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

Nano Barahona Cauê Dobbin Sebastián Otero

November 21, 2025

Abstract: We study the rapid expansion of Brazil's private online higher-education sector and its effects on market structure and college enrollment. Exploiting regional and field-specific variation in online education penetration, we find that online programs expand access for older students but divert younger students from higher-quality in-person programs. Greater competition lowers tuition prices but also reduces the supply of in-person degrees. Using an equilibrium model of college education, we show that in the absence of online programs, total enrollment would be 14 percent lower, while in-person enrollment would rise by 33 percent. On net, aggregate labor-market value added declines by 1.4 percent. Online education raises value added for older students, who benefit from increased access, but lowers it for younger students, who shift toward lower-return online options. Counterfactual policies that restrict online enrollment to older cohorts could increase value added for younger students without reducing gains for older cohorts.

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, Karthik Muralidharan, 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, University of Michigan, FGV Sao Paulo, INSPER, PUC Rio, Arizona State University, UC Berkeley, UCLA, Georgetown, CEMFI, IDB, World Bank, NBER Education, RIDGE Forum, SITE, O-Lab Conference, BSE Summer Forum, Northeastern Economics of Education Workshop, and the Bay Area IO Fest for useful comments. We thank Joaquín Fuenzalida for contributions during the early stages of this project. 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. Otero: Columbia University and NBER, so2699@columbia.edu.

1. Introduction

Over the past two decades, rapid advances in digital technology have transformed numerous industries, expanding the provision of goods and services through online channels. Sectors traditionally reliant on physical, face-to-face interactions—such as healthcare and education—have increasingly adopted hybrid or fully virtual models of delivery. This shift has been particularly pronounced in higher education, where institutions have expanded online degree offerings to meet growing demand for more affordable and flexible learning opportunities (Deming et al., 2012; Aucejo et al., 2024). In the Unites States, for example, roughly 15% of undergraduate students were enrolled exclusively in distance education in 2019, rising to 24% by 2022, a trend accelerated by the COVID-19 pandemic (NCES, 2022).

Expanding education 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 instruction can alter the nature of the learning experience, compromising its quality (Bettinger et al., 2017; Garrett et al., 2022). Consequently, while online education can boost total enrollment, it may also divert students from high-quality, in-person instruction toward potentially lower quality online alternatives. This is particularly problematic in higher education, where quality is difficult to assess (Hastings et al., 2015; Larroucau et al., 2024).

These concerns are amplified when considering equilibrium effects. 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). Yet, 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 acute in markets with limited in-person institutions, where students may be left with lower-quality online alternatives as their only option.

This paper studies the expansion of online higher education in Brazil, the world's largest market for online degrees. 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 affects market expansion, student diversion, tuition prices, and availability of in-person programs. We then embed these findings in an equilibrium model of college education to quantify the aggregate effects of online education and to 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 context in which online education constitutes a significant share of the education market. In Brazil, fully remote programs have grown rapidly, accounting for 17% of all new undergraduate enrollments in 2010, 44% by 2019, and reaching 65% after the COVID-19 pandemic. Second, isolating the causal impact of online education on local market outcomes from nationwide time trends requires variation in online penetration across local markets. Brazilian regulation provides such variation by requiring online programs to establish local physical hubs, which students must attend periodically to take exams and for administrative purposes, generating geographic differences in online market entry. In addition, 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 compile 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, enabling us to compute the labor-market 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 program quality.

Our analysis focuses on the private sector, which accounts for 82% of incoming undergraduate students between 2010 and 2019, and virtually all online programs. During this period, private sector enrollment grew from 1.7 million to 3 million students, with 89% of this growth concentrated in 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 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 compare 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. Tuition fees for online programs are 44% lower than for in-person programs. In terms of quality, online programs exhibit lower dropout rates but are associated with a 42% lower value added, measured by students' gains in labor market earnings after attending a given degree program.

To study the consequences of the online education expansion in Brazil, we estimate a linear model that regresses changes in outcomes—such as student enrollment, market structure, and tuition—on changes in the number of online degrees between 2010 and 2019. The unit of analysis is the interaction between a commuting zone and a field of study. We begin by estimating the model using OLS. This approach relies on a parallel trends assumption, which requires that outcomes in regions and fields with higher online degree entry would have followed the same trend as those with lower entry in the absence of differential online expansion. 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*). The instrument exploits the tendency of institutions to expand their online programs in regions closer to their headquarters and assumes that a region's proximity to an institution is uncorrelated with unobserved region-specific trends in fields where the institution offers in-person degrees (i.e., exogenous *shares*).

Both the OLS and SSIV estimates yield qualitatively consistent results: expanding online education increases online enrollment, reduces in-person enrollment, and raises overall college enrollment. This pattern reveals two offsetting effects of online education. On one hand, it expands access for students who might not otherwise have attended college. On the other hand, it diverts students from in-person programs, often toward lower-quality online alternatives. Our findings suggest that for each additional online student, 51% are new to higher education, while 49% would have otherwise enrolled in an in-person degree. We show the online education expansion primarily benefits older students by increasing their college enrollment and value added. By contrast, younger students—who are more likely to enroll in college in the first place—experience a stronger diversion to online options and a reduction in value added. Additionally, this expansion heightens competition, forcing local in-person institutions to lower prices. As competition intensifies, in-person programs are less likely to persist in the market, accelerating the diversion toward online degrees.

Estimating the linear model provides a transparent method for recovering marginal causal effects, but it relies 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. 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 substantial non-linear effects. To address these limitations, we develop a demand and supply model of college education that incorporates rich substitution patterns and accounts for equilibrium responses, enabling us to assess the effects of online education expansion under alternative 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 choose whether to enter a given market, which degrees to offer, and what prices to charge. To operate in a market, an institution must establish either a campus for in-person degrees or a physical hub for online degrees. Institutions' decisions unfold in two stages. In the first stage, they 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. They form expectations about their competitors' entry decisions and adjust their strategies accordingly—potentially choosing not to open 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 sources of variations in the data. To capture substitution patterns across fields of study and delivery modes, we exploit market-level differences in degree availability induced by the shift-share instrument. The estimates suggest that in-person and online degrees are close substitutes: when an in-person degree closes, 56% of students switch to another in-person option, 18% move online, and 24% opt out of college. To estimate price elasticities, we instrument for prices using contemporaneous prices of the same degree offered by the same institution in other regions, which serve as proxies for cost-shifters (Hausman et al., 1994). These leave-one-out mean prices serve as instruments because institution-specific cost shocks affect prices across all markets where a degree is offered. 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 -3.2 for in-person degrees and -1.3 for online degrees, consistent with estimates in the literature (Barahona et al., 2025; Armona and Cao, 2024).

For the supply model, we estimate 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' distances to various regions, which creates variation in competitive pressure that affects profits without changing fixed costs. We estimate a median entry elasticity with respect to own profits of 0.5 and find that opening an in-person campus is 2.4 times more costly than establishing an online hub.

We use our estimated model to quantify the impact of online education expansion on student enrollment, market structure, tuition fees, value added, and consumer expenditure and surplus. To assess how supply-side equilibrium effects shape these outcomes, we simulate three progressively more flexible counterfactuals in which we remove online education. Each scenario is benchmarked against a baseline counterfactual that reflects the status quo, in which online education remains available.

In the first counterfactual, we remove online education without allowing supply-side responses (i.e., holding degree offerings and prices fixed). Under this scenario, 57% of students enrolled in online programs switch to in-person programs, while the remaining 43% exit the market, resulting in a 15.1% decline in total enrollment.[1] Because online degrees provide greater value added than the outside option of no college, 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. At the same time, the substantial tuition gap between online and in-person programs causes total expenditures on higher education to rise as in-person enrollment increases.

Under the second counterfactual, we allow institutions to respond by adjusting degree prices. As a result, tuition for in-person degrees rises by 6.7%, leading to a further decline in enrollment and total value added, as well as an increase in total expenditure.

Finally, under the third counterfactual, we allow institutions to adjust both tuition and degree offerings. Relative to the previous scenario, the supply of in-person degrees rises by 9.3%, attracting new students and increasing total enrollment by 3.1%. The total value added

[1] These results suggest that the linear model underestimates diversion from in-person degrees, partly because it restricts substitution to occur only within fields of study. The equilibrium model allows for broader substitution patterns across fields and delivery modes.

under this counterfactual is 1.2% higher than under the first counterfactual, highlighting 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 1.4% lower. A revealed-preference analysis, assuming full rationality, shows that consumer surplus declines by 14%, suggesting that students strongly value more affordable and flexible options, even if they provide lower value added.

We next examine the distributional consequences of online education expansion by evaluating its impact on value added across different student age cohorts. Under the third counterfactual, which fully incorporates supply-side responses, younger students (18–25 years old) benefit from increased access to in-person programs that were previously unavailable due to the presence of online options. In contrast, older students (26–45 years old), 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 increases access for older students, it reduces in-person options, pushing younger students into lower-quality alternatives. Building on this insight, we explore potential government policies aimed at targeting online education to those who benefit most. Specifically, we consider a policy in which online education is available only to students above the age of 25. Our results show that this approach would increase value added for younger cohorts without reducing outcomes for older cohorts, increasing total value added by 1% relative to the baseline scenario, in which online education is available to all age cohorts.

Our findings shed light on how disruptive technologies, such as online services, can reshape competition and market dynamics. While they expand choice and reduce costs, the benefits may be unevenly distributed. In markets with imperfect information, some consumers may unintentionally switch from higher-quality options, leaving them worse off—particularly 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 technologies, balancing innovation with quality preservation.

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

Our research is also related to the literature examining 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), as well as on labor market outcomes (Deming et al., 2016; Hoxby, 2018; Fabregas and Navarro-Sola, 2024). Our findings are consistent with this work, highlighting the role of online education as a preferable alternative to no education, though less favorable than in-person options. Furthermore, our study contributes to the growing 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. Evidence also suggests that online healthcare services can enhance efficiency by providing faster, shorter consultations and improving the matching of doctors and patients (Dahlstrand et al., 2024; Dahlstrand, 2024).

We also contribute to the broader literature analyzing 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 (Barahona et al., 2025, 2023; Armona and Cao, 2024), on instructional levels (Bau, 2022), and on institutions' participation in voucher programs (Sanchez, 2023). More closely related to our work, three papers assess the impact of competition on market structure. Bodéré (2023) studies policies expanding high-quality enrollment in preschools in Pennsylvania, Dinerstein and Smith (2021) assess the impact of increased public school funding in New York, and Dinerstein et al. (2023) investigate the expansion of public schools in the Dominican Republic. We study how introducing a new delivery format affects market structure by modeling institutions' entry decisions and degree-portfolio choices.

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

2. Setting and data

In this section, we provide an overview of Brazil's higher education and online education regulatory landscape. We then present the data sources and summarize key descriptive statistics on the expansion of online education.

2.1. *Online higher education landscape in Brazil*

Brazil's higher education sector has expanded substantially over the past decade, with new undergraduate enrollment increasing from roughly 2.2 million students in 2010 to 3.6 million in 2022. A key driver of this growth has been the rising popularity of online degree programs, predominantly offered by for-profit private institutions.[2] The shift toward online education has

[2] Private institutions account for 95% of the online education market, with 79% of this share held by for-profit institutions. Furthermore, the market is highly concentrated, with seven institutions collectively accounting for 73% of total online enrollments.

been striking: in 2010, only 17% of new students chose online programs; by 2019, this share had risen to 44%, and it reached 65% following the COVID-19 pandemic.[3]

Online degree programs in Brazil, known as "Educação a Distância," offer remote versions of traditional undergraduate programs and must adhere to the same curriculum and duration standards.[4] Diplomas make no distinction between whether a degree was earned online or in person, effectively placing both modes on equal footing in terms of formal recognition. However, despite their lower cost, online programs are often perceived as lower in quality, a concern that continues to challenge Brazilian policymakers (Bertolin et al., 2023).

Online programs must include in-person sessions for essential activities such as 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 components of online education, placing geographic limits on the reach of these programs. Although in-person sessions are required for certain activities, all instruction is 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 older students—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). Finally, 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.[5]

Some fields of study face restrictions on being offered online. Specifically, Law, Medicine, and Psychology require special authorization from regulatory bodies, including the National Bar Association and the National Health Council. To date, no online programs in these fields have been approved. By contrast, disciplines such as Business and Education have experienced

[3] Brazil's share of fully remote undergraduate enrollments is substantially higher than in other countries. Although reliable cross-country data on fully online degrees are scarce, official data from 2019 (pre-COVID-19) indicate 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.

[4] Students may transfer credits between online and in-person formats within the same institution if they change modalities.

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

substantial growth in online education. In 2019, these two fields accounted for 78% of total online enrollment, in contrast to their 31% share of in-person enrollment.

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, the year the degree was introduced, and 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*: Because universities are not required to report tuition fees to the supervising authority, we rely on four distinct data sources to obtain this information. The first two sources come from Brazil's government fellowship and loan programs, PROUNI and FIES, via administrative records from the National Education Fund (FNDE), which track payments made to students, enabling us to estimate 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, we outline the methodology used to combine these sources into a unified tuition price for each degree program. Using this approach, we are able to recover year-specific tuition prices for 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 considerably in age: in 2010, approximately 62% had finished high school more than a year prior, and 25% were 25 or older.

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

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, defining the potential market for higher education students.[6] We also use administrative data from DSCOM ("Dados do Setor de Comunicações") from the Ministry of Science and Technology to calculate regional internet penetration rates over time, and data from IBGE ("Instituto Brasileiro de Geografia e Estatística") on municipality-level GDP per capita.

2.3. *Data definitions*

We next define the units of analysis and samples used throughout the article.

2.3.1. *Regions and markets*: A central aspect of our analysis is the definition of regions that 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 shared characteristics. 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. 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, using "firms" and "institutions" interchangeably. In our data, we distinguish two types of firms: those expanding their services into new regions (*expanding institutions*) and those operating in the same locations each year (*local institutions*). Most local institutions operate at a single location, with a few present at two to five different locations.[7]

2.3.3. *Degrees*: Throughout this article, we use two definitions of "degrees." The first, *degree programs*, refers to 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 (morning, afternoon, night, or full-time). For online degree programs, those sharing the same administrative code may be linked to different physical hubs, making them available in multiple regions. The dataset includes 35,420 unique degree programs.

To reduce dimensionality, we define a second type of degrees by aggregating similar degree programs within the same institution. A *degree* is defined as the combination of all degree programs offered by the same institution, within the same field of study, and delivered in the

[6] We use the cohort-level population size to extrapolate to the other years.

[7] During our study period, there was a wave of acquisitions of higher education institutions by educational groups in Brazil (Garcia and de Azevedo, 2019; Teodorovicz, 2025). We only have information about ownership structure in 2019, which we use to assign universities to firms across the whole sample.

same format.[8] For example, in our dataset, a degree might include all degree programs in the field of Business (including Administration, Accounting, Marketing, and Economics), offered by Estácio, a for-profit educational company in Brazil, and taught in person. A degree can be offered in multiple regions and years.

2.3.4. *Sample*: Our analysis focuses 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 than the public sector, which is highly selective and considered higher quality (Barahona et al., 2023). Consequently, the vast majority of individuals in our market definition do not have the option to attend public universities, which limits substitution between the two sectors. Public institutions, which rely entirely on government funding (federal or state) are oversubscribed and therefore less influenced by market forces.

We limit our analysis to the years 2010, when microdata first became available, through 2019, the final year before the COVID-19 pandemic. To avoid including very small markets, 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 each year, and degree-region pairs with fewer than five students in each year. After these exclusions, our final sample consists of 110 regions over a 10-year period, covering 474 institutions, 3,824 unique degrees, 14,319 degree-region pairs, and 88,108 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 the construction of a measure of quality for degree programs. We approximate program quality using value-added estimates derived from data on student enrollment, test scores, a rich set of covariates, and administrative wage records. To compute value added, we track all students who took the university entrance exam (ENEM) between 2010 and 2013, assigning them to degree programs based on their initial college enrollment—or to an outside option if they did not enroll. We then follow these individuals in the labor market, measuring their salary reported in the administrative wage records (RAIS) in 2023—more than 10 years after taking ENEM. This approach allows us to quantify the contribution of each program to subsequent labor market outcomes.

Our value-added model allows for selection on both observable and unobservable characteristics, as well as for heterogeneous returns across age cohorts. The methodology for calculating value added, described in Appendix C, incorporates four key features. First, we control for a rich set of observable characteristics, including ENEM scores, household income at the time of

[8] We classify programs into 10 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.

taking ENEM, and age, among other demographics.[9] Second, we control for unobserved student preferences for different programs and the outside option by estimating a two-step control function model that uses distance between the individual's residence and the program's location as an excluded instrument, as is common in the literature (Card, 1993; Carneiro et al., 2011; Eisenhauer et al., 2015; Mountjoy, 2022). Third, we normalize all measures relative to the outside option of not attending college. Finally, we use empirical Bayes methods to improve the precision of the estimates for smaller programs (Abdulkadiroğlu et al., 2020; Angrist et al., 2023).

We present summaries of our value-added estimates in Online Appendix C, Figure C.1. On average, enrolling in an in-person or online degree program increases future wages by 18% and 7%, respectively, relative to not attending college. We find substantial differences across age cohorts. For the youngest cohort (18–20), the returns to enrolling in an in-person degree program are roughly 3 times larger than those to enrolling in an online program. These differences narrow for older cohorts and disappear entirely for the oldest group (35–45). Finally, we document considerable heterogeneity across fields of study, with Medicine programs exhibiting the highest returns.

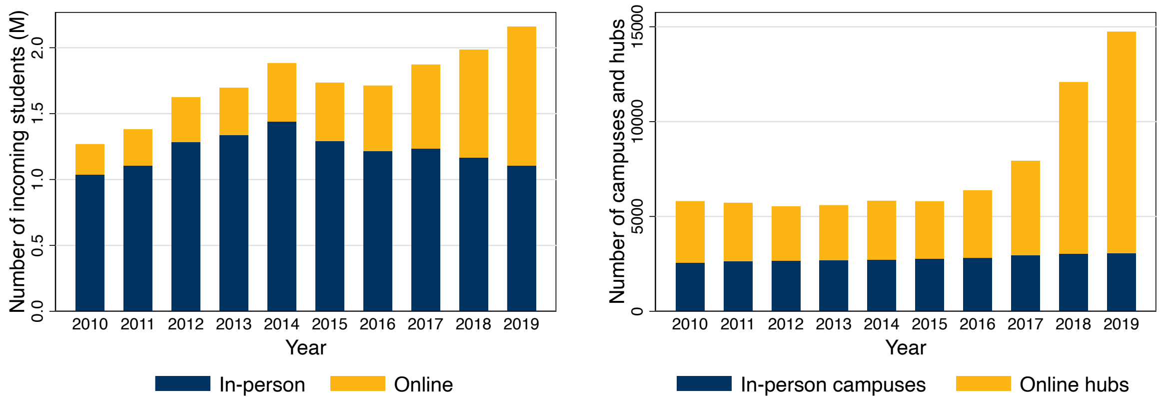
Our value-added estimates serve as a proxy for degree program quality, which we use to evaluate whether students enroll in higher- or lower-quality programs following the online expansion. Two caveats are worth noting. First, value added is estimated using students who took ENEM. Taking ENEM is not required for admission to private-sector institutions, and only 35% of enrolled students do so. Second, wage outcomes come from RAIS, which records earnings only for workers in the formal labor market, thereby omitting informal workers—a group that is disproportionately composed of individuals with lower education levels. Despite these limitations, the estimates provide an informative benchmark for assessing how the online expansion reallocated students across programs of different quality.

2.5. *Descriptives*

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

[9] Covariates include gender; flexible age dummies; race dummies; ENEM scores in the five subjects (math, language, social sciences, natural sciences, and a written essay); ENEM year dummies; household income dummies at the time of taking ENEM; a dummy for whether the individual attended a public high school; a dummy for whether the individual graduated from high school that year; the log-GDP per capita of the individual's municipality of residence; and a constant.

similar trends in the number of in-person campuses and online hubs. Between 2010 and 2015, the numbers were comparable, with a marked rise in online hubs following the policy reforms.



(a) Student enrollment

(b) Campuses and hubs

Figure 1: Expansion of online education

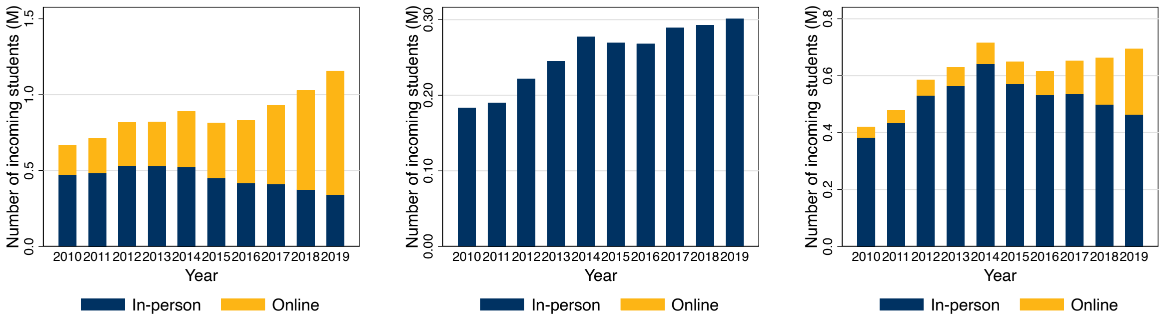
Notes: This figure presents trends in the Brazilian private-sector higher education sector. Panel (a) shows the number of incoming students in private institution for in-person (blue) and online (yellow) degrees over time. Panel (b) shows the number of in-person campuses (blue) and online hubs (yellow) 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 experienced 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, while it continued to rise in fields where online education is restricted.

Finally, we highlight that online degree programs are particularly popular among older students. There is a clear difference in the age distribution of students enrolled in online versus in-person programs. Figure 3 presents online and in-person enrollment by age cohort for 2010 and 2019.[10] While in-person programs are dominated by younger students, online programs attract a majority of students aged 26 and older. By 2019, approximately 64% of online students 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

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



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

Figure 2: Enrollment in online education by field of study

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

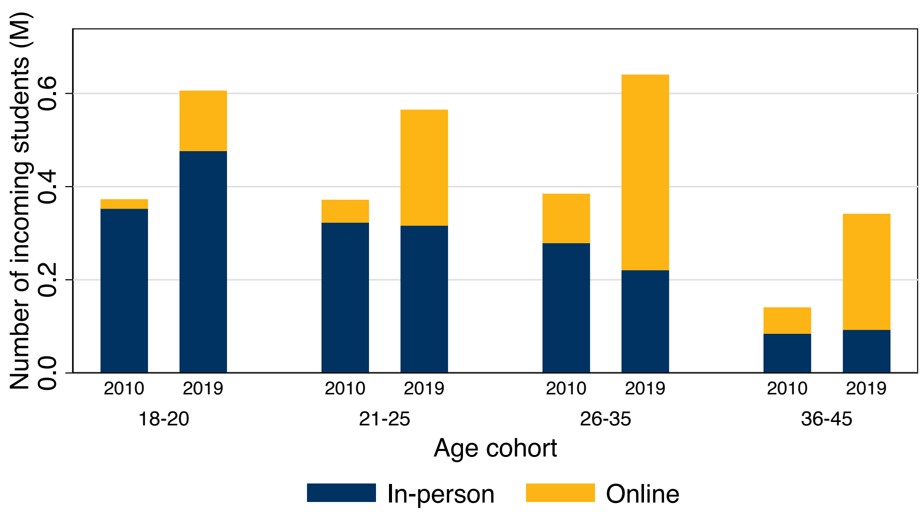


Figure 3: Enrollment in online education by age cohort

Notes: This figure presents enrollment levels in the Brazilian private-sector higher education sector across age cohorts in 2010 and 2019. For each age cohort, the left and right bars represent enrollment in private institutions in 2010 and 2019, respectively, with in-person enrollment shown in blue and online enrollment shown in yellow.

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 required hours to complete the program, (2) the log of tuition price, (3) first-year dropout rates, and (4) value added as estimated in Section 2.4. All regressions are weighted by the number of students enrolled in program j, and standard errors are clustered at the university-major level.

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. Both online and in-person programs require a similar number of hours, as expected under legal mandates. In column (2), we observe that online programs are 44% (i.e., 0.57 log-points) less expensive than their in-person counterparts. Column (3) indicates that online programs have dropout rates that are 0.021 percentage points lower than their in-person counterparts (over a base of 0.20). Lastly, column (4) reveals that the value added of online programs is, on average, 0.076 log-points lower than that of in-person programs (over a base of 0.18 for in-person programs).

Table 1: Comparison between online and in-person programs

	log total hours	log prices	dropout rate	value added
	(1)	(2)	(3)	(4)
Online	-0.003	-0.573	-0.021	-0.076
	(0.005)	(0.030)	(0.011)	(0.006)
Obs.	542,295	547,079	547,079	251,518
Mean dep var. (levels)	3220	4431	0.20	0.15
Region FE	Yes	Yes	Yes	Yes
University-major-year FE	Yes	Yes	Yes	Yes

Notes: This table reports the coefficient from a linear regression of each outcome on an indicator for online degree programs, as specified in Equation (1). All regressions are weighted by the number of students in each degree and include region fixed effects and university-program-year fixed effects. Column (1) reports the log of the number of hours required to complete the program; Column (2) reports the log of tuition price; Column (3) reports first-year dropout rates; and Column (4) reports the value added estimates from Section 2.4.

These findings shed light on the potential trade-offs between in-person and online education. Online education can be offered at significantly lower cost, making tuition more affordable. Lower cost and greater flexibility may slightly increase first-year persistence rates. However, online programs generally provide lower value added compared to in-person programs. Despite this, both online and in-person degree programs, on average, deliver higher value added than the alternative of not attending college (see Online Appendix C, Figure C.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 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 described 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 corresponding change in one of the following outcomes: (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, and (iv) the average price of in-person degrees. The error term ε_{ra} captures unobserved shocks at the region-field level that influence the trend in y_{ra}. The coefficient ϕ is the key parameter of interest.[11]

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 requires a no-interference assumption implying that changes in the number of online degrees offered in region r and field of study a affect only the outcomes in that particular region and field, 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 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 the coefficient ϕ when estimating Equation (2) using OLS. Next, we introduce a shift-share instrumental variable approach to address potential identification threats in the OLS specification. 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 equivalent to a parallel trends assumption, which states that the trajectory of outcomes for regions and fields of study with lower online degree growth represents the hypothetical outcomes in higher-growth regions and fields had they experienced lower online growth (Callaway et al., 2024). Standard tests for this assumption—based on pre-trends evaluations—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.

[11] 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} denote the outcome of interest and the number of online degrees, respectively, in region r, field of study a, and year t; and δ_{ra}, δ_t, and ε_{rat} represent region-field fixed effects, year fixed effects, and region-field-year specific shocks, respectively.

3.2. *Shift-share instrumental variables*

We address endogeneity concerns from 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.[12] Our *shares* variable is constructed from a combination of three components—described in detail below—that predict the likelihood of online entry. For the *shift*, we use the rapid expansion of online degrees, driven by growing demand, advancements in online technology, and the policy changes discussed in Section 2.1.

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

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

where ΔN_f^o represents the *shift*, capturing the total number of online degrees introduced by institution f between 2010 and 2019. We allow this shift to be correlated with the distribution of shocks across regions and fields, ε_{ra}. The variable $z_{fr}z_{fa}z_a$ corresponds to the exogenous *shares* that predict the regions r and fields of study a in which institution f is likely to expand. These shares are derived from three sources: (i) differences in regions' exposure to potential entrants based on the distance between the regions and institutions' headquarters, z_{fr}; (ii) the institutions' propensity to expand in different fields of study given their specialization across fields based on their 2010 in-person offerings, z_{fa}; and (iii) an indicator for whether online education is permitted in a given field of study, z_a.[13] We begin by detailing the *shift*, followed by an explanation of each component of the *share*.

Figure 4 visualizes each component. In each panel, we order institutions by the location of their headquarters from northwest to southeast, based on the IBGE encoding system. We describe 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 as the change in the total number of online degrees that each institution offered across all regions and fields of study in 2010 and 2019:

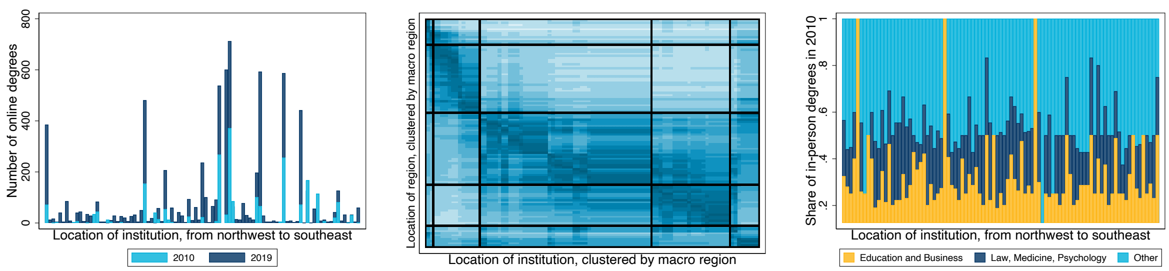
$$\Delta N_f^o \quad = \quad \sum_r \sum_a \left(N_{fra,2019}^o - N_{fra,2010}^o \right), \tag{4}$$

[12] 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.

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

where $N^o_{fra,t}$ equals 1 if institution f offers a degree in field of study a in region r in year t.[14] In total, our sample includes 474 institutions, of which 93 expanded their online offerings between 2010 and 2019.

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



(a) Online degrees by firm (N^o_f)

(b) Prob. of online regional entry by firm (z_{fr})

(c) Share of degrees across field of study by firm (z_{fa})

Figure 4: Visualization of shift-share instrument components

Notes: This figure illustrates the main patterns in the data driving the 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 2010 and 2019 (N^o_f). Institutions are arranged by their headquarters' region from northwest to southeast based on the IBGE encoding system. ΔN^o_f 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 is one of the 110 regions in the sample and each column corresponds to one of the 93 institutions that expanded their online presence between 2010 and 2019. Institutions and regions are are arranged from northwest to southeast based on their location. Black lines separate Brazil into five macro-regions, so that all cells in diagonal blocks represent region-institution pairs within the same macro-region. Cells are shaded in blue, with lighter shades indicating lower exposure (lower likelihood of entry) 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 expanding institutions and are arranged from northwest to southeast by headquarters' region.

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 analysis shows that institutions are more likely to establish online hubs—and consequently offer degrees—in regions closer to their headquarters, likely reflecting lower costs of launching and maintaining nearby operations.[15] We document this pattern estimating the following regression:

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

[14] Our results are robust to using a leave-one-out estimate of ΔN^o_f. Because our identification strategy relies on having exogenous shares, a leave-one-out estimator is unnecessary in this setting.

[15] 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.

where Entered_{fr} equals 1 if institution f had entered region r by 2010 and 0 otherwise, and d_{fr} represents the distance between the headquarters of institution f and region r.[16] 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 statistically 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%. Detailed results are reported in Online Appendix A, Table A.2.

We use the estimates from Equation (5) to predict the likelihood that institution f will open an online hub in region r, based solely on 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.[17] Our measure of exposure is defined as:

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

Figure 4(b) illustrates the exposure instrument, z_{fr}. The y-axis represents 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 (lower likelihood of entry) and darker shades indicating higher exposure. 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 estimates the likelihood that each institution enters a given region. However, institutions may differ in their propensity to expand across fields of study. To capture this, we use institutions' 2010 offerings to predict likely fields of expansion, under the assumption 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. Specifically, we estimate the likelihood of expansion

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

[17] 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.

in each field as:

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

where $N^{\iota}_{fra,2010}$ takes the value of 1 if institution f offers an in-person degree in field a in region r in 2010, and 0 otherwise.

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). Some institutions concentrated their offerings in fields such as Education and Business, while others prioritized areas like Engineering or Math. Our instrument predicts more intense online expansion in field a in regions with nearby institutions that were already specializing in that field in 2010.

3.2.4. *Fields of study regulatory constraints (z_a)*: Regulations prohibit online education in Law, Medicine, and Psychology. We use this constraint to build an indicator variable, z_a, which equals 1 for fields of study permitted to expand online and 0 otherwise. The variable z_a in Equation (3) ensures that our instrument predicts zero online growth in restricted fields such as Law, Medicine, and Psychology.

3.3. *Identification*

As is standard in linear models using instrumental variables, two assumptions must hold for a valid causal interpretation. First, the instrument must be relevant; that is it must have predictive power over the endogenous variable. We test this by estimating a first-stage regression of the endogenous variable, ΔN^o_{ra}—the change in the number of online degrees offered in region r and field a—on the SSIV, z_{ra}, yielding an F-statistic of 185. The coefficient from this regression is reported in Table 2, Panel B, Column (1). Second, the exclusion restriction must hold. Specifically, the exposure-shares product, $z_{fr}z_{fa}z_a$, 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, define $\bar{\epsilon}_{fr} = \mathbb{E}[z_{fa}z_a\varepsilon_{ra}|fr]$, which captures the tendency of institution f to expand in region r because of anticipated region-specific shocks $\{\varepsilon_{ra}\}_a$. For example, if institution f specializes in business degrees and region r is expected to experience rising demand for online business education, then $\bar{\epsilon}_{fr}$ will be high. Using the law of iterated expectations, the exclusion restriction can be restated as $\mathbb{E}[z_{fr}z_{fa}z_a\varepsilon_{ra}] = \mathbb{E}[z_{fr}\bar{\epsilon}_{fr}] = 0$. This condition holds if the distance between institutions and regions—the basis for our exposure shares—is uncorrelated with unobserved, region-specific shocks affecting the fields in which those institutions traditionally specialize.

In Online Appendix D, we follow Goldsmith-Pinkham et al. (2020) and examine the identifying assumptions more directly. First, we compute Rotemberg weights to identify the firms that contribute most to the estimator and show that our results are robust when instruments are constructed from each of these firms separately. Second, we exploit the acceleration of online education following the 2016 reforms by re-estimating the model separately for 2010–2014 and 2015–2019, periods with lower and higher online-degrees entry, respectively. Despite large differences in the IV first-stage coefficient across the two periods, the IV estimates remain stable. These exercises provide supportive evidence for the validity of our identification assumption.

3.4. *Results*

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

3.4.1. *Overall 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 Table 2, Columns (2) and (3), respectively. These outcomes are expressed relative to market size, defined as the number of individuals aged 18–45 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 0.386 students per 1,000 individuals in the market, equivalent to a 14% increase in total enrollment when the number of online degrees in that field rises by 50%. In contrast, Column (3) shows that this increase in online degrees reduced in in-person enrollment in the same field by 0.2 students per 1,000 individuals, corresponding to a 7% decline in total enrollment under the same 50% increase in online degree availability.[18]

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, 51% are new to higher education, whereas 49% would otherwise have enrolled in an in-person degree.[19] These forces generate an ambiguous effect on total value added. Market expansion increases value added, as newly enrolled students enter degrees with positive value compared to the outside option of no college, which we nor-

[18]The average number of online degrees in a given region and field of study increased from 3.10 in 2010 to 9.91 in 2019, representing an increase of 220%.

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

malize to have zero value added. In contrast, market diversion reduces value added by shifting students from higher value-added in-person degrees to lower value-added online alternatives. Using 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 8.7% due to higher online enrollment, while the corresponding decline in in-person enrollment reduces value added by 7.6%. These effects largely offset each other, resulting in a net increase in total value added of 1.02%.[20]

Table 2: Effect of an additional online degree on enrollment, degree availability, and tuition

	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					
Δ in online degrees		0.342	-0.192	-0.163	-0.009
		(0.033)	(0.026)	(0.028)	(0.002)
Panel B: SSIV					
shift-share instrument	1.726				
	(0.127)				
Δ in online degrees		0.386	-0.190	-0.185	-0.014
		(0.035)	(0.033)	(0.038)	(0.003)
Panel C: Mean of the dependant variable in 2010 and 2019					
2010	3.10	0.80	3.41	11.43	1.59
2019	9.91	3.06	3.24	12.44	1.61
Obs.	1,100	1,100	1,100	1,100	893

Notes: This table reports estimates of the linear model in Equation (2), $\Delta y_{ra} = \phi \Delta N^o_{ra} + \varepsilon_{ra}$, for several outcomes, y_{ra}, using two identification strategies. Panel A presents OLS estimates. Panel B reports 2SLS estimates using the shift-share instrument described in Section 3.2. Panel C reports the mean of each dependent variable (in levels) in 2010 and 2019. Column (1) shows the 2SLS first-stage regression of the change in the number of online degrees on the shift-share instrument. Columns (2)–(5) report coefficients for the number of online students relative to market size, the number of in-person students relative to market size, the number of in-person degrees, and the log price of in-person degrees. Columns (1)–(4) use the full set of region-field pairs; Column (5) uses the 893 region-field pairs with at least one in-person degree in both 2010 and 2019. Standard errors clustered at the region-field level are shown in parentheses.

We next analyze the consequences of the online expansion on the availability and pricing of in-person degrees. Column (4) shows that regions and fields experiencing larger online expansion saw a decline in the number of in-person degrees. Specifically, for each additional online degree, there are 0.185 fewer in-person degrees relative to a 2010 baseline of 11.43. These effects stem from both higher exit rates of in-person programs and the deterrence of new program entry, which together exacerbate the diversion from in-person to online alternatives. As a result,

[20] 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 0.096 and 0.172, respectively.

some students who prefer a particular in-person program may be forced to switch to an online alternative when their preferred program exits the market. Finally, Column (5) shows that the average price of in-person degrees dropped by 1.4% for each additional online degree introduced in a region-field pair. This supports the idea that online degrees intensify local competition, reducing prices and deterring new in-person program entry.

In Online Appendix A, Table A.3, we show that our main results are robust to controlling for regional changes in internet penetration and GDP per capita.

3.4.2. *Effects by age cohorts*: The online education expansion may have heterogenous effects across age cohorts. Evidence in other contexts shows that online programs disproportionately attract older individuals, who often face higher costs and challenges 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 enroll in an in-person degree over an online degree than those aged 36–45. Differences in preferences across age cohorts may lead to heterogeneous patterns of market expansion and student diversion that we examine next.

Table 3 presents estimates of enrollment changes across age cohorts. The results indicate substantial substitution from in-person to online degrees among younger cohorts. For students aged 18–21, each additional online enrollment is associated with 0.70 fewer students in 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, the corresponding substitution is 0.54 students (ratio of the coefficients in Column 6 to Column 2, Panel B), and it declines further for older cohorts: 0.44 for ages 26–35 and 0.26 for ages 36–45 (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 2% among students aged 18–20 but increases value added by 25% among students aged 36–45.[21]

Overall, Tables 2 and 3 suggest that online education has the potential to expand access to higher education, particularly for older cohorts. At the same time, online education diverts students from in-person alternatives. Increased competition from online programs lowers in-person tuition but reduces the number of in-person degrees offered. Consequently, the average value added in the economy declines, and students who prefer in-person learning may find themselves pushed into online alternatives when their preferred programs exit the market.

Our analysis so far relies on a linear model that assumes no interaction between degrees across fields within a region. This assumption prevents the model from separating market growth driven by new students entering college from shifts caused by students reallocating across fields. Moreover, the linear approach is unsuitable for out-of-sample counterfactuals, where competition, pricing, and entry/exit decisions interact nonlinearly.[22] We address these

[21]The average value added of online and in-person degrees in 2010 for the 18–20 cohort is 0.102 and 0.243, respectively, while for the 35–45 cohort it is 0.100 and 0.077, respectively.

[22]Nonlinear effects are important for 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

Table 3: Effect of an additional online degree on enrollment, by age cohort

	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								
Δ in online degrees	0.330	0.412	0.369	0.262	-0.344	-0.237	-0.172	-0.082
	(0.045)	(0.043)	(0.036)	(0.024)	(0.057)	(0.037)	(0.030)	(0.012)
Panel B: SSIV								
Δ in online degrees	0.385	0.465	0.414	0.290	-0.268	-0.249	-0.183	-0.076
	(0.054)	(0.045)	(0.037)	(0.025)	(0.078)	(0.044)	(0.034)	(0.016)
Panel C: Mean of the dependant variable in 2010 and 2019								
2010	0.74	0.83	1.05	0.90	6.99	2.97	1.47	0.70
2019	3.85	3.39	3.35	2.80	9.02	2.85	1.11	0.60
Obs.	1,100	1,100	1,100	1,100	1,100	1,100	1,100	1,100

Notes: This table reports estimates of the linear model in Equation (2), $\Delta y_{ra} = \phi \Delta N^o_{ra} + \varepsilon_{ra}$, for changes in the number of in-person and online students across four age cohorts (18–20, 21–25, 26–35, 36–45), using two identification strategies. Panel A presents OLS estimates. Panel B reports 2SLS estimates using the shift-share instrument described in Section 3.2. Panel C reports the mean of each dependent variable (in levels) in 2010 and 2019. Columns (1)–(4) show coefficients for changes in online students, and Columns (5)–(8) show coefficients for changes in in-person students, relative to market size for the corresponding age cohort. Standard errors clustered at the region-field level are shown in parentheses.

limitations in the next section.

4. Model

We develop and estimate an equilibrium model of the Brazilian college education market, covering both in-person and online formats. This model addresses two key limitations of linear models. First, by explicitly modeling demand, it allows for flexible substitution across fields of study. Second, by modeling supply, it enables counterfactual analysis that incorporates 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 counterfactual policies that restrict online education to older cohorts.

4.1. *Setup*

The model comprises institutions (firms) $f \in \mathcal{F}$, that offer undergraduate degrees, and students $i \in \mathcal{I}$ who choose which degree to attend, if any. We define degrees as the combination of the offering firm f, the field of study $a \in \mathcal{A}$, and the delivery mode (in-person or online), and index them by $j \in \{\mathcal{F} \times \mathcal{A} \times \{0, 1\}\}$. Geographical regions are indexed by $r \in \mathcal{R}$ and years by $t \in \mathcal{T}$. We define markets as the intersection of a region r and a year t. A single degree j can be

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.

available in multiple markets, and a market may offer multiple degrees. In each market, a firm f offers a product bundle $\mathcal{J}_{frt} \in \mathcal{B}_f$, which consists of one or more degrees. The empty bundle $\mathcal{J}_{frt} = \emptyset$ represents the option of offering no degree and is available to every firm.[23]

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 observe the pre-existing market structure, $\mathcal{J}_{frt_0}$, and all relevant characteristics of demand as well as marginal and fixed costs of potential entrants up to idiosyncratic shocks. They also observe their own private-information fixed-costs shocks. Institutions' entry decisions determine the market structure. After entry decisions are made, demand and marginal cost shocks are realized. In the second stage, institutions compete à 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, institutions' pricing decisions; third, institutions' optimal bundle choice.

4.2. *Demand*

In each year t and region r, a potential student $i \in \mathcal{I}_{rt}$, choose either to enroll in one of the degrees available $j \in \mathcal{J}_{rt} \equiv \cup_f \mathcal{J}_{frt}$, or not to enroll at all. Each degree j is characterized by its offering institution f, its field of study a, and a vector of characteristics $x_j = [x_j^{(1)}, x_j^{(2)}]$ defined below.

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

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

where p_{jrt} denote tuition, $x_j^{(1)}$ is a vector that indicates delivery mode (online or in-person; i.e., $x_j^{(1)} \equiv [x_{j1}^{(1)}, x_{j2}^{(1)}] \in \{[1,0], [0,1]\}$), w_{jrt} is a demand shifter based on region r's internet penetration in year t interacted with delivery mode, 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 cohort $b(i)$. The term ϵ_{ijrt} represents a consumer-specific demand shock following a generalized extreme value distribution, consistent with a nested logit model, with nests defined by the degree's field of study and intra-nest correlation ρ.

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

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

where $x_j^{(2)}$ includes 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 hours); and the degree's

[23]The set of available bundles, $\mathcal{B}_f$, is firm-specific to allow for firm-level specialization. For instance, some institutions may be unable to offer degrees in fields requiring specialized technology, such as Medicine.

STEM load, calculated as the share of degree programs within the degree that are STEM. The term δ_j captures degree-specific utility components, δ_{ra} captures region-field-specific factors, δ_{ta} represents year-field-specific components, δ_{to_j} captures online-year factors, allowing for differential yearly demand shifts between online and in-person education. We refer to these terms as the 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, which is calculated as:

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

where $\mathbf{p_{rt}}$ denotes the market vector of degree prices, and $\Theta_{rt} = \{i \in \mathcal{I}_{rt} : u_{ijrt} \geq u_{ikrt}, \ \ \forall k \in \mathcal{J}_{rt}\}$ denotes the set of potential students in region r and year t who choose 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 instrumental variables and micro-moments to identify the model parameters.[24]

Instruments for prices. A key challenge in demand estimation is price endogeneity, as institutions 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 not available, proxies can serve as practical alternatives. Following Hausman et al. (1994), we use the contemporaneous prices of the same degree in other regions as a cost-shifter 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 where degree j is available other than r. This instrument exploits the idea that variation in firm-level costs impacts prices across all markets where a degree is offered. The primary identification assumption is that while costs shocks 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, as well as the nesting parameter ρ. To identify these parameters, we construct instruments that shift the choice set $\mathcal{J}_{rt}$ but are uncorrelated with demand shocks. Following Section 3.2, we build two shift-share instruments that predict the number of online degrees, z^o_{rta}, and in-person degrees, z^ι_{rta}, offered from a given field of study in each market:

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

where z_{fr} represents the likelihood of institution f expanding to region r, based on the distance

[24] 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) and Conlon and Gortmaker (2025).

to its headquarters (Section 3.2.2); z_{fa} captures the institution's propensity to expand into different fields (Section 3.2.3); z_a an indicator equal to zero for fields where online instruction is not allowed (Section 3.2.4); 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.

We then use functions of these instruments to construct four additional instruments that predict: i) the number of degrees offered on the same field of study and mode of delivery, $z^m_{jrt} = x^{(1)}_{j1} z^o_{rta} + x^{(1)}_{j2} z^\iota_{rta}$, where $x^{(1)}_{j1}$ and $x^{(1)}_{j2}$ denote whether degree j is online or in-person; ii) the total number of degrees offered in the same mode of delivery across all fields, $\bar{z}^m_{jrt} = \sum_a \left(x^{(1)}_{j1} z^o_{rta} + x^{(1)}_{j2} z^\iota_{rta} \right)$; iii) the total number of online degrees offered, $\bar{z}^o_{rt} = \sum_a z^o_{rta}$; and iv) the total number of in-person degrees offered, $\bar{z}^\iota_{rt} = \sum_a z^\iota_{rta}$. Together, these six instruments shift consumer choice sets and provide variation to identify the variance of β_i, the random coefficients on the online and in-person indicators, and the nesting parameter ρ.

Finally, following Gandhi and Houde (2019), we construct differentiation instruments based on the distance between a degree's predicted price and other degrees' predicted prices. The first instrument captures the average absolute predicted price difference, $z^{gh1}_{jrt} = \frac{1}{|\mathcal{J}_{rt}|-1} \sum_{k \in \mathcal{J}_{rt}} |\hat{p}_{jrt} - \hat{p}_{krt}|$, and the second instrument measures the number of degrees whose predicted prices lie within a narrow band, $z^{gh2}_{jrt} = \sum_{k \in \mathcal{J}_{rt}} \mathbb{1}\{|\hat{p}_{jrt} - \hat{p}_{krt}| < \frac{1}{10}\sigma_{\hat{p}}\}$, where $\hat{p}_{jrt}$ is the predicted price based on the reduced-form pricing equation constructed using the price instrument and fixed effects, and $\sigma_{\hat{p}}$ is the standard deviation of $\hat{p}_{jrt}$.[25] These instruments generate additional variation that helps identify the variance of the random coefficient on price, α_i.

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 empirical counterparts. We use eight moments defined by the probability that students from each of four age cohorts defined above choose to enroll in-person or online. Formally, these moments are defined as $\mathbb{E}[\sum_{j \in \mathcal{J}_{rt}} s_{ijrt} x^{(1)}_{jk} | i \in b]$ for $k \in \{1, 2\}$, where s_{ijrt} is the probability that individual i chooses degree j, $x^{(1)}_{jk}$ is the k-th element of $x^{(1)}_j$ (i.e., online or in-person), and b represents age cohorts corresponding to the age cohorts 18–20, 21–25, 26–35, and 36–45.

Estimator. The estimator proposed by Berry et al. (1995) yields 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 components of δ_{jrt} from Equation (9) as nuisance parameters that might be imprecisely estimated. This is typically inconsequential because market shares and price elasticities depend only on δ_{jrt} and not on the individual components of δ_{jrt}. In our setting, however, we are interested in recovering demand for degrees that are not offered in certain markets in the data and for which we do not observe estimates of δ_{jrt}. This requires reliable estimates of the components of δ_{jrt}.

To estimate the demand model, we proceed in two steps. In the first step, we follow Berry

[25] The reduced form regression is given by $p_{jrt} = \zeta_p z^p_{jrt} + \zeta_w w_{jrt} + z_{jrt}\zeta_z + \zeta_j + \zeta_{ra} + \zeta_{ta} + \zeta_{to_j} + \varepsilon_{jrt}$, where z^p_{jrt} is the instrument for price, $z_{jrt} = [z^o_{rta}, z^\iota_{rta}, z^m_{jrt}, \bar{z}^m_{jrt}, \bar{z}^o_{rt}, \bar{z}^\iota_{rt}]$ is the set of choice shifters, 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, respectively.

et al. (1995) and estimate the parameters from Equation (8) using the generalized method of moments. We absorb the components of δ_{jrt} via fixed effects. Due to collinearity, the parameter on $x_j^{(2)}$, δ, is absorbed by δ_j during estimation. We use a total of seventeen moments. Nine moments are 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^p_{jrt}, z^o_{rta}, z^ι_{rta}, z^m_{jrt}, $\bar{z}^m_{jrt}$, $\bar{z}^o_{rt}$, $\bar{z}^\iota_{rt}$, z^{gh1}_{jrt}, z^{gh2}_{jrt}. The remaining eight moments are the 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 component 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 components of δ_{jrt}, which we then use to project utility into degree-markets, whether or not they are observed in the data.[26] This second step is in the spirit of Abdulkadiroğlu et al. (2020) and Andrews et al. (2024), and serves to shrink imprecise estimates, thereby reducing noise in the profit function and improving the estimation of the entry game.

4.2.2. *Results*: The estimated parameters are reported in Online Appendix A, Table A.4. To summarize, Table 4 presents median own-price elasticities and diversion ratios for 2010 and 2019. We estimate median own-price elasticities of approximately -3.2 for in-person degrees and -1.3 for online degrees, consistent with prior estimates of demand for in-person degrees in the U.S. (Armona and Cao, 2024) and Brazil (Barahona et al., 2025). Regarding diversion ratios, by 2019, a marginal increase in the price of an in-person degree induces approximately 56% of students who leave that degree to switch to another in-person program, 18% to switch to an online program, and 24% to exit higher education altogether. Older cohorts exhibit stronger preferences for online degrees (see Online Appendix A, Table A.5). Finally, we find that higher internet penetration increases demand for both online and in-person degrees, which we exploit as an additional source of variation to estimate entry fixed costs in the next subsections.

4.3. *Institutions' pricing decisions*

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

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

[26] 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.

Table 4: Median own-price elasticities and diversion ratios by degree type and year

	$t = 2010$		$t = 2019$	
	In-person	Online	In-person	Online
Median own-price elasticity:	-3.02	-1.24	-3.23	-1.26
Median diversion ratios:				
To in-person:	0.66	0.36	0.56	0.17
To online:	0.04	0.35	0.18	0.58
To outside option:	0.26	0.26	0.24	0.25

Notes: The table reports median own-price elasticities and diversion ratios across products and markets for 2010 and 2019, separately for in-person and online degrees. Diversion ratios measure the fraction of students among those who decide to leave degree j in response to an increase in tuition that would switch to 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 $\mathcal{K}$ is 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 in-person degrees j that are in-person with $\mathcal{K}$ defined as the set of all online degrees. Since the reported diversion ratios are based on medians, they are not constrained to sum to one.

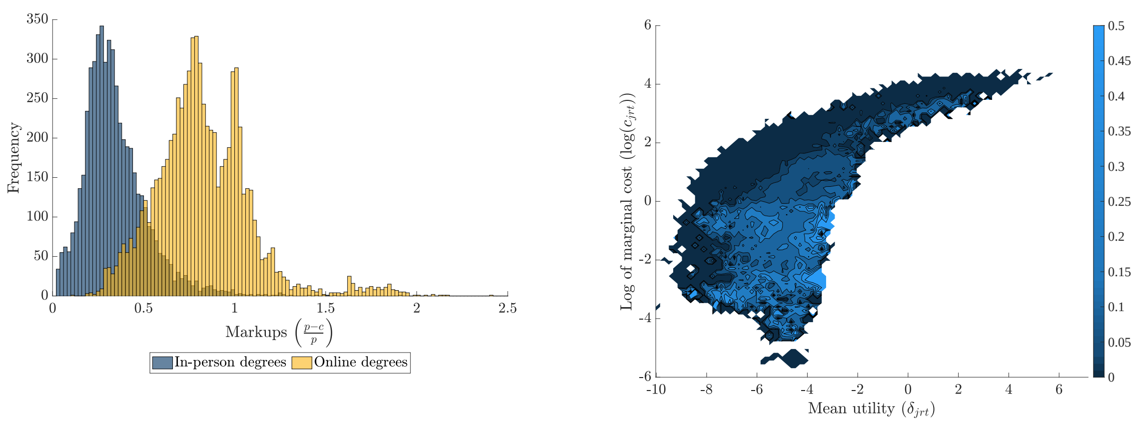
Marginal costs are modeled as:

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

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 headquarter's location, as in Equation (5). The term z^p_{jrt} represents the price instrument used to estimate demand, $x_j^{(2)}$ denotes the degree characteristics from Equation (9), γ_j captures degree-specific components of the cost function, γ_{ta} represent year-field-specific components, γ_{to} reflects online-year-specific factors that allow for differential shifts in marginal costs between online and in-person degrees. Lastly, ω_{jrt} is a degree-region-year idiosyncratic cost shock. Institutions compete à 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, absorbing the components 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 are interested in recovering cost estimates for degrees that may not be offered in certain markets in the data, we require reliable estimates for each component of γ_{jrt}. In the third step, we estimate a mixed-effects Bayesian hierarchical model to recover these component, assuming $\gamma_j \sim \mathcal{N}(0, \varsigma_j^2)$, $\gamma_{ra} \sim \mathcal{N}(0, \varsigma_{ra}^2)$, $\gamma_{ta} \sim \mathcal{N}(0, \varsigma_{ta}^2)$, $\gamma_{to} \sim \mathcal{N}(0, \varsigma_{to}^2)$, and $\omega_{jrt} \sim \mathcal{N}(0, \varsigma_\omega^2)$ random coefficients, and γ_x a non-random parameter that allows for flexible correlation between γ_{jrt} and $x_j^{(2)}$. We estimate this model via maximum likelihood and use the resulting posterior means to estimate marginal costs for all degree-markets, regardless of whether they are observed in the data.

4.3.2. *Results*: Figure 5(a) presents the distribution of markups—defined as the ratio of price minus marginal cost to price—for in-person and online degrees. The corresponding estimated parameters are reported in Online Appendix A, Table A.6. We find that online degrees exhibit higher markups compared to in-person degrees, consistent with institutions pricing online alternatives on a more inelastic region of the demand curve. However, despite having larger markups, online degrees have lower prices due to substantially lower marginal costs. The average annual tuition is \$5655 for in-person degrees and \$1792 for online degrees, while the corresponding average estimated marginal costs are \$4208 and \$454, respectively.



(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) reports the distribution of markups, defined as the ratio of the difference between prices and marginal costs to prices, for in-person (blue) and online (yellow) degrees. Panel (b) illustrates the relationship between the determinants of demand and marginal costs and the empirical probability that a degree is offered in a given market. The x-axis corresponds to the components of δ_{jrt}, and the y-axis shows the logarithm of predicted marginal cost. Shading indicates the empirical probability that degrees with given characteristics were offered in 2019, with lighter colors denoting higher probabilities. Unshaded areas correspond to combinations of mean utility and marginal cost for which no potential degree exists. Entry probabilities are directly estimated from the data and do not rely on the entry model.

We also find that degrees that face higher demand—measured by their mean utility, δ_{jrt}—and lower marginal costs, are offered in a larger number of regions in 2019. Figure 5(b) shows a heatmap illustrating the relationship between degrees' demand, marginal cost, and their empirical probability of being offered in 2019. Higher probabilities are concentrated in the south-east region of the map. This finding suggests that, in the data, institutions tend to select the more profitable degrees when deciding which bundles to offer, consistent with economic theory.[27]

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 adopt a static approach to the problem for two reasons. First, adjustment frictions make the timing of institutions' responses to economic shocks uncertain. To avoid modeling such frictions, we compare medium-run changes in market structure over a nine-year span. Second, in our setting, many institutions offer multiple degrees per market, resulting in a highly-dimensional state space and renders the problem computationally infeasible.[28] This static approach parsimoniously captures the key trade-offs institutions face when choosing degree offerings.

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

$$FC_{fr}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) = \text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) + \text{Infrastructure}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) + \sigma_\varepsilon \varepsilon_{fr\mathcal{J}_{frt}}, \qquad (13)$$

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

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

$$\text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) \quad = \quad \mathbb{1}\{\mathcal{J}_{frt} \neq \emptyset\} \cdot \kappa_0 + \sum_{k \in \mathcal{J}_{frt} \cap \mathcal{J}_{frt_0}} \kappa^{\text{old}}_{a(k)o(k)} + \sum_{k \in \mathcal{J}_{frt} \setminus \mathcal{J}_{frt_0}} \kappa^{\text{new}}_{a(k)o(k)}, \qquad (14)$$

where κ_0 captures an operational fixed cost that is incurred when at least one degree is offered, $\kappa^{\text{old}}_{a(k)o(k)}$ denotes the cost of maintaining an existing degree, and $\kappa^{\text{new}}_{a(k)o(k)}$ denotes the cost of opening a new degree using existing infrastructure. Note that these costs vary only by field of study, indexed by $a(k)$, and by delivery mode, indexed by $o(k)$. We restrict κ_0, κ^{old}_{ao}, and κ^{new}_{ao}

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

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

to be non-negative, ensuring that offering degrees is costly.

The second component captures the cost of opening a new campus or hub needed to offer degrees in region r:

$$\text{Infrastructure}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) = \sum_{k\in\{\text{campus,hub}\}} \mathbb{1}\{\text{New}_{frk}\}\left(\chi_0^k + g(d_{fr})'\chi^k\right), \tag{15}$$

where $\mathbb{1}\{\text{New}_{frk}\}$, for $k \in \{\text{campus}, \text{hub}\}$, is an indicator that equals 1 if $\mathcal{J}_{frt}$ includes an in-person or online degree (for campus and hub, respectively), while $\mathcal{J}_{frt_0}$ does not. This indicator captures the transition in a firm's offerings in region r, reflecting the need to invest in new campus infrastructure for in-person degrees, or to develop an online hub for online degrees if the firm had not previously offered degrees of that type. The vector $g(d_{fr})$ is the distance function between the firm's headquarters and the region r, as defined in Equation (5). To allow for infrastructure-specific differences in cost, χ_0^k and χ^k vary between constructing a campus or developing an online hub. As before, we impose the restriction that $\left(\chi_0^k + g(d_{fr})'\chi^k\right)$ 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. Given this information, firm f solves:

$$\Pi_{frt} = \max_{\mathcal{J}_{frt}\in\mathcal{B}_f} \mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})] - FC_{fr}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}), \tag{16}$$

where $\pi_{frt}(\mathcal{J}_{frt})$ is defined by Equation (11) and the expectation is taken over the distribution of other firms' idiosyncratic shocks, $\varepsilon_{f'r\mathcal{J}_{f'rt}}$, as well as 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 firms' bundle choices across markets induced by instruments that shift firms' expected profits but do not enter the fixed cost function. 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 E.1, we show that higher growth in internet penetration is associated with higher online entry. The exclusion restriction is that while greater internet penetration may increase demand for online programs, it does not affect the 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 are lower in regions located near the headquarters of competing institutions that offer similar degrees. However, we allow fixed costs to depend only on firm f's distance to region r and not on the distances of its competitors.[29]

[29] Our 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).

A common concern related to discrete games with incomplete information is the potential for multiple equilibria, complicating estimation and counterfactual analysis. Our identification argument follows the framework for inference in incomplete-information games of Aradillas-López (2020), in which point identification can be achieved by assuming that the underlying equilibrium selection mechanism is degenerate—that is, the data come from a single equilibrium—without having to assume which equilibrium is chosen.[30] We follow the literature and adopt the same assumption (Seim, 2006; Bajari et al., 2010; Bhattacharya et al., 2025; Atal et al., 2025).

We use the firm's maximization problem to recover bundle choice probabilities. From Equation (16), the probability that firm f chooses bundle $\mathcal{J}_{frt}$ is

$$\phi_{\mathcal{J}_{frt}} = \frac{\exp(\frac{1}{\sigma_\varepsilon}\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})] - \text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) - \text{Infrastucture}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}))}{\sum_{\mathcal{J}\in\mathcal{B}_f}\exp(\frac{1}{\sigma_\varepsilon}\mathbb{E}[\pi_{frt}(\mathcal{J})] - \text{Degrees}(\mathcal{J}|\mathcal{J}_{frt_0}) - \text{Infrastucture}(\mathcal{J}|\mathcal{J}_{frt_0}))}. \tag{17}$$

If expected profits were observed and independent of 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 (17) are a function of the choice probabilities, $\phi_{\mathcal{J}_{frt}}$, on the left-hand side of the same equation. Earlier literature addresses this problem by implementing a nested fixed-point likelihood estimator (Seim, 2006). In our setting, this approach is infeasible due to the high computational burden—it requires solving for an equilibrium for each candidate set of parameters in each market. To circumvent this issue, we adopt a two-step estimation procedure, following Sweeting (2009).

In the first step, we compute choice probabilities directly from the data. We construct a proxy variable that approximates expected profits in a parsimonious way, using demand and marginal cost estimates evaluated at $\xi_{jrt} = 0$ and $\omega_{jrt} = 0$, respectively. We then estimate an analogue of Equation (17), replacing $\mathbb{E}[\pi_{frt}(\mathcal{J})]$ with this proxy variable, and predict the probability that each bundle is offered in the data. We denote the estimated probabilities by $\hat{\phi}_{\mathcal{J}frt}$.

We use the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, to compute 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 by solving the static equilibrium game defined in Equation (11). We simulate 10,000 draws per market and integrate over all draws to obtain expected variable profits for each bundle. Finally, we fit a random forest prediction model to reduce noise and produce 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 $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$ to recover the fixed-cost parameters via maximum

[30] Multiple equilibria are more prevalent in complete-information games, where assuming a degenerate equilibrium selection mechanism is not enough for identification (Aradillas-López, 2020). Note that counterfactual analysis still requires assumptions about which equilibrium is selected.

likelihood, replacing $\mathbb{E}[\pi_{frt}(\cdot)]$ with $\hat{\mathbb{E}}[\pi_{frt}(\cdot)]$ in Equation (17). Further details on the estimation procedure are provided in Online Appendix E.2.

4.4.2. *Results*: We report the estimated parameters in Online Appendix A, Table A.7. The median entry elasticity with respect to own profits is 0.5, implying that the probability of offering a given bundle increases by 0.5% for each percentage increase in expected profits. We find that adding a new degree to the bundle is more costly than continuing to offer an existing degree. For many fields of study, offering an existing degree does not entail an additional fixed cost (i.e., $\kappa_{ao}^{\text{old}} = 0$), which explains why, conditional on keeping their campus or hub open, institutions continue offering degrees in 2019 that they were already offering in 2010. We also find that the cost of building new infrastructure increases with the distance between regions and firms' headquarters. For a region located 1200 kilometers from an institution's headquarters—the median distance in the data—the cost of establishing a new in-person campus or online hub is \$60 and \$25 per person in the market, respectively. The full distribution of fixed costs associated with opening a new campus or hub as a function of distance is shown in Online Appendix A, Figure A.1.

5. Counterfactuals

We use the model to evaluate 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 enable in-person programs to remain viable while preserving online access for those who benefit most.

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

Using our model, we compare the 2019 status quo, referred to as the *Baseline* (BL), in which online education is available, with counterfactual scenarios in which online education is absent. To isolate the respective contributions of supply and demand, we construct a series of increasingly flexible counterfactuals. We describe these counterfactuals below and summarize them in Table 5.

1. No supply-side responses (NS): We remove online degrees while holding institutions' tuition and program offerings fixed. The change in value added when moving from BL to to NS depends on two factors: (i) differences in value added between online degrees, in-person degrees, and the outside option, and (ii) the extent to which students switch from online to in-person degrees or the outside option.

2. Price responses (PR): Institutions are allowed to adjust prices optimally, while their program offerings remain fixed. This scenario isolates the effects of reduced competition, which may lead to higher prices and displaced students from high-quality in-person degrees.

3. Equilibrium (EQ): Institutions have full flexibility to adjust both prices and degree offerings according to the equilibrium model from Section 4. Under EQ, we expect the availability of in-person degrees to expand, boosting enrollment in these programs and increasing total value added. Online Appendix E.3 describes the algorithm used to compute the market equilibrium.

Table 5: Counterfactual scenarios

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 adjust prices
Equilibrium (EQ)	(PR) + firms adjust degree offerings

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

We examine seven outcome categories: (i) the number of online and in-person degrees, (ii) total college enrollment, separately for in-person and online degrees, (iii) the total value added produced in the economy, (iv) total expenditure on higher education, (v) the average price of in-person degrees, (vi) total variable profits, and (vii) revealed-preference consumer surplus.[31] Online Appendix E.4 provides detailed formulas for the last five outcome measures.

Table 6 presents the outcomes under the Baseline scenario (BL) and the three counterfactuals. Under BL, an average market offers 41 online degrees and 37.1 in-person degrees. Total college enrollment reaches 5.6%, with 64.6% of students attending in-person and 35.4% attending online. The total value added per enrolled student is 0.172 log points, and students spend an average of US$5011 in tuition. On the firm side, the average price of in-person degrees is approximately US$6100 per year, and average industry profits per enrolled student are US$1553. Consumer surplus per enrolled student is $5038 higher than in a world without higher education.

Under the NS scenario, where online degrees are removed, and institutions cannot adjust prices or degree offerings, total enrollment declines by 15.1%, falling from 5.6% to 4.7%. This decline highlights the role of online education in expanding access for students who might otherwise not attend college. At the same time, because online programs can attract students who would have otherwise enrolled in in-person degrees, restricting online offerings increases in-person enrollment by 31.4%, from 3.6% to 4.7%. The reduction in overall enrollment is accompanied by a corresponding 2.4% increase in total student expenditure and a 14.7% decrease in firms' profits.

The effect on total value added is ambiguous when online education is restricted. On one hand, because online degrees generate more value than not attending college, the decline in overall enrollment could reduce overall value added. On the other hand, because online degrees typically provide less value added than in-person degrees, the shift toward higher in-person enrollment could increase total value added. In the NS scenario, total value added decreases

[31] Our consumer surplus measure assumes that students make informed enrollment decisions that maximize their welfare. This assumption contrasts with policymakers' concerns about the difficulties students may face in identifying low-quality degrees. We offer value added as a complementary normative outcome.

Table 6: Effects of online education expansion

	BL (1)	NS (2)	PR (3)	EQ (4)	$\Delta_{\text{EQ/BL}}$ (5)
Number of online degrees:	41	0	0	0	—
Number of in-person degrees:	37.1	37.1	37.1	40.5	9.3%
Total college enrollment (%):	5.58	4.73	4.65	4.79	-14.1%
In-person enrollment (%):	3.6	4.73	4.65	4.79	33.1%
Total value added (log points):	0.172	0.167	0.165	0.169	-1.4%
Total expenditure (US$):	5011	5132	5130	5256	4.9%
Value added per dollar spent (log points/US$):	0.032	0.03	0.029	0.03	-8.1%
Average price of in-person degrees (US$):	6147	6147	6559	6322	2.9%
Total variable profits (US$):	1553	1324	1364	1388	-10.6%
Consumer surplus (US$):	5038	4276	4197	4331	-14%

Notes: This table reports values of key outcomes for each counterfactual scenario from Table 5. Values are averages across markets, weighted by market size. Outcomes include: the number of online and in-person degrees, total college enrollment and in-person enrollment (% students), total value added (log points), total expenditure (US$ per enrolled student at BL), value added per dollar spent (log points per US$), the average price of in-person degrees (US$ per year), total variable profits (US$ per enrolled student at BL), and consumer surplus (US$ per enrolled student at BL). Detailed formulas for constructing the last three outcome measures are provided in Online Appendix E.4. Columns (1)–(4) report outcomes under each counterfactual from Table 5, averaged across 1,000 simulations. Column (5) reports the average percent change between the *Equilibrium* and *Baseline* counterfactuals, given by $\Delta_{\text{EQ/BL}} = \frac{EQ-BL}{BL}$.

by 2.5%. Revealed-preferences consumer surplus falls by 15.1% due to the reduced number of available options.

Under the PR scenario, where institutions can respond to the absence of online degrees by adjusting the prices of in-person programs, tuition increases by 6.7%. The higher prices discourage enrollment in higher-quality degrees, leading to a 3.9% decline in total value added.

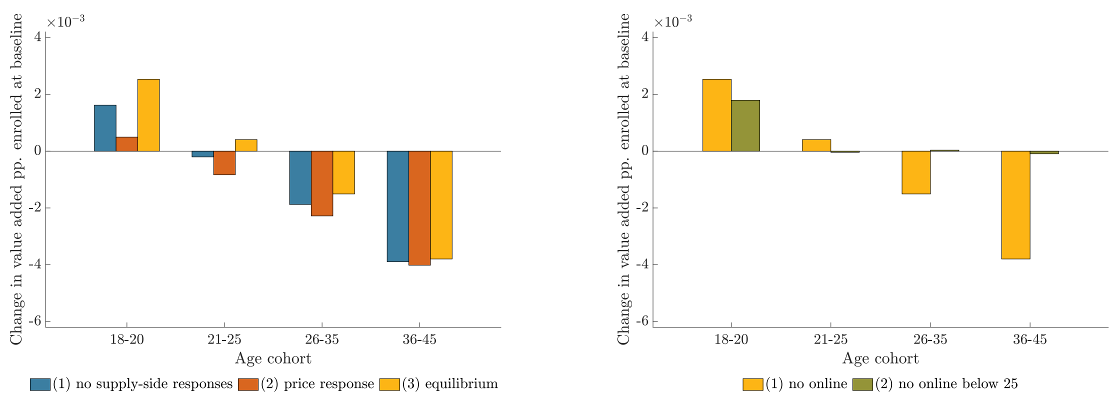
Under the EQ scenario, in which institutions can adjust their degree offerings, the supply of in-person options expands by 9.3%, from an average of 37.1 to 40.5 degrees per market. Moving from the BL to EQ, total enrollment falls by 14.1%, from 5.6% to 4.8%, while in-person enrollment rises by 33.1%, from 3.6% to 4.8%. Total value added declines slightly, by 1.4%, to 0.169 log points. Total expenditure increases by 4.9%, reducing value added per dollar spent by 8.1%. The average tuition price for in-person degrees rises by 2.9% relative to BL. Nevertheless, higher markups are insufficient to offset the profit losses resulting from the absence of online offerings, leading to an average 10.6% decline in total variable profits. Revealed-preferences consumer surplus decreases by 14%.

These results highlight the benefits and trade-offs associated with the expansion of online education. The availability of online options increases college enrollment, lowers total expenditure, and raises consumer surplus, measured as the area under the demand curve. While online expansion diverts some students from high value-added in-person degrees, the resulting increase in total enrollment generates a modest rise in total value added. Finally, 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 potential declines in total value added.

5.2. *Unpacking value added changes by age cohort*

The aggregate changes in total value added discussed above mask important heterogeneity across age cohorts. Older individuals, who might not attend college in the absence of online education, are the primary beneficiaries of its expansion. By contrast, younger individuals are more often diverted from higher-quality, in-person programs. Such diversion can be particularly costly if online education induces the exit, or deters the entry of in-person degrees that young individuals would otherwise prefer, even when online options are available.

Figure 6(a) illustrates the heterogeneous effects by age cohort, showing changes in total value added relative to BL. When online education is restricted and supply-side adjustments are not allowed, the 18–20 cohort experiences an increase in value added, whereas the 26–45 cohorts experience significant decline. Allowing institutions to adjust prices reduces value added across all cohorts. Under the Equilibrium counterfactual, in which institutions can adjust their degree offerings, new in-person options enable the 18–25 cohorts to achieve net value-added gains.



(a) Decomposition of value-added changes by age cohort

(b) Impact of targeted online education ban on ages 18–25

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

Notes: This figure presents changes in total value added by age cohort under alternative counterfactuals. Panel (a) reports changes in value added relative to *Baseline* (BL) for all counterfactuals presented in Table 5. Panel (b) reports the changes in value added relative to BL for the *Equilibrium* counterfactual (EQ) and for an additional counterfactual restricting access to online education for students aged 18–25. In both panels, total value is divided by the number of students enrolled under BL.

Online Appendix A, Figure A.2, reports heterogenous results for total enrollment, value added per dollar spent in higher education, and revealed-preferences consumer surplus. The expansion of online education benefits all age cohorts across these three measures, with large gains for individuals aged 25 and older and small gains for individuals aged 18–20.

5.3. *Targeted policies*

We consider 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 restricting access to online education for students aged 18–25.[32] By directing younger cohorts toward in-person programs, the policy maintains sufficient profitability for in-person degrees, allowing them to coexist with online options. Consequently, younger students can pursue high-value-added in-person degrees, while older students continue to benefit from expanded online access. Our main results, shown in Figure 6(b), indicate that, relative to BL, this targeted policy increases value added among young cohorts without negatively affecting the older cohorts, increasing total value added by 1%. In Online Appendix A, Figure A.2, we show that restricting online education in this targeted manner has a smaller negative effect on other outcomes of interest.

6. Conclusion and discussion

This paper studies the equilibrium effects of the rapid growth of online college education in Brazil. We highlight two key findings. First, the growth of online degrees has a dual impact: it broadens access to higher education for new students while diverting potential students away from higher-quality in-person programs. Second, increased availability of online degrees intensifies competition, lowering prices for in-person programs and deterring new entry.

Using an equilibrium model of demand and supply, we quantify the impact of online education on total value added and document substantial heterogeneity across age cohorts. A ban on online degrees would benefit younger cohorts by preserving access to higher-quality, in-person programs, while older cohorts, who gain 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 for younger students without reducing gains for older ones relative to an unrestricted baseline.

These results illustrate 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 programs become less available. Policymakers could mitigate these effects by restricting online access for groups that would gain more from traditional degrees, thereby sustaining demand for incumbent high-quality institutions and preventing their exit.

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-

[32] Restricting online access by age could induce strategic delays in enrollment, which our model does not capture. We abstract from these dynamics to illustrate the potential benefits of targeted access restrictions.

invariant; however a surge in online degrees could saturate the labor market, reducing returns to college education. Modeling higher education as a market equilibrium that interacts with labor market outcomes offers a promising direction for future research. Second, technological advancements may further enhance the cost-effectiveness and quality of online education, potentially altering some conclusions of this study. Finally, non-pecuniary benefits—such as trust, networking, and social skills—are more effectively developed in in-person settings. These social benefits may be difficult to replicate online, suggesting that traditional college experiences offer unique non-monetary advantages, particularly for younger students.

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*.

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

Anatolyev, S. (2013). Instrumental variables estimation and inference in the presence of many exogenous regressors. *The Econometrics Journal 16*(1), 27–72.

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.

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.

Barahona, N., C. Dobbin, and S. Otero (2025). Equilibrium price responses to targeted student financial aid. Technical report, National Bureau of Economic Research.

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.

Bhattacharya, V., G. Illanes, and M. Padi (2025). Fiduciary duty and the market for financial advice. Technical Report 4.

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.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–2781.

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

Conlon, C. and J. Gortmaker (2025). Incorporating micro data into differentiated products demand estimation with pyblp. *Journal of Econometrics*, 105926.

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.

Eisenhauer, P., J. J. Heckman, and E. Vytlacil (2015). The generalized roy model and the cost-benefit analysis of social programs. *Journal of Political Economy 123*(2), 413–443.

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).

Garcia, C. P. and P. F. de Azevedo (2019). Should competition authorities care about conglomerate mergers? *International Journal of Industrial Organization 66*, 78–118.

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.

Hastings, J., C. A. Neilson, and S. D. Zimmerman (2015). The effects of earnings disclosure on college enrollment decisions. Technical report, National Bureau of Economic Research.

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

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics 3*(2), 211–255.

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.

Kolesár, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens (2015). Identification and inference with many invalid instruments. *Journal of Business & Economic Statistics 33*(4), 474–484.

Larroucau, T., I. Rios, A. Fabre, and C. Neilson (2024). College application mistakes and the design of information policies at scale. *Working paper*.

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*.

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

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.

Teodorovicz, T. (2025). Corporate ownership and workforce reconfiguration: Evidence from private higher education. Technical report.

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 Sebastián Otero

November 21, 2025

Appendix A: Additional Figures and Tables

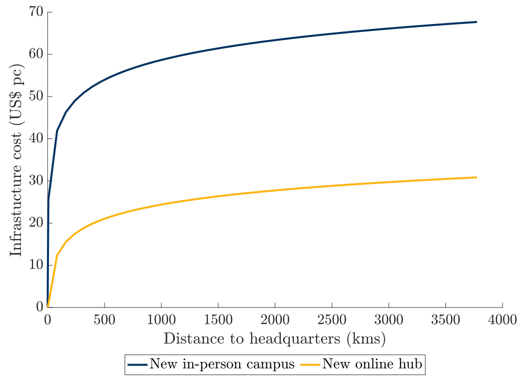
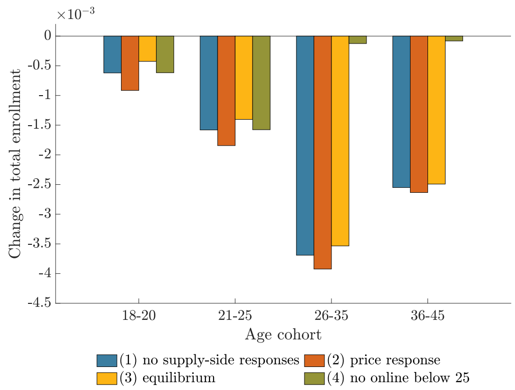
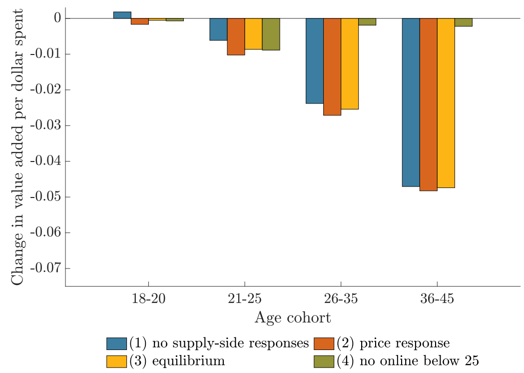


Figure A.1: Infrastructure cost as a function of distance

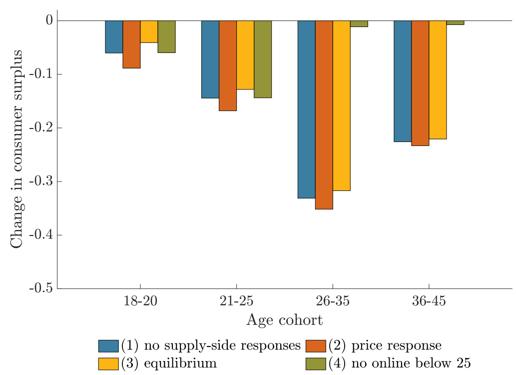
Notes: This figure plots the cost of building a new campus or hub as a function of the distance between the firm's headquarters and a given region. Costs are defined as $\chi_0^k + g(d_{fr})'\chi^k$ for $k \in \{\text{campus, hub}\}$.



(a) Total enrollment



(b) Value added per dollar



(c) Consumer surplus

Figure A.2: Counterfactual changes in various outcome by age cohort

Notes: This figure presents changes across several outcomes by age cohort under different counterfactual scenarios. We report changes in outcomes relative to *Baseline* (BL) for all counterfactuals presented in Table 5, as well as the hypothetical counterfactual discussed in Section 5.3. Panel (a) displays the changes in total enrollment. Panel (b) shows changes in value added per US$1000 of higher education spending. Panel (c) presents changes in total consumer surplus measured as the area under the demand curve.

Table A.1: Degree programs in each field of study

Field of study	**Degree programs**
Arts, Humanities, and Social Sciences	Audiovisual Production, Communication and Journalism, Design, Fine Arts, International Relations, Library Science, Literature and Translation, Music, Photography, Political Science and Sociology, Theater, Theology
Business	Accounting, Administration, Business Management, Commercial Management, Economics, Entrepreneurship, Financial Management, Human Resources Management, Information Technology Management, International Business, Logistics, Marketing, Public Administration, Public Relations, Real Estate Management
Education	Biological Sciences, English Language and Literature, Geography, History, Mathematics, Pedagogy, Philosophy, Physical Education, Portuguese and English Language Studies, Visual Arts, Chemistry
Engineering	Agribusiness, Agronomy, Architecture and Urbanism, Biomedical Engineering, Building Construction, Civil Engineering, Computer Engineering, Control and Automation Engineering, Electrical Engineering, Environmental Engineering, Industrial Mechatronics, Mechanical Engineering, Chemical Engineering, Telecommunications Engineering, Aircraft Maintenance, Chemical Processes
Health Sciences	Biochemistry and Pharmacy, Nursing, Nutrition, Occupational Therapy, Optics and Optometry, Pharmacy, Physiotherapy, Radiology, Speech Therapy, Veterinary Medicine
Law	Law
Math, Computer Science, and Natural Sciences	Applied Mathematics, Biomedicine, Biotechnology, Computer Science, Data Management, Digital Games, Geology, Information Security, Information Systems, Internet Systems, Networks and Telecommunications, Statistics, Industrial Chemistry, Oceanography, Systems Analysis and Development
Medicine	Medicine, Dentistry
Psychology	Psychology
Services	Environmental Management, Aesthetics and Cosmetics, Event Management, Gastronomy, Hospitality and Tourism, Occupational Safety, Public Safety, Social Work

Notes: This table summarizes the aggregation of degree programs into broader areas of study, classified according to the International Standard Classification of Education (ISCED) codes.

Table A.2: Relationship between distance and the probability of having entered a market with online degrees by 2010

	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 reports estimates from 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} is the distance between the headquarters of institution f and region r. The vector $g(d_{fr}) = [\log(1 + d_{fr}), H_{fr}]'$, where $H_{fr} \in \{0, 1\}$ indicates whether the headquarters of f are located in region r ($d_{fr} = 0$), to account for cases with zero distance. The regression is estimated using all 37 firms offering at least one online degree in 2010.

Table A.3: 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					
Δ in online degrees		0.352	-0.194	-0.171	-0.008
		(0.033)	(0.027)	(0.028)	(0.002)
Panel B: SSIV					
shift-share instrument	1.733				
	(0.125)				
Δ in online degrees		0.389	-0.187	-0.191	-0.014
		(0.032)	(0.034)	(0.038)	(0.003)
Panel C: Mean of the dependant variable in 2010 and 2019					
2010	3.10	0.80	3.41	11.43	1.59
2019	9.91	3.06	3.24	12.44	1.61
Obs.	1,100	1,100	1,100	1,100	893

Notes: This table presents the results from estimating an extended version of Equation (2), given by $\Delta y_{ra} = \phi \Delta N^o_{ra} + \beta \Delta X_r + \varepsilon_{ra}$, where Δy_{ra} and ΔN^o_{ra} are defined as in Equation (2), and ΔX_r are additional controls, which include the change in log internet penetration and regional log GDP per capita. Panel A presents OLS estimates. Panel B reports 2SLS estimates using the shift-share instrument described in Section 3.2. Panel C reports the mean of each dependent variable (in levels) in 2010 and 2019. Column (1) shows the 2SLS first-stage regression of the change in the number of online degrees on the shift-share instrument. Columns (2)–(5) report coefficients for the number of online students relative to market size, the number of in-person students relative to market size, the number of in-person degrees, and the log price of in-person degrees. Columns (1)–(4) use the full set of region-field pairs; Column (5) uses the 893 region-field pairs with at least one in-person degree in both 2010 and 2019. Standard errors clustered at the region-field level are shown in parentheses.

Table A.4: Demand parameters

Panel A: BLP parameters									
Price coef.		Price RC		Online RC		In-person RC		Nest param.	
$\bar{\alpha}$	-1.57	σ_α	0.003	σ^o	0.002	σ^p	0.001	ρ	0.743
	(0.345)		(4.087)		(20.681)		(40.503)		(0.041)
Dem. shifter		Dem. shifter		Age 21–25		Age 26–35		Age 36–45	
ψ^0	0.11	ψ^o	-0.07	μ_2^p	-0.932	μ_3^p	-1.536	μ_4^p	-2.155
	(0.024)		(0.022)		(3.733)		(7.287)		(19.519)
Age 21–25		Age 26–35		Age 36–45					
μ_2^o	-0.542	μ_3^o	-0.761	μ_4^o	-1.114				
	(1.664)		(5.978)		(2.471)				
Panel B: Mixed-effects model parameters									
Degree r.e.		Region-area r.e.		Year-area r.e.		Year-online r.e.		Demand shock	
σ_j	0.456	σ_{ra}	0.642	σ_{ta}	0.811	σ_{to}	0.664	σ_ξ	0.421
	(0.006)		(0.015)		(0.064)		(0.111)		(0.001)
Hours f.e.		STEM f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
δ_1	0.244	δ_2	0.067	δ_3	0.076	δ_4	0.012	δ_5	0.077
	(0.013)		(0.046)		(0.012)		(0.0004)		(0.041)
Const. f.e.									
δ_0	-12.712								
	(0.344)								

Notes: This table reports estimated parameters from the demand model described in Section 4.2. Panel A presents parameters estimated in the first step, following Berry et al. (1995). Panel B presents parameters estimated in the second step using a mixed-effects Bayesian hierarchical model to recover the mean utility components of δ_{jrt}. Standard errors are reported in parentheses. Random and fixed effects are indicated for the corresponding variables.

Table A.5: Median diversion ratios by age cohort

	$t = 2010$		$t = 2019$	
	In-person	Online	In-person	Online
Median diversion ratios for cohort 18–20:				
To in-person:	0.73	0.72	0.69	0.54
To online:	0.01	0.04	0.05	0.24
To outside option:	0.25	0.23	0.23	0.23
Median diversion ratios for cohort 21–25:				
To in-person:	0.68	0.57	0.57	0.25
To online:	0.02	0.15	0.15	0.50
To outside option:	0.26	0.25	0.25	0.24
Median diversion ratios for cohort 26–35:				
To in-person:	0.60	0.31	0.34	0.08
To online:	0.07	0.38	0.38	0.66
To outside option:	0.26	0.26	0.25	0.25
Median diversion ratios for cohort 36–45:				
To in-person:	0.51	0.15	0.18	0.03
To online:	0.16	0.53	0.56	0.70
To outside option:	0.26	0.27	0.25	0.26

Notes: This table reports median diversion ratios across products and markets by age cohort for in-person and online degrees in 2010 and 2019. Diversion ratios measure the fraction of students among those who decide to leave degree j in response to an increase in tuition that would switch to 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 $\mathcal{K}$ is 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 for age cohort 36–45 is 0.16, 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 $b = \{36\text{–}45\}$. Since the reported diversion ratios are based on medians, they are not constrained to sum to one.

Table A.6: Marginal cost parameters

Panel A: Main parameters									
Demand instrument		log-distance		same-region					
γ_z	0.124	γ_{g_1}	0.156	γ_{g_2}	0.731				
	(0.013)		(0.057)		(0.371)				
Panel B: Mixed-effects model parameters									
Degree r.e.		Region-area r.e.		Year-area r.e.		Year-online r.e.		Cost shock	
ς_j	0.506	ς_{ra}	0.289	ς_{ta}	0.308	ς_{to}	1.014	ς_ω	0.543
	(0.007)		(0.009)		(0.034)		(0.164)		(0.001)
Hours f.e.		STEM f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
γ_1	0.095	γ_2	-0.338	γ_3	0.031	γ_4	0.007	γ_5	0.238
	(0.014)		(0.049)		(0.015)		(0.0004)		(0.046)
Const. f.e.									
γ_0	-7.29								
	(0.406)								

Notes: This table reports the estimated parameters from the marginal cost model described in Section 4.3. Panel A presents parameters estimated in the first step via OLS. Panel B presents parameters estimated in the second step from a mixed-effects Bayesian hierarchical model to recover the marginal cost components of γ_{jrt}. Standard errors are reported in parentheses. Random and fixed effects are indicated for the corresponding variables.

Table A.7: Fixed-cost parameters

Panel A: Infrastructure fixed costs											
New campus						New hub					
χ_0^c	1.175	χ_H^c	-1.175	χ_d^c	0.676	χ_0^h	-0.907	χ_H^h	0.907	χ_d^h	0.483
	(1.504)		(1.504)		(0.221)		(0.84)		(0.837)		(0.115)
Panel B: Degree fixed costs											
κ_0	0.93	$\kappa_{\iota 1}^{old}$	0	κ_{o1}^{old}	0	$\kappa_{\iota 2}^{old}$	0	κ_{o2}^{old}	0	$\kappa_{\iota 3}^{old}$	0
	(0.141)		(0)		(0)		(0)		(0)		(0)
κ_{o3}^{old}	0	$\kappa_{\iota 4}^{old}$	0	κ_{o4}^{old}	0	$\kappa_{\iota 5}^{old}$	0	κ_{o5}^{old}	0	$\kappa_{\iota 6}^{old}$	0
	(0)		(0)		(0)		(0)		(0)		(0)
$\kappa_{\iota 7}^{old}$	0	κ_{o7}^{old}	0	$\kappa_{\iota 8}^{old}$	0	$\kappa_{\iota 9}^{old}$	0	$\kappa_{\iota 10}^{old}$	0	κ_{o10}^{old}	0
	(0)		(0)		(0)		(0)		(0)		(0)
$\kappa_{\iota 1}^{new}$	0.676	κ_{o1}^{new}	0	$\kappa_{\iota 2}^{new}$	0.738	κ_{o2}^{new}	0	$\kappa_{\iota 3}^{new}$	0.565	κ_{o3}^{new}	0
	(0.184)		(0)		(0.267)		(0)		(0.197)		(0)
$\kappa_{\iota 4}^{new}$	0	κ_{o4}^{new}	0	$\kappa_{\iota 5}^{new}$	0	κ_{o5}^{new}	0	$\kappa_{\iota 6}^{new}$	0.139	$\kappa_{\iota 7}^{new}$	0.436
	(0)		(0)		(0.016)		(0)		(0.162)		(0.133)
κ_{o7}^{new}	0	$\kappa_{\iota 8}^{new}$	0	$\kappa_{\iota 9}^{new}$	0	$\kappa_{\iota 10}^{new}$	0	κ_{o10}^{new}	0	σ_ε	0.01
	(0)		(0.005)		(0)		(0.016)		(0)		(0.001)

Notes: This table reports estimated fixed-cost parameters from the entry model described in Section 4.4. Panel A shows infrastructure fixed-costs parameters from Equation (15). Panel B shows degrees fixed-costs parameters from Equation (14). Standard errors are computed via bootstrap with 100 repetitions, clustering at the market level.

Appendix B: Tuition Fees

To construct program-level prices, we combine four data sources. The first two sources come from Brazil's government fellowship and loan programs, PROUNI and FIES. We use administrative records from the National Education Fund (FNDE), which track government payments to students in these programs, enabling estimation of 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. Together, the data sources contain information for 59.6% of all degree program-years, covering 81.9% of total enrollment, with at least one data point for 95.5% of all programs, covering 95.5% of total enrollment.

Using these datasets, we construct prices as follows: We use information from all sources to recover an average posted price by program, campus or hub, and year, by regressing log-prices on program-region-year and source fixed effects. Controlling for source fixed effects account for persistent differences across data sources and enables us to recover a program-region-year price that are consistent across sources. When information for a given year is missing, we impute values by regressing the predicted price on program-region and year fixed effects, and summing the relevant coefficients. In cases where both year and region information are unavailable, we regress the predicted price on program and year fixed effects, using the sum of these coefficients to impute the missing values. Using this approach, we are able to recover year-specific tuition prices for approximately 95.5% of degree program-years, covering 98.5% of total enrollment.

Appendix C: Value-Added Estimation

C.1. *Value-added model*

Our value-added model captures the effect of enrolling in different degree programs on students' future earnings. We consider all ENEM (university entrance exam) takers from years $\tau \in \{2010, 2011, 2012, 2013\}$ 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 in 2023—the latest available year—using administrative wage records (RAIS). The model accounts for selection on both observable and unobservable characteristics and allows for heterogeneous returns across age cohorts $b \in \{18\text{–}20, 21\text{–}25, 26\text{–}35, 36\text{–}45\}$. All values are normalized relative to the outside option of not attending college, which may vary by region to capture local labor market conditions.

Let Y_{ij} denote the potential outcome for individual i enrolling in degree program j. The potential outcome equation is given by

$$Y_{ij} = X_i'\beta_{b(i)} + Z_j'\gamma_{b(i)} + \delta_{r(i)} + \delta_{j|r(i)} + \varepsilon_{ij}, \tag{C.1}$$

where Y_{ij} is the potential log-wage in 2023; X_i includes individual characteristics: gender, flexible age dummies, race dummies, ENEM scores in five subjects (math, language, social sciences, natural sciences, and an essay), ENEM year dummies, household income dummies at the time of taking ENEM, a dummy for attending a public high school, a dummy for graduating from high school that year, the log-GDP per capita of the municipality of residence, and a constant; Z_j includes degree-program characteristics: online or in-person status, program age, average incoming students' score, average first-job salary of graduates, program length (hours), a STEM dummy, and dummies for program field of study; $r(i)$ denotes the student's region; and $\delta_{r(i)}$ and $\delta_{j|r(i)}$ are region and program-region-specific intercepts. The $\beta_{b(i)}$ parameters vary flexibly by age cohort. For online/in-person status, $\gamma_{b(i)}$ varies by age cohort; other program characteristics are constant across age cohorts.

Let $S_i \in \{0, 1, 2, \ldots, J\}$ denote the degree program in which student i enrolls, with $S_i = 0$ indicating non-enrollment for at least the next nine years. We exclude students who: (i) enroll in a federal degree program, (ii) enroll outside their region of residence, or (iii) initially do not enroll but later enter a program in subsequent years.

Our empirical analysis uses two strategies to estimate the parameters of Equation (C.1). The first strategy relies on a standard selection-on-observables assumption:

$$\mathbb{E}[\varepsilon_{ij} \mid X_i, S_i] = 0. \tag{C.2}$$

Assumption (C.2) requires that the controls in X_i absorb all sources of selection bias when comparing outcomes across degree programs and the outside option. In other words, conditional on these characteristics, potential outcomes are mean-independent of program enrollment S_i.

The second and preferred strategy leverages information on individuals' program choices to relax the selection-on-observables assumption and allow for match effects in unobserved preferences. Specifically, we estimate a two-step control function model using the distance between individuals' home municipalities and program locations as an excluded instrument, capturing unobserved individual tastes for each degree program.

In the first step, we summarize individual preferences by fitting random-utility models with parameters that vary by age cohort b and ENEM year τ. Individual i's utility from enrolling in program j is:

$$u_{ij} = \delta_{b(i)\tau(i)j} - \pi_{b(i)\tau(i)z(j)} \log(1 + d_{ij}) + \eta_{ij}, \tag{C.3}$$

where d_{ij} is the distance between the student's home municipality and program location. The parameter $\delta_{b\tau j}$ represents the mean utility of program j for students in age cohort b and ENEM year τ, while $\pi_{b\tau z}$ is a distance parameter (or "cost") that varies by age cohort, ENEM year, and program delivery mode $z \in \{\text{in-person}, \text{online}\}$. Distance serves as a preference shifter excluded from the outcome equation, facilitating identification of the selection parameters. Unobserved tastes η_{ij} are modeled as independent type I extreme value random variables, conditional on X_i and the full vector of distances $d_i = (d_{i1}, ..., d_{iJ})'$. We estimate this preference model by maximum likelihood and use the results to construct control functions for the value-added model.

In the second step, we parameterize the relationship between unobserved tastes and potential outcomes as:

$$\mathbb{E}[\varepsilon_{ij} \mid X_i, \eta_i, S_i] = \sum_{k=1}^{J} \phi_{b(i)a(k)z(k)} (\eta_{ik} - \mu_\eta) + \psi_{b(i)} (\eta_{ij} - \mu_\eta) \mathbb{1}\{S_i = j\}, \tag{C.4}$$

where ϕ_{baz} are field-of-study–mode-of-delivery–specific coefficients varying by age cohort, and $\mu_\eta = \mathbb{E}[\eta_{ij}]$ is Euler's constant. Functional-form assumptions of this type are common in multinomial selection models with many alternatives, where nonparametric identification requirements are otherwise very stringent (Abdulkadiroğlu et al., 2020). The coefficients ϕ_{baz} capture the effect of individuals' preferences for field a and delivery mode z within age cohort b that is common across all potential outcomes. For example, students with strong preferences for certain fields (e.g., medicine) may systematically differ in ability, influencing long-run earnings regardless of where they enroll. The parameter ψ_b captures an additional match effect of program-specific preferences on the potential outcome of enrolling in that specific program. Together, these parameters allow for rich heterogeneity linking preferences to unobserved match effects in outcomes, permitting both "selection in levels" (through ϕ_{baz}) and "selection in gains" (through ψ_b).

Equation (C.4) implies that observed outcomes can be written as:

$$Y_i = X_i'\beta_{b(i)} + Z_{j(i)}'\gamma_{b(i)} + \delta_{r(i)} + \delta_{j(i)|r(i)} + \sum_{k=1}^{J} \phi_{b(i)a(k)z(k)}\lambda_{ik} + \psi_{b(i)}\lambda_{ij(i)} + \tilde{\varepsilon}_i, \tag{C.5}$$

where $\lambda_{ik} = \mathbb{E}[\eta_{ik} - \mu_\eta \mid X_i, d_i, S_i]$ is the mean logit taste for program k conditional on covariates, distances, and the individual's choice, and $j(i) = \sum_j j \cdot \mathbb{1}\{S_i = j\}$ denotes the degree chosen by individual i. Identification of the selection parameters ϕ_{baz} and ψ_b comes from variation in distance between students' homes and degree programs. Intuitively, if students willing to travel long distances to attend in-person medicine programs perform better than expected—given their observed characteristics—across all degree programs, we infer $\phi_{baz} > 0$ for that field and mode. If these students perform better only at their preferred program but not elsewhere, we infer $\psi_b > 0$.

Many degree programs have few enrollees, which can lead to noisy value-added estimates. To address this, we fit a nested random-effects model and report empirical Bayes estimates that shrink imprecise degree effects toward the overall mean, conditional on Z_j. Specifically, we assume:

$$\delta_r \sim \mathcal{N}(0, \sigma_r^2), \quad \delta_{j|r} \sim \mathcal{N}(0, \sigma_{\delta_{j|r}}^2), \quad \tilde{\varepsilon}_i \sim \mathcal{N}(0, \sigma_{\tilde{\varepsilon}}^2),$$

all mutually independent and independent of (X_i, Z_j). We estimate the model via restricted maximum likelihood and recover a degree program's value added as:

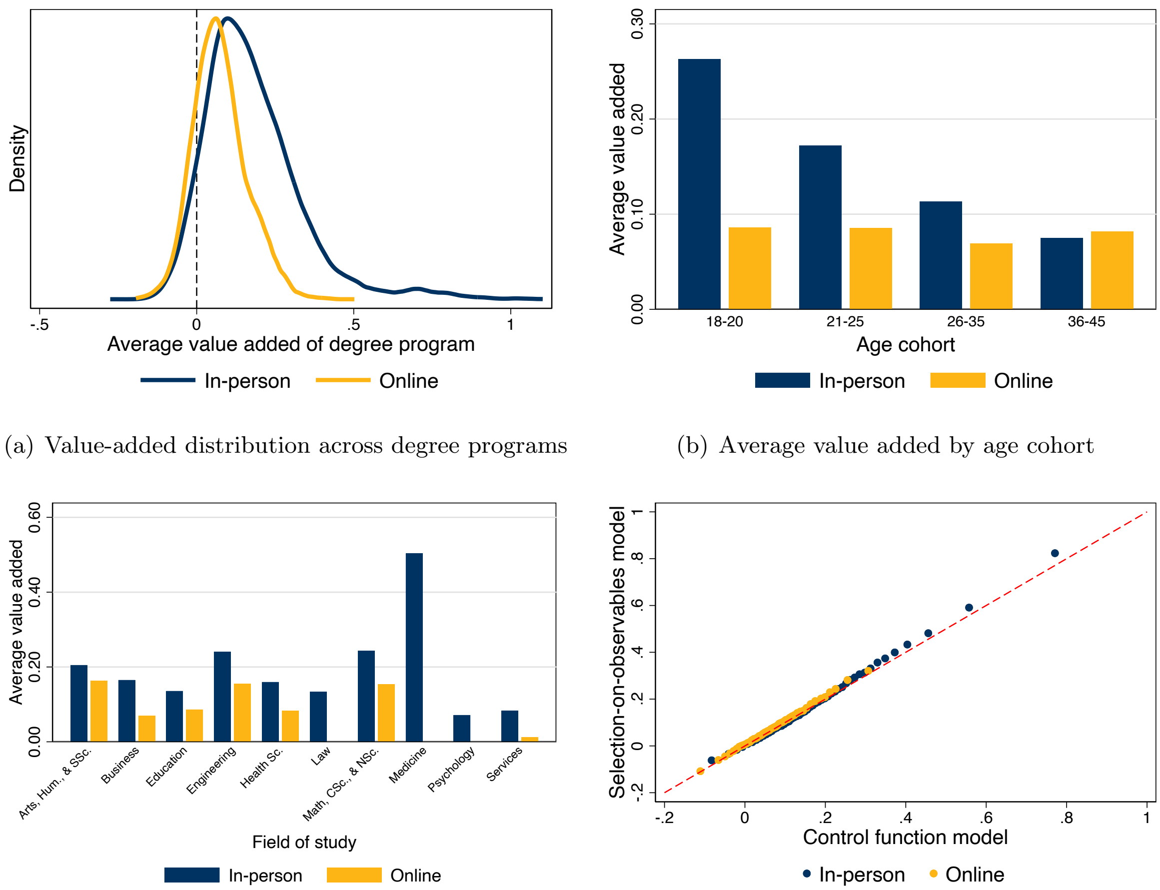
$$VA_{jr} = Z_j'\hat{\gamma}_b + \hat{\delta}_{j|r} - \hat{\delta}_{0|r},$$

where $\hat{\delta}_{j|r}$ and $\hat{\delta}_{0|r}$ are the posterior means of $\delta_{j|r}$ and $\delta_{0|r}$, respectively.

We are able to recover value-added estimates for 28% of program-region pairs in the data, representing 85% of total enrollment. Restricting to pairs that existed during 2011–2014, coverage rises to 82% of pairs and 98% of enrollment.

C.2. *Results*

Figure C.1 summarizes our value-added estimates. Panel (a) presents the distribution of average value added for in-person and online degree programs. We observe substantial overlap between the two types, although in-person programs exhibit more mass at higher and more positive values. Many of these high-value programs are in Medicine, which is not available in the online format. Panel (b) reports the average value added for in-person and online degrees by age cohort. Returns to in-person programs vary markedly across age cohorts, whereas returns to online programs are more stable. For the youngest cohort (18–20), enrollment in an in-person degree program yields substantially higher value than enrollment in an online program. These differences diminish for older cohorts and disappear entirely for the oldest cohort (36–45).[1] Panel (c) shows average value added for in-person and online degrees by field of study, with Medicine programs exhibiting the highest returns. Finally, Panel (d) compares estimates from: the selection-on-observables model and the control-function model. The selection-on-observables model produces slightly higher value-added estimates, particularly for programs with the highest value added.



(a) Value-added distribution across degree programs

(b) Average value added by age cohort

(c) Average value added by field of study

(d) Selection-on-observables vs control-function model

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

Notes: This figure summarizes program-level value-added estimates. Panel (a) shows the distribution of average value added across existing degree programs, weighting by enrollment shares and averaging over age groups. Panel (b) reports the average value added by age cohort for in-person and online degrees, averaging over programs using enrollment shares as weights. Panel (c) shows average value added by field of study for in-person and online degrees, averaging across programs and age cohorts using enrollment shares as weights. Panel (d) compares the value-added estimates from the selection-on-observables and control-function models. Each dot represents the average value added under the selection-on-observables model for programs grouped into bins based on their value added under the control-function model. Blue markers correspond to in-person programs; yellow markers correspond to online programs.

[1]Because these are averages, the differences could be driven by variation in the composition of degree choices across age cohorts. Results are quantitatively similar when we use common weights across groups.

C.3. *Degree-level aggregation and imputation for counterfactual analysis*

The estimator described above recovers value-added estimates for *degree programs*. For the model in Section 4 and the counterfactuals in Section 5, we work with data at the *degree* level. We calculate degree-level value added as the weighted average of the value added of all programs composing a given degree, using each program's enrollment share as weights. These calculations are performed separately by age cohort. Our counterfactual analysis requires estimates of demand, marginal cost, and value-added not only for existing degrees but also for degrees that could potentially enter the market. The main text explains how we recover demand and marginal cost parameters; here, we describe how we recover value added. The notation follows that of Section 4.

To obtain reliable estimates for degrees not observed in the data, we estimate a simple model using the program-level value-added measures described above. For each degree-region-age cohort combination, value added is given by:

$$\mathrm{VA}_{jrb} = o_j \nu_{ob} + x_j^{(2)} \nu_{xb} + \nu_{jb} + \nu_{rab} + \varsigma_{jrb} \tag{C.6}$$

where VA_{jrb} is the weighted average of the value added of all degree programs composing degree j in region r for age cohort b; o_j is an indicator for online degrees; $x_j^{(2)}$ are the degree-level characteristics from Equations (8) and (9) in the main text; ν_{jb} are degree-specific components of value added; ν_{rab} are field-of-study-by-region components; and ς_{jrb} is a degree-region idiosyncratic shock. As before, the outside option is normalized to zero within each region. The model is estimated separately for each age cohort.

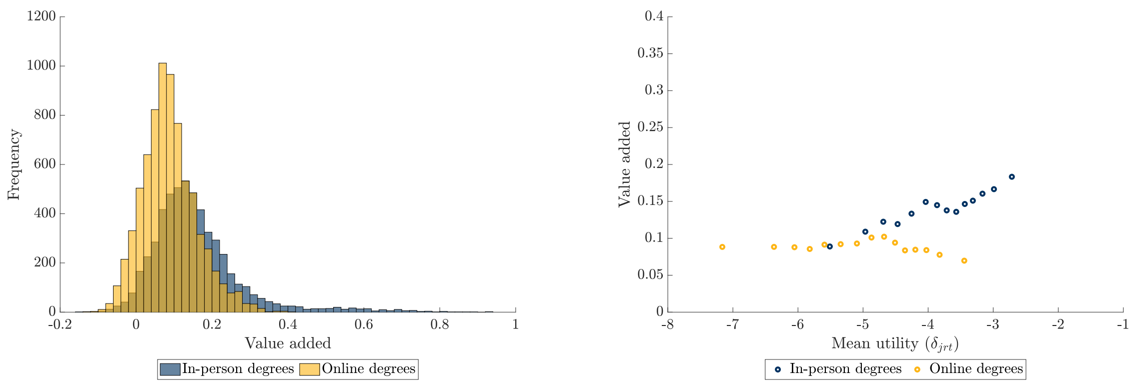
Following Section 4, we estimate the value-added model using a mixed-effects Bayesian hierarchical model, assuming $\nu_{jb} \sim \mathcal{N}(0, \vartheta_{jb}^2)$, $\nu_{rab} \sim \mathcal{N}(0, \vartheta_{rab}^2)$, and $\varsigma_{jrb} \sim \mathcal{N}(0, \vartheta_{\varsigma b}^2)$, with (ν_{ob}, ν_{xb}) treated as fixed parameters. The model is estimated by maximum likelihood. Posterior means of each component of value added are used to impute value added for all degree-market pairs, whether or not they are observed in the data.

Table C.1 reports the estimated parameters. Figure C.2, Panel (a), presents the distribution of value added for online and in-person degrees. In-person programs have an average value added of 0.158, compared with 0.088 for online programs. Figure C.2, Panel (b), shows that students' preferences for degrees are positively but weakly correlated with value added. The correlation is especially low for online degrees, which tend to be newer and for which students may have limited information on quality.

Table C.1: Value-added parameters

Panel A: Mixed-effects model parameters for cohort 18–20									
Degree r.e.		Region-area r.e.		VA shock		Const. f.e.		Online f.e.	
ϑ_{jb}	0.02	ϑ_{rab}	0.068	$\vartheta_{\varsigma b}$	0.058	ν_{0b}	-2.076	ν_{ob}	-0.117
	(0.002)		(0.002)		(0.001)		(0.039)		(0.003)
Hours f.e.		STEM f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
ν_{x_1b}	0.024	ν_{x_2b}	-0.007	ν_{x_3b}	-0.004	ν_{x_4b}	0.0013	ν_{x_5b}	0.191
	(0.002)		(0.005)		(0.003)		(5e-05)		(0.006)
Panel B: Mixed-effects model parameters for cohort 21–25									
Degree r.e.		Region-area r.e.		VA shock		Const. f.e.		Online f.e.	
ϑ_{jb}	0.02	ϑ_{rab}	0.07	$\vartheta_{\varsigma b}$	0.057	ν_{0b}	-2.173	ν_{ob}	-0.047
	(0.002)		(0.002)		(0.001)		(0.038)		(0.003)
Hours f.e.		STEM f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
ν_{x_1b}	0.027	ν_{x_2b}	-0.006	ν_{x_3b}	-0.003	ν_{x_4b}	0.0013	ν_{x_5b}	0.192
	(0.002)		(0.005)		(0.003)		(5e-05)		(0.006)
Panel C: Mixed-effects model parameters for cohort 26–35									
Degree r.e.		Region-area r.e.		VA shock		Const. f.e.		Online f.e.	
ϑ_{jb}	0.021	ϑ_{rab}	0.07	$\vartheta_{\varsigma b}$	0.057	ν_{0b}	-2.187	ν_{ob}	-0.022
	(0.002)		(0.002)		(0.001)		(0.038)		(0.003)
Hours f.e.		STEM f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
ν_{x_1b}	0.027	ν_{x_2b}	-0.004	ν_{x_3b}	-0.002	ν_{x_4b}	0.0013	ν_{x_5b}	0.191
	(0.002)		(0.005)		(0.003)		(5e-05)		(0.006)
Panel D: Mixed-effects model parameters for cohort 36–45									
Degree r.e.		Region-area r.e.		VA shock		Const. f.e.		Online f.e.	
ϑ_{jb}	0.022	ϑ_{rab}	0.069	$\vartheta_{\varsigma b}$	0.057	ν_{0b}	-2.089	ν_{ob}	0.016
	(0.002)		(0.002)		(0.001)		(0.04)		(0.004)
Hours f.e.		STEM f.e.		Degree age f.e.		Av. score f.e.		Av. wages f.e.	
ν_{x_1b}	0.026	ν_{x_2b}	-0.001	ν_{x_3b}	-0.001	ν_{x_4b}	0.0012	ν_{x_5b}	0.182
	(0.002)		(0.005)		(0.003)		(5e-05)		(0.006)

Notes: This table reports the estimated parameters from the mixed-effects Bayesian hierarchical model described in Section C.3. Standard errors are reported in parentheses. Random and fixed effects are indicated for the corresponding variables.



(a) Distribution of value added

(b) Relationship between preferences and value added

Figure C.2: Estimated parameters from the value-added model

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

Appendix D: Shift-share Tests

In this appendix, we examine and address common concerns regarding the identifying assumptions underlying the use of shift-share instruments. We conduct two sets of tests.

In Section D.1, we unpack the shift-share estimator and demonstrate its robustness to alternative estimators. Our instrument captures exposure to online expansion from several firms. It can be shown that our shift-share IV estimate is equivalent to a weighted average of the IV estimates that would arise from a series of IV regressions regressions using, one at a time, the exposure shares of each firm as a separate instrument. We begin by identifying the firms that contribute most to the estimator, and show that results are robust to estimate the model using each of these firms separately. We then further validate our results by showing robustness to several alternative estimators, as proposed by Goldsmith-Pinkham et al. (2020).[2]

In Section D.2, we test the exclusion restriction by examining whether the instrument is uncorrelated with unobserved differential trends in outcomes. We document that most online expansion occurs in the second half of our sample. Consequently, our instrument predicts larger increases in online degree during this period. Nonetheless, the IV estimator cannot reject the null hypothesis that the treatment effect is constant across both sample periods, lending confidence to our identifying assumptions.

D.1. *Rotemberg weights decomposition and alternative estimators*

Following Goldsmith-Pinkham et al. (2020), we decompose the shift-share estimator into a weighted sum of the just-identified IV estimators that use each firm's exposure share, $z_{fra} \equiv z_{fr} z_{fa} z_a$, as a separate instrument. The weights—commonly referred to as Rotemberg weights—indicate how much each firm's instrument contribute to the overall estimate. Intuitively, they also highlight how sensitive the overidentified estimate of ϕ is to potential misspecification (i.e., endogeneity) in any individual instrument, and thus which firms' identifying assumptions are most important to evaluate. Under the assumptions of homogeneous treatment effects, correct model specification, and valid exclusion restrictions, each just-identified estimator should converge to the same value. We show that the just-identified estimators corresponding to the firms with the largest Rotemberg weights produce estimates similar to the overall shift-share IV estimate. Finally, we show that our results are robust to using several alternative estimators.

Goldsmith-Pinkham et al. (2020) show that the estimator from Equation (2) in the main text,

$$\Delta y_{ra} \quad = \quad \phi \Delta N^{o}_{ra} + \varepsilon_{ra}, \tag{D.1}$$

can be written as $\hat{\phi} = \sum_f \hat{\alpha}_f \hat{\phi}_f$, where $\hat{\phi}_f$ is the just-identified IV estimator obtained by using z_{fra} as an instrument, and $\hat{\alpha}_f$ are the Rotemberg weights, normalized so that $\sum_f \hat{\alpha}_f = 1$.

[2] Under the null of constant effects, a researcher can consider alternative estimators which combine multiple instruments. A divergence between estimators can be interpreted as a failure of the exclusion restriction.

These weights indicate the contribution of each firm to the overall shift-share IV estimate. Consequently, if a particular instrument is misspecified, α_f measures how much that misspecification translates into bias in the estimator. For example, if α_f is small, then bias in the fth instrument has little impact on the overall estimate. Goldsmith-Pinkham et al. (2020) recommend paying particular attention to firms with large α_f, as violations of the exclusion restriction for these firms could meaningfully affect the IV estimate.

We compute the Rotemberg weights and report them in Table D.1. Panel A presents the sum, mean, and share of weights for firms with positive and negative weights, showing that only a small fraction of these weights are negative. We find a small share of negative weights in our setting.[3] Panel B reports the correlations between these weights ($\hat{\alpha}_f$), the shift component of the instrument (ΔN_f^o), and the variance of the share component of each firm's instrument ($\text{Var}(z_{fra})$). We find that the weights are strongly correlated with the shift component, indicating that firms with the highest expansion in online degrees contribute the most to the overall IV estimate.

In Panel C, Column (1), we present the Rotemberg weights for the nine firms with the largest values, which together account for 62.8% of the estimator's total positive weight. These are the nine firms with the largest number of online degrees in 2019, as shown in Figure 4(a) of the main text. Because $\hat{\alpha}_f$ measures sensitivity to misspecification, this implies that our IV estimator is most sensitive to deviations from the identifying assumptions related to the instruments of these nine firms. Our identification assumption requires that $\mathbb{E}[z_{fra}\varepsilon_{ra}] = 0$ for each firm. As discussed in Section 3, this condition holds if the distance between regions and a given firm's headquarters is uncorrelated with region-specific unobserved shocks, $\{\varepsilon_{ra}\}_a$, associated with the fields of study in which the firm specializes in in-person education. A potential threat to our identification assumption arises if firms specializing in certain fields are systematically located near regions where those fields are expected to grow faster. This concern would be particularly severe if all firms were located in the same region and specialized in the same field, because then differential growth around that region in that specific field could induce substantial bias in the estimator. For example, if all firms were located in São Paulo and specialized in Education, and nearby regions were experiencing faster growth in demand for Education programs, the IV estimator would be severely biased. Examining the Rotemberg weights allows us to assess the geographic dispersion and field specialization of the most influential firms, providing reassurance about the plausibility of the identifying assumptions.[4]

We find that the nine firms with the largest Rotemberg weights are spread across five states in Brazil: three in the South, five in the Southeast, and one in the North. This geographical dispersion suggests that the distance between each firm and the regions in our data is unlikely

[3] Under the assumption of constant treatment effects, the presence of negative weights does not pose a conceptual concern.

[4] Another potential concern is that firms may have deliberately located their headquarters in regions expected to experience faster growth in their fields of specialization. We view this as unlikely, as all firms established their headquarters well before the study period (the average founding year is 1994).

to be systematically correlated with the same unobserved shock. The firms also specialize in different fields of study. For example, in Education, the least specialized firm offered 5 percent of its degrees in this field in 2010, while the most specialized offered 14 percent. Similar patterns hold for other fields. The highest specialization share is in Business, in which one firm in the state of Santa Catarina offered 20 percent of its degrees in this field. Taken together, these facts suggest that the moment condition for each firm is unlikely to be affected by the same unobserved shock.

We next estimate Equation (D.1) using the shares z_{fra} of each of the nine firms as instruments separately, and report the resulting coefficients in columns (2)–(5) of Table D.1, Panel C. The coefficients are stable across all nine firms, despite being located in different regions and specializing in different fields, indicating that the results are unlikely to be driven by unobserved shocks specific to any city or field. In Figure D.1, we plot the coefficients from these regressions for all 93 firms, which exhibit similar patterns. Finally, in Table D.1, Panel D, we report the average estimated coefficient, $\hat{\phi}_f$, from separate regressions for firms with positive and negative weights, showing that the coefficients remain fairly stable even for firms with negative Rotemberg weights.

Under homogeneous treatment effects, Goldsmith-Pinkham et al. (2020) recommend that researchers consider alternative estimators that combine the moment conditions in potentially more efficient ways. In particular, they suggest estimating the model using all firms' shares as separate instruments (overidentified TSLS). Because the overidentified TSLS estimator is biased in finite samples, three alternative estimators with better properties under many instruments are recommended: the Modified Bias-corrected TSLS (MBTSLS) estimator from Anatolyev (2013) and Kolesár et al. (2015), the Limited Information Maximum Likelihood (LIML) estimator, and the HFUL estimator from Hausman et al. (2012). Comparing these estimates, along with the shift-share IV estimate, provides a useful diagnostic for misspecification: agreement across estimators strengthens confidence in the identifying assumptions (Goldsmith-Pinkham et al., 2020).[5] Table D.2 reports the estimated parameter of Equation (2), showing that all estimators perform similarly, with the exception of HFUL in Column (1).[6]

D.2. *Treatment effects under differential growth rates*

In applications where a sharp policy change in period t_0 triggers the shift, researchers can often test for pretrends. Our context does not allow for such a test, as the expansion of online degrees began prior to our sample period. However, as discussed in Section 2, government reforms introduced in 2016 accelerated the growth of the online higher education sector. We exploit this differential expansion before and after the reforms to validate our instrumental variable

[5] Differences in underlying assumptions means these estimators may yield different estimates. The LIML estimator, as discussed in Hausman et al. (2012), is inconsistent under heteroskedasticity and many instruments. The HFUL estimator is consistent under both heteroskedasticity and many instrument asymptotics. The properties of the MBTSLS estimator under heteroskedasticity have not yet been established in the literature.

[6] The HFUL estimator also performs differently in the examples from Goldsmith-Pinkham et al. (2020).

Table D.1: Rotemberg weights and firm-level IV estimates

Panel A: Rotemberg weights for firms with positive and negative weights.

	Sum	Mean	Share
Negative	-0.051	-0.005	0.046
Positive	1.051	0.013	0.954

Panel B: Correlations among firm-level variables.

	$\hat{\alpha}_f$	ΔN^o_f	Var(z_{fra})
$\hat{\alpha}_f$	1		
ΔN^o_f	0.954	1	
Var(z_{fra})	-0.072	-0.079	1

Panel C: Weights and IV estimates, ϕ_f, for firms with the largest Rotemberg weights.

	$\hat{\alpha}_f$ (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)
Firm 1	0.135	0.282	-0.170	-0.167	-0.017
Firm 2	0.087	0.343	-0.192	-0.154	-0.017
Firm 3	0.074	0.298	-0.158	-0.198	-0.010
Firm 4	0.072	0.342	-0.221	-0.183	-0.007
Firm 5	0.067	0.440	-0.261	-0.210	-0.008
Firm 6	0.064	0.340	-0.179	-0.174	-0.018
Firm 7	0.063	0.330	-0.165	-0.202	-0.011
Firm 8	0.052	0.313	-0.199	-0.162	-0.017
Firm 9	0.044	0.610	-0.189	-0.234	-0.014

Panel D: Average estimated ϕ_f using firms with positive and negative weights.

	(2)	(3)	(4)	(5)
Negative	0.368	-0.268	-0.182	-0.005
Positive	0.430	-0.198	-0.192	-0.012

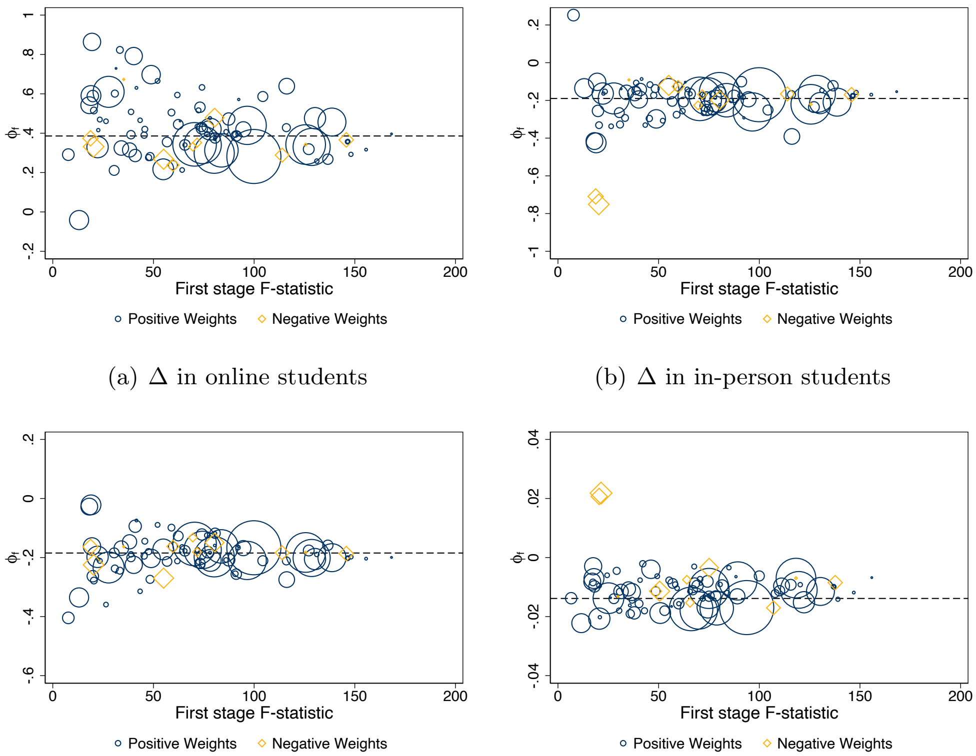
Notes: Panel A reports the sum, mean, and share of Rotemberg weights for firms with positive and negative weights. Panel B shows correlations among the Rotemberg weights, $\hat{\alpha}_f$, the change in the number of online degrees, ΔN^o_f, and the variance of the shares, z_{fra}, within each firm. Panel C, Column (1), reports the Rotemberg weights, $\hat{\alpha}_f$, for the nine firms with the largest weights. Columns (2)–(5) report the estimated parameter, $\hat{\phi}_f$, from Equation (2), by instrumenting with each firm's share, z_{fra}, separately for each of the outcomes of interest. Panel D, columns (2)–(5) show the average estimated coefficients, $\hat{\phi}_k$, from separate regressions for firms with positive and negative Rotemberg weights.

approach. Specifically, we estimate the following variation of Equation (2),

$$\Delta_t y_{ra} \quad = \quad \phi_t \Delta_t N^o_{ra} + \varepsilon^t_{ra}, \tag{D.2}$$

where Δ_t, for $t \in \{2010\text{–}2014, 2015\text{–}2019\}$, denotes changes over the periods 2010–2014, and 2015–2019, respectively.[7]

[7] We divide the sample into these two periods for two reasons. First, government reforms introduced in 2016 facilitated the expansion of new hubs beginning in 2015 (see Figure 1(b)). Second, this partition yields two periods of equal length, which facilitates comparison.



(a) Δ in online students

(b) Δ in in-person students

(c) Δ in in-person degrees

(d) Δ in log-price of in-person degrees

Figure D.1: Estimated coefficients ϕ_f using shares from each firm as separate instruments

Notes: This figure reports the estimated parameter, $\hat{\phi}_f$, from Equation (2), obtained by instrumenting with the firm-level shares, z_{fra}, separately for each outcome variable. The horizontal axis shows the first-stage F-statistic of the instrument, and the vertical axis shows the estimated parameter value. Bubble size corresponds to the value of the Rotemberg weight. Blue circles indicate estimates coming from firms with positive weights, and yellow diamonds indicate estimates coming from firms with negative weights. The dashed line shows the shift-share IV estimate from the main text (Table 2, Panel B).

Between 2010 and 2014, the expansion of online degrees was slower than during 2015 to 2019. Consequently, we should expect our instrument to predict a smaller increase in the number of online degrees in the earlier period—that is, a smaller first-stage coefficient. What is crucial, however, is how the outcomes respond. Under the exclusion restriction, any correlation between the instrument and the outcomes must operate through the instrument's effect on the endogenous variable—in this case, the number of online degrees. Therefore, if the exclusion restriction holds, the reduced-form relationship between the instrument and the outcomes should be stronger when the first stage is stronger. Since the IV estimator in Equation (D.2) is the ratio of the reduced-form coefficient to the first-stage coefficient, this implies that the second-stage estimates should be constant across periods. In other words, under constant treatment effects, correct functional form, and valid identifying assumptions, the shift-share IV estimator should converge to the same value in both periods. Observing this pattern would further support the credibility of our identification strategy.

Table D.2: Robustness to alternative estimators

	Δ in online students (1)	Δ in in-person students (2)	Δ in in-person degrees (3)	Δ in log-price of in-person degrees (4)
OLS	0.341	-0.192	-0.163	-0.009
	(0.034)	(0.026)	(0.028)	(0.002)
Shift-share IV	0.386	-0.190	-0.185	-0.013
	(0.035)	(0.030)	(0.036)	(0.003)
Overid. TSLS	0.358	-0.201	-0.179	-0.009
	(0.037)	(0.026)	(0.027)	(0.002)
MBTSLS	0.360	-0.202	-0.180	-0.009
	(0.037)	(0.026)	(0.027)	(0.002)
LIML	0.443	-0.213	-0.194	-0.009
	(0.056)	(0.028)	(0.028)	(0.002)
HFUL	1.070	-0.207	-0.203	-0.007
	(0.052)	(0.012)	(0.016)	(0.002)

Notes: This table reports estimates of the effect of an additional online degree on several outcomes of interest. Columns (1)–(4) present 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)–(3) use all region-field pairs; column (4) restricts to the 893 region-field pairs with at least one in-person degree in both 2010 and 2019. The OLS row reports ordinary least squares estimates, replicating Table 2, Panel A. The Shift-share IV row reports estimates from the preferred shift-share instrumental variable approach, replicating Table 2, Panel B. The Overid. TSLS row uses each firm share as a separate instrument. The MBTSLS row uses the estimator from Anatolyev (2013) and Kolesár et al. (2015). The LIML row shows estimates using the limited information maximum likelihood estimator. Finally, the HFUL row uses the HFUL estimator from Hausman et al. (2012). Standard errors, shown in parentheses, are obtained by bootstrap.

We test this prediction and present the results in Table D.3. We begin by presenting results using our shift-share instrument, and then using the shares of the nine firms with the highest Rotemberg weights as separate instruments.[8] Column (1) shows the growth in the number of online degrees during each period, first aggregating across all firms and then separately for each of the nine most influential firms. We find that the expansion of online degrees was nearly five times larger during the second period. Column (2) reports the first-stage regression coefficient for the change in the number of online degrees against the instrument, and the t-test comparing the coefficient from both periods. As expected, the first-stage coefficient is significantly larger in the second period, a pattern that holds when instrumenting with each firm's share separately. Columns (3)–(6) present the IV estimates of the parameter ϕ_t from Equation (D.2), along with t-tests comparing estimates across periods. Despite substantial and significant differences in

[8]When estimating the model using each firm's instrument separately, we multiply the shares by the total growth in online degrees to keep the first-stage parameter scale comparable to the main specification (i.e., the instrument is given by $\tilde{z}_{frt} = z_{fr} z_{fa} z_a \sum_f \Delta N^o_f$). Since $\sum_f \Delta N^o_f$ is a constant, results in Columns (3)–(6) are invariant to this adjustment.

the instruments' first-stage coefficients, the differences in the IV estimates across periods are generally small and statistically insignificant. We interpret this as further evidence supporting the validity of our exclusion restriction.

Table D.3: Robustness to alternative sample periods

IV	years	ΔN_f^o (1)	Δ in online degrees (2) coeff.	t-stat	Δ in online students (3) coeff.	t-stat	Δ in in-person students (4) coeff.	t-stat	Δ in in-person degrees (5) coeff.	t-stat	Δ in log-price of in-person degrees (6) coeff.	t-stat
SS-IV	10–14	881	0.359	6.93	0.568	1.87	0.103	2.27	-0.313	0.43	-0.033	2.12
			(0.056)		(0.121)		(0.115)		(0.145)		(0.012)	
	15–19	3975	1.255		0.368		-0.196		-0.245		-0.007	
			(0.107)		(0.032)		(0.029)		(0.035)		(0.002)	
Firm 1	10–14	50	0.309	3.80	0.341	0.79	0.074	1.70	-0.305	0.11	-0.038	2.86
			(0.049)		(0.105)		(0.124)		(0.155)		(0.010)	
	15–19	479	0.827		0.277		-0.191		-0.286		-0.006	
			(0.109)		(0.039)		(0.045)		(0.049)		(0.002)	
Firm 2	10–14	68	0.245	3.61	0.434	0.78	0.195	2.13	-0.289	0.08	-0.041	2.58
			(0.053)		(0.140)		(0.162)		(0.195)		(0.013)	
	15–19	255	0.699		0.340		-0.231		-0.272		-0.007	
			(0.103)		(0.036)		(0.052)		(0.053)		(0.002)	
Firm 3	10–14	71	0.245	6.73	0.435	2.25	-0.085	0.38	-0.407	1.47	-0.024	1.63
			(0.037)		(0.086)		(0.092)		(0.124)		(0.011)	
	15–19	248	0.904		0.278		-0.128		-0.206		-0.006	
			(0.107)		(0.032)		(0.029)		(0.034)		(0.002)	
Firm 4	10–14	192	0.290	5.89	0.450	1.78	-0.012	2.24	-0.156	1.13	-0.007	0.19
			(0.035)		(0.087)		(0.082)		(0.098)		(0.009)	
	15–19	52	0.759		0.328		-0.251		-0.281		-0.006	
			(0.078)		(0.034)		(0.032)		(0.036)		(0.002)	
Firm 5	10–14	13	0.289	4.91	0.591	1.27	-0.107	1.27	-0.256	0.07	-0.018	1.43
			(0.042)		(0.134)		(0.090)		(0.112)		(0.010)	
	15–19	337	0.839		0.448		-0.239		-0.264		-0.003	
			(0.102)		(0.069)		(0.024)		(0.033)		(0.002)	
Firm 6	10–14	133	0.264	3.81	0.440	0.83	0.063	1.53	-0.374	0.54	-0.046	2.76
			(0.061)		(0.142)		(0.138)		(0.172)		(0.014)	
	15–19	132	0.778		0.337		-0.180		-0.265		-0.006	
			(0.113)		(0.036)		(0.044)		(0.054)		(0.002)	
Firm 7	10–14	74	0.374	7.36	0.447	1.99	0.163	3.34	-0.202	0.63	-0.019	1.05
			(0.048)		(0.077)		(0.091)		(0.105)		(0.012)	
	15–19	453	1.098		0.322		-0.216		-0.278		-0.006	
			(0.105)		(0.027)		(0.030)		(0.036)		(0.002)	
Firm 8	10–14	64	0.274	3.45	0.384	0.64	0.153	1.98	-0.275	0.04	-0.039	2.60
			(0.059)		(0.129)		(0.156)		(0.189)		(0.012)	
	15–19	142	0.738		0.317		-0.234		-0.283		-0.007	
			(0.104)		(0.036)		(0.052)		(0.052)		(0.002)	
Firm 9	10–14	43	0.083	6.43	2.303	2.73	-0.014	0.62	-0.605	1.51	-0.047	1.15
			(0.030)		(0.666)		(0.203)		(0.270)		(0.033)	
	15–19	189	0.792		0.490		-0.141		-0.185		-0.009	
			(0.134)		(0.041)		(0.021)		(0.028)		(0.002)	

Notes: This table reports estimates of the effect of an additional online degree on our outcomes of interest from Equation (D.2). Column (1) presents the change in the number of online degree by firm f (and by all firms for the SS-IV panel). Column (2) reports the first-stage coefficient from regressing the endogenous variable, $\Delta_t N^o_{ra}$, on the instrument. Columns (3)–(6) report 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 (3)–(5) use all region-field pairs, while column (6) restricts the sample to region-field pairs with at least one in-person degree at the beginning and end of each period. The table includes 10 panels. The SS-IV presents estimates using the shift-share estimator from the main article. The remaining panels, labeled "Firm i", are the estimates from using the instrument derived from each of the nine firms with the largest Rotemberg weights. Each panel presents results for the periods 2010–2014 and 2015–2019, along with a t-statistic testing for differences in coefficients between the two periods. Regressions using differences between 2019 and 2010 are reported in Table 2, Panel B, and Table D.1, Panel C.

Appendix E: Model Details

E.1. *Internet penetration and degree entry*

In Section 4, we use internet penetration as an excluded variable to estimate the entry fixed-cost parameters. 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:

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

where ΔI_r 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 each of 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 E.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 E.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				
Δ internet penetration	3.372 (0.875)	-0.516 (0.323)	1.856 (0.660)	-1.205 (0.271)
Panel B: Mean of the dependant variable in 2010 and 2019				
2010	3.10	11.43	0.80	3.41
2019	9.91	12.44	3.06	3.24
Obs.	1,100	1,100	1,100	1,100

Notes:

E.2. *First-step prediction model of entry probabilities*

We estimate the entry fixed costs described in Section 4.4 using the two-step estimator developed by Sweeting (2009). From the firm's maximization problem, the probability that firm f chooses bundle $\mathcal{J}_{frt}$ is given by Equation (17) of the main text:

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

A key challenge in estimation is that the expected profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$, are unknown, and they depend on the entry probabilities, $\phi_{\mathcal{J}_{frt}}$, which in turn depend on $\mathbb{E}[\pi_{frt}(\mathcal{J})]$. To overcome this challenge, we proceed as follows: First, we compute a proxy variable, $\tilde{\pi}_{frt}(\mathcal{J}_{frt})$, that approximates expected profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$, in a parsimonious way. Second, we estimate an analogue of Equation (17), replacing $\mathbb{E}[\pi_{frt}(\mathcal{J})]$ with our proxy variable, $\tilde{\pi}_{frt}(\mathcal{J}_{frt})$, to obtain first-step entry probabilities, $\phi_{\mathcal{J}_{frt}}$. Third, we use the estimated first-step entry probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, to simulate market structures and estimate the expected variable profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$. Finally, we use the resulting estimated expected profits, $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$, to estimate the entry fixed costs in the estimator's second step. In this appendix, we provide more details about this procedure.

1. *Compute profit proxies*: We approximate a bundle's potential profits by summing the profits of each degree if offered individually, holding competitors' bundles fixed, and assuming supply and demand shocks are zero. Formally:

$$\tilde{\pi}_{frt}(\mathcal{J}_{frt}) = \sum_{j \in \mathcal{J}_{frt}} \hat{\pi}_{jrt}, \tag{E.2}$$

where

$$\hat{\pi}_{jrt} = \max_{p_{jrt}} (p_{jrt} - \tilde{c}_{jrt}) \tilde{s}(p_{jrt}, p_{-jrt}) \tag{E.3}$$

represents single-product profits when firm f offers only degree j, competitors offer the bundles observed in the data, $\tilde{c}_{jrt}$ and $\tilde{s}(\cdot)$ are the marginal-cost and demand functions when supply and demand shocks, ω_{jrt} and ξ_{jrt}, are equal to zero, and firms compete in prices a la Nash-Bertrand. We use $\tilde{\pi}_{frt}(\mathcal{J}_{frt})$ to approximate expected profits, $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$, in a parsimonious way without having to integrate over the distribution of supply and demand shocks or knowing the equilibrium entry probabilities, $\phi_{\mathcal{J}_{frt}}$.

2. *Estimate first-step entry probabilities*: We specify the first-step entry probabilities as a function of the profit proxies, $\tilde{\pi}_{frt}(\mathcal{J}_{frt})$, and a linear combination of the fixed-cost determinants from Equation (13), replacing the unknown expected profits, $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$, with their proxies, $\tilde{\pi}_{frt}(\mathcal{J}_{frt})$, in Equation (17). Formally, we model the probability that firm f offers bundle $\mathcal{J}_{frt}$ as:

$$\phi_{\mathcal{J}_{frt}} = \frac{\exp(\theta \tilde{\pi}_{frt}(\mathcal{J}_{frt}) - \text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}) - \text{Infrastucture}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0}))}{\sum_{\mathcal{J} \in \mathcal{B}_f} \exp(\theta \tilde{\pi}_{frt}(\mathcal{J}) - \text{Degrees}(\mathcal{J}|\mathcal{J}_{frt_0}) - \text{Infrastucture}(\mathcal{J}|\mathcal{J}_{frt_0}))}, \tag{E.4}$$

where $\text{Degrees}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0})$ and $\text{Infrastructure}(\mathcal{J}_{frt}|\mathcal{J}_{frt_0})$ are defined as in the main text and estimated jointly with θ. We estimate Equation (E.4) via maximum likelihood to recover the first-step entry probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$.

3. *Estimate expected variable profits*: We use the estimated first-step probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$,

to compute the expected variable profits, $\mathbb{E}[\pi_{frt}(\mathcal{J})]$. For each market, we simulate the bundle choices of the focal firm and its competitors, $\{\mathcal{J}_{frt}, \mathcal{J}_{f'rt}\}$, based on the estimated probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, and draw 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 described in Equation (11). Each market is simulated 10,000 times, and the expected variable profits for each bundle are obtained by integrating over all simulation draws. To reduce noise and get stable predictions for bundles with low choice probability, a random forest model is applied to the simulated values. 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 $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$.

4. *Estimate the entry model*: In the second step, we estimate the fixed-cost parameters via maximum likelihood, replacing $\mathbb{E}[\pi_{frt}(\cdot)]$ in Equation (17) with $\hat{\mathbb{E}}[\pi_{frt}(\cdot)]$.

Figure E.1, Panel (a), presents a binscatter fit comparing the first-step profit proxies, $\tilde{\pi}_{frt}(\mathcal{J}_{frt})$, estimated in Equation (E.2) with the estimated expected variable profits, $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$, computed using the first-step entry probabilities from Equation (E.4). Although the binscatter points lie slightly below the yellow 45-degree line, the correlation between the two estimates is 0.962, indicating that the first-step profit proxies capture the effects of demand and marginal cost components on profits relatively well.[9]

In Figure E.1, Panel (b), we compare the first-step probability estimates of Equation (E.4) with the equilibrium probabilities obtained in the second step from Equation (17). The correlation between first- and second-step probabilities is 0.99, indicating that the two-step estimator converges rapidly to the final estimated probabilities, thus avoiding the need to use a nested fixed-point algorithm, as in Seim (2006).

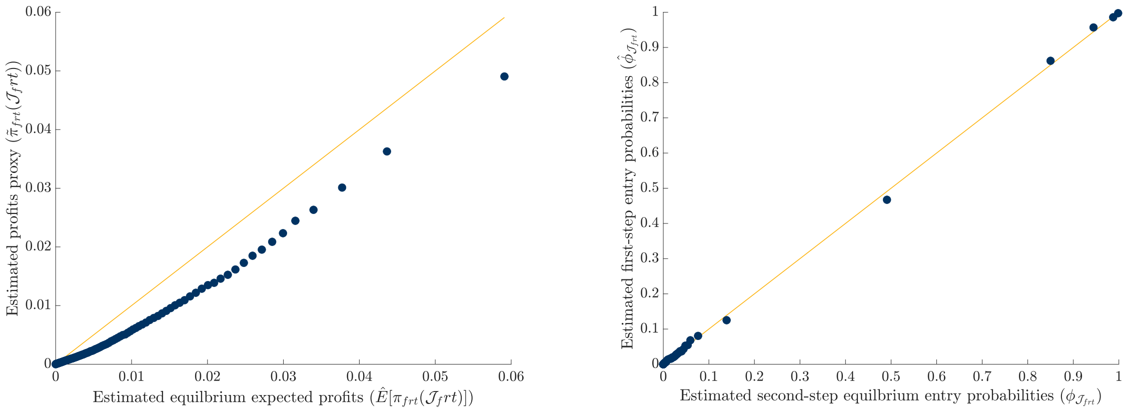
E.3. *Counterfactual Simulation Details*

To implement our counterfactual analysis in Section 5, we simulate equilibrium outcomes using our estimated model. The dataset includes all firms and the respective bundles $\mathcal{J}_{frt} \in \mathcal{B}_f$ that they can offer in each market, together with the estimated profits of offering each bundle, $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$, 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 $i-1$, denoted $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{i-1}$. For the first iteration, use the expected profits estimated from the data.

   (a) Draw the private-information fixed-cost shocks, $\varepsilon_{fr\mathcal{J}_{frt}}$, for all potential bundles, $\mathcal{J}_{frt}$.

[9] Any proportional differences between the two variables will be absorbed by θ in Equation (E.4), which implies that a high correlation between both variables is more important than matching the 45-degree line in the binscatter of Figure E.1.



(a) Estimated proxies of bundles' profits vs estimated equilibrium profits

(b) Estimated first-stage entry probabilities vs estimated equilibrium entry probabilities

Figure E.1: Comparison of first-step and second-step estimates

Notes: This figure presents binscatter plots comparing the first- and second-step estimates from our two-step algorithm. Blue dots show the average value of the vertical-axis variable within each bin. In each panel, the horizontal axis is divided into 100 equally-sized bins. The yellow line indicates the 45-degree line. Panel (a) presents the relationship between the first-step profit proxies estimated in Equation (E.2) and the estimated expected variable profits, $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J})]$, computed using the first-step entry probabilities from Equation (E.4) and integrating over the distributions of ω_{jrt} and ξ_{jrt} via simulations. The correlation between the two estimates is 0.962. Panel (b) presents the relationship between the first-step entry probabilities, $\hat{\phi}_{\mathcal{J}_{frt}}$, and the equilibrium probabilities obtained in the second step from Equation (17). The correlation between the two estimates is 0.99.

(b) Solve the firms' entry problem from Equation (16) using the fixed-cost draws and $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{i-1}$ to determine which degrees are offered in the market.

(c) Draw 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 $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]$. Denote the estimate $\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J}_{frt})]$.

2. Update the vector of expected profits to $\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_i = \sqrt{\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{i-1}\hat{\mathbb{E}}[\pi_{frt}(\mathcal{J}_{frt})]}$ and calculate the mean square difference between iteration i and $i-1$ as $MSD_i = \sum\left(\mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_i - \mathbb{E}[\pi_{frt}(\mathcal{J}_{frt})]_{i-1}\right)^2$, where the sum is over all markets, firms, and bundles.

3. Iterate until $MSD_i < \overline{MSD}$ for $\overline{MSD}$ small.

E.4. *Outcomes of Interest*

In this appendix, we provide the formulas used to calculate the main outcomes of interest reported in Section 5 of the main text. For each counterfactual k, we calculate

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

where VA_{jr} is the value added of degree j in region r, p_{jr}^k 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_{jr}^k is the market share of degree j in region r under counterfactual k, and s_{jr}^{BL} is the market share of degree j in region r under the baseline counterfactual. Note that the denominator is constant across counterfactuals and serves to normalize all outcomes on a per-student basis.