parallel trends assumption, which requires that outcomes in regions and fields with lower online degree entry would have followed the same trend as those in higher online degree entry areas had they experienced similar online degree entry. To address potential bias from unobserved shocks, we also implement a shift-share instrumental variable (SSIV) approach (Bartik, 1991; Goldsmith-Pinkham et al., 2020), combining predetermined institution headquarter locations (*shares*) with online sector growth (*shift*). Our instrument builds on the observation that institutions tend to expand their online programs in regions closer to their headquarters and assumes that a region's proximity to an institution is not correlated with unobserved region-specific shocks in fields where the institution offers in-person degrees (i.e., exogenous *shares*).

Both the OLS and SSIV approaches yield qualitatively consistent results: expanding online education increases online enrollment, reduces in-person enrollment, and raises overall college enrollment. This highlights the dual effect of online education. On one hand, it expands access for students who might not otherwise have attended college. On the other hand, it pulls students away from in-person programs, often toward lower-quality online alternatives. This shift heightens competition, forcing local in-person institutions to lower prices and reduce profits. As competition intensifies, in-person programs are less likely to persist in the market, accelerating the shift toward online degrees. As a result, the value added attained by the average student declines. We then present an analysis for different age cohorts, showing that this expansion primarily benefits older students by increasing their college enrollment and value added. Younger students, who are more likely to enroll in in-person programs initially, experience a stronger diversion to online options and a reduction in value added.

Estimating the linear model provides a transparent method for recovering marginal causal effects, but it rests on a strong no-interference assumption, which requires that changes in the number of online degrees in a given region and field only affect outcomes of that specific field of study. This assumption is violated if degrees across fields are substitutes. Moreover, the linear structure may not capture out-of-sample counterfactuals effectively, where competition and supply-side responses, such as price changes and entry or exit decisions, can lead to significant non-linear effects. To address these limitations, we develop a supply and demand model for college education that incorporates rich substitution patterns and accounts for equilibrium responses, allowing us to assess the impact of online education expansion under different counterfactual scenarios.

Our equilibrium model consists of students and educational institutions. On the demand side, students decide whether to attend college and in which degree to enroll. On the supply side, institutions decide whether to enter a particular market, which degrees to offer, and what prices to charge. To operate in a market, institutions must establish either a campus for in-person degrees or a hub for online degrees. Their decision-making occurs in two stages. In the first stage, institutions simultaneously choose in which regions to operate and which degrees to offer, after observing their fixed entry costs and taking into account pre-existing offerings (Seim, 2006; Atal et al., 2025). They form expectations about their competitors' entry decisions and adjust their strategies accordingly, potentially opting out of opening an in-person campus to avoid fixed costs if they anticipate competitors will expand their online offerings. In the second stage, institutions compete on prices.

We estimate demand by leveraging different variations in the data. To estimate substitution between degrees in different fields of study and delivery modes, we use market-level differences in degree availability induced by the shift-share instrument. Our estimates suggest that in-person and online degrees are close substitutes: when an in-person degree closes, 67% of students switch to another in-person option, 18% move online, and 14% opt out of college. To estimate price elasticities, we use contemporaneous prices of the same degree in other regions as proxies for cost-shifters. These leave-one-out mean prices serve as an instrument for prices (Hausman et al., 1994). The mechanics behind this instrument are that firm-level cost variations influence prices in all markets where a degree is offered, and the key identification assumption is that, although costs for degrees from the same institution are correlated across markets, demand shocks are not. We estimate median own-price elasticities of approximately -2.9 for in-person degrees and -1.1 for online degrees, which aligns with findings from the literature (Dobbin et al., 2021; Armona and Cao, 2024).

For the supply model, we recover entry elasticities with respect to profits using two instruments that affect profits but do not influence institutions' fixed costs. First, we exploit regional variation over time in internet penetration as a demand shifter for online education. Second, we use differences in competitors' distance to various regions, which creates varying levels of competition that impact profits without affecting institutions' fixed costs. We estimate a median entry elasticity with respect to own profits of 1.2 and find that in-person campuses are 2.3 times more expensive to open than online hubs.

We use our estimated model to quantify the impact of online education expansion on student enrollment, market structure, tuition fees, value added, and consumer's expenditure and surplus. To examine how supply-side equilibrium effects shape these outcomes, we simulate three progressively more flexible counterfactuals where we remove the supply of online education. We benchmark each of these scenarios against a baseline counterfactual that reflects the status quo, where online education exists.

In the first counterfactual scenario, we examine the effects of removing online education without accounting for supply-side responses (i.e., fixing degree offerings and prices). Under this scenario, two-thirds of students enrolled in online programs would switch to in-person programs, while the remaining third would exit the market. This shift results in a 14% decline in total enrollment. Because online degrees provide greater value-added than the outside option, the loss of students leaving the market outweighs the gains from those switching to in-person programs, leading to a net decline in total value-added. However, given the significant tuition gap between online and in-person programs, the increase in in-person enrollment raises total expenditures on higher education.

Under the second counterfactual, we allow institutions to respond by adjusting the prices of their degrees. As a result, tuition fees for in-person degrees increase by 13.6%, causing a further reduction in enrollment and total value added, and increased expenditure.

Finally, under the third counterfactual, we allow institutions to adjust both tuition fees and degree offerings. Relative to the previous scenario, the supply of in-person degrees under this scenario expands by 17.3%, attracting new students and increasing total enrollment by 4.3%.

The total value added under this counterfactual is 3% higher than under the first counterfactual, underscoring the importance of accounting for equilibrium responses. Compared to the status quo—where online education is available—total value added in the absence of online education is 3.4% higher. However, a revealed-preference analysis assuming full rationality shows that consumer surplus declines by 11.9%. This suggests that students may strongly prefer more affordable and flexible options, even if they offer lower value added.

We then examine the distributional consequences of online education expansion by evaluating its effect on value added across different student age cohorts. Under the first counterfactual, which removes online education while ignoring supply-side responses, no cohort experiences significant value-added gains, and older students (ages 35-45) see a decline in value added. When supply-side responses are fully accounted for under the third counterfactual, younger students benefit from increased access to in-person programs that were previously limited due to the presence of online options. However, older students, who have strong preferences for online degrees, are worse off and tend to exit the market.

Our findings highlight the uneven effects of online education. While its expansion broadens access for older students, increasing their opportunities for higher education, it reduces in-person options, pushing younger students into lower-quality alternatives. Using this insight, we explore potential government policies aimed at aligning online education with those who benefit most. Specifically, we examine a policy where online education is available only to students above the age of 25. Our results show that this approach would increase value added for all cohorts compared to the baseline scenario, where online education is available to all age groups.

Our findings shed light on how disruptive technologies like online services can reshape competition and market dynamics. While they expand choice and lower costs, benefits may be unevenly distributed. In markets with imperfect information, some consumers may unknowingly switch from higher-quality options, leaving them worse off—especially as declining demand erodes traditional alternatives. To mitigate these risks, policymakers could ensure access to established options for affected groups while allowing others to adopt new technology, balancing innovation with quality preservation.

This paper adds to the growing literature studying the effects of introducing more accessible, lower-quality options in educational markets. Research on community colleges shows they democratize access to higher education but also divert some students from four-year institutions (Rouse, 1995, 1998; Mountjoy, 2022). Likewise, studies on online education underscore its potential to expand access or draw students away from higher-quality alternatives (Deming et al., 2012; Goodman et al., 2019). Additional research also emphasizes the competitive pressure online degrees exert on tuition prices of traditional in-person programs (Deming et al., 2015). Our paper advances this line of work by developing an equilibrium framework that accounts for market expansion, market diversion, price changes, and endogenous degree offerings.

Our research is also related to the literature that examines the effects of online forms of education on student learning and academic progression (Figlio et al., 2013; Bettinger et al., 2017; Kofoed et al., 2024), and on labor market outcomes (Deming et al., 2016; Hoxby, 2018; Fabregas and Navarro-Sola, 2024). Our findings are consistent with this literature, highlighting the role of online education as a preferable alternative to no education, albeit less favorable than in-person options. Furthermore, our study contributes to the expanding literature on the market effects of online services beyond education, particularly telemedicine. Zeltzer et al. (2023) show that telemedicine can reduce overall healthcare spending without compromising diagnostic accuracy or outcomes. Additionally, evidence suggests that online healthcare services can enhance efficiency by offering faster and shorter consultations and improving the matching of doctors and patients (Dahlstrand et al., 2024; Dahlstrand, 2024).

We also contribute to the broader literature that analyzes education policy using equilibrium models of imperfect competition. These models have been used to study the effects of educational policy on pricing and quality in secondary schools (Neilson et al., 2013; Allende, 2019) and colleges (Dobbin et al., 2021; Barahona et al., 2023; Armona and Cao, 2024), on instructional levels (Bau, 2022), and on institutions' decisions to participate in voucher programs (Sanchez, 2023). More closely related to our work, three papers assess the impact of competition on market structure. All three papers focus on the impact of improved public sector offerings on private institutions' entry and exit decisions. Bodéré (2023) explores the effects of higher-quality public preschools in Pennsylvania, while Dinerstein and Smith (2021) assess the impact of increased public school funding in New York. Similarly, Dinerstein et al. (2023) investigates the expansion of public schools in the Dominican Republic. Our contribution to this literature is to provide evidence of increased competition driven by the private sector itself through the introduction of a new delivery format into the market.

The paper is organized as follows. Section 2 provides institutional background on Brazil's higher education sector, discusses the data, and presents descriptive statistics on the growth of online education. Section 3 presents the results from the linear model used to estimate the causal effects of expanding online education on various outcomes. In Section 4, we introduce and estimate the equilibrium model, and in Section 5, we use it for counterfactual analysis. Finally, Section 6 concludes the paper.

2. Setting and data

In this section, we describe the Brazilian higher education and online education regulatory landscape. We then describe the data sources and provide several descriptive statistics about the expansion of online education.

2.1. *Online higher education landscape in Brazil*

Brazil's higher education system has experienced significant expansion over the past decade, with new undergraduate enrollment growing from approximately 2.2 million students in 2010 to 3.6 million in 2022. A crucial factor underlying this growth has been the rising popularity of online degree programs, predominantly offered by for-profit private institutions.[1] The shift towards online education has been dramatic: in 2010, 17% of new students chose online programs, rising to 44% by 2019, and soaring to 65% after the COVID-19 pandemic.[2]

Online degree programs in Brazil, referred to as "Educação a Distância", offer remote versions of traditional in-person undergraduate diplomas and are required to adhere to the same curriculum and duration standards.[3] Diplomas make no distinction between whether a degree was earned online or in person, theoretically placing both modes on equal footing. However, despite their lower cost, online programs are often perceived as being of inferior quality, a concern that continues to challenge Brazilian policymakers (Bertolin et al., 2023).

Online programs are required to include in-person sessions for essential activities like assessments and laboratory work, which must be conducted either at the institution's main campus or at designated local hubs. These hubs are decentralized centers created to support the face-to-face elements of online education, placing geographic limits on the reach of these programs. Despite the in-person requirements, all instruction remains fully remote. Most institutions provide live, synchronous classes to facilitate real-time interaction between students and instructors. In addition, 78% of institutions offer asynchronous resources—such as pre-recorded lectures, reading materials, and interactive quizzes—giving students more flexibility to engage with course content on their own schedule (ABED, 2018).

The growth of online education has been driven by several key factors. First, there's growing demand stemming from the flexibility these programs offer, enabling more students—particularly an older demographic—to pursue higher education while balancing other responsibilities (El Galad et al., 2024). In 2019, 71% of new online students were over the age of 24, compared to just 32% in traditional in-person programs. Second, the widespread improvement in internet infrastructure across Brazil has significantly facilitated access to online degrees, even in previously underserved regions. In 2010, just 40% of Brazilian households had internet access (IBGE, 2010), but by 2019, this figure surged to 83% (PNAD, 2019). Lastly, government reforms introduced in 2016 have streamlined the accreditation process for new online programs and granted institutions greater autonomy to establish new hubs. This regulatory evolution has made it considerably easier for educational institutions to manage, expand, and diversify their online offerings, contributing to the sector's overall growth.[4]

Some fields of study face restrictions on being offered online. Specifically, Law, Medicine,

[1] Private institutions make up 95% of the total online education market, with 79% of that share held by for-profit institutions. Furthermore, this market is heavily concentrated, with seven institutions dominating 73% of the entire online education market.

[2] Brazil's share of fully remote undergraduate enrollments is significantly high compared to other countries. Although reliable cross-country data on fully online degrees is generally scarce, official data from 2019, prior to the COVID-19 pandemic, indicates fully remote enrollment shares of approximately 13% in Australia, 14% in India, 17% in Mexico, 8% in the United Kingdom, and 15% in the United States.

[3] Students have the option to transfer credits between online and in-person formats within the same institution if they change modalities.

[4] See Resolução CNE/CES N1, de 11 de março de 2016, Decreto 9.057, de 25 de maio de 2017, and Portaria Normativa 11, de 20 de junho de 2017.

and Psychology require special authorization from regulatory bodies such as the National Bar Association and the National Health Council. To date, no online programs in these fields have been approved. In contrast, disciplines like Business and Education have seen significant growth in online education. In 2019, these two fields accounted for 77% of total online enrollment, in contrast to their 31% share in in-person programs.

2.2. *Data*

2.2.1. *Higher education census*: This dataset encompasses several layers of information. First, it includes institution-level details, such as ownership status and the parent firm. Second, it captures program characteristics, including detailed categories of the field of study, the delivery mode (online or in-person), the required number of hours for graduation, and the year the degree was introduced. Additionally, it provides information on the hubs associated with each online program. Third, it contains student-level data for all enrolled students, along with their demographics, enabling us to track each student's educational path.

2.2.2. *Tuition fees*: Universities are not required to report tuition fees to the supervising authority, so we rely on four distinct data sources to gather this information. The first two sources come from Brazil's government fellowship and loan programs, PROUNI and FIES. We utilize administrative records from the National Education Fund (FNDE), which track the payments made by the government to students in these programs, allowing us to estimate the tuition fees at participating institutions. The third source is a nationally representative survey conducted by Hoper, a consultancy specializing in higher education. The fourth source is administrative data from QueroBolsa, Brazil's largest degree search platform. In Appendix B.1, we outline the methodology used to combine these sources into a unified tuition price for each degree program. As a result, we are able to recover year-specific tuition prices for approximately 95.5% of degree program-years, covering 98.5% of total enrollment.

2.2.3. *Test scores*: We have access to detailed data for all students who took ENEM, Brazil's university entrance exam. This standardized test is high stakes, as it determines eligibility for financial aid and is used for admissions by several public universities. The data include scores for each section of the exam, along with responses to a comprehensive socioeconomic background questionnaire. College applicants who take the entrance exam vary significantly in age; for example, in 2010, about 62% had finished high school more than a year before taking the test, and 25% were 25 or older.

2.2.4. *Matched employer-employee records*: Finally, we integrate the previously mentioned data sources with matched employer-employee annual administrative records (RAIS) from the Ministry of Labor. This dataset includes detailed worker and firm-level variables, such as salaries, contracted hours, hiring and firing dates, and occupation of Brazil's formal labor market. We use these data to construct earning profiles for each program and student, covering both online and in-person formats. These profiles allow us to calculate value added measures for each program in the system.

2.2.5. *Additional auxiliary sources*: We use data from the 2010 Brazilian Population Census to estimate the size of various age cohorts within each region, allowing us to define the potential market for higher education students. We also use administrative data from DSCOM ("Dados do Setor de Comunicações") from the Ministry of Science and Technology to calculate rates of internet penetration across different regions over time.

2.3. *Data definitions*

Next, we provide several definitions for the units of analysis and samples that we use throughout the article.

2.3.1. *Regions and markets*: A central aspect of our analysis is defining regions that allow us to segment educational markets. Brazil consists of 5,568 municipalities, which we group into 137 meso regions—an administrative division from the Brazilian National Bureau of Statistics (IBGE) that clusters municipalities based on proximity and common features. These meso regions serve as our definition of local markets, which we refer to as *"regions"* throughout the paper. We define a *"market"* as the intersection of a region and a year. The market size is determined by the number of 18- to 45-year-old residents without a college degree living in the region for that year.

2.3.2. *Firms*: We define a firm as a company that may own multiple universities, and we use the terms "firms" and "institutions" interchangeably. In our data, we observe two types of firms: those that are expanding their services into other regions and others that operate in the same locations every year. We refer to the former as *expanding* institutions and the latter as *local* institutions. Most local institutions operate in a single location with a few of them operating in two to five different locations.

2.3.3. *Degrees*: Throughout this draft, we use two different definitions of "degrees." The first is *degree programs*, which are distinct undergraduate programs sharing the same administrative code. These programs are offered by the same university, have identical curricula, and are delivered in the same format—either in person or online. For in-person degree programs, the only variation may be the schedule, such as morning, evening, day, or full-time. For online degree programs, those sharing the same administrative code may be associated with different physical hubs, allowing them to be available in multiple regions. The dataset includes 35,527 unique degree programs.

To reduce the data's dimensionality, we introduce a second definition of degrees by aggregating similar degree programs within the same institution. We define a *degree* as the combination of all degree programs offered by the same institution, within the same field of study, and delivered in the same format.[5] For example, in our dataset, a degree might include all degree programs in the field of "Business" (such as Administration, Accounting, Marketing, and Economics), offered by Anhanguera Educacional, a for-profit educational company in Brazil, and taught in person. A degree can be offered in several regions and years.

2.3.4. *Sample*: Our analysis centers on the private sector, which accounts for nearly all online programs and roughly 82% of total enrollment, including both online and in-person students. This sector is predominantly non-selective and is often perceived as lower in quality compared to the public sector, which is higher quality and highly selective (Barahona et al., 2023). Consequently, there is limited substitution between the two sectors. Public institutions, which rely entirely on government funding—whether federal or state—are also less influenced by market forces.

We also limit our analysis to the years 2010, when microdata first became available, and 2019, the final year before the onset of the COVID-19 pandemic. To avoid including very small markets in our analysis, we apply the following restrictions: we exclude 27 regions with fewer than 50,000 residents aged 18-45, institutions with fewer than 500 students nationwide in any given year, and degree-region pairs with fewer than five students in any year. After these exclusions, our final sample consists of 110 regions over a 10-year period, covering 474 institutions, 4,119 unique degrees, 15,206 degree-region pairs, and 92,625 degree-region-years. 340 institutions are local and 134 are expanding, out of which 93 expanded online.

2.4. *Value added*

An important component of our analysis is computing a quality measure for *degree programs*. We approximate program quality using value added measures derived from data on student enrollment, test scores, a comprehensive set of covariates, and administrative wage records. To calculate value added, we track all students taking the university entrance exam (ENEM), assigning them to specific degree programs based on their initial college enrollment or to an outside option if they do not enroll. We then follow these individuals in the labor market, recording their highest salary reported in the administrative wage records (RAIS) several years post-exam.

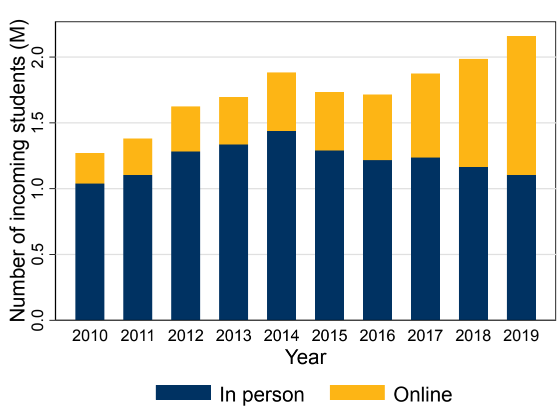
Our approach employs a standard selection-on-observables model (Rothstein, 2010; Angrist et al., 2017; Ainsworth et al., 2023), which compares labor market earnings across degree programs while controlling for test scores and a wide range of student characteristics. Our methodology for calculating value added, outlined in Appendix B.2, incorporates three key features.

[5] In total, we have 11 fields of study based on the International Standard Classification of Education (ISCED) codes. Details on the categorization are provided in Online Appendix A, Table A.1.

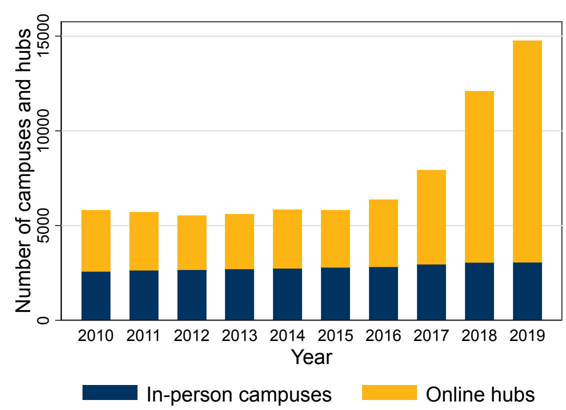
First, we normalize all measures relative to the outside option of not attending college. Because we care about the outside options that students face in their specific local markets, we calculate value added separately by region. As a result, degree programs' value added can differ across regions if local market conditions favor some programs over others or if the outside option of not attending college is different. Second, we analyze value added outcomes using 2010 test takers and follow them seven years post-application (using ENEM 2010 to RAIS 2017 data), providing a stable, time-invariant measure. Finally, following Angrist et al. (2023), we use empirical Bayes methods to improve the precision of estimates for smaller programs.

2.5. *Descriptives*

2.5.1. *Trends in the Brazilian private-sector higher education sector*: We present the trends of the overall expansion of online education from 2010 to 2019 in Figure 1. Panel 1(a) shows the enrollment trends of incoming students for both online and in-person education. Between 2010 and 2014, college enrollment steadily increased, with the share of online education remaining relatively stable, representing approximately 23% of total enrollment by 2014. However, following the policy reforms around 2016 that liberalized access, described in Section 2.1, a notable shift occurred: online education saw significant growth while in-person enrollment began to decline. By 2019, over half of incoming students at private universities were enrolled in online education. Panel 1(b) illustrates similar trends in the number of in-person campuses and online hubs in Brazil. Between 2010 and 2015, the number of campuses and hubs remained comparable, with a marked rise in the number of online hubs following the policy reforms.



(a) Student enrollment



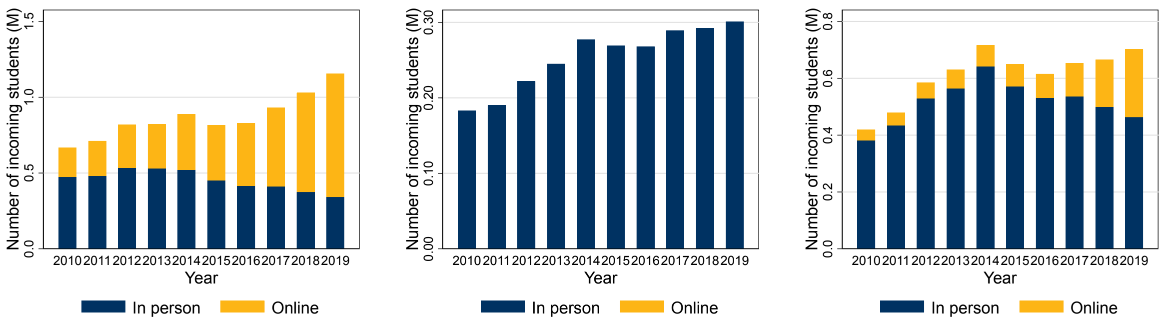
(b) Campuses and hubs

Figure 1: Expansion of online education

Notes: This figure presents the trends of the Brazilian private-sector higher-education sector. Panel (a) shows the number of incoming students in private institution for in-person (blue) and online (yellow) degrees across years. Panel (b) shows the number of open in-person campuses (blue) and online hubs (yellow) across all regions in Brazil.

The expansion of the online education sector varied across fields of study. As shown in

Figure 2, the most substantial growth occurred in programs related to Business and Education, partly due to their suitability for online delivery. In contrast, fields such as Law, Medicine, and Psychology saw no comparable expansion, as legal restrictions prevent institutions from offering these programs online. Moreover, in-person enrollment declined in fields that experienced substantial online growth but continued to rise in those where online education is restricted.



(a) Education & Business (b) Law, Medicine, & Psychology (c) Other areas

Figure 2: Expansion of online education enrollment by categories of field of study

Notes: This figure presents enrollment trends of the Brazilian private-sector higher-education sector across fields of study. Panel (a) shows the number of incoming students in private institution for in-person (blue) and online (yellow) degrees in Education and Business, Panel (b) for degrees in Law, Medicine, and Psychology, and Panel (c) for degrees in other fields of study.

Finally, we highlight that online degree programs are especially popular among older students. There is a noticeable difference in the age distribution of students enrolled in online versus in-person programs. In Figure 3, we show online and in-person enrollment segmented by age group for 2010 and 2019.[6] While in-person programs are dominated by younger students, online programs attract a majority of students aged 26 and older. By 2019, around 64% of online students in these fields were aged 26 or older, compared to just 28% in in-person programs.

2.5.2. *Comparison between online and in-person programs*: To examine the practical differences between online and in-person programs, we compare equivalent *degree programs* offered by the same university, differing only in their mode of delivery. We estimate the following regression:

$$Y_{jrt} = \beta \cdot o_j + \delta_{m(j)t} + \delta_r + \varepsilon_j, \tag{1}$$

where Y_{jrt} is an outcome in degree program j in region r in year t, $o_j \in \{0, 1\}$ indexes whether the program is online, $\delta_{m(j)t}$ is a university-major and year-specific fixed effect (e.g., Economics at the Universidade Norte do Paraná in 2011), and δ_r is a region fixed effect. We examine four primary outcomes: (1) the log of the number of hours required to complete the program, (2) the

[6]Throughout our analysis, we split students into four age groups: 18-20, 21-25, 26-35, and 36-45. In 2010, each of the first three groups represent 30% of total college enrollment, and the last group represents the remaining 10%. In terms of the total population, the first group represents 12%, the second 24%, the third 38% and the fourth 24%.

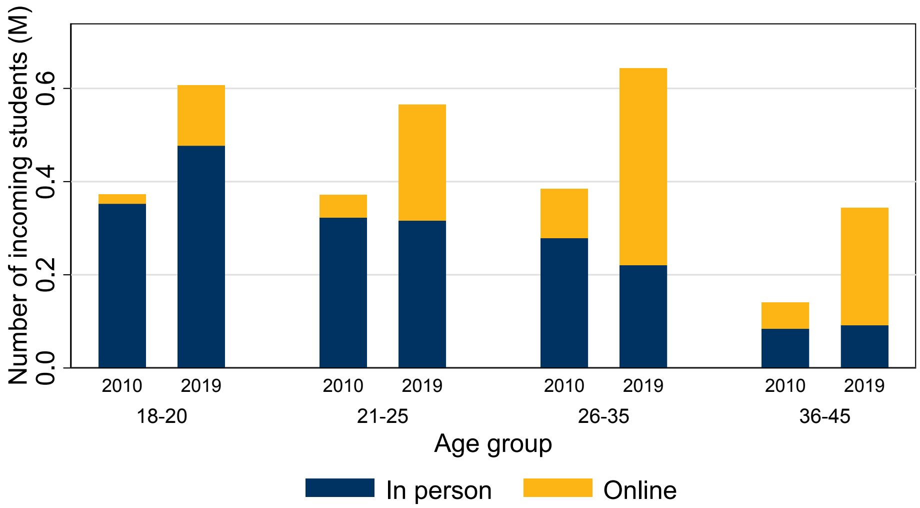


Figure 3: Expansion of online education enrollment by age group

Notes: This figure presents enrollment levels in the Brazilian private-sector higher-education sector across age groups in 2010 and 2019. For each age group, the left and right bars show enrollment in private institutions for in-person (blue) and online (yellow) degrees in 2010 and 2019, respectively.

log of the tuition price, (3) first year dropout rates, and (4) value added as defined in Section 2.4. Since our value-added measure does not vary over time, we estimate the corresponding regression using a cross-section of the data for $t = 2011$ (corresponding to 2010 ENEM takers). All regressions are weighted by the number of students enrolled in program j.

The results are presented in Table 1. Column (1) reports the findings for the log of the required number of hours to complete the degree, showing that both online and in-person programs require a similar number of hours, which is expected as it is mandated by law. In column (2), we observe that online programs are 0.6 log-points less expensive than their in-person counterparts. Column (3) shows that online programs have dropout rates that are 0.021 percentage points lower (over a base of 0.20) than their in-person counterparts. Lastly, column (4) reveals that the value added of online programs is, on average, 0.057 log-points lower than that of in-person programs (over a base of 0.14).

These findings shed light on the potential trade-offs between in-person and online education. Online education, for example, can be offered at a significantly lower cost, making tuition more affordable. Its flexibility and reduced fees may contribute to slightly higher first-year persistence rates. However, it generally provides a lower value added compared to in-person programs. Nonetheless, both online and in-person degrees, on average, deliver higher value added than the alternative of not attending college (see Online Appendix A, Figure A.1).

3. The Effects of Online Entry on Market Outcomes

In this section, we estimate the effects of introducing an additional online degree on various market outcomes using a linear model. To achieve this, we compare changes in outcomes between

Table 1: Comparison between online and in-person programs

	log total hours (1)	log prices (2)	dropout rate (3)	value added (4)
Online	-0.003	-0.573	-0.021	-0.057
	(0.005)	(0.030)	(0.011)	(0.007)
Obs.	544,920	549,704	549,704	17,540
Mean dep var. (levels)	3220	4430	0.20	0.14
Region FE	Yes	Yes	Yes	Yes
University-major-year FE	Yes	Yes	Yes	Yes

Notes: This table reports the coefficient of a linear regression of each outcome against an indicator variable for whether the degree is online or not, reported in Equation (1). All regressions are weighted by the degrees' number of students and control for region fixed effects and university × program name × year fixed effects. Column (1) uses the log-number of hours to complete the program as an outcome, Column (2) uses the log of the tuition price, Column (3) shows first-year dropout rates, and Column (4) uses value added. To calculate value added, we run a regression of log income on a rich set of student characteristics and degree fixed effects. We use the estimate of the degree fixed effect as the measure of value added (see Section 2.4). Columns (1), (2), and (3) use all years. Column (4) uses only 2011, which is the year for which we estimate value added.

2010 and 2019 across regions and fields of study with varying levels of exposure to the growth of online degrees. Throughout this section, we use the *"degree"* definition as outlined in Section 2.3.

Specifically, we estimate the following structural equation:

$$\Delta y_{ra} \quad = \quad \phi \Delta N^o_{ra} + \varepsilon_{ra}, \tag{2}$$

where ΔN^o_{ra} denotes the change in the number of online degrees offered in region r and field of study a between 2010 and 2019, and Δy_{ra} represents the change during the same period in the following outcomes of interest: (i) the number of online students relative to market size, (ii) the number of in-person students relative to market size, (iii) the total number of in-person degrees, (iv) and the average price of in-person degrees. The error term ε_{ra} captures unobserved shocks at the region-field level that influence the trend of y_{ra}. The coefficient ϕ is the key parameter of interest.[7]

Estimating a linear model presents both advantages and limitations. On the one hand, it offers simplicity and transparency. On the other hand, it relies on strong assumptions. In particular, it imposes a no-interference assumption implying that changes in the number of online degrees offered in region r and field of study a only impact the outcomes in that particular region and field of study, without influencing outcomes in other regions or fields—a condition known as the Stable Unit Treatment Value Assumption (SUTVA). This assumption may be violated if

[7] Equation (2) can be derived by taking differences between 2019 and 2010 of the following structural equation: $y_{rat} = \phi N^o_{rat} + \delta_{ra} + \delta_t + \varepsilon_{rat}$, where y_{rat} and N^o_{rat} are the outcome of interest and the number of online degrees in region r, field of study a, and year t, and δ_{ra}, δ_t, and ε_{rat} are region-field of study fixed effects, year fixed effects, and region-field-year specific shocks.

degrees across different fields are close substitutes. In Section 4, we extend our model to account for potential market-level interactions.

We propose two alternative estimation strategies for the linear model, each based on different assumptions that identify the parameter of interest. First, we outline the assumptions required for a causal interpretation of ϕ when estimating Equation (2) using OLS. Next, we introduce a shift-share instrumental variable approach to address potential threats to the identification in the OLS regression. Finally, we present and compare the results from both estimation strategies.

3.1. *Ordinary Least-Squares Regression*

The parameters estimated through OLS can be interpreted as causal under the conditional independence assumption $\mathbb{E}[\varepsilon_{ra}|\Delta N^o_{ra}] = 0$. Since Equation (2) is written in differences, this assumption is implied by a parallel trends assumption, which states that the trajectory of outcomes for regions and fields of study with lower online degree growth must represent the outcomes in higher-growth regions and fields had they experienced lower online growth (Callaway et al., 2024). Unfortunately, standard tests for this assumption—based on evaluations of pre-trends—are not feasible, as online education was already widespread and growing at the start of our sample. To address this concern, we implement a shift-share instrumental variable approach.

3.2. *Shift-share instrumental variables*

We tackle the issue of endogeneity in institutions' entry decisions by exploiting quasi-random variation in regions' exposure to online entry. Following Goldsmith-Pinkham et al. (2020), we implement a shift-share instrumental variables (SSIV) design—or "Bartik instruments"—with exogenous shares.[8] Our *shares* variable is constructed from a combination of three factors: (i) differences in regions' exposure to potential entrants based on the distance between the regions and institutions' headquarters (ii) institutions' specialization across different fields of study based on their 2010 in-person offerings, and (iii) an indicator variable that captures whether online education is allowed in a given field of study.[9] For the *shift*, we use the rapid expansion of online degree programs, driven by a combination of factors, including growing demand, advancements in online technology, and the policy changes mentioned in Section 2.1.

Building on this, we define the following shift-share instrument:

$$z_{ra} = \sum_f \underbrace{\Delta N^o_f}_{\text{Shift}} \cdot \underbrace{(z_{fr} z_{fa} z_a)}_{\text{Share}}, \tag{3}$$

where ΔN^o_f represents the *shift*, capturing the total number of online degrees introduced by

[8] Goldsmith-Pinkham et al. (2020) implement SSIV under the assumption that the shares are exogenous. In contrast, Borusyak et al. (2021) and Borusyak and Hull (2023) provide a framework for SSIV where identification is achieved through an exogenous shift.

[9] As discussed in Section 2.1, regulation prohibits the fields of Law, Medicine, and Psychology from offering remote education.

institution f between 2010 and 2019. We allow this shift to be correlated with the distribution of shocks ε_{ra}. The variable $z_{fr}z_{fa}z_a$ corresponds to the exogenous *shares* that predicts the region r and field of study a where institution f is likely to expand. As described above, these shares are derived from three sources: (i) exposure to potential entrants, z_{fr}, (ii) the institutions' propensity to expand in different fields of study, z_{fa}, and (iii) regulatory constraints within fields of study, z_a. We begin by detailing the *shift*, followed by an explanation of each component of the *share*.

A visualization of each component is provided in Figure 4. In each panel, we order institutions by the location of their headquarters from northwest to southeast, based on the IBGE encoding system. We explain each panel in detail as we introduce the different components of Equation (3).

3.2.1. *Institutions' expansion of online degrees (ΔN_f^o)*: The first component of the instrument in Equation (3) is the shift term. We estimate ΔN_f^o by calculating the total number of online degrees that each institution opened in any region or field of study between 2010 and 2019:

$$\Delta N_f^o \quad = \quad \sum_r \sum_a \left(N^o_{fra,2019} - N^o_{fra,2010} \right), \tag{4}$$

where $N^o_{fra,t}$ takes the value of 1 if institution f offers a degree in field of study a in region r in year t.[10] In total, we have 474 institutions, out of which 93 expanded their online offerings between 2010 and 2019.

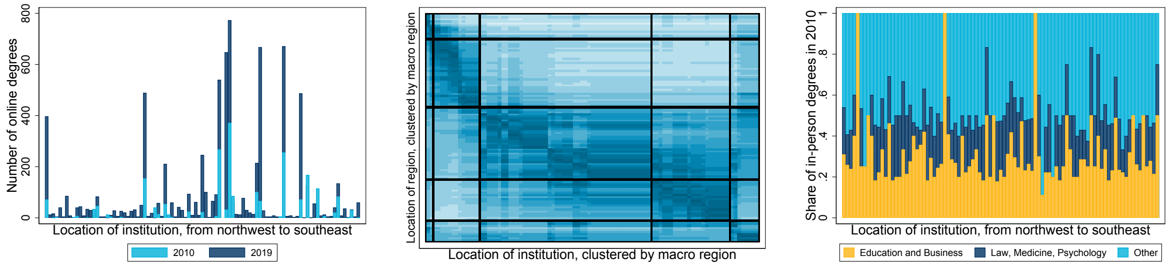
Figure 4(a) illustrates the variation in overall online degree expansion across institutions. On the x-axis, each column represents one of the 93 institutions that expanded online between 2010 and 2019, organized by their headquarters' location region. There is significant heterogeneity across institutions. Our instrument predicts more intense online expansion in regions located near institutions that expanded more aggressively.

3.2.2. *Exposure to potential entrants (z_{fr})*: We leverage the differential exposure of regions to institutions' online expansion, driven by their distance from each institution's headquarters. Our empirical analysis shows that institutions are more likely to establish online hubs—and consequently offer degrees—in regions closer to their headquarters, likely due to the lower costs of launching and maintaining nearby operations.[11] We document this pattern through the following regression analysis, using data from 2010:

$$\text{Entered}_{fr} = g(d_{fr})'\gamma + \delta_f + \delta_r + \eta_{fr}, \tag{5}$$

[10] Our results are robust to using a leave-one-out estimate of ΔN_f^o. Because our identification strategy relies on having exogenous shares, a leave-one-out estimator is not necessary in this setting.

[11] These lower costs might be attributed to factors such as easier management and coordination enabled by proximity, reduced setup and operational expenses, and a deeper understanding of local needs and challenges near the headquarters.



(a) Online degrees by firm (N_f^o) (b) Prob. of online regional entry by firm (z_{fr}) (c) Share of degrees across field of study by firm (z_{fa})

Figure 4: Expansion of online education enrollment by categories of field of study

Notes: This figures illustrates the main patterns in the data driving the different components of the shift-share instrument. Panel (a) shows the total number of online degrees offered by each of the 93 institutions that expanded their online presence between 2010 and 2019 in the years 2010 and 2019 (N_f^o). Institutions are arranged by their headquarters' location region from northwest to southeast based on the IBGE encoding system. ΔN_f^o corresponds to the difference between the 2010 and 2019 bars. Panel (b) is a heat map representing the exposure of regions to potential entrants (z_{rf}). Each row in the vertical axis is one of the 110 regions in our data where institutions can enter and are arranged from northwest to southeast, following Panel (a). Each row in the horizontal axis is one of the 93 institutions that expanded their online presence between 2010 and 2019 and are arranged from northwest to southeast based on their headquarters' location region. The black lines separate Brazil in five macro regions, such that all cells in the diagonal blocks represent region-institution pairs that are located in the same macro region. Each region-institution pair is shaded in blue, with lighter shades indicating lower exposure, or a lower likelihood that institution f will enter region r, and darker shades indicating higher exposure. Panel (c) shows the share of in-person degrees offered by each institution in 2010 (z_{fa}). Each bar represents one of the 93 institutions that expanded their online presence between 2010 and 2019 and are arranged from northwest to southeast based on their headquarters' location region.

where Entered_{fr} equals 1 if institution f had entered region r by 2010 and 0 otherwise; d_{fr} represents the distance between the headquarters of institution f and region r.[12] We define the vector $g(d_{fr}) = [\log(1 + d_{fr}), H_{fr}]'$, where $H_{fr} \in \{0, 1\}$ indicates whether the headquarters of institution f are located in region r (i.e., $d_{fr} = 0$), to account for cases where the distance is zero. Our results show a negative and significant relationship between distance and the probability of entry. Specifically, a 10% increase in the distance between a region and an institution's headquarters reduces the probability of entry by 4%. More detailed results can be found in Online Appendix A, Table A.2.

We use the estimates from Equation (5) to determine the likelihood that institution f will open an online hub in region r, based solely only the distance between the institution and the region. This likelihood is calculated by predicting entry using distance while excluding fixed effects, thereby removing potentially endogenous variation from unobserved regional characteristics. We then normalize the likelihood so that it sums to one across all regions.[13] Our measure

[12] Our results are robust to estimating exposure using a different year.

[13] Equation (6) can be interpreted as the probability that institution f opens an online hub in region r, assuming it opens exactly one hub, with the probability depending only on the distance between the institution's headquarters and the region.

of exposure is defined as:

$$z_{fr} \quad = \quad \frac{g(d_{fr})'\hat{\gamma}}{\sum_r g(d_{fr})'\hat{\gamma}}. \tag{6}$$

Figure 4(b) illustrates the exposure instrument, z_{fr}. The y-axis represents each of the 110 regions in our sample, organized by the five macro-regions defined by the IBGE, with black lines separating each macro-region. Within each macro-region, regions are arranged from northwest to southeast, based on the IBGE encoding system. On the x-axis, each column corresponds to one of the 93 institutions that expanded their online presence between 2010 and 2019, ordered by the location of their headquarters, using the same strategy as for the regions and Figure 4(a). Each region-institution pair is shaded in blue, with lighter shades indicating lower exposure, or a lower likelihood that institution f will enter region r, and darker shades indicating higher exposure. We observe that exposure tends to be higher for region-institution pairs that are geographically close and within the same macro-region.

3.2.3. *Institutions' propensity to expand in different fields of study (z_{fa})*: Section 3.2.2 focuses on estimating the likelihood that each institution enters a given region. However, different institutions might have different likelihoods of expanding into different fields of study. To address this, we use institutions' 2010 offerings to predict their likely fields of expansion, assuming that institutions initially specializing in certain areas may find it easier to expand within those fields. Since many institutions offered few or no online degrees in 2010, we use the intensity of their in-person instruction to predict where they are most likely to expand online. We estimate the likelihood of expansion in each field based on the share of in-person degrees offered in that field in 2010. Specifically, we calculate:

$$z_{fa} \quad = \quad \frac{\sum_r N^{\iota}_{fra,2010}}{\sum_a \sum_r N^{\iota}_{fra,2010}}, \tag{7}$$

where $N^{\iota}_{fra,2010}$ takes the value of 1 if institution f offers an in-person degree from field of study a in region r in 2010.

Figure 4(c) illustrates the resulting shares by institution, aggregated into broader groups of study areas. On the x-axis, each column represents one of the 93 institutions that expanded their online presence between 2010 and 2019, ordered by the location of their headquarters, as in Figures 4(a) and 4(b). We observe that some institutions concentrated more on fields like Education and Business, while others prioritized areas such as Engineering or Math. Our instrument predicts more intense online expansion in field of study a in regions with nearby institutions that were already specializing in that area of study in 2010.

3.2.4. *Areas of study regulatory constraints (z_a)*: Throughout our sample, regulations prohibit online education in Law, Medicine, and Psychology. We use this constraint to build an indicator variable, z_a, which equals 1 for areas of study allowed to expand online and 0 for those that are not. We incorporate z_a into Equation (3), ensuring that our instrument predicts zero online growth in areas of study such as Law, Medicine, and Psychology.

3.3. *Identification*

As is standard in linear models using instrumental variables, two assumptions must hold for the correct interpretation of causal effects. First, the instrument needs to be relevant, meaning it must have predictive power over the endogenous variable. We test this by conducting a first-stage regression between the SSIV, z_{ra}, and the endogenous variable, ΔN^o_{ra}—the change in the number of online degrees offered in region r and field a—yielding an F-statistic of 201. The coefficient from this regression is presented in Table 2, Panel B, Column (1). Second, the exclusion restriction must hold. Specifically, the exposure shares, z_{fra}, must be uncorrelated with the structural error term, ε_{ra}. Formally, this requires that $\mathbb{E}[z_{fr}z_{fa}z_a\varepsilon_{ra}] = 0$ for all expanding firms (i.e., $\forall f$ where $\Delta N^o_f \neq 0$).

To build intuition around the identification assumption, it helps to define $\bar{\epsilon}_{fr} = \mathbb{E}[z_{fa}z_a\varepsilon_{ra}|fr]$, which represents the propensity of institution f to expand in region r due to anticipated region-field-specific shocks, ε_{ra}. For example, if institution f specializes in business degrees in 2010, and region r is expected to see an increase in demand for online business degrees, institution f might be more likely to offer online degrees in that region, making $\bar{\epsilon}_{fr}$ larger. Using the law of iterated expectations, the identification assumption can be reformulated as $\mathbb{E}[z_{fr}z_{fa}z_a\varepsilon_{ra}] = \mathbb{E}[z_{fr}\bar{\epsilon}_{fr}] = 0$ for all expanding firms. This condition holds if the distance between regions and a given institution is uncorrelated with region-specific unobserved demand shocks, ε_{ra}, related to the field of study in which the institution specializes in in-person education.

3.4. *Results*

Table 2 presents the results for both identification strategies. Panel A displays the OLS results, while Panel B shows the SSIV results. We find that the OLS and SSIV strategies yield qualitatively similar results, though the IV estimates are larger in magnitude. We interpret this as evidence that endogeneity concerns about online entry are small, as institutions might have difficulty anticipating future demand shocks when deciding to expand their online portfolio. This suggests that most online expansion decisions are driven by other factors, such as costs—which are captured by our shift-share instrument—and further influenced by the 2016 policy reforms that facilitated rapid expansion to close locations that might have lower expansion costs. Given the similarity of the results, we focus our discussion on the SSIV, though the conclusions are the qualitatively the same for OLS.

3.4.1. *Average effects*: We begin by examining the impact of introducing an additional online degree in a specific field and region on both online and in-person enrollment, with results presented in Columns (2) and (3), respectively. These outcomes are reported relative to market size, defined as the number of individuals aged 18 to 45 without a college degree in that region. Column (2) indicates that each additional online degree introduced between 2010 and 2019 in a given field increased online enrollment by 0.391 students per 1,000 individuals in the market. This is equivalent to a 14% increase in total enrollment when the number of online degrees in that field is raised by 50%. In contrast, Column (3) shows that this increase in online degrees led to a reduction in in-person enrollment in that field of study by 0.2 students per 1,000 individuals, which corresponds to a 7% decline in total enrollment under the same 50% increase in online degree availability.

These results reveal two opposing forces: online degrees expand the market by attracting new students to college while simultaneously diverting students from in-person programs. Our findings show that for each additional online student, 52% are new to higher education, while 48% would have otherwise enrolled in an in-person degree.[14] These forces create an ambiguous effect on total value added. Market expansion increases value added, as newly enrolled students enter programs with positive value compared to the outside option of no college, which we normalize to have zero value added. However, market diversion reduces value added by shifting students from higher value-added in-person programs to lower value-added online alternatives. Based on the average value added differences between online and in-person degrees, a back-of-the-envelope calculation suggests that increasing the number of online degrees by 50% raises value added by 7.6% due to the increase in online enrollment, while the corresponding decline in in-person enrollment reduces value added by 7.7%. These effects largely cancel out, resulting in a net reduction in total value added of 0.04%.[15]

We next analyze the consequences of the online expansion on the availability and pricing of in-person degrees. In Column (4), we show that the number of in-person degrees decreased in regions and fields with larger online growth. Specifically, for each additional online degree, there are 0.184 fewer in-person degrees relative to a 2010 baseline of 10.98. These effects stem from both an increase in degree exit and entry deterrence, which exacerbates the diversion from in-person degrees to online alternatives. As a result, even students who prefer to enroll in a specific in-person program may be forced to switch to an online alternative when their preferred in-person option exits the market. Finally, in Column (5), we show that the average price of in-person degrees dropped by 1.4% for each additional online degree introduced in that region and field of study. This supports the idea that online degrees intensify local competition, driving down prices and deterring new in-person program entry.

[14] Note that the linear model does not allow for cross-field effects, thereby excluding the possibility of expansion coming from other fields of study.

[15] We calculate changes in total value added using the formula: $\Delta \text{VA} = \frac{\Delta s^o \cdot \text{VA}^o + \Delta s^\iota \cdot \text{VA}^\iota}{s^o \cdot \text{VA}^o + s^\iota \cdot \text{VA}^\iota}$. In this expression, Δs^o and Δs^ι represent the coefficients from Columns (2) and (3) of Table 2, Panel B. The terms s^o and s^ι denote the average number of online and in-person students attending college in 2010 for every 1,000 individuals aged 18 to 45 living in that region, as reported in Columns (2) and (3) of Table 2, Panel C. Additionally, VA^o and VA^ι represent the average value added of online and in-person degrees in 2010, given by 0.078 and 0.164, respectively.

Table 2: Effects of introducing an additional online degree

	Δ in online degrees (1)	Δ in online students (2)	Δ in in-person students (3)	Δ in in-person degrees (4)	Δ in log-price of in-person degrees (5)
Panel A: OLS regression					
Δ in online degrees		0.344 (0.033)	-0.190 (0.026)	-0.148 (0.031)	-0.009 (0.002)
Panel B: IV regression					
shift-share instrument	1.723 (0.122)				
Δ in online degrees		0.391 (0.035)	-0.187 (0.033)	-0.184 (0.037)	-0.014 (0.003)
Panel C: Average value of the dependant variable in levels in 2010 and 2019					
2010	2.82	0.72	3.10	10.98	5.63
2019	9.48	2.79	2.94	12.19	6.21
Obs.	1,210	1,210	1,210	1,210	957

Notes: This table shows the results from estimating the linear model from Equation (2), $\Delta y_{ra} = \phi \Delta N^o_{ra} + \varepsilon_{ra}$, for different outcomes, y_{ra}, using two different identification strategies. Panel A shows the results estimating the model via OLS regressions. Panel B estimates the model via 2SLS using the shift-share design described in Section 3.2. Panel C shows the average value of the outcome, y_{ra}, for years 2010 and 2019. Column (1) shows the 2SLS first-stage regression of the total number of online degrees against the shift-share instrument. Columns (2)-(5) present the regression coefficients for the number of online students relative to market size, the number of in-person students relative to market size, the total number of in-person degrees, and the average price of in-person degrees. Columns (1)-(4) are estimated using all region-field of study pairs, while column (5) uses the 954 region-field of study pairs for which there is at least one in-person degree in 2010 and 2019.

3.4.2. *Effects by age groups*: The expansion of online education may affect age groups differently. Evidence shows that online programs tend to attract older individuals, who often face greater challenges and costs in attending in-person classes (Goodman et al., 2019; Aucejo et al., 2024). In our context, by 2010, college enrollees aged 18-20 were 12 times more likely to pursue an in-person degree over an online degree than students aged 36-45. Differences in preferences across age cohorts can lead to heterogeneous patterns of market expansion and diversion that we analyze next.

In Table 3, we estimate enrollment changes across different age cohorts. We find high rates of market diversion from in-person to online degrees among younger cohorts. For students aged 18-21, each additional online enrollment leads to 0.67 students leaving the in-person sector (ratio of the coefficients in Column 5 to Column 1, Panel B), indicating that online education acts as a substitute of in-person alternatives. For students aged 21-25, 0.52 students leave in-person programs for every new online student (ratio of the coefficients in Column 6 to Column 2, Panel B). For older cohorts (ages 26-35 and 36-45), this substitution drops to 0.43 and 0.26 students, respectively (calculated similarly using Panel B coefficients). A back-of-the-envelope calculation suggests that increasing the number of online degrees by 50% decreases value added by 1.6% among students aged 18-20 but increases value added by 8.9% among students aged 36-45.

Table 3: Effects of introducing an additional online degree on enrollment by age groups

	Δ in online students				Δ in in-person students			
	18-20	21-25	26-35	36-45	18-20	21-25	26-35	36-45
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: OLS regressions by cohort								
Δ in online	0.333	0.414	0.370	0.263	-0.332	-0.235	-0.172	-0.081
degrees	(0.044)	(0.043)	(0.036)	(0.024)	(0.058)	(0.037)	(0.030)	(0.012)
Panel B: IV regressions by cohort								
Δ in online	0.393	0.472	0.419	0.295	-0.262	-0.247	-0.181	-0.076
degrees	(0.054)	(0.045)	(0.038)	(0.025)	(0.079)	(0.044)	(0.034)	(0.016)
Panel C: Average value of the dependant variable in levels in 2010 and 2019								
2010	0.67	0.75	0.96	0.81	6.35	2.70	1.34	0.63
2019	3.50	3.08	3.05	2.55	8.20	2.59	1.01	0.55
Obs.	1,210	1,210	1,210	1,210	1,210	1,210	1,210	1,210

Notes: This table shows the results from estimating the linear model from Equation (2), $\Delta y_{ra} = \phi \Delta N^{o}_{ra} + \varepsilon_{ra}$, for the change in the number of in-person and online students from different age groups, using two different identification strategies. Panel A shows the results estimating the model via OLS regressions. Panel B estimates the model via 2SLS using the shift-share design described in Section 3.2. Panel C shows the average value of the outcome, y_{ra}, for years 2010 and 2019. Columns (1)-(4) show the regression coefficients for the the number of online students relative to market size for age cohorts 18-20, 21-25, 26-35, and 36-45. Columns (5)-(8) show the regression coefficients for the number of in-person students relative to market size for the same age cohorts.

Overall, the results from Tables 2 and 3 indicate that online education has the potential to expand access to higher education, especially among older cohorts. However, online education also diverts students from in-person alternatives. Increased competition drives down in-person tuition but reduces the number of in-person degrees. Consequently, the average value added in the economy declines, and students who would prefer in-person learning may find themselves pushed into online alternatives as their preferred programs disappear.

This section's analysis relies on a linear model that assumes no interaction between degrees in different fields within the same region. This limitation is important because it prevents the model from distinguishing between overall market growth driven by new students entering college and shifts caused by students switching from other fields of study. Moreover, the linear approach is unsuitable for out-of-sample counterfactuals, where competition, price changes, and entry/exit decisions can have nonlinear effects.[16] We address these concerns in the next section.

[16] Nonlinearity is crucial when considering entry and exit decisions. While small increases in online competition may have minimal impact, extrapolating this effect to larger scales could misleadingly imply that substantial increases in competition would also have no effect. Therefore, the relationship between the number of online competitors and market structure outcomes cannot be accurately represented by simple linear extrapolation.

4. Model

We develop and estimate an equilibrium model of the Brazilian college education market, covering both in-person and online formats. This model overcomes two key limitations of linear models: first, by explicitly modeling demand, it flexibly accounts for substitution across areas of study; second, by modeling supply, it supports counterfactual analyses that incorporate price adjustments and entry/exit decisions in equilibrium. Using this framework, we assess the impact of expanded online education on value added and explore targeted policies that restrict online education to older cohorts.

4.1. *Setup*

The model consists of institutions (i.e., firms) $f \in \mathcal{F}$ that offer undergraduate college degrees. Students can choose to major in a particular field of study $a \in \mathcal{A}$ by enrolling in either an in-person or online degree. A degree is defined as the combination of the firm offering it, the field of study it belongs to, and whether it is in-person or online, and is represented by $j \in \{\mathcal{F} \times \mathcal{A} \times \{0, 1\}\}$. Geographical regions are denoted by $r \in \mathcal{R}$ and years by $t \in \mathcal{T}$. Markets are the intersection of regions and years. A single degree j can be available in multiple markets, and a market may offer multiple degrees. Each firm offers a product bundle (i.e., a combination of degrees) $\mathcal{J}_{frt} \in \mathcal{B}_f$ in each market, where $\mathcal{J}_{frt} = \emptyset$ represents the option of not offering any degree and is available for every firm.[17]

The model has two stages. In the first stage, institutions decide whether to operate in a given market and select a bundle to offer by comparing expected profits to bundle-specific entry fixed costs. At this stage, institutions know the pre-existing market structure, $\mathcal{J}_{frt_0}$, all relevant characteristics of demand and marginal and fixed costs of potential entrants up to idiosyncratic shocks to demand and marginal costs, and their own private-information fixed-costs shocks. Institutions' entry choices determine the market structure. After entry decisions are made, demand and marginal cost shocks are realized. In the second stage, institutions compete a la Bertrand by setting tuition prices for the degrees in their chosen bundle, and demand is realized.

In the following sections, we present the components of the model in reverse order. First, we outline the demand model. Second, we discuss the institutions' pricing decisions. Third, we detail the institutions' optimal bundle choice.

4.2. *Demand*

In each year t and region r, a potential student, $i \in \mathcal{I}_{rt}$, decides whether to enroll in a degree $j \in \mathcal{J}_{rt}$, where $\mathcal{J}_{rt} = \cup_f \mathcal{J}_{frt}$ are the degrees available in year t and region r, or not to enroll at all. A degree j is characterized by the institution f that offers it, the field of study a it belongs to, and a vector of characteristics $x_j = [x_j^{(1)}, x_j^{(2)}]$ that we define below.

[17] The set of available bundles, $\mathcal{B}_f$, is firm specific to allow for firm specialization. For instance, some institutions may lack the necessary technology to offer degrees in certain fields, such as Medicine.

The utility student i gets from enrolling in degree j in region r in year t is:

$$u_{ijrt} = -\alpha_i p_{jrt} + x_j^{(1)}\beta_i + \psi w_{jrt} + \delta_{jrt} + \epsilon_{ijrt}, \tag{8}$$

where p_{jrt} denote the tuition price, $x_j^{(1)}$ is a vector that indicates whether the degree is online or in-person (i.e., $x_j^{(1)} \in \{[1,0],[0,1]\}$), w_{jrt} is a demand shifter based on region r's internet penetration in year t interacted with whether degree j is in person or online, and δ_{jrt} captures degree-market-specific characteristics constant across individuals. We allow individuals to have heterogenous preferences over tuition prices and delivery mode, assuming $[\log(\alpha_i), \beta_i]' \sim \mathcal{N}(\mu_{b(i)}, \Sigma)$, where $\mu_{b(i)}$ depends on the student's age bin $b(i)$. The term ϵ_{ijrt} represents a consumer-specific demand shock following a generalized extreme value distribution, consistent with a nested logit model. The nests are defined by the degree's area of study, and the intra-nest correlation is denoted by ρ.

We further decompose the degree-market-specific utility as follows:

$$\delta_{jrt} = x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{to_j} + \xi_{jrt}, \tag{9}$$

where $x_j^{(2)}$ include a constant, the age of the program, the average score of incoming students, the average wages of graduate students, the length of the degree (in number of hours required), and the degree's stem load, calculated as the share of degree programs comprising the degree that are stem. The term δ_j captures degree-specific components of utility, δ_{ra} captures area of study and region-specific factors, δ_{ta} represents area of study and year-specific components, δ_{to_j} account for online-year factors, allowing for differential yearly demand shifts between online and in-person education. We refer to these terms as the mean utility components of δ_{jrt}. Finally, ξ_{jrt} denotes a degree-region-year-specific idiosyncratic demand shock.

We denote s_{jrt} as the share of potential students from region r in year t who choose to enroll in degree j, and is calculated as:

$$s_{jrt}(\mathbf{p_{rt}}) = \int_{i\in\Theta_{rt}} di, \tag{10}$$

where $\mathbf{p_{rt}}$ denotes the market vector of degree prices, and $\Theta_{rt} = \{i \in \mathcal{I}_{rt} : u_{ijrt} \geq u_{ikrt}, \ \forall k \in \mathcal{J}_{rt}\}$ captures the set of potential students in region r and year t who choose to enroll in degree j.

4.2.1. *Identification and estimation*: We estimate the demand model with yearly data from 2010 to 2019 using the generalized method of moments introduced by Berry et al. (1995), combining both instrumental variables and micro-moments to identify the model parameters.[18]

Instruments for prices. A key challenge in demand estimation is price endogeneity, as institutions

[18]For a comprehensive review of related literature, see Berry and Haile (2016), and for a detailed guide on best practices, including those we adopt, refer to Conlon and Gortmaker (2020).

may set prices in response to unobserved demand shocks ξ_{jrt}. Ideally, we would rely on cost-shifters that affect pricing decisions but are independent of demand shocks. When these shifters are difficult to observe, proxies serve as a practical alternative. Following Hausman et al. (1994), we use the contemporaneous prices of the same degree in other regions as a proxy, defined as $z^p_{jrt} = \frac{1}{|\mathcal{R}(j)|-1} \sum_{r' \neq r} p_{jr't}$, which represents the average price of degree j in year t in regions other than r. The intuition behind this instrument is that variation in firm-level costs impacts prices across all markets where the degree is offered. The primary identification assumption is that while costs for degree j are correlated across regions, demand shocks are not.

Instruments for substitution patterns. The demand model incorporates individual heterogeneity through random coefficients on α_i and β_i and the nesting parameter ρ. To identify these parameters, we use instruments that shift the choice set, $\mathcal{J}_{rt}$, but are unrelated to demand. To construct these instruments, we employ shift-share type instruments similar to the one outlined in Section 3.2. Specifically, we build two instruments designed to predict the number of online degrees, z^o_{rta}, and the number of in-person degrees, z^ι_{rta}, offered from a given field of study in each market. Each instrument is calculated as follows:

$$z^o_{rta} = \sum_f z_{fr} z_{fa} z_a N^o_{ft} \quad , \quad z^\iota_{rta} = \sum_f z_{fr} z_{fa} N^\iota_{ft}$$

where z_{fr} represents the likelihood of institution f expanding to region r, based on the distance between the institution's headquarters and the region as defined in Section 3.2.2; z_{fa} captures the institution's propensity to expand into different fields of study as outlined in Section 3.2.3, z_a is the indicator variable (defined in Section 3.2.4) that takes a value of zero for fields of study where online instruction is not allowed, and N^o_{ft} and N^ι_{ft} represent the total number of online and in-person degrees offered by institution f in year t across all regions and fields of study.

These instruments shift the total number of online and in-person degrees available in region r, year t and field of study a, respectively, providing variation to identify the variance of β_i, the random coefficients on the two indicator variables for online or in-person degrees, and the nesting parameter ρ. Finally, following Gandhi and Houde (2019), we construct differentiation instruments as functions of the distance between degree j's predicted price and other degrees' predicted prices given by $z^{gh1}_{jrt} = \frac{1}{|\mathcal{J}_{rt}|-1} \sum_{k \in \mathcal{J}_{rt}} |\hat{p}_{jrt} - \hat{p}_{krt}|$ and $z^{gh2}_{jrt} = \frac{1}{|\mathcal{J}_{rt}|-1} \sum_{k \in \mathcal{J}_{rt}} \mathbb{1}\{|\hat{p}_{jrt} - \hat{p}_{krt}| < \sigma_{\hat{p}}\}$, where $\hat{p}_{jrt}$ is the predicted price based the reduced-form pricing equation constructed from the price instrument and fixed effects and $\sigma_{\hat{p}}$ is the standard deviation of $\hat{p}_{jrt}$.[19] These instruments generate variation that helps identify the variance of the random coefficient on price, α_i, and the nesting parameter ρ.

Micro-moments for age heterogeneity. We incorporate additional micro-moments to discipline the age heterogeneity parameters, $\mu_{b(i)}$, by matching moments predicted by the model with their

[19] The reduced form regression is given by $p_{jrt} = \zeta_p z^p_{jrt} + \zeta_w w_{jrt} + \zeta_j + \zeta_{ra} + \zeta_{ta} + \zeta_{to_j} + \varepsilon_{jrt}$, where z^p_{jrt} is the instrument for price, and ζ_j, ζ_{ra}, ζ_{ta}, and ζ_{to_j} are degree, region-field of study, year-field of study, and year-online fixed effects.

empirical counterparts. We use eight moments defined by the probability that students from each of four age bins defined above choose to enroll in person or online.[20]

Estimator. The estimator proposed by Berry et al. (1995) produces consistent estimates of $\mu_{b(i)}$, Σ, ψ, and δ_{jrt}, which together recover the distribution of α_i, β_i, market shares, and price elasticities. The estimator treats the mean utility components of δ_{jrt} from Equation (9) as nuisance parameters that might be unprecisely estimated. This is typically not an issue because market shares and price elasticities depend on δ_{jrt} and not on the mean utility components of δ_{jrt}. However, in our setting, we are interested in recovering demand for degrees that are not offered in certain markets in the data and for which we don't have estimates of δ_{jrt}. This requires reliable estimates of the mean utility components of δ_{jrt}.

To estimate the demand model, we proceed in two steps. In the first step, we follow Berry et al. (1995) and Petrin (2002), and estimate the parameters from Equation (8) using the generalized method of moments. We absorb the mean utility components of δ_{jrt} via fixed effects. Due to collinearity, $x_j^{(2)}$ is absorbed by δ_j during the estimation. We use a total of thirteen moments, five moments defined by $\mathbb{E}[\xi_{jrt}|Z_{jrt}] = 0$, where ξ_{jrt} is the demand shock from Equation (9), and Z_{jrt} are the instruments z_{jrt}^{p}, z_{rt}^{o}, z_{rt}^{ι}, z_{rta}^{a}, and z_{jrt}^{gh}, and the eight micro-moments defined above.

In the second step, we impose additional structure on Equation (9) and estimate a mixed-effects Bayesian hierarchical model to recover each of the mean utility components of δ_{jrt}. We assume that $\delta_j \sim \mathcal{N}(0, \sigma_j^2)$, $\delta_{ra} \sim \mathcal{N}(0, \sigma_{ra}^2)$, $\delta_{ta} \sim \mathcal{N}(0, \sigma_{ta}^2)$, $\delta_{to} \sim \mathcal{N}(0, \sigma_{to}^2)$, and $\xi_{jrt} \sim \mathcal{N}(0, \sigma_{\xi}^2)$ are random coefficients, and δ is a non-random parameter that allows for flexible correlation between δ_{jrt} and $x_j^{(2)}$. We estimate the model via maximum likelihood to recover posterior means of the mean utility components of δ_{jrt} that we use to project utility into degree-markets, regardless of whether they are offered in the data or not.[21] This second step is in the spirit of Abdulkadiroğlu et al. (2020) and Andrews et al. (2024), and serves to shrink imprecise estimates, reducing noise in the profit function and improving the estimation of the entry game.

4.2.2. *Results*: We present the estimated parameters in Online Appendix A, Table A.3. To summarize our results, we present median own-price elasticities and diversion ratios for 2010 and 2019 in Table 4. We estimate median own-price elasticities of approximately -2.9 for in-person degrees and -1.1 for online degrees, which aligns with findings from the literature that estimates demand for in-person degrees in the U.S. (Armona and Cao, 2024) and Brazil (Dobbin

[20]Formally, the moments are defined as $\mathbb{E}[\sum_{j \in \mathcal{J}_{rt}} s_{ijrt} x_j^{(1)}(k) | i \in b]$ for $k \in \{1, 2\}$, where q_{ijrt} is the probability that individual i chooses degree j, $x_j^{(1)}(k)$ is the k-th element of $x_j^{(1)}$ (i.e., online or in-person), and b represents age bins corresponding to the age groups 18-20, 21-25, 26-35, and 36-45.

[21]In a standard fixed-effects model, estimating δ when x_j and δ_j are collinear is problematic because δ_j would be treated as a fixed parameter, absorbing variation in x_j. However, in a mixed-effects framework, estimation is possible. The mixed model is estimated using maximum likelihood, where the key idea is that δ_j is integrated out in the likelihood function, leaving only the fixed parameters δ and variance components of the random effects to be estimated. Intuitively, within-j variation in δ_{jrt} helps estimate δ, while between-j variation helps estimate σ_j^2.

et al., 2021).[22] We find that, by 2019, when the price of an in-person degree marginally increases, 67% of students who leave that degree would enroll in another in-person program, 18% would switch to an online program, and 14% would exit higher education altogether. Additionally, we find that older cohorts show stronger preferences for attending online degrees (see Online Appendix A, Table A.4). Finally, our findings indicate that higher internet penetration increases demand for both online and in-person degrees, which we use as an additional source of variation to estimate entry fixed cost in the next subsections.

Table 4: Elasticity and diversion ratios

	$t = 2010$		$t = 2019$	
	In-person	Online	In-person	Online
Median own-price elasticity:	-2.70	-1.10	-2.91	-1.14
Median diversion ratios:				
To in-person:	0.78	0.34	0.67	0.16
To online:	0.04	0.49	0.18	0.70
To outside good:	0.15	0.15	0.14	0.14

Notes: The table reports median own-price elasticities and diversion ratios across products and markets for 2010 and 2019 for in-person and online degrees. We calculate diversion ratios as the share of students that decide to stop attending degree j upon an increase in tuition price that would switch to either an in-person degree, an online degree, or the outside option. Formally, we calculate diversion ratios as $D_{j\mathcal{K}} = \left(\left|\frac{\partial s_j}{\partial p_j}\right|\right)^{-1}\left(\sum_{k\in\mathcal{K}/j}\frac{\partial s_k}{\partial p_j}\right)$, where $\mathcal{K}$ the set of all degrees that are either in-person, online, or the outside option. For example, the diversion ratio from in-person degrees to online degrees in 2010 is 0.04, which is given by the median of $D_{j\mathcal{K}}$ across all degrees j that are in-person and for $\mathcal{K}$ the set of all online degrees.

4.3. *Institutions' pricing decision*

The variable profits of institution f in region r and year t, given its choice of degree offerings $\mathcal{J}_{frt}$, are expressed as:

$$\pi_{frt}(\mathcal{J}_{frt}) = \max_{\{p_{jrt}\}_{j\in\mathcal{J}_{frt}}} \sum_{j\in\mathcal{J}_{frt}} (p_{jrt} - c_{jrt}) \cdot s_{jrt}(\mathbf{p}_{rt}). \tag{11}$$

The log of marginal cost c_{jrt} is given by:

$$\log(c_{jrt}) \;=\; g(d_{jrt})'\gamma_g + \gamma_z z^p_{jrt} + \underbrace{x_j^{(2)}\gamma_x + \gamma_j + \gamma_{ra} + \gamma_{ta} + \gamma_{to} + \omega_{jrt}}_{\gamma_{jrt}}, \tag{12}$$

where the vector $g(d_{fr}) = [\log(1 + d_{fr}), H_{fr}]'$ captures the log-distance between institution f's headquarters and region r, as well as an indicator for whether region r corresponds to the

[22] To the best of our knowledge, no studies have reported price elasticities for online degrees.

headquarter's location, as in Equation (5). The term z^p_{jrt} represents the price instrument used to estimate demand, $x^{(2)}_j$ denotes the degree characteristics from Equation (9), γ_j captures degree-specific components of the cost function, γ_{ta} represent area of study-year specific components, γ_{to} reflects online-year specific components that allow for differential trends in demand for online and in-person education. Lastly, ω_{jrt} is a degree-region-year specific idiosyncratic supply shock. Institutions compete a la Bertrand.

4.3.1. *Estimation*: We estimate the cost parameters in three steps. First, we recover c_{jrt} for all degrees offered in the data by inverting the firms' first-order conditions. Second, we estimate γ_g and γ_z via OLS and absorb the elements of γ_{jrt} as fixed effects. Analogous to the demand model estimation, the components of γ_{jrt} are treated as nuisance parameters and are not estimated precisely. Because we need to estimate costs for degrees that may not be offered in certain markets in the data, we require reliable estimates for each component of γ_{jrt}. We do this in a third step, where we estimate a mixed-effects Bayesian hierarchical model where

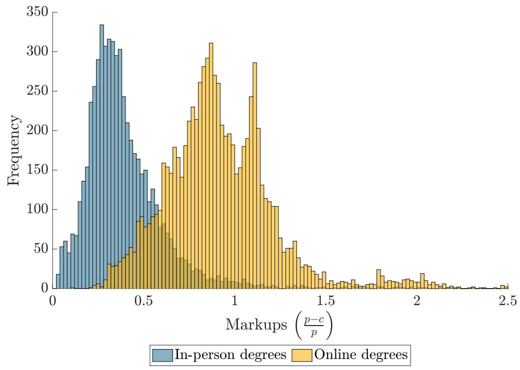
$$\gamma_{jrt} \quad = \quad x^{(2)}_j \gamma + \gamma_j + \gamma_{ra} + \gamma_{ta} + \gamma_{to} + \omega_{jrt}, \tag{13}$$

with $\gamma_j \sim \mathcal{N}(0, \varsigma^2_j)$, $\gamma_{ra} \sim \mathcal{N}(0, \varsigma^2_{ra})$, $\gamma_{ta} \sim \mathcal{N}(0, \varsigma^2_{ta})$, $\gamma_{to} \sim \mathcal{N}(0, \varsigma^2_{to})$, and $\omega_{jrt} \sim \mathcal{N}(0, \varsigma^2_{\omega})$ random coefficients, and γ a non-random parameter that allows for flexible correlation between γ_{jrt} and $x^{(2)}_j$. We estimate the model via maximum likelihood to recover posterior means of each component of γ_{jrt} that we use to estimate marginal costs for all degree-markets, regardless of whether they are offered in the data or not.

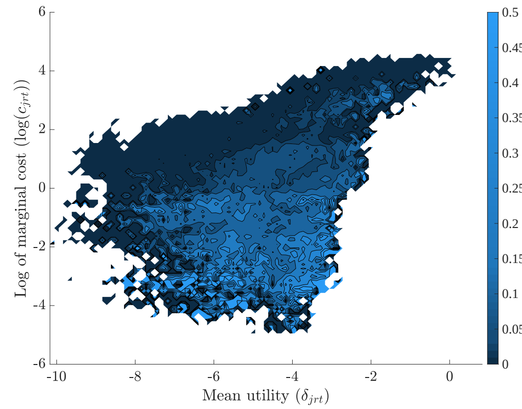
4.3.2. *Results*: We present the distribution of markups—defined as the ratio of price minus marginal cost to price—for in-person and online degrees in Figure 5(a). We present the corresponding estimated parameters in Online Appendix A, Table A.5. We find that online degrees have higher markups compared to in-person degrees, consistent with institutions pricing their online alternatives at a more inelastic part of the demand curve. However, despite having larger markups, online degrees' prices are lower due to lower marginal costs. The average price of in-person and online degrees is \$5650 and \$1821 per year, respectively, and the average estimated marginal cost is \$3966 and \$303 per year, respectively.

We also find that degrees that face higher demand—through a higher value of the mean utility, δ_{jrt}—and a lower marginal cost, are offered in more regions in 2019. Figure 5(b) shows a heatmap with the relationship between degrees' demand, marginal cost, and their empirical probability of being offered in 2019. Higher probabilities are more prevalent in the south-east region of the map. This finding suggests that, in the data, institutions are choosing more profitable degrees when deciding which bundles to offer, which is consistent with economic theory.[23]

[23] This result is not mechanically driven from the model. Entry probabilities are estimated directly from the data, while demand and marginal costs are estimated from market shares and prices from existing degrees.



(a) Distribution of markups



(b) Probability of a degree being offered in each region in 2019

Figure 5: Estimated parameters from demand and pricing decision

Notes: The figure summarizes the estimated demand and marginal cost functions. Panel (a) shows the distribution of markups, defined as the ratio between the difference between prices and marginal costs and prices, for in-person and online degrees. In-person degree are depicted in blue and online degrees in yellow. Panel (b) shows the relationship between demand and marginal costs determinants and the empirical probability that a given degree is offered in each market. The x-axis represents the mean utility components of δ_{jrt} of each degree and the y-axis the logarithm of their predicted marginal cost. The shaded area represents the empirical probability that degrees with given characteristics are offered in each region in 2019. Lighter colors reflect larger probabilities. Uncolored areas are combinations of mean utility and marginal costs for which no potential degree exists. Note that entry probabilities are directly estimated from the data and are agnostic about the underlying entry model.

4.4. *Institutions' choice of degrees' offerings*

In this subsection, we model institutions offerings decisions in 2019, denoted by $\mathcal{J}_{frt}$, based on the existing degree offerings in 2010, denoted by $\mathcal{J}_{frt_0}$. We take a static approach to the problem for two reasons. First, there are adjustment frictions that make the timing of institutions' responses to economic shocks uncertain. We avoid modeling such frictions by comparing medium-run changes in market structure over a span of nine years. Second, in our setting, there are many institutions that can offer multiple degrees in each market, which makes the state space highly dimensional and the problem intractable.[24] Our static approach captures the main trade-offs institutions face when choosing offerings in a parsimonious way.

Institution f's fixed cost of offering a given bundle, $\mathcal{J}_{frt}$, depends on three key components: (i) the composition of the bundle and whether its degrees were part of the previous bundle, $\mathcal{J}_{frt_0}$, (ii) the infrastructure investments required to offer the degrees in the bundle, which depend on the need for an in-person campus or online hub, the existence of prior infrastructure, and the distance from f's headquarters to region r, and (iii) a private-information cost shock only known

[24] Bodéré (2023) develops an approximation method for high-dimensional dynamic games with single-product firms. Our setting features multi-product firms, which complicates the approximation of the state space due to the need to account for interactions across multiple products.

by institution f. The fixed cost function is parametrized as follows:

$$FC_{fr}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) = \text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) + \text{Infrastructure}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) + \sigma_\varepsilon \varepsilon_{fr\mathcal{J}_{frt}}, \tag{14}$$

where $\varepsilon_{fr\mathcal{J}_{frt}}$ is a firm-region-bundle specific idiosyncratic shock that is only observed by institution f and that we assume to follow an extreme value type I distribution. These shocks, unobserved by the researcher, help to rationalize firms' bundle choices observed in the data, and can include firms' private information about the economic returns of offering certain degrees or expanding into specific regions.

The first component of the fixed cost equation represents the cost associated with maintaining existing degrees or opening new ones:

$$\text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) = \mathbb{1}\{\mathcal{J}_{frt} \neq \emptyset\} \cdot \kappa_0 + \sum_{k \in \mathcal{J}_{frt} \cap \mathcal{J}_{frt_0}} \kappa^{\text{old}}_{a(k)o(k)} + \sum_{k \in \mathcal{J}_{frt} \setminus \mathcal{J}_{frt_0}} \kappa^{\text{new}}_{a(k)o(k)}, \tag{15}$$

where κ_0 captures an operational fixed cost that turns on when at least one degree is offered, $\kappa^{\text{old}}_{a(k)o(k)}$ represents the cost of maintaining an existing degree, and $\kappa^{\text{new}}_{a(k)o(k)}$ denotes the cost of opening a new degree using the existing infrastructure. Note that these costs only vary at the field of study level, indexed by $a(k)$ for each degree k, and whether the degree is in person or online, indexed by $o(k)$. We restrict κ_0, κ^{old}_{ao} and κ^{new}_{ao} to be non-negative so that offering degrees is more expensive than not offering them.

The second component reflects the cost associated with opening a new campus or hub necessary to host a given degree:

$$\text{Infrastructure}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) = \sum_{k \in \{\text{campus,hub}\}} \mathbb{1}\{\text{New}_{frk}\} \left(\chi^k_0 + g(d_{fr})'\chi^k\right), \tag{16}$$

where $\mathbb{1}\{\text{New}_{frk}\}$, for $k \in \{\text{campus, hub}\}$, is an indicator variable that equals 1 if $\mathcal{J}_{frt}$ includes an in-person or online degree (respectively), while $\mathcal{J}_{frt_0}$ does not. This variable captures the transition in a firm's offerings in region r, reflecting the need to invest in building the required campus infrastructure for in-person degrees if the firm previously did not offer any in-person programs, or in developing online hubs if the firm had not previously provided online degrees. The vector $g(d_{fr})$ is the distance function between the firm's headquarters and the region, as defined in Equation (5). To account for differences in cost, we allow χ^k_0 and χ^k to vary between building a new campus or a new online hub. As before, we impose the restriction that $\left(\chi^k_0 + g(d_{fr})'\chi^k\right)$ is non-negative to ensure that building new infrastructure is more costly than utilizing existing infrastructure.

Firms decide which degree bundles to offer after observing the region-bundle-specific idiosyncratic shocks, $\varepsilon_{fr\mathcal{J}_{frt}}$, but before the demand and supply shocks, ξ_{jrt} and ω_{jrt}, are realized.

Their objective is to maximize expected profits, as defined by the firm's problem:

$$\Pi_{frt} \quad = \max_{\mathcal{J}_{frt} \in \mathcal{B}_f} \quad \mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})] - FC_{fr}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}), \tag{17}$$

where $\pi_{frt}(\mathcal{J}_{frt})$ is given by Equation (11) and the expectation is taken over the distribution of all other firms' idiosyncratic shocks, $\varepsilon_{f'r\mathcal{J}_{f'rt}}$, and over the demand and supply shocks, ξ_{jrt} and ω_{jrt}. We assume the players' strategies form a Bayesian Nash equilibrium.

4.4.1. *Identification and estimation*: To identify the fixed cost parameters, we exploit variation in the firms' bundle choices across different markets. Nonparametric identification of the fixed costs requires instruments that shift firms' expected profits but do not enter the fixed cost function directly. We use two sets of profit shifters. The first set of instruments corresponds to the demand shifters, w_{jrt}, based on region r's internet penetration in year t interacted with whether degree j is in person or online as outlined in Section 4.2. In Online Appendix C.1, we show that higher internet penetration growth is associated with higher online entry. The identification assumption is that while greater internet penetration may increase demand for online programs, it does not affect firms' fixed cost of offering different degrees. The second set of instruments is the vector of distances between region r and institution f's competitors. Firm f's expected profits will be lower in regions that are close to the headquarters of other firms that offer degrees similar to firm f. However, we only allow fixed costs to depend on firm f's distance to region r and not that of its competitors.[25]

Our identification follows the framework for inference in incomplete-information games from Aradillas-López (2020). In their framework, point identification can be obtained by assuming that the underlying equilibrium selection mechanism is degenerate (i.e., the data come from a single equilibrium), without having to assume which equilibrium is chosen (Seim, 2006; Bajari et al., 2010; Atal et al., 2025). We use the firm's maximization problem to recover bundle choice probabilities. From Equation (17), the probability that firm f chooses bundle $\mathcal{J}_{frt}$ is given by

$$\phi_{\mathcal{J}_{frt}} = \frac{\exp(\frac{1}{\sigma_\varepsilon}\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})] - \text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) - \text{Infrastucture}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}))}{\sum_{\mathcal{J}\in\mathcal{B}_f}\exp(\frac{1}{\sigma_\varepsilon}\mathbb{E}[\pi_{frt}(\mathcal{J})] - \text{Degrees}(\mathcal{J}|\mathcal{J}_{frt_0}) - \text{Infrastucture}(\mathcal{J}|\mathcal{J}_{frt_0}))}. \tag{18}$$

If expected profits were observed and did not depend on competitors' strategies, estimation of the fixed cost parameters via maximum likelihood would be straightforward. However, the expected profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$, on the right-hand side of Equation (18) are a function of the choice probabilities, $\phi_{\mathcal{J}_{frt}}$, on the left-hand side of the same equation. Because of that, a common concern related to discrete games with incomplete information is the potential for multiple equilibria (i.e., multiple solutions of Equation (18)), which complicates both estimation and counterfactuals.[26] To avoid this issue in the estimation stage, we follow Sweeting (2009) and

[25]Our identification assumption could be violated if competitors' entry into markets affect firms' fixed costs by changing local input prices (e.g., by increasing the cost of acquiring new online hubs).

[26]Multiple equilibria are more prevalent in complete-information games, where assuming a degenerate equi-

estimate the parameters in two steps.

In the first step, we compute choice probabilities directly from the data. Specifically, we estimate a multinomial choice logit model using all own and competitors' drivers of variable profits, which include the demand mean utilities, δ_{jrt}, evaluated at $\xi_{jrt} = 0$, the demand shifters, w_{jrt}, the marginal costs, c_{jrt}, evaluated at $\omega_{jrt} = 0$, and the determinants of fixed costs from Equation (14), to predict the probability that each bundle is offered in the data. We denote the estimated probabilities by $\hat{\phi}_{\mathcal{J}_{frt}}$. We provide more details about the prediction model in Online Appendix C.2.

We use the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, to estimate the expected variable profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$. To do that, we randomly draw own and competitors' bundle choices, $\{\mathcal{J}_{frt}, \mathcal{J}_{f'rt}\}$, using the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, and demand and supply shocks, ξ_{jrt} and ω_{jrt} from their respective empirical distributions. For each draw, we compute variable profits solving for the static equilibrium game given by Equation (11). We do this 10,000 times per market. We compute the expected variable profits for each bundle choice by integrating over all simulation draws. Finally, we fit a random forest prediction model to reduce noise and get stable predictions for bundles that have low probability of being chosen. This gives us viable estimators of $\mathbb{E}[\pi_{frt}(\mathcal{J})]$ for all bundles $\mathcal{J} \in \mathcal{B}_f$ for all firms f. We denote the estimated expected variable profit by $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$.

In the second step, we use the estimated expected variable profits, $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$, and the equilibrium conditions on entry probabilities from Equation (18) to recover the fixed cost parameters via maximum likelihood, replacing $\mathbb{E}[\pi_{frt}(\cdot)]$ with $\hat{\mathbb{E}}[\pi_{frt}(\cdot)]$ in Equation (18).

4.4.2. *Results*: We present the estimated parameters in Online Appendix A, Table A.6. We estimate a median entry elasticity with respect to own profits of 1.2, meaning that the probability of offering a given bundle increases by 1.2% for each percentage increase in expected profits. We also find that adding a new degree to the bundle has larger cost than continuing to offer an existing degree. For many areas of study, offering an existing degree does not have an additional fixed cost (i.e., $\kappa_{ao}^{\text{old}} = 0$), which explains why, conditional on keeping their campus or hub open, institutions keep offering degrees in 2019 that they were already offering in 2010. We also find important costs of building new infrastructure that are increasing in the distance between regions and firms' headquarters. The cost of building a new in-person campuses or online hub in a region 1200 kilometers away from the institution's headquarters—the median distance in the data—is \$42 and \$18 per person living in the market, respectively. We show the full distribution of fixed costs associated with opening a new campus or hub in every region as a function of distance in Online Appendix A, Figure A.2.

librium selection mechanism is not enough for identification (Aradillas-López, 2020). Note that counterfactual analysis still requires assumptions about which equilibrium is selected.

5. Counterfactuals

We use the model to assess the impact of introducing online education, isolating the effects of demand, pricing, and program offerings. Additionally, we explore policies that restrict online education for specific groups, examining how these measures could allow in-person programs to remain in the market while ensuring online access for those who benefit most.

5.1. *The effects of the online education expansion: the role of supply and demand*

Using our model, we compare the 2019 status quo, referred to as *Baseline* (BL), where online education exists, with counterfactuals where online education does not exist. To isolate supply and demand effect, we consider increasingly flexible counterfactuals. We describe these below and summarize them in Table 5.

1. No supply-side responses (NS): we remove online degrees while restricting institutions from adjusting tuition or program offerings. Value added under NS depends on two factors: (i) the value added differences between online degrees, in-person degrees, and the outside option, and (ii) the extent to which students switch from online degrees to in-person programs or to the outside option.

2. Price responses (PR): institutions are allowed to set prices optimally, though their program offerings stay constant. This scenario helps measure the effects of reduced competition, which may lead to increased prices and potentially displace students from high-quality in-person degrees.

3. Equilibrium (EQ): In the final counterfactual, institutions have the flexibility to decide which degrees to offer in each market, corresponding to the equilibrium model outlined in Section 4. Under EQ, the availability of in-person degrees is expected to expand, potentially boosting enrollment in these programs and increasing total value added. We summarize the counterfactuals in Table 5. We present the algorithm we use to find the market equilibrium of each counterfactual in Online Appendix C.3.

Table 5: Counterfactual exercises

Counterfactual	**Description**
Baseline (BL)	Free entry of online education
No supply-side responses (NS)	No online education & no supply-side responses
Price responses (PR)	(NS) + firms choose prices
Equilibrium (EQ)	(NS) + (PR) + firms choose degree offerings

Notes: This table summarizes the main counterfactuals simulated in Section 5.

We examine six different outcome categories: (i) the number of online and in-person degrees, (ii) total college enrollment levels, separated by in-person and online enrollment, (iii) the total value added produced in the economy, (iv) total expenditure in higher education, (iv) the average price of in-person degrees, (v) total variable profits, and (vi) revealed-preferences consumer surplus.[27] Detailed formulas for constructing the last five outcome measures are provided in Online Appendix C.4.

Results are presented in Table 6. Under the Baseline scenario (BL) with the presence of online education, an average market has 47.8 online degrees and 41.7 in-person degrees offered. Total college enrollment reaches 5.9%, with 57.4% of students attending in person and 42.6% attending online. The total value added per enrolled student under BL is 0.137 log-points and student spend an average of $4211 in tuition. On the firm side, the average price of in-person degrees is approximately US$6500 per year, and average industry profits per enrolled student are US$1742. Finally, consumer surplus per enrolled student is $9604 higher than in a world without higher education. These numbers compare well to the data, reported in Table 6, Column (0). Differences arise from the fact that our counterfactuals take an average over 10,000 simulations, while the observed data correspond to a single realization of shocks.

Table 6: The effects of the online education expansion

	DT (0)	BL (1)	NS (2)	PR (3)	EQ (4)	$\Delta_{\text{EQ/BL}}$ (5)
Number of online degrees:	72.1	47.8	0	0	0	—
Number of in-person degrees:	52.3	41.7	41.7	41.7	48.9	17.3%
Total college enrollment:	6.3%	5.9%	5.1%	5%	5.2%	-11.9%
In-person enrollment:	3.2%	3.4%	5.1%	5%	5.2%	53.4%
Total value-added:	0.131	0.137	0.138	0.136	0.142	3.4%
Total expenditure:	4134	4211	4448	4502	4590	9%
Value-added per dollar spent:	0.033	0.033	0.031	0.03	0.031	-6.3%
Av. price of in-person degrees:	5375	6530	6530	7420	6721	2.9%
Total profits:	—	1742	1504	1589	1618	-7.1%
Consumer surplus:	—	9604	8261	8107	8464	-11.9%

Notes: This table reports the value of different outcomes for each counterfactual from . We report the average value across markets, weighted by market size, for seven different outcomes: the number of online and in-person degrees, total college enrollment and in-person enrollment, the total value added produced in the economy, the average price of in-person degrees, and total variable profits. Detailed formulas for constructing the last three outcome measures are provided in Online Appendix C.4. Column (0) reports the values in the original data. Columns (1)-(4) present the value for each counterfactual from Table 5, averaged across 1,000 simulations. Column (5) reports the average value for the percentual changes between the *Equilibrium* counterfactual and the *Baseline* counterfactual, given by $\Delta_{\text{EQ/BL}} = \frac{EQ-BL}{BL}$.

When online degrees are removed from the market, and institutions are unable to adjust prices or alter their degree offerings (i.e., under the NS scenario), total enrollment declines by

[27] Our consumer surplus measure assumes that students make informed decisions and enroll in degrees that maximize their welfare. This assumption contrasts with policymakers' concerns regarding the challenges students may face in identifying low-quality degrees.

14% (dropping from 5.9% to 5.1%). This decline underscores the role of online education in broadening access for students who might otherwise not attend college. Additionally, because online programs can attract students who might have chosen in-person degrees, restricting the supply of online education leads to a 49.8% increase in in-person enrollment (rising from 3.4% to 5.1%). This drop in overall enrollment is accompanied by a corresponding 5.6% increase and 13.6% decrease in consumer expenditure and firms' profits, respectively.

The impact on value added is ambiguous when online education is restricted. On one hand, since online degrees provide higher value added than not attending college, the reduction in total enrollment in higher education could lower overall value added. On the other hand, because online degrees generally have lower value added than in-person programs, the shift toward greater in-person enrollment could raise total value added. Overall, we observe a small increase in value added of 0.4%. Consumer surplus, measured by revealed preferences, decreases by 14% due to the lower number of available options.

When institutions are allowed to respond to the absence of online degrees by adjusting the prices of their in-person programs (i.e., under the PR scenario), tuition prices increase by 13.6%. This rise in prices pushes students away from higher-quality degrees, resulting in a decrease in total value added of 1%.

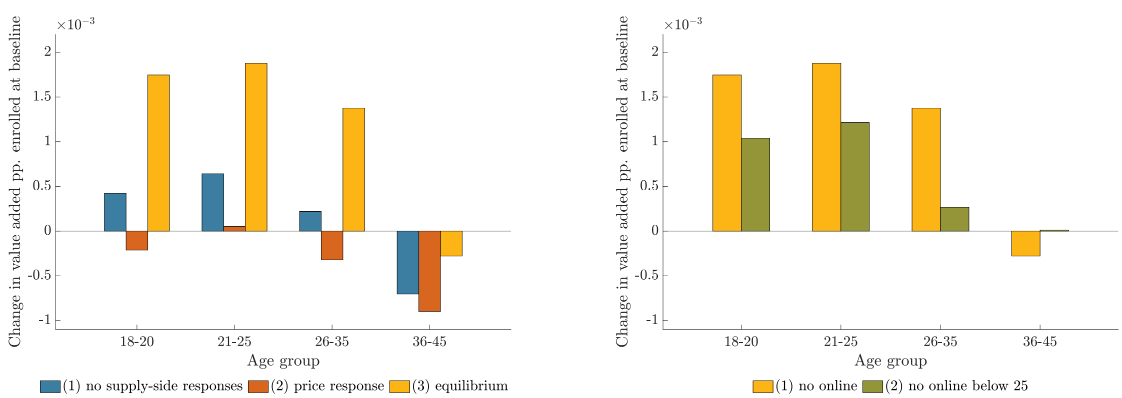
When institutions are allowed to adjust their degree offerings (i.e., under the EQ scenario), they expand the supply of in-person options by 17.3%, increasing from an average of 41.7 degrees to 48.9 degrees per market. Moving from the baseline counterfactual to the full equilibrium scenario, total enrollment decreases by approximately 11.9% (from 5.9% to 5.2%), while in-person enrollment rises by 53.4% (from 3.4% to 5.2%). In terms of value added, the equilibrium scenario shows an improvement relative to the baseline, with an increase of 3.4%, reaching 0.142 log-points. However, total expenditure increases by 9%, decreasing the value added per dollar spent by 6.3%. Additionally, the average tuition price for in-person degrees rises by 2.9% relative to the baseline counterfactual. However, the increased markups are insufficient to fully offset the profit losses resulting from the absence of online degree offerings, leading to an average profit decrease of 7.1%. Finally, consumer surplus, measured by revealed preferences, decreases by 11.9%.

These results highlight the benefits and risks associated with the expansion of online education. On the one hand, the availability of online options increases college enrollment, reduces total expenditure, and increases consumer surplus, defined as the area below the demand curve. On the other hand, the online expansion diverts students from high value-added in-person alternatives, resulting in a reduction in total value added. Even if students are not fully informed about the quality of the degrees they choose, the flexibility and cost savings offered by online education may offset the decline in total value added.

5.2. *Unpacking value added changes by age groups*

The overall changes in total value added from the previous subsection mask important heterogeneity between those who benefit and those who lose out. Older individuals, who might not attend college without online education, are the primary winners of its expansion. In contrast, younger individuals are more often diverted from higher-quality, in-person programs. Diversion from those degrees can be particularly detrimental if online education induces exit or deters entry of in-person degrees that young individuals would prefer, even in the presence of online options.

Figure 6(a) illustrates the heterogeneous effects on value added by cohort, showing changes in total value added for each age group relative to the baseline counterfactual. When online education is restricted and supply-side adjustments are not allowed, the 18-25 age cohort experiences a slight increase in value added, while the 36-45 cohort faces significant decreases. Allowing institutions to adjust prices results in value added reductions across all cohorts. Only in the equilibrium counterfactual, where institutions can adjust their degree offerings, the 18-35 age group sees substantial gains in total value added.



(a) Value added changes decomposition

(b) Targeted ban to 18-25 years old students

Figure 6: Changes in total value added by age group under each counterfactual

Notes: This figure presents changes in total value added for different age groups under different counterfactuals. In panel (a), we show changes in value added relative to the *Baseline* counterfactual for all counterfactuals presented in Table 5. In panel (b) we show the changes in value added relative to the *Baseline* counterfactual for the *Equilibrium* counterfactual and an additional counterfactual in which we restrict access to online education for students aged 18-25. In both panels, total value is divided by the number of enrolled students in the *Baseline* couterfactual.

In Online Appendix A, Figure A.3, we show heterogenous results for total enrollment, value added per dollar spent in higher education, and revealed-preferences consumer surplus. We find that the expansion of online education benefits all age groups in these three dimensions, with large benefits for individuals above 25 years old and small benefits for individuals 18-20 years old.

5.3. *Targeted policies*

Next, we explore hypothetical policies designed to harness the benefits of online education for older students while preserving value added for younger cohorts. Specifically, we examine a policy that restricts access to online education for students aged 18-25. By directing younger cohorts toward in-person programs, this policy leaves sufficient profits for in-person degrees, enabling them to coexist with online options. Consequently, younger students can pursue high-value-added in-person degrees, while older students continue to benefit from the expanded access offered by online education. Our main results, shown in Figure 6(b), indicate that, compared to the baseline counterfactual, this targeted policy raises value added across all cohorts. In Online Appendix A, Figure A.3, we show that restricting online education in this targeted way has a smaller negative effect on other outcomes of interest.

6. Conclusion and discussion

In this paper, we examine the equilibrium effects of the rapid growth of online college education in Brazil, highlighting two key findings. First, the expansion of online degrees creates a dual impact: it broadens access to higher education for new students but simultaneously diverts potential students away from in-person options. Second, the increased availability of online degrees intensifies competition, which drives down prices for in-person programs and deters new in-person entrants.

Using an equilibrium model of supply and demand, we quantify the impact of online education on total value added and find significant heterogeneity across age groups. A ban on online degrees would benefit younger cohorts by preserving access to higher-quality, in-person programs, while older cohorts, who benefit from the flexibility and availability of online options, would face reduced college access. A targeted policy restricting online education for younger students would increase overall value added across all cohorts compared to an unrestricted baseline.

Our analysis highlights how introducing lower-cost alternatives can reshape competition and market structure, with potential adverse effects in settings with imperfect information. Consumers may unknowingly substitute lower-quality online degrees for traditional options, particularly as in-person providers become less available in the market. Policymakers could help counter these effects by restricting online education access for groups that would gain more from traditional degrees, thereby sustaining demand for incumbent high-quality institutions and preventing them from leaving the market.

While this paper addresses critical aspects of online education's recent expansion, several open questions remain. First, this study assumes that degree value added is fixed and policy invariant, though a surge in online degrees could saturate the labor market, reducing returns to college education. Modeling college education as an equilibrium of supply and demand interacting with labor market outcomes presents an exciting path for future research. Second, technological advancements could further enhance the cost-effectiveness and quality of online education, possibly altering some conclusions of this study. Lastly, non-pecuniary benefits—such as trust, networking, and social skills—are more effectively developed in in-person settings, where students engage in face-to-face interactions. These social benefits may be harder to replicate in online formats, suggesting that traditional college experiences offer unique non-monetary advantages, especially valuable for younger students, which could further support this study's findings.

References

Abdulkadiroğlu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2020). Do parents value school effectiveness? *American Economic Review 110*(5), 1502–1539.

ABED (2018). Censo ead.br: relatório analitico da aprendizagem a distância no brasil 2017. *National Center for Education Statistics*.

Ainsworth, R., R. Dehejia, C. Pop-Eleches, and M. Urquiola (2023). Why do households leave school value added on the table? The roles of information and preferences. *American Economic Review 113*(4), 1049–82.

Allende, C. (2019). Competition under social interactions and the design of education policies. *Working Paper*.

Andrews, I., T. Kitagawa, and A. McCloskey (2024). Inference on winners. *The Quarterly Journal of Economics 139*(1), 305–358.

Angrist, J., P. Hull, and C. Walters (2023). Methods for measuring school effectiveness. *Handbook of the Economics of Education 7*, 1–60.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2017). Leveraging Lotteries for School Value-Added: Testing and Estimation. *The Quarterly Journal of Economics 132*(2), 871–919.

Aradillas-López, A. (2020). The econometrics of static games. *Annual Review of Economics 12*(1), 135–165.

Armona, L. and S. Cao (2024). Redesigning federal student aid in sub-baccalaureate education. *Available at SSRN 4300755*.

Atal, J. P., J. I. Cuesta, and M. Sæthre (2025). Quality regulation and competition: Evidence from pharmaceutical markets. Technical report, National Bureau of Economic Research.

Aucejo, E. M., A. S. Perry, and B. Zafar (2024). Assessing the costs of balancing college and work activities: The gig economy meets online education. Technical report, National Bureau of Economic Research.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010). Estimating static models of strategic interactions. *Journal of Business & Economic Statistics 28*(4), 469–482.

Barahona, N., C. Dobbin, H. Ho, S. Otero, and C. Yannelis (2023). Skin in the game: colleges' financial incentives and student outcomes. *Working Paper*.

Barahona, N., C. Dobbin, and S. Otero (2023). Affirmative action in centralized college admissions systems. Technical report, Columbia University.

Barrow, L., W. T. Morris, and L. Sartain (2024). The expanding landscape of online education: Who engages and how they fare. *Journal of Labor Economics 42*(S1), S417–S443.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute.

Bau, N. (2022). Estimating an equilibrium model of horizontal competition in education. *Journal of Political Economy 130*(7), 1717–1764.

Berry, S. and P. Haile (2016). Identification in differentiated products markets. *Annual review of Economics 8*, 27–52.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Bertolin, J., T. McCowan, and H. Bittencourt (2023). Expansion of the distance modality in brazilian higher education: Implications for quality and equity. *High Educ Policy 36*, 231–249.

Bettinger, E. P., L. Fox, S. Loeb, and E. S. Taylor (2017). Virtual Classrooms: How Online College Courses Affect Student Success. *American Economic Review 107*(9), 2855–2875.

Bodéré, P. (2023). Dynamic spatial competition in early education: An equilibrium analysis of the preschool market in pennsylvania. *Working Paper*.

Borusyak, K. and P. Hull (2023). Nonrandom exposure to exogenous shocks. *Econometrica 91*(6), 2155–2185.

Borusyak, K., P. Hull, and X. Jaravel (2021). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies 89*(1), 181–213.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2024). Difference-in-differences with a continuous treatment. Technical report, National Bureau of Economic Research.

Conlon, C. and J. Gortmaker (2020). Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics 51*(4), 1108–1161.

Dahlstrand, A. (2024). Defying distance? the provision of services in the digital age. Technical report, University of Zurich.

Dahlstrand, A., N. Le Nestour, and G. Michaels (2024). Online versus in-person services: Effects on patients and providers. Technical report, Institute of Labor Economics (IZA), Bonn.

Deming, D. J., C. Goldin, and L. F. Katz (2012). The for-profit postsecondary school sector: Nimble critters or agile predators? *Journal of Economic Perspectives 26*(1), 139–64.

Deming, D. J., C. Goldin, L. F. Katz, and N. Yuchtman (2015). Can online learning bend the higher education cost curve? *American Economic Review 105*(5), 496–501.

Deming, D. J., M. Lovenheim, and R. W. Patterson (2016). The competitive effects of online education. Working Paper 22749, National Bureau of Economic Research.

Deming, D. J., N. Yuchtman, A. Abulafi, C. Goldin, and L. F. Katz (2016). The value of postsecondary credentials in the labor market: An experimental study. *American Economic Review 106*(3), 778–806.

Dinerstein, M., D. Morales, C. Neilson, and S. Otero (2023). The equilibrium effects of public provision in education markets. *Working paper*.

Dinerstein, M. and T. D. Smith (2021). Quantifying the supply response of private schools to public policies. *American Economic Review 111*(10), 3376–3417.

Dobbin, C., N. Barahona, and S. Otero (2021). The equilibrium effects of subsidized student loans. Technical report, Working paper, Stanford University.

El Galad, A., D. H. Betts, and N. Campbell (2024). Flexible learning dimensions in higher education: aligning students' and educators' perspectives for more inclusive practices. In *Frontiers in Education*, Volume 9, pp. 1347432. Frontiers Media SA.

Fabregas, R. and L. Navarro-Sola (2024). Broadcasting education at scale: Long-term labor market impacts of television-based schools. Working paper.

Figlio, D., M. Rush, and L. Yin (2013). Is it live or is it internet? experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics 31*(4), 763–784.

Gandhi, A. and J.-F. Houde (2019). Measuring substitution patterns in differentiated-products industries. *NBER working paper* (w26375).

Garrett, R., B. Simunich, R. Legon, and E. E. Fredericksen (2022). Tracking online learning from mainstream acceptance to universal adoption. Technical report.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

Goodman, J., J. Melkers, and A. Pallais (2019). Can Online Delivery Increase Access to Education? *Journal of Labor Economics 37*(1), 1–34. Publisher: The University of Chicago Press.

Hausman, J., G. Leonard, and J. D. Zona (1994). Competitive analysis with differenciated products. *Annales d'Economie et de Statistique*, 159–180.

Hoxby, C. M. (2018). Online postsecondary education and labor productivity. In C. R. Hulten and V. A. Ramey (Eds.), *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, Volume 2, pp. 401–460. Chicago: University of Chicago Press. Conference held in October 2015, published in December 2018.

IBGE (2010). Household internet access. *Population Census*.

Kofoed, M. S., L. Gebhart, D. Gilmore, and R. Moschitto (2024). Zooming to class? experimental evidence on college students' online learning during covid-19. *American Economic Review: Insights 6*(3), 324–40.

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

NCES (2022). Digest of education statistics 2022, table 311.15. Technical report, U.S. Department of Education.

Neilson, C. et al. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. *Job Market Paper 48*.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy 110*(4), 705–729.

PNAD (2019). Household internet access. *Pesquisa Nacional por Amostra de Domicílios*.

Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement*. *The Quarterly Journal of Economics 125*(1), 175–214.

Rouse, C. E. (1995). Democratization or diversion? the effect of community colleges on educational attainment. *Journal of Business & Economic Statistics 13*(2), 217–224.

Rouse, C. E. (1998). Do two-year colleges increase overall educational attainment? evidence from the states. *Journal of Policy Analysis and Management 17*(4), 595–620.

Sanchez, C. (2023). Equilibrium consequences of vouchers under simultaneous extensive and intensive margins competition. Technical report, Working Papers.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics 37*(3), 619–640.

Sweeting, A. (2009). The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *The RAND Journal of Economics 40*(4), 710–742.

Zeltzer, D., L. Einav, J. Rashba, and R. D. Balicer (2023). The Impact of Increased Access to Telemedicine. *Journal of the European Economic Association 22*(2), 712–750.

Online Appendix for:

The Effects of Widespread Online Education on Market Structure and Enrollment

(Not for publication)

Nano Barahona   Cauê Dobbin   Joaquín Fuenzalida   Sebastián Otero

March 9, 2025

Appendix A: Additional Figures and Tables

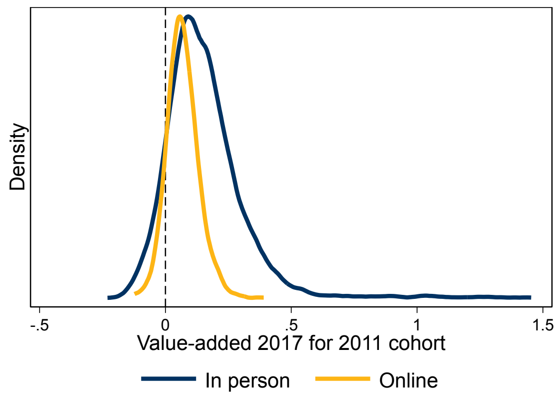


Figure A.1: Value-added distribution for in-person and online degree programs

Notes: This figure shows the distribution of value added for existing degree programs. The blue lines represents the distribution of in-person degree programs and the yellow line the distribution of online degree programs.

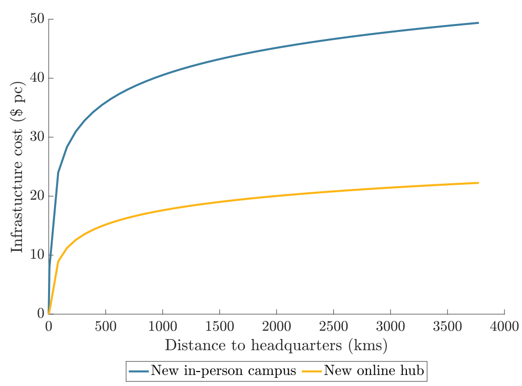
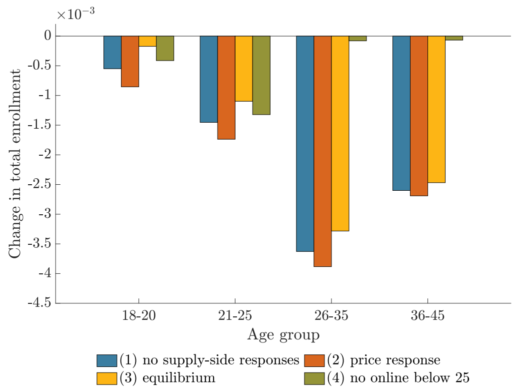
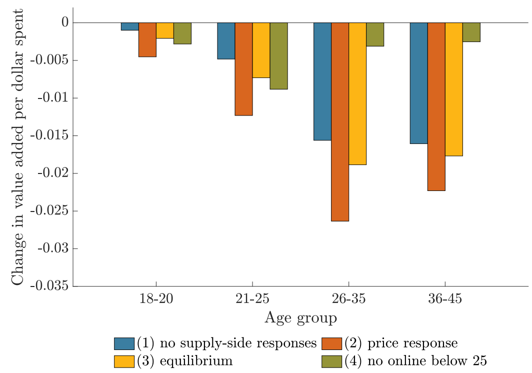


Figure A.2: Cost of building infrastructure as a function of distance

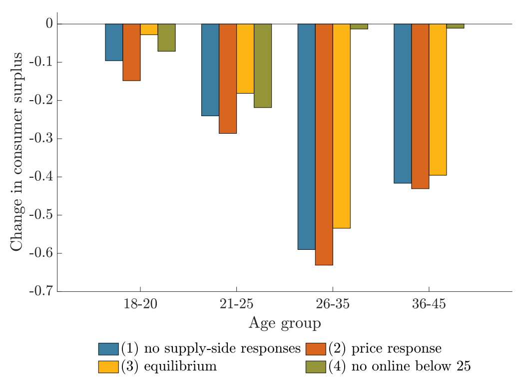
Notes: This figure shows the infrastructure cost of building a new campus or hub as a function of the distance between the firm's headquarter and any given region, given by $\chi_0^k + g(d_{fr})'\chi^k$ for $k \in \{\text{campus, hub}\}$.



(a) Total enrollment



(b) Value added per dollar



(c) Consumer surplus

Figure A.3: Changes in various outcome by age group under each counterfactual

Notes: This figure presents changes in various outcomes for different age groups under different counterfactuals. We show changes in value added relative to the *Baseline* counterfactual for all counterfactuals presented in Table 5 and the hypothetical counterfactual from Section 5.3. In Panel (a), we show the changes in total enrollment. In Panel (b), we show changes in value added for every $1000 dollars spent in higher education. In Panel (c), we show changes in total consumer surplus measured as the area below the demand curve.

Table A.1: Areas of Study and Degrees

Area of Study	**Type**	**Degree programs**
Arts, Humanities, and Social Sciences	Bachelor	Library, Political Science, Social Communication, Design, Journalism, International Relations
	Technical	Interior Design, Fashion Design, Product Design, Graphic Design, Photography, Digital Marketing, Audiovisual Production, Multimedia Production
Business	Bachelor	Administration, Accounting, Economics, Public Administration, Social Communication, Advertisement and Marketing, Public Relations, Executive Secretariat (Secretariado Executivo)
	Technical	Foreign Trade, Entrepreneurship, Commercial Management, Quality Management, Information Techonology Management, Human Resources Management, Security Management, Public Health Management, Legal Management, Information System Management, Financial Management, Hospital Management, Public Management, Logistics, Marketing, Real Estate, Management Processes, Secretariat (Secretariado)
Education	Bachelor	Visual Arts, Biological Sciences, Physical Education, Philosophy, Geography, History, Mathematics, Psychopedagogy, Chemistry
	?	Visual Arts, Biological Sciences, Social Sciences, Physical Education, Geography, History, (Letras), Mathematics, Pedagogy, Chemistry, Sociology
Engineering	Bachelor	Agronomy, Architecture and Urbanism, Civil Engineering, Computer Engineering, Software Engineering, Electric Engineering, Mechanical Engineering, Chemical Engineering
	Technical	Agribusiness, Agribusiness Management, Industrial Mechatronics (Mecatronica Industrial), Mechanical Manufacturing
Health Sciences	Bachelor	Nursing, Pharmacy, Physiotherapy, Veterinary Medicine, Speech Therapy, Nutrition, Occupational Therapy
	Technical	Radiology
Services	Bachelor	Aeronautical Sciences, Social Service, Tourism
	Technical	Beauty and Personal Image, Aesthetics and Cosmetics, Gastronomy, Environmental Management, Tourism Management, Public Security, Juridic Services
Others	Bachelor	Environmental Engineering, Production Engineering, Theology
	Technical	Industrial Automation, Social Educator, Management of Industrial Production
Law	Bachelor	Law
Medicine	Bachelor	Medicine, Odontology
Psychology	Bachelor	Psychology
Math, Computer and Natural Sciences	Bachelor	Biomedicine, Computer Science, Information Systems
	Technical	Analysis and System Development, *Banco de Dados*, *Jogos Digitais*, *Redes de Computadores*, Information Security, *Sistemas para Internet*

Notes: This table summarizes the aggregation of different degree programs into broader areas of study. We categorize programs based on the International Standard Classification of Education (ISCED) codes.

Table A.2: Relationship between distance and probability of opening an online degree for selected years

	entered_{fr} (1)
$\log(1 + d_{fr})$	-0.088
	(0.008)
H_{fr}	-0.146
	(0.076)
Obs.	4070
Regions	110
Firms	37
Mean Dep. Var	0.22

Notes: This table shows the estimated parameters from estimating Equation (5), given by $\text{Entered}_{fr} = g(d_{fr})'\gamma + \delta_f + \delta_r + \eta_{fr}$, where Entered_{fr} equals 1 if institution f had entered region r by 2010 and 0 otherwise; d_{fr} represents the distance between the headquarters of institution f and region r. We define the vector $g(d_{fr}) = [\log(1+d_{fr}), H_{fr}]'$, where $H_{fr} \in \{0, 1\}$ indicates whether the headquarters of institution f are located in region r (i.e., $d_{fr} = 0$), to account for cases where the distance is zero. We estimate the regression using all 37 firms that offer at least one online degree in 2010.

Table A.3: Demand parameters

Panel A: BLP parameters									
Price coef.		Price RC		Online RC		In-person RC		Nest param.	
$\bar{\alpha}$	-2.214	σ_{α}	-0.002	σ^{o}	-0.112	σ^{p}	0.087	ρ	0.852
	(1.111)		(40.041)		(60.217)		(73.554)		(0.248)
Dem. shifter		Dem. shifter		Age 21-25		Age 26-35		Age 36-45	
ψ^{0}	0.099	ψ^{o}	-0.04	μ_2^{p}	-0.906	μ_3^{p}	-1.465	μ_4^{p}	-2.095
	(0.148)		(0.785)		(15.233)		(14.835)		(25.36)
Age 21-25		Age 26-35		Age 36-45					
μ_2^{o}	-0.63	μ_3^{o}	-0.856	μ_4^{o}	-1.263				
	(5.577)		(14.237)		(1.725)				
Panel B: Mixed model parameters									
Degree r.e.		Region-area r.e.		Year-area r.e.		Year-online r.e.		Demand sock	
σ_{j}	0.237	σ_{ra}	0.663	σ_{ta}	0.855	σ_{to}	0.467	σ_{ξ}	0.316
	(0.004)		(0.014)		(0.063)		(0.08)		(0.001)
Hours f.e.		Stem f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
δ_1	0.119	δ_2	0.019	δ_3	0.044	δ_4	0.006	δ_5	0.049
	(0.007)		(0.023)		(0.007)		(0.0002)		(0.022)
Const. f.e.									
δ_0	-9.249								
	(0.208)								

Notes: This table reports the estimated parameters from the demand model presented in Section 4.2. Panel A presents the parameters estimated in the first step following Berry et al. (1995). Panel B presents the parameters estimated in the second step from the mixed-effects Bayesian hierarchical model that we use to recover each of the mean utility components of δ_{jrt}.

Table A.4: Diversion ratios by age group

	$t = 2010$		$t = 2019$	
	In-person	Online	In-person	Online
Median diversion ratios for cohort 18-20:				
To in-person:	0.84	0.80	0.81	0.54
To online:	0.00	0.06	0.04	0.33
To outside good:	0.14	0.14	0.14	0.13
Median diversion ratios for cohort 21-25:				
To in-person:	0.80	0.61	0.69	0.25
To online:	0.01	0.23	0.14	0.61
To outside good:	0.15	0.15	0.14	0.14
Median diversion ratios for cohort 26-35:				
To in-person:	0.73	0.30	0.41	0.06
To online:	0.06	0.50	0.40	0.78
To outside good:	0.15	0.15	0.14	0.15
Median diversion ratios for cohort 36-45:				
To in-person:	0.61	0.12	0.22	0.02
To online:	0.18	0.68	0.61	0.82
To outside good:	0.15	0.16	0.15	0.15

Notes: This table reports diversion ratios across products and markets for 2010 and 2019 for in-person and online degrees for each age group. We calculate diversion ratios as the share of students from each age group that decide to stop attending degree j upon an increase in tuition price that would switch to either an in-person degree, an online degree, or the outside option. Formally, we calculate diversion ratios as $D_{jb\mathcal{K}} = \left(\left|\frac{\partial s_{jb}}{\partial p_j}\right|\right)^{-1}\left(\sum_{k\in\mathcal{K}/j}\frac{\partial s_{kb}}{\partial p_j}\right)$, where $\mathcal{K}$ the set of all degrees that are either in-person, online, or the outside option, and s_{jb} is the share of students among those in group age b that decide to attend degree j. For example, the diversion ratio from in-person degrees to online degrees in 2010 for age group 36-45 is 0.2, which is given by the median of $D_{jb\mathcal{K}}$ across all degrees j that are in-person and for $\mathcal{K}$ the set of all online degrees, with $b = \{36\text{-}45\}$.

Table A.5: Marginal cost parameters

Panel A: Main parameters									
Demand instrument		log-distance		same-region					
γ_{z_1}	0.152	γ_{z_2}	0.151	γ_{z_3}	0.748				
	(0.399)		(1.04)		(1.961)				
Panel B: Mixed model parameters									
Degree r.e.		Region-area r.e.		Year-area r.e.		Year-online r.e.		Cost sock	
σ_j	0.539	σ_{ra}	0.267	σ_{ta}	0.101	σ_{to}	0.989	σ_ω	0.56
	(0.008)		(0.009)		(0.033)		(0.16)		(0.001)
Hours f.e.		Stem f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
δ_1	0.187	δ_2	-0.12	δ_3	-0.014	δ_4	0.007	δ_5	0.257
	(0.012)		(0.038)		(0.015)		(0.0004)		(0.045)
Const. f.e.									
δ_0	-7.756								
	(0.388)								

Notes: This table reports the estimated parameters from the marginal cost model presented in Section 4.3. Panel A presents the parameters estimated in the first step via OLS. Panel B presents the parameters estimated in the second step from the mixed-effects Bayesian hierarchical model that we use to recover each of the marginal cost components of γ_{jrt}.

Table A.6: Fixed cost parameters

Panel A: Infrastructure costs											
New campus						New hub					
χ_0^c	-0.795 (1.093)	χ_H^c	0.795 (1.093)	χ_d^c	0.953 (0.165)	χ_0^h	-0.94 (0.917)	χ_H^h	0.94 (0.916)	χ_d^h	0.5 (0.125)
Panel B: Degree costs											
κ_0	1.057 (0.15)	$\kappa_{\iota 1}^{old}$	0 (0)	κ_{o1}^{old}	0.17 (0.334)	$\kappa_{\iota 2}^{old}$	0 (0)	κ_{o2}^{old}	0 (0)	$\kappa_{\iota 3}^{old}$	0 (0)
κ_{o3}^{old}	0 (0)	$\kappa_{\iota 4}^{old}$	0 (0)	κ_{o4}^{old}	0 (0)	$\kappa_{\iota 5}^{old}$	0 (0)	κ_{o5}^{old}	0 (0)	$\kappa_{\iota 6}^{old}$	0 (0)
$\kappa_{\iota 7}^{old}$	0 (0)	κ_{o7}^{old}	0 (0)	$\kappa_{\iota 8}^{old}$	0 (0)	$\kappa_{\iota 9}^{old}$	0 (0)	κ_{o9}^{old}	0 (0)	$\kappa_{\iota 10}^{old}$	0 (0)
$\kappa_{\iota 11}^{old}$	0 (0)	κ_{o11}^{old}	0 (0)	$\kappa_{\iota 1}^{new}$	0.852 (0.201)	κ_{o1}^{new}	0 (0)	$\kappa_{\iota 2}^{new}$	1.039 (0.295)	κ_{o2}^{new}	0 (0)
$\kappa_{\iota 3}^{new}$	0.874 (0.213)	κ_{o3}^{new}	0 (0)	$\kappa_{\iota 4}^{new}$	0 (0)	κ_{o4}^{new}	0 (0)	$\kappa_{\iota 5}^{new}$	0 (0.043)	κ_{o5}^{new}	0 (0.06)
$\kappa_{\iota 6}^{new}$	0.345 (0.207)	$\kappa_{\iota 7}^{new}$	0.515 (0.137)	κ_{o7}^{new}	0 (0)	$\kappa_{\iota 8}^{new}$	0 (0)	$\kappa_{\iota 9}^{new}$	0.049 (0.093)	κ_{o9}^{new}	0 (0.014)
$\kappa_{\iota 10}^{new}$	0 (0)	$\kappa_{\iota 11}^{new}$	0 (0.064)	κ_{o11}^{new}	0 (0)	σ_ε	0.007 (0.001)				

Notes: This table reports the estimated fixed-cost parameters from the entry model presented in Section 4.4. Panel A presents the infrastructure costs parameters from Equation (16). Panel B presents the degrees fixed costs parameters from Equation (15). Standard errors are estimated via Bootstrap with 100 repetitions and clustering at the market level.

Appendix B: Data Construction

In this Appendix, we provide details on some key variables construction process.

B.1. *Tuition fees*

To construct program-level prices, we combine four different data sources. The first two sources come from Brazil's government fellowship and loan programs, PROUNI and FIES. We utilize administrative records from the National Education Fund (FNDE), which track the payments made by the government to students in these programs, allowing us to estimate the tuition fees at participating institutions. The third source is a nationally representative survey conducted by Hoper, a consultancy specializing in higher education. The fourth source is administrative data from QueroBolsa, Brazil's largest degree search platform.

Using these datasets, we construct prices as follows: First, we use the information from all sources to recover an average posted price by program, campus or hub, and year. This is done by regressing log-prices on program-region-year and source fixed effects. Controlling for source fixed effects helps account for persistent differences across data sources and allows us to recover a program-region-year price. In cases where information for a certain year is missing, we run a regression for the predicted price on program-region and year fixed effects, imputing the missing values based on the sum of the coefficients. Finally, in situations where information on both year and region is unavailable, we regress the predicted price on program and year fixed effects, filling the missing values using the sum of these coefficients.

B.2. *Value added*

We track all university entrance exam (ENEM) participants and assign them to degree programs based on their initial college enrollment, or to an outside option if they do not enroll. We then follow these individuals in the labor market seven years after the exam, recording their highest salary from labor market administrative records (RAIS).

Our value added estimation relies on a standard selections-on-observables model (Rothstein, 2010; Angrist et al., 2017), which compares the labor market earnings of students across various programs while controlling for test scores and an extensive range of student characteristics. All values are normalized relative to the outside option of not attending college. We allow the outside option to vary by region to account for local labor market conditions.

Specifically, we estimate the following model for each region r:

$$\log(Y_i) = \sum_j \mathrm{VA}_{jr} \cdot D_{ij} + X_i' \beta_r + \varepsilon_i \tag{B.1}$$

where Y_i represents the wage income of student i, $D_{ij} \in \{0, 1\}$ indicates whether student i is enrolled in degree program j, X_i includes students' characteristics such as gender, age, ENEM score, and a constant, and ε_i captures other determinants of income that are uncorrelated with school enrollment. The term VA_{jr} denotes the region-specific value added of each program. We normalize the value added of the outside option to zero in each region and estimate the model via OLS.

We follow Angrist et al. (2023) and use empirical Bayes shrinkage methods to improve the precision of estimates. The estimates $\widehat{\text{VA}}_{jr}$ from estimating Equation (B.1) via OLS are unbiased but noisy measures of the underlying program-specific value added. We investigate the distribution of VA_{jr} using the following hierarchical model:

$$\begin{aligned} \widehat{\text{VA}}_{jr}|\text{VA}_{jr} &\sim \mathcal{N}(\text{VA}_{jr}, s_{jr}^2) \\ \text{VA}_{jr} &\sim \mathcal{N}(\mu_k, \sigma_k^2) \end{aligned}$$

where ω_{jr}^2 is the sampling variance of the estimator $\widehat{\text{VA}}_{jr}$, while μ_k and σ_k^2 are the hyperparameters that govern the distribution of the latent parameters across programs that we allow to be different for in-person and online degrees (i.e., $k \in \{\text{in-person}, \text{online}\}$). Method of moments estimates of these hyperparameters are given by:

$$\begin{aligned} \hat{\mu}_k &= \frac{1}{J_k}\sum_{j\in k}\widehat{\text{VA}}_{jr} \\ \hat{\sigma}_k^2 &= \frac{1}{J_k}\sum_{j\in k}\left[\left(\hat{\text{VA}}_{jr} - \hat{\mu}_{VA}\right)^2 - s_{jr}^2\right] \end{aligned}$$

where J_k is the number of programs offered in format k.

The final step in the empirical Bayes estimation is to construct posteriors for the value added of each program. Given the model above, the posterior means are given by

$$\text{VA}_{jr}^* = \left(\frac{\sigma_{k(j)}^2}{\sigma_{k(j)}^2 + s_j^2}\right)\widehat{\text{VA}}_{jr} + \left(\frac{s_j^2}{\sigma_{k(j)}^2 + s_j^2}\right)\hat{\mu}_{k(j)}. \tag{B.2}$$

By shrinking the noisy estimate of VA_{jr} toward the prior mean, the posterior mean reduces variance, with more shrinkage for programs with noisier estimates. We use the posterior means in the analysis from Section 2.5.2.

To estimate the model, we use the universe of ENEM 2010 applicants from the 2011 admissions cycle and use student labor market earnings in 2017, seven years after their enrollment decisions. Because we use only one student cohort, our program value added measure remains constant over time. We are able to estimate value added for 22% of the program-region pairs that exist in the data, which covers 73% of enrollment across all years. Among program-region pairs that existed in 2011, we cover 86% of program-region pairs and 98% of total enrollment.

B.2.1. *Degree-level aggregation and inputation for counterfactual analysis*: The counterfactual analysis from Section 5 requires value-added estimates for all degrees that can potentially enter in a market. To get reliable estimates for degree that we do not observe in the data, we estimate a simple model using the program-level value-added measures presented above. In our model, each degree-region pair is associated with a specific value added given by

$$\mathrm{VA}_{jr} = x_j^{(1)}\nu_o + x_j^{(2)}\nu_x + \nu_j + \nu_{ra} + \varsigma_{jr} \tag{B.3}$$

where VA_{jr} is the (weighted) average of the posterior mean of the estimated value-added of all degree programs that compose degree j, $x_j^{(1)} = [\iota_j, o_j]$ and $x_j^{(2)}$ are the degree-level characteristics from Equations (8) and (9) from the main text, ν_j are degree specific components of value added, ν_{ra} are field of study-region specific components, and ς_{jr} is a degree-region specific idiosyncratic shock. As before, the outside option is normalized to be zero in each region.

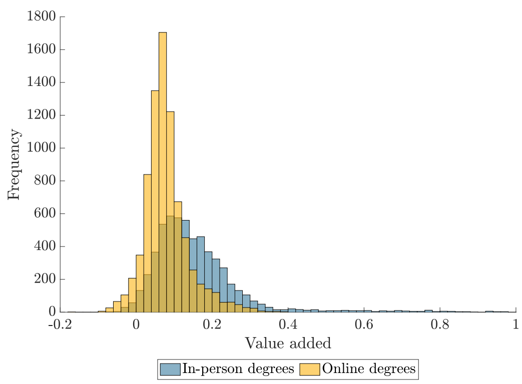
Following Section 4, we estimate the value added model using mixed-effects Bayesian hierarchical model with $\nu_j \sim \mathcal{N}(0, \vartheta_j^2)$, $\nu_{ra} \sim \mathcal{N}(0, \vartheta_{ra}^2)$, and $\varsigma_{jr} \sim \mathcal{N}(0, \vartheta_\varsigma^2)$, and (ν_o, ν_x) non-random parameters. We estimate the model via maximum likelihood to recover posterior means of each component of the value added model that we use to impute value added for all degree-markets, regardless of whether they are offered in the data or not.

We present the estimated parameters in Table B.1. In Figure B.1, Panel (a), we show the distribution of value added for online and in-person degrees. We find that in-person and online degrees have an average value added of 0.157 and 0.079, respectively. Finally, in Figure B.1, Panel (b), we show that students' preferences over degrees have a positive but weak correlation with value added. The correlation is particularly low for online degrees, which tend to be newer and students might be less informed about their quality.

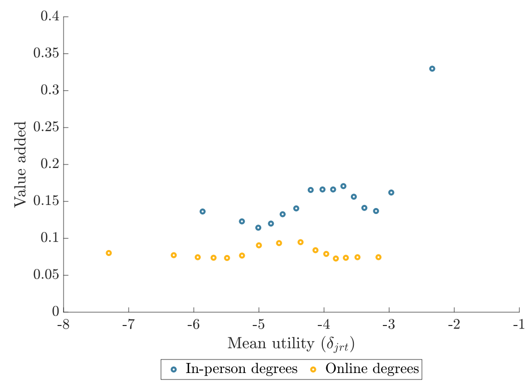
Table B.1: Value added parameters

Panel A: Mixed model parameters									
Degree r.e.		Region-area r.e.		VA shock		Const. f.e.		Online f.e.	
σ_j	0.057	σ_{ra}	0.068	σ_{ta}	0.068	δ_0	-1.237	δ_o	-0.041
	(0.002)		(0.003)		(0.001)		(0.056)		(0.007)
Hours f.e.		Stem f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
δ_1	0.066	δ_2	-0.026	δ_3	-0.015	δ_4	0.001	δ_5	0.103
	(0.003)		(0.007)		(0.004)		(0.0001)		(0.008)

Notes: This table reports the estimated fixed-cost parameters from the model presented in Section B.2.1.



(a) Distribution of value added



(b) Relationship between preferences and value added

Figure B.1: Estimated parameters from value added model

Notes: The figure summarizes the estimation results of the value added model. Panel (a) shows the distribution of value added across all in-person and online degrees that are offered in the data. Panel (b) shows the relationship between degrees' mean utility components of δ_{jrt} and the estimated value added. In both panels, in-person degree are depicted in blue and online degrees in yellow

Appendix C: Empirical Model Details

C.1. *Internet penetration and degree entry*

In Section 4, we use internet penetration as an excluded variable to estimate the entry fixed cost. This appendix presents the reduced-form relationship between internet penetration and changes in enrollment and degree availability for in-person and online education. Specifically, we estimate the following equation:

$$\Delta y_{ra} = \phi \Delta I_r + \varepsilon_{ra} \tag{C.1}$$

where ΔI_{ra} denotes the change in the number of internet accesses per person in region r between 2010 and 2019, and Δy_{ra} represents the corresponding change in the following outcomes: (i) the total number of online degrees, (ii) the total number of in-person degrees, (iii) the share of online students relative to market size, and (iv) the share of in-person students relative to market size. We estimate the model using OLS, with results summarized in Table C.1. Our findings indicate that increased internet penetration predicts a rise in the number of online degrees and a small, statistically insignificant decline in the number of in-person degrees. Additionally, internet penetration is associated with increased online enrollment and decreased in-person enrollment.

Table C.1: Effects of internet expansion on degree availability and enrollment

	Δ in online degrees (1)	Δ in in-person degrees (2)	Δ in online students (3)	Δ in in-person students (4)
Panel A: OLS regression				
Δ internet penetration	3.383 (0.800)	-0.221 (0.323)	1.705 (0.607)	-1.094 (0.246)
Panel B: Average value of the dependant variable in levels in 2010 and 2019				
2010	2.82	10.98	0.72	3.10
2019	9.48	12.19	2.79	2.94
Obs.	1,210	1,210	1,210	1,210

Notes:

C.2. *Entry probabilities prediction model*

We estimate the entry fixed costs from Section 4.4 using the two-step estimator developed in Sweeting (2009). In the first step, we estimate a multinomial choice logit model using all own and competitors' drivers of variable profits, and the determinants of fixed costs from Equation (14), to predict the probability that each bundle is offered in the data. To do that, we start by computing proxies of a bundle's value by estimating three different measures. The first measure proxies for expected the utility that a consumer perceives when facing the opportunity of choosing a given degree j that belongs to bundle $\mathcal{J}_{frt}$, and is given by

$$\text{log-sum}(\vec{\delta}_{\mathcal{J}_{frt}}) = \log\left(1 + \sum_{j\in\mathcal{J}_{frt}} \exp(x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{to_j})\right)$$

where the terms inside the exponential are the posterior means of the mean utility components of δ_{jrt} from Equation (9).

The second measure proxies for expected the utility that a consumer perceives when facing the opportunity of choosing a given degree j that belongs to bundle $\mathcal{J}_{frt}$ but taking into account the marginal cost of each associated degree, and is given by

$$\text{log-sum}(\vec{\delta}_{\mathcal{J}_{frt}}, \vec{c}_{\mathcal{J}_{frt}}) = \log\left(1 + \sum_{j\in\mathcal{J}_{frt}} \exp(x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{to_j} - \bar{\alpha}c_{jrt}|_{\omega_{jrt}=0})\right)$$

where $c_{jrt}|_{\omega_{jrt}=0}$ corresponds to the marginal cost from Equation (12) evaluated at $\omega_{jrt} = 0$, and $\bar{\alpha}$ is the average price sensitivity across al students.

The third measure proxies for the total market share that firm f would capture offering bundle $\mathcal{J}_{frt}$ and is given by

$$\hat{s}(\vec{\delta}_{\mathcal{J}_{frt}}, \vec{c}_{\mathcal{J}_{frt}}) = \frac{\sum_{j\in\mathcal{J}_{frt}} \exp(x_j^{(2)}\delta + \delta_j + \delta_{ra} + \delta_{ta} + \delta_{to_j} - \bar{\alpha}c_{jrt}|_{\omega_{jrt}=0})}{1 + \sum_{\mathcal{J}}\sum_{k\in\mathcal{J}} \exp(x_k^{(2)}\delta + \delta_k + \delta_{ra} + \delta_{ta} + \delta_{to_k} - \bar{\alpha}c_{krt}|_{\omega_{krt}=0})}$$

where the numerator sums over all degrees that belong to bundle $\mathcal{J}_{frt}$, and the denominator adds over all degrees available in market rt.

We use the linear combination of $\Theta_{\mathcal{J}_{frt}} = \left[\text{log-sum}(\vec{\delta}_{\mathcal{J}_{frt}}), \text{log-sum}(\vec{\delta}_{\mathcal{J}_{frt}}, \vec{c}_{\mathcal{J}_{frt}}), \hat{s}(\vec{\delta}_{\mathcal{J}_{frt}}, \vec{c}_{\mathcal{J}_{frt}})\right]$ to approximate the variable profits that firm f would receive after entering with bundle $\mathcal{J}_{frt}$, and a linear combination of the fixed cost determinants from Equation (14) to approximate for the fixed costs associated with offering bundle $\mathcal{J}_{frt}$. Specifically, we approximate the probability that bundle $\mathcal{J}_{frt}$ is offered by

$$\phi_{\mathcal{J}_{frt}} = \frac{\exp(\Theta'_{\mathcal{J}_{frt}}\beta + \text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) + \text{Infrastucture}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}))}{\sum_{\mathcal{J}\in\mathcal{B}_f} \exp(\Theta'_{\mathcal{J}}\beta + \text{Degrees}(\mathcal{J}|\mathcal{J}_{frt_0}) + \text{Infrastucture}(\mathcal{J}|\mathcal{J}_{frt_0}))}. \tag{C.2}$$

where

$$\Theta'_{\mathcal{J}}\beta = \beta_1\text{log-sum}(\vec{\delta}_{\mathcal{J}}) + \beta_2\text{log-sum}(\vec{\delta}_{\mathcal{J}}, \vec{c}_{\mathcal{J}}) + \beta_3\hat{s}(\vec{\delta}_{\mathcal{J}}, \vec{c}_{\mathcal{J}}).$$

We estimate Equation (C.2) via maximum likelihood and use the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, to estimate the expected variable profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$, as described in the main text.

C.3. *Counterfactual Simulation Details*

To implement our counterfactual analysis in Section 5, we simulate equilibrium outcomes using our estimated model. We begin with a dataset that includes all firms and the respective bundles $\mathcal{J}_{frt} \in \mathcal{B}_f$ they can offer in each market, together with the estimated profits of offering each bundle, $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$, as estimated from the data, and the estimated entry probabilities, $\phi_{\mathcal{J}_{frt}}$. For each market, we proceed as follows:

1. Take the vector of expected profits from iteration $t-1$ and call it $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1}$. For the first iteration, use the expected profits estimated from the data.
   (a) Take draws of the private-information fixed cost shocks, $\varepsilon_{fr\mathcal{J}_{frt}}$, for all potential bundles, $\mathcal{J}_{frt}$.
   (b) Solve the firms' problem from Equation (17) using the fixed cost draws and $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1}$ to determine which degrees are offered in the market.
   (c) Take draws from demand and marginal cost shocks, ξ_{jrt} and ω_{jrt}, solve for optimal pricing and compute variable profits using Equation (11) for all firms.
   (d) Repeat steps (a)-(c) 10,000 times and use the simulated variable profits to estimate $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$. Call the estimate $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J}_{frt})]$.
2. Update the vector of expected profits to $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_t = \sqrt{\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1}\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J}_{frt})]}$ and calculate the mean square difference between iteration t and $t-1$ as $MSD_t = \sum\left(\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_t - \mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{t-1}\right)^2$, where the sum is over all markets, firms, and bundles.
3. Iterate until $MSD_t < \overline{MSD}$ for $\overline{MSD}$ small.

C.4. *Outcomes of Interest*

In this appendix, we present detailed formulas to calculate the main outcomes of interest presented in Section 5 of the main text. For each counterfactual k, we calculate

$$
\begin{aligned}
\text{Total value-added}^k &= \frac{\sum_r M_r \sum_j s_{jr}^k \mathrm{VA}_{jr}}{\sum_r M_r \sum_j s_{jr}^{BL}} \\
\text{Total expenditure}^k &= \frac{\sum_r M_r \sum_j s_{jr}^k p_{jr}^k}{\sum_r M_r \sum_j s_{jr}^{BL}} \\
\text{Av. price of in-person degrees}^k &= \frac{\sum_r M_r \frac{\sum_{j:\iota_j=1} p_{jr}^k}{\sum_{j:\iota_j=1} 1}}{\sum_r M_r} \\
\text{Total profits per capita}^k &= \frac{\sum_r M_r \sum_j s_{jr}^k (p_{jr}^k - c_{jr})}{\sum_r M_r \sum_j s_{jr}^{BL}} \\
\text{Total consumer welfare}^k &= \frac{\sum_r M_r CS_r^k}{\sum_r M_r},
\end{aligned}
$$

where VA_{jr} is the value added of degree j in region r, p^k_{jr} is the price of degree j in region r under counterfactual k, c_{jr} is the marginal cost of degree j in region r, M_r is the market size for region r, s^k_{jr} is the market share of degree j in region r under counterfactual k, and s^{BL}_{jr} is the market share of degree j in region r under the baseline counterfactual. Note that the denominator is constant across counterfactuals and we use it to normalize the values at a per student base.