Effects of Competition and Policy in Brazilian Higher Education

Evaluation of social and market outcomes of the expansion of private higher education in Brazil.

Word count 31,212

Israel Gottschalk

A thesis presented for the degree of

Doctor of Philosophy

School of Economics and Centre for Competition Policy
University of East Anglia
United Kingdom

March 2022

Abstract

Higher education has expanded in most developed and developing countries as part of a shift from elitist to mass education systems. Since national budgets are constrained and demand is ever increasing, countries have resorted to a combination of private provision and public policies to promote equal access to higher education - such as tuition financing and scholarships - while relying on market competition to deliver on enrollment expansion and education quality while keeping fees at bay. In this thesis we investigate the effects of tuition financing on student enrollments and migration from cities of different sizes, and we also look at the effects of provider competition on city-level higher education outputs, such as enrollments and education quality.

Access Condition and Agreement

Contents

List of Figures

List of Tables

List of Acronyms

ENEM - Exame Nacional do Ensino Medio - National Examination of Secondary Education

ENADE - Exame Nacional de Desempenho dos Estudantes - National Examination of Student Achievement

FIES - Fundo de Financiamento Estudantil - Student Financing Fund

FNDE - Fundo Nacional de Desenvolvimento da Educacao

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anisio Teixeira - National Institute of Educational Studies and Research Anisio Teixeira

MEC - Ministerio da Educacao - Ministry of Education

PROUNI - Programa Universidade para Todos - University for All Program

SERES -Secretaria de Regulacao e Supervisao da Educacao Superior - Secretariat of Regulation and Supervision of Higher Education

SINAES - Sistema Nacional de Avaliacao da Educacao Superior - National System of Higher Education Evaluation

Acknowledgements

This PhD was a long ride full of ups and downs. More than finding answers to research questions, doing research is mostly about getting lost and trying to find a way. It is through embracing the unknown that we get access to the new and unexpected. In this rocky road, the support and understanding of supervisors, friends and family becomes a lifeline. I thank them for their patience, love and support.

My supervisors Farasat Bokhari, Kai-Uwe Kuhn and Sean Ennis, who have offered me much more than guidance, but also friendship and inspiration. Also Corrado di Maria who has been the mentor of my unlikely academic career.

My wife Carolina Gottschalk, my love, my shelter and intellectual drive.

My family Joao, Marion and Raquel Gottschalk, my companions of life from beginning to end, partners in every enterprise.

My friends Deanna Karapetyan and Vicenc Guasch, with whom I shared all the necessary and unnecessary complications of this journey.

Chapter 1

Introduction

1.1 The global expansion of private higher education

Higher education has been historically seen as a public good. Positive externalities mean that the socially desirable number of higher education graduates is higher than what would be observed if individuals were to bear the full costs of their degrees. This has led governments to either foot the bill in the form of public-funded universities or subsidize tuition fees.

But in the past few decades, higher education has been increasingly viewed as a private good due to the larger portion of the benefits from a degree accruing to the individual who obtained it. In this sense, the transfer of the costs of higher education from the taxpayer to the graduate seemed like a fairer deal (Altbach et al., 2009).

Add to this the constraints on funding in the face of increased demand and massification of higher education and we have a comprehensive picture of the recent expansion of higher education in most of the developed and developing world. The result was a shift from public to private provision, transferring the costs of education to individuals and relying on markets to provide the vertical and horizontal product differentiation to match the demand while ensuring prices are right.

But this shift from public to private provision brought a significant change to higher education systems which policymakers are still struggling to understand. Private provision of education subject to the market laws of demand and supply looks very different from the public provision. Also, the policy interventions which are required to guarantee fairness of access for students of different income levels generate further changes in the landscape which can add to the complexity of the picture.

According to Altbach et al. (2009), among the worldwide trends for higher education financing in the decade from 2010 were the increasing demand for higher education, rising costs, the inability of governments' tax revenues to keep pace with the rising costs of funding higher education, the increase in liberalization (free markets and private sector-driven provision) and the increased importance of financial assistance.

The research in this thesis investigates two aspects of this shift to mass private higher education systems. First we look at the effects of higher education financing on the geographical imbalances of tertiary students. We investigate whether tuition fee financing leads to increased imbalances of enrollments between rural and urban locations and to the relocation of students from small to larger cities.

Second, we investigate the broader outcomes of higher education under different degrees of competition. Using an HHI measure of concentration at the city level as a proxy for competition, we model fees, enrollments, and quality in geographical locations with different levels of competition.

Although the conclusions from these investigations can be generalized, we use Brazil as our case study throughout this thesis. We do this because the country gathers several qualities of a good subject of study: Brazil is a country with relatively abundant and publicly accessible data, especially in higher education; it also went through a liberalization of the education sector in the early 1990s which gave rise to a boom in enrollments and providers which extended through the 2000s. Also, the geographical heterogeneity of the country meant large variances in the data while the expansion occurred.

Brazil is also an interesting case due to the several public interventions in the private higher education sector in the 2000s. Many of those interventions are very large and their implementation was as rampant as their decline, which provides a researcher with a rich environment to examine their effects. This is the case of FIES, the tuition financing scheme which, in less than 10 years, went from negligible numbers to financing nearly half of all private sector enrollments, and back to where it started.

1.2 Higher education in Brazil

To give some context to the research conducted through this thesis, here are a few fundamental aspects of Brazilian higher education which are worth keeping in mind.

1.2.1 Provision

The Brazilian higher education sector consists of 2,448 higher education providers, split between public and private institutions. Roughly 75% of the 8,290,911 students are enrolled in the private sector, while the public sector responds for little over 25% according to the Higher Education Census (INEP, 2018).

In the public sector, larger federal universities are the typical provider. They respond for most of the total research output and are associated with higher quality. Placements in public higher education programs are remarkably competitive given the associated high standards, the capacity constraints, and the absence of fees. Conversely, private providers are typically smaller, have lower quality standards and do not produce research, except for some niche higher-end providers.

1.2.2 Admission

Students looking to be admitted into higher education programs are required to undertake an admission examination. Most providers adopt the National Secondary School Examination (ENEM) as the main criteria for admissions. This exam is held annually in multiple locations in the country at the exact same time. In 2018, the examination was available in 1,728 municipalities, almost one third of all municipalities in the country.

1.2.3 Assessment, regulation and supervision

These responsibilities are shared between two departments under the Ministry of Education: The National Institute of Educational Research (INEP) carries out the quality assessment of higher education programs in a triennial cycle, evaluating programs in different subjects every year. One of the main instruments of quality assessment is the National Examination of Student Achievement (ENADE), a subject-specific examination applied to students in their senior year which in this research will serve as our output measurement of student achievement. The quality assessment results inform the operation of the Secretariat of Regulation and Supervision of Higher Education (SERES), responsible for the enforcement of minimum quality standards, the accreditation of providers and authorization of programs.

1.2.4 Funding and financing

Public funding and financing are key features of the landscape of Brazilian higher education. Federal, state and municipal governments are responsible for funding public providers. For the private sector, though, the role of public funds is also substantial. According to IDados (2016) over a third of all students enrolled in private higher education had their tuition paid through the financing scheme FIES (22%) or by the scholarship program PROUNI (8%) in 2018.

1.3 The structure of this thesis

After a brief background to Brazil provided in Chapter 1, we revisit the economics of education literature in Chapter 2, focusing on the areas which are of particular relevance for the work we conducted on the effects of competition on student and social outcomes, and on the effects of loan programs on student migration. These are the topics for the research we presented in Chapter 3 on domestic student migration and Chapter 4 on competition among universities. Our last chapter 5 presents a conclusion and directions for further research.

Chapter 2

The economics of education literature

The literature in the economics of education is a rapidly expanding field of research. This can be observed in the breath of subjects and the depth of the academic literature, as well as in the geographical expansion of its coverage - primarily concentrated in US schools - to virtually all countries.

A seminal champion and contributor of this field of research is Eric Hanushek, whose framework of analysis of education as a process of inputs and outputs (see Hanushek (1979)) is central to a large proportion of the literature in this field. Hanushek has also served as a hub of the wider research in this area through his organization of the Handbook of the Economics of the Education, which in 2023 entered its 7th edition (see Hanushek et al. (2023)).

According to Hanushek et al. (2000), the research in economics of education has several natural overlaps with labor economics, public finance, and growth theory, which have become fertile grounds for the expansion of the literature beyond the effectiveness and efficiency of schools. The two papers in this thesis, which are focused not on schooling but on higher education, belong in that broader field of economics of education, both of them sitting in the crossroads with adjacent economics fields of academic research. In Chapter 4 we examine the effects of competition on higher education outputs, including a proxy of market structure among education inputs. This paper, therefore, is well placed within the economics of education literature as well as in the industrial organization literature, joining a crescent body of research now elegantly named the Industrial Organization of Education (see Belfield and Levin (2002)).

Chapter 3 is in the intersection between economics of education and development economics, looking at the effects of a policy to finance university tuition in patterns of student relocation between Brazilian cities.

As this thesis is centered on the intersections of economics of education with development economics - Chapter 3 - and with the economics of industrial organization - Chapter 4 - we first give an overview of the general development of the literature in the common field of economics of education, and then explore the academic literature dedicated to the intersection with other academic fields.

2.1 The literature on education inputs and outputs

The empirical literature in economics of education has, at its core, the canon dedicated to exploring the relationship of education inputs and outputs - the effects of school characteristics on student outputs and outcomes. Beginning with the Coleman Congressional Report of 1966 in the U.S., most of the empirics pointed to nil or modest impacts of schools in student outcomes, placing far greater importance on the family environment of students, their socio-economic background and individual ability. This was later corroborated by the results of a major UK report entitled Fifteen Thousand Hours Rutter (1982).

Since the 1980's, though, this field of research saw a progressive improvement of results. This was mainly driven by the increasing availability of data on the US and beyond, which enabled researchers to not only obtain better estimates using traditional methodologies, but also - and especially - to deploy more effective and robust models using newly available data.

A framework which gained a strong foothold since then is the education production function's view of the education process in terms of inputs and outputs.

2.1.1 Hanushek's production functions

Hanushek's first paper investigating the relationship between student achievement (output) and the inputs to the educational process through education production functions considers schools as the catalyst of the transformation of education inputs, such as educational resources and pupil characteristics, into outputs, measured in examination scores, attainment and other observable outcomes (see Hanushek (1979)). This approach has since become ubiquitous in the economics of education literature, with different degrees of success in implementation.

In his paper *The Economics of Schooling: Production and Efficiency in Public Schools*, Hanushek (1986) offers a review of 147 studies which employ educational production functions to examine the relationship of educational inputs and outputs. Hanushek divides inputs into family, peer and school variables. Among the school-related factors, the literature has traditionally examined teacher experience, salary and rank, as well as student-teacher ratios and expenditure per pupil, but with little success in determining their impact on outcomes.

Among those studies examined by Hanushek, papers such as Hanushek (1971), Hanushek (1986), Murnane (1975), Armor et al. (1976) and Murnane and Phillips

(1981) are unequivocal in their conclusion that teachers and schools differ dramatically in their effectiveness, contrary to the impression left by the Coleman Report and other subsequent studies. According to Hanushek, these faulty impressions are mainly derived from the difficulty in measuring components of effectiveness and true effectiveness. This is an issue which we concentrated many efforts to address in our Chapter on the effects of competition on student achievement. Understanding that measures of characteristics of teachers and schools are flawed and as such are poor indicators of the true effects of schools, we introduce a new methodology to estimate measures of education quality which can help address the traditional measurement issues.

In more recent studies applying the same input and output functions to schooling in a cross-country comparison, Wedel (2021) estimates that an additional hour of instruction time leads to an increase of 0.03 standard deviations in students' test scores.

2.1.2 OLS regressions and frontier estimations

In an inventory check of the economic literature on the relationship between resource allocation and pupil attainment, Vignoles et al. (2000) survey the methodologies and the pitfalls that have underpinned this field of research. According to them, the education production functions had become ubiquitous in the literature after Hanushek's work and the literature split between different implementation of that approach: regression models to ascertain how schools with higher (lower) educational resources perform against the mean, and frontier estimation approaches, such as DEA and SFA, that evaluate the performance of schools in comparison to the production frontier.

Following Vignoles et al. (2000), the regression techniques are generally parametric models used to estimate 'average statistical behaviour', such as the multilevel approach implemented in Mayston and Jesson (1999), the simpler school-effect models provided in Creemers and Reezigt (1996) and Reynolds et al. (1996) which we discuss below, and the wider aforementioned literature on the education production functions which is the basis of our Chapter 4 on competition and education outputs. Regression analysis has been used to ascertain whether schools with higher resource levels also have higher performance levels, in relation to the average performance of all schools. These regression models generally also require the user to specify a particular relationship between the chosen inputs and the outputs of interest.

The second primary method is known as frontier estimation. This technique can either be parametric, which uses stochastic frontier regressions to specify the functional form of the stochastic production function, or non-parametric, such as Data Envelopment Analysis (DEA). The main goal of these approaches is to assess the performance of schools concerning the production frontier.

A stochastic frontier analysis, as provided in Cooper and Cohn (1997), can identify the schools with the best possible outcomes for a given level of inputs, which are on the frontier of the educational production set. Several papers provide comparisons between these two methods such as Thanassoulis (1993) and some have championed frontier methods as more superior to OLS regressions in this context, such Ganley and Cubbin (1992), although Vignoles et al. (2000) argues that advantages and disadvantages of different strategies are very context-specific.

DEA analysis can help identify efficient and inefficient schools and even estimate the potential efficiency gains that would be had in the case where all schools were as efficient as the best-performing ones in the sample. DEA has a shortcoming, though, as it does not give a quantitative estimate of the impact of any particular input, which is an important factor for policymakers. Some of the studies employing DEA therefore use a second-stage regression analysis to investigate the factors that affect school efficiency. Good examples of this are Bradley et al. (1999) and Kirjavainen and Loikkanent (1998), and a more recent study championing the value of DEA models for policymakers looking at school efficiency is presented by López-Torres and Prior (2022).

For all studies,Vignoles et al. (2000) point to three major methodological problems associated with the empirical works: First, the lack of an established consensus theory from which appropriate models can be constructed; second, the technical and empirical problems found in the existing research, especially the endogeneity of school funding and student selection. Finally, the shortcomings hatched by poor data on school characteristics and students, often limited and aggregated at the district level.

These shortcomings lead the literature into focusing into the effects of school funding on student outputs, where data at district-level was available and which did not require a specific theoretical framework for how school inputs affected student outputs, especially in the late 1990s. Papers such as Dewey et al. (2000), Maranto et al. (2000), Figlio (1997) and Gupta et al. (1999) point to positive and significant effects of per-pupil spending on educational outcomes. Effects range from small effects of the order of 0.1% in the case of Dewey et al. (2000), to larger estimates ranging from 2.5% to 6.4% in the case of Figlio (1997).

A more recent paper, Jackson et al. (2015) examines the effects of school funding on wages and years of education. Event-study and instrumental variable models show that a 10% increase in per-pupil spending each year for all twelve years of public school leads to 0.27 more completed years of education, 7.25% higher wages, and a 3.67% reduction in the annual incidence of adult poverty, with effects much more pronounced for children from low-income families. See Hyman (2017) for a similar paper with corroborating results.

The literature looking into the effects of school financing is quite active. Holmlund et al. (2010) investigates the effects of school spending in England, where school expenditure has increased by about 40% since 2000. Holmlund et al. (2010) finds that that the increase in school expenditure had a consistently positive effect on primary school outcomes, with higher effects for students who come from economically disadvantaged backgrounds.

Given the lack of data, another strand of the literature concentrated in trying to identify a measure of general school effects, according to Vignoles et al. (2000). In practice, this meant incorporating a set of dummy variables for good' or 'bad' schools in regressions to measure the individual effect of schools in a standard regression of outcomes, such as standardised test scores, controls for family background and the set of school dummies. Results indicated that schools have a significant impact on outcomes. Based on a set of countries including the UK, Creemers and Reezigt (1996) concluded that between 10-20% of the variance in student achievement is explained by school factors. According to Reynolds et al. (1996), this range is around 8-12%, and larger in primary school compared to secondary.

These results contradicted the conclusions of the early Coleman Congressional Report of 1966 and Rutter's *Fifteen Thousand Hours*, and now convincingly indicated that school-specific effects matter. Yet when researchers progressed from identifying school effects to analysing the effect of particular inputs, such as family background and educational expenditure, the results were disappointing.

2.1.3 Industrial organization of education

Although there is an established empirical literature on the economics of education and on the industrial organization of schools, the literature on the industrial organization of higher education is largely missing. As a result, market regulation in higher education markets has often lacked clear scientific guidance.

In an effort of similar magnitude to that of Vignoles et al. (2000), Belfield and Levin (2002) provide an extensive review of the literature directed at the effects of competition on schooling in the U.S. context. In general, the inconclusive results in the early economics of education literature flagged the forthcoming challenges that would test this field of research: if the general schooling effects themselves were hard to establish in the education production functions, the effects of school competition on student outcomes were even harder to pin down, since competition should only have an effect on students' achievement *through* school effects. Furthermore, researchers found another source of endogeneity on the market structure controls: School quality is a joint determinant of market shares, and not all of its dimensions can be controlled for in the regressions.

2.2 Industrial organization of education

The industrial organization of education refers to the study of how market structure, competition, and regulation affect the delivery and outcomes of education. It is an interdisciplinary field that draws on economics, sociology, psychology, and education.

Some of the main topics studied in the industrial organization of education include market structure and competition in education markets, the impact of school funding and resource allocation on student outcomes, the role of government in regulating education markets and the evaluation of education policies and interventions, among others.

2.2.1 Industrial organization of schooling

The literature on competition in education relies heavily on studies looking at school inputs and outputs, largely based on the education production functions championed by Hanushek (1986). Borland and Howsen (1992) inaugurated the canon by pioneering the inclusion of a Herfindahl-Hirschman Index (HHI) into the consolidated input-output models to verify for the effect of market concentration in student achievement. The authors use a 2SLS model and data on 170 school districts in Kentucky from 1989-1990, estimating simultaneous equations on the impact of HHI on student achievement and teacher salary to find statistically non-significant effects of competition. In a replication of the study a year later, but now converting the continuous HHI variable into a binary control of high and low competition levels, Borland and Howsen (1993) found a very small and positive effect, but now significant at the 5% level.

To address the endogeneity of market structure, Borland and Howsen use school switching regimes to instrument for the level of competition in this second paper, and include a measurement of students' cognitive ability to control for the endogenous selection of students into schools, which represented a significant improvement on the previous literature, but still fell short of controlling for prior education quality.

Beyond the challenges posed by the endogenous nature of market structure, studies on the effects of education inputs on outcomes have met further sources of omitted variable biases. According to Hanushek (1986), student achievement is a product of not only the current but also prior education received by the student, not to mention other unobserved education features, as well as other factors external to the provider such as parental education and innate ability. To address this, Hanushek (1986) posited the inclusion of students' characteristics and previous attainment as controls in the production function estimation, turning it into a value-added model capable of separating the effects of a specific level of education from the students' baseline ability and education. Adding to the aforementioned endogeneity problems, Mayston (1996) also point out to the potential endogeneity of resource allocation among pupils within schools. He argues that the level of resources directed at each pupil will be endogenously determined if schools optimize their allocation of resources according to the ability of students, which creates an important endogeneity bias especially when researchers are solely modeling the supply side of the market in single, reduced form equations.

From the second half of the 1990s on, many studies were dedicated to improve on the results and the identification strategies laid out by the Hanushek and those first efforts by Borland and Howsen . According to a meta analysis of the literature conducted by Belfield and Levin (2002), 38% of 206 studies looking into the the effect of market competition on school outcomes have found positive but modest statistically significant results, while a trivial number of less than 5% of studies indicated negative effects of competition. Also, a large portion of the literature was devoted to investigate the cross effects of private provision on public school performance, such as Simon and Lovrich (1996) using controls for parental education, and Smith and Meier (1995) and Wrinkle et al. (1999) using lagged enrollment in private schools to address endogeneity. These studies found small negative impact of private school enrollments in public school performance but, as later demonstrated by Maranto et al. (2000), results were sensitive to income distribution.

2.2.2 Industrial organization of higher education

While the competition literature on schooling sprung in the 1980's, it was not until the 1990s that it set foot in higher education. Contrary to the empirically oriented literature devoted to schools, the higher education stream is more abundant in theoretical papers, dedicated to modeling student selection and the effects of provider competition, including the seminal paper by Fernandez and Gali (1999) that effectively pioneered the education incursion of the well established literature dedicated to matching problems in Becker (1973), Sattinger (1995) and many others.

Theoretical models

A number of theoretical papers have been dedicated to the sorting or students among universities. These include tournament matching models such as in Fernandez and Gali (1999), and build into more complex models that are able to account for an increasing array of factors which make up education markets. For instance Romero and Rey (2004) introduces different types of universities - public and private - and the role of university fees, leading the way for Del Rey and Estevan (2015) to later introduce the issue of imperfect capital markets, which is particularly pressing in developing economies. The paper by De Fraja and Iossa (2002), which includes mobility costs into students' objective functions is of particular interest for this thesis as it is not only relevant to our Chapter 4 on Competition, but also to our Chapter 3 on student migration.

Starting from the earlier works, Fernandez and Gali (1999) develop a matching model to examine the efficiency of tournaments and markets in assigning students to schools under imperfect capital markets. Borrowing constraints, in this case, can prevent efficient market allocations under fees because of the affordability requirements - students may have the *willingness* to pay, but not the *ability* to pay - so that public agents may consider a tournament mechanism instead, so that students and schools are matched on ability and quality only. In principle, efficient matching occurs when the best quality education is assigned to those higher ability individuals who can make a more efficient use of the resources.

Fernandez and Gali (1999) introduces the idea that fees impact selection under imperfect capital markets. But the financial burden of higher education is not limited to tuition and other directly associated costs. De Fraja and Iossa (2002) for instance has included mobility costs in their theoretical modelling, but did not account for the fact that wealthier students may be more mobile than poor students given the same mobility costs. Also, the same applies to the opportunity costs of higher education. Poor students cannot afford to forgo their wages to pursue a degree lest their family subsistence may be jeopardized. In fact, this explains why the vast majority of Brazilian students prefer the evening shift to study (see descriptive statistics).

Considering that under tournaments providers need to signal their quality to higher ability students, De Fraja and Iossa (2002) develop a theoretical model where two symmetric universities located in two towns compete by selecting the share of their budget dedicated to teaching and research, as well as the exam threshold for the admission of students. Assuming that fees are fixed, universities maximize their reputation while students maximize their utility from higher education - a function of their ability, the quality of the education and the mobility costs. De Fraja and Iossa (2002) find that if students' mobility costs are high, providers do not face competition and choose the same admission standards, admit the same number of students and invest the same amount in research. At intermediary mobility costs the equilibrium is asymmetric: one university becomes an elite institution and the other sets lower admission standards. At low mobility costs, there is no pure strategy equilibrium.

But considering that De Fraja and Iossa (2002) only allow for the matching of students and providers through entry examinations, a market with both fees and examinations as matching variables - such as the Brazilian case - escapes the scope of the paper. Allowing for the competition of asymmetric providers under fees and examinations, Romero and Rey (2004) model the interaction of a public and free provider and a private and fee-charging provider. The authors assume that the objective function of the free public university is to maximize public surplus while private maximize profits. The absence of fees, though, allows the public provider to act as a monopoly by cream-skimming on ability. According to Romero and Rey (2004), there is a unique equilibrium in which the public institution is of higher quality than the private.

Del Rey and Estevan (2015) follow Romero and Rey (2004) matching students and providers on examinations and fees under imperfect financial markets, in an approximation of the Brazilian context. In this paper, students incur in financial costs to pass the entry examinations, which are decreasing in ability. Considering that the entry exam threshold in the higher quality public, free universities is higher than that of lower quality and fee-charging private ones, students weigh the costs of passing the examination against the fees of the private provider. In the presence of borrowing constraints, lower ability individuals may prefer the lower quality private provider if the admission costs are sufficiently high.

In a paper investigating the student selection dynamics of higher education provision in Brazil, Del Rey and Estevan (2015) develop a theoretic model of the interaction of private and public providers. Since public providers are of higher quality and free but much harder to gain acceptance into, they conclude that the effective demand is simultaneously determined by students' willingness to pay for the fees and ability to pass the entry examinations into public universities. They find that private provision will only exist when providers are able to charge sufficiently low fees and will only attract those lower ability students that face higher personal costs of passing the public admissions test. The results stemming from our empirical study confirm the conclusions of Del Rey and Estevan (2015).

Empirical literature

Our main reference in the empirical literature of industrial organization of higher education, Hoxby (1997) uses a very similar methodology to the following Chapter 4. Writing on the U.S. college sector and using panel data on 1,121 baccalaureate granting colleges from 1940 to 1991, Hoxby (1997) uses both theoretical and empirical approaches to the dynamics of competition in the tertiary education sector, but did not establish a working mathematical model of competition to frame the empirical section. Hoxby (1997) proxies competition with geographic integration by constructing a measure of concentration based on the proportion of local students to students from other states. The author finds that markets under more stringent competition present higher average quality and higher tuition, as well as greater between-college variation in student quality.

On one hand, the author predicts that increased market integration should give rise to higher quality because investments in quality have higher returns in markets with open trade. Higher quality, therefore, comes at a higher tuition fee. On the other hand, competition leads to fee increases because it disrupts monopsony rents of the labor supply, so that providers need to compete for the staff.

But Hoxby's measurement of competition is not entirely flawless. The first one of them being that local students are defined as the ones that were born in the same location, disregarding those students who may have been born elsewhere but lived in the state prior to college education. But this is minor. The second and main flaw inherent to this is that mobility is likely to be correlated with ability - those students of higher ability may well be more willing to move towns in pursuit of more quality or a program that better suits them. Furthermore, the paper does not control for the selection of students coming in. In fact, it could be claimed that the proxy used for competition is very much also a proxy for the quality of education. Since wealthier and higher ability students are also more mobile, this index may well be capturing the attraction of the best students and not the competition effect on the SATs.

Even though many aspects of the Brazilian higher education sector are modeled in the literature, still there is no paper to model symmetric providers competing on fees and exams simultaneously, such as the private sector in Brazil. Neither the extremely complex case where symmetric and asymmetric providers interact in the same market, some of them free and some of them charging fees. This opens an opportunity for the development of a new theoretical model where symmetric providers compete in fees and examinations.

Effects of mergers on education output

A stream of literature which has been emerging more recently is the one dedicated to observe the effects of mergers on education output. This relates closely to the literature looking to the effects of concentration albeit using different empirical approaches.

Looking at the impacts of mergers in private higher education, Garcia and de Azevedo (2019) use a difference-in-difference approach to investigate the effects of horizontal and conglomerate mergers on education fees, quality and quantity.

They find that conglomerate mergers lead to an improvement in education quality and an increase in fees and freshmen students. In the case of horizontal mergers, they found that even though the effect on quality was positive, fees also increased, resulting in reduced number of students accepted.

In a closely related paper, Russell (2017) analyzes 107 mergers between 2001 and 2013 involving public and private non-profit higher education providers in the US, and found that the average merger increases tuition and fees by 7% compared to non-merging institutions in the same state.

Although the literature in mergers and acquisitions among education institutions is limited, both of these papers present relevant results for our investigation into the effects of competition in higher education output. Both Garcia and de Azevedo (2019) and Russell (2017) observe increase in fees and reduction in enrollments where horizontal mergers take place. Horizontal mergers are those with the potential to reduce competition in markets. Our research also points to higher fees and lower student enrollments in more concentrated markets.

2.3 Student financing and domestic migration

In our Chapter 3, we examine the effects of student financing covering tuition fees on the domestic migration of students in Brazil. This paper is inserted in the economics of education literature by way of its investigation of student financing policies.

Student financing is a major part of the higher education landscape in most developed economies. In the UK, the US and Europe, a large proportion of undergraduate students are funded by a financing scheme to cover their tuition fees, and Brazil is none different. However a study on whether and how student financing affects students' relocation decisions is at large. Some previous literature explored whether state-level financial aid affects the retention of students (see Mixon (1992) and Tuckman (1970) for studies in the U.S.), but the effects of the much more ample and far reaching tuition fee loans on student flows is still to be investigated.

2.3.1 Determinants of domestic migration

The migration literature is ample and far reaching, with a significant proportion of works dedicated to the causes and effects of international brain drain, a specially attractive topic for development economists concerned with the relation of human capital movement and the income gap between developed and developing countries. For a good review of papers examining education-led brain drain, both domestic and internationally see Browne (2017).

Early theoretical works on domestic migration

Domestic migration flows within countries is also very robust, despite having received somewhat less attention in the more recent years. The seminal article on rural to urban brain drain by Harris and Todaro (1970) being the most prominent theoretical framework in the area. [1]. The earliest empirical studies of determinants of regional migration, mostly theoretical pieces based on models from natural sciences, include Stouffer (1940) and Stewart (1948). More recent research papers look at the causes of rural-urban brain drain, such as Elder et al. (2015), and some also include education as a cause for migration, such as Punch and Sugden (2013).

Stouffer (1940) outlines a theoretical framework to explain how intervening opportunities can affect human mobility and provides some examples to illustrate the concept while Stewart (1948) based his theoretical framework on the concept of 'de-

[1]See (Giannoccolo, 2009) for a comprehensive literature review on the topic of brain drain, specially focused on the international dimension of human capital transfer from developing to developed countries, its causes and consequences.

mographic gravitation', which suggests that people are drawn to places with larger and more diverse populations. The paper argues that this concept can be used to explain a wide range of demographic phenomena, including migration patterns, urban growth, and the formation of social networks.

In an early empirical push Hagerstrand (1957) presents a survey-based study on the factors that influence human migration using Sweden as a case study. The methodology involves a survey of a sample of Swedish 'migration fields', which are defined as regions within which a significant proportion of the population has moved from one location to another. The author collected data on the characteristics of these migration fields, including their size, population density, economic structure, and social organization, and concluded that migration within a given area is influenced by factors including economic opportunities, social networks, and cultural ties.

Empirical literature on domestic migration

Following Stewart (1948), a number of papers invested in developing 'gravity models', fitting multiple regressions of rates of migration between paired locations as functions of their characteristics such as education and employment levels, the distances that separate them, among other variables. The 1960s and 1970s saw a number of research papers employing place-to-place migration models on census data in different countries and population strata. Seminal to this stream of literature is the contribution of Levy and Wadycki (1974) implementing seemingly unrelated regressions to a stratified sample for education levels in Venezuela, Beals et al. (1967) on the effect of income in inter-regional migration in Ghana, Kono and Shio (1965) on the inter-municipal migration in Japan, Schultz (1971) on the rural-urban migration in Colombia and Sahota (1968) on the labour migration between agricultural and industrial locations.

Schultz (1982) is an early reference to our Chapter 3 on student migration. Schultz (1982), estimated a logit model of migration probabilities for Venezuelan regions using a single aggregate cross section of data obtained from the 1961 Venezuelan Census, which contains only males, in an early precedent to our research using logit regressions at the student level. Schultz's modelling is broadly based on the previous gravity models implemented in the early literature as he estimates gross rates of migration between regions as a function of the distance between locations, and conditioning characteristics of regions. Schultz found a tendency for more educated men to leave their birthplace, and to be less deterred by distance. He also found that average wage rates at destination were associated with higher migration for all education levels, and higher wages at their hometown acted made higher educated workers more sticky.

2.3.2 Student migration

The literature in the more general issue of domestic migration serves as an important background to our Chapter 3 on student migration as it serves as the foundation for the literature looking at the more specific issue of determinants of student's decisions to relocate. In this section we look into the references within the domestic migration literature which have examined the specific case of students.

A few researchers have previously examined the determinants of national *student* migration, starting with the seminal work of Tuckman (1970). The earlier works in this body of literature, though, are limited to the U.S. interstate mobility with data aggregated at the state level (see Tuckman (1970); Mak and Moncur (2003); McHugh and Morgan (1984); Mixon (1992)), with a few papers also dedicated to student mobility within specific states (see Alm and Winters (2009)).

While Tuckman (1970) is specifically interested in the importance of fees and college capacity as determinants of inter-state migration, our Chapter 3 is dedicated to the effects of student financing on migration flows between cities so that we can examine whether student financing is increasing the divide between rural and urban areas. Though Tuckman (1970) also included a control for state-aid available to students in their model of outmigration, because their research was more concerned with the effects of tuition fees in the state of origin than with the effects of aid, no attempt was made to refine the state-aid control variable. Unsurprisingly, Tuckman (1970) did not find any significant effects of aid on the number of students migrating from a U.S. state.

Variations of Tuckman (1970) have extended the empiric research of student migration, especially the more recent papers of Mixon (1992) including destination characteristics, Mak and Moncur (2003) adding other explanatory factors on the place of origin and Alm and Winters (2009) performing a state-level study on the importance of travel distances in college migration in the state of Georgia. A few papers have also examined the effects of policy interventions on internal human capital flows in the U.S. Murphy (1969) looked at the statistical evidence supporting the view that qualified scientists left the U.S. Midwest for locations where federal funds for research and development were more concentrated. Murphy concluded that the Midwest suffered a net outflow of high-skilled workers which was partly due to public intervention.

De Fraja and Iossa (2002), theoretical literature in matching student to university, provides an interesting insight for our paper in migration by including mobility costs in their theoretical modeling of student and university matching.

In a slightly different exercise, Gibbons and Vignoles (2012) analyse the determinants of higher education participation. They find that the geographical distance between the place of study and of origin has little or no impact on the decision to participate, although it does have a strong influence on the institutional choice.

Our Chapter 3 differs from the existing literature in a few important points. First and foremost, this is the first paper to analyze city-to-city migration, where the previous literature has used state-to-state models of migration. Also this is the first time that we can test for policy implication on migration patterns. This is because our dataset contains year-by-year observations over a 10 year period where previous papers have worked with Census data, where data collections normally occur in 10-year intervals.

Chapter 3

Higher education financing and rural to urban brain-drain: a case study in Brazil

Abstract

Tuition financing for higher education allows cash constrained students to choose from universities away from their places of origin, which may increase the brain drain from rural to urban centers. Using a large Brazilian student-level dataset, we implement a logit model of the probability that a student will relocate given financing. We find that students from smaller cities and from large metropolises are more likely to relocate under financing, while those from medium-sized cities are more likely to stay in their place of origin. Testing for the differential effect by individual characteristics, we find that financing raises the likelihood of relocation disproportionately among students from private schools, white and female. Using a multinomial logit model of destination choice, we find that students from smaller

locations are more likely to relocate to other small cities and less likely to relocate to large metropolises. In fact, students from large metropolises are more likely to migrate to smaller cities under financing, indicating a reverse brain drain from the policy. Finally, we implement 2SLS models of the effects of financing on enrollments on city-level panels, and we find that although the relocation is more likely among financed students, the scheme has increased both the number of students enrolling in their places of origin as well as away. We also find that the policy helped expand the supply of higher education to locations previously deprived of access. We conclude that although the financing scheme increased the mobility of students especially from smaller cities this did not amount to brain drain.

3.1 Introduction

A large body of research has been dedicated to examine the determinants of domestic rural to urban migration, such as the seminal paper by Harris and Todaro (1970), the empirical work on US interstate migration produced byTuckman (1970) and other more recent empirical papers such as Mixon (1992), Mak and Moncur (2003) and Cooke and Boyle (2011). All of these papers point to a dominant concern: that for several reasons, rural locations continue to lose workers to urban ones.

Empirical works such as the early research by Schultz (1982) also confirmed another long-held suspicion: that mobility was positively correlated with degree of education. Schultz (1982) found an increased probability for more educated men to leave their birthplace and to be less deterred by distance. This means that rural locations are not only losing population, but maybe even more importantly, being depleted of the key human capital resources necessary to make good local government decisions, start and expand business that will innovate and create jobs to help close the income gap between rural and urban locations, and help raising the standards of living outside of cities.

Our research in this chapter looks a little further into this issue. We examine whether a Brazilian government policy to provide loans to cover tuition fees for higher education programs is increasing the probability that students will migrate away from smaller, rural locations. As these students may or may not come back to their hometowns, the policy may be contributing to deepening the rural-urban human capital gap and, therefore, regional development gaps.

3.1.1 Why students relocating can be important

The development gap between rural and urban areas is a long-dated and persistent issue feeding into political polarization and societal division, extending beyond the income gap and into political affiliations and social attitudes (see Eagles (1986) for one of a list of articles looking into this issue).

Although this divide is visible geographically, location is in fact a proxy for deep-rooted demographic differences, chief among them education levels. In Brazil, source of our data, large metropolises held 44% of all higher education students but only 24% of the country's population in 2019. Conversely, non-metropolitan locations accounted for 23.5% of students and 45% of the population.

Because at least some of the relocating students will be retained in the workforce in those places where they graduate (Groen, 2004; Parsad et al., 2005), policies directed at the expansion of access to higher education can either exacerbate or diminish the demographic differences between geographical areas by creating incentives for students to enroll and to relocate (Cooke and Boyle, 2011).

According to Groen (2004), in a paper looking at the effect of US merit-aid programs to develop and retain college-educated workers in a state, a student's decision to study away from her home state decreases the probability of working in her home state by 9% and increases the probability of working in her college state by 6%. The 3% difference relates to students who went on to work other states that are neither their home state nor their college state.

3.1.2 How finance can affect migration

In this chapter, we examine the migration effects of the Student Financing Fund (FIES), a large tuition loan scheme for classroom-based higher education students in Brazil. Introduced in 2010 and rapidly expanded, half of all private students were supported by the scheme in 2014. In the subsequent years, though, reforms lead to its reversion back to its initial 2010 level. The objective of this expansion was to increase enrollments and access to higher education supply.

The distribution of higher education opportunities is often described as *skewed* and *hierarchical* across regions, counties and urban areas, leading to the geographic concentration of skilled workers which is an important determinant of regional economic growth (Tight, 2007; Storper and Scott, 2009; Cooke and Boyle, 2011). Student financing can contribute to deepen these geographical inequalities in two ways. First, financing can increase the number of local students in their places of origin as the education investment becomes more affordable. Second, student financing mitigates income constraints so that destination alternatives are more numerous for those joining higher education. Students may therefore prefer to relocate to locations with more differentiated programs and better prospects of career opportunities. Our guiding research questions therefore are the following: Are financed students more likely to relocate from rural areas? Does student financing increase the likelihood of relocation from smaller rural areas to larger urban ones? What are the effects of the financing policy on the number of local students across city sizes?

The previous literature on U.S. interstate migration has consistently found patterns of migration from rural to urban states (Mixon, 1992; Mak and Moncur, 2003; Cooke and Boyle, 2011), but there are so far no attempt at explaining domestic migration among cities. Our research investigates whether students from rural locations are more likely to migrate and the choices of destination of students. Destinations of choice are is important to determine whether there is a brain drain because even though loan financing may have a significant effect on student propensity to migrate this may not be detrimental for rural locations if students are migrating among rural locations, or if a larger effect is observed in the inverse flow from large to small cities. Brain drain or not, financing may have a distinguishable marginal effect on migration but still be beneficial if the local enrollments growth outpace the effect on migration.

According to Tuckman (1970), student migration may be explained either by an investment or a consumption theory of demand. In the first case, students may migrate to increase the expected returns from their education. These returns may differ if obtaining higher education in smaller vs larger cities differs in terms of access to labour market, choice of subject, curriculum, etc. A student making a decision regarding the location of her university will presumably migrate when her expected returns from migration exceed the costs. In the case of a consumption theory of demand, students may decide to relocate because of the amenities offered in a given location such as access to better entertainment and culture. But regardless of the theory which better encapsulates the individual decision to migrate, for a number of students, the option to relocate is not accessible despite benefits exceeding costs because of income constraints. For those students, tuition financing can ease those constraints and facilitate migration.

From an investment perspective, a student from a small city may prefer to relocate to a larger city to pursue her higher education degree because of the better professional opportunities she will be exposed to during the course of her higher education (Cooke and Boyle, 2011). That option, though, may not be available to her because of her budget constraints. Studying away from home is more costly for many reasons, especially when students move to larger cities. Housing costs and living costs are generally higher, but also tuition fees are also more expensive. According to our data, the mean tuition fees for the 507 programs in cities with populations of less than 50,000 was R$709.76 per month in 2019. For the 1,079 programs in cities larger than 500,000, mean tuition was R$1008,63.

Several papers have previously examined the determinants of regional student migration since the seminal paper of Tuckman (1970), in which students weigh up the costs of migration against the benefits of relocating. These studies have considered the effects of tuition fees, quality and variety of educational opportunities on students decisions (Mak and Moncur, 2003), the economic climate in the destination location (McHugh and Morgan, 1984; Baryla Jr and Dotterweich, 2001) and the distance between the place of origin and destination Alm and Winters (2009). Although some of these papers have included the availability of financial aid in the destination location among the determinants of migration decisions, the literature on the effects of the comprehensive national programs of tuition financing is still missing. Furthermore, while the literature has largely established what are the characteristics of cities that serve as net senders and net receivers of students, it has not yet recognized that these factors are unevenly distributed among smaller and larger cities.

So although the established literature on student migration is large and prolific with a number of papers focused on the individual and geographical determinants of student mobility, this is the first paper to investigate the effects of student financing on individual location decisions and to explore the urban and rural divide as a determinant of relocation.

In the present investigation, we estimate the effects of student financing on students' individual propensities to migrate and their destination choices, and how it amounts to the aggregate patterns of enrollments at home and away from home across cities of different population sizes. Our hypothesis is that student financing increases the probability of migration among students out of rural locations, and that those students are more likely to relocate to larger urban centers. This follows directly from the previous literature looking at the US interstate migration of students, which points to students migrating out of rural states into urban (Mixon, 1992; Mak and Moncur, 2003; Cooke and Boyle, 2011).

We also posit that although students are more likely to relocate, financing should also lead to a surge in local students both in rural and urban areas. If so, financing may be fostering a brain-drain of students at the same time that it increases human capital formation locally in smaller rural areas. In this case, the effects of financing may be beneficial for regional development on balance, as pointed by Tight (2007); Storper and Scott (2009); Cooke and Boyle (2011).

3.1.3 What we do in this paper

To verify whether student financing increases the probability of migration among students out of rural locations, we implement probability models on the effects of tuition financing on students' decision to relocate. We use student-level data from the 2010-2019 Brazilian Higher Education Censuses which carry information on students' place of origin and of study, their participation in the financing scheme, and individual characteristics, to implement a logit model of the individual's choice to enrol at home or away from their place of origin. We find that student financing increased the students' propensity to migrate from small locations and national metropolises, and reduced the propensity to migrate for students of medium to large-sized cities.

But an increased probability of migration does not necessarily mean that students are more prone to relocate from smaller rural areas to larger urban ones. Using a multinomial logit regression, we model students' location choices given their decision to enroll. In practice, we therefore treat the location decision as sequential and conditional on a students' decision to enroll. We observe that students from smaller locations are in fact not likely to relocate to larger cities, but to other small cities, indicating that while student mobility is increased, the pattern is not of a brain drain from rural places. Moreover, students from larger urban locations are more likely to relocate to smaller cities given the financing, pointing to a reversed brain drain. These results lead us to reject the brain drain hypothesis.

In the second part of the paper we estimate the aggregate enrollment effects of financing to gauge its effects on the contingents of students, following the initial objectives of the expansion of the scheme. Arranging the dataset into panels, we construct city-level counts of student enrollments at home and away from home. We convert out financing treatment variable to the number of students supported by the scheme within a city. In this estimation, though, we must grapple with the issue of the simultaneity of enrolling and financing. To do so, we construct a measurement of the probability that a student from a particular city would be successful in her application for finance based on two interacted components: the proportion of the population in the city which is eligible to the financing scheme given its income-test requirement, and the total funds available for allocation under the program in a given year. We find that student financing increases the number of students relocating to other cities, as well as the number of students enrolling locally.

Finally, we estimate the effect of financing on higher education supply, more specifically, on the number of providers across geographical locations. Our results demonstrate that financing also increased the number of higher education providers in rural areas, especially in those cities without provision. This indicates that the policy was not only successful in delivering its objective of increasing the number of students where supply was available, but also on its objective of expanding the access to higher education where it was not previously available. This finding is a key conclusion of this paper because of its policy implications in face of the many proposals to promote the geographical dispersion of universities and subsidize relocation costs (Gibbons and Vignoles, 2012; Tight, 2007; Jepsen and Montgomery, 2009).

This paper is setup as follows. First, we examine the general setup of the student financing program in Section 3.3. In Section 3.4 we verify the patterns of individuals' propensity to migrate as well as the aggregate mobility patterns, and we offer a more detailed description of the data in our two data-sets. Our empirical strategy to identify the effects of financing at the individual and at the city-levels is presented in 3.5 followed by the presentation of our results in Section 3.6. The final section 3.7 concludes.

3.2 Literature

Student financing is a major part of the higher education landscape in most developed economies. In the UK, the US and Europe, a large proportion of undergraduate students are funded by a financing scheme to cover their tuition fees, and Brazil is none different. However a study on whether and how student financing affects students' relocation decisions is still missing. Some previous literature explored whether state-level financial aid affects the retention of students (see Mixon (1992) and Tuckman (1970) for studies in the U.S.), but the effects of the much more ample and far reaching tuition fee loans on student flows is still to be investigated.

The migration literature is ample and far reaching, with a significant proportion of works dedicated to the causes and effects of international brain drain, a specially attractive topic for development economists concerned with the relation of human capital movement and the income gap between developed and developing countries. Human capital flows within countries have received less attention, despite most of migration taking place within borders, and the important governance consequences of regional inequalities.[1]

The literature on national migration determinants, although not as prolific as the international brain-drain canon, is also very robust. Some of the earliest empirical studies of determinants of regional migration include Hagerstrand (1957), employing the earlier gravity models in Stouffer (1940) and Stewart (1948).

Following Stouffer (1940); Stewart (1948); Hagerstrand (1957), a number of papers employed these gravity models, which consist of fitting multiple regressions of rates of migration between paired locations as functions of their characteristics such as education and employment levels, the distances that separate them, among other variables. In fact the 1960s and 1970s saw a number of research papers employing place-to-place migration models on census data in different countries and population strata. Seminal to this stream of literature is the contribution of Levy and Wadycki (1974) implementing seemingly unrelated regressions to a stratified sample for education levels in Venezuela, Beals et al. (1967) on the effect of income in inter-regional migration in Ghana and Kono and Shio (1965) on the inter-municipal

[1]See (Giannoccolo, 2009) for a comprehensive literature review on the topic of brain drain, specially focused on the international dimension of human capital transfer from developing to developed countries, its causes and consequences.

migration in Japan, Schultz (1971) on the rural-urban migration in Colombia and Sahota (1968) on the labour migration between agricultural and industrial locations. In a similar fashion,Schultz (1982), estimated a logit model of migration probabilities for Venezuelan regions in an early precedent to our initial models at individual level.

A few researchers have previously examined the determinants of national student migration, starting with the seminal work of Tuckman (1970). The earlier works in this body of literature, though, are limited to the U.S. interstate mobility with data aggregated at the state level (see Tuckman (1970); Mak and Moncur (2003); McHugh and Morgan (1984); Mixon (1992)), with a few papers also dedicated to student mobility within specific states (see Alm and Winters (2009)). While Tuckman (1970) is specifically interested in the importance of fees and college capacity as determinants of inter-state migration, we are interested in the effects of student financing on migration flows between cities so that we can examine whether student financing is increasing the divide between rural and urban areas. Though Tuckman (1970) also included a control for state-aid available to students in their model of outmigration, because their research was more concerned with the effects of tuition fees in the state of origin than with the effects of aid, no attempt was made to refine the state-aid control variable. Unsurprisingly, Tuckman (1970) did not find any significant effects of aid on the number of students migrating from a U.S. state.

Variations of Tuckman (1970) have extended the empiric research of student migration, especially the more recent papers of Mixon (1992) including destination characteristics, Mak and Moncur (2003) adding other explanatory factors on the place of origin and Alm and Winters (2009) performing a state-level study on the importance of travel distances in college migration in the state of Georgia. A few papers have also examined the effects of policy interventions on internal human capital flows in the U.S. Murphy (1969) looked at the statistical evidence supporting the view that qualified scientists left the U.S. Midwest for locations where federal funds for research and development were more concentrated. Murphy concluded that the Midwest suffered a net outflow of high-skilled workers which was partly due to public intervention.

In a slightly different exercise, Gibbons and Vignoles (2012) analyse the determinants of higher education participation. They find that the geographical distance between the place of study and of origin has little or no impact on the decision to participate, although it does have a strong influence on the institutional choice.

The present paper differs from the existing literature in a few important points. Because our dataset is a collection of all yearly higher education freshmen over a 10 year-period so that we can, using a few assumptions, observe the yearly changes in the probabilities and contingents of migration given a policy treatment.Also, this is the first paper to analyze city-to-city migration, where the previous literature has used state-to-state models of migration.

Furthermore, the dataset employed in this paper is a very large census of all individuals enrolling in higher education, which allows significant flexibility in the choice of modelling alternatives and confidence in the results. We will explore this by using a collection of modelling strategies to construct a complete picture of the effects of financing on student migration across rural and urban settings.

3.3 The tuition financing scheme

The Student Financing Fund (FIES) is a public financing scheme designed to expand access to classroom-based higher education by financing undergraduate students in private institutions. Initially set up in 1975, FIES became an important feature of Brazilian higher education after its reform in 2010, when the funds available were substantially expanded and annual interest rates were cut from 6.5% to 3.4% pa. The post reform period saw the number of new students joining the financing program grow tenfold from 75,754 in 2010 to 732,700 in 2014.

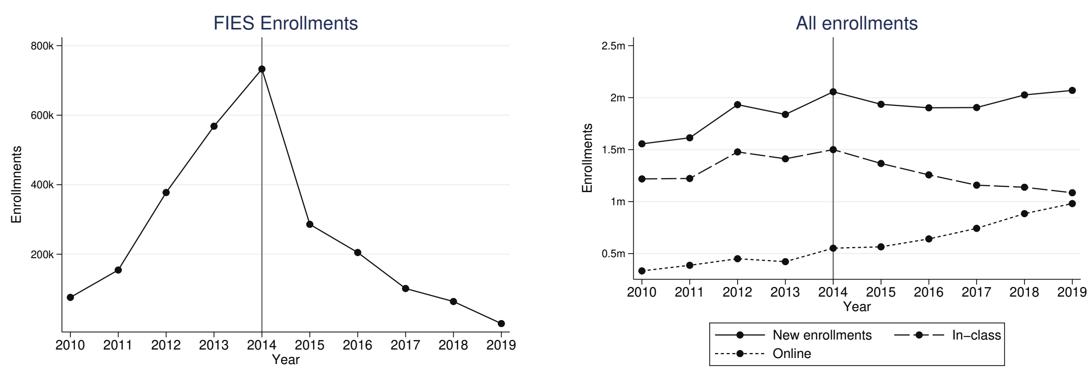
Between 2010 and 2014, FIES became a dominant element of the private higher education environment. In 2014, for instance, over half of all in-class freshmen students joined higher education under the scheme. From 2011 to 2012, new contracts surged by nearly 200 thousand p.a., which is roughly equivalent to the growth in classroom-based enrollments in the same period of time.

In June 2015, the government announced a reduction of the eligible household income threshold for new students [2] which in practice halved the maximum threshold. The result was a sharp drop in the number of new students, dropping in 2016 to only 28% of two years before.

Figure 3.1: New FIES contracts and enrollments

(a) New contracts

(b) Freshmen enrollments



The pattern of expansion and contraction of FIES in 3.1a finds a similar pattern of surge and fall in classroom-based enrollments in 3.1b. We will later attempt to

[2]Portaria Normativa N 10, de 31 de Julho de 2015

identify whether these correlations are due to factors external to FIES or due to an effect of the financing scheme. Our main interest, though, is to evaluate the differential effects of the scheme on locations small and large.

3.4 Data

In this section we present the two data-sets employed in this investigation as well as the descriptive statistics of all variables.

3.4.1 Student-level data

The first dataset used in this paper was constructed out of student-level data from the 2010-2019 Higher Education Censuses[3], which is the period we observe the rise and fall of the financing program FIES. The Higher Education Census is carried out yearly by the National Institute for Educational Research (INEP) under the Brazilian Ministry of Education.

Our main dataset consists of 8.2 million classroom-based freshmen out of the 12.8 million students in the original dataset. We kept in our dataset only those students from cities where higher education supply was available. In cities where there is no provision, all students must be enrolled away from their place of origin by construction. Also, because of missing data, 3.9 million observations were dropped from the sample. This was an issue which raised some concern regarding any systematic differences between students who were reported versus those who were not. To address this concern, we present the descriptive statistics of the full sample in the appendix, which are not significantly different from the ones below.

[3]Our dataset starts in 2010 because for the years previous to 2010 the city of origin of students is not available.

Table 3.1: Descriptive statistics of student-level variables

		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Migrate (d)	Mean	.38	.37	.38	.37	.37	.35	.37	.38	.39	.39
	Std Dev	.49	.48	.49	.48	.48	.48	.48	.49	.49	.49
	Min	0	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1	1
FIES (d)	Mean	.08	.12	.22	.25	.32	.2	.14	.12	.1	.09
	Std Dev	.27	.32	.41	.43	.47	.4	.35	.32	.3	.29
	Min	0	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1	1
Age	Mean	25	25	25	25	25	25	24	25	25	25
	Std Dev	7.4	7.4	7.3	7.4	7.3	7.4	7.4	7.5	7.8	8.1
	Min	12	13	13	13	13	15	14	14	14	15
	Max	92	100	88	91	92	84	88	86	88	92
White (d)	Mean	.27	.26	.27	.26	.34	.39	.42	.43	.39	.43
	Std Dev	.44	.44	.44	.44	.47	.49	.49	.5	.49	.49
	Min	0	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1	1
Female (d)	Mean	.55	.55	.55	.55	.55	.54	.54	.55	.56	.57
	Std Dev	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
	Min	0	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1	1
Public education (d)	Mean	.6	.61	.62	.65	.64	.6	.62	.64	.73	.74
	Std Dev	.49	.49	.48	.48	.48	.49	.49	.48	.44	.44
	Min	0	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1	1
Population origin (log)	Mean	13.33	13.44	13.45	13.59	13.58	13.59	13.60	13.56	13.50	13.58
	Std Dev	1.77	1.78	1.76	1.75	1.73	1.71	1.72	1.72	1.62	1.66
	Min	8.8	7.9	8.4	8.8	8.8	8.1	8.1	8.9	8.1	8.1
	Max	16.22	7 16.24	16.24	16.28	16.29	16.29	16.30	16.30	16.31	16.32
	Obs	288,699	320,193	773,329	1,163,889	1,236,180	1,137,571	1,043,836	957,594	512,737	528,886

From the roughly 8 million observations kept in our sample, we see that about half of them are enrolling in locations different from their place of origin and 21% of all students have signed up to the financing program. Students' mean age is 25 in the year they join university, the younger student in our sample is 12 years old, and the oldest is 100 years old. Both numbers have to be approached with suspicion. Our initial idea was to drop from the sample students who were unreasonably young or old, but since that would require us to define subjectively what we consider to be too young or old while the effect in our overall result is negligible, we decided to keep all the observations.

The majority of the students in the sample have been schooled in public institutions. According to the Ministry of Education, 87.4% of high-school students in

Brazil were enrolled in public schools. The fact that they represent only roughly two-thirds of students in tertiary education is consistent with the fact that students from public schools are less likely to progress into higher education than those of private.

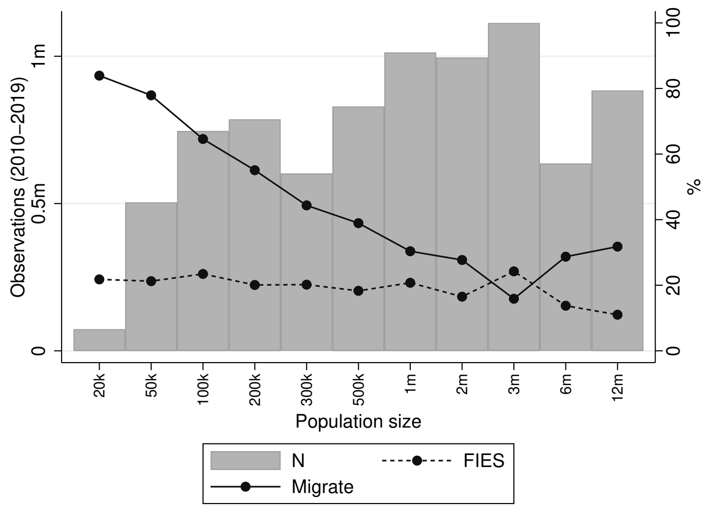
In terms of race, roughly one-third of our subjects identified as white, in contrast with 47.7% of the total population in the 2010 population Census. It is unlikely that whites are underrepresented in higher education (the contrary is more likely given the correlation of race and income). This could potentially be explained because of the specific incentives for students to declare themselves non-white in higher education: several programs to promote the inclusion of minority students consider race as a criterion for scholarships, fee discounts and even access to higher education.

In the case of gender, around 55% of students in our sample were recorded as female, a proportion similar to that of the general population (52.2% in 2019). The case of over-representation here is more realistic although not very pronounced, since there are little incentives for misreporting and the fact that the phenomenon of the more-than-proportionate presence of women in higher education is well documented (see Broecke and Hamed (2008) for an interesting report in gender gaps in participation for the UK).

The mean of the population size of students' place of origin tells us that larger cities have disproportionately more students enrolled in higher education. Out of the 5,565 cities in Brazil, only seven cities were of more than 2 million people, which represented 15% of the national population. Even so, 50% of the students in our sample were original from those cities. This points to the large gap that is already existing in the access to higher education between rural and urban locations.

In the graphs below we display the patterns of our dependent variable and our treatment variable in the student-level regressions. This is important because it will give us an idea of what our data looks like and the functional form we will have to adopt to better fit the data.

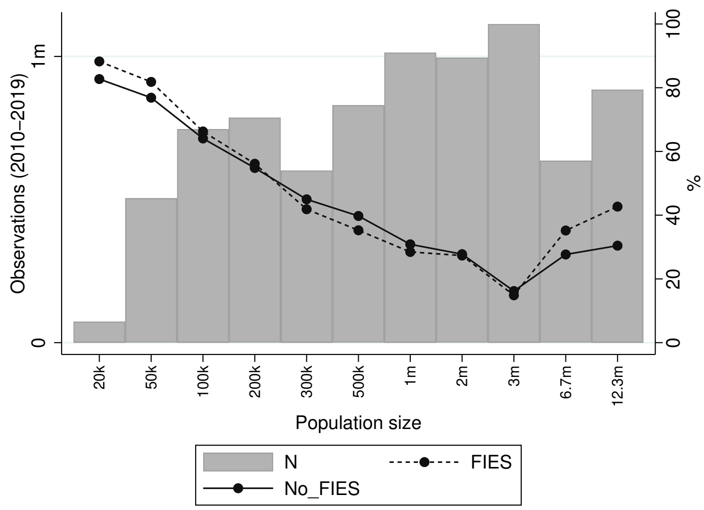
Figure 3.2: Observations, % of FIES and migration by city size (all years)



In Figure 3.2 above, we observe that the probability of migration is inversely proportional to the size of the city as per our hypotheses. Students from cities with population under 20,000 have almost certain probability of studying away. This is because those locations rarely actually offer higher education, so students who want to undertake higher education have no choice but to study away from home. As cities become larger, though, higher education becomes progressively more available at home and students are less likely to study away from home. That is the case until we reach the two last city categories. At that level, though, we must warn the reader that there is only Rio de Janeiro has a population of 6.7 million, and only Sao Paulo has a population of 12.3 million. The two last categories of cities, therefore, are not categories but in fact two specific cities which are very different from the rest.

In terms of students recipients of the loan financing scheme, the proportion is somewhat constant across city sizes. Roughly 20% of students benefit from the program regardless of the size of the city. Observing the proportion of students with and without FIES who migrate in Figure 3.3, we find a small difference across city sizes. In smaller cities, students under the financing scheme are marginally more likely to migrate than those not under the scheme. For larger cities, though, we observe the opposite: students with financing are then less likely to migrate. This trend continues until we reach Rio and Sao Paulo, where students under the program appear to be more likely to migrate.

Figure 3.3: Left scale: frequency; right scale:% of FIES and % migration; horizontal axis: city size (log scale; all years)



The overall pattern of the migration probability in these graphs is a key determinant of the functional form we adopt to estimate the models described in Equation 3.1. The models as presented are a good fit for a linear change in the individual migration propensity in population. Although some degree of non-linearity is allowed in our student-level models since we are estimating non-parametric logit models, the general functional form of the model specification should allow for the increase of the migration probability given finance in cities with population larger than 3 million. This is why we introduce the squared term of the population variable in our estimations.

To adjust for the two larger cities being very detached from the rest of the distribution, instead of using the population variable in its natural form, we instead included its log transformation to reduce the gap between these two outliers and the rest of the cities. The alternative to this would have been to drop the observation in these very large cities and to estimate linear models for those cities separately. We also present these alternative model specifications in the Appendix.

Although these graphs show only small differences between treated and non-treated students, we still do not have an overall view of whether the presence of the policy is in fact causing a systemic change in the propensities to migrate for all students. This we will examine in the city-level investigations. This brings us to the central issue which we aim to address in our city-level models. Given its sheer scale, it is reasonable to expect FIES to have implications on the general patterns of enrollments.

3.4.2 City-level data

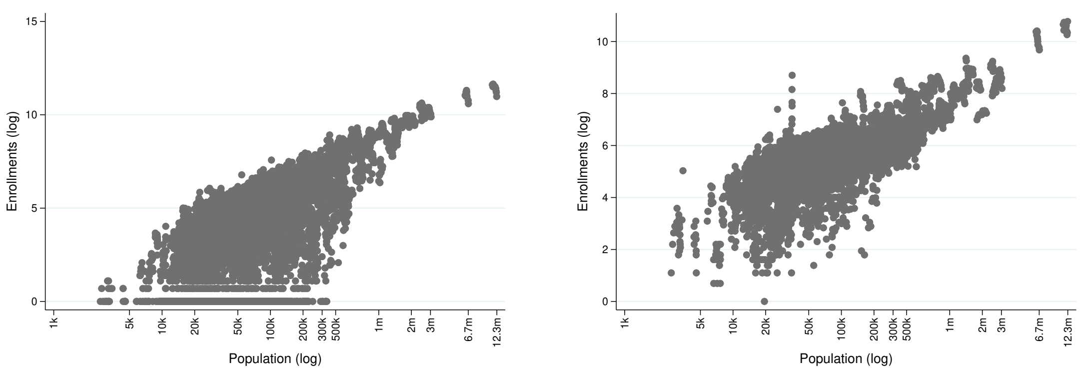
From the student-level dataset we aggregated the data into the 5,570 Brazilian municipalities based on the 7-digit city codes obtained from the Brazilian statistics bureau (IBGE), and once again, we kept only the 871 municipalities where we observed provision of higher education. To each city-year combination we then added information on the number of freshmen students from that city who have enrolled in their place of birth and away, the number of students who signed up to the financing program, and the total enrollments in online and classroom-based education. The number of observations in our study is therefore 8,710, or 871 municipalities over 10 years.

This first graph in Figure 3.4 below is a plot of classroom-based enrollments for all years. Because of the skewness of the city-level observations in populations sizes and, consequently, enrollments, we have used the log-transformation of those variables to approximate these distributions to the normal.

To allow for a log-transformation we added 1 student to all locations with no provision of higher education - which would therefore observe 0 home enrollments. This is why we observe a number of municipalities - some of them rather large - lined up along the horizontal axis of the graph on the left.

Figure 3.4: Classroom-based enrollments (all years)

(a) Home (b) Away from home

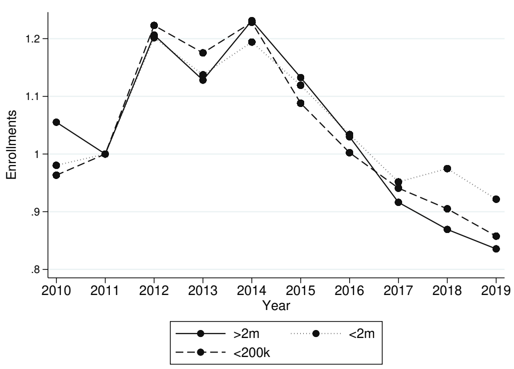


Differently from the student-level models, these graphs points us to a more linear relationship between log-enrollments and log-population, which is the reason behind why we will not adopt the squared-term of the population variable in this investigation or its interactions with the treatment variable.

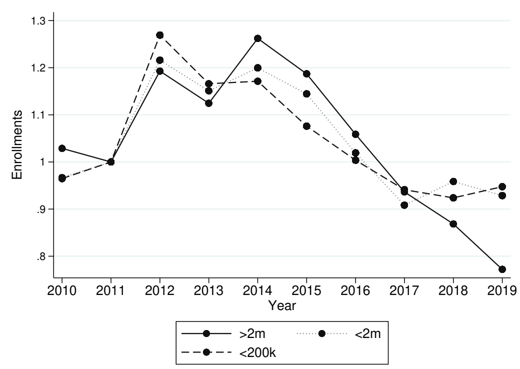
Now looking at the effect of the scheme over time, we observe the patterns of enrollments of students at home and away in larger and smaller cities. We see that both enrollments at home and away expand and contract alongside with FIES. For large cities, though, it is the number of home enrollments which is more visibly correlated with FIES fluctuations, while for smaller cities, it's the number of students enrolled away from home.

Figure 3.5: Enrollments by city type at home and away (2011 baseline)

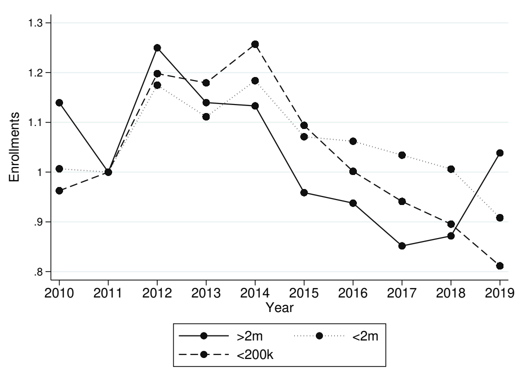
(a) All enrollments



(b) Enrollments at home



(c) Enrollments away from home



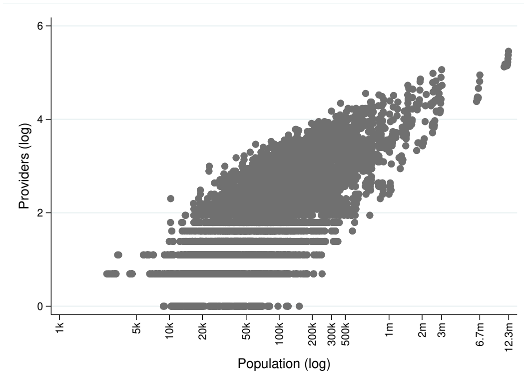
Regarding our main treatment variable for this section, although the Higher Education Census identifies individually those who are financially supported by FIES, there is a discrepancy between the total number of yearly new contracts from the Census and from the FIES operator FNDE. This discrepancy occurs due to providers' failure to report FIES students in the Census. For the sake of accuracy, we constructed the numbers of FIES contracts in each city directly from the FIES operator database instead of adding up the individual enrollments from the student-level dataset. This did not alter the results in any significant way, though the magnitude of the effect using the operator's database is slightly more conservative.

Table 3.2: Descriptive statistics of city-level variables

		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
All enrollments (log)	Mean	5.8	5.8	6	6	6.1	6	5.9	5.8	5.8	5.8
	Std Dev	1.4	1.4	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
	Min	1.1	1.6	1.6	1.6	1.6	1.4	1.6	0	1.1	.69
	Max	12	12	12	12	12	12	12	12	12	12
Home enrollments (log)	Mean	3.8	3.8	4.1	4	4	3.9	3.9	3.9	3.9	3.9
	Std Dev	2.5	2.5	2.5	2.5	2.6	2.6	2.5	2.4	2.5	2.5
	Min	0	0	0	0	0	0	0	0	0	0
	Max	12	12	12	12	11	11	11	11	11	11
Enrollments away (log)	Mean	5.4	5.4	5.6	5.6	5.7	5.6	5.5	5.4	5.4	5.3
	Std Dev	1.1	1.1	1.1	1	1	1	1	1	1.1	1.1
	Min	.69	1.1	1.6	1.6	.69	1.1	1.6	0	1.1	.69
	Max	11	10	11	11	11	10	10	10	10	11
FIES (log)	Mean	3	3.7	4.5	4.9	5.2	4.3	4.1	3.4	2.9	.2
	Std Dev	1.5	1.5	1.6	1.6	1.5	1.5	1.4	1.4	1.4	.45
	Min	0	0	0	0	.69	0	.69	0	0	0
	Max	8.5	8.9	10	11	11	9.7	8.9	8	7.5	2.9
Providers (log)	Mean	1.5	1.6	1.6	1.6	1.6	1.6	1.6	2.1	2.4	2.7
	Std Dev	.8	.82	.81	.81	.8	.81	.83	.87	.85	.86
	Min	0	0	0	0	0	0	0	0	0	0
	Max	5.1	5.1	5.2	5.2	5.1	5.2	5.2	5.3	5.4	5.5
Class-based access (d)	Mean	.81	.81	.83	.82	.8	.8	.81	.82	.83	.84
	Std Dev	.39	.39	.38	.39	.4	.4	.39	.38	.37	.37
	Min	0	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1	1
Online access (d)	Mean	.72	.71	.71	.73	.73	.72	.76	.81	.89	.92
	Std Dev	.45	.45	.45	.45	.44	.45	.43	.39	.32	.27
	Min	0	0	0	0	0	0	0	0	0	0
	Max	1	1	1	1	1	1	1	1	1	1
Population (log)	Mean	11	11	11	11	11	11	11	11	11	11
	Std Dev	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
	Min	7.9	7.9	7.9	8	8	8.1	8.1	8.1	8.1	8.1
	Max	16	16	16	16	16	16	16	16	16	16
	Obs	871	871	871	871	871	871	871	871	871	871

The descriptive statistics of our city-level variables give us a few insights into our data. First of all, enrollments away from home are typically more numerous than enrollments at home, although naturally the enrollments at home has a larger standard deviation to reflect that many locations have no provision of higher education and thus, home enrollments are zero. Standard deviations are in fact high for all variables, which indicates the high level of heterogeneity among cities in the sample. The maximum of all variables pertain to Sao Paulo.

Figure 3.6: Number of providers (all years)

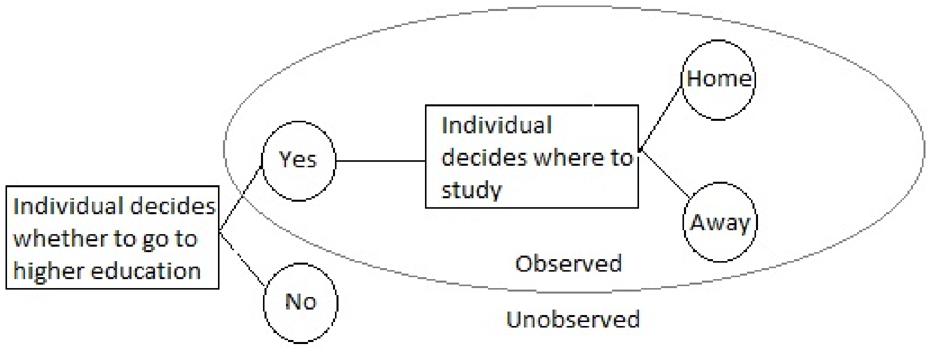


3.5 Empirical strategy

We treat the enrollment and location decisions as sequentially determined. First, students decide whether they will enroll in higher education. Second, conditional on their decision to enroll, students decide upon their location and university. Because we are interested on students' decisions to stay at home or migrate we only estimate the second part of the decision process, which is conditional on the first.

This is convenient in our case as we do not observe students which do not participate in higher education - our dataset only includes those who have enrolled at a higher education provider.

Figure 3.7: Student's decision tree



Because we do not observe from the data those students who did not enroll in higher education, our model is in fact capturing the decision to migrate given the decision to enroll. But the effects of the financing scheme certainly impact the decision to enroll as well. We can, in fact, separate the effects of the scheme in two parts: an enrollment effect and a relocation effect. The enrollment effect is the impact of the availability of financing on the probability that a student will enroll in higher education. The relocation effect is the change in the probability that a student will enroll away from home as opposed to at home. In this chapter, we are interested in the second of those.

To answer our first question of whether financing increased the likelihood of migration disproportionately among student from smaller cities we estimate a logit model of students' propensity to migrate from cities of different population sizes for recipients and non-recipients of student financing to determine the effect of the scheme on student's decisions. Similarly to Schultz (1982), we used a binary 'migrate or stay' outcome variable in a logit model along with local and individual controls to model the students' decision to migrate. We also examine the differential effect of finance in gender, race, age and on those previously educated in private vs public schools.

But since the probability of migration across city sized tells us little about the patterns of migration, we implemented a multinomial logit model of discrete choice of destination. This approach has a precedent in Krieg (1993) where the author investigates the influence of race and education on choices of migration destinations using American Census micro-data. Also in De Jong (2000), where the author fits a multinomial logit model to a large sample of individuals in Thailand to explain determinants of permanent and temporary migration, and of intentions to migrate.[4]

To capture the aggregate effects of the policy on enrollments at the city-level, though, we restructured our data into a panel of municipalities and constructed measures of aggregate enrollments for each location, in a similar strategy to that of Tuckman (1970); Alm and Winters (2009); Mak and Moncur (2003); McHugh and Morgan (1984); Mixon (1992) and others.

3.5.1 Individual relocation choices

To answer our first research question we estimate a logit model of the differential effect of financing on individual propensity to migrate for students of different city sizes. To capture whether students have relocated, we used students' places of birth and the location of their higher education provider to construct a binary variable M_{ict} to indicate whether a student i from a city c is enrolled away from their place of origin in period t.

Just as we observe students who have enrolled away from their places of origin, we can also observe from the data which students participate in the financing scheme.

[4]A comparison of the explanatory power of multinomial probit models at individual level and the previously employed place-to-place models using aggregate data can be found in Kanaroglou and Ferguson (1998)

Using a binary treatment variable $FIES_i$, therefore, we will try to establish whether those enrolled in the financing program are more or less likely to have enrolled away from home.

The first caveat from our treatment variable is whether it is capturing the effects of the financing scheme or some other confounding effect. The first of these possible confounding effects we must consider is that obtaining financing is itself a proxy for income. Students in families of higher income may be less likely to obtain financing either because they do not need it or because their income exceeds the requirements of the scheme. These students may be more likely to relocate to larger cities because they have the means to do so regardless of financing. Similarly, those of lower income may be more likely to seek and obtain financing, and less likely to migrate. In this case, failing to control for students' household income would lead to an omitted variable bias in the estimated treatment effect. Unfortunately we do not have information on income from the dataset, so we have used the proxies for income we had available such as a binary indicator W_i for whether the student is white $W_i = 1$, $W_i = 0$ if otherwise, and PS_i is a binary indicator for whether the student was schooled in a public institution $PS_i = 1$, $PS_i = 0$ if private.

This concern could be alleviated if the income threshold of the program were restrictive enough to limit access to a homogeneous income group which we then could define as our sample, but that is not the case. In fact, according to our calculations, 90% of the population of small cities were eligible to FIES in 2014, and 65% for large metropolitan cities (see Table 3.4). Overall, we believe that biases stemming from this source are, if not entirely controlled for, not particularly concerning because they would in fact lead to more conservative estimates of our effects and therefore, are unlikely to be the main driver of our results.

To help address our concerns that our outcome variable will be affected by relocation decisions both related and unrelated to higher education we included A_i the age of the student in our model. The assumption here is that the probability that a student will relocate for any reasons outside of their higher education choice is cumulative on age e.g. a 30 year old is more likely to have changed location over the course of their life than a 20 year old. We would, therefore, like to identify departures from that normal pattern of probabilities of migration over age, and whether taking part on the scheme has any differential effects on that pattern. We also include a dummy F_i for whether the individual is female $F_i = 1$, $F_i = 0$ if male.

Finally, we define P_{ct} as the population of the students' city of origin as a continuous measure of city size in period t. This variable is a control for the inherent probability that a student will move away from a city given its size. Smaller cities, in general, offer less opportunities of higher education to students and less choices of institutions and programs and therefore, smaller cities are likely associated with a higher probability of moving away. To capture whether financing has a differential effect across city sizes, we include in our model specification the interaction of our finance treatment variable with the population size. Our results therefore will be determined jointly by the FIES treatment and its interaction with population size.

To control for fluctuations in the contingents of students enrolling at home and away which are uncorrelated with the policy, we include time-fixed effects ϕ_t in all of our models. We do not include city-fixed effects in our regressions because as our population variable is in fact a proxy for all the characteristics of cities that come along with city size, city-fixed effects would serve to capture the variance in the outcome variable which we would like to be captured in the population variable.[5]

[5] We include in the Appendix alternative model specifications using city fixed effects and city-year fixed effects to resolve any concerns that this strategy is not sufficient to account for enrollment trends over time which differ across cities.

We estimate our student-level models using a logit model where our dependent variable is a binary variable M_i for whether the student has enrolled in a university away from her place of origin. So our model specification is the following:

$$
\begin{aligned}
P(M_{ict} = 1|x) = G(&\beta_1 + \beta_2 FIES_i + \beta_3 FIES_i * \log(P_{ct}) + \beta_3 \log(P_{ct}) \\
&+\beta_4 PS_i + \beta_5 A_i + \beta_5 W_i + \beta_5 F_i + \phi_t + \varepsilon_{it})
\end{aligned} \tag{3.1}
$$

We are mostly interested in the differential effects of $FIES_i$ on the probability of relocation for students from cities of different population sizes P_{ct}, which is captured by the interaction term $FIES_i$* P_{ct}.

Although our logit model can help us determine how financing affects the propensity to migrate across locations, we still cannot claim that there is a brain drain because we do not know about the destination of those students. Even if students from certain locations are more likely to relocate due to the policy, the policy might still be neutral from the human capital perspective if it is also receiving students from other locations.

To determine the effect of the financing program on students' choices of location, we estimate a multinomial logit model similar to the logit specified above but now replacing the binary dependent variable with a categorical variable $P(M_{ict} = m|x)$ where m is a choice set of the migration alternatives a student is faced with.

$m = \{0, 1, 2, 3\}$ such that

$$\begin{cases} m(_i) = 0, \text{ stay} \\ m(_i) = 1, \text{ migrate to a city with population <200k} \\ m(_i) = 2, \text{ migrate to a city with 5m>population>200k} \\ m(_i) = 3, \text{ migrate to a city with population>5m} \end{cases}$$

So that our multinomial-logit model specification becomes:

$$\begin{aligned} P(M_{ict} = m|x) = G(&\beta_1 + \beta_2 FIES_i + \beta_3 FIES_i * \log(P_{ct}) + \beta_3 log(P_{ct}) \\ &+ \beta_4 PS_i + \beta_5 A_i + \beta_5 W_i + \beta_5 F_i + \phi_t + \varepsilon_{it}) \end{aligned} \tag{3.2}$$

The interaction term in this model now should indicate the differential effects of receiving financing on the probability that students from different city sizes will relocate to cities of specific city sizes. To facilitate interpretation, we will include the marginal effects following the regression results.

3.5.2 Aggregate effects at city-level

In Section 3.6.1 we focused on the effect of financing on the probability that a student will migrate given the size of its place of origin. In this Section we gauge the aggregate effects of the policy on the student population in cities small and large. Our objective is to now determine whether the size of the student contingents is also changing due to the policy. In other words, while the first part of the empirical strategy was dedicated to determining the effects of financing on the proportions of students migrating, this section is focused on gauging the effects on the number of students enrolling in and away from their places of origin.

Because the dependent variable now is a geographical and time-variant continuous variable, we now use a linear model with a simpler specification. Our dependent variable here is defined generically as E_{ct} and denote counts of enrollments at home and away from home for each city-year combination. We use this more generic notation to avoid repeating the specifications for models which are otherwise identical.

By extension, our treatment variable $FIES_{ct}$ will now be the the number of students in a city-year which are supported by the program. Similarly to the individual models, our variables of interest here are both $FIES_{ct}$ and its interaction term with the continuous measure of population size P_{ct}, which should capture the differential effects of the program by city sizes. We also include time-fixed effect T_t as per the individual models. So our model specification is the following:

$$\begin{aligned}\log(E_{ct}) = \beta_1 + \beta_2 \log(FIES_{ct}) + \beta_3 \log(FIES_{ct}) * \log(P_{ct}) + \beta_3 \log(P_{ct}) \\ + \beta_4 T_t + \varepsilon_{ct}\end{aligned} \tag{3.3}$$

Our aggregate models, though, have a methodological caveat: our treatment variable $FIES_{ct}$ is now endogenous. At the city-level we need a measure of FIES intensity to capture whether more financial support leads to more enrollments, and if it affects the patterns of at-home and away enrollments. The first response to this was to aggregate the binary FIES indicator to the city-level, but this leads to a tricky problem of simultaneity. At the individual level, we made an assumption that the FIES treatment was exogenous to the students' decision to relocate. But since we are now trying to model also the decision to enroll, obtaining finance is in fact endogenous to the enrollment decision because more financing leads to more enrollment, and more enrollment leads to more financing.

To address this issue, we have devised an exogenous city-level predictor of the number of students that participate in the financing scheme by interacting the yearly amount of funds made available for financing by the federal government for the incoming cohort, and the city-level share of the population eligible to the scheme given the income threshold. This instrument satisfies both the relevance criterion, since it has a direct effect on the number of students which obtain financing in a given city; and the exclusion restriction as it does not have a direct effect on the student numbers except through the financing itself and is exogenous to students' decisions.

3.5.3 Instrument

As we have previously established, any counts of student enrollments aggregated at the city-level are endogenous with financing. To address that, in our regressions at the city-level we will instrument the number of students receiving financing to address that endogeneity. This instrument, which can be interpreted as the probability that a student from a given city will obtain financing, is in fact an interaction of two exogenous determinants. The first of them is the yearly budget allocated to financing new students by the central government. We can see that the allocation of new funds from Table 3.3 follows the same patterns we observe in the total number of new students enrolled in the scheme in Figure 3.1. The shortcoming here is that although this gives us a nice measure of how likely a student is to get financing over time, this does not tell us anything about local heterogeneity.

Table 3.3: Yearly income thresholds for access and new fund allocation

Year	Income per family (m.w.e.)	Income per capita (m.w.e.)	Fund expansion (BRL billion)
2010	20	-	1.18
2011	20	-	1.04
2012	20	-	5.15
2013	20	-	5.79
2014	20	-	10.66
2015	-	2.5	4.87
2016	-	2.5	1.29
2017	-	2.5	0.12
2018	-	3	-0.04
2019	-	3	-0.13

To make our instrument a city-level predictor, we interacted the new funds allocated with the estimated proportion of the local population eligible to the scheme, taking into account the local distribution of income and that the income test for eligibility changed along the period covered in our data. This gives us an exogenous instrument which, just like the number of FIES students, varies in time and across locations.[6]

Table 3.4: Proportion of the population eligible to FIES in 2014

Population Size	Obs	Population eligible(%)
<20k	887	.88
<50k	2,691	.87
<100k	2,290	.86
<200k	1,410	.85
<300k	573	.83
<500k	448	.82
<1m	245	.8
<2m	99	.79
<3m	46	.8
<7m	10	.72
10m	10	.71
Total	8,710	.86

So our first stage specifications for both the endogenous FIES headcount per city

[6]For a detailed step-by-step explanation of how we constructed this instrument refer to Section 3.8.2 in the Appendix.

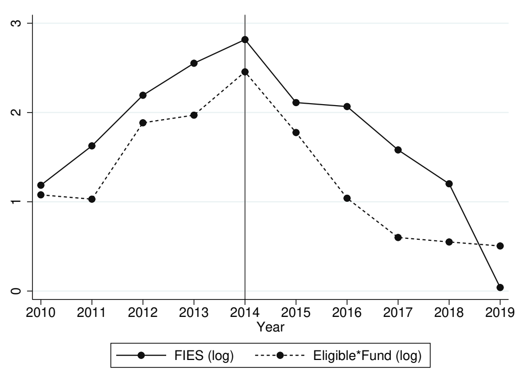
and its interaction with the exogenous population variable are as follows:

$$\log(FIES_{ct}) = \beta_1 + \beta_2 \log(Z_{ct}) + \beta_3 \log(Z_{ct}) * \log(P_{ct}) + \beta_4 T_t + \varepsilon_{ct} \tag{3.4}$$

$$\begin{aligned}\log(FIES_{ct}) * \log(P_{ct}) = \beta_1 + \beta_2 \log(Z_{ct}) + \beta_3 \log(Z_{ct}) * \log(P_{ct}) \\ + \beta_4 T_t + \varepsilon_{ct}\end{aligned} \tag{3.5}$$

To display how our instrument compares to the actual number of new FIES students year on year, the graph in Figure 3.8 shows the aggregate pattern of our instrument versus the instrumented variable.

Figure 3.8: Mean instrument over time



3.5.4 Instrument validity

As previously exposed, our instrument expresses the probability that a student from a given city will obtain financing through the FIES programme. This is exogenously determined by the central government as a combination of two factors: the budget allocation for the programme in a given year and the income threshold. The income threshold is used in combination with a city's distribution of income to determine the probability that a student from a given city will be under the threshold.

This instrument was devised to satisfy all of the conditions for a valid instrument, as proposed by Angrist and Pischke (2009) We discuss our instrument in light of the conditions below:

i) Relevance: Our instrument is related to the endogenous variable of interest as we would expect a higher allocation of resources to the programme in combination with a higher eligibility to result in higher enrollments.

ii) Exogeneity: All of the components of our instrument are exogenous. The yearly budget is exogenously determined by central government, as well as the maximum income threshold.

iii) Monotonicity: Our instrument affects the endogenous variable in a monotonic fashion. In other words, we do not expect that an increase in budget and the maximum threshold would reduce enrollments in any situation.

iv) Exclusion restriction: The instrument must not have any direct effect on the outcome variable, other than through its effect on the endogenous variable. Since our instrument is fundamentally based on the rules of the financing programme FIES, we do not expect that changes in the rules of the programme would affect enrollments in higher education in any other way except through the programme itself.

3.5.5 First stages

Table 3.7 below presents the results from the first-stage regressions we will employ throughout our city-level estimations. Because in all of our city-level models we only change the outcome variables, the first-stage regressions are the same for all models.[7]

Table 3.5: First stages

	FIES	FIES*Pop
Z (log)	1.02***	-8.15***
	(0.218)	(2.54)
Z (log) x Population (log)	1.02***	3.78***
	(0.139)	(0.126)
Population (log)	0.81***	11.39
	(0.029)	(0.41)
Constant	-13.22***	-128.52***
	(0.297)	(5.12)
Observations	8,710	8,710
F-test of excl. inst	159.18	366.30
R-squared	0.81	0.84

Standard errors in parentheses
Time fixed-effects included not displayed
Standard errors clustered at municipal level
* p<0.10, ** p<0.05, *** p<0.010

Our instrument appears to be a good exogenous predictor for the city-level number of students financed by the scheme as shown by the F-test of the excluded instrument in both its interacted and non-interaction forms. The R-squared shows that the first-stage estimation has a good fit to the underlying data.

3.6 Results

In this section we present the results of the estimations from the student-level and the city-level regressions. We start with the student-level results and in the following subsection we move onto the city-level.

[7]Although the first-level regressions are the same for all city-level models, we do split the sample in our final section. The first-levels for those regressions are presented in the Appendix.

3.6.1 Effects of FIES on individual relocation choices

Table 3.6 below presents the results of the IV models expressed in Equation 3.1. These models include time fixed effects which are displayed in the full results in Table 3.16 of the Appendix.

In Column 1, we find that our binary treatment for whether a student has received financing has a small and significant negative effect on a student's propensity to migrate, once controlled for city, school and individual characteristics. This is an indication that the scheme has a negative mean effect on all students' propensity to migrate.

Once we include the squared term of the population control in Column 2, which in practice improves the fit of the population control for larger cities, we find a small positive and significant effect of FIES on the mean propensity to migrate.

In Column 3, we add the interaction term of the financing dummy with the population size to capture whether the program has a differential effect depending on the size of the city. We find a positive effect of financing on the student's propensity to migrate, but decreasing in population size. In this model, though, we have not included the squared term of the population variable.

In Column 4 we include the interaction of the treatment dummy with the log-population variable and with its squared term. In this model, which is the model of best-fit, we find that that the effect of FIES from Column 3 persists and is in fact much stronger for small cities, but reduces with the population size up until a certain threshold, then reverts. This is similar to what we observed in Figure 3.3.

Table 3.6: Individual propensity to migrate (logit) - IV

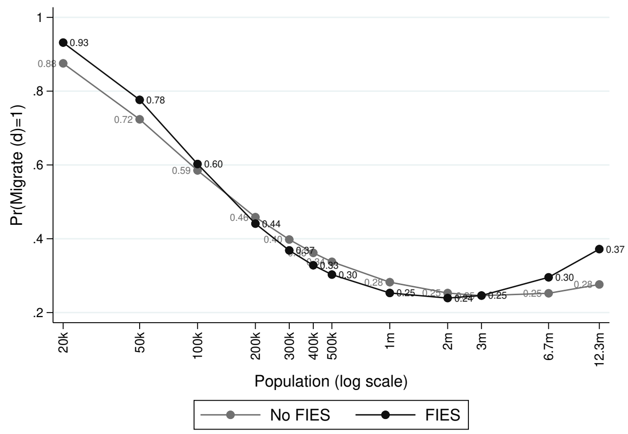
Migrate (d)	1	2	3	4
FIES (d)	-0.0081***	0.023***	0.67***	12.2***
	(0.002)	(0.002)	(0.017)	(0.136)
FIES (d) × Population (log)			-0.052***	-1.85***
			(0.001)	(0.020)
FIES (d) × Population (log)^2				0.069***
				(0.001)
Population (log)	-0.39***	-3.74***	-0.38***	-3.43***
	(0.000)	(0.007)	(0.001)	(0.008)
Population (log)^2		0.12***		0.11***
		(0.000)		(0.000)
Public schooling (d)	0.031***	-0.019***	0.031***	-0.020***
	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.019***	0.019***	0.019***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)
White (d)	0.15***	0.093***	0.15***	0.094***
	(0.002)	(0.002)	(0.002)	(0.002)
Female (d)	-0.0069***	-0.011***	-0.0067***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	4.15***	26.3***	4.03***	24.3***
	(0.008)	(0.050)	(0.008)	(0.055)
Observations	8183748	8183748	8183748	8183748

Standard errors in parentheses
Time-fixed effects included.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A Hausman (1978) test - with the null hypothesis that the specified endogenous regressors can actually be treated as exogenous - yielded a coefficient of 221.211, with a Chi-square P-value of 0.000, which confirms our suspicions that we should reject the null hypothesis and consider our regressors endogenous. For guidance on implementation of the Hausman test in an environment with fixed effects and clustered standard errors see Hayashi (2000).

Using the model in Column 4 we draw the predicted probability of migration for students from cities of different sizes.

Figure 3.9: Prediction the propensity to migrate for students with and without financing



The overall probability that a student will move away from her place of origin is very high for smaller cities, and financing increases this probability by 5% for students from cities with population of 20,000, and 6% for those from cities with population of 50,000 compared to students which are not financed by FIES. This is an indication that student financing has a differential effect on the propensity to migrate on students from cities of different population sizes.[8]

From that point on, as population increases, we find the opposite marginal effect: students with financing are less likely to relocate for populations above 200,000. This is further evidence of increased regional imbalances caused by student financing.

These results initially confirm our first hypothesis that student financing increases the probability of migration among students out of rural locations, but with a caveat. Following from the patterns we previously observed in Figure 3.3, we observe distinct marginal effects of the treatment in the two very large cities. We posit that because

[8]We did not include the confidence intervals in this graph because they were small and got confused with the graph projections. We do provide the confidence intervals for these projections in Table 3.17 in the Appendix

Sao Paulo and Rio contain such a disproportionate population compared to other cities, it could be the case that the surge in demand from FIES in those cities is overwhelming the supply and therefore students seek placement in higher education in smaller cities. To answer that we will test whether students from large cities are migrating to smaller cities in our multinomial logit model in the following section, and also how the number of providers responds to the financing program in the final section of this paper.

Although we have so far used the terms migration and relocation interchangeably, deducing that students have migrated because they enrolled away from their place of origin can be inaccurate. This is because students may commute to a different city to study and return to their hometowns on a daily basis if these are not too distant. In our case, our outcome variable would not be able to differentiate these students from those who have actually moved away from home for studying.

As a robustness check, we have included two alternative specifications of our main model in Column 4 in the Appendix, where our outcome variable is a dummy for whether students have enrolled at least 50 geographical km away from their place of origin and a second one for whether students enrolled away from their home region. We also include the same model run in OLS because of usual concerns about interaction terms in logit models, although such concerns are normally around the interpretation of the coefficients of interactions which is among the reasons why we calculate and display marginal effects.

As an alternative functional specification to the inclusion of the squared-log of population we also split the sample and run a separate model for those very large cities Sao Paulo and Rio. This is shown in Table 3.19 of the Appendix where we separate cities with population larger than and smaller than 5 million, effectively running a separate model just for the two largest cities. Results obtained in these models confirm what we can observe in our main regressions. We have also included in the Appendix a version of these logit models where we implement city dummies. Finally, we also included the OLS version of these models, and to also display meaningful R-squared statistics which reflect the best fit of our models as we change the specification across columns. Results of the OLS are shown in Table 3.20.

In the following models we now interact the financing treatment with other individual characteristics of the student.

Table 3.7: Individual propensity to migrate (logit)

Migrate (d)	1	2	3	4
FIES (d)	12.3***	12.2***	12.2***	0.040***
	(0.136)	(0.136)	(0.136)	(0.007)
FIES (d) × Population (log)	-1.87***	-1.84***	-1.85***	
	(0.020)	(0.020)	(0.020)	
FIES (d) × Population (log)^2	0.070***	0.069***	0.069***	
	(0.001)	(0.001)	(0.001)	
FIES (d) × Public schooling (d)	-0.047***			
	(0.004)			
FIES (d) × White (d)		0.018***		
		(0.004)		
FIES (d) × Female (d)			0.016***	
			(0.004)	
FIES (d) × Age				-0.00069*
				(0.000)
Public schooling (d)	-0.012***	-0.020***	-0.020***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.019***	0.019***	0.019***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)
White (d)	0.094***	0.091***	0.094***	0.093***
	(0.002)	(0.002)	(0.002)	(0.002)
Female (d)	-0.011***	-0.011***	-0.014***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)
Population (log)	-3.43***	-3.43***	-3.43***	-3.74***
	(0.008)	(0.008)	(0.008)	(0.007)
Population (log)^2)	0.11***	0.11***	0.11***	0.12***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	24.2***	24.3***	24.3***	26.3***
	(0.055)	(0.055)	(0.055)	(0.050)
Observations	8183748	8183748	8183748	8183748

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We find that the financing program has significant effects on the provision of opportunities of mobility for specific groups. Financing yields a larger effect on females compared to males. Financing, in this case, helps closing the opportunity gap that generally is observed between sexes, but it also benefits whites at a larger rate than other racial groups, and students from private schools more than those from public. Also, younger students, who are generally more mobile, are more likely to relocate given the treatment. In Table 3.8 below we display the predicted effect of financing on the probability that a student of certain characteristics will migrate keeping other factors constant.

Table 3.8: Marginal effects of FIES by characteristic

	Schooling	Race	Gender	Age
	Private	Non-white	Male	17
FIES (d)	0.0140***	0.0065***	0.0056***	0.0091***
	(0.001)	(0.000)	(0.001)	(0.001)
	0.0045***	0.010***	0.0090***	0.0074***
	(0.001)	(0.001)	(0.001)	(0.000)
	Public	White	Female	25

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The final model on this section in Table 3.9 is a multinomial logit of the decision of students. The dependent variable is a categorical variable representing four discrete choices: the baseline (0) is to stay in the city of origin, and the three alternative choices are discrete variations of the migration decision: to migrate to a small city (1), migrate to a medium-sized city (2), and to migrate to a large city (3). Although the model below is presented in three columns, these are in fact the results of one model only with each of the three alternative migration choices in each column.

Table 3.9: Multinomial logit model of student choice

Migrate:	to small <200k	to medium >200k	to large >5m
Baseline: stay at home			
FIES (d)	12.9***	10.5***	6.88***
	(0.178)	(0.147)	(0.564)
FIES (d) × Population (log)	-2.00***	-1.57***	-1.13***
	(0.027)	(0.022)	(0.086)
FIES (d) × Population (log)^2	0.077***	0.058***	0.042***
	(0.001)	(0.001)	(0.003)
Population (log)	-4.61***	-3.54***	2.63***
	(0.012)	(0.009)	(0.028)
Population (log)^2	0.15***	0.12***	-0.12***
	(0.000)	(0.000)	(0.001)
Age	0.0091***	0.021***	0.026***
	(0.000)	(0.000)	(0.000)
White (d)	0.32***	-0.016***	0.19***
	(0.003)	(0.002)	(0.004)
Female (d)	-0.0053*	-0.017***	0.011**
	(0.002)	(0.002)	(0.004)
Constant	31.5***	24.2***	-18.1***
	(0.076)	(0.059)	(0.184)
Observations		8183748	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

And the margins predictions which follows from the multinomial probit in the paper here were calculated out of the multinomial logit model above.

Table 3.10: Multinomial logit model of student choice: marginal effects of financing

Migrate:	to small <200k	to medium >200k	to large >5m
Baseline: stay at home			
20k	0.032***	0.00028	-0.0023***
	(0.003)	(0.003)	(0.000)
100k	0.011***	0.031***	-0.023***
	(0.001)	(0.001)	(0.000)
500k	-0.00054	-0.0096***	-0.026***
	(0.000)	(0.000)	(0.000)
1m	0.0017***	-0.012***	-0.020***
	(0.000)	(0.000)	(0.000)
2m	0.0069***	-0.0073***	-0.013***
	(0.000)	(0.000)	(0.000)
6.7m	0.030***	0.018***	-0.0037***
	(0.000)	(0.001)	(0.000)
12.3m	0.057***	0.041***	-0.0017***
	(0.001)	(0.001)	(0.000)
Observations		8183748	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These results of the marginal effects computation show that financing increases the probability that a student will migrate to a small city especially for students from very small cities and for those in large metropolises. These results also directly contradict our second hypothesis that those students from rural locations were more likely to relocate to larger cities. What we find instead is that those students are actually relocating to other rural areas. This is a strong indication that there is no brain-drain from smaller cities due to financing, just an increase in mobility among those cities. In fact, at this stage what we see is a brain drain in the opposite direction: from large to smaller cities.The question of whether these observed effects are due to an expansion of supply in smaller cities which is not observed in larger cities will be answered in our final model in this paper.

This is consistent with our conjecture that the increased migration probability from large to small cities occurs because of a surge in demand which is not met by an increase in supply. This hypothesis will be tested in the following sections where we model the effect of FIES on the geographically aggregated enrollments and the number of providers at the city-level.

On concerns about the IIA assumption of multinomial logit

Our choice of a multinomial logit model to estimate the individuals' mean propensity to migrate to locations of different population sizes may raise concerns with respect to the independence of irrelevant alternatives (IIA) assumption in multinomial logit models.

Suppose a student is deciding between two cities to study in: City A and City B. The student evaluates the pros and cons of each city and decides that City A is the better choice. According to the IIA assumption, the relative preference between City A and City B should not change if a third, irrelevant option is added to the set of choices.

For example, if a third city, City C, is added to the set of choices but is not relevant to the decision-making process (e.g. because it is too far away), the student's preference order between City A and City B should not change. However, if the introduction of City C does affect the student's preference order and they now prefer City B over City A, this would be a violation of the IIA assumption.

In our testing, this in fact proved to be an issue by the Hansen test which displayed a p-value of zero. As a robustness check of our logit model, therefore, we also ran probit models as they do not make the same IIA assumption. These are included in the Appendix of this chapter.

From an empirical perspective, though, our logit and probit estimations do not

Table 3.11: Hausman tests of IIA assumption

Ho	chi2	df	P>chi2
0	3089.905	32	0.000
1	1970.947	32	0.000
2	44513.137	32	0.000
3	5192.841	32	0.000

Table 3.12: suest-based Hausman tests of IIA assumption

Ho	chi2	df	P>chi2
0	59855.645	36	0.000
1	7660.881	36	0.000
2	87241.078	36	0.000
3	20925.567	36	0.000

differ substantially. According to Kropko (2007) this is not unusual. In simulations, Kropko (2007) found that multinomial logits nearly always provide more accurate results than probits, even when the IIA assumption is severely violated.

3.6.2 City-level effects of FIES

In the previous section we examined the effects of student financing on the individuals' relocation decisions and destination choices under financing. From the policy-maker's perspective, though, a broader view of the aggregate effects of the policy allow for a better understanding of how it affects the contingents of higher education students in rural and urban locations, and therefore whether the financing scheme has an overall positive or negative effect on the local human capital stock and regional imbalances of human capital.

This section will also help us elucidate why we are observing a surge in the probability of small to small migration, and from large to small cities. The first models of aggregate enrollments at city-level will give us a sense of the impact of the policy on the total contingent of students and whether the effect is driven by changes in enrollments at home or of enrollments away from home. The final regressions on this section will estimate the effects of the financing scheme on the local number of providers, giving us a sense of whether and how the policy affects the supply of higher education in rural and urban areas.

But as we previously discussed in Section 3.5, at the city-level our treatment variable and our outcome variable are endogenous due to a reverse causation: the number of students enrolled and financed locally are simultaneously determined. To address this issue we start this section with the first-stage regressions we have used in our 2SLS models which we present in the following subsections.

3.6.3 Effects of FIES on enrollments at home vs away

Using the model specification we have previously laid out in Equation 3.3 and the instrumenting of the endogenous treatment we presented in Table 3.7, we start by running a 2SLS regression of aggregate enrollments.

Table 3.13: Classroom-based enrollments

	OLS			IV		
	All	Home	Away	All	Home	Away
FIES (log)	0.37***	0.72***	0.33***	0.89***	1.57***	0.87***
	(0.032)	(0.097)	(0.027)	(0.091)	(0.264)	(0.085)
FIES (log) x Population (log)	-0.024***	-0.039***	-0.023***	-0.053***	-0.086***	-0.053***
	(0.003)	(0.007)	(0.002)	(0.006)	(0.016)	(0.006)
Population (log)	1.03***	1.38***	0.81***	1.03***	1.25***	0.80***
	(0.024)	(0.049)	(0.024)	(0.040)	(0.085)	(0.034)
Constant	-5.91***	-12.4***	-3.87***	-6.54***	-11.8***	-4.34***
	(0.264)	(0.505)	(0.268)	(0.433)	(0.793)	(0.361)
Observations	8710	8710	8710	8710	8710	8710

Standard errors in parentheses
Time fixed-effects estimation
Error terms clustered by city
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Complementing our student-level results, we find that student financing indeed increases enrollments away from home, but what we can now observe too is that financing does not do so *at the expense of* enrollments at home, but in fact it increases enrollments both at home and away.

Because our student-level data is truncated and we only observe students who have enrolled in higher education, the previous student-level regressions give us a picture of the probability of migration given financing but they do not give us a measure of whether more students are enrolling because of the scheme. In these regressions, though, we can now observe that although the probability of relocation increases with FIES, the impact in total enrollments in a city is in fact driven by the surge in the number of students enrolled in their places of origin, and that effect is particularly strong for the small rural locations. In other words, not only the financing program does not cause a brain drain from small to large cities, but it also means a surge in the total number of enrollments especially in smaller cities and especially at home. These results answer our third research question and confirm our third hypothesis, which posits that financing is at the root of a surge in local students both in rural and urban areas.

3.6.4 Effects of FIES on supply of higher education

In this final section of this paper, we estimate the effect of FIES on the number of higher education providers across rural and urban locations. Our previous models have shown that the scheme has fostered the expansion of enrollments and of mobility of students, favouring small cities in particular. In this final portion of our paper we look into whether this expansion of access has also come in the form of more supply of higher education in rural and urban areas.

In the results we display below in Table 3.14, we have used the number of higher education providers in a given city as our dependent variable in the same model specifications we have adopted throughout the city-level enrollments investigation.

The three models we present in OLS and IV versions are distinct in terms of their samples. The first uses all cities which had provision of higher education in any of the years of our investigation. The second includes only the cities with at least one provider, and the last one is only for cities which had at least one year of no provision between 2010-2019.

Table 3.14: Effects on the coverage of higher education

	OLS			IV		
	All	With provision	No provision	All	With provision	No provision
FIES (log)	0.094**	0.019	0.51***	0.26**	0.046	1.22***
	(0.036)	(0.035)	(0.117)	(0.078)	(0.069)	(0.286)
FIES (log) x Population (log)	-0.0082**	-0.0030	-0.050***	-0.024***	-0.0096*	-0.13***
	(0.003)	(0.003)	(0.011)	(0.005)	(0.005)	(0.022)
Population (log)	0.68***	0.64***	0.60***	0.80***	0.85***	0.94***
	(0.014)	(0.014)	(0.050)	(0.027)	(0.068)	(0.097)
Constant	-6.00***	-5.48***	-5.43***	-7.25***	-7.67***	-8.76***
	(0.153)	(0.150)	(0.518)	(0.268)	(0.739)	(0.918)
Observations	8710	7121	1589	8710	7121	1589

Standard errors in parentheses
Fixed-effects estimation
Error terms clustered by city
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

What we observe from our results in Table 3.14 is that for the full sample, there is increase in the number of providers in a city, but only for small cities. In larger cities (where log-population is above ten), the effect of the policy on the supply goes from neutral to the consolidation of providers. This is further confirmed by the model we estimate only on the cities with provision. There is a small and slightly significant negative effect of the policy on the number of providers for those cities with provision, but that only effect only holds for larger cities. This indicates that our hypothesis that students from large metropolises are more likely to relocate because of supply constraints is correct.

In the last column of Table 3.14 we estimate the effects of financing in those cities where there was no provision of higher education in at least one of the years of our sample. We find a strong and significant positive effect of financing on the number of providers in cities with no provisions. This is important from a policy perspective because this means that FIES is fostering the expansion of the access to higher education to rural locations where there was previously no supply. This also indicates that the hypotheses that students from larger cities were more likely to relocate to small ones was because of an increase in supply in smaller cities and a squeeze in larger ones is correct.

3.7 Conclusion

In this Chapter we have looked at the effects of the rise and demise of a large classroom-based student financing scheme on regional higher education imbalances in Brazil. More specifically, we looked at the differential effects of the policy on student migration, enrollments and supply coverage in cities of different sizes.

From our student-level models we found that students from smaller cities were more likely to relocate if given financing, but that the destination choices for those students were in fact other smaller cities. We also found that students from larger cities were more prone to relocate to smaller cities under the financing scheme, indicating that our hypothesis o brain drain from rural to urban areas was largely incorrect. We found instead a brain drain from large to small cities, and an increased mobility among smaller cities.

In our city-level aggregate models we have shown that the total number of enrollments is increasing with the scheme, but especially the enrollments at home and in smaller rural locations. This indicates that the policy is increasing mobility and the number of students from smaller cities at the same time.

3.8 Appendix

3.8.1 Descriptive statistics of the original data and dataset employed

In the Table below, we display the initial sample size and the reasons why data was dropped until we obtained to our final sample.

Original data	12,834,595
FIES (d)	-966,979
Public schooling (d)	-2,237,097
Without provision	-1,446,771
Final dataset	8,183,748

And in the following table, we show the original descriptive statistics of our variables in the full sample.

Table 3.15: Descriptive statistics of student-level variables from original data

	Obs	Mean	Std Dev	Min	Max
Migrate(d)	12,818,255	.48	.5	0	1
Age	12,834,595	25	7.6	12	110
White(d)	12,834,595	.33	.47	0	1
Female(d)	12,834,595	.55	.5	0	1
Population origin	12,834,595	2.2	3.6	.00079	12

3.8.2 Instrument step-by-step

On Table 3.3, along with the information on the budget allocation for new students coming into higher education (the first component of our instrument), we also display the maximum income threshold set by the central government in minimum wage equivalents (m.w.e.). Our second component of the instrument is, therefore, the estimated proportion of the population in a city which fits under the income ceiling.

To estimate this we needed two elements: the income threshold and a cumulative function of the income distribution among the population of each city, so that we could calculate the cutoff point of the function where the income per capita reaches the ceiling. The first is not hard to obtain. Although the criterion over which the maximum threshold is set changed along the years we studied, we cleared that hurdle without major complications. While for the years 2015-2019 the maximum threshold was set per capita, for the years of 2010 through to 2014 the maximum income was set per family, which we have then converted to per capita income in our dataset by dividing the family income by the mean household size in the city according to the 1991 Demographic Census.

The cumulative function of the income distribution is slightly more complex. This is because this function is naturally not available from the income information recorded by the national bureau of statistics IBGE. What is provided, though, is the mean income per quintile of the population. So in this case we can estimate the cumulative function that better fits the data points that were provided, and then we can, using that function, calculate the cutoff point.

This approach, though, required that we fit to each city's income distribution data a non-linear functional form which never exceeded 1 and which increased at lower rates in higher quintiles because the income gaps between quintiles becomes larger towards the top. Also, as this is a cumulative function, we fit to the data an exponential function of the type $y_{ct} = \beta_1(1 - \beta_2^{x_{ct}}) + \epsilon$ which satisfies all of our functional requirements.

Fitting a function and calculating the cutoff points at the income thresholds, we find that for smaller cities with population under 20,000 - which comprise over half of all cities in our sample - 90% of the population was eligible for financing (see Table 3.4). This is a much higher proportion than we expected. Estimating the same function for the larger metropolises, though, we found that around 64% of the population was included under the income threshold. These results require some comment. Although it is unclear whether the policymarkers were aware, the income thresholds for the program were in fact quite high. Also, the discrepancies between the relative stringency of the thresholds in towns and cities is also an interesting case of how nationally-set policies tend to affect cities of different sizes in very distinct degrees.

3.8.3 Alternative specifications and robustness checks

Displaying the coefficients of time-fixed effects hidden in the main table

Table 3.16 below displays the full results of Table 3.6, now including the coefficients for the year fixed effects which we omitted from the main results table for simplicity.

Table 3.16: Individual propensity to migrate - displaying year FE

Migrate (d)	1	2	3	4
FIES (d)	-0.0081***	0.023***	0.67***	12.2***
	(0.002)	(0.002)	(0.017)	(0.136)
Population (log)	-0.39***	-3.74***	-0.38***	-3.43***
	(0.000)	(0.007)	(0.001)	(0.008)
FIES (d) × Population (log)			-0.052***	-1.85***
			(0.001)	(0.020)
FIES (d) × Population (log)^2				0.069***
				(0.001)
Population (log)^2		0.12***		0.11***
		(0.000)		(0.000)
Public schooling (d)	0.031***	-0.019***	0.031***	-0.020***
	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.019***	0.019***	0.019***	0.019***
	(0.000)	(0.000)	(0.000)	(0.000)
White (d)	0.15***	0.093***	0.15***	0.094***
	(0.002)	(0.002)	(0.002)	(0.002)
Female (d)	-0.0069***	-0.011***	-0.0067***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)
2011	-0.011	-0.018**	-0.012*	-0.018**
	(0.006)	(0.006)	(0.006)	(0.006)
2012	0.032***	0.041***	0.030***	0.041***
	(0.005)	(0.005)	(0.005)	(0.005)
2013	0.063***	0.069***	0.062***	0.069***
	(0.005)	(0.005)	(0.005)	(0.005)
2014	0.028***	0.044***	0.027***	0.045***
	(0.005)	(0.005)	(0.005)	(0.005)
2015	-0.047***	-0.022***	-0.048***	-0.023***
	(0.005)	(0.005)	(0.005)	(0.005)
2016	0.018***	0.048***	0.018***	0.047***
	(0.005)	(0.005)	(0.005)	(0.005)
2017	0.067***	0.098***	0.066***	0.098***
	(0.005)	(0.005)	(0.005)	(0.005)
2018	0.087***	0.16***	0.085***	0.16***
	(0.005)	(0.005)	(0.005)	(0.005)
2019	0.100***	0.16***	0.098***	0.16***
	(0.005)	(0.005)	(0.005)	(0.005)
Constant	4.15***	26.3***	4.03***	24.3***
	(0.008)	(0.050)	(0.008)	(0.055)
Observations	8183748	8183748	8183748	8183748

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The Table 3.17 below displays the individuals' propensity to migrate given in the baseline and treatment scenarios, which we have used to produce Figure 3.9. Note that in this table we display the standard errors of the estimate, so that we can test whether the predictions for the probability of relocating with and without FIES are statistically different from each other.

Table 3.17: Predicted probability of individual propensity to migrate (logit)

	No FIES	With FIES
20k	0.88***	0.93***
	(0.000)	(0.001)
50k	0.72***	0.78***
	(0.000)	(0.001)
100k	0.59***	0.60***
	(0.000)	(0.001)
200k	0.46***	0.44***
	(0.000)	(0.001)
300k	0.40***	0.37***
	(0.000)	(0.001)
400k	0.36***	0.33***
	(0.000)	(0.001)
500k	0.34***	0.30***
	(0.000)	(0.001)
1m	0.28***	0.25***
	(0.000)	(0.000)
2m	0.25***	0.24***
	(0.000)	(0.000)
3m	0.25***	0.25***
	(0.000)	(0.000)
6.7m	0.25***	0.30***
	(0.000)	(0.001)
12.3m	0.28***	0.37***
	(0.000)	(0.001)
Observations	8183748	8183748

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Alternative specification of the migration dummy - 50km radius

The results on Table 3.18 below are alternative specifications of the binary dependent variable we used throughout the paper. As previously discussed, our most simple version of the migration dummy presents a few caveats, some of which we can address with alternative dependent variables.

In the first of these models I use a different specification of the dependent variable which considers only those who are studying at least 50km away from their places of origin in geographical distance to have moved. This variable was constructed by the latitude and longitude coordinates of places of origin and of destination and then using the Stata geodist package to calculate the distance between the two points in the globe. This specification of this variable is more robust to those students who commute between adjacent cities. Results do not differ substantially.

The second model is a more robust specification to address our concerns that the data for individual FIES participation may be missing systematically by city size. Our suspicion lies in that smaller cities generally host smaller higher education providers with less rigid administrative controls which end up under-reporting the participation in FIES. To address this, we weight the treatment variable by the probability of FIES not being reported for each city of destination and year. This probability is calculated as 1 minus the proportion of all enrollments in that city for whom the FIES variable was present in the full sample. This method gives more importance to observations in cities with a smaller probability of FIES reporting, which should correct for any under-representation of those cities in the restricted vs the full sample.

In these two cases we still have very similar effects of the treatment which do not alter our conclusions.

Table 3.18: Individual propensity to migrate - alternative dependent variable specifications

Migrate (d)	Dist	Mic	OLS
FIES (d)	9.24***	8.30***	16.3***
	(0.138)	(0.139)	(0.187)
FIES (d) × Population (log)	-1.40***	-1.26***	-2.37***
	(0.021)	(0.021)	(0.028)
FIES (d) × Population (log)^2	0.053***	0.047***	0.085***
	(0.001)	(0.001)	(0.001)
Population (log)	-3.32***	-3.84***	-3.45***
	(0.008)	(0.008)	(0.008)
Population (log)^2	0.11***	0.13***	0.11***
	(0.000)	(0.000)	(0.000)
Public schooling (d)	-0.18***	-0.11***	-0.019***
	(0.002)	(0.002)	(0.002)
Age	0.027***	0.024***	0.019***
	(0.000)	(0.000)	(0.000)
White (d)	0.073***	0.14***	0.094***
	(0.002)	(0.002)	(0.002)
Female (d)	-0.012***	-0.011***	-0.011***
	(0.002)	(0.002)	(0.002)
Constant	22.3***	25.8***	24.3***
	(0.056)	(0.057)	(0.054)
Observations	8183748	8180890	8183748

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Alternative specification of the population variable - split sample

On Table 3.19 below we show the results of the alternative model specification which was available for us to deal with the fact that Rio and Sao Paulo are outliers in the population scale but represent a significant portion of the number of students. In the main results we present in Table 3.6 we used a log-transformation of the population variable to address the skewedness of the data and we also included its squared term to account for the non-linearity of the propensity to migrate we observe in Figure

3.2.

An alternative functional form is to skip the log-transformation and run the model for the separate linear portions of the data. The model we present in Table 3.19 answers to that specification, with the sample now split between Rio and Sao Paulo in Coluns 3 and 4, and all the rest of the cities in Columns 1 and 2.

Table 3.19: Individual propensity to migrate - split sample

Migrate (d)	<5m	<5m	>5m	>5m
FIES (d)	-0.033***	1.07***	0.51***	-4.21***
	(0.002)	(0.021)	(0.005)	(0.278)
FIES (d) × Population		-0.087***		0.29***
		(0.002)		(0.017)
FIES (d) × Population		-0.087***		0.29***
		(0.002)		(0.017)
Population	-0.63***	-0.61***	0.15***	0.10***
	(0.001)	(0.001)	(0.006)	(0.007)
Public schooling (d)	-0.12***	-0.12***	0.34***	0.34***
	(0.002)	(0.002)	(0.004)	(0.004)
Age	0.021***	0.021***	0.013***	0.013***
	(0.000)	(0.000)	(0.000)	(0.000)
White (d)	0.050***	0.051***	0.097***	0.098***
	(0.002)	(0.002)	(0.004)	(0.004)
Female (d)	-0.0087***	-0.0083***	-0.023***	-0.022***
	(0.002)	(0.002)	(0.004)	(0.004)
2011	-0.0015	-0.0033	-0.20***	-0.20***
	(0.006)	(0.006)	(0.013)	(0.013)
2012	0.100***	0.097***	-0.42***	-0.42***
	(0.005)	(0.005)	(0.011)	(0.011)
2013	0.20***	0.19***	-0.62***	-0.62***
	(0.005)	(0.005)	(0.011)	(0.011)
2014	0.20***	0.20***	-0.76***	-0.76***
	(0.005)	(0.005)	(0.011)	(0.011)
2015	0.13***	0.13***	-0.79***	-0.79***
	(0.005)	(0.005)	(0.011)	(0.011)
2016	0.20***	0.20***	-0.72***	-0.72***
	(0.005)	(0.005)	(0.011)	(0.011)
2017	0.27***	0.26***	-0.75***	-0.75***
	(0.005)	(0.005)	(0.011)	(0.011)
2018	0.22***	0.22***	-0.21***	-0.20***
	(0.006)	(0.006)	(0.012)	(0.012)
2019	0.10***	0.097***	0.25***	0.25***
	(0.006)	(0.006)	(0.012)	(0.012)
Constant	7.08***	6.85***	-3.24***	-2.55***
	(0.010)	(0.011)	(0.101)	(0.108)
Observations	6664112	6664112	1519636	1519636

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We see that these regressions run on the split sample confirm what we find on the main results using the log-squared specification. For smaller cities, the financing treatment is positive on the propensity to migrate and decreasing as the population increases. For the two largest cities, FIES is also positive on migration but now increasing with population.

Alternative model - linear probability OLS instead of logit

The OLS versions of the main logit regressions are also displayed below in Table 3.20. We can confirm the validity of the coefficients of the interactions of the discrete variables from the logit models, and we can also observe the improvement of the fit of these models as we ass the squared-log form of population, although this improvement is modest as expected. The three sets of results that follow are iterations of model specifications using different sets of fixed effects: time fixed effects, two-way fixed effects (time and city), and time-city fixed effects (interaction of time and city dummies). All models' standard errors are clustered within cities.

Table 3.20: Individual propensity to migrate - OLS

	1	2	3	4
FIES (d)	-0.0033***	0.0050***	0.15***	2.07***
	(0.000)	(0.000)	(0.003)	(0.023)
Population (log)	-0.085***	-0.80***	-0.083***	-0.74***
	(0.000)	(0.001)	(0.000)	(0.001)
Public schooling (d)	0.0081***	-0.0036***	0.0083***	-0.0037***
	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.0041***	0.0040***	0.0041***	0.0040***
	(0.000)	(0.000)	(0.000)	(0.000)
White (d)	0.035***	0.019***	0.035***	0.020***
	(0.000)	(0.000)	(0.000)	(0.000)
Female (d)	-0.0014***	-0.0023***	-0.0014***	-0.0022***
	(0.000)	(0.000)	(0.000)	(0.000)
Population (log) × Population (log)		0.026***		0.024***
		(0.000)		(0.000)
FIES (d) × Population (log)			-0.012***	-0.32***
			(0.000)	(0.003)
FIES (d) × Population (log) × Population (log)				0.012***
				(0.000)
Constant	1.42***	6.12***	1.40***	5.77***
	(0.001)	(0.009)	(0.002)	(0.010)
Observations	8183748	8183748	8183748	8183748

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Alternative model - linear probability OLS instead of logit with city fixed-effects

The following Table 3.21 displays the results of the same OLS model as above, although here we have included city-fixed effects as well as time-fixed effects. Our results from the previous logit and OLS models are preserved only for our last model specification, where we include the interaction of the FIES dummy with the squared of the population log.

Important to observe that the effects of the race and gender dummies, which were flipping signs in the previous models, are now stable once we include city dummies.

Table 3.21: Individual propensity to migrate - OLS with city and time-fixed effects

	1	2	3	4
FIES (d)	0.019***	0.019***	-0.074***	1.65***
	(0.000)	(0.000)	(0.003)	(0.021)
FIES (d) × Population (log)			0.0070***	-0.26***
			(0.000)	(0.003)
FIES (d) × Population (log) × Population (log)				0.0100***
				(0.000)
Population (log)	0.16***	0.95***	0.16***	0.91***
	(0.010)	(0.043)	(0.010)	(0.043)
Population (log) × Population (log)		-0.031***		-0.030***
		(0.002)		(0.002)
Public schooling (d)	-0.0092***	-0.0091***	-0.0094***	-0.0095***
	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.0035***	0.0035***	0.0034***	0.0034***
	(0.000)	(0.000)	(0.000)	(0.000)
White (d)	0.025***	0.025***	0.025***	0.025***
	(0.000)	(0.000)	(0.000)	(0.000)
Female (d)	-0.0043***	-0.0043***	-0.0044***	-0.0043***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-1.89***	-6.81***	-1.86***	-6.43***
	(0.128)	(0.297)	(0.128)	(0.297)
Observations	8183748	8183748	8183748	8183748

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Alternative model - split samples for home and away

In this model we test an alternative specification of the logit and the OLS models. In this specification we split the samples by studying at home and away for each location, and our outcome variable is the total number of students studying at home in one regression, and the number of students studying away in the other.

The financing program increases enrollments at home and away for all locations of all sizes. Students are more likely to stay at home due to financing, and the larger the city the closer a student is to 50/50 chance of studying at home vs away.

Table 3.22: Classroom-based enrollments

	Home	Away	Home	Away
FIES (log)	0.72***	0.33***	1.57***	0.87***
	(0.097)	(0.027)	(0.264)	(0.085)
FIES (log) x Population (log)	-0.039***	-0.023***	-0.086***	-0.053***
	(0.007)	(0.002)	(0.016)	(0.006)
Population (log)	1.38***	0.81***	1.25***	0.80***
	(0.049)	(0.024)	(0.085)	(0.034)
2011	-0.17***	-0.014	-0.40***	-0.16***
	(0.036)	(0.010)	(0.081)	(0.023)
2012	-0.19***	0.14***	-0.68***	-0.16***
	(0.058)	(0.014)	(0.162)	(0.045)
2013	-0.42***	0.077***	-1.03***	-0.30***
	(0.068)	(0.016)	(0.195)	(0.055)
2014	-0.50***	0.13***	-1.20***	-0.30***
	(0.074)	(0.018)	(0.220)	(0.063)
2015	-0.37***	0.059***	-0.78***	-0.19***
	(0.056)	(0.014)	(0.133)	(0.039)
2016	-0.31***	-0.0038	-0.66***	-0.22***
	(0.052)	(0.015)	(0.112)	(0.034)
2017	-0.13*	-0.036*	-0.25***	-0.12***
	(0.052)	(0.016)	(0.060)	(0.021)
2018	0.012	-0.030	0.055	-0.0094
	(0.064)	(0.017)	(0.067)	(0.018)
2019	0.75***	0.053	1.64***	0.59***
	(0.123)	(0.029)	(0.307)	(0.078)
Constant	-12.4***	-3.87***	-11.8***	-4.34***
	(0.505)	(0.268)	(0.793)	(0.361)
Observations	8710	8710	8710	8710

Standard errors in parentheses
Fixed-effects estimation
Error terms clustered by city
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Alternative model - multinomial probits to address IIA concerns

In this paper we used a multinomial logit model to estimate the individuals' mean propensity to migrate to locations of different sizes in Table 3.9. This raised concerns with respect to the multinomial logit's independence of irrelevant alternatives (IIA) assumption, which proved to be an issue in our application as demonstrated by the formal Hansen test which displayed a p-value of zero.

As a robustness check of our logit model, therefore, we also ran probit models as they do not make the same IIA assumption. From an empirical perspective, though, our logit and probit estimations do not differ substantially. For a more thorough discussion of using probit models to address IIA concerns in logit models and the differences in estimates can be found in Kropko (2007).

Table 3.23: Multinomial probit model of student choice

Migrate:	to small <200k	to medium >200k	to large >5m
Baseline: stay at home			
FIES (d)	9.16***	7.95***	6.50***
	(0.125)	(0.112)	(0.264)
FIES (d) × Population (log)	-1.41***	-1.19***	-1.04***
	(0.019)	(0.017)	(0.040)
FIES (d) × Population (log)^2	0.054***	0.044***	0.039***
	(0.001)	(0.001)	(0.002)
Population (log)	-3.31***	-2.75***	0.71***
	(0.008)	(0.007)	(0.014)
Population (log)^2	0.11***	0.092***	-0.039***
	(0.000)	(0.000)	(0.001)
Age	0.0081***	0.017***	0.018***
	(0.000)	(0.000)	(0.000)
White (d)	0.22***	-0.0069***	0.13***
	(0.002)	(0.001)	(0.002)
Female (d)	-0.0040*	-0.014***	0.0029
	(0.002)	(0.001)	(0.002)
Constant	22.6***	18.8***	-5.22***
	(0.053)	(0.045)	(0.092)
Year (d)	Yes	Yes	Yes
Observations		8,183,748	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

And the margins predictions which follows from the multinomial probit in the paper here were calculated out of the multinomial logit model above.

Table 3.24: Multinomial probit model of student choice: marginal effects of financing

Migrate:	to small <200k	to medium >200k	to large >5m
Baseline: stay at home			
5k	0.014***	-0.0074*	-0.00029***
	(0.004)	(0.004)	(0.000)
10k	0.017***	0.014***	-0.0022***
	(0.003)	(0.003)	(0.000)
100k	0.011***	0.027***	-0.024***
	(0.001)	(0.001)	(0.000)
500k	-0.00017	-0.0083***	-0.025***
	(0.000)	(0.001)	(0.000)
1m	0.0022***	-0.011***	-0.020***
	(0.000)	(0.000)	(0.000)
2m	0.0076***	-0.0067***	-0.013***
	(0.000)	(0.000)	(0.000)
6.7m	0.031***	0.016***	-0.0036***
	(0.000)	(0.001)	(0.000)
12.3m	0.056***	0.036***	-0.0013***
	(0.001)	(0.001)	(0.000)
Observations		8,183,748	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 4

Effects of provider competition in higher education outputs: the Brazilian case

How does competition affect higher education outputs? Should policymakers worry about the effects of competition on student achievement, tuition fees and, more broadly, social benefits and costs? In this paper we investigate the relationship of market structures and enrollments, fees, quality, student ability and achievement in higher education. Using an IV identification strategy in a cross-section of Brazilian students and programs we demonstrate that in more competitive markets fees are lower and enrollments are higher, but mean student achievement is also diminished. But this is misleading. Using sequential results from fresher and senior-year examinations we show that the lower mean achievement is due to the expansion of enrollments especially among students of lower initial ability, and the vertical differentiation of programs. Finally, we demonstrate that because tuition is reduced and enrollment is increased in all tiers of ability, competition in fact leads to socially desirable outcomes.

4.1 Introduction

As national higher education systems shift to the mass provision of education, sector regulators have increasingly relied on provider competition to increase enrollments and maintain quality while keeping tuition fees at bay. This has stirred some debate among educators and policymakers concerned that competition, while effectively expanding enrollments and lowering tuition fees, is detrimental for student achievement.[1] This narrative has also gained ground among regulators of higher education concerned about decreasing returns to university degrees and an increasing number of institutions operating at lower quality standards. [2][3]

This view has been especially enabled by the lack of clear guidance from academic research on the effects of competition on higher education thus far. According to Vignoles et al. (2000), Belfield and Levin (2002) and others, these shortcomings have persisted due to constraints in data quality and quantity, and the lack of a clear theoretical framework to explain how competition affects education outcomes. This latter issue, in fact, has resulted in misspecifications of several of the econometric models of previous research papers and, not surprisingly, in mixed results. Furthermore, the literature has ignored the effects of expanding education systems, where the means

[1]See Carvalho (2013) for an example of that argument in the Brazilian context, Shumar (2013) for the US and Wilkinson and Wilkinson (2020) for the UK.

[2]The UK Office for Students consultation on raising the minimum quality standards on higher education: https://www.officeforstudents.org.uk/news-blog-and-events/press-and-media/regulator-plans-tougher-minimum-standards-in-higher-education

[3]Brazilian Minister of Education announces a ban on new medical programs to deter pervasive effects of competition: https://oglobo.globo.com/brasil/abertura-de-novos-cursos-de-medicina-ficara-suspensa-por-cinco-anos-22560806

of student achievement and institutional quality are shifting concomitantly because of sample changes.

More generally, the industrial organization research focused on education suffers from the compounded hurdles of both economics of education and industrial organization. From the economics of education canon, researchers working to establish the relationship of education inputs and outputs struggled with the biases from the omitted characteristics of students, schools and cities, which are especially difficult to address when student data is unavailable (see Hanushek (1986)). From the industrial organization side, as well as the omitted unobservables, researchers also faced the simultaneity of market structures and education quality, tuition fees and enrollments.

In this paper, we take a step back from the literature to analyze not just whether and how competition affects education outputs, but also through which avenues that effect occurs. We show that educators and policymakers have been misled by the negative effects of competition at the means of student achievement and education quality because as markets expand, participation increases disproportionately among students with lower ability and programs of lower quality. Looking at the distribution of student performance, we demonstrate that most students gain from increased competition: tuition fees decrease for all, the quality of high ability graduates increases while lower ability students gain access into higher education.

With 2,152 providers and 24,955 programs scattered in 502 cities, Brazil is an interesting setting to study the effects of competition in higher education due to its large and disperse higher education sector and its heterogeneous market configurations. Also, because participation rates are relatively low, competition can have a much more distinguishable effect on participation. The fundamental lessons learned in this paper, though, can be generalized to most expanding national education systems.

Brazilian higher education also offers a unique opportunity for the identification of the effects of competition on student achievement: higher education students are required to take a national secondary education exam as a means of selection into their programs. Additionally, the regulator conducts an examination for senior-year students as part of the quality control exercise. Through these sequential tests we are able to disentangle the effects of competition on final scores from all individual characteristics and prior education by assuming that all of these dimensions are captured in the entry test.

Following Borland and Howsen (1992) and the more extensive literature on the effects of competition in education, we compute a traditional Herfindahl-Hirschman Index to capture the effects of competition on enrollments, tuition fees, student achievement and finally, a specially constructed measure of the social benefits of higher education. According to Evans et al. (1993) regressions of market structure and outputs suffer from both a omitted bias - due to omitted student and local factors correlated with education outcomes - and a simultaneity issue as market structure and outputs are jointly determined.

To help us take care of omitted local factors affecting student outcomes, we make a simple assumption. Differently from schooling, focus of most of the research in this field, a defining characteristic of higher education is the horizontal product differentiation of courses. This is crucial in our paper because if we consider that not all subjects are substitutes, then it follows that multiple product markets can coexist inside each geographical market. This is a reasonable assumption and it allows us to use the within and cross-city heterogeneity as sources of identification while taking care of local specificities with dummies. This is possible if we assume that programs are comparable once their subject and city characteristics are held constant. This strategy was not available to the authors of the main references of this paper. Borland and Howsen (1992, 1993, 1996) on schooling and Hoxby (1997) on higher education relied on market structure proxies calculated at the geographical level, which means that local dummies would be collinear with market structures.

Besides the issue of the unobserved local factors, we still need to address the simultaneity of market structure and outputs. To address this issue, we used a 2SLS approach instrumenting current HHI with its eight-year lag. This approach too was not available to previous researchers because of the serial autocorrelation of past and present concentration measures in time-series environments. In our case, though, the correlation of error terms over time is not an issue since we are working in cross-sectional data, although we are able to retrieve older versions of most variables.

To test our identification strategy, we begin our empirical estimation section looking into the effect of our market structure proxy on tuition fees (prices) and on freshmen enrollments (quantity). Standard microeconomic theory predicts that, holding all other factors constant, increased competition should move the market equilibrium toward lower prices and higher output.

Once our identification strategy is validated, we move to examine the effect of competition on student achievement. To do so, we include the HHI in regressions of students' senior-year scores following Borland and Howsen (1992). Our results show that concerns over the effects of competition on higher education are not without reason: we find a negative and significant effect of competition on student achievement. Our results, at this stage, support the view that in more competitive markets fees are lower, enrollments are higher, and student achievement is reduced.

This effect of competition on achievement, though, is partly due to an ability [4] composition effect, and partly to an education quality effect. Controlling for these

[4] the term *ability* is used in this paper to conflate students' innate abilities and all of their prior formal and informal education

two channels, we show that competition has no direct effect on student achievement.

To control for student ability, we introduce in our achievement regression the individual entry scores from the National High School Examination, which students use to apply for higher education. By assuming that entry test results are the product of all observable and unobservable student characteristics and prior education, we claim that all of the unobserved student characteristics are taken care of by our ability control.

By extension, a similar approach can be used to control for the quality of higher education. If entry scores are a good proxy for student characteristics and prior education, then any contribution towards achievement in higher education must be a product of the observable and unobservable quality of higher education. In that case, controlling for student ability, the final examination results themselves are the best measure of education quality.

Because we cannot use students' final scores as outcome and explanatory variables in the same regression. We therefore construct our quality measure using the mean final test results of students from the previous cohort. To avoid any residual problems stemming from selection effects which may have been at play in the lagged cohort and are not taken care of by individual entry scores, we instrumented the past mean achievement using an indicator of academic staff qualification. This novel approach to controlling for education quality addresses one of the persistent challenges in the empirical literature in economics of education, which is the bias introduced by unobservable education quality in student achievement models.

Knowing that competition reduces tuition fees and expands student participation, we then examine how the composition of student ability and the quality of programs change across market structures. Using regressions similar to the ones on fees and enrollments, we see that in more competitive markets students are of lower mean ability, as measured by their entry test scores. Running the same regressions on the top and bottom entry scores of students selected in each market, though, we obtain an interesting result: we find that competition shifts both the upper and the lower bounds of the entry scores in opposite directions, but with the bottom effect dominating the top.

But as we saw from the results from previous regressions, the ability composition is not the only avenue through which market structure can affect student achievement. Examining how competition affects program quality we find a similar pattern to what we observe in the effects of competition on ability. More competitive markets observe lower mean program quality, but with the top and the bottom quality programs shifting in opposite directions. This is an indication of vertical differentiation under competition.

At this stage the hypothesis of eroding quality standards appears to be overly simplistic to describe how competition affects higher education. In fact, it seems that most students are better off in more competitive markets since participation is expanded, tuition fees are lower for all and program quality is better suited to students' ability. Our results paint a similar but slightly different picture from Hoxby (1997) overall. Where Hoxby finds evidence that competition leads to the selection of students of higher ability measured by GPA, we find that in more competitive markets average student entry scores are lower. A similar divergence occurs with respect to education quality. Hoxby found mean education quality increasing with competition, and our results point in the opposite direction. But when it comes to the vertical differentiation of the market, we find a similar pattern to what Hoxby (1997) observed. Programs are more stratified with competition while student ability is more homogeneous within programs and more disperse within markets.

In our final section, we examine the overarching question of whether competition in education is socially desirable. To do this, we construct a measure of the total social benefit of higher education in a given market as well as a measure of the total social cost to be used as output variables in market-level regressions in an exercise akin to a cost-benefit analysis of the effects of competition in higher education outputs. This follows from previous works discussing measures of education output that incorporate both the number of students and the quality of education such as Hanushek and Woessmann (2015) and Unesco's Education at a Glance (UNESCO, 2018).

Hanushek and Woessmann (2015) explores the relationship between education and economic growth and includes a chapter on the use of test scores as a measure of educational quality and argue that such scores should be combined with measures of enrollment and completion rates to get a more accurate picture of education output, in a similar manner to what we do in this section. Similarly, Unesco's Education at a Glance (UNESCO, 2018) includes a measure of education output that takes into account both the number of students and the quality of education by using indicators such as graduation rates and standardized test scores.

As indicated by Hanushek and Woessmann (2015) and (UNESCO, 2018), we constructed is a measure of the total benefit of higher education in a given city-subject combination. It is the log of mean final score of observed students in the program multiplied by the number of freshmen students, so that we have a program-level benefit measurement which we then sum across all programs in the same subject and city.

As our measure of total costs, though, we use pure and simply an estimation of the aggregate amount of tuition fees paid to institutions by all students.

We find that competition increases the total benefits as well as total costs, as expected, but benefits outpace costs, an indication that more competitive markets get more benefits from higher education at a lower cost.

The remaining sections of this paper are a description of the data in Section 4.2, the identification strategy in Section 4.3, the effects of competition on student achievement in Section 4.4 and on total social benefits and costs in Section 4.5. Section 4.6 concludes.

4.2 Data

In this section, we first introduce the data we use to construct our variables and discuss their advantages and limitations.

To begin our investigation on the impact of competition on student achievement, we need a measure for achievement and a measure for competition. We measure achievement using the students' score in the comprehensive National Examination of Student Development (ENADE) taken in the year they graduate from higher education. This examination feeds into the triennial quality assessment exercise conducted by the National Institute of Educational Research (INEP).

Because INEP carry out the assessment in three-year cycles, only a subset of subjects is assessed per year, so a complete cross section of all higher education is comprised of three years. For our analysis, we use the data from the 2015-2017 assessment cycle, which gives us a full cross-section of observations for all subjects and higher education institutions.[5]

[5] Although data for previous cycles is available, we do not have enough years to structure our data as a panel because although the regulator has publicized data for the last 8 years, this in fact means we only have two full cycles or, in other words, two periods.

4.2.1 A measure of achievement (ENADE)

The ENADE (Exame Nacional de Desempenho dos Estudantes - National Examination of Student Achievement) is a standardized test conducted by the Instituto Nacional de Estudos e Pesquisas Educacionais AnÃsio Teixeira (INEP - National Institute of Research and Educational Studies Anisio Teixeira), which is an agency of the Brazilian Ministry of Education.

The test is administered to students in undergraduate programs, and is used to evaluate the quality of education provided by these institutions.The test consists of a general knowledge section, which assesses the student's understanding of topics such as current events, citizenship, and environmental issues, and a specific knowledge section, which assesses the student's understanding of the specific subject area in which they are enrolled.

The ENADE is conducted every three years, with a rotating cycle of courses that are evaluated each year. Because we only have data from one 3-year cycle, in our cross-section students enrolled in the same subject have taken the same exam in the same year, so that our within-subject results are not influenced by variations on the difficulty of exams over time.

This makes measured achievement comparable across different geographical markets with different competition conditions. However, the achievement measure will differ from subject to subject because it is a weighted mean of the two components of the ENADE exam: the General Test (25%), which is common for all subjects examined in the same year, and the Subject-Specific Test (75%). To deal with differences in the Subject-Specific Test across subjects, we normalize the distribution for each subject so that we implicitly use an achievement measure which is relative to the performance of others taking the same subject. Effectively, we look at the impact of competition in a market on the achievement of a student in that market relative to the national distribution of achievement in the studentâs subject.

As is well known, test scores at the end of higher education are not a good measure for how much value higher education adds for a specific student (Hanushek, 1979). Student achievement, is not just a product of higher education, but the compounded result of higher education, prior schooling, family and individual characteristics as well as many other observed and unobserved factors.

4.2.2 A measure of prior ability (ENEM)

The Brazilian ENEM (Exame Nacional do Ensino Medio - Nationaion Examination of High School) is a standardized test that assesses the knowledge and skills of Brazilian high school students. The like the ENADE, the test is conducted by INEP, which is an agency of the Brazilian Ministry of Education.

The ENEM is administered annually, typically in November, and is used by universities and other higher education institutions in Brazil as a criterion for admission. The test is also used as a means of assessing the quality of secondary education in the country.

The test consists of four sections: Languages, Codes, and their Technologies; Natural Sciences and their Technologies; Human Sciences and their Technologies; and Mathematics and their Technologies. In addition, there is a written essay section that assesses the student's ability to articulate and support an argument.

To isolate the effects of higher education on achievement we use ENEM as a measure of students' prior achievement at the point of entry into higher education as a control variable. Although this test was taken by students of all subjects in our dataset, students may still have taken their entry test in different years, so these scores were also normalized inside the year of examination. To distinguish this from our higher education achievement measure, we will refer to it as student ability. This terminology should, however, not be confused with innate ability but rather as referring to a signal of endowment at the start of higher education.

Our dataset, therefore, consists of 324,616 matched scores of individual senior-year examinations and pre-entry test scores, publicly available from the 2015-2017 quality assessment cycle. [6]. The data is truncated so that we only observe the initial scores for those students who have taken the final exam. Students enrolled in public universities were removed from the sample because public and private higher education are not directly comparable. While private education is funded through tuition fees paid by students, public education is funded by taxpayers. Private providers therefore compete for students on quality and fees. Public universities though, are higher quality institutions operating at capacity, with no incentive to respond to competition from private providers. We also excluded students enrolled in online programs for similar reasons. Online programs attract a different set of students, with a broader catchment area which does not fit our geographical market definition, so online programs are not directly comparable to brick and mortar.[7]

Since entry into higher education and subject choices are endogenous, these choices themselves convey unobserved information about studentsâ ability which impact the final scores. These are factors that may be related with the cultural environment where a student grew up, the availability of higher education where a student lives, or other cultural factors that may facilitate learning in a higher education context. To capture these generally, we include in the dataset information on the subject of the student's program and the town or city where it is located.

[6]https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/idd

[7]A detailed list of the different compiled data-sets is provided in Table 4.13 of the Appendix.

4.2.3 A measure of education quality

Student achievement is also a product of the quality of the institution a student attends. However, good quality measurements have been a challenge in the literature because education quality depends on unobservables which are not readily controlled for by standard measures of institutional quality such as student/teacher ratios or salaries (Goldhaber and Brewer, 1997). In this paper, we propose to use the average scores obtained by students in a program in the previous assessment cycle as a more inclusive measure of quality. With this approach we intend to bring in a quality proxy to address both observable and unobservable quality, which is an issue the previous literature has wrestled with (Vignoles et al., 2000). There may be a concern that this measure is essentially endogenous because better institutions also attract better students on average, which affects achievement. To deal with this issue we have used more traditional measures of quality for higher education institutions, specifically the qualification of the academic staff, as instruments for average student achievement in the prior cycle.

4.2.4 A measure of competition (HHI)

The central focus of this study is on the impact of competition on student achievement and the quality of higher education. Measuring the degree of competition is notoriously difficult, though. Traditionally, the empirical literature has used concentration measures as a proxy for the degree of competition in a market. However, the literature on the endogeneity of market structure has questioned this approach because this implies that market structure and outcomes are jointly determined. In addition, in the 1980s previously observed strong relationships between concentration and margins appeared to break down in cross-industry regressions.

In our context, we can approximate actual markets by identifying a subject-city [8] pair as the relevant market in which a higher education institution competes. We assume for the purposes of this study that two higher education providers are only competitors if they both offer the same subject in the town or city. In this context using a concentration measure still makes sense because of two reasons. First, some of the problems of the empirical literature on the relationship between concentration and outcomes arose in cross-industry regressions. Cross-industry regressions typically rely on aggregated data across many actual markets. This type of averaging confounds the actual impact of concentration. Second, when geographical markets can be narrowly defined, dummy variables for location and subject effects can largely address concerns about omitted variables.

One could be concerned that looking at concentration at the level of the municipality could be too narrow a measure for competition because students may choose between courses in different cities. However, we do not observe a large level of mobility of students at the time of choosing a higher education institution. Although some students do move away from the cities they grew up in, most class-based students live with their parents (61%) or partners (26%), an indication that they did not move residence to study. Also, and in contrast to students in other countries, Brazilian students attend university primarily in the evening (80%) and work during the day (64%), so that location choices are heavily dependent on convenience. It should, therefore, not be surprising that, for 24% of students, proximity to work or home was the primary reason for their choice of higher education institution.[9] Even if there is some competitive pressure from higher education institutions in other municipalities, it appears that the impact of concentration at the municipal level will

[8]For simplicity we refer to locations as towns and cities interchangeably.

[9]Data from the student questionnaire of the 2017 national quality assessment (ENADE).

be a good proxy for the level of competition in the market, because this determines how much competitive pressure varies from municipality to municipality.

At the subject level there is an even stronger ambiguity over which subjects should be included in a concentration measure to capture the degree of competition. Not all courses offered within a geographical location are substitutes. But there are often different flavours of courses that students might often substitute between. The OECD 6-digit classification of subjects which is assigned to each course in the Higher Education Census is helpful in giving guidance as to such a choice. Table 4.1 gives an example of how different categorizations differ for economics subjects. Economics is classified as part of the Social Sciences (1-digit) and within that the behavioural sciences (2-digit). The economics subject appears at the 3-digit level indicating that this is generally a *major* in the studies, while we could think of finer qualifications as referring to *minors*.

Table 4.1: OECD subject classification

1-digit	2-digit	3-digit	6-digit
Social sciences	Behavioral sciences	Economics	Economics Business Economics - Finance Business Economics - International Business Business Economics - Public Policy Economics - Environment Economics - Development

Table 4.1 also indicates that there are ambiguities in these definitions. Economics and Business Economics, for example, are often different in terms of student population and academic organization, but are distinguished only below the 3-digit level. We nevertheless define the subject at the 3 first digits of the OECD classification,

acknowledging that typically within the 3-digit level there is a significant degree of substitution (and even a higher mid-course switching level than between 3-digit subjects). Therefore, when we refer to *subject* in the rest of the paper we refer to the 3-digit OECD classification. When we refer to a *program* we mean the group of courses in the subject offered by the same provider in a specific city. It should be noted that, as for all differentiated goods, a perfect market delineation is not possible. But it is also not necessary, because the degree of substitutability at the course/city is so important that a concentration measure at that level appears appropriate as a measure that distinguishes the difference in competitive pressure at the subject/city level.

As the measure of concentration at the city-subject level, we follow the literature using the Herfindahl-Hirschman Index. Market shares are constructed by calculating the number of freshmen initially enrolled in a program (university-subject-city) as a proportion of all freshmen in the same subject-city.

$$HHI_{sc} = \sum_{u=1}^{nsc} \left(\left(\frac{freshmen_{usc}}{\sum_{u=1}^{nsc} freshmen_{sc}} \right)^2 \right) \tag{4.1}$$

We provide below an example of the HHI index calculation for one specific subject-city combination.

So that $\text{HHI}_{sc} \in [0, 1]$ where 1 is a monopoly in the subject-city combination.

The descriptive statistics of the HHI and other variables at market-level are displayed below

Table 4.2: HHI computation from market shares

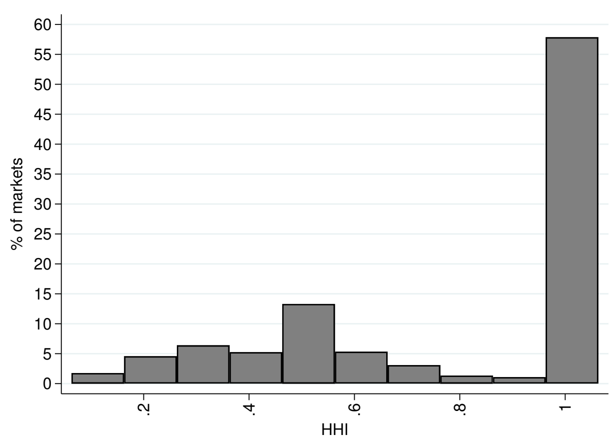
Year	Location	Subject	Provider ID	New enrollments	Market share	Market share^2	HHI
2015	Campo Grande	Law	270	699	0.24	0.06	0.38
			1308	176	0.06	0.00	
			10001	296	0.10	0.01	
			10003	1614	0.55	0.30	
			10010	143	0.05	0.00	

Table 4.3: Descriptive statistics of HHI and lagged HHI

	Year	Obs	Mean	Std Dev	Min	Max
Enrollments	2017	4,988	287	1,084	1	48,024
Fees	2017	4,988	898	649	93	7,164
HHI	2017	4,988	.78	.29	.063	1
HHI (lag)	2009	4,988	.85	.26	.044	1

The mean HHI is typically very high, which reflects the fact that over 50% of our markets are in fact monopolies, so the source of our identification will mostly be the differences in student achievement and other market outcomes between markets with a single provider and ones with more than one.

Figure 4.1: Distribution of HHI



We get a wide distribution for the HHI except for a considerable mass of monopoly markets. Over 55% of the 4,988 markets in our sample are monopolies.This is an important feature of the Brazilian higher education sector: highly competitive markets are few and appear in larger cities where many providers operate. Monopolistic markets are associated with smaller geographically isolated locations where only one program per discipline is financially sustainable.

The main results in this paper are tested in alternative specifications in the Appendix where we add a dummy for monopolies to ascertain that our results are not entirely driven by the distinction between monopolistic and non-monopolistic markets.

Before we examine the effects of competition on student achievement, we test our identification strategy on the effects of competition on enrollments (quantity) and tuition fees (prices), which are largely undisputed in the microeconomics literature as well as among policymakers. Tuition fees are constructed from the national student financing program FIES data. This dataset brings information on each monthly transaction from the public purse to providers to cover student fees. Although the information is at the individual level, it does not contain student IDs common to other datasets, so we can only construct fees at the program level, which we then aggregate to the market level. Finally, our measure of student numbers is obtained by adding up freshmen enrollments for all courses in the same subject and city in the Higher Education Census dataset.

One issue concerning tuition fees from the FIES dataset is that payments for students enrolled in the same program can vary due to studentsâ choice of modules and attendance regime, which cannot be observed in the data set. To obtain a more robust or typical measure for program fee, we calculated the fees based on the median of the observed monthly fees per program. We have checked whether using means or medians would alter results significantly, but we did not find any important differences. Out of the 11,006 programs we had initially in our sample, 10,263 of them had at least one student who was a beneficiary of FIES. Fees at the market level are then calculated as the mean of these course-level tuition fees. Figure 4.2 shows the distribution of monthly fees and across programs.

From those 385,291 initial students and 11,006 programs, we were left with 324,616 (84%) students and 8,469 programs due to either missing quality information, or data on fees. Among those students who have been excluded, there is a slight bias towards lower achievement since completion of the Census and participation in the national financing program are proxies for institutional quality. This is not concerning because these observations would have made our results stronger, since the biased sample would not favour our results

Table 4.4: Descriptive statistics of program-level variables

	Year	Obs	Mean	Std Dev	Min	Max
Fees	2017	8,469	855	602	93	7,164
Final scores (lag)	2017	8,469	43	7.3	7.1	74
Staff MSc (lags)	2012	2,343	69	21	0	100
	2013	2,933	73	20	0	100
	2014	3,193	4.3	17	0	100
	All	8,469	46	38	0	100

Figure 4.2: Fees by markets

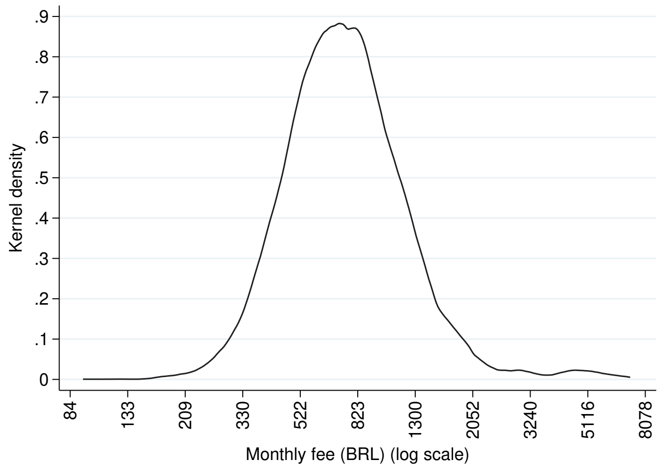
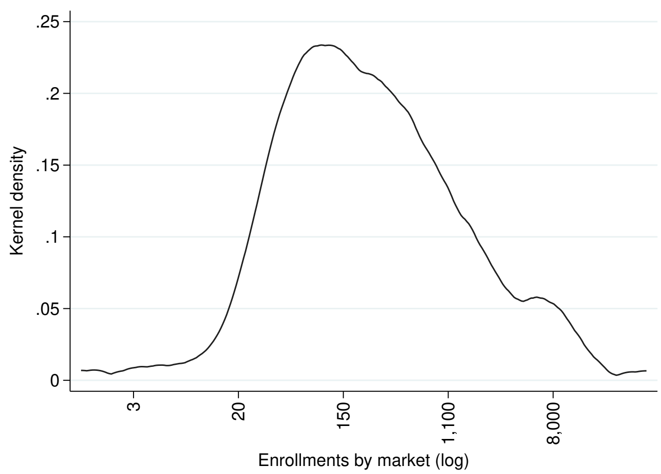


Figure 4.3: Enrollments by markets



The above distributions are displayed in log scale due to their long tails on the right-hand side. In a typical higher education market, between 100 and 1000 students enrol in higher education per year. Mean fees are around R$800 per month (roughly $150) - close to the minimum monthly income in Brazil.

Table 4.5: Descriptive statistics of variables at student level

	Year	Obs	Mean	Std Dev	Min	Max
Final scores	2015	142,980	46	13	0	93
	2016	71,154	48	14	0	93
	2017	110,482	44	13	0	97
	All	324,616	45	13	0	97
Entry scores	2009	25,938	53	7.9	31	82
	2010	79,224	56	7.7	33	81
	2011	96,007	52	6.7	31	79
	2012	71,018	53	7.1	33	79
	2013	33,644	51	6.1	34	76
	2014	16,696	52	5.6	21	78
	2015	1,908	53	6	33	82
	2016	181	56	7.1	40	74
	All	324,616	53	7.2	21	82

From the descriptive statistics of the variables in Table 4.5, we can see that our sample of student-level final scores is split into the three years of the evaluation cycle. Entry scores, though, are split over many more years because students in their senior year can take longer or shorter to get to their final year, depending on the length of the program, personal circumstances and other factors. Because programs in Brazil are typically 4-years long, the core of our students have taken their entry tests between 2010 and 2012.

4.3 Identification

In the previous section we detailed the variables we employ in the regressions of this paper. In this section we lay out the identification strategy we implement to isolate and capture the effects of market structure on student achievement. Our goal is to ensure that the identifying variation is coming solely from the differences in market structure in each subject-city combination.

Higher education providers may behave differently in competitive markets in comparison to providers in less competitive ones. While the monopolist's profit-maximizing decision is to produce until average costs equal marginal costs, under Cournot-Nash competition firms are driven to an equilibrium where prices equal marginal costs. In other words, standard microeconomics predict that in equilibrium monopolistic markets display higher prices (tuition fees), lower quantities (enrollments) and quality which can be either higher or lower than the social optimum.

In previous papers such as Hanushek (1971), Hanushek (1986), Murnane (1975), Armor et al. (1976), Murnane and Phillips (1981) and others, among the school variables, researchers have concentrated on the effect of measures of education quality on the achievement of students. In this paper, though, we go further to recognize that many other non-quality related factors feed into observed students results. For example, lower tuition fees may lead to higher enrollments as students with a lower valuation of education join higher education, which may in turn lead to lower achievement.

Therefore, in our main model of student achievement we will examine the effects of competition on students' scores on the senior-year results through two channels: the education quality effect and the student selection effect. We contend that if we have good controls for these two effects of competition in our models, then there should be no effect of competition on student achievement.

To control for students' ability at the point of entry into higher education, we use the scores each individual has obtained in the entry test examinations. We assume that those results are a product of all factors which have contributed towards the students' education up until that moment, including family and individual characteristics, and prior formal and informal schooling.

To control for education quality in a way to include both observable and unobservable elements, we used the previously described mean of scores that the students enrolled in the same program have obtained in the previous evaluation cycle, instrumented by the contemporary proportion of academic staff.

Regarding our measurement of competition, though, the general literature has long established that OLS estimators including a proxy of market concentration may be biased due to endogeneity issues (Evans et al., 1993). These can arise from two sources, the first of them being omitted local and subject-specific factors which can influence student achievement, such as the quality of local schools and how subjects may attract students of different abilities. To mitigate these, we use dummies at city and subject levels to control for confounding effects coming from city and program specificities which may be picked up by the HHI. This approach relies on the assumption that the marginal effects of HHI are the same across cities and subjects, which is a strong assumption. With this strategy we intend to capture the average effect of HHI across subjects.

Our approach to include two-way dummies to capture location and subject-specific heterogeneity holds similarities with the use of two way fixed effects (TWFE) in panel environments. This topic has motivated some recent discussion, especially with respect to the equivalence of TWFE models with difference-in-differences models in Imai and Kim (2021), Wooldridge (2021) and Cunningham (2021) among others. Our approach is similar to the TWFE strategy in the sense that we include two sets of dummies to control for variations in two dimensions of our data, which approximates our estimation to a panel data modelled with TWFE.

Our data, though, is not structured as a panel as time is not one of the dimensions of our modelling. This can be puzzling as there are year variables in our dataset. We remind the reader that although our data is sourced over several years, it is compiled to form only one complete cross-section.

Some of the advantages and shortcomings of the TWFE persist, though. Among the advantages, our FE estimator can adjust for unobserved location-specific and subject-specific confounders at the same time. An important disadvantage, though, is that the ability of the FE model to simultaneously adjust for these two types of unobserved confounders critically relies upon the assumption of linear additive effects as discussed by Imai and Kim (2021).

To address this shortcoming, Wooldridge (2021) proposes an estimator obtained from a pooled ordinary least squares regression that includes unit-specific time averages and time-period specific cross-sectional averages, which Wooldridge (2021) coins as the two-way Mundlak (TWM) regression. For the purposes of this paper, though, we make the assumption that the effects of HHI are linear in parameters for simplicity, as we do not observe a non-monotonic relationship between market concentration and our variables of interest in our test for alternative model specifications.

The dummy approach may well address omitted subject and city effects, but not problems of simultaneity. These issues are expected because firms' performances feed back into market structure causing a simultaneous equation bias where performance determines concentration and concentration determines performance. To address this issue, we implement an instrumental variable approach using an older version of the HHI as instrument. We follow the literature relying on lags as exogenous predictors of current market structure, such as Smith and Meier (1995) and Wrinkle et al. (1999) in competition among schools. To construct market shares that are not endogenous with unobserved market, provider or program characteristics, we recovered market shares from the year before any of the students in our sample entered higher education - 2009 - and use them as predictors for current market shares.

The case for lags as valid instruments relies on the argument that because the lag precedes the contemporary value in time, causality can only flow in one direction. Since there is normally a high degree of correlation between observations in time, lags should be strong predictors of their current values (Wang and Bellemare, 2019). The same argument for lags as valid instruments, though, draws the attention to a potential caveat. If observations have a high degree of auto-correlation, then arguably residuals do too. In that case, lags carry with them the endogeneity which was the reason for instrumenting in the first place and, therefore, do not fulfil the exclusion restriction criterion for valid instruments.

This issue has been thoroughly exposed in the seminal paper by Hausman (1978) and is the central motivation for the conception of the Newey-West estimator (Newey and West, 1987). These concerns, though, are directed at time-series environments. In a cross-section setup though, instrumenting with lags should not suffer from this shortcoming because we do not use repeated observations over time. This is why this instrumenting approach is particularly appropriate in our case: although we work in a cross-section we are still able to construct lagged HHIs, which is unusual.

$$F_{sc} = \alpha + \beta HHI_{sc} + \sum_{c=1}^{567} \theta_c C_c + \sum_{s=1}^{88} \gamma_s S_s + \varepsilon_{sc} \tag{4.2}$$

$$E_{sc} = \alpha + \beta HHI_{sc} + \sum_{c=1}^{567} \theta_c C_c + \sum_{s=1}^{88} \gamma_s S_s + \varepsilon_{sc} \tag{4.3}$$

Where F_{sc} is the median fee of the market (subject-city), E_{sc} is the total freshmen enrollments in a given market and C_c and S_s are sets of dummies for city and subject.

$$C_c = \begin{cases} 1 \text{ if market is in city } k \\ 0 \text{ otherwise} \end{cases} \qquad S_s = \begin{cases} 1 \text{ if market is in subject } k \\ 0 \text{ otherwise} \end{cases}$$

From the results of these regressions in Table 4.6, we find evidence that HHI is a meaningful measure of market power. Our results go along with the predictions and we see strong negative (positive) effects of competition (concentration) on median market fees and a positive (negative) effect of competition (concentration) on enrollments.

Table 4.6: Market-level fees and enrollments

	Fees			Enrollments		
	OLS	IV	FS	OLS	IV	FS
HHI	61.9**	298.5***		-252.0***	-1378.3***	
	(26.792)	(85.590)		(68.115)	(267.260)	
HHI (2009)			0.42***			0.42***
			(0.032)			(0.032)
Subject	Yes	Yes	No	Yes	Yes	No
City	Yes	Yes	No	Yes	Yes	No
Observations	4988	4988	4988	4988	4988	4988
F			174.65			174.65

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* p<0.10, ** p<0.05, *** p<0.010

Table 4.6 also presents the first stages of the 2SLS regressions. We obtain an F-statistic of 174.65 for the instrument in the first regression. This is comfortably exceeding the rule of thumb of 10 proposed by Staiger and Stock (1997) for instrument strength, but also not so high as to raise concerns that market structures may be too immutable over time so that our instrument is in fact a linear approximation of the contemporary HHI. In Table 4.17 of the Appendix we display an alternative specification of this model where we include dummies for monopolistic markets to ascertain that our results are not entirely driven by the distinction between monopolistic vs non-monopolistic markets.

If any concerns with the instrumenting persist, we expect that failing to correct for simultaneity should cause a downward bias to our estimates once controlling for subject and local effects, so that our results should be conservative estimates of the real effects. This is because the unobserved effects of competition are likely to be absorbed by the other control variables that we have included in our model.

4.4 Competition and student achievement

Education policymakers are concerned that market competition is leading to the commodification of higher education - lower fees and a surge in enrollments but compensated by lower student achievement. Our objective in this section is to verify whether the policymakers' concerns are well placed, using the identification strategy we previously laid out. We start by looking into the effects of competition on student achievement and we show that there are indeed negative effects, which occur through changes in student composition and education quality.

To further investigate these effects, we then move to examine how market structure affects the composition of ability of higher education students. According to our results, competition is negative on student ability. This is an indication that in more competitive markets, where fees are lower and enrollments higher, lower ability students are over-represented in higher education.

We then examine the effects of competition on program quality, and we find that mean program quality is lower in more competitive markets, supporting the view that competition leads to commodification and race to the bottom. Going beyond the effects at the means, though, we observe that competition broadens the entire distribution of program quality, and this vertical differentiation leads to both higher upper quality and lower bottom quality.

To examine the effects of competition on student achievement, we use a similar specification to Equations 4.2 and 4.3. We run a model of students' final scores FS_{iusc}, our proxy for student achievement, on our market structure index HHI_{sc}, alongside with the usual set of two-way dummies. We then repeat this estimation controlling for the initial ability of students by including the students' scores in the entry assessment ES_{iusc} and in the third model we include our measure of program quality Q_{usc}.

$$FS_{isc} = \alpha + \beta_1 ES_{isc} + \beta_2 Q_{usc} + \beta_3 HHI_{sc} + \sum_{k=1}^{567} \theta_k C_c + \sum_{k=1}^{88} \gamma_k S_s + \varepsilon_{sc} \tag{4.4}$$

And for the subject and city dummies:

$$C_k = \begin{cases} 1 \text{ if student is in city } k \\ 0 \text{ otherwise} \end{cases} \qquad S_k = \begin{cases} 1 \text{ if student is in subject } k \\ 0 \text{ otherwise} \end{cases}$$

In regressions OLS 1 and IV 1 on Table 4.7 below we observe that competition is negative on student achievement (or conversely, that concentration is positive). This is an indication that policymakers have a reason to be concerned. But market competition itself is not an input to education. If market structure affects achievement, it should occur *through* changes on the student composition and on the quality of education.

Table 4.7: Effect of market competition on student achievement

	OLS 1	OLS 2	OLS 3	IV 1	IV 2	IV 3
HHI	0.081**	0.065**	0.045	0.34***	0.20***	0.0011
	(0.033)	(0.030)	(0.030)	(0.090)	(0.052)	(0.065)
Student ability		0.58***	0.57***		0.58***	0.56***
		(0.009)	(0.010)		(0.009)	(0.011)
Program quality			0.091***			0.19***
			(0.006)			(0.066)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	324616	324616	324616	324616	324616	324616

Standard errors in parentheses
Standard errors clustered at market level
* $p<0.10$, ** $p<0.05$, *** $p<0.010$

In regressions OLS 2 and IV 2 of Table 4.7 we include the control for student ability. We observe that ability is positively correlated with achievement, and that the magnitude of the HHI coefficient is significantly reduced. This is an indication that, at least in part, the negative effect of competition on achievement is in fact due to an effect of competition on student composition, which is the first confounding factor the policymaker must bear in mind [10].

In regressions OLS 3 and IV 3 of Table 4.7 we add the quality control. Now the effect of competition is virtually zero and non-significant. This is an indication that if the effects of student composition and education quality are well controlled for, market structure dose not have an effect on student achievement. Examining the literature on the effects of competition on the achievement of school pupils, Belfield and Levin (2002) refers to the inconclusive results as a shortcoming of the literature. But if our hypothesis is correct, then it may just be the case that the most accurate research papers in this field are precisely those which find no significant effects.

In their first paper, Borland and Howsen (1992, 1993) found only just significant

[10] First stages of these models are provided at Table 4.27 along with first stages for all other regressions in the Appendix.

positive effects of competition on pupil achievement using a similar specification to Equation 4.4. In another regression which they consider to be their main, they include pupil ability and school quality controls together with the market structure proxy, and found, unsurprisingly, unconvincing results. Their estimates, though, could be an indication that their model was in fact well specified and controls were well constructed. But because their data is at the school district level, the two-way dummy approach we implemented was not feasible for them. In that sense, their market structure variable may well be picking up the confounding effects of unobserved local heterogeneity.

As robustness checks, we have also tested alternative model specifications for the model in Equation 4.4. First, throughout the paper we have used standardized final scores and entry scores to make sure that these measures are internally consistent. The final scores were standardized inside each subject and year groups to take into account the fact that tests are different by year and subject. Entry scores, similarly, are different every year so test may be slightly harder/easier in given years.

Our first alternative specification shows alternative results to the ones we displayed in Table 4.7, but which corroborate our story. Controlling for student ability and education quality, there are no significant effects of market competition on student achievement. Those results are presented in Table 4.19 of the appendix. Also, this indicates that our dummies for years and subjects are being effective in controlling for specific effects at those levels.

The results of our second set of alternative specifications are presented in Table 4.20 of the appendix. In those regressions we tested alternative specifications of the quality variable but still instrumenting the HHI measure. In the first alternative, in regression 3, we entered the lagged mean student scores directly into the model without instrumenting. In the second, we entered the present variable of academic qualification into the equation as a measure of quality. In both these models, we see that the coefficient of HHI is greatly eroded from the inclusion of these controls, but we still observe a residual effect we are only able to control once we include the instrumented quality variable as we presented in Table 4.7.

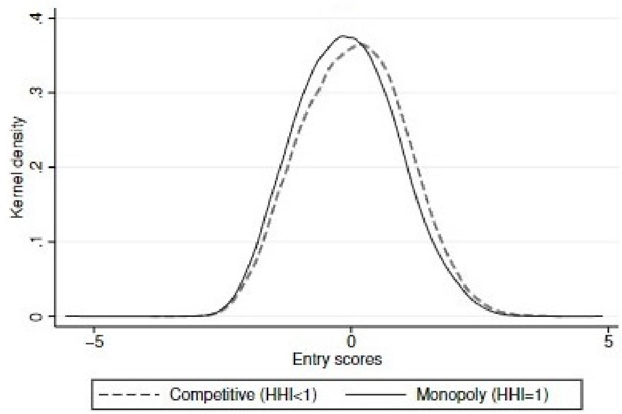
Finally, because our sample is dominated by single-supplier markets, we also added a dummy for monopolistic markets to ascertain that not all of our results are driven by the distinction between monopolistic vs non-monopolistic markets in Table 4.18. Results hold to show that competition has uniform effects in the spectrum of non-monopolistic markets.

4.4.1 Effects of competition on the composition of ability

As we examined in the previous section, concerns over the effects of market competition on student achievement are genuine, but the relationship between market structure, student composition and education quality requires further investigation. In this section we look into how market structure affects student composition.

From the previous results in Table 4.7, we can deduce that competition has a negative effect on the mean ability of freshmen students. This is a straightforward interpretation of the change in the HHI coefficient once we include the ability control. But plotting the distribution of ability in monopolistic and competitive markets presents a puzzling picture. In more competitive markets, providers do not seem to recruit students of lower ability - in fact freshmen students in more competitive markets are of higher ability than in non-competitive markets.

Figure 4.4: Kernel density estimate of entry scores in monopolies vs competitive markets



This occurs because monopolistic markets are different from competitive ones in several ways - not just market structure. Towns which can only sustain one higher education provider are generally smaller, lower income rural locations where primary and secondary education may be of lower quality, access to culture more scarce, among other factors, all of which point in the direction of lower student initial ability, and all of them have little to do with competition. To verify how students' mean ability varies with competition controlling for city specificities, we implement a regression of students' ability with city - and subject - dummies in the same manner we have done previously.

$$ES_{iusc} = \alpha + \beta_1 HHI_{sc} + \sum_{k=1}^{567} \theta_k C_c + \sum_{k=1}^{88} \gamma_k S_s + \varepsilon_{iusc} \tag{4.5}$$

The results in Table 4.8 are in line with the predictions of our previous model. Keeping local factors constant, more competitive markets recruit students of lower mean ability. This could be a direct effect of the reduction of fees and the expansion of access to higher education resulting from competition.

Table 4.8: Effect of competition on student ability

	(OLS) All	(OLS) Top 10	(OLS) Bottom 10	(IV) All	(IV) Top 10	(IV) Bottom 10
HHI	0.029	-0.48***	0.61***	0.25**	-1.16***	1.60***
	(0.028)	(0.044)	(0.055)	(0.116)	(0.127)	(0.170)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	324616	47643	47643	324616	47643	47643

Standard errors in parentheses

Standard errors clustered at municipal and subject levels

* $p<0.10$, ** $p<0.05$, *** $p<0.010$

We also implement the model to two sub-samples of the data so that we can observe how the tails of the ability distribution change. To do so we separate the top and bottom 10 entry test results from each market and run the same regression as we have run at the means. The results show that in more competitive markets, the 10 bottom ability students are of lower mean ability in comparison to students in less competitive markets. At the top 10 though, students' mean ability is also higher, which means that competition improves access for students of both high and low ability, but slightly more so for those on the lower part of the distribution. This is what we would expect from an expansion of access driven by the reduction of fees. First stages of these regressions are included in the appendix in Table 4.22.

Now if we run the model of student achievement from Equation 4.4 splitting the sample into students of higher and lower ability, we see a similar pattern. Students of higher ability have higher final scores in more competitive markets. For students of lower ability, competition has a negative effect on their achievement, but these students are in fact better off because their alternative would have been not to participate in higher education in a less competitive market. Table 4.23 presents the

OLS and the 2SLS versions of these regressions, first stages are identical to those in Table 4.27.

Table 4.9: Effect of competition on students' achievement

	OLS All	OLS Top 10	OLS Bottom 10	IV All	IV Top 10	IV Bottom 10
HHI	0.081**	-0.22***	0.36***	0.34***	-0.51***	0.97***
	(0.033)	(0.038)	(0.048)	(0.090)	(0.120)	(0.121)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	324616	47643	47643	324616	47643	47643

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* $p<0.10$, ** $p<0.05$, *** $p<0.010$

We also tested quantile regressions which could serve either as our main models or alternatives to our split-sample regressions of the effects of market concentration at different points of the distribution of our dependent variables. The results are presented in Section 4.9.1 of the Appendix.

4.4.2 Competition and market differentiation

In the previous section we observed the effects of competition on the composition of student ability. In this section we turn to the effects of competition on the vertical differentiation of programs under competition. In the identification section we have shown how HHI is a good measure of competition because we see that enrollments and tuition fees respond to it as predicted by microeconomics. So at this stage we know that more competitive markets display lower mean fees than less competitive ones. But fierce price competition can lead programs to vertically differentiate.

To verify how programs position themselves in the quality spectrum under competition, we will run OLS and 2SLS models of education quality on HHI. The dependent variable here, though, is worth some attention.

In previous models we have used the lagged mean results of students in the final year assessment of the previous evaluation cycle as a control for both the observed and the unobserved quality of a program. Even though in those regressions we were not particularly concerned with the student selection component carried by that variable since we were controlling for individual student ability, we nonetheless instrumented that variable with the proportion of academics with a postgraduate degree in that program to mitigate possible biases.

In this regression, though, we are interested in the quality of a program itself, so we are looking to use our measure of quality as a dependent variable. To do so, we re-run the first stage of the regression of quality as we demonstrate in Table 4.27 and we use the predicted values of the mean scores of students in the previous final tests as a measure of quality, just as we did in the achievement regressions.

$$Q_{usc} = \alpha + \beta_1 HHI_{sc} + \sum_{k=1}^{567} \theta_k C_c + \sum_{k=1}^{88} \gamma_k S_s + \varepsilon_{usc} \tag{4.6}$$

The results displayed in Table 4.10 show a negative effect of competition on program quality in the IV regression including all programs. But just as we learned in the previous sections, in the case of course quality there is also a slightly more complex picture below the surface. If we keep in our sample only the top quality program within each market and we run the same model as in Equation 4.6, we observe that in more competitive markets the top quality program is of higher quality than its counterpart in the less competitive market. On the lower tail, the lowest quality program is of lower quality in the more competitive market. This is an interesting finding because this means that while those lower ability students joining higher education because of lower fees are finding programs to match their ability, those students of higher ability, who were likely to be in higher education in any case, find programs of higher quality in the more competitive markets. First stage regressions are displayed in Table 4.24 in the appendix.

Table 4.10: Effects of competition on top and bottom quality programs by market

	OLS All	OLS Top quality	OLS Bottom quality	IV All	IV Top quality	IV Bottom quality
HHI	0.12	-0.70***	0.84***	1.26***	-1.16***	2.95***
	(0.077)	(0.093)	(0.108)	(0.191)	(0.210)	(0.248)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8469	4638	4602	8469	4638	4602

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* $p<0.10$, ** $p<0.05$, *** $p<0.010$

We also tested quantile regressions which could serve either as our main models or alternatives to our split-sample regressions of the effects of market concentration at different points of the distribution of our dependent variables. The results are presented in Section 4.9.1 of the Appendix.

In another alternative specification of this model, we run the exact same specification of Equation 4.6, but now instead of the lagged mean of students we use the qualification of the academic staff as the quality outcome variable, which we previously used as an instrument, which we considered to be a control for the observable portion of the quality. We display the results of this model specification in Table 4.21. The story is similar, although in those regressions we do not capture any significant effects on the top quality programs. That is also an interesting finding because it indicates that while lower quality programs are differentiating towards lower quality in observable measurements, top quality programs are in fact only differentiating themselves towards better quality in unobservable quality, which we do not capture in this model specification.

On Table 4.11, we display the results of the fee regression from Equation 4.2 now also applied to the top and bottom quality programs of each market. We find that fees are lower in more competitive markets for all programs both at the top and the lower ends of the quality distribution. The interesting finding here is that not only those students of lower ability are better off in more competitive markets - since their alternative would have been no higher education at all - but those students of higher ability are also better off given that they can find programs of higher quality, and that higher education now comes at lower tuition fees. First stage regressions are displayed in Table 4.25 in the appendix, although they are identical to the ones of the quality regressions immediately above. We include them just as a spot check on our model specifications.

Table 4.11: Fee differentiation

	OLS All	OLS Top quality	OLS Bottom quality	IV All	IV Top quality	IV Bottom quality
HHI	47.2*	73.9**	78.7***	311.4***	345.1***	350.8***
	(25.966)	(30.835)	(27.927)	(79.612)	(67.044)	(65.585)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8469	4638	4602	8469	4638	4602

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* p<0.10, ** p<0.05, *** p<0.010

4.5 Social cost and benefit analysis (CBA) of competition in higher education

In the previous sections we explored how market structure affects tuition fees, enrollments, the ability composition of freshmen students and the vertical differentiation of programs. We have shown that although the concerns over a potential watering down of student achievement are legitimate, it is in fact a red herring in a more complex picture. In more competitive markets, fees are lower, enrollments are higher, and the additional access benefits students of lower ability searching for lower quality programs.

In this section, we try to bring all of these effects together. We construct market-level measures of total benefit and total cost from higher education to use as our dependent variables. Our objective is to estimate the effect of competition on these aggregate measures to give the policymaker a clear guidance on the social desirability of competition in higher education.

Among previous works discussing measures of education output that incorporate both the number of students and the quality of education are Hanushek and Woessmann (2015) and Unesco's Education at a Glance (UNESCO, 2018). Hanushek and Woessmann (2015) explores the relationship between education and economic growth and includes a chapter on measuring the quality of education. The authors discuss the use of test scores as a measure of educational quality and argue that such scores should be combined with measures of enrollment and completion rates to get a more accurate picture of education output, in a similar manner to what we do in this section.

Unesco's Education at a Glance (UNESCO, 2018) provides an overview of education systems around the world and includes a section on education output. The report includes a measure of education output that takes into account both the number of students and the quality of education by using indicators such as graduation rates and standardized test scores.

As indicated by Hanushek and Woessmann (2015) and (UNESCO, 2018), we constructed is a measure of the total benefit of higher education in a given city-subject combination. It is the log of mean final score of observed students in the program multiplied by the number of freshmen students, so that we have a program-level benefit measurement which we then sum across all programs in the same subject and city. We implement log-linear models here to simplify the interpretation so that a one unit change in our treatment can be read as a percentage effect on the outcome variable.

$$log(TB_{sc}) = \sum_{usc=1}^{u} E_{usc} \left(\frac{\sum_{iusc=1}^{n} FS_{iusc}}{n_{usc}} \right) \tag{4.7}$$

To calculate the log of total cost measure we multiply the program fees we constructed by the number of students enrolled, so we obtain a program-level cost measurement we can sum across all programs in the same subject-city to proxy the total investment in higher education in a market.

$$log(TC_{sc}) = \sum_{usc=1}^{u} (F_{usc} * E_{usc}) \tag{4.8}$$

If the expansion of higher education due to competition is not a process where increased enrollments are compensated by lower achievement, then we should see that competition is positive on total benefits. In terms of social costs we can have a similar interpretation in terms of the comparison of lower fees and increased enrollments.

Our interest, though, is in the comparison of the effects of competition on total benefits and total costs. We would expect that both of them would be increased with the expansion from competition, but ideally we would want the effect on total benefits to be higher than the effects on total costs if competition is to be welcomed into education.

Table 4.12: Social benefits

	OLS Benefit	OLS Cost	IV Benefit	IV Cost
HHI	-1.29***	-1.21***	-3.33***	-3.10***
	(0.146)	(0.135)	(0.346)	(0.333)
Subject	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes
Observations	4988	4988	4988	4988

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From these results on Table 4.12, we see that from a monopolistic market structure (HHI=1) to perfect competition (HHI≈0), the social benefits would increase by 333% while costs would increase by 310%. So that in our estimations, total benefits would outpace total costs. We display the first-stage regressions for this model on Table 4.26 of the appendix.

The effects of competition on social benefits and costs are driven by the expansion of access, and that expansion is not being compensated by lower achievement. Furthermore, the rate of expansion of social benefit from higher education due to competition is higher than the expansion of costs, which means that not only more competitive markets are getting more social benefits from education, but their costs are lower per unit of social benefit.

4.6 Conclusion

In this chapter, we demonstrated that education practitioners concerned with a so-called *commodification* of higher education due to market competition have been misled by a red herring. Competition indeed leads to lower mean tuition fees, higher enrollments, lower student achievement and program quality, but looking at the effects of competition at the means can be overly simplistic. What we find is that competition in fact expands participation to students of high and low ability students, but this effect is more pronounced among lower ability students. A similar effect is found in the quality of programs where competition drives a vertical differentiation effect where the range of quality is expanded in both directions. We find that all students are better off since they pay lower fees, have improved participation and find programs which suit their quality preferences.

We also find that competition has an overall positive social effect in education output. Total benefits from education are increased at a larger proportion than total costs, meaning that more competitive markets reap larger benefits from education at a lower cost.

4.7 Appendix

4.8 Data sources

Perhaps one of the most important contributions of this research has been the production of very large and rich datasets on the Brazilian higher education alongside with the Stata code that was used to compile them. We made thes datasets publicly available on the data repository available on this webpage.

We have used a number of datasets contained in the above repository. The original sources of the data are provided below.

Data	n. of data sets	Source
Programs	4	INEP/HE Census
Providers	4	INEP/HE Census
Academic staff	3	INEP/HE Census
Quality of programs	3	INEP/CPC
Students results	3	INEP/IDD
Fees	3	FNDE/FIES

Table 4.13: datasets employed

The original data on the Higher Education Census (all years) can be found here.

All of the other indicators produced by Instituto Nacional de Pesquisas Educacionais Anisio Teixeira (INEP) - the main government body in charge of higher educaiton institutional evaluation - are provided in one website.

The table below provides a little more empirical evidence that studying close to home is in fact a primary concern of students. According to the student questionnaire which is attached to the annual ENADE end-of-program evaluation, 20% of students in private higher education pointed to proximity to home as the primary reason for their choice of program and 10% pointed to fees. Other 40% pointed to quality and reputation of programs.

4.9 Alternative specifications

4.9.1 Quantile regressions

Throughout the paper, additionally to our regressions at the means, we also split our observations into subsamples to observe the effects of our concentration measure at different parts of the distribution without the constraints of the linear function.

Quantile regressions are designed to do exactly that. Quantile regressions can be used to obtain prediction intervals for the conditional distribution of our response variables ENADE, ENEM, and program quality in equations 4.4,4.5 and 4.6 (see Lehmann (1974) for the seminal paper).

Unlike traditional regression models that focus on the mean or median of the dependent variable, quantile regressions enable us to examine how the relationships between the independent and dependent variables differ at different points in the distribution. This makes them useful in situations where the relationships of interest may vary in different parts of the distribution, or when the data exhibits non-normal distributions, such as skewed or heavy-tailed data (Davino et al., 2013).

Quantile regressions are better than our split-sample approach in the sense that they employ the full sample on the estimation of the response function instead of only a subsample as in our approach. The results are displayed below.

This first quantile regression estimates the effect of the treatment variable at different parts of the distribution of students' entry scores. This should be compared to the split-sample results displayed at Table 4.5.

Table 4.14: Alternative specification - Quantile regression of student entry scores.

	OLS			IV		
	1st decile	Median	9th decile	1st decile	Median	9th decile
HHI	0.030	0.029***	0.027	0.26***	0.25***	0.25***
	(0.022)	(0.011)	(0.020)	(0.054)	(0.028)	(0.051)
Observations	324616	324616	324616	324616	324616	324616

Standard errors in parentheses

Standard errors clustered at municipal and subject levels

* $p<0.10$, ** $p<0.05$, *** $p<0.010$

The results from this model are very different from what we obtained from our split-sample approach, and would lead to a different story. From the IV results above, we see that in all three estimates concentration is positive on student entry scores with very similar coefficients. Compare this with our split-sample results, where concentration was negative on the entry scores for top students, and positive for entry scores of the bottom students.

This is because quantile regressions and our split-sample estimates are very different strategies. First, in our split-sample regressions we consider the distribution of the response variable within each city and subject combination, the same specification of our HHI variable, which is also calculated for each subject-city pair. Quantile regressions on the other hand consider the distribution of the outcome variable across the whole sample.

In the regressions we used in the paper, therefore, we are not examining the effect of the treatment on specific intervals of the *overall* distribution of the dependent variable, but *on specific intervals of subsamples of the dependent variable within markets* i.e. we are examining the effects of concentration for observations ranked *within* each city-subject combinations.

Second, the split-sample regressions estimate the effect of concentration at the top and bottom *counts* of students within each city-subject while the quantile regressions estimate the effect of the variable at different percentiles of the distribution of the outcome variable across all observations.

We also include below the quantile model for student's final scores and for our proxy of program quality.

Table 4.15: Alternative specification - Quantile regression of student final scores.

	OLS			IV		
	1st decile	Median	9th decile	1st decile	Median	9th decile
HHI	0.081***	0.081***	0.082***	0.34***	0.34***	0.34***
	(0.023)	(0.013)	(0.024)	(0.059)	(0.032)	(0.060)
Observations	324616	324616	324616	324616	324616	324616

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* $p<0.10$, ** $p<0.05$, *** $p<0.010$

Table 4.16: Alternative specification - Quantile regression of student final scores.

	OLS			IV		
	1st decile	Median	9th decile	1st decile	Median	9th decile
HHI	0.078***	0.21***	0.36***	0.66***	1.05***	1.50***
	(0.020)	(0.012)	(0.023)	(0.050)	(0.029)	(0.058)
Observations	324616	324616	324616	324616	324616	324616

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* $p<0.10$, ** $p<0.05$, *** $p<0.010$

Finally, because intervals are based on the conditional distribution of response variables, quantile regressions would not be applicable to our model on fees in equation 4.11, as the sample split in that particular case was done based not on the response variable but on the distribution of a third variable on program quality.

4.9.2 Robustness checks

In the Table 4.17 below, we present an alternative model specification to that displayed in in Section 4.3. Because we have observed in Figure **??** that the distribution of our calculated HHI for all subject city-combinations is skewed with many HHI=1, we include an alternative specification of the model 4.3 whose results are presented in Table 4.6 adding dummies for those markets where there is only one provider present. We find that our results change slightly, but that our IV specification is in fact robust to that constraint, which indicates that our results are not driven by the monopoly markets.

Table 4.17: Alternative specification - Market-level fees and enrollments with dummy for monopolistic markets.

	OLS	IV	FS	OLS	IV	FS
HHI	65.2	697.4***		-1236.5***	-4001.6***	
	(45.742)	(234.966)		(194.545)	(781.794)	
HHI (2009)			0.17***			0.17***
			(0.021)			(0.021)
Subject	Yes	Yes	No	Yes	Yes	No
City	Yes	Yes	No	Yes	Yes	No
Observations	4988	4988	4988	4988	4988	4988

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* p<0.10, ** p<0.05, *** p<0.010

The results presented below are from a similar robustness check on our regressions of student achievement regression whose results are shown in Table 4.7. In this case, once again we observe some changes in the sizes of coefficients, but still not enough to change our conclusions or to suspect that our results are being driven by the monopoly markets.

Table 4.18: Alternative specification of the achievement models - dummy for monopolistic markets

	OLS	OLS	OLS	IV	IV	IV
HHI	0.098*	0.086*	0.059	0.58***	0.32***	-0.015
	(0.055)	(0.045)	(0.046)	(0.177)	(0.101)	(0.116)
Student ability		0.58***	0.57***		0.58***	0.56***
		(0.009)	(0.010)		(0.009)	(0.010)
Program Quality			0.091***			0.18***
			(0.006)			(0.066)
Monopoly (d)	-0.012	-0.015	-0.0100	-0.21**	-0.11**	0.015
	(0.031)	(0.026)	(0.022)	(0.082)	(0.049)	(0.049)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	324616	324616	324616	324616	324616	324616

Standard errors in parentheses
Standard errors clustered at market level
* p<0.10, ** p<0.05, *** p<0.010

In the alternative specification below, we provide a robustness check for the case where our results presented in Table 4.7 are sensitive to changes we have done to the variables such as the standardization of ENADE and ENEM, our measures of achievement and ability.

Table 4.19: Alternative specification of the achievement model - non standardized measures of achievement and ability

	OLS	OLS	OLS	IV	IV	IV
HHI	1.13***	0.88**	0.64*	4.33***	2.31***	0.27
	(0.372)	(0.335)	(0.353)	(1.197)	(0.830)	(1.008)
Student entry score		1.04***	1.02***		1.04***	1.01***
		(0.028)	(0.030)		(0.028)	(0.031)
Program quality			1.11***			1.95**
			(0.079)			(0.802)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	324616	324616	324616	324616	324616	324616

Standard errors in parentheses

Standard errors clustered at market level

* $p<0.10$, ** $p<0.05$, *** $p<0.010$

In the model below we display an alternative specification of the achievement model presented in Table 4.7 where we include the staff qualification variable itself into the equation, and not as an instrument to the program quality variable as we did to obtain in the results we used. We can see that in this case, the staff qualification is not as strong a control for program quality as the instrumented quality control we used.

Table 4.20: Alternative specification of the achievement models - alternative specifications of the education quality control

	IV	IV	IV	IV	IV
HHI	0.34***	0.20***	0.20***	0.100**	0.19***
	(0.090)	(0.052)	(0.052)	(0.046)	(0.051)
Student ability		0.58***	0.58***	0.57***	0.58***
		(0.009)	(0.009)	(0.010)	(0.009)
Lagged final scores				0.091***	
				(0.006)	
Staff with MSc					0.023***
					(0.008)
Subject	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes
Observations	324616	324616	324616	324616	324616

Standard errors in parentheses
Standard errors clustered at market level
* p<0.10, ** p<0.05, *** p<0.010

In the Table 4.20 below, the three first regressions are repetitions of the IV models we presented in Section 4.4 to serve as comparisons to the two last regressions here presented. The second last regression uses the lagged final scores directly as a control, without instrumenting. The last regression includes instead the proportion of staff with a postgraduate degree as a measure of program quality.

Table 4.21: Effects of competition on top and bottom quality programs by market - staff qualification

	OLS	OLS	OLS	IV	IV	IV
	All	Top quality	Bottom quality	All	Top quality	Bottom quality
HHI	0.082	-0.0047	0.18**	0.58***	0.22	1.04***
	(0.076)	(0.092)	(0.078)	(0.187)	(0.198)	(0.192)
Subject	Yes	Yes	Yes	Yes	Yes	Yes
City	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8469	4638	4602	8469	4638	4602

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* p<0.10, ** p<0.05, *** p<0.010

4.10 First stages

The following tables present the first stages of the IV models we implemented throughout the paper, except for the first stages of our main model on student achievement in Table 4.7, which is presented in the main body of the paper in Table 4.27.

The first stages below relate to the results of our model on student ability presented in Table 4.8.

Table 4.22: First stages - Effect of competition on student ability

	All	Top 10	Bottom 10
HHI (2009)	0.34***	0.42***	0.42***
	(0.045)	(0.032)	(0.032)
Observations	324616	47643	47643
F excl inst	58.19	167.98	167.93

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* $p<0.10$, ** $p<0.05$, *** $p<0.010$

The first stages below relate to the results of our model on student achievement presented in Table 4.23.

Table 4.23: First stages - Effect of competition on students' achievement

	All	Top 10	Bottom 10
HHI (2009)	0.34***	0.42***	0.42***
	(0.045)	(0.032)	(0.032)
Observations	324616	47643	47643
F excl inst	58.19	167.98	167.93

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* p<0.10, ** p<0.05, *** p<0.010

The first stages below relate to the results of our model on program quality presented on Table 4.10.

Table 4.24: First stages - Effects of competition on program quality

	All	Top quality	Bottom quality
HHI (2009)	0.38***	0.43***	0.42***
	(0.020)	(0.018)	(0.019)
Observations	8469	4638	4602
F excl inst	354.73	551.76	521.66

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* p<0.10, ** p<0.05, *** p<0.010

The first stages below relate to the results of our model on program fees presented on Table 4.11.

Table 4.25: First stages - Fee differentiation

	All	Top quality	Bottom quality
HHI (2009)	0.38***	0.43***	0.42***
	(0.020)	(0.018)	(0.019)
Observations	8469	4638	4602
F excl inst	354.73	551.76	521.66

Standard errors in parentheses
Standard errors clustered at municipal and subject levels
* p<0.10, ** p<0.05, *** p<0.010

The first stages below relate to the results of our model on social benefits presented on Table 4.12.

Table 4.26: First stage - Social benefits

	First stages - Benefit	Cost
HHI (2009)	0.42***	0.42***
	(0.032)	(0.032)
Observations	4988	4988
F excl inst	174.65	174.65

Standard errors in parentheses

Standard errors clustered at municipal and subject levels

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

And finally the first stages of our main regression of student final scores on controls including our HHI measurement on Table 4.7

Table 4.27: First stages of main student output regression

	IV 1	IV 2	IV 3 - HHI	IV 3 - Quality
HHI (2009)	0.34***	0.34***	0.34***	0.34***
	(0.045)	(0.045)	(0.045)	(0.077)
Student ability		-0.00028	-0.00038	0.085***
		(0.001)	(0.001)	(0.013)
Staff with MSc			0.0021	0.12***
			(0.002)	(0.030)
Observations	324616	324616	324616	324616
F excl inst	58.19	58.27	29.35	24.23

Standard errors in parentheses

Standard errors clustered at market level

* p<0.10, ** p<0.05, *** p<0.010

4.11 Other empirical evidence

Table 4.28: Why did you choose this provider? ENADE student questionnaire, 2016

	Public	Private				
	Total	Total	Medicine	Social services	Social services higher 25% scores	Social services lower 25% scores
Gratuity	41.22	1.54	0.61	3.32	3.59	3.91
Fee	0.79	9.13	2.22	16.61	15.65	20.49
Close to residence	9.77	19.67	16.04	16.63	16.08	16.92
Close to work	0.22	1.9	0.13	2.25	2.05	2.71
Access	1.52	7.5	1.47	10.76	9.62	11.00
Quality/ reputation	38.81	39.36	35.65	28.01	30.37	23.11
Only offer	2.68	2.85	23.91	1.29	1.20	1.72
Scholarship possibility	0.62	7.92	7.65	9.19	10.18	7.82
Other	4.36	10.14	12.31	11.91	11.25	12.33
N	47,117	87,650	7,199	9,895	2,328	2,328

Chapter 5

Conclusion and directions for further research

As part of a shift from elitist to mass education systems, the past century has seen an expansion in higher education provision in most developed and developing countries. As national budgets are constrained and demand is continuously expanding, countries have resorted to a combination of private provision and tuition financing to fuel a sustainable expansion while relying on market competition to keep fees at bay.

In this thesis we investigate two important aspects of this expansion: the effects of tuition financing on student enrollments and migration from cities of different sizes, and we also look at the effects of provider competition on city-level higher education outputs, such as enrollments and education quality.

After a brief background to Brazil provided in Chapter 1, we revisit the economics of education literature in Chapter 2, focusing on the areas which are of particular relevance for the work we conducted on the effects of competition on student and social outcomes, and on the effects of loan programs on student migration. These are the topics for the research we presented in Chapter 3 on domestic student migration and Chapter 4 on competition among universities.

5.1 Student migration

In this chapter we have looked at the effects of the rise and demise of a large classroom-based student financing scheme on regional higher education imbalances in Brazil. More specifically, we looked at the differential effects of the policy on student migration, enrollments and supply coverage in cities of different sizes. From our student-level models we found that students from smaller cities were more likely to relocate if given financing, but that the destination choices for those students were in fact other smaller cities. We also found that students from larger cities were more prone to relocate to smaller cities under the financing scheme, indicating that our hypothesis o brain drain from rural to urban areas was largely incorrect. We found instead a brain drain from large to small cities, and an increased mobility among smaller cities. In our city-level aggregate models we have shown that the total number of enrollments is increasing with the scheme, but especially the enrollments at home and in smaller rural locations. This indicates that the policy is increasing mobility and the number of students from smaller cities at the same time.

5.1.1 Further research

Our research into student migration relates a canon of literature which has seen little action in recent years, and which could profit greatly from its interaction with other fields of research such as economics of education and development economics.

A proportion of the literature we revisited in Chapter 3 is focused on the theoretical modelling of migration. But a theoretical framework for students' decision making is still absent from our paper. The next steps for this chapter are going to be the development of this framework so that our paper can be better grounded and the empirical modelling can be calibrated. A potential start for this work is likely provided by De Fraja and Iossa (2002) in their theoretical modeling of student and university matching.

5.2 Effects of competition on education outcomes

In this chapter, we demonstrate that education practitioners concerned with a so-called commodification of higher education due to market competition may have been misled by a red herring. Competition indeed leads to lower mean tuition fees, higher enrollments, lower student achievement and program quality, but looking at the effects of competition at the means can be overly simplistic. We find that competition contributes to expand participation for students of high and low ability, although this effect is more pronounced among lower ability students, which virtually drives down the mean ability of the pool of students. A similar effect is found in the quality of programs where competition contributed to a vertical differentiation effect where the range of quality is expanded in both directions as more providers enter the market, but the mean quality of the sector is lowered.

Overall, we find that all students are better off since they pay lower fees, have improved participation and find programs which are better suited to their quality preferences. We also find that competition has an overall positive effect in education output for the society as a whole. Total benefits from education are increased at a larger proportion than total costs, meaning that more competitive markets reap larger benefits from education at a lower cost.

5.2.1 Further research

Our chapter on the effects of competition on student outcomes will be reworked into a publishable paper, and that will require considerable refinement. Differently from our chapter on student migration, though, this paper may not demand any additional working parts, which means we should start to consider the next topics for our research into this field.

As we mentioned in previous chapters, Brazil presents an important opportunity for future economic research as data is fairly abundant, public and free, and researchers outside of Brazil have yet to tap into the resources. One of these is the RAIS (Relacao Anual de Informacoes Sociais - Annual Report of Social Information datasets, which provide a registry of all formal workers in Brazil, their levels of education, their industry and position title, among several other variables.

This dataset provides an opportunity for us to expand our investigation of the effects of competition on education social outcomes. We would like to provide insight into whether the increase in competition in higher education in a given location would lead to observable and measurable effects in the labour market, such as on the range and quality of jobs in that location, possibly drawing from the more recen research conducted by Hanushek and Woessmann (2017)

5.3 Datasets

Perhaps one of the most important contributions of this research has been the production of very large and rich datasets on the Brazilian higher education alongside with the Stata code that was used to compile them. These datasets are publicly available on The data repository available on my personal webpage.

Bibliography

Alm, J. and Winters, J. V. (2009). Distance and intrastate college student migration. *Economics of Education Review*, 28(6):728–738.

Altbach, P. G., Reisberg, L., and Rumbley, L. E. (2009). *Trends in global higher education: Tracking an academic revolution*. UNESCO.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Armor, D. et al. (1976). Analysis of the school preferred reading program in selected los angeles minority schools.

Baryla Jr, E. A. and Dotterweich, D. (2001). Student migration: Do significant factors vary by region? *Education economics*, 9(3):269–280.

Beals, R. E., Levy, M. B., and Moses, L. N. (1967). Rationality and migration in ghana. *The Review of Economics and Statistics*, pages 480–486.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political economy*, 81(4):813–846.

Belfield, C. R. and Levin, H. M. (2002). The effects of competition between schools on educational outcomes: A review for the united states. *Review of Educational research*, 72(2):279–341.

Borland, M. V. and Howsen, R. M. (1992). Student academic achievement and the degree of market concentration in education. *Economics of education review*, 11(1):31–39.

Borland, M. V. and Howsen, R. M. (1993). On the determination of the critical level of market concentration in education. *Economics of Education Review*, 12(2):165–169.

Borland, M. V. and Howsen, R. M. (1996). Competition, expenditures and student performance in mathematics: A comment on couch et al. *Public Choice*, 87(3-4):395–400.

Bradley, S., Johnes, G., and Millington, J. (1999). School choice, competition and the efficiency of secondary schools in england. *Discussion Paper. Lancaster Management School.*

Broecke, S. and Hamed, J. (2008). Gender gaps in higher education participation: An analysis of the relationship between prior attainment and young participation by gender, socio-economic class and ethnicity.

Browne, E. (2017). Evidence on education as a driver for migration.

Carvalho, C. H. A. d. (2013). The commodification of higher education and the market strategies of for-profit institutions. *Revista Brasileira de Educação*, 18(54):761–776.

Cooke, T. J. and Boyle, P. (2011). The migration of high school graduates to college. *Educational Evaluation and Policy Analysis*, 33(2):202–213.

Cooper, S. T. and Cohn, E. (1997). Estimation of a frontier production function for the south carolina educational process. *Economics of Education Review*, 16(3):313–327.

Creemers, B. P. and Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School effectiveness and school improvement*, 7(3):197–228.

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Davino, C., Furno, M., and Vistocco, D. (2013). *Quantile regression: theory and applications*, volume 988. John Wiley & Sons.

De Fraja, G. and Iossa, E. (2002). Competition among universities and the emergence of the elite institution. *Bulletin of economic Research*, 54(3):275–293.

De Jong, G. F. (2000). Expectations, gender, and norms in migration decision-making. *Population studies*, 54(3):307–319.

Del Rey, E. and Estevan, F. (2015). University competition: An application to brazil. *Working paper*.

Dewey, J., Husted, T. A., and Kenny, L. W. (2000). The ineffectiveness of school inputs: a product of misspecification? *Economics of education review*, 19(1):27–45.

Eagles, C. W. (1986). Urban-rural conflict in the 1920s: A historiographical assessment. *The Historian*, 49(1):26–48.

Elder, S., de Haas, H., Principi, M., and Schewel, K. (2015). *Youth and rural development: Evidence from 25 school-to-work transition surveys*. ILO Geneva, Switzerland.

Evans, W. N., Froeb, L. M., and Werden, G. J. (1993). Endogeneity in the concentration–price relationship: Causes, consequences, and cures. *The Journal of Industrial Economics*, pages 431–438.

Fernandez, R. and Gali, J. (1999). To each according toâ¦? markets, tournaments, and the matching problem with borrowing constraints. *The Review of Economic Studies*, 66(4):799–824.

Figlio, D. N. (1997). Teacher salaries and teacher quality. *Economics Letters*, 55(2):267–271.

Ganley, J. A. and Cubbin, J. S. (1992). *Public sector efficiency measurement: Applications of data envelopment analysis*. Elsevier Science Inc.

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

Giannoccolo, P. (2009). The brain drain: a survey of the literature. *Università degli Studi di Milano-Bicocca, Department of Statistics, Working Paper*, (2006-03):02.

Gibbons, S. and Vignoles, A. (2012). Geography, choice and participation in higher education in england. *Regional science and urban economics*, 42(1-2):98–113.

Goldhaber, D. D. and Brewer, D. J. (1997). Why don't schools and teachers seem to matter? assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, pages 505–523.

Groen, J. A. (2004). The effect of college location on migration of college-educated labor. *Journal of Econometrics*, 121(1-2):125–142.

Gupta, M. S., Tiongson, M. E., and Verhoeven, M. (1999). *Does higher government spending buy better results in education and health care?* International Monetary Fund.

Hagerstrand, T. (1957). Migration and area. *Lund Studies in Geography*, (Series B, 13).

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, 61(2):280–288.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of human Resources*, pages 351–388.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of economic literature*, 24(3):1141–1177.

Hanushek, E. A., Machin, S., and Woessmann, L. (2000). Economics of education. *International Encyclopedia of the Social & Behavioral Sciences*, pages 4201–4208.

Hanushek, E. A. and Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth.* MIT Press.

Hanushek, E. A. and Woessmann, L. (2017). School resources and student achievement: A review of cross-country economic research. *Cognitive abilities and educational outcomes: A festschrift in honour of jan-eric gustafsson*, pages 149–171.

Hanushek, E. A., Woessmann, L., and Machin, S. J. (2023). *Handbook of the economics of education.* Elsevier.

Harris, J. and Todaro, M. (1970). Migration, unemployment, and development: A two-sector analysis. *American Economic Review*, 60:126–142.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, pages 1251–1271.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Holmlund, H., McNally, S., and Viarengo, M. (2010). Does money matter for schools? *Economics of education review*, 29(6):1154–1164.

Hoxby, C. M. (1997). How the changing market structure of us higher education explains college tuition.

Hyman, J. (2017). Does money matter in the long run? effects of school spending on educational attainment. *American Economic Journal: Economic Policy*, 9(4):256–280.

IDados (2016). Financiamento do ensino superior: Fies e prouni. *https://blog.idados.id/financiamento-do-ensino-superior-fies-e-prouni*, (20-04-2019).

Imai, K. and Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, 29(3):405–415.

INEP (2018). Results from the higher education census. *http://portal.inep.gov.br/web/guest/microdados*, (2019-04-29).

Jackson, C. K., Johnson, R. C., and Persico, C. (2015). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. Technical report, National Bureau of Economic Research.

Jepsen, C. and Montgomery, M. (2009). Miles to go before i learn: The effect of travel distance on the mature person's choice of a community college. *Journal of Urban Economics*, 65(1):64–73.

Kanaroglou, P. S. and Ferguson, M. R. (1998). The aggregated spatial choice model vs. the multinomial logit: an empirical comparison using migration microdata. *Canadian Geographer/Le Géographe Canadien*, 42(3):218–231.

Kirjavainen, T. and Loikkanent, H. A. (1998). Efficiency differences of finnish senior secondary schools: an application of dea and tobit analysis. *Economics of Education Review*, 17(4):377–394.

Kono, S. and Shio, M. (1965). *Inter-prefectural migration In Japan, 1956 and 1961: migration stream analysis*, volume 3. Asia Publishing House.

Krieg, R. G. (1993). Black-white regional migration and the impact of education: a multinomial logit analysis. *The Annals of regional science*, 27(3):211–222.

Kropko, J. (2007). *Choosing between multinomial logit and multinomial probit models for analysis of unordered choice data*. PhD thesis, The University of North Carolina at Chapel Hill.

Lehmann, E. (1974). Statistical methods based on ranks.

Levy, M. B. and Wadycki, W. J. (1974). Education and the decision to migrate: An econometric analysis of migration in venezuela. *Econometrica: journal of the Econometric Society*, pages 377–388.

López-Torres, L. and Prior, D. (2022). Long-term efficiency of public service provision in a context of budget restrictions. an application to the education sector. *Socio-Economic Planning Sciences*, 81:100946.

Mak, J. and Moncur, J. E. (2003). Interstate migration of college freshmen. *The Annals of Regional Science*, 37(4):603–612.

Maranto, R., Milliman, S., and Stevens, S. (2000). Does private school competition harm public schools? revisiting smith and meier's: The case against school choice. *Political Research Quarterly*, 53(1):177–192.

Mayston, D. J. (1996). Educational attainment and resource use: Mystery or econometric misspecification? *Education Economics*, 4(2):127–142.

Mayston, D. J. and Jesson, D. (1999). *Linking educational resourcing with enhanced educational outcomes*. DfEE.

McHugh, R. and Morgan, J. N. (1984). The determinants of interstate student migration: A place-to-place analysis. *Economics of Education Review*, 3(4):269–278.

Mixon, F. G. (1992). Factors affecting college student migration across states. *International Journal of Manpower*.

Murnane, R. J. (1975). The impact of school resources on the learning of inner city children.

Murnane, R. J. and Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research*, 10(1):83–100.

Murphy, T. P. (1969). Political economy and geographic distribution of federal funds for research and development: The midwest case. *Administrative Science Quarterly*, pages 426–441.

Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787.

Parsad, B., Gray, L., and Tsapogas, J. (2005). *Interstate Migration Patterns of Recent Recipients of Bachelor's and Master's Degrees in Science and Engineering: Special Report*. National Science Foundation, Division of Science Resources Statistics.

Punch, S. and Sugden, F. (2013). Work, education and out-migration among children and youth in upland asia: changing patterns of labour and ecological knowledge in an era of globalisation. *Local Environment*, 18(3):255–270.

Reynolds, D., Sammons, P., Stoll, L., Barber, M., and Hillman, J. (1996). School effectiveness and school improvement in the united kingdom. *School effectiveness and school improvement*, 7(2):133–158.

Romero, L. and Rey, E. d. (2004). Competition between public and private universities: quality, prices and exams.

Russell, L. (2017). Market power effects of college and university mergers. *Massachussetts Institute of Technology, Cambridge, US*.

Rutter, M. (1982). *Fifteen thousand hours: Secondary schools and their effects on children*. Harvard University Press.

Sahota, G. S. (1968). An economic analysis of internal migration in brazil. *Journal of political economy*, 76(2):218–245.

Sattinger, M. (1995). Search and the efficient assignment of workers to jobs. *International Economic Review*, pages 283–302.

Schultz, T. P. (1971). Rural-urban migration in colombia. *The Review of Economics and Statistics*, pages 157–163.

Schultz, T. P. (1982). Lifetime migration within educational strata in venezuela: estimates of a logistic model. *Economic Development and Cultural Change*, 30(3):559–593.

Shumar, W. (2013). *College for sale: A critique of the commodification of higher education*. Routledge.

Simon, C. A. and Lovrich, N. P. (1996). Private school enrollment and public school performance. *Policy Studies Journal*, 24(4):666–675.

Smith, K. B. and Meier, K. J. (1995). Public choice in education: Markets and the demand for quality education. *Political Research Quarterly*, 48(3):461–478.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, page 557â586.

Stewart, J. Q. (1948). Demographic gravitation: evidence and applications. *Sociometry*, 11(1/2):31–58.

Storper, M. and Scott, A. J. (2009). Rethinking human capital, creativity and urban growth. *Journal of economic geography*, 9(2):147–167.

Stouffer, S. A. (1940). Intervening opportunities: a theory relating mobility and distance. *American sociological review*, 5(6):845–867.

Thanassoulis, E. (1993). A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments. *Journal of the operational research society*, 44(11):1129–1144.

Tight, M. (2007). The (re) location of higher education in england (revisited). *Higher Education Quarterly*, 61(3):250–265.

Tuckman, H. P. (1970). Determinants of college student migration. *Southern Economic Journal*, pages 184–189.

UNESCO (2018). *Education at a Glance*. UNESCO Publishing, Paris.

Vignoles, A., Levacic, R., Walker, J., Machin, S., and Reynolds, D. (2000). *The relationship between resource allocation and pupil attainment: a review*. Number 2. Centre for the Economics of Education, London School of Economics and â¦.

Wang, Y. and Bellemare, M. F. (2019). Lagged variables as instruments.

Wedel, K. (2021). Instruction time and student achievement: The moderating role of teacher qualifications. *Economics of Education Review*, 85:102183.

Wilkinson, L. and Wilkinson, M. (2020). Value for money and the commodification of higher education: front-line narratives. *Teaching in Higher Education*, pages 1–17.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.

Wrinkle, R. D., Stewart Jr, J., and Polinard, J. (1999). Public school quality, private schools, and race. *American Journal of Political Science*, pages 1248–1253.