Conceição, Otávio Canozzi; Oliveira, Rodrigo Carvalho; Souza, André Portela

Working Paper

The impacts of studying abroad: Evidence from a government-sponsored scholarship programme in Brazil

WIDER Working Paper, No. 2023/49

Provided in Cooperation with:
United Nations University (UNU), World Institute for Development Economics Research (WIDER)

Suggested Citation: Conceição, Otávio Canozzi; Oliveira, Rodrigo Carvalho; Souza, André Portela (2023) : The impacts of studying abroad: Evidence from a government-sponsored scholarship programme in Brazil, WIDER Working Paper, No. 2023/49, ISBN 978-92-9267-357-4, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki, https://doi.org/10.35188/UNU-WIDER/2023/357-4

This Version is available at:
https://hdl.handle.net/10419/283745

UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2023/49

The impacts of studying abroad

Evidence from a government-sponsored scholarship programme in Brazil

Otavio Conceição,[1] Rodrigo Oliveira,[2] and André Portela Souza[1]

March 2023

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: This paper investigates the impact of the Science without Borders (*Ciência sem Fronteiras*) (CSF) programme on participants' post-graduation enrolment, employment, and entrepreneurship. The programme was launched in 2011 to increase students' human capital and interest in science and postgraduate education studies through a substantial increase in scholarships for Brazilians to carry out part of their undergraduate studies abroad. We exploit variation in the approval rate across CSF selection calls for the same destination country and year and combine 17 public and private administrative records to track CSF candidates' outcomes up to eight years after the call. The main results suggest that the programme did not achieve its goals of increasing approved student enrolment in postgraduate education programmes in Brazil. Even though the programme could have improved student skills and acted as a market signalling, we do not find effects on the probability of working in the formal labour market, or as formal entrepreneurs. Using detailed data from one top university, we show that approved students graduate more often, but take longer to graduate, which may have negative impacts on their labour market outcomes. Finally, although we cannot rule out that students moved to a foreign country after the programme, we show that the likelihood of this event may have decreased over time.

Key words: study abroad, education, Brazil, Ciência sem Fronteiras

JEL classification: H00, I23, J01

Acknowledgements: We thank Antonio Leon, Bruno Ferman, Breno Braga, Claudio Ferraz, Edson Severnini, Kelly Gonçalves, Guilherme Lichand, Luiz Felipe Fontes, Marco Tulio França, Murilo Sepulvida, Nancy Chau, Jessica Miranda, Patricia Justino, Paulo Jacinto, Willian Adamczyk, and the participants of academic seminars and international conferences for their valuable comments. This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil* (CAPES) – Finance Code 001. It was also approved by the Getúlio Vargas Foundation (FGV) IRB (Approval Letter No. 45/2021). Otavio acknowledges financial support from CAPES, FGV-EESP, and FAPESP - Grant No. 2021/00470-3. We are very grateful to the higher education institutions we partnered with and to CAPES, CNPq, and RFB for providing data for this research. The paper is based on Otavio's PhD thesis.

[1] Sao Paulo School of Economics – FGV, Brazil, otavio.canozzi@gmail.com, andre.portela.souza@fgv.br [2] UNU-WIDER, Finland, oliveira@wider.unu.edu

This study is published within the UNU-WIDER project Academic excellence.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-357-4

https://doi.org/10.35188/UNU-WIDER/2023/357-4

Typescript prepared by Siméon Rapin.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

The number of students enrolled in higher education institutions abroad reached 5.3 million in 2017, more than double the total in 2000 (UIS 2023). Studying abroad is an opportunity to access high-quality university education that is scarce in some developing countries. Foreign students' share in U.S. higher education more than doubled between 1980 and 2018. They represent 5% of all bachelor's degree enrolments, 18% of all master's degree enrolments, and 13% of all doctoral degree enrolments. The increase in international students is also a phenomenon at universities in Europe, Australia, Canada, the UK, China, and Índia (Bound et al. 2021). However, in most cases, studying abroad is exclusive to a minority elite who can pay for the education or who went to the best private elementary and high schools in their home country.

International student mobility programmes have risen sharply in previous decades as a way to provide access to a short-term experience abroad for students enrolled in higher-education institutions in their home country. These programmes aim to boost students' human capital, expand their cultural perspectives, and create international networks among partner countries. In developed countries, this kind of programme works mostly as a cultural exchange, as in the case of the most well-known exchange programme worldwide—the ERASMUS Programme in Europe—which also aims to contribute to European integration. In low- and middle-income countries with low science and education investments, these programmes are more critical to enhancing the national quality of human capital in technical fields such as science, technology, engineering, and mathematics (STEM), viewed as the base for economic development. Although many developing countries have launched international student mobility programmes, little is known about the impacts of these policies on individual employment, post-graduation enrollment, and entrepreneurship.[1]

This paper investigates the impacts of a massive government-sponsored study abroad programme in Brazil, named Ciência sem Fronteiras (CSF)—Science without Borders—on postgraduate education, the labour market, and entrepreneurship. We are the first paper to provide causal estimates of such programmes on those outcomes for developed and developing countries. We built a new and unique data set by combining seventeen administrative records from different sources that allow tracking students up to eight years after the application to the programme. Most data sets are not public and were obtained through formal requests to the Brazilian authorities using the Access to Information Law (LAI).[2] The records were linked using the candidates' names and the six intermediary digits of the taxpayer identification number in Brazil.

Our first main data set contains information about all applicants and winners for the programme in Brazil through formal requests to the Ministry of Science and Technology and the Ministry of Education. The second main data set has detailed academic records from enrolled students in thirteen Brazilian federal universities that applied to CSF, which were obtained directly through formal requests to each university. Even though we requested data to all Brazilian federal universities that participated in the programme, only thirteen answered. The final sample contains 19,245 students who applied to the programme, with 50.03% being approved.

Our sample covers all Brazilian macro-regions and almost 50% of the Brazilian states. The pool of universities is geographically diverse, with nine ranked among the thirty institutions with the highest number of CSF applicants. Out of the 102 calls launched by the programme, we have information about candidates that applied for 97 calls. After merging the programme information with students' academic

[1] This is the case for Mexico's Proyecta 100,000, Colombia's COLFUTURO, Chile's Becas Chile, Argentina's Bec.Ar, Kazakhstan's Bolashak and Saudi Arabia's King Abdullah Scholarship Program (KASP).

[2] http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm

records, we use three other administrative data sets to track the student's presence in the formal labour market, postgraduate education in Brazil, and the formal entrepreneurship market after applying to the programme.

The Brazilian government launched the programme in 2011 to promote student and professional exchange in the STEM fields. The programme was initially designed to last four years, operating through a substantial increase in scholarships for Brazilians to carry out part of their undergraduate studies in a foreign institution located in a developed country, mainly in Europe and North America. CSF increased the number of Brazilian students abroad by eightfold in only four years (McManus and Nobre 2017; Senado Federal 2015; Cruz and Eichler 2021), providing more than 90,000 scholarships for Brazilians to attend foreign universities.

In comparison, Brazil offered 13,819 undergrad, Ph.D., and post-doc scholarships between 1987 and 2000 (Mazza 2009). Consequently, the programme represented a unique inflection in the historical pattern of Brazilian policies on international student exchange.[3] The programme was expensive. It cost US$ 2.72 billion (US$ 27.200 per student), which represents five times the average cost to maintain a student in a public university per year and is equivalent to the cost of a nationwide school meal programme that benefits 39 million students (FAPESP 2017).[4] It also represents almost fifteen times the budget of the main research funding agency of the Ministry of Science and Technology in 2016. To put it into an international context, the European Commission spent € 14 billion on the ERASMUS programme between 2014 and 2020.[5]

Measuring the impact of study abroad programmes is not trivial because of the selection into the programme. Some students are more self-motivated, receive more parental incentives, and have more financial assets and preparation, such as language proficiency. Therefore, to estimate the causal impacts of the programme on the participants' life prospects, we develop a novel instrumental variable approach that exploits variation in the approval rate across CSF calls for the same destination country and launching year. The instrument is as good as random because when a specific call for scholarships was in place, the students could not predict whether new calls would be launched and whether they would be more competitive than the current ones. They also did not know if there would be more calls for the same destination country in the same year. Both the programme schedule and budget were not public. Instead, the number of approved candidates in each call was defined ex-post. It was mainly driven by budgetary availability that should not be related to the determinants of candidates' approval.

The main results suggest that the programme did not achieve its main goals of increasing the presence of CSF beneficiaries in postgraduate education programmes, in the formal labour market, and as entrepreneurs up to six years after each call. Approved applicants are 12.5 percentage points (p.p.) less likely to pursue a postgraduate degree in a Brazilian university in the first three years after the call and 7.1 p.p. less likely between four to six years after the call. They are also 3.8 p.p. less likely to have a formal job in the first three years after the call, but we do not find a statistically significant effect between four to six years after the call. The results for formal entrepreneurship are similar to those for labour market outcomes.

[3] Appendix Figures A1 and A2 provide some descriptive evidence about the Brazilian government's spending on science and technology and the number of undergraduate scholarships granted by the Brazilian government between 2000 and 2020. The period 2011–17, when the CSF undergraduates were attending foreign universities, is clearly an inflection in the historical pattern.

[4] https://exame.com/brasil/o-corte-do-ciencia-sem-fronteiras-em-numeros/

[5] https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1326

To understand the mechanisms behind our findings, we explore the Federal University of Bahia (UFBA) data, the only university that provided the students' full academic records.[6] UFBA is the second-largest university in the northeast region of Brazil and one of the fifteen most prominent universities in the country. We show that students exposed to the programme are more likely to graduate, even though they take more time to finish their university degrees. Therefore, delayed graduation seems to be an important mechanism of our results, as it postpones students' exposure to the labour market, which is an important explanation for labour market outcomes in developing countries (Le Barbanchon et al. 2023).

One may think that another mechanism arises because the programme can increase the number of students who didn't come back to Brazil or decided to return to a developed country to study or work after the programme. This is known as brain drain and is a concern in the context of the ERASMUS programme. Even though we cannot rule out the brain drain issue, we provide some evidence that its occurrence will decrease over time. Our results indicate that from the perspective of the programme implementer, the policy has not achieved its goals.

This paper innovates in two ways. First, we are the first paper to estimate the causal impact of a massive government-sponsored study abroad programme in a developing country. Despite the existence of some literature about study abroad programmes, most papers focus on European countries and the ERASMUS programme (Giorgio 2021; Netz and Cordua 2021). In addition, only three papers provide causal estimates of studying abroad. For instance, they study international labour mobility (Parey and Waldinger 2011; Di Pietro 2012), brain drain (Oosterbeek and Webbink 2011), and acquisition of language skills (Sorrenti 2017). These papers use the same instrumental variable proposed by (Parey and Waldinger 2011), while we propose a new instrument exploring CSF-specific features.

Second, this is the first paper to estimate the causal impacts of a massive government-sponsored study abroad programme by tracking post-university outcomes using different administrative records covering academic, labour market, and entrepreneurship outcomes. As in Oosterbeek and Webbink (2011), we look at administrative data on applicants of a study abroad programme instead of mobile versus non-mobile students. This novel data allows us to mitigate self-selection by comparing similar students who applied to the same programme in the same year, major, and home university. In contrast, the existing empirical evidence typically uses nationally representative surveys of graduates that are non-mandatory, have low response rates, and have self-selected respondents, which can lead to biased estimates.[7]

Our findings contribute to the small literature in economics that investigates the individual-level impacts of carrying out part of university-level studies abroad (Parey and Waldinger 2011; Oosterbeek and Webbink 2011; Di Pietro 2012; Meya and Suntheim 2014; Di Pietro 2015; Sorrenti 2017; Liwiński 2019b). Di Pietro (2019) and Czarnitzki et al. (2021) highlight that, yet the widespread belief in the positive effects of student exchange programmes, the empirical evidence is scarce. The vast majority of studies rely on correlations to report a positive effect of exchange on students' characteristics, career outcomes, language skills acquisition, and labour market mobility (e.g., Sutton and Rubin 2004; Cammelli et al. 2006; European Commission 2014; UK HE International Unit 2015; European Commission 2019).

Although the literature is unequivocal in pointing out that students who conduct part of their studies abroad are more likely to emigrate permanently (Parey and Waldinger 2011; Oosterbeek and Webbink 2011; Di Pietro 2012), especially to the foreign country in which they studied, there is mixed evidence

[6] We have access to all information, including students' entrance exam scores, GPA, graduation year, some socioeconomic characteristics, and students academic history.

[7] For instance, Parey and Waldinger (2011), Rodrigues (2013), and Schnepf and d'Hombres (2018) report response rates of 25%, 30%, and 18% (in the UK survey conducted three years after graduation), respectively. In these surveys, respondents are likely to be highly motivated and interested, as they are non-mandatory (Rodrigues 2013).

about the impacts on finding a formal job in the home country. Additionally, most of the studies rely on association instead of causal inference. Student exchange does not enhance job prospects among graduates from Poland (Liwiński 2019b), Italy (Orrù 2014), Spain (Pinto 2022), and thirteen other European countries (Rodrigues 2013). Other studies, in contrast, indicate that mobility favors employability in the short-run—one year after graduation (Li 2016), between three and four years (Di Pietro 2015; Schnepf and d'Hombres 2018), and six years after graduation (Iriondo 2020). The evidence is less controversial for earnings, as it indicates the existence of a wage premium for studying abroad (Rodrigues 2013; Li 2016; Liwiński 2019a; Iriondo 2020).

2 The Science without Borders programme

2.1 Main features

The Brazilian Ministry of Science and Technology and the Ministry of Education launched the Science without Borders programme in July 2011, designed to last initially for four years (2011-2014). The main goals were to promote the internationalization of Brazilian science, boost innovative research, and increase the competitiveness of local companies (Brazil 2010). CSF was designed as a big-push scholarship programme targeted at different levels of tertiary education, with a focus on undergraduates, which accounted for 79% (73,353) of the scholarships.[8] Undergraduates approved for a CSF scholarship were given the opportunity to attend an academic year in a foreign university in a fully funded fashion. The idea was that by exposing students from the STEM fields to an experience abroad, the programme could induce a large share of participants to work in the same fields in Brazil. Appendix A provides a detailed explanation of the CSF, including a description of the other modalities (e.g., for graduate students) of the programme.

Two research funding agencies, CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*), an agency linked to the Ministry of Education, and CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*), an agency linked to the Ministry of Science and Technology, were the leading programme implementers. They were responsible for selecting the scholarship recipients and values, priority areas, and the foreign higher education institutions that would receive the Brazilian students. CAPES was responsible for implementing approximately 70% of the scholarships (McManus and Nobre 2017). Although students from public and private universities were eligible for the programme, only students from the priority areas (mainly the majors in STEM-related fields) were eligible for a scholarship. The CSF undergraduate modality, which is the focus of this paper, was officially canceled in 2017 after most beneficiaries had returned to Brazil.

In addition to a monthly stipend, the programme fully financed several expenses, such as airfare, housing allowance, health insurance, installation aid, and aid for educational materials. The scholarship values and other benefits varied according to the destination country, students' major, destination city, and the academic year the candidate would attend in the foreign university. Both the destination countries and the foreign universities were chosen based on their academic excellence, measured by intellectual production and training focused on the labour market (Brazil 2010).[9] In some cases, short-term language courses were provided before the start of university-related activities. Because of the shortage of bilin-

[8] When the programme was launched, the Brazilian government stated that the programme goal was to provide approximately 100,000 scholarships for undergraduate and postgraduate students.

[9] The full list of destination countries is composed of the following nations: Austria, Australia, Belgium, Canada, China, Germany, Finland, France, Hungary, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Poland, Portugal, South Korea, Spain, Sweden, United Kingdom, and USA.

gual undergraduates, the Brazilian government launched a CSF-derived programme, Language without Borders, intended to boost the learning of foreign languages among university students.

The CSF costs turned out to be very high, estimated at over US$ 2.72 billion (BRL 15 billion in 2022) or US$ 27,200 thousand (BRL 150,000) per recipient on average (FAPESP 2017). This amount represents at least five times the average expenditure necessary to maintain a student in a public university during one year in Brazil (2016), estimated at approximately US$ 3.800/year (BRL 21.000) (INEP 2016), and is equivalent to fifteen times the budget of CNPq in 2016 (FAPESP 2017).

Some descriptive evidence and documentation suggest the programme was created and implemented abruptly with very limited planning (Knobel 2012; Aveiro 2014; Manços and Coelho 2017; Saldanha et al. 2019; Granja and Carneiro 2021). Just a few days after the official visit of Barack Obama to Brazil in March 2011, Brazilian President Dilma Rousseff announced the CSF as a government educational priority. Despite the longstanding role of federal agencies in managing scholarship programmes, there was no consultation process or public deliberation on the programme priorities or design (Sá 2016). In Brazilian newspaper opinion pieces, it is common to find that the programme was characterized as a presidential initiative.

2.2 Selection process

Programme calls. Undergraduates applying for a CSF scholarship were selected through nationwide public calls. Between 2011 and 2014, the programme launched 102 calls grouped into nine rounds. The calls from the same rounds were open simultaneously and had similar rules regarding the selection process and eligibility criteria for undergraduates. Each call was exclusively for a single destination country, and students could apply for only one call per round. It is important to highlight that students did not know the host university. The calls specified only the destination country. CAPES and CNPq partner agencies abroad were in charge of the allocation of applicants to the host institutions.[10]

The selection process had two phases: (i) a local and decentralized competition, and (ii) a national-level competition only among those who were approved in the first phase. The sequence of events always followed the same steps regardless of the call's destination country and year. First, the federal government launched calls for different countries through the programme website. Then, students could apply for a scholarship by filling out forms in the CSF platform. After that, each Brazilian university could launch its own selection process to validate the individual candidates, respecting the programme rules. Only students who had their application approved by their home university moved to the national-level phase, which were managed by CAPES and CNPq.[11]

On the national stage, the score on the National High School Exam (*Exame Nacional do Ensino Médio*, ENEM) was used as the unique criterion to order candidates so that students with the highest scores on the exam had priority in the scholarship offers. In the case of a tie in the ENEM score, preference was given to candidates: i) with prizes in scientific Olympics, in Brazil or abroad, or ii) who had received or were receiving scientific or technological scholarships from CAPES, CNPq, or any other state foundation for research support.[12]

---

[10] These foreign agencies were typically long-standing partners of CAPES and CNPq, such as DAAD in Germany, IIE in the USA, and UNIBO in Italy. Unfortunately, CAPES and CNPq have never published information about the criteria that guided the foreign partner agencies in their allocation decisions.

[11] To have their students considered for the national stage of the CSF selection process, the Brazilian higher education institutions had to first sign accreditation agreements with CAPES and CNPq.

[12] ENEM is a nationwide standardized exam whose score is used to apply for a seat in public universities in Brazil. Most of the Brazilian federal universities currently use the ENEM score in their admissions process.

More than twenty countries have received CSF fellows, and each destination country had, on average, three different calls between 2011 and 2014. Students who applied for more than one call from the same round were automatically dismissed for the first call. Approved students were entitled to receive the scholarship only once. Therefore, a student could be approved more than once to the programme in different years (if they declined the CSF scholarship at least once), but could not be granted the scholarship more than once. Appendix Figure A3 presents the timeline of the CSF calls for undergraduates.

Eligibility. To be eligible for a scholarship, students must meet the following criteria: i) be Brazilian; ii) be enrolled in an university major compatible with the programme priority areas in a CSF-accredited Brazilian institution; iii) have a score of at least 600 points in the ENEM (as of the 2009 edition of the exam); (iv) present good academic performance (according to the criteria defined by the home university in the first stage of the CSF selection process); v) have completed at least 20% and at most 90% of the credits from their major curriculum at the time of application; and vi) demonstrate language proficiency through a minimum score in international standardized exams such as TOEFL (*Test of English as a Foreign Language*) and IELTS (*International English Language Testing System*).[13]

Returning to Brazil after the scholarship. As a mandatory rule of the programme, CSF fellows had to sign a commitment letter containing all their obligations, including: i) return to Brazil up to thirty days after the end of the scholarship and stay in the country for at least the same duration of the scholarship, and ii) reimburse the total amount corresponding to the expenses incurred on their behalf in case of non-compliance with other CSF rules (Senado Federal 2015).[14] Therefore, completing the degree in Brazil was not a requirement, but students were required to stay in Brazil for the duration of the scholarship. It is worth mentioning that there was no quality assessment of students' behaviour. Also, students had no obligation to enroll and be approved in a minimum number of courses abroad.

3 Data, sample selection, and record linkage

In this paper, we use seventeen administrative data sets from different sources. Most of the data sets are not public and were obtained through formal requests to the Brazilian authorities using the Access to Information Law (LAI). The records were linked using the candidates' names and their taxpayer identification number in Brazil called *Cadastro da Pessoa Física, CPF*, which is an 11-digit sequence issued by the Brazilian Internal Revenue Service (*Receita Federal do Brasil, RFB*). Due to confidentiality restrictions under the Brazilian law, we asked universities for the six intermediary digits of CPF (e.g., ***.123.456-**) of each student in order to maximize the likelihood of being given access to their data since the 11-digit CPF is a personally identifiable information.

3.1 Data sources

CSF candidates registry. This data set contains the list of all candidates who applied for at least one of the 102 CSF calls launched between 2011 and 2014. CAPES and CNPq gave us access to the candidates' full names, their home university at the time of application, the calls they applied for, and whether they were approved in each call. More than 142,000 undergraduates applied to the programme in at least one call between 2011 and 2014. CAPES and CNPq also provided us with the total number of applicants and approved candidates in each call. In addition, they provided the ENEM score of the last approved

[13] The ENEM score is calculated using the item response theory (IRT), and its range is such that the maximum score is 1,000 points and the minimum score varies from year to year, but the typical average is around 280 points.

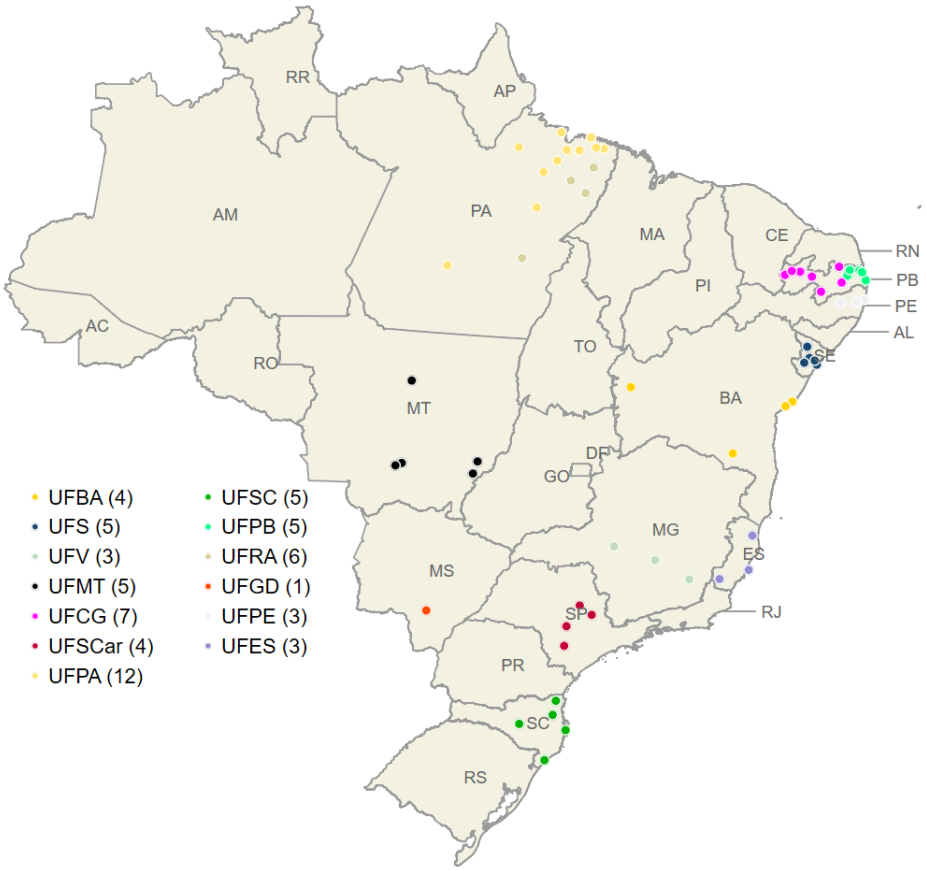
[14] Returning to Brazil without the consent of CAPES during the period of fellowship could cause the cancellation of the scholarship (CAPES 2015).

student, but only for a subset of calls. The data contains all Brazilian CSF-accredited universities with at least one candidate for the programme. These data sets do not contain any information about candidates' socioeconomic characteristics, or academic records, such as university majors, entrance exam scores, etc.

Universities records. We formally requested data for all 63 Brazilian federal universities, but we managed to obtain the data for 13, representing almost 20% of them. For each university, we request the following variables: i) student's full name; ii) gender; iii) age; iv) race; v) six intermediary digits of CPF; vi) student's university major; vii) admission semester and year, and viii) entrance exam score used for admission in the corresponding major.

Figure 1 presents the complete list of universities and their distribution in the Brazilian territory. The pool of universities is geographically sizeable and diverse, and nine of them are ranked among the 30 institutions with the highest number of CSF applicants. Additionally, the Federal University of Bahia (UFBA) granted access to their students' entire academic records, which include grades, the time between admission and graduation, and whether they graduated or not.

Figure 1: Spatial distribution of the campuses of the thirteen universities in our sample



Note: each dot represents a city in which at least one of the 13 universities has a campus. The number close to the university name corresponds to the number of municipalities in which the university has at least one campus.

Source: authors' elaboration. Map based on data from Brazilian Bureau of Statistics (IBGE).

Formal labour market (RAIS). The formal labour market information comes from the Brazilian matched employer-employee data set called RAIS (*Relação Anual de Informações Sociais*), which is the result of a mandatory annual survey administered by the Ministry of Economy. Every registered firm in Brazil is obliged to fill out RAIS yearly. The data have detailed information about employees' socioeconomic characteristics and labour earnings. Firms must submit the full name and the 11-digit CPF for every listed worker. We use RAIS data to retrieve information on whether each CSF applicant

had a formal job, their wage, job tenure, and whether the occupation was considered technical according to the classification proposed by Araújo et al. (2009). Data for labour market outcomes are available until 2020. RAIS data are not public and were accessed through an institutional agreement.

Postgraduate education (SUCUPIRA). The variables related to the postgraduate education in Brazil stem from information collected by CAPES. Every year, CAPES collects information about students, postgraduate programmes, and universities in Brazil. The information is organized in an online platform called SUCUPIRA. The postgraduate programmes must submit detailed information about their students, faculty staff, research projects, and academic production. SUCUPIRA public data sets contains the full name and the six intermediary digits of CPF of each graduate student. Data for postgraduate education outcomes are available until 2020.

Formal entrepreneurship (RFB). The data about formal entrepreneurial activity in Brazil are public and collected by the Brazilian Internal Revenue Service (*Receita Federal do Brasil*, RFB).[15] The data sets contain information about all firms registered in the country. Every registered firm has a tax identification number and a corresponding legal representative or a set of business partners.[16] For each firm owner or partner listed in RFB data sets, there is information about her full name and six intermediary digits of CPF. We use this data set to create information about: i) the year in which the firm was created; ii) the exact date when individuals became a formal partner or owner, which allows us to determine whether they are founders of the company; and iii) firm economic sector. Data for formal entrepreneurship outcomes are available until July 2021.

3.2 Sample selection

The sample is composed of CSF applicants regularly enrolled in Brazilian universities. We pooled data from the different universities and aggregated similar majors. Candidates from the same university, major, and admission semester/year were considered a common cohort, regardless of their college campus.

We focus on candidates that applied for a single call. This allows us to have a one-to-one map between the call each candidate applied for and her outcomes after that specific call. The main challenge of considering multiple-call candidates is that there are different periods in which their approval status can change, and it might not be constant after the first call. It means that there is no obvious one-to-one map between candidates and their career outcomes.

Table B1 shows that 84% of the candidates in our sample applied for only one call, 14% applied for two calls, and less than 2% applied for three or more calls. In particular, 73% of the candidates applied for a single call and were associated with only one major in their home university. Appendix E presents the details of the data preparation, while Appendix Table E11 presents the definition of each variable.

---

[15] A particularly important feature of the public version of RFB data is that they only contain information about individuals that are formally considered business partners of societies or partnerships. It means that for firms that are not registered as partnerships (or societies), which is the case of firms formed by a single owner (sole proprietorship), there is no information about the legal person (individual) attached to it. More than 60% of the business entities in Brazil are registered under a sole proprietorship regime, and thus this is an important limitation of the public data sets. To overcome this limitation, we requested RFB a special data extraction to obtain three variables related to the individuals linked to firms registered under the legal nature *Empresário Individual*, CONCLA code 213-5, under which more than 90% of the sole-proprietorship firms in Brazil are registered.

[16] For firms, the tax identification number is a 14-digit sequence called *Cadastro Nacional da Pessoa Jurídica*, CNPJ), which is also issued by RFB.

3.3 Record linkage

As explained before, we use seventeen individual-level data sets administered by different institutions. Most contain the individuals' full name and their six intermediary digits of CPF (in the format ***.123.456-**). This allows us to conduct a two-variable probabilistic record linkage to merge them. The details about the linkage and each data set are described in Appendices F and G.

4 Descriptive statistics and empirical strategy

4.1 Descriptive statistics

Panel A of Table 1 presents the number of candidates, the percentage of approved applicants, and the percentage of applicants found in the administrative data sets used to create the outcome variables by candidates' first call year. The sample contains 19,245 candidates (column 2), representing approximately 14% of the total number of CSF candidates who applied for calls targeting undergraduates. Of these candidates, 50.03% (column 3) were approved in the first call they applied for. As to the first application year, the vast majority applied for the first time for calls launched in 2012 (33.1%) or 2013 (43.8%). Columns 5, 7, and 9 in Panel A show the number of candidates found in the RAIS (formal labour market), SUCUPIRA (graduate studies), and RFB (entrepreneurship) data sets. The correspondent percentage of the total candidates by applicant's first call year are displayed in columns 6, 8, and 10. On average, 33.1% of the students were found in the formal labour market, 27.7% in a postgraduate programme in Brazil, and 23.4% as formal firm owners or partners.

Columns 11 and 12 show that 35% of candidates were not found in any outcome data set. There are two explanations for this result. The first is that we can track students until 2020 (postgraduate education), 2020 (formal labour market), and mid-2021 (formal entrepreneurship) depending on the outcome data set. The second is that most students applied to the programme in 2013 and 2014. Therefore, they may not have completed their bachelor's degree until 2020/2021. This argument is reinforced by the fact that while only 29.7% of the candidates that applied in 2011 were not found in any data set, this number grew to 40.8% among 2014 applicants. Unfortunately, only the Federal University of Bahia data has graduation information. Therefore, we will use this data set to shed light on this hypothesis.

Panel B of Table 1 shows the number of candidates, and the percentage of approved applicants by home university and region in which the university is located. Not surprisingly, the largest universities in our sample had the highest number of students applying to the programme. However, there is a geographical dispersion in the top positions of the ranking, with two universities from the northeast (UFPE, 16.5%, and UFBA, 14.6%), the poorest region of Brazil, and two from the southeast (UFSCar, 11.8%, and UFV, 10.1%), and UFSC from the south region with 18.3% of the candidates. It reinforces the argument that our sample is representative nationwide.

Table 2 complements the analysis by showing the cross-tabulation of candidates' career paths. Column 1 presents the number of candidates found in each data set. Most were found only in one outcome data set. For example, of the 6,366 candidates with a formal job in the RAIS data set, 51.7% were found only in this data set, while among the 4,955 candidates enrolled in a postgraduate programme in the SUCUPIRA data set, 54% were found only in the SUCUPIRA data set.

Panel A of Table 3 presents some summary statistics of the main variables for both approved and not approved candidates in their first CSF candidature. Unfortunately, we only have a few pre-call covariates at the individual level due to the restrictions of data protection laws. Our measures of competitiveness are the variables *ratio* and *ratio top 25th pctle*. Approved candidates are, on average, more likely to be male, perform better in the entrance exam score in their corresponding university major, and enroll in

an engineering major. Importantly, they are more likely to have applied for a less competitive call than their unapproved counterparts. Approved candidates applied for calls with an average approval rate of 36%, while not approved candidates applied for calls with an average approval rate of 27%.

Panel B of Table 3 presents the mean and standard deviation of the outcome variables in the post-call period. The means indicate that approved candidates are less likely to have a formal job and more likely to have smaller job tenure. They were also less frequently found in RFB data sets, and the difference is even larger for enrolment in a Brazilian postgraduate programme (- 5 p.p., on average).

Appendix Figure A4 shows the distribution of the difference between candidates' call year and their admission year at home university. Approximately 70% of the candidates applied to the programme in the first three years after entering college. The overall pattern is, therefore, consistent with the eligibility criteria. Appendix Figure A5, in turn, displays the distribution of candidates across majors within the different home universities. engineering majors are prevalent in most universities, followed by health sciences. Appendix Table B2 presents the top 30 universities according to the number of CSF candidates. The vast majority are federal universities, and nine are in our data set. Finally, Appendix Figure A9 shows that most of the candidates applied to the same countries, such as the U.S., Portugal, U.K., Spain, and Canada.

Table 1: Number of candidates and proportion found in the formal labour market, postgraduate education and formal entrepreneurship market by first call's year and home university

				Labour market		Postgraduate education		Entrepreneurship		Neither/nor	
Call's year/University	No. students	Approved (%)	Prop.	No.	%	No.	%	No.	%	No.	%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Number of candidates by first call's year											
Total	19,245	50.0	-	6,366	33.1	5,332	27.7	4,509	23.4	6,730	35.0
2011	639	60.4	3.3	226	35.4	228	35.7	145	22.7	190	29.7
2012	6,371	47.6	33.1	2,383	37.4	2,193	34.4	1,478	23.2	1,908	29.9
2013	8,429	55.1	43.8	2,660	31.6	2,162	25.6	1,958	23.2	3,079	36.5
2014	3,806	41.1	19.8	1,097	28.8	749	19.7	928	24.4	1,553	40.8
Panel B. Number of candidates by home university											
UFBA	2,825	48.9	Northeast	848	30.0	641	22.7	765	27.1	1,062	37.6
UFCG	1,250	48.5	Northeast	366	29.3	439	35.1	237	19.0	455	36.4
UFES	1,431	46.4	Southeast	458	32.0	439	30.7	355	24.8	466	32.6
UFGD	273	41.4	Central West	103	37.7	100	36.6	58	21.2	77	28.2
UFMT	742	35.2	Central West	253	34.1	201	27.1	212	28.6	234	31.5
UFPA	506	46.4	North	143	28.3	197	38.9	72	14.2	176	34.8
UFPB	501	56.9	Northeast	123	24.6	140	27.9	105	21.0	214	42.7
UFPE	3,181	49.2	Northeast	1,095	34.4	837	26.3	729	22.9	1,148	36.1
UFRA	285	33.7	North	80	28.1	148	51.9	46	16.1	70	24.6
UFS	477	47.0	Northeast	128	26.8	141	29.6	92	19.3	193	40.5
UFSC	3,537	49.4	South	1,220	34.5	910	25.7	1,005	28.4	1,176	33.2
UFSCar	2,282	58.5	Southeast	894	39.2	578	25.3	466	20.4	761	33.3
UFV	1,955	57.1	Southeast	655	33.5	561	28.7	367	18.8	698	35.7

Note: data related to the labour market and postgraduate education are available until 2020, while data related to entrepreneurship are available until July 2021. The candidates were considered as found if there exist at least one registry in the corresponding dataset-specific time period. Percentages do not add up to 100 across columns 'Labour market', 'Postgraduate education', 'Entrepreneurship', and 'Neither/nor' because there are candidates found in more than one data set.

Table 2: Number and percentage of applicants that were found in at least one outcome data set

			RAIS	RFB	SUCUPIRA
		Total	Formal employed	Firm ownership	Postgrad. student
RAIS	Formal employed	6,366	3,292	1,554	1,520
			[51.7%]	[24.4%]	[23.9%]
RFB	Firm ownership	4,509	1,554	2,024	931
			[34.5%]	[44.9%]	[20.6%]
SUCUPIRA	Postgrad. student	5,332	1,520	931	2,881
			[28.5%]	[17.5%]	[54.0%]

Note: each row corresponds to a different data set and the column 'Total' indicates the number of candidates found in the corresponding data set. Relative percentages are shown in square brackets. The table should be read as follows. To illustrate, let us consider the example of the first row, which refers to the formal labour market. It says that we found 6,366 candidates in the RAIS data set. Out of them, 3,292 (51.7%) were found solely in this data set. In contrast, of the 6,366 candidates, 1,554 (24.4%) were also found to be a firm owner or partner, and 1,520 (23.9%) were also found to be a student enrolled in a Brazilian postgraduate programme.

Table 3: Summary statistics by candidates' approval status

	Approved		Not Approved			
	mean	sd	mean	sd	p-value	Prop. missings
Panel A. General variables						
Male	0.58	(0.49)	0.50	(0.50)	0.000	0.0
Entrance exam score	0.56	(0.32)	0.45	(0.35)	0.000	11.6
Ratio	0.36	(0.14)	0.27	(0.14)	0.000	0.0
Ratio top 25th pctle.	0.34	(0.47)	0.19	(0.40)	0.000	0.0
Engineering	0.53	(0.50)	0.38	(0.49)	0.000	0.0
Health Sciences	0.10	(0.30)	0.17	(0.37)	0.000	0.0
Other majors	0.37	(0.48)	0.45	(0.50)	0.000	0.0
Panel B. Outcome variables						
Formal employed	0.31	(0.46)	0.33	(0.47)	0.070	0.0
Firm ownership	0.20	(0.40)	0.22	(0.42)	0.000	0.0
Postgrad. student	0.25	(0.44)	0.30	(0.46)	0.000	0.0
Found in any outcome data set	0.61	(0.49)	0.65	(0.48)	0.000	0.0
Observations	9,629		9,616			

Note: columns 'mean' and 'sd' display the mean and the standard deviation, respectively, while the column 'p-value' shows the p-value of a mean equality t-test, and the column 'Prop. missings' shows the proportion of missings (as a percentage) for the corresponding variable. In panel B, we consider the whole post-call period for all variables and only events (e.g., job spells, postgraduate programme enrolments or connections with firms as owner or partner) that started after the candidate's first call year. The variables are described in detail in Appendix Table E11.

4.2 Empirical strategy

Ideally, to identify the effect of the CSF programme, the only difference between the approved and not approved candidates would be their approval status in a specific CSF call. Nevertheless, both the students' selection into the programme and the selection of approved candidates were not random. CAPES and CNPq selected students using the shortlist of candidates sent by all universities participating in the programme, and each student's ENEM score. One may think that we could use this information to compare students around the threshold of each call, but CNPq and CAPES only made available the score threshold for a subset of calls. Besides, we do not have information for all students, but only for the thirteen universities described before.

We thus leverage the characteristics of the calls and students' observable characteristics to estimate the causal effects of the programme using an instrumental variable approach. Our primary goal is to compare applicants from the same home university, major, and admission year with similar university entrance exam scores and applied for CSF calls for the same destination country and call's launching year. Therefore, our identification strategy allows us to compare the approved students in a less competitive call who would not have been approved had they applied for a more competitive call for the same country, and year. Our instrumental variable (IV) excludes the case of students that would be always approved, and those who would never be approved. This is possible because of the random competitiveness of the calls. In particular, to investigate the effects of the CSF programme, we estimate the following equation:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 Approved_{i,c,y,d} + \beta_2 Entrance\ exam\ score_{i,m,s,u} + \beta_3 Male_i \\ &\quad + \beta_4 Dup\ major_{i,m} + \alpha_s + \pi_u + \theta_m + \mu_y + \psi_d + \varepsilon_i \end{aligned} \tag{1}$$

where Y is one of the outcomes related to the labour market, postgraduate education, or entrepreneurship for each student i in a call c; s stands for the student's admission year at college; m a given university major; u a given home university; y the year in which call c was launched; and d the destination country associated with call c. $Approved_{i,c,y,d}$ is a dummy variable for whether the student was approved in call c; $Entrance\ exam\ score_{i,m,s,u}$ is the normalized entrance examination score used by student i for admission in university major m, year s and home university u; $Male_i$ is a dummy variable for whether the student is a male; $Dup\ major_i$ is a dummy variable for whether the student has ever enrolled in more than one major in his home university u; α_s is a cohort (admission year) fixed effect; π_u is a home university fixed effect; θ_m is a university major fixed effect; μ_y is a call's year fixed effect; ψ_d is a destination country fixed effect; and ε_i is an error term.

Our main interest is identifying β_1, the programme's intention-to-treat (ITT) effect. According to the programme documents, only 3% of the approved candidates decided not to enroll in the programme. Therefore, given the high take-up rate, we should expect very minor differences between the ITT and the average treatment effect on the treated (ATT). Even though equation 1 controls for some observable and unobservable characteristics, it does not control for some unobserved factors that are likely to affect both the approval status and the outcome variables. Some of these factors include students' ability not correlated with the entrance exam score, foreign language proficiency, motivation, career goals, and interest in the university major. These factors imply that the OLS estimates are likely to yield upward biased estimates of β_1. To circumvent the potential bias, we leverage on the CSF selection process, and the competitiveness of each call to create an instrument for $Approved_{i,c,y,d}$.

First, we create a measure of the competitiveness of each call, which is calculated by dividing the total number of approved candidates by the total number of applicants at the Brazilian level. This measure considers all Brazilian universities that participated in the programme. Candidates from the 13 universities in our sample are counted in both the numerator and denominator of the national-level approval rate of a particular call. Second, we create the discounted version of the above measure, which

excludes the candidates from these 13 universities. The discounted approval rate is not affected by the approval status of the candidates in our sample. The main rationality for creating the discounted version is that it is affected only by the approval status of applicants from universities other than the ones in our sample. Therefore, it avoids reverse causality and a possible mechanical effect in the first stage.

To identify the causal effect of being approved for the programme on the outcomes of interest, we use the discounted-version approval rate (approval rate, henceforth) of a particular call as an instrument for the variable $Approved_{i,c,y,d}$ in equation 1. Because the instrument accounts for each call's competitiveness, the intuition for the instrument is that if many applicants from other universities were approved in a given call, it is more likely that a given candidate from one of the 13 universities in our sample is also approved, regardless of the programme eligibility criteria. Therefore, the more competitive a call, which means fewer applicants from other universities being approved, the less likely it is for a given applicant from one of the 13 universities in our sample to be granted a CSF scholarship. The first stage regression is estimated as:

$$\begin{aligned} Approved_i &= \gamma_0 + \gamma_1 Ratio_{c,y,d} + \gamma_2 Entrance\ exam\ score_{i,m,s,u} + \gamma_3 Male_i \\ &\quad + \gamma_4 Dup\ major_{i,m} + \lambda_s + \nu_u + \rho_m + \Pi_y + \tau_d + {}_i \end{aligned} \tag{2}$$

where $Ratio_{c,y,d}$ is the instrument, λ_s is a cohort fixed effect, ν_u is a home university fixed effect, ρ_m is an university major fixed effect, Π_y is a call's year fixed effect, τ_d is a destination country fixed effect, and ${}_{i,c,y,d}$ is an error term.

It is important to highlight that students could not predict the competitiveness of each call. First, students did not know the number of available slots in each call. Second, students didn't have information about the number of applicants, or the quality of the competitors—such as their ENEM score—and they could not predict if there would be another call to the same destination country to be launched in the same year.

As shown in Section 4.1, there are differences in the approval status across universities between majors and admission cohorts for each university. The inclusion of fixed effects accounts for these differences. The inclusion of gender increases the precision of point estimates, while the entrance examination score accounts for students' observed cognitive ability when they entered college. Finally, we include an indicator variable for whether the candidate was associated with more than one major in his home university to account for students who had ever enrolled in a second major at the time they applied for a CSF call. This is important because these students could use course credits from their previous major for the new one, changing their probability of being approved into the programme.

The call's year and destination country fixed effects are most relevant in our setting. Since the variation takes place at the call level, we cannot compare candidates who applied for the same call. However, we can compare candidates that applied for different calls, but for the same destination country. Moreover, we can also include dummies for the call's launching year, which enables us to compare candidates who applied for calls launched in the same year. Including these fixed effects restricts the regression sample to candidates who applied for destination countries with more than one CSF call per year. Standard errors are clustered at the call level because the treatment variation takes place at this level.

We will discuss the validity of the IV in more detail in subsection 4.3. In Appendix Table B10 we provide evidence that our results are robust to other fixed effects combinations. More specifically, we use a more flexible specification with fixed effects for home university, admission year, and university major-call-destination country-call-year combinations. The results are qualitatively similar.

4.3 Instrumental variable validity

To be valid, the instrumental variable must satisfy the relevance condition and the exclusion restriction. The first condition requires that the instrument affects the endogenous variable, while the second condition requires it to affect the outcomes of interest only through its intermediary effect on the endogenous variable.

Relevance condition. The relevance condition requires that the IV should affect the endogenous variable. Table 4 presents the first stage estimates for different samples: (i) considering only the first call of each candidate, (ii) considering only the last call of each candidate, and (iii) restricting to single-call candidates. For (i) and (ii), we also include fixed effects for the number of different calls that each candidate applied for in order to control for a possible participation bias. The F-statistics associated with the instrument in each sample are 32.4, 45.4, and 43.2, respectively. Our point estimates are statistically significant and with values around 1.2 percentage points. It means that for each one percentage point increase in the call approval rate, an applicant's probability of approval increases, on average, by 1.2 percentage points.

Exclusion restriction. The exclusion restriction requires that the programme's effect comes from the competitiveness of the call, and not from applicants' characteristics. Intuitively, the only reason for which a given call's approval rate is correlated with applicants' future career outcomes is because the competitiveness of the call they applied for may have affected their chance of approval into the programme. Therefore, the exclusion restriction would be violated if candidates could predict which calls would be less competitive for the same destination country and year and decide to apply only for the least competitive calls in order to increase their chances of approval. If this kind of prediction happens, the degree of competition in a given call would be correlated with students' overall ability, or there would be other characteristics potentially correlated with the outcome variables. We argue that students predicting the competitiveness of the call do not happen because of the the programme setting.

First, because at the time of a specific call, except for the programme implementer, no one could know whether more calls would be launched in the future. It would be even more difficult to predict whether news calls for the same destination country would be launched, and in the same year, and whether they would be more competitive than the current calls. Importantly for our strategy, there was no public access to the programme schedule indicating which calls would be launched and the number of scholarships to be offered in each call.

Second, to provide some evidence on the programme impacts on pre-treatment covariates that should not be different between approved and not approved candidates induced by our instrument, we show in Appendix Table B7 the estimates of our model on some personal characteristics from UFBA candidates (i.e., the only ones for which we have information). These characteristics include (i) students' age at the time of taking the UFBA entrance exam; (ii) whether she was resident of the Metropolitan Region of Salvador, the state capital of Bahia; (iii) whether either the mother or father had a college degree; (iv) whether the student declared to be single; (v) whether she declared to be financially dependent; and (vi) whether she declared to have attended a vocational track during high school. Reassuringly, the point estimates are small and not statistically significant, which suggests that these characteristics are balanced between the approved and not approved candidates induced by our instrument.

Range of the instrument. Appendix Table B8 presents the list of calls for each destination country that had at least two different calls in the same year. The table shows a sizable difference in the approval rate across calls for the same destination country and year. For instance, for the U.S., the first call in 2011 had an approval rate equal to 11.6%, 36.6% in 2012 , 41.9% in 2013 , and 29.8% in 2014. Appendix Figure A7 complements the previous evidence by showing that there is a substantial approval rate difference across within-group calls.

Even though we have no information about the expected number of scholarships that would be offered in each call, Appendix Table B8 shows that for 22 out of the 46 calls for which we have information on the ENEM score of the last approved candidate, the requirement was not binding (i.e., the score was lower than 600 points).[17]. Interestingly, it was binding for the calls launched in 2014, the last year in which the calls were launched (see Appendix Table B8). This is consistent with an institutional learning on the part of the programme-executing agencies, and also reflects the fact that the programme was implemented abruptly and with limited planning. We hence argue that the main driver of this variation in the approval rate was related to budgetary issues that are orthogonal to any latent determinants of candidates' approval status, conditional on our fixed effects and covariates.

Compliers, always-takers, and defiers. Our identification strategy allows us to identify the programme's ITT impact only on those approved candidates induced by our instrument and not on all approved candidates in our sample. Consequently, the Local Average Treatment Effect (LATE) is the average treatment effect of having been offered a CSF scholarship for a subpopulation of the applicants, defined as the compliers. The complier in our setting is an approved candidate in a less competitive call who would not have been approved had she applied for a more competitive call within the same group (i.e., same destination country and year). In particular, we expect the complier to be a lower-middle student in the ENEM's score distribution. It means that the IV might work to compare students in the neighbourhood of a lower-middle performance, and if this is the case, the LATE that we can arguably identify is an informative estimand from a policy standpoint.

Because the *Ratio* variable is continuous, it is more difficult to grasp the magnitudes of the first stage. It also makes it more difficult to identify who would be our compliers. Therefore, we create an alternative specification where the IV is an indicator variable for whether the call is amongst the 25% of calls with the highest approval rates. The results (columns 2, 4, and 6 in Table 4) indicate that the candidates who applied for a less competitive call were, on average, between 14.3 and 15.4 p.p. more likely to be approved in the programme, conditional on our fixed effects and covariates. Additionally, Appendix Figure A8 depicts the point estimates and their 95% confidence interval when we replace the indicator variable for the 25% least competitive calls with indicators for other thresholds varying from the 50th percentile to the 95th percentile. The results are qualitatively similar and even more significant in magnitude between the 50th and 70th percentiles.

The always-takers are the candidates with the highest entrance exam scores. They would always be selected for the programme and are not part of our analysis. The defier is a candidate not approved who applied for a less competitive call and that would have been approved had she applied for a more competitive call. We argue that the existence of defiers is not likely in our setting.

Just-identified case. In addition, we applied the inference procedure for the just-identified case relative to a single instrumental variable recently proposed by Lee et al. (2022). Based on the first-stage F statistic, the method adjusts the Two-Stage Least Squares (2SLS) t-ratio inference. The authors propose an adjustment factor to the 2SLS standard errors to calculate an adjusted t-ratio. Table 4 presents in parentheses the adjusted t-ratio for the 95% confidence interval. The results do not corroborate the hypothesis of a weak IV based on the Lee et al. (2022) procedure because the adjusted t-statistic of the coefficient of interest is above the threshold for significance at the 1 percent level ($t > 2.51$).

Robustness to different sets of fixed effects. Appendix Table B9 shows that our first-stage results are robust to the inclusion of different sets of fixed effects both in our preferred sample with single-call candidates and the sample that considers the first call of each candidate. The only specification for which the adjusted t-stat of Lee et al. (2022) in both samples is statistically significant solely at the 10 percent level (threshold = 1.645) is the fully saturated one, which includes dummies for each

[17] Recall from Section 2 that candidates were ordered according to their ENEM score.

cohort-home university-major-destination country-call year combination. Although this specification can be considered theoretically superior because it compares very similar individuals, there seems to exist an important trade-off between the possibly most appropriate specification and statistical power, as reflected by the relatively reduced number of observations. In particular, our data is such that we have, at most, only four different candidates within each cell composed of the aforementioned fixed effect with an average of only two candidates, which reinforces the idea of low statistical power in this specification (the increased standard errors also reflect that).

Table 4: First stage estimates

Dependent variable: Approved	First call		Last call		Single-call candidates	
	(1)	(2)	(3)	(4)	(5)	(6)
Ratio	1.183***		1.165***		1.206***	
	[0.208]		[0.173]		[0.183]	
	(4.829)		(4.835)		(4.513)	
Ratio top 25th pctle.		0.143***		0.154***		0.154***
		[0.041]		[0.039]		[0.040]
		(2.243)		(2.714)		(2.618)
Entrance exam score	0.186***	0.187***	0.203***	0.205***	0.215***	0.218***
	[0.018]	[0.018]	[0.017]	[0.017]	[0.018]	[0.017]
Obs	17,007	17,007	16,999	16,999	14,271	14,271
R2	0.06	0.01	0.05	0.01	0.06	0.01
No. clusters	98	98	97	97	97	97
F-stat of Instrument	32.49	12.45	45.43	15.40	43.21	14.77

Note: this table presents the results of the first stage of the 2sls estimation. More precisely, the estimation of equation 2. All regressions include fixed effects for home university, major, admission year, call's year and destination country, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at home university. In columns 1 to 4, we also control for the number of CSF calls the candidate applied for. Clustered robust standard errors at the call level are shown in square brackets. In parentheses, we present the 0.05 tF statistic from Lee et al. (2022). The number of observations displayed in each column is calculated excluding singletons. The explanatory variables are described in detail in Appendix Table E11. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

5 Results

This section shows the results of the IV estimation of equation 1. First, we present the results of the CSF programme on postgraduate education enrolment, second on the labour market, and third on entrepreneurship. The second subsection presents the analysis of the mechanisms using the detailed UFBA data set, and finally, the last subsection analyzes potential attrition.

As shown in the red curve in Figure A10, the difference between the call date and the end of the scholarship varies from 12 to 36 months. Therefore, in the first three years after applying for a call, we should expect a negative probability of finding students in the data sets. Because of that, we estimate the effect of CSF year-by-year after the call and also on different pooled year samples. First, we consider a pooled sample of one to three years after the call, which should reflect a lock-in effect because of the time between the call results and the end of the scholarship. Second, we consider a pooled sample of four to six years after the call, representing the period after the end of the scholarship.

Finally, because we can only observe a sub-sample of applicants seven and eight years after the application, the number of clusters in a year-by-year estimation for seven and eight years after the application is equal to 71 and 30, respectively. The sample size also reduces to 10,881 and 4,807. Therefore, we provide only an estimation of a pooled sample of seven to eight years after the call to try to identify some evidence of a medium-run effect of the programme.

5.1 Main results

Table 5 displays the results for the probability of being enrolled in a postgraduate programme in Brazil. Columns 1 to 6 show the estimates for each year after applying to the programme, from the first to sixth year. Columns 7 and 8 present estimates for the two pooled periods for which we can observe the whole sample. Column 9 shows the results for a reduced pool of candidates who we can observe for up to eight years after the application. In general, the results suggest that approved applicants have a lower probability of enrolling in a Brazilian postgraduate programme than their not approved counterparts. Approved candidates are, on average, 12.5 p.p. less likely to enroll in a postgraduate programme in the first three years after the call, and 7.1 p.p. less likely to enroll in a postgraduate programme between four to six years after the call. We do not find significant effects for the reduced sample of applicants who can be observed for a longer period. As expected, the results are stronger in the first three years after the call, when some students were not in Brazil. The year-by-year estimations presented in columns 1 to 6 show a negative result in the whole period, but point estimates are not statistically significant for the fifth and sixth years.

Table 6 presents the programme effects on formal employability. In panel A, we present the impact on the probability of having a formal job only for contracts that started after the programme call, while in panel B, we present the impact on the probability of having a formal contract independent of when it started. The estimates suggest that the programme was unsuccessful in increasing the presence of approved students in the formal labour market after the lock-in period of up to three years after the call. Column 7 suggests a negative effect but not significant, and column 9 suggests a significant negative effect in the long term. The results are very similar in both panels, indicating that most jobs started after the applicant's call year.

Table 7 shows the effects on formal firm ownership from one to six years after the call and for the pooled samples. Panel A presents the results on the probability of being a formal firm owner or partner only after the candidate applied for a given call. Panel B shows the results relative to firm ownership regardless of when the firm started. Therefore, panel A considers only firms that the candidates started, while panel B also includes firms that were not created by the candidates themselves. The first situation typically refers to sole-proprietorship companies, while the latter commonly refers to when candidates become business partners of a registered organization.

One might be concerned that candidates may come to these firms through succession (family firms) rather than through effective creation. This risk is mitigated in panel A since succession is unlikely to occur precisely after students apply for a CSF call. Panels A and B show similar point estimates, indicating that the firms to which candidates are connected were self-started. Our results thus suggest that CSF did not contribute to promote formal entrepreneurial activities among approved candidates.

Table 5: Effects on postgraduate education enrolment

	+1 year	+2 years	+3 years	+4 years	+5 years	+6 years	Pooled +1 to +3 years	Pooled +4 to +6 years	Pooled +7 to +8 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Approved	-0.028**	-0.085***	-0.118***	-0.074*	-0.032	-0.005	-0.125***	-0.071**	-0.002
	[0.013]	[0.024]	[0.036]	[0.038]	[0.031]	[0.024]	[0.036]	[0.028]	[0.024]
Mean control dep. var	0.02	0.07	0.14	0.20	0.21	0.20	0.14	0.27	0.21
Obs	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271	10,881
No. clusters	97	97	97	97	97	97	97	97	71

Note: this table presents the 2SLS estimation results of equation 1. The dependent variable in columns 1 to 6 is a binary variable for whether the student was enrolled in a postgraduate education programme in the corresponding N-year window where N $\in \{1,2,3,4,5,6\}$. The dependent variable in column 7 is a binary variable for whether the student was enrolled in a postgraduate education programme in any year between the first and third after the call's one, while that in column 8 is the same binary variable but that considers the period between the fourth and sixth year after the application. The dependent variable in column 9 follows the same logic, but considers the period between the seventh and eighth year after the application. All regressions include fixed effects for home university, university major, admission year, call's year and destination country, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at home university. Clustered robust standard errors at the call level are shown in square brackets. Data for postgraduate education are available until 2020. 'Mean control dep. var' shows the mean of the dependent variable for not approved candidates. The number of observations displayed in each column is calculated excluding singletons. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Effects on having a formal job

	+1 year	+2 years	+3 years	+4 years	+5 years	+6 years	Pooled +1 to +3 years	Pooled +4 to +6 years	Pooled +7 to +8 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Only contracts started after the call's year									
Approved	-0.012	-0.044***	-0.022	-0.034**	0.011	-0.084***	-0.038**	-0.042	-0.049*
	[0.008]	[0.014]	[0.017]	[0.015]	[0.022]	[0.025]	[0.019]	[0.028]	[0.029]
Mean control dep. var	0.02	0.05	0.09	0.10	0.13	0.17	0.10	0.24	0.25
Obs	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271	10,881
No. clusters	97	97	97	97	97	97	97	97	71
Panel B. All contracts independently of when they started									
Approved	-0.017	-0.049***	-0.026	-0.036**	0.006	-0.083***	-0.044**	-0.045	-0.046
	[0.012]	[0.015]	[0.018]	[0.017]	[0.023]	[0.024]	[0.020]	[0.029]	[0.030]
Mean control dep. var	0.04	0.06	0.09	0.10	0.13	0.17	0.12	0.25	0.26
Obs	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271	10,881
No. clusters	97	97	97	97	97	97	97	97	71

Note: this table presents the 2SLS estimation results of equation 1. The dependent variable in columns 1 to 6 is a binary variable for whether the student was in the formal labour market in the corresponding N-year window where N $\in \{1,2,3,4,5,6\}$. The dependent variable in column 7 is a binary variable for whether the student was in the formal labour market in any year between the first and third after the call's one, while that in column 8 is the same binary variable but that considers the period between the fourth and sixth year after the application. The dependent variable in column 9 follows the same logic, but considers the period between the seventh and eighth year after the application. All regressions include fixed effects for home university, university major, admission year, call's year and destination country, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at home university. Clustered robust standard errors at the call level are shown in square brackets. Data for labour market outcomes are available until 2020. 'Mean control dep. var' shows the mean of the dependent variable for not approved candidates. The number of observations displayed in each column is calculated excluding singletons. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7: Effects on being a firm owner or partner

	+1 year	+2 years	+3 years	+4 years	+5 years	+6 years	Pooled +1 to +3 years	Pooled +4 to +6 years	Pooled +7 to +8 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Only firms started up after the call's year									
Approved	-0.006	-0.021***	-0.011	-0.012	-0.021	0.017	-0.038**	-0.017	-0.038***
	[0.007]	[0.007]	[0.011]	[0.013]	[0.015]	[0.015]	[0.016]	[0.021]	[0.014]
Mean control dep. var	0.01	0.02	0.02	0.04	0.05	0.05	0.05	0.12	0.09
Obs	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271	10.881
No. clusters	97	97	97	97	97	97	97	97	71
Panel B. All firms independently of when they started									
Approved	-0.006	-0.021***	-0.011	-0.012	-0.021	0.017	-0.038**	-0.017	-0.038**
	[0.007]	[0.007]	[0.011]	[0.013]	[0.015]	[0.015]	[0.016]	[0.021]	[0.014]
Mean control dep. var	0.01	0.02	0.02	0.04	0.05	0.05	0.05	0.12	0.09
Obs	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271	10,881
No. clusters	97	97	97	97	97	97	97	97	71

Note: this table presents the 2SLS estimation results of equation 1. The dependent variable in columns 1 to 6 is a binary variable for whether the student became a sole-proprietorship firm owner or entered an existing society as a business partner in the corresponding N-year window where N $\in \{1,2,3,4,5,6\}$. The dependent variable in column 7 is a binary variable for whether the student became a sole-proprietorship firm owner or entered an existing society as a business partner in any year between the first and third after the call's one, while that in column 8 is the same binary variable but that considers the period between the fourth and sixth year after the application. The dependent variable in column 9 follows the same logic but considers the period between the seventh and eighth year after the application. All regressions include fixed effects for home university, university major, admission year, call's year, and destination country, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at home university. Clustered robust standard errors at the call level are shown in square brackets. Data for formal entrepreneurial activity are available until July 2021. 'Mean control dep. var' shows the mean of the dependent variable for not approved candidates. The number of observations displayed in each column is calculated excluding singletons. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

5.2 Mechanisms

In the previous section, we show that the CSF programme had a negative impact on being enrolled in a postgraduate programme in Brazil and on having a formal job, and a null impact on formal entrepreneurship. A natural next step would be to investigate what happened with students' academic trajectories after the programme. Unfortunately, only one out of the 13 universities in our sample provided detailed information about students' academic records, including the information whether students graduated and when.

We then turn to the Federal University of Bahia (UFBA), the largest university in the northeast region and one of the top fifteen universities in Brazil, to understand the programme's ITT impact on graduation rates. The UFBA sample contains 2,044 candidates and has information relative to graduation until the second semester of 2021. Column 1 of Table 8 shows that approved candidates are 18.5 p.p. more likely to graduate, but are also 23.1 p.p. less likely to graduate on time. Therefore, the programme does not reduce the probability of graduation.

Delayed graduation can affect employment and earnings through different channels. (Le Barbanchon et al. 2023), for example, show that students in high school and college who worked while studying in Chile had higher earnings due to the accumulation of work experience. (Häkkinen 2006) finds a similar result for Finnish students enrolled in college. In a setting where employers cannot observe students' productivity throughout the curriculum or GPA, work experience can act as a signal of motivation and productivity (Pallais 2014; Carranza et al. 2022). In addition, according to the vocational training literature (Alfonsi et al. 2020; Card et al. 2018), work experience can provide students with soft skills that are also important for employment and earnings. Therefore, the delayed exposure to professional work environments might in part explain the negative effects of the programme.

Columns 3 to 8 of Table 8 present the estimation results of the main outcomes for the UFBA sample. We show that the results do not systematically differ from what we observe for the whole sample. Columns 3 and 4 show that the programme did not increase the presence of students in postgraduate education programmes in Brazil. Columns 5 and 6 suggest a negative, and strong effect, of CSF on UFBA students' presence in the labour market. Surprisingly and different from the results for the whole sample, there is a positive effect on the probability of being a firm owner.

Finally, one may think that some approved candidates decided to return to a foreign country to either study or work after the CSF programme. This is an important issue in the context of the ERASMUS programme. In the case of CSF, however, we argue that it should not be a major issue. The first reason is that, according to UFBA data, approved candidates took longer to graduate, which means that after returning to Brazil, they had to spend more time at college. Note, for instance, that 20% of UFBA approved candidates did not graduate by December 2021. Moreover, the estimates from column 1 of Table 8 show a 23.7-percent increase in the likelihood of graduating compared to the control group, which would imply a graduation rate of 96% among the treated group. Therefore, students could migrate, but if they did so, they moved to a foreign country after finishing their undergraduate studies in Brazil. In this case, we would expect the brain drain to be more intense in the first three years after the call.

The second reason is that approved candidates had the obligation to stay in Brazil for at least the same duration of the scholarship. If they received a 12-month scholarship, they could not leave Brazil during the 12 months after the scholarship ended. If they did so, they could face a criminal charge and be required to return the scholarship value. In addition, Brazilians typically do not have a work or study permit to live in the U.S., Australia, or European countries. Students willing to obtain a visa normally must go through a long and expensive application process, which means that there are non-negligible barriers to international mobility. Therefore, our setting differs significantly from that of the high mobility of students within Europe (Parey and Waldinger 2011; Oosterbeek and Webbink 2011).

We are able to find 65% of the 19,245 candidates in the administrative registries within at least one year after the call. The other 35% must be in at least one of the following situations: i) unemployed, working informally, or looking for a job; ii) studying; iii) having started a business after July 2021 or entered a postgraduate programme after 2020; or iv) living in a foreign country. One may think that the best students, in terms of the entrance exam score, were more likely to leave Brazil after the programme. Although we cannot fully rule out this hypothesis, we provide evidence in Table B6 that there is a positive correlation between students' ability, measured by their home university entrance exam score, and the probability of being found in any outcome data set. Brain drain is commonly associated with the movement of high-skilled and likely the most talented workers, and thus seems to be less of a concern in our setting.

Table 8: Effects on graduation, on-time graduation, and the main outcomes for candidates enrolled at UFBA

	Graduation	On-time graduation	Postgrad. Pooled +1 to +3	Postgrad. Pooled +4 to +6	Formal emp. Pooled +1 to +3	Formal emp. Pooled +4 to +6	Firm owner Pooled +1 to +3	Firm owner Pooled +4 to +6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Second stage								
Approved	0.185*	-0.231***	-0.018	-0.073	-0.116***	-0.264***	-0.064	0.138**
	[0.104]	[0.051]	[0.054]	[0.074]	[0.032]	[0.070]	[0.043]	[0.064]
Mean control dep. var	0.78	0.18	0.08	0.19	0.10	0.24	0.07	0.16
Obs	2,044	2,044	2,044	2,044	2,044	2,044	2,044	2,044
No. clusters	85	85	78	78	78	85	85	85
Panel B. First stage								
Ratio	1.051***	1.055***	1.051***	1.051***	1.051***	1.072***	1.072***	1.072***
	[0.157]	[0.159]	[0.157]	[0.157]	[0.157]	[0.152]	[0.152]	[0.152]
F-stat of Instrument	44.54	43.98	44.54	44.54	44.54	49.85	49.85	50.85

Note: this table presents the 2SLS estimation results of equation 1. The dependent variable in column 1 is an indicator of whether the UFBA candidate graduated in his university major, while that in column 2 is an indicator of whether he graduated on time. The dependent variables in columns 3 to 5 refer to the pooled post-call period. The pooled post-call period encompasses up to six years for those that applied in 2014, seven years for those that applied in 2013, eight years for those that applied in 2012 and nine years for those that applied in 2011. All regressions include fixed effects for major, admission year, call's year, and destination country, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at UFBA. Clustered robust standard errors at the call level are shown in square brackets. 'Mean control dep. var' shows the mean of the dependent variable for candidates that were not approved. The number of observations displayed in each column is calculated excluding singletons. The dependent variables are described in detail in Appendix Table E11. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

5.3 Attrition

A potential criticism of the analysis is the possibility of having a larger sample of candidates from the control group because some of the approved students would decide not to return to Brazil after leaving the country. To provide some evidence on this matter, we estimate the same model presented in 1 on the probability of being found in any outcome data set separately for the first to the sixth year after applying for a CSF call. The intuition is that if the effect is null or not significant, there is no meaningful difference in terms of differential attrition.

As shown in Figure A10, the difference between the call date and the end of the scholarship is of approximately three years. Therefore, we should expect a negative probability of finding approved students in any outcome data set in the first three years. But as time goes on, the probability should decrease. This is precisely the result found in Table 9. It also presents some evidence that the probability of finding not approved candidates is higher than that of finding approved candidates. We restrict the analysis to year-by-year because the pooled sample estimation can mask students who appear in the data set in only one year, for example. The results indicate that there is a negative probability of finding students in any outcome data set in the first three years of the programme. In the fourth year, this effect reduces by 21%. In the fifth year, it becomes non-significant, and in the sixth year, it becomes slightly significant, but its effect size reduces by 60% compared to that of the third year. We also add the estimation for the seven and eight years after the call, separately. The results suggest the same pattern, a reduction until the parameter gets non-significant in the eighth year.

The results in Table 9 suggest that our estimations in Tables 5, 6, and 7 are downward biased if the not-found approved students are most productive, which would imply easier access to the labour market and postgraduate education programmes. However, Appendix Table B6 points to a negative association between entrance exam scores, our best measure of productivity, and being found in any outcome data set among approved candidates.

Finally, one may also think that even though we found 65% of students in at least one year, the number of individuals found is low year-to-year. Table 9 shows that it varies from 6% in the first year after the call to 39% in the sixth year, and achieves 44% seven years after application. It is important to consider that Brazil faced its worst recession period in the last thirty years between 2014 and 2019, which was exacerbated in 2020 because of the COVID-19 pandemic.[18] Appendix Figure A11 shows that most of the UFBA-CSF candidates graduated during the recession. Therefore, the low number of students found in the outcome data sets could also be explained by graduating during a recession, which may harm students' career perspectives (von Wachter 2020; Oreopoulos et al. 2012).

[18] In 2014, Brazil ended a period of positive economic growth. By 2019, before the COVID-19 pandemic, the GDP decreased by more than 7%. See more: https://www.nytimes.com/2014/08/30/business/international/brazil-fell-into-recession-in-first-half-of-year.html and https://www.imf.org/en/Publications/WP/Issues/2018/01/12/Investment-in-Brazil-From-Crisis-to-Recovery-45557.

Table 9: Effects on the probability of finding the candidate in any outcome data set

	+1 year	+2 years	+3 years	+4 years	+5 years	+6 years	+7 years	+8 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Approved	-0.043**	-0.137***	-0.143***	-0.114***	-0.031	-0.062*	-0.057*	-0.03
	[0.017]	[0.024]	[0.034]	[0.041]	[0.035]	[0.033]	[0.031]	[0.057]
Mean control dep. var	0.05	0.14	0.24	0.31	0.36	0.39	0.44	0.42
Obs	14,271	14,271	14,271	14,271	14,271	14,271	10,881	4807
No. clusters	97	97	97	97	97	97	71	30

Note: this table presents the 2SLS estimation results of equation 1. The dependent variable in columns 1 to 6 is a binary variable for whether the candidate was found in any outcome data set in the corresponding N-year window where N $\in \{1,2,3,4,5,6\}$. The dependent variable in column 7 is a binary variable for whether the student was found in any outcome data set in any year between the first and third after the call, while that in column 8 is the same binary variable but it considers the period between the fourth and sixth year after the application. The dependent variable in column 9 follows the same logic, but considers the period between the first and sixth year after the application. The dependent variable in column 10 is a binary variable for whether the student was found in any outcome data set after the call's year. The pooled post-call period encompasses up to six years for those that applied in 2014, seven years for those that applied in 2013, eight years for those that applied in 2012 and nine years for those that applied in 2011. All regressions include fixed effects for home university, major, admission year, call's year and destination country, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at home university. Clustered robust standard errors at the call level are shown in square brackets. 'Mean control dep. var' shows the mean of the dependent variable for not approved candidates. The number of observations displayed in each column is calculated excluding singletons. *** p<0.01, ** p<0.05, * p<0.1.

6 Conclusion

This paper estimates the intention-to-treat effects of the Science without Borders programme in Brazil in many outcomes, covering career decisions relative to the formal labour market, postgraduate education, and formal entrepreneurial activity. The programme provided scholarships for undergraduates to attend an academic year in universities located in developed countries, mostly in Europe and North America. To deal with the potential bias caused by self-selection into the programme, we propose an instrumental variable approach. The identification strategy assumes that variations in the approval rate across programme selection calls are orthogonal to any latent determinants of the outcome variables of interest.

Our main result indicates that the programme did not achieve its main goals. We found a negative impact on the probability of enrolling in a Brazilian postgraduate programme and no effect on the probability of having a formal job, or being a formal firm owner or partner. It means that one of the medium-term expected impacts of government-sponsored study abroad programmes, namely to increase the local pool of highly trained individuals, might not be an easy-to-achieve goal in the context of a developing country. Despite the negative results, we believe that study abroad programmes can be good policies and have the potential to enhance the national level of human capital. However, these programmes need to be carefully designed and implemented.

In our setting, some programme characteristics seem to have a substantial influence on our conclusions. First, we highlight the abundant availability of scholarships and the possibly loose selection of candidates to meet the audacious goal of sending 100,000 students abroad in just four years. Second, the focus on undergraduates was not backed up by international evidence or successful policy experiences. These issues, especially important for developing countries, may hinder the possible positive impacts of study abroad programmes that countries worldwide implement to enhance their national knowledge base in fields considered a top priority for economic development. Furthermore, the mechanism section suggests that the negative results arise not because the programme harmed students' human capital, but because it delayed students' entrance into the labour market. Therefore, the long-term effects of the programme may differ from the observed results.

References

Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., and Vitali, A. (2020). 'Tackling youth unemployment: Evidence from a labor market experiment in Uganda'. *Econometrica*, 88(6): 2369–414. https://doi.org/10.3982/ECTA15959

Araújo, B. d., Cavalcante, L. R., and Alves, P. (2009). 'Variáveis proxy para os gastos empresariais em inovação com base no pessoal ocupado técnico-científico disponível na Relação Anual de Informações Sociais (RAIS)'. *Radar: Tecnologia, Produção e Comércio Exterior 5*. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea). https://repositorio.ipea.gov.br/handle/11058/5431

Aveiro, T. (2014). 'O programa ciência sem fronteiras como ferramenta de acesso à mobilidade internacional'. *Tear: Revista de Educação, Ciência e Tecnologia*, 3(2): 1–21. https://doi.org/10.35819/tear.v3.n2.a1867

Blasnik, M. (2010). 'RECLINK: Stata module to probabilistically match records'. Statistical Software Components S456876. Boston: Boston College Department of Economics.

Borusyak, K., and Jaravel, X. (2018). 'The distributional effects of trade: Theory and evidence from the united states'. Working Paper.

Bound, J., Braga, B., Khanna, G., and Turner, S. (2021). 'The Globalization of Postsecondary Education: The Role of International Students in the US Higher Education System'. *Journal of Economic Perspectives*, 35(1): 163–84. https://doi.org/10.1257/jep.35.1.163

Brazil (2010). *Ciência sem fronteiras: um programa especial de mobilidade internacional em ciãncia, tecnologia e invoção.* (unpublished)

Cammelli, A., Ghiselli, S., and Mignoli, G. P. (2006). 'Study Experience Abroad: Italian Graduate Characteristics and Employment Outcomes'. In M. Byram and F. Dervin (eds), *Students, Staff and Academic Mobility in Higher Education.* Cambridge: Cambridge Scholars Publishing.

CAPES (2015). *Manual para Bolsistas: Graduação Sanduíche.* Brasília: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

Card, D., Kluve, J., and Weber, A. (2018, 10). 'What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations'. *Journal of the European Economic Association*, 16(3): 894–931. https://doi.org/10.1093/jeea/jvx028

Carranza, E., Garlick, R., Orkin, K., and Rankin, N. (2022, November). 'Job search and hiring with limited information about workseekers' skills'. *American Economic Review*, 112(11): 3547–83. https://doi.org/10.1257/aer.20200961

Castro, C. d. M., Barros, H., Ito-Adler, J., and Schwartzman, S. (2012). 'Cem mil bolsistas no exterior'. *Interesse Nacional*, 4(17): 25–36.

Chaves, V. L. J., and Castro, A. M. D. (2016). 'Internacionalização da educação superior no brasil: programas de indução à mobilidade estudantil'. *Revista Internacional de Educação Superior*, 2(1): 118–37. https://doi.org/10.22348/riesup.v2i1.7531

Christen, P. (2012). *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*. Springer: New York. https://doi.org/10.1007/978-3-642-31164-2

CNI (2015). *Fortalecimento das engenharias.* Brasília: Confederação Nacional da Indústria.

Cruz, V. X., and Eichler, M. L. (2021). 'Bolsas capes de mobilidade acadêmica internacional 1952-2019: um estudo a partir dos contextos de internacionalização da educação superior'. *Revista Brasileira de Pós-Graduação*, 17(37): 1–25. https://doi.org/10.21713/rbpg.v17i37.1768

Czarnitzki, D., Joosten, W., and Toivanen, O. (2021). 'International student exchange and academic performance'. ZEW Discussion Paper 21-011. Mannheim: ZEW - Leibniz Centre for European Economic Research. https://doi.org/10.2139/ssrn.3792802

De Negri, F. (2021). 'Políticas Públicas para Ciência e Tecnologia no Brasil: Cenário e Evolução Recente'. Nota Técnica 92. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea). https://doi.org/10.38116/ntdiset92

Di Pietro, G. (2012). 'Does studying abroad cause international labor mobility? Evidence from Italy'. *Economics Letters*, 117(3): 632–35. https://doi.org/10.1016/j.econlet.2012.08.007

Di Pietro, G. (2015). 'Do study abroad programs enhance the employability of graduates?' *Education Finance and Policy*, 10(2): 223–43. https://doi.org/10.1162/EDFP_a_00159

Di Pietro, G. (2019). 'University study abroad and graduates' employability'. *IZA World of Labor*: 109v2. https://doi.org/10.15185/izawol.109.v2

European Commission (2014). *Effects of mobility on the skills and employability of students and the internationalisation of higher education institutions*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2766/75468

European Commission (2019). *Erasmus+ higher education impact study*. Luxembourg: Publications Office of the European Union. https://doi.org/10.2766/162060

FAPESP (2017). 'Experiência encerrada: O programa de intercâmbio Ciência sem Fronteiras, que gastou R$ 13,2 bilhões, a maior parte com bolsas de graduação no exterior, deixa de existir'. *Revista Pesquisa FAPESP*, 256(-): 27–29.

Giorgio, D. P. (2021). 'Studying Abroad and Earnings: A Meta-Analysis'. *Journal of Economic Surveys*, 36(4): 1096–129. https://doi.org/10.1111/joes.12472

Granja, C., and Carneiro, A. M. (2021). 'O programa ciência sem fronteiras e a falha sistêmica no ciclo de políticas públicas'. *Ensaio: Avaliação e Políticas Públicas em Educação*, 29(110): 183–205. https://doi.org/10.1590/S0104-40362020002801962

Grochocki, L. F. M., and Guimarães, J. A. (2017). 'A contribuição do programa capes/brafitec para a internacionalização dos cursos de graduação em engenharia no Brasil'. *Revista de Ensino de Engenharia*, 36(1): 72–84. https://doi.org/10.5935/2236-0158.20170007

Gusso, D. A., and Nascimento, P. M. (2014). 'Evolução da formação de engenheiros e profissionais técnico-científicos no brasil entre 2000 e 2012'. Texto para Discussão 1982. Brasília: Instituto de Pesquisa Econômica Aplicada (Ipea).

Häkkinen, I. (2006). 'Working while enrolled in a university: does it pay?' *Labour Economics*, 13(2): 167–89. https://doi.org/10.1016/j.labeco.2004.10.003

INEP (2016). *Indicadores Financeiros Educacionais.* Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

Iriondo, I. (2020). 'Evaluation of the impact of Erasmus study mobility on salaries and employment of recent graduates in Spain'. *Studies in Higher Education*, 45(4): 925–43. https://doi.org/10.1080/03075079.2019.1582011

Knobel, M. (2012). 'Brazil: Brazil Seeks Academic Boost by Sending Students Abroad'. In G. Mihut, P. G. Altbach, and H. d. Wit (eds), *Understanding Higher Education Internationalization. Global Perspectives on Higher Education* (pp. 147–49). Rotterdam: SensePublishers. https://doi.org/10.1007/978-94-6351-161-2_32

Le Barbanchon, T., Ubfal, D., and Araya, F. (2023). 'The effects of working while in school: evidence from Uruguayan lotteries'. *American Economic Journal: Applied Economics*, 15(1): 383–410. https://doi.org/10.1257/app.20210041

Lee, D. S., McCrary, J., Moreira, M. J., and Porter, J. (2022). 'Valid t-ratio Inference for IV'. *American Economic Review*, 112(10): 3260–90. https://doi.org/10.1257/aer.20211063

Li, J. (2016). *The Impact of Study Abroad on Student Academic Achievement, Global Perspectives and Labor Market Outcomes: Evidence from US undergraduate students* (Doctoral dissertation, Columbia University). https://doi.org/10.7916/D85X28ZH

Liwiński, J. (2019a). 'Does it pay to study abroad? Evidence from Poland'. *International Journal of Manpower*, 40(3): 525–55. https://doi.org/10.1108/IJM-11-2017-0305

Liwiński, J. (2019b). 'Does studying abroad enhance employability?' *Economics of Transition and Institutional Change*, 27(2): 409–23. https://doi.org/10.1111/ecot.12203

Manços, G., and Coelho, F. (2017). 'Internacionalização da Ciência Brasileira: Subsídios para Avaliação do Programa Ciência sem Fronteiras'. *Revista Brasileira de Políticas Públicas e Internacionais*, 2(2): 52–82.

Martins, J. (2015). *Programa Ciência Sem Fronteiras no contexto da política de internacionalização da educação superior brasileira.* (Masters thesis, Department of Education, Federal University of Mato Grosso)

Mazza, D. (2009). 'Intercâmbios acadêmicos internacionais: bolsas Capes, CNPq e Fapesp'. *Cadernos de Pesquisa*, 39(137): 521–47. https://doi.org/10.1590/S0100-15742009000200010

McManus, C., and Nobre, C. A. (2017). 'Brazilian Scientific Mobility Program-Science without Borders - Preliminary Results and Perspectives'. *Anais da Academia Brasileira de Ciências*, 89(): 773–86. https://doi.org/10.1590/0001-3765201720160829

Meya, J., and Suntheim, K. (2014). 'The second dividend of studying abroad: The impact of international student mobility on academic performance'. cege Discussion Paper 215. Göttingen: Centre for European, Governance and Economic Development Research (cege). https://doi.org/10.2139/ssrn.2501317

Netz, N., and Cordua, F. (2021, Jul.). 'Does studying abroad influence graduates' wages? A literature review'. *Journal of International Students*, 11(4): 768—89. https://doi.org/10.32674/jis.v11i4.4008

OECD (2015). *OECD Science, Technology and Industry Scoreboard 2015: Innovation for Growth and Society*. Paris: OECD Publishing. https://doi.org/10.1787/20725345

OECD (2019). 'Country note: Brazil'. *Education at a Glance 2019*. Paris: OECD.

Oosterbeek, H., and Webbink, D. (2011). 'Does studying abroad induce a brain drain?' *Economica*, 78(310): 347–66. https://doi.org/10.1111/j.1468-0335.2009.00818.x

Oreopoulos, P., Von Wachter, T., and Heisz, A. (2012). 'The short-and long-term career effects of graduating in a recession'. *American Economic Journal: Applied Economics*, 4(1): 1–29. https://doi.org710.1257/app.4.1.1

Orrù, E. (2014). *Student mobility policies in the European Union: the case of the Master and Back programme: private returns, job matching and determinants of return migration* (Doctoral dissertation, The London School of Economics and Political Science (LSE)). http://etheses.lse.ac.uk/id/eprint/942

Pallais, A. (2014). 'Inefficient hiring in entry-level labor markets'. *American Economic Review*, 104(11): 3565–99. https://doi.org/10.1257/aer.104.11.3565

Parey, M., and Waldinger, F. (2011). 'Studying abroad and the effect on international labour market mobility: Evidence from the introduction of ERASMUS'. *The Economic Journal*, 121(551): 194–222. https://doi.org/10.1111/j.1468-0297.2010.02369.x

Pinto, F. (2022). 'The effect of university graduates' international mobility on labour outcomes in Spain'. *Studies in Higher Education*, 47(1): 26–37. https://doi.org/10.1080/03075079.2020.1725877

Raffo, J. (2020). 'MATCHIT: Stata module to match two datasets based on similar text patterns'. Statistical Software Components S457992. Boston: Boston College Department of Economics.

Rodrigues, M. (2013). *Does student mobility during higher education pay? Evidence from 16 European countries*. Luxembourg: Publications Office of the European Union,. (Joint Research Centre Scientific and Policy Reports, Report EUR 26089) https://doi.org/10.2788/95642

Sá, C. M. (2016). 'The rise and fall of Brazil's Science Without Borders'. *International Higher Education*, Spring(85): 17–18. https://doi.org/10.6017/ihe.2016.85.9241

Saldanha, C., de Oliveira, L., Aburachid, V., and Denardi, A. (2019). 'PROGRAMA CIÊNCIA SEM FRONTEIRAS: um retrospecto da política de estímulo à ciência, tecnologia e inovação'. *Revista de Políticas Públicas*, 23(2): 675–94. https://doi.org/10.18764/2178-2865.v23n2p675-694

Schnepf, S. V., and d'Hombres, B. (2018). 'International mobility of students in Italy and the UK: does it pay off and for whom?'. IZA Discussion 12033. Bonne: Institute of Labor Economics (IZA). https://doi.org/10.2139/ssrn.3318807

Senado Federal (2015). *Avaliação de Politica Publica: Programa Ciência Sem Fronteiras.* (Relatório da Comissão de Ciência e Tecnologia)

Sorrenti, G. (2017). 'The Spanish or the German apartment? Study abroad and the acquisition of permanent skills'. *Economics of Education Review*, 60(October): 142–58. https://doi.org/10.1016/j.econedurev.2017.07.001

Sutton, R. C., and Rubin, D. L. (2004). 'The GLOSSARI project: Initial findings from a system-wide research initiative on study abroad learning outcomes'. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10(1): 65–82. https://doi.org/10.36366/frontiers.v10i1.133

UIS (2023). 'Education Data'. UIS database. Montreal: UNESCO Institute for Statistic (UIS). Available at: http://data.uis.unesco.org/ (accessed 1 March 2022)

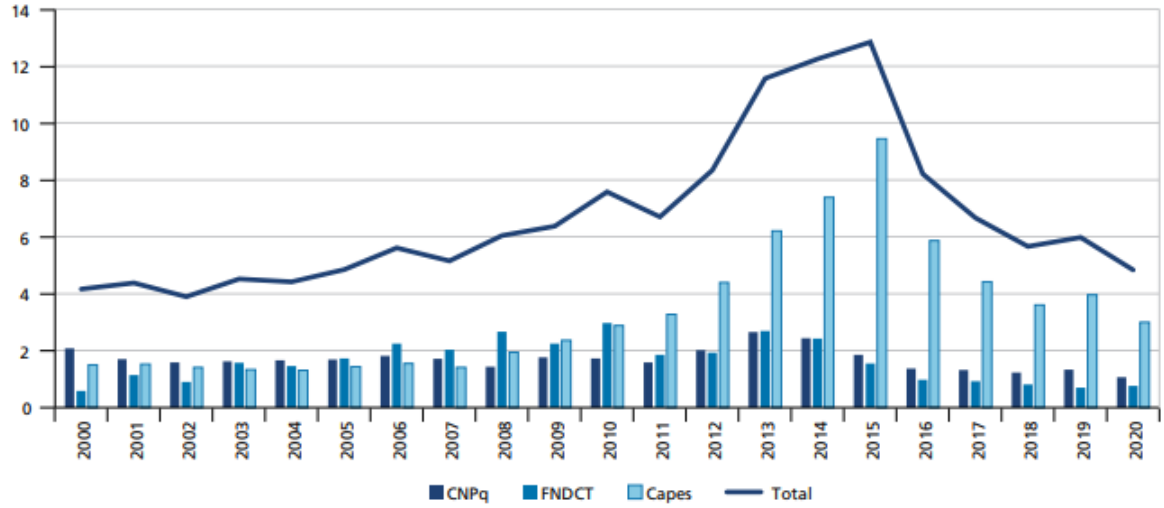
UK HE International Unit (2015). *Gone international: mobile students and their outcomes: Report on the 2012/13 graduating cohort.* London: Go International programme based at the UK Higher Education International Unit.

von Wachter, T. (2020). 'The persistent effects of initial labor market conditions for young adults and their sources'. *Journal of Economic Perspectives*, 34(4): 168–94. https://doi.org/10.1257/jep.34.4.168

Appendix[19]

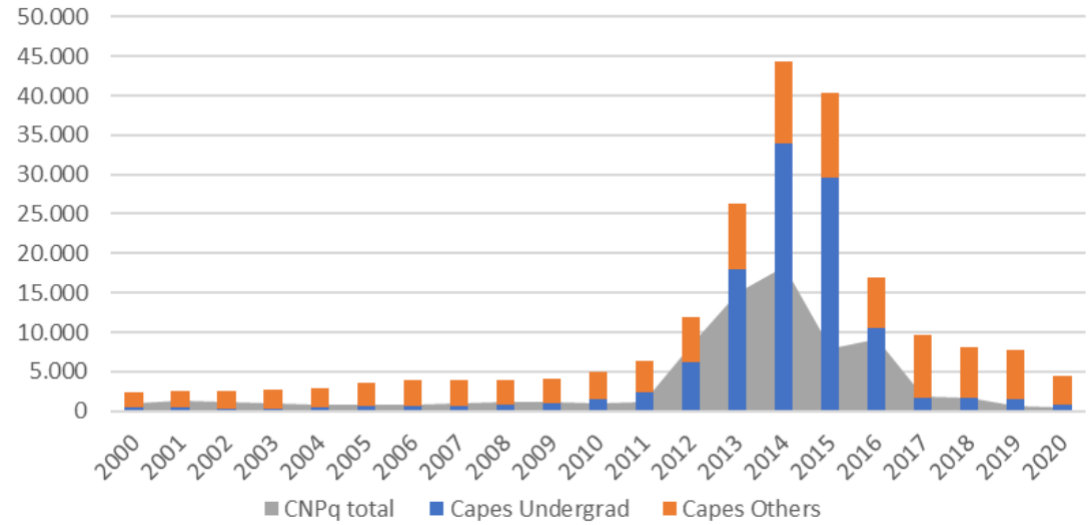
A Figures

Figure A1: Government spending with science and technology in Brazil (in BRL millions)



Note: BRL stands for Brazilian reais. CAPES and CNPq are the main research funding agencies of the Brazilian Ministry of Education, and Ministry of Science and Technology, respectively. FNDCT stands for the National Fund for Scientific and Technological Development.
Source: Figure 3 in De Negri (2021). Available at: https://doi.org/10.38116/ntdiset92.

Figure A2: Number of government-sponsored undergraduate scholarships per year in Brazil

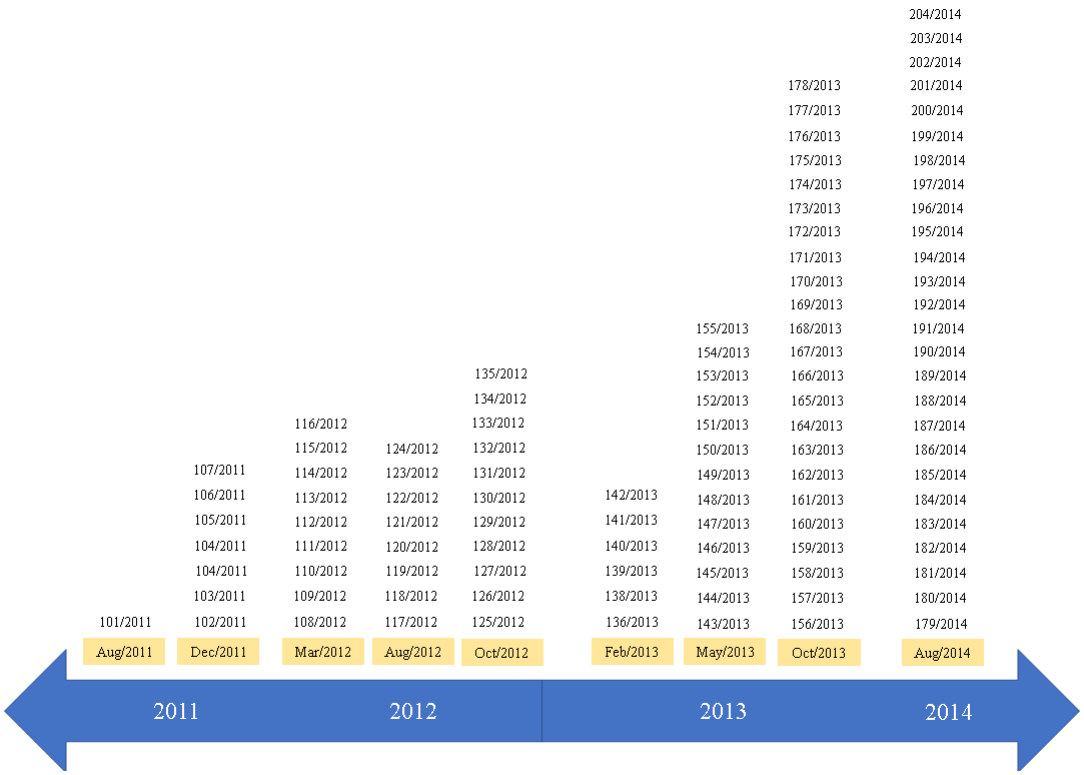


Note: 'CNPq total' corresponds to the number of scholarships granted by CNPq both for undergraduate and graduate students, while 'CAPES Undergrad' corresponds to the number of scholarships granted by CAPES for undergraduates, and 'CAPES Other' shows the number of scholarships granted by CAPES for students other than undergraduates.

Source: used courtesy of Fernanda De Negri.

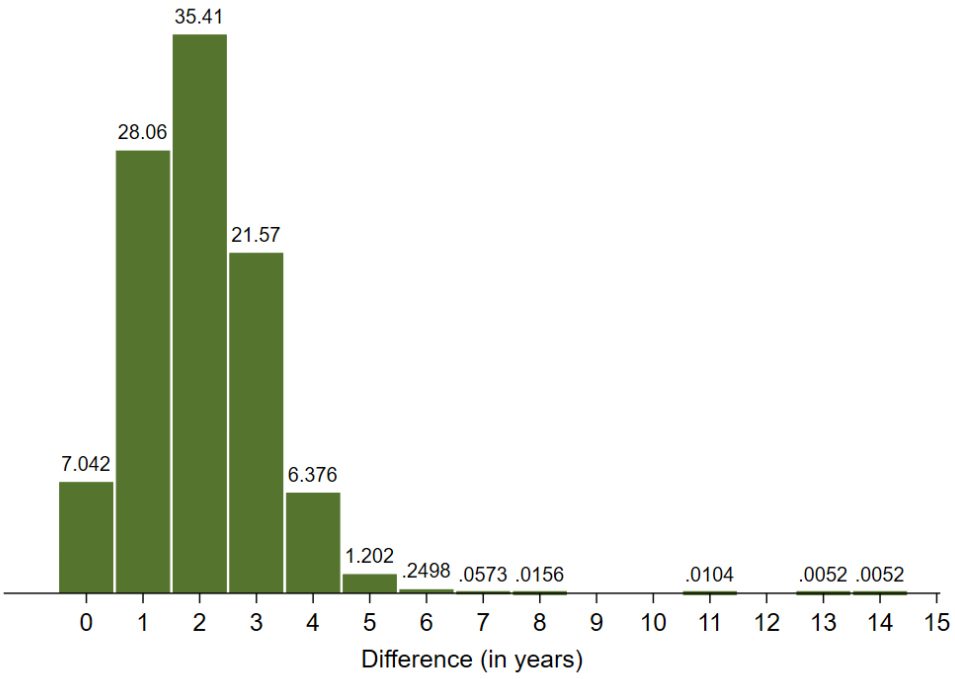
[19] Appendix formatted but not checked

Figure A3: CSF calls by launching year and month



Source: this figure is based on Table 1 from Martins (2015), which compiled information about CSF calls for undergraduates. Call No. 137/2012 was not launched.

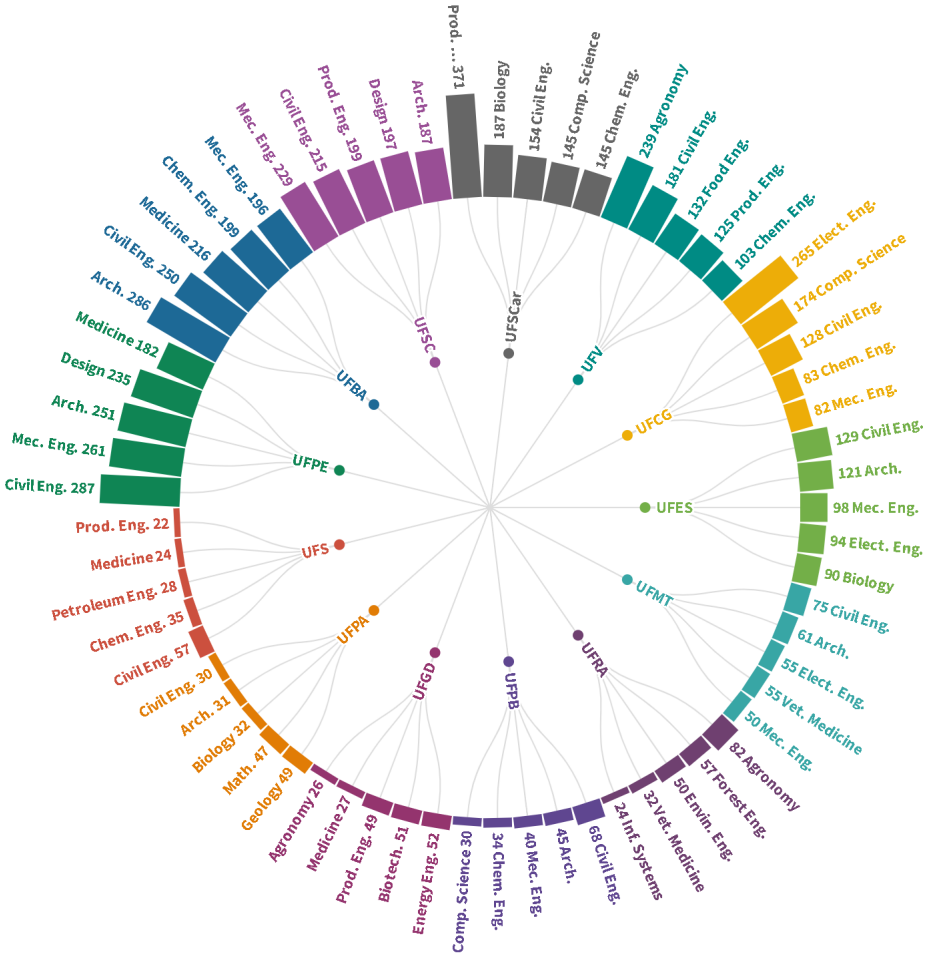
Figure A4: Distribution of the difference between candidates' call year and their admission year at home university



Note: the values depicted above bars are percentages.

Source: authors' elaboration.

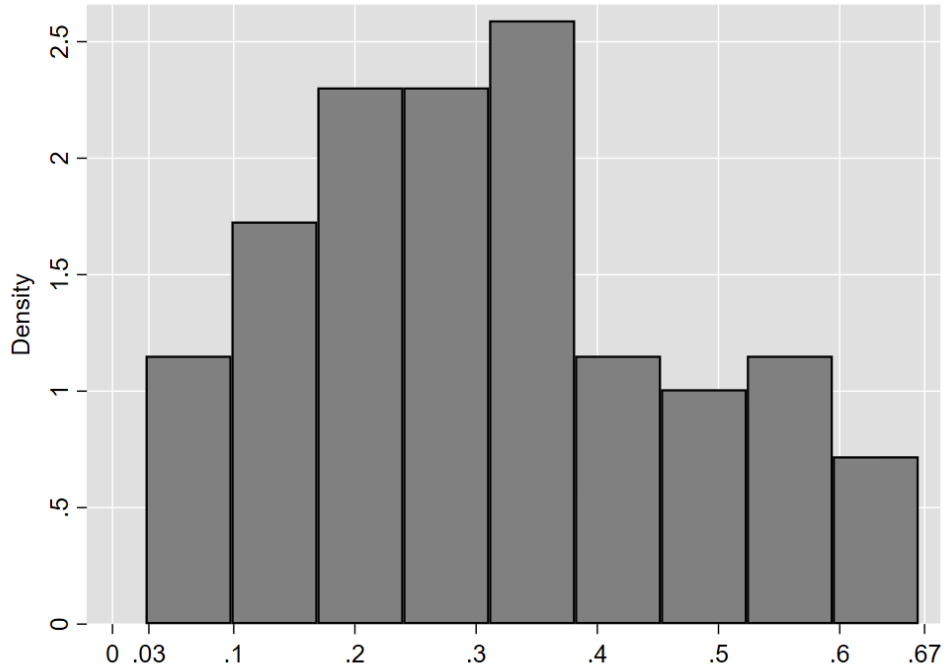
Figure A5: Number of candidates by major in each home university



Note: we restrict the figure to show only the top five most frequent majors in each home university. Each bar represents a given major and the number next to the bar label indicates the number of candidates in the corresponding major-home university combination.

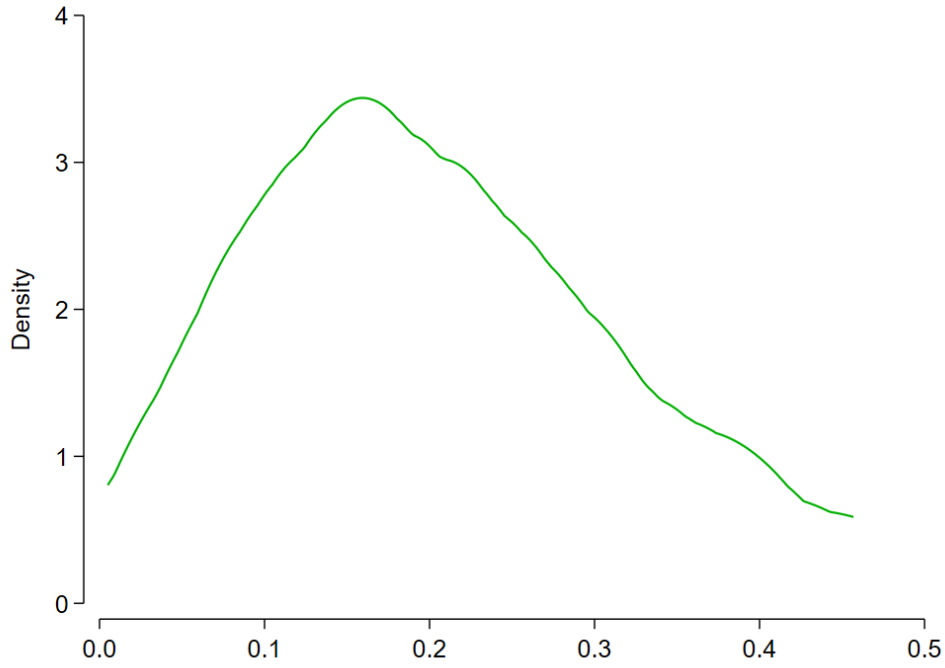
Source: authors' elaboration.

Figure A6: Distribution of calls' approval rate



Source: authors' elaboration.

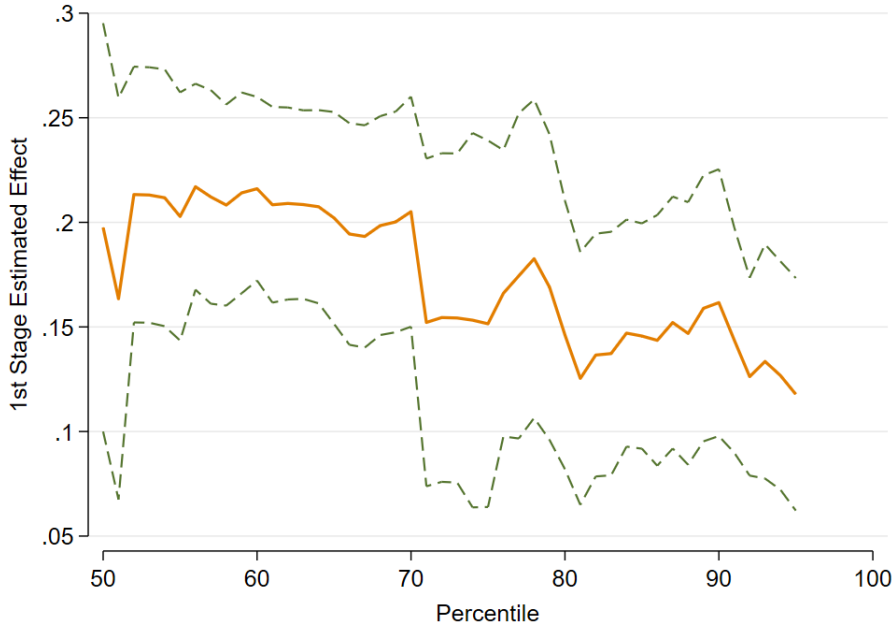
Figure A7: Distribution of the difference in the approval rate across calls for the same destination country and launching year



Note: the variable whose distribution is displayed is the difference in the discounted-version approval rate of calls for the same destination country and year. The difference is calculated as the difference between the highest and the lowest approval rate of calls for the same destination country and year. To produce such figure, we created a data set in which each observation is a combination call's destination country-year.

Source: authors' elaboration.

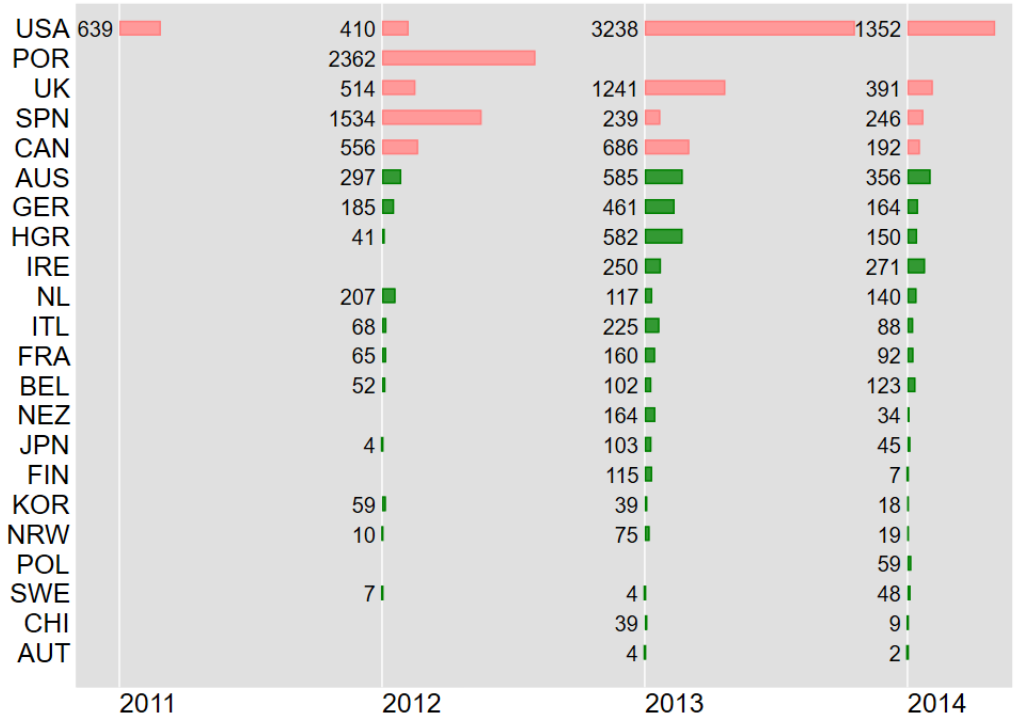
Figure A8: First stage estimated effect using different moments of the distribution of candidate's call approval rate among single-call applicants



Note: the solid line represents the point estimates while the dashed lines represent the 95% confidence intervals. Each estimate is obtained from a specification that follows column 5 of Table B9. The variable 'Percentile' is a binary variable for when the value of the approval rate of a given call is greater than the value of the approval rate associated with the percentile #N where N takes every integer value between 50 and 95.

Source: authors' elaboration.

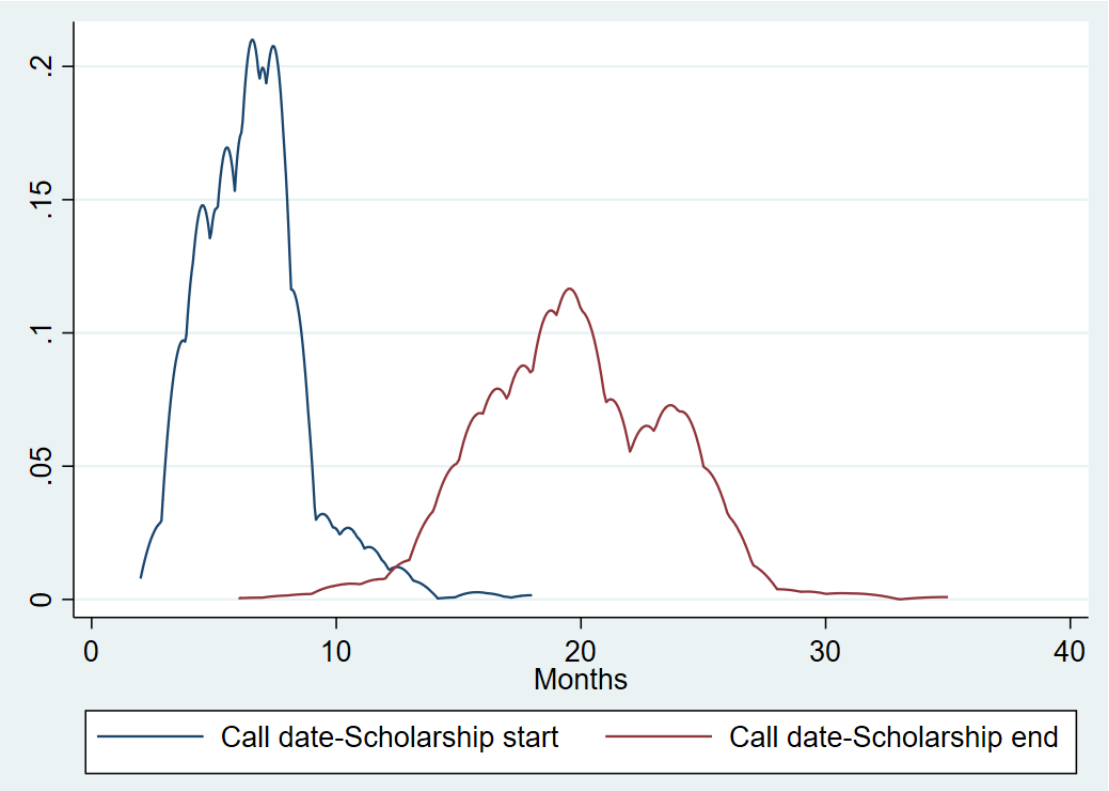
Figure A9: Number of candidates by applicant's first call destination country and launching year



Note: the five destination countries with the highest number of candidates are highlighted in orange.

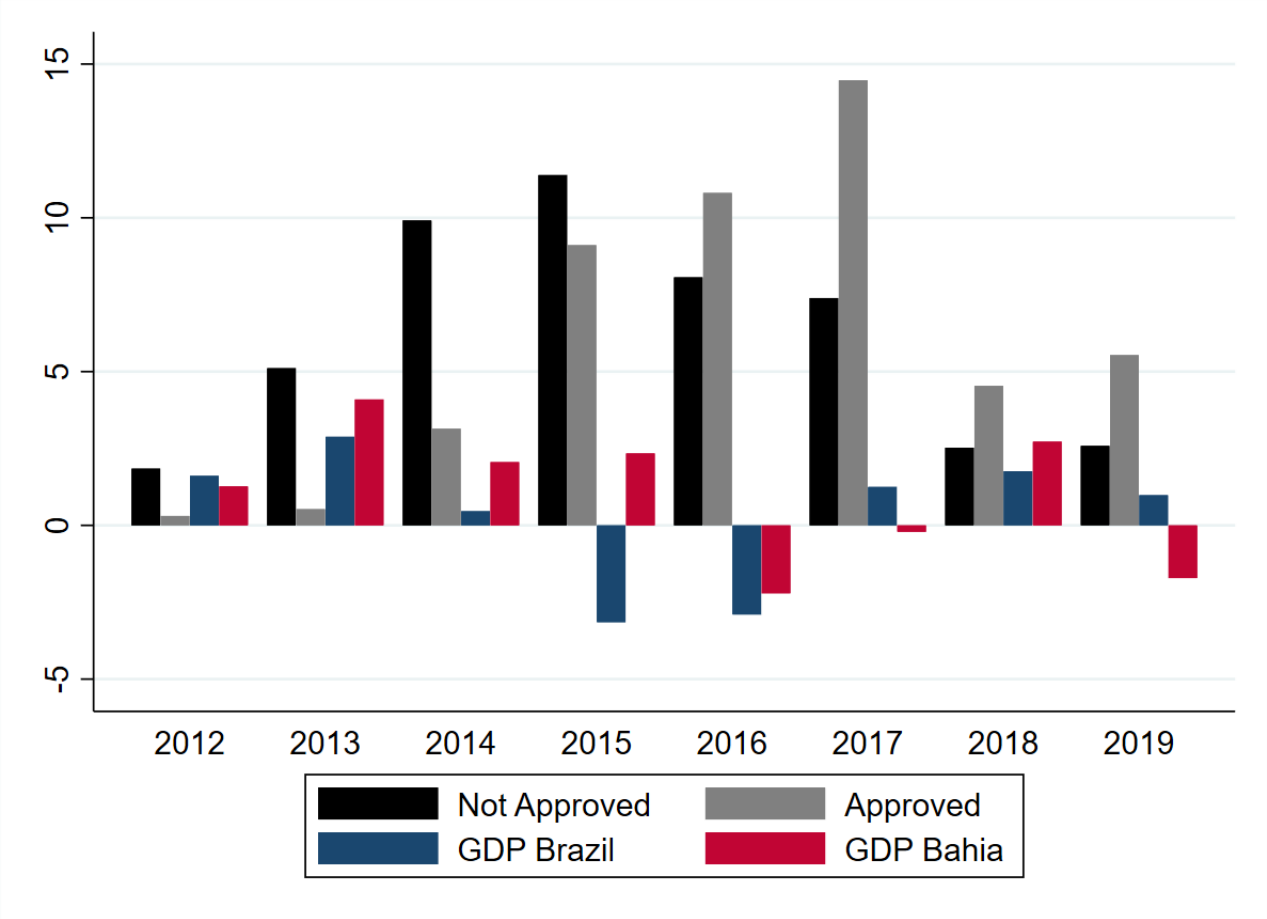
Source: authors' elaboration.

Figure A10: Distribution of the difference between the first call's launching year-month and the CSF scholarship start and end among UFBA candidates



Note: the blue curve is the difference in months, between the call date and the scholarship start date. The red curve is the difference in months, between the call date and the scholarship end date.
Source: the figure was produced with data relative to 828 approved UFBA applicants for which we have information about both the start and end dates of the CSF scholarship.

Figure A11: Distribution of the graduation years among UFBA candidates by candidates' approval status



Source: authors' elaboration.

B Tables

Table B1: Number of candidates by number of calls they applied for and number of majors in which they have ever been enrolled at home university

		No. of majors					
		1	2	3	4	5	Total
No. of calls	1	14,106	1,809	198	31	7	16,151
		[73%]	[9%]	[1%]	[0%]	[0%]	[84%]
	2	2,333	333	33	9	2	2,710
		[12%]	[2%]	[0%]	[0%]	[0%]	[14%]
	3	299	44	7	1	0	351
		[2%]	[0%]	[0%]	[0%]	[0%]	[2%]
	4	17	8	1	0	0	26
		[0%]	[0%]	[0%]	[0%]	[0%]	[0%]
	5	7	0	0	0	0	7
		[0%]	[0%]	[0%]	[0%]	[0%]	[0%]
	Total	16,762	2,194	239	41	9	19,245
		[87%]	[11%]	[1%]	[0%]	[0%]	

Note: percentages relative to the total number of candidates are shown in square brackets.

Table B2: Home universities with the highest number of CSF candidates

Ranking	University	No. of candidates
1º	Federal University of Minas Gerais (UFMG)	5,571
2º	University of São Paulo (USP)	5,494
3º	University of Brasília (UnB)	4,252
4º	Federal University of Santa Catarina (UFSC)	3,680
5º	Federal University of Rio de Janeiro (UFRJ)	3,414
6º	Federal Technological University of Paraná (UTFPR)	3,408
7º	Federal University of Ceará (UFC)	3,398
8º	Federal University of Pernambuco (UFPE)	3,219
9º	Federal University of Bahia (UFBA)	2,894
10º	São Paulo State University (Unesp)	2,842
11º	Federal University of Paraná (UFPR)	2,600
12º	Federal University of Rio Grande do Norte (UFRN)	2,529
13º	Pontifical Catholic University of Minas Gerais (PUC-Minas)	2,507
14º	State University of Campinas (Unicamp)	2,506
15º	Fluminense Federal University (UFF)	2,376
16º	Federal University of São Carlos (UFSCar)	2,322
17º	Federal University of Viçosa (UFV)	2,287
18º	Federal University of Rio Grande do Sul (UFRGS)	2,180
19º	Federal University of ABC (UFABC)	1,740
20º	Federal University of Ouro Preto (UFOP)	1,502
21º	Federal University of Goiás (UFG)	1,497
22º	Federal University of Itajúba (UNIFEI)	1,466
23º	Federal University of Espírito Santo (UFES)	1,456
24º	Federal University of São João Del Rei (UFSJ)	1,435
25º	Federal University of Santa Maria (UFSM)	1,410
26º	Federal University of Juiz de Fora (UFJF)	1,387
27º	Federal University of Uberlândia (UFU)	1,274
28º	Federal University of Pará (UFPA)	1,264
29º	Federal University of Campina Grande (UFCG)	1,260
30º	Federal University of Alagoas (UFAL)	1,256

Table B3: Most frequent postgraduate programmes among CSF candidates

Postgraduate programme	N	Cum	Pct (%)	Cum Pct (%)
Electrical engineering	297	297	5.5%	5.5%
Mechanical engineering	260	557	4.8%	10.4%
Civil engineering	254	811	4.7%	15.1%
Computer sciences	251	1,062	4.7%	19.8%
Chemical engineering	207	1,269	3.9%	23.6%
Chemistry	170	1,439	3.2%	26.8%
Production engineering	83	1,522	1.5%	28.3%
Physics	80	1,602	1.5%	29.8%
Architecture and urbanism	79	1,681	1.5%	31.3%
Materials engineering	66	1,747	1.2%	32.5%
Other	3,625	5,372	67.5%	100.0%

Note: the frequencies were calculated with the data set at the level of combinations candidate-postgraduate programme for the period between 2013 and 2019. Consequently, there might exist more than one postgraduate programme associated with a given student.

Table B4: Most frequent occupations among CSF candidates

Occupation	N	Cum	Pct (%)	Cum Pct (%)
Administrative assistant	359	359	9.6%	9.6%
System analyst	239	598	6.4%	16.0%
Manager	128	726	3.4%	19.5%
Salesperson	115	841	3.1%	22.6%
Pharmacist	83	924	2.2%	24.8%
Short-courses instrutor	67	991	1.8%	26.6%
Business analyst	62	1,053	1.7%	28.3%
Production engineer	50	1,103	1.3%	29.6%
Graphic designer	46	1,149	1.2%	30.8%
Civil engineer	45	1,194	1.2%	32.0%
Other	2,533	3,727	68.0%	100.0%

Note: the frequencies were calculated with the data set at the level of combinations candidate-occupations for the period between 2013 and 2018. Consequently, there might exist more than one occupation associated with a given student. We only considered contracts that started after the candidates' call year.

Table B5: Most frequent (main) economic activity of firms connected to CSF candidates

Economic activity	N	Cum	Pct (%)	Cum Pct (%)
Engineering services	254	254	4.5%	4.5%
Professional and managerial development training	220	474	3.9%	8.4%
Unspecified instructional activities	188	662	3.3%	11.8%
Graphical design services	184	846	3.3%	15.0%
Specialized administrative support services	156	1,002	2.8%	17.8%
Retail sale of clothing and accessories	139	1,141	2.5%	20.3%
Outpatient medical activity restricted to consultations	139	1,280	2.5%	22.8%
Sales promotion	119	1,399	2.1%	24.9%
Emergency care activities	119	1,518	2.1%	27.0%
Repair of computers and peripheral equipment	107	1,625	1.9%	28.9%
Other	4,000	5,625	71.1%	100.0%

Note: the frequencies were calculated with the data set at the level of combinations candidate-economic activities for the period until July 2021. Consequently, there might exist more than one economic activity associated with a given student. We only considered firms that started up after the candidates' call year

Table B6: Heterogeneous effects on the probability of finding the candidate in any outcome data set using a linear regression

	+1 year	+2 years	+3 years	+4 years	+5 years	+6 years	Pooled +1 to +3 years	Pooled +4 to +6 years	Pooled +1 to +6 years	Pooled post-call period
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Approved	-0.026***	-0.084***	-0.116***	-0.096***	-0.062***	-0.033**	-0.130***	-0.062***	-0.074***	-0.066***
	[0.008]	[0.013]	[0.013]	[0.016]	[0.018]	[0.015]	[0.014]	[0.016]	[0.016]	[0.010]
Entrance exam score	-0.002	-0.01	-0.008	0.02	0.021	0.016	-0.005	0.026	0.009	<0.001
	[0.006]	[0.007]	[0.010]	[0.014]	[0.016]	[0.017]	[0.010]	[0.017]	[0.016]	[0.014]
Approved x Entrance exam score	-0.002	0.012	0.038***	0.025	0.038	0.041*	0.029*	0.04	0.045*	0.052***
	[0.007]	[0.010]	[0.014]	[0.018]	[0.024]	[0.024]	[0.015]	[0.027]	[0.026]	[0.017]
Mean control dep. var	0.05	0.14	0.24	0.31	0.36	0.39	0.27	0.54	0.58	0.65
Obs	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271	14,271
No. clusters	97	97	97	97	97	97	97	97	97	97

Note: this table presents the OLS estimation results of an equation similar to equation 1 without covariates. All regressions include fixed effects for home university, major, admission year, call's year and destination country. The dependent variable in columns 1 to 6 is a binary variable for whether the candidate was found in any outcome data set in the corresponding N-year window where N $\in \{1,2,3,4,5,6\}$. The dependent variable in column 7 is a binary variable for whether the student was found in any outcome data set in any year between the first and third after the call's one, while that in column 8 is the same binary variable but that considers the period between the fourth and sixth year after the application. The dependent variable in column 9 follows the same logic, but considers the period between the first and sixth year after the application. The dependent variable in column 10 is a binary variable for whether the student was found in any outcome data set after the call's year. The pooled post-call period encompasses up to 6 years for those that applied in 2014, 7 years for those that applied in 2013, 8 years for those that applied in 2012 and 9 years for those that applied in 2011. Clustered robust standard errors at the call level are shown in square brackets. 'Mean control dep. var' shows the mean of the dependent variable for not approved candidates. The number of observations displayed in each column is calculated excluding singletons. The explanatory variables are described in detail in Appendix Table E11. *** p<0.01, ** p<0.05, * p<0.1.

Table B7: The effect of CSF on pre-treatment covariates using UFBA data

	Age Age	Metropolitan Region of Salvador	Mother or father with a college degree	Single Single	Financially dependent Financially dependent	Attended vocational track in high school
	(1)	(2)	(3)	(4)	(5)	(6)
Approved	0.181	0.067	0.020	-0.085	0.019	0.007
	[0.532]	[0.114]	[0.092]	[0.083]	[0.072]	[0.0445]
Mean dep. var	18.92	0.626	0.198	0.847	0.545	0.072
Obs	1,566	1,566	1,566	1,566	1,566	1,566
No. clusters	80	80	80	80	80	80

Note: this table presents the 2SLS estimation results of equation 1 for placebo outcomes. All regressions include fixed effects for home university, major, admission year, call's year and destination country, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at home university. Clustered robust standard errors at the call level are shown in square brackets. 'Mean control dep. var' shows the mean of the dependent variable for not approved candidates. The number of observations displayed in each column is calculated excluding singletons. *** p<0.01, ** p<0.05, * p<0.1.

Table B8: List of calls for destination countries that had at least two different calls in the same year

Year	Dest. country	Call no.	No. approved	No. applicants	Approval rate	Disc. approval rate	ENEM's threshold score
2011	USA	101/2011	930	7,997	11.6%	10.3%	241
		102/2011	864	16,256	5.3%	3.4%	-
2012	Australia	112/2012	611	1,560	39.2%	39.1%	527
		119/2012	713	1,176	60.6%	59.9%	-
		125/2012	35	117	29.9%	30.0%	-
	Belgium	110/2012	28	163	17.2%	17.2%	686
		111/2012	30	183	16.4%	17.7%	623
	Canada	108/2012	179	1,022	17.5%	17.0%	393
		109/2012	765	1,350	56.7%	56.4%	465
		120/2012	1,538	2,564	60.0%	60.2%	420
		124/2012	67	447	15.0%	15.4%	-
	Netherlands	116/2012	366	1,057	34.6%	35.4%	574
		122/2012	373	661	56.4%	61.0%	-
	Portugal	113/2012	1,541	12,126	12.7%	12.8%	679
		127/2012	8,215	28,191	29.1%	25.5%	-
	South Korea	114/2012	173	464	37.3%	36.4%	568
		121/2012	132	252	52.4%	54.1%	-
	Spain	115/2012	1,678	9,918	16.9%	17.1%	269
		126/2012	443	1,524	29.1%	28.5%	462
	USA	117/2012	1,565	4,272	36.6%	33.8%	-
		131/2012	120	616	19.5%	18.4%	-
		132/2012	158	748	21.1%	18.3%	536
2013	Australia	148/2013	614	1,877	32.7%	31.6%	646
		153/2013	193	1,208	16.0%	15.6%	497
		167/2013	682	1,306	52.2%	53.0%	546
		172/2013	405	1,117	36.3%	35.4%	628
	Austria	139/2013	14	135	10.4%	7.6%	613
		166/2013	15	64	23.4%	20.3%	551
	Belgium	140/2013	72	421	17.1%	16.3%	629
		141/2013	11	177	6.2%	5.7%	-
		175/2013	23	160	14.4%	15.6%	-
		176/2013	42	113	37.2%	38.1%	-
	Canada	147/2013	716	1,897	37.7%	36.3%	-
		149/2013	608	2,431	25.0%	25.5%	581
		152/2013	49	705	7.0%	7.1%	636
		168/2013	667	1,259	53.0%	52.8%	585
		171/2013	72	562	12.8%	12.1%	-
	China	136/2013	226	664	34.0%	30.7%	243
		163/2013	80	383	20.9%	18.8%	600

(continues)

Table B8: List of calls associated with destination countries that have at least two different calls in the same year (cont.).

Year	Dest. country	Call no.	No. approved	No. applicants	Approval rate	Disc. approval rate	ENEM's threshold score
2013 (cont.)	Finland	142/2013	58	304	19.1%	18.7%	672
		154/2013	20	359	5.6%	5.4%	-
		173/2013	32	136	23.5%	22.0%	-
	Germany	144/2013	959	1,827	52.5%	49.5%	-
		157/2013	1,474	2,388	61.7%	58.7%	-
	Hungary	146/2013	1,443	3,097	46.6%	43.6%	545
		164/2013	337	2,540	13.3%	12.7%	603
	Ireland	138/2013	532	1,739	30.6%	25.7%	236
		162/2013	983	2,200	44.7%	40.9%	288
	Japan	145/2013	220	532	41.4%	37.8%	601
		165/2013	135	751	18.0%	16.4%	600
	New Zeland	155/2013	64	1,155	5.5%	5.3%	667
		174/2013	97	368	26.4%	25.6%	662
	South Korea	150/2013	56	219	25.6%	24.6%	-
		169/2013	78	166	47.0%	48.3%	635
	UK	151/2013	1,864	4,775	39.0%	39.1%	403
		170/2013	2,659	4,263	62.4%	61.4%	330
	USA	143/2013	7,386	17,634	41.9%	37.6%	-
		156/2013	6,874	22,104	31.1%	29.1%	406
2014	Australia	184/2014	1,000	3,062	32.7%	31.9%	664
		185/2014	900	2,220	40.5%	40.8%	676
	Belgium	186/2014	75	210	35.7%	36.8%	658
		187/2014	148	861	17.2%	19.0%	686
	Canada	188/2014	539	1,618	33.3%	33.4%	676
		189/2014	108	741	14.6%	14.3%	661
		204/2014	11	403	2.7%	2.5%	620
	USA	180/2014	4,828	16,197	29.8%	28.7%	-
		196/2014	51	1,083	4.7%	4.4%	602

Table B9: Different specifications for the first stage regression

Dependent variable: Approved								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Single-call candidates								
Ratio	1.043***	0.959***	0.720***	1.085***	1.206***	0.669***	1.284***	1.386***
	[0.226]	[0.221]	[0.174]	[0.238]	[0.183]	[0.155]	[0.234]	[0.463]
	(3.528)	(3.245)	(3.022)	(3.474)	(5.017)	(3.238)	(4.563)	(1.606)
Entrance exam score		0.223***	0.236***	0.248***	0.215***	0.274***	0.253***	0.234***
		[0.033]	[0.023]	[0.020]	[0.018]	[0.023]	[0.020]	[0.037]
Admission year (cohort) FE	No	No	Yes	Yes	Yes	No	No	No
Major FE	No	No	Yes	Yes	Yes	No	No	No
Home university (HEI) FE	No	No	Yes	Yes	Yes	No	No	No
Dest. country FE	No	No	No	Yes	Yes	No	No	No
Call year FE	No	No	No	No	Yes	No	No	No
Cohort-major-HEI FE	No	No	No	No	No	Yes	Yes	No
Dest. country-call year FE	No	No	No	No	No	No	Yes	No
Cohort-major-HEI-dest. country-call year FE	No	No	No	No	No	No	No	Yes
Obs	16,151	14,272	14,271	14,271	14,271	14,215	14,215	9,489
R2	0.09	0.08	0.05	0.06	0.06	0.04	0.04	0.04
No. clusters	97	97	97	97	97	97	97	82
F-stat of Instrument	21.24	18.85	17.18	20.81	43.21	18.65	30.08	8.97
Panel B. First-call of all candidates								
Ratio	0.881***	0.814***	0.644***	0.998***	1.183***	0.612***	1.290***	1.510***
	[0.224]	[0.220]	[0.178]	[0.237]	[0.208]	[0.160]	[0.253]	[0.482]
	(2.729)	(2.443)	(2.338)	(2.743)	(4.829)	(2.595)	(4.122)	(1.771)
Entrance exam score		0.186***	0.200***	0.213***	0.186***	0.229***	0.218***	0.205***
		[0.034]	[0.025]	[0.021]	[0.018]	[0.027]	[0.021]	[0.038]
Admission year (cohort) FE	No	No	Yes	Yes	Yes	No	No	No
Major FE	No	No	Yes	Yes	Yes	No	No	No
Home university (HEI) FE	No	No	Yes	Yes	Yes	No	No	No
Dest. country FE	No	No	No	Yes	Yes	No	No	No
Call year FE	No	No	No	No	Yes	No	No	No
Cohort-major-HEI FE	No	No	No	No	No	Yes	Yes	No
Dest. country-call year FE	No	No	No	No	No	No	Yes	No
Cohort-major-HEI-dest. country-call year FE	No	No	No	No	No	No	No	Yes
Obs	19,245	17,007	17,007	17,007	17,007	17,007	17,007	11,845
R2	0.07	0.06	0.04	0.06	0.06	0.04	0.05	0.05
No. clusters	98	98	98	98	98	98	98	85
F-stat of Instrument	15.49	13.67	13.02	17.76	32.49	14.62	26.09	9.82

Note: all regressions include fixed effects for home university, major, admission year, call's year and destination country, and control for gender and whether the candidate has ever enrolled in more than one major at home university. In panel B, we also control for the number of CSF calls the candidate applied for. Clustered robust standard errors at the call level are shown in square brackets. In parentheses, we present the 0.05 tF statistic from Lee et al. (2022). The number of observations displayed in each column is calculated excluding singletons. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B10: Main results considering an alternative set of fixed effects

	+1 year	+2 years	+3 years	+4 years	+5 years	+6 years	Pooled +1 to +3 years	Pooled +4 to +6 years	Pooled +7 to +8 years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(4)
Panel A. Postgraduate education enrolment									
Approved	-0.024*	-0.077***	-0.113***	-0.068	-0.031	0.007	-0.120***	-0.054	0.076
	[0.014]	[0.025]	[0.043]	[0.048]	[0.039]	[0.032]	[0.043]	[0.033]	[0.067]
Mean dep. var	0.02	0.07	0.20	0.15	0.27	0.17	0.15	0.27	0.17
Obs	13,575	13,575	13,575	13,575	13,575	13,575	13,575	13,575	13,575
No. clusters	93	93	93	93	93	93	93	93	63
Panel B. Formal employability									
Approved	-0.015*	-0.058***	-0.040**	-0.055***	-0.005	-0.051*	-0.054***	-0.032	-0.224**
	[0.008]	[0.015]	[0.018]	[0.017]	[0.018]	[0.001]	[0.019]	[0.031]	[0.090]
Mean dep. var	0.02	0.06	0.09	0.10	0.13	0.00	0.10	0.24	0.23
Obs	13,575	13,575	13,575	13,575	13,575	13,575	13,575	13,575	
No. clusters	93	93	93	93	93	93	93	93	63
Panel C. Firm ownership									
Approved	-0.008	-0.020**	0.000	-0.010	-0.038**	0.013	-0.026	-0.035	-0.080
	[0.005]	[0.008]	[0.011]	[0.012]	[0.017]	[0.020]	[0.017]	[0.024]	[0.050]
Mean dep. var	0.01	0.02	0.02	0.04	0.05	0.05	0.05	0.12	0.10
Obs	13,575	13,575	13,575	13,575	13,575	13,575	13,575	13,575	13,575
No. clusters	93	93	93	93	93	93	93	93	63

Note: this table presents the 2SLS estimation results of equation 1. The dependent variable in columns 1 to 6 is a binary variable for whether the student became a sole-proprietorship firm owner or entered an existing society as a business partner in the corresponding N-year window where N $\in \{1,2,3,4,5,6\}$. The dependent variable in column 7 is a binary variable for whether the student became a sole-proprietorship firm owner or entered an existing society as a business partner in any year between the first and third after the call's one, while that in column 8 is the same binary variable but that considers the period between the fourth and sixth year after the application. The dependent variable in column 9 follows the same logic, but considers the period between the seventh and eighth year after the application. All regressions include fixed effects for home university, admission year and university major-destination country-call year combinations, and control for gender, normalized entrance exam score, and whether the candidate has ever enrolled in more than one major at home university. Clustered robust standard errors at the call level are shown in square brackets. Data for formal entrepreneurial activity are available until July 2021. 'Mean control dep. var' shows the mean of the dependent variable for not approved candidates. The number of observations displayed in each column is calculated excluding singletons. *** p$<$0.01, ** p$<$0.05, * p$<$0.1.

C Detailed description of the Science without Borders programme

The CSF programme was designed as a project composed of three broad arms, each of them with a different goal: i) to complement the education of Brazilian students through international experience in foreign universities, ii) encourage young Brazilian talented graduates to return to their country of origin and stimulate their retention in the local labour market, and iii) attract renowned foreign researchers to carry out their activities in partnership with the Brazilian academic community.

For achieving these goals, the programme created seven grant modalities. Five of them were related to the first arm of programme, while the other ones were left with a single modality each. The five modalities linked to the first arm were: (i) sandwich undergraduate, (ii) sandwich doctorate, (iii) postdoctorate, (iv) full doctorate and (V) full master's. The second arm was implemented through scholarships for young talents and the third was operationalized through scholarships for special visiting researchers. Out of the three arms, the most prominent one was the first, which awarded scholarships for both undergraduate and graduate students to study abroad for one (sandwich undergraduate, post-doctorate and sandwich doctorate), two (full master's degree) or four years (full doctoral degree). Out of 92,880 scholarships granted under CSF by 2016, 91,601 were dedicated to the first arm, 504 to the second and 775 to the third one (Chaves and Castro 2016).

The high concentration of grants on the undergraduate students was the most remarkable features of the programme. Administrative data show that 73,353 (79%) scholarships were awarded for the sandwich undergraduate modality, 9,685 (10.4%) for the full doctorate, 4,652 (5%) for the post-doctorate, 3,353 (3.6%) for the sandwich doctorate and, finally, 558 (0.6%) for full master's (Chaves and Castro 2016; McManus and Nobre 2017). In particular, this scholarship modality was much larger than previous programmes, although it was not the first focused on undergraduates[20]. The programme BRAFITEC (*Brasil-France Ingénieurs Technologie*), for instance, which was created in 2002 and that stands out for its longevity, has awarded 5,220 scholarships for Brazilian engineering students to attend French universities by 2016 (Grochocki and Guimarães 2017).

To put into perspective, between 1987 and 2000, CAPES offered a total of 6,089 study-abroad scholarships including all grant modalities and CNPq, between 1986 and 1999, financed a total of 7,730 study-abroad scholarships (Mazza 2009), while CSF granted, for the same targeted population, more than five times the sum of these amounts from CAPES and CNPq in just a few years. The undergraduate sandwich modality was the most celebrated one also because it stood as a very singular opportunity for young people from all socioeconomic backgrounds to study abroad with public resources in a fully funded fashion. As demonstrated by Castro et al. (2012), the trend among Brazilian research funding agencies between 1997 and 2009 was a gradual increase in the number of fellowships, with a significant reduction in the number of full PhD fellowships (about 1,300 per year between 2001 and 2002, and 800 between 2007 and 2009), the elimination of master's fellowships, and an important increase in the number of sandwich PhD fellowships.

The focus on the STEM fields, at the expense of non-STEM areas, whose students were not eligible for the CSF, was defined considering the Brazilian priorities set by the programme's creators, namely the Brazilian Ministry of Education and the Brazilian Ministry of Science and Technology. In 2012, Brazil had the lowest percentage of STEM degrees among 40 countries (OECD 2015). Graduation rates in Engineering degrees were also very low: 45% considering several cohorts from 2003 to 2011 in both public and private higher education institutions (CNI 2015). Historically, Brazil's undergraduates

[20] The first government-sponsored study abroad programme for undergraduates was launched in 1997 by CAPES and enabled, jointly with similar initiatives, the international mobility of 710 Brazilian engineering students between 1997 and 2001 (Grochocki and Guimarães 2017).

were concentrated on the majors of Administration, Law and Education (Gusso and Nascimento 2014; OECD 2019), which is also true at the postgraduate level. Data from 2011 show that the main area of CAPES scholarships for full doctorate was the Social Sciences such as Arts, Law, Psychology and History (McManus and Nobre 2017). As a whole, the tertiary education in Brazil used to produce a reduced number of professionals in STEM fields before the CSF programme, especially in Engineering.

D Federal universities

The set of Brazilian universities from our sample is very heterogeneous both in terms of spatial distribution and prominence. The next subsections provide some details about each universities we partnered with. The information about enrolled students and entrants per year was extracted from the Higher Education Census 2017, administered by INEP[21].

D1 UFBA

The Federal University of Bahia (UFBA) is considered the largest and most influential university in the state of Bahia and one of the most important in the country, especially in the Northeast. It has more than 36 thousand enrolled students and 7.1 thousand entrants per year. The campuses are located in the capital, Salvador, and three municipalities of the countryside: Vitória da Conquista, Camaçari and Barreiras. Bahia also has other five federal universities: Federal University of the South of Bahia (UFSB), Federal University of Recôncavo da Bahia (UFRB), Federal University of Western Bahia (UFOB), Federal University of Afro-Brazilian Lusophony (UNILAB) and Federal University of the São Francisco Valley (UNIVASF)[22].

D2 UFPE

The Federal University of Pernambuco (UFPE) is the largest public university in the state of Pernambuco. It has more than 32 thousand enrolled students and 7.4 thousand entrants per year. The campuses are located in the capital, Recife, and two municipalities of the countryside: Caruaru and Vitória de Santo Antão. Pernambuco also has other three federal universities: Federal University of Agreste de Pernambuco (UFAPE), Federal Rural University of Pernambuco (UFRB) and Federal University of the São Francisco Valley (UNIVASF).

D3 UFPB

The Federal University of Paraíba (UFPB) is the largest public university in the state of Paraíba. It has more than 29 thousand enrolled students and 7.7 thousand entrants per year. The campuses are located in the capital, João Pessoa, and four municipalities of the countryside: Areia, Banananeiras, Rio Tinto and Mamanguape. Paraíba also has another federal university: the Federal University of Campina Grande (UFCG).

D4 UFCG

The Federal University of Campina Grande (UFPB) has more than 15 thousand enrolled students and 4.2 thousand entrants per year. The campuses are located in the countryside, one in the second-largest city of Paraíba, Campina Grande, and other six municipalities: Pombal, Patos, Sousa, Cajazeiras, Cuité and Sumé.

D5 UFS

The Federal University of Sergipe (UFS) is the unique federal university in the state of Sergipe. It has more than 26 thousand enrolled students and 4.8 thousand entrants per year. The campuses are located

[21] The National Institute for Educational Studies and Research Anísio Teixeira (INEP) is a special research agency linked to the Ministry of Education and which is responsible for assessing the quality of postgraduate education programmes in Brazil.

[22] UNIVASF is a multi-state federal university that also has campuses in municipalities located in the states of Pernambuco and Piauí.

in the capital, Aracaju, and four municipalities of the countryside: Laranjeiras, Lagarto, Itabaiana and Nossa Senhora da Glória.

D6 UFES

The Federal University of Espirito Santo (UFS) is the unique federal university in the state of Espírito Santo. It has more than 24 thousand enrolled students and 4.6 thousand entrants per year. The campuses are located in the capital, Vitória, and two municipalities of the countryside: Alegre and São Mateus.

D7 UFV

The Federal University of Viçosa (UFV) is the fourth-largest federal university in the state of Minas Gerais. It has more than 13 thousand enrolled students and 3.8 thousand entrants per year. The campuses are located in three municipalities of the countryside: Viçosa, Rio Paranaíba and Florestal. The other ten federal universities in the state are: Federal University of Minas Gerais (UFMG), Federal University of Uberlândia (UFU), Federal University of Juiz de Fora (UFJF), Federal University of São João Del Rei (UFSJ), Federal University of Ouro Preto (UFOP), Federal University of Lavras (UFL), Federal University of Itajúba (UNIFEI), Federal University of Jequitinhonha and Mucuri Valleys (UFVJM), Federal University of Triângulo Mineiro (UFTM) and Federal University of Alfenas (UNIFAL).

D8 UFSCar

The Federal University of São Carlos (UFSCar) is the largest federal university in the state of São Paulo. It has 13.1 thousand enrolled students and 2.7 thousand entrants per year. The campuses are located in four municipalities of the countryside: São Carlos, Araras, Botucatu and Buri. São Paulo also has two other federal universities: the Federal University of ABC (UFABC) and the Federal University of São Paulo (UNIFESP), which are similar in the number of students: 12.7 thousand and 11.2 thousand, respectively.

D9 UFSC

The Federal University of Santa Catarina (UFSC) is the largest public university in the state of Santa Catarina. It has more than 30 thousand enrolled students and 7.9 thousand entrants per year. The campuses are located in the capital, Florianópolis, and four municipalities of the countryside: Blumenau, Araranguá, Joinville and Curitibanos. Santa Catarina also has campuses of a multi-state federal university called Federal University of the South Frontier (UFFS), which has campuses in municipalities of two other states: Rio Grande do Sul and Paraná.

D10 UFMT

The Federal University of Mato Grosso (UFMT) is the largest public university in the state of Mato Grosso. It has more than 21 thousand enrolled students and 6.1 thousand entrants per year. The campuses are located in the capital, Cuiabá, and four municipalities of the countryside: Barra do Garças, Pontal do Araguaia, Sinop and Várzea Grande. Mato Grosso also has another federal university: the Federal University of Rondonópolis, which is much smaller: 5.3 thousand students.

D11 UFGD

The Federal University of Grande Dourados (UFGD) is the largest federal university in the state of Mato Grosso do Sul. It has more than 20 thousand enrolled students and 8.1 thousand entrants per year. There are two campuses in the same city, located in the countryside: Dourados. Mato Grosso do Sul also has

another federal university: the Federal University of Mato Grosso do Sul, which is similar in size: 18.9 thousand students.

D12 UFPA

The Federal University of Pará (UFPA) is the largest federal university in the state of Pará. It has more than 38 thousand enrolled students and 7.7 thousand entrants per year. The campuses are located in the capital, Belém, and eleven municipalities of the countryside: Abaetetuba, Altamira, Ananindeua, Bragança, Breves, Cametá, Capanema, Castanhal, Salinópolis, Soure and Tucuruí. Pará also has three other federal university: Federal Rural University of Amazônia (UFRA), Federal University of South and Southeast Pará (UFESSPA) and Federal University of Western Pará (UFOPA).

D13 UFRA

The Federal Rural University of Amazônia (UFRA) is the second-largest federal university in the state of Pará. It has more than 6 thousand enrolled students and 1.5 thousand entrants per year. The campuses are located in the capital and four municipalities of the countryside: Capanema, Capitão Poço, Paragominas, Parauapebas and Tomé-Açu.

E Data preparation and variables definition

To handle with missings in the universities academic records related to the variable for gender, we took advantage of a toolkit called genderBR[23], written in R language, that predicts the gender of the individual based on its first name[24]. Using the genderBR, we were able to determine the gender of all applicants in our sample.

To create a cross-university normalized entrance examination score, we set to one the score of the highest-performing student of a given home university, major, and admission year, and to zero the score of the lowest-performing student in the same group (i.e., home university, major and admission year). The students with a median performance were rated with a score between (0,1), given by the following formula:

$$\text{normalized score} = \frac{\text{score} - \text{minimum score}}{\text{maximum score} - \text{minimum score}} \tag{E1}$$

where *score* is the entrance examination score of a given student, *minimum score* is the score of the lowest-performing student, and *maximum score* is that of the highest-performing student.

[23] https://github.com/meirelesff/genderBR

[24] The genderBR package was constructed based on the Brazilian Population Census 2010 and uses data on the number of females and males with the same name in Brazil, or in a given Brazilian state, and calculates the proportion of female's uses of it. The function then classifies a name as male or female only when that proportion is higher than a given threshold (e.g., female if proportion > 0.9, or male if proportion $<= 0.1$); proportions below these thresholds are classified as missing.

Table E11: Variables definition and sources

Variable	Definition	Source
Approved	Dummy variable = 1 if the student was approved in the CSF call for which he applied, and 0 otherwise	CNPq/CAPES
Entrance exam score	Normalized score used for home university admission. We normalized the entrance examination score so that the score lies in the interval between 0 and 1. The normalization was undertaken within both university, major and admission year and is such that the lowest score is set to 0 and the highest one is set to 1.	CNPq/CAPES
Ratio	CSF call's approval rate excluding applicants from the home universities in our sample given by the ratio of the total number of approved candidates from other home universities to the total number of applicants from other home universities (not in our sample)	CNPq/CAPES
Ratio top 25th pctle.	Dummy variable = 1 if the CSF call's approval rate excluding applicants from the home universities in our sample is among the 25% highest ones in the universe of calls in our sample and 0 otherwise	CNPq/CAPES
ENEM's threshold score	ENEM score of the last approved candidate of a given call	CNPq/CAPES
Male	Dummy variable = 1 if the student is a male, and 0 otherwise	Universities' records
Graduated	Dummy variable = 1 if the student has graduated in his major at UFBA considering data until the second semester of 2021	Universities' records
On-time graduation	Dummy variable = 1 if the student has graduated in his major at UFBA in the period expected by the major curriculum considering data until the second semester of 2021	Universities' records
Engineering	Dummy variable = 1 if the student's major at home university is in the field of Engineering, and 0 otherwise	Universities' records
Health Sciences	Dummy variable = 1 if the student's major at home university is considered related to Health Sciences, and 0 otherwise. The list is composed of the following majors: Medicine, Biomedicine, Pharmacy, Nursing, Physiotherapy, Speech Therapy, Gerontology, Nutrition, Psychology, Dentistry, Collective Health and Occupational Therapy	Universities' records
Other majors	Dummy variable = 1 if the student's major at home university is neither from the field of Engineering nor from the Health Sciences, and 0 otherwise	Universities records
Formal employed	Dummy variable = 1 if the student was in the formal labour market for at least one month during the corresponding period, and 0 otherwise	RAIS
Hourly wage	Hourly wage in Brazilian reais (BRL) of the job with the highest hourly wage in the corresponding period	RAIS
Monthly wage	Monthly wage in Brazilian reais (BRL) of the job with the highest hourly wage in the corresponding period	RAIS
ln(Hourly wage +1)	Natural logarithm of the hourly wage of the job with the highest hourly wage in the corresponding period	RAIS
Technical occupation	Dummy variable = 1 if the student was in the formal labour market in an occupation considered technical according to the classification proposed by Araújo et al. (2009) for at least one month during the corresponding period, and 0 otherwise	RAIS
Open-ended contract	Dummy variable = 1 if the student was in the formal labour market under an open-ended contract (i.e., RAIS codes 10 or 20) for at least one month during the corresponding period, and 0 otherwise	RAIS
Job tenure (in months)	Number of months in the same job for the longest-lasting job in the corresponding period	RAIS

(continues)

Table E11: Variables definition and sources (cont.).

Variable	Definition	Source
Public contract	Dummy variable = 1 if the student was in the formal labour market under the public sector regime called 'estatutário' (i.e., RAIS codes 30 or 31) for at least one month during the corresponding period, and 0 otherwise	RAIS
Public institution	Dummy variable = 1 if the student was in the formal labour market in a public institution (i.e., RAIS codes 1015, 1023, 1031, 1040, 1058, 1066, 1074, 1082, 1104, 1112, 1120, 1139, 1147, 1155, 1163, 1171, 1180, 1198, 1201, 1210, 1228, 1236, 1244, 1252, 1260, 1279, 2011 or 2038) for at least one month during the corresponding period, and 0 otherwise	RAIS
Firm ownership	Dummy variable = 1 if the student started up a sole-proprietorship formal firm or entered an existing society as a business partner in the corresponding period, and 0 otherwise	RFB
Postgrad. student	Dummy variable = 1 if the student was enrolled in a Brazilian postgraduate programme in the corresponding period, and 0 otherwise	SUCUPIRA

F Record linkage preparation

In this appendix, we explain the procedures used to prepare the outcome data sets and implement the probabilistic record linkages.

F1 RAIS

We began with eight different RAIS data sets, one for each year between 2013 and 2020 with information about all formal workers registered in Brazil. If we had access to the 11-digit CPF of each CSF candidate, there would be no major issue in finding them in RAIS using this information. However, our data only contains the full name and some intermediary digits of CPF for every CSF applicant that was a regular student of the universities we partnered with.

The first challenge to implement the matching procedure is to unify the full name of every worker listed in RAIS, as there could exist two reasons for duplicates. The first occurs when we observe the same individual at least two times in the same year. In each occurrence, the individual has the same CPF, but different names. The only reason for a given individual having more than one observation in a matched employer-employee data set in the same year is having more than one formal job. Thus, in the case of an individual who worked for only one firm in a given year, we expect only a single observation in RAIS in the corresponding year (and, consequently, a unique full name associated with a given set of CPF digits). The second problem happens when we observe different name spellings for the same CPF in different RAIS years. This might happen because the same employer can mistype or shorten the name of the same worker in a given year of RAIS, and not in other years, or because the individual changed jobs across RAIS years and the new employer registered his name with a different spelling.

Both sources of errors stem from the fact that employers are responsible for filling the RAIS information, including typing the full name of every worker. Importantly, there are no checks for spelling mistakes or any other alike on the part of the Ministry of Economy, which administers RAIS data. As a consequence, there could exist multiple unique names of workers with the same 11-digit CPF that differ by a few letters, which is typical of misspellings. For illustrative purposes, Table F12 shows a fictitious yet representative case of the first type of variation relative to RAIS 2018. As can bee seen, it is reasonable to assume that all observations refer to the same individual as the first name is clearly the same ('ANTONIO CARLOS') and the surname varies but typically contains most of the same letters that appear in other observations[25].

Table F12: Misspelling cases

Obs. number	Worker's full name	11-digit CPF
1	ANTONIO CARLOS DE SOUZA MATOS	02220392050
2	ANTONIO CARLOS DE SOUZA	02220392050
3	ANTO NIO CARLOS BASTOS	02220392050
4	ANTONIO CARLOS DE SOUSA BASTOS	02220392050
5	ANTONIO CARLOS DE SOUZA BASRTOS	02220392050
6	ANTONIO CARLOS DE SOUZA BASTOS	02220392050

Note: both the worker's full name and 11-digit CPF are fictitious. The true case that inspired this was found in the data set relative to RAIS 2018.

[25] Brazilian names traditionally have two surnames, with the first referring to the mother's paternal and the second to the father's paternal one. Out of the six observations, we observe either 'SOUZA' or 'SOUSA' in five of them. Replacing 'Z' with 'S' is a very common mistake when people type Brazilian names. This means that 'SOUZA' or 'SOUSA' is very likely to be one of the surnames. 'MATOS' or 'BASTOS' also seems to be another surname, as these words appear in five observations. In this case, we need to replace two letters to turn one word into another (e.g., replace 'M' with 'B' and include a 'S' between 'A' and 'T').

We followed a multi-step procedure to create a pooled RAIS data set containing all the unique combinations of workers' full names and intermediary digits of CPF considering observations for the period 2013-2020[26]. This pooled database is composed of 354,706,251 observations. The most critical step of the procedure was the one intended to deal with the first type of variation, i.e., within year. For each Brazilian-level RAIS database referring to a given year between 2013 and 2020, we proceeded as follows.

We saved two separate data sets: one for observations such that the 11-digit CPFs in the corresponding year were associated with a single worker's full name, and another for those 11-digit CPFs that were associated with more than one full name. The latter is the one for which we propose the first procedure. For approximately 98% (average) of the CPFs, there were only two distinct full names. For each CPF of this separate data set, we then created two variables: (i) the longest and (ii) the shortest worker's full name associated with that 11-digit CPF. At this point, our goal was to define whether a set of observations with the same CPF did refer to the same person. To accomplish this goal, we took advantage of a matching technique[27] that calculates a similarity score between the longest full name and shortest full name of each 11-digit CPF. The algorithm returns a numeric variable that ranges from 0 to 1. A similarity score of 1 implies a perfect similarity according to the string matching algorithm that we used (bigram) and decreases when the match is less similar (Raffo 2020).

Once we had the similarity scores for each set of full names linked to a given CPF, we defined rules to determine whether those observations referred to the same individual or not. We removed all CPFs that, according to our criteria (described in the next paragraph), referred to more than one individual within the same RAIS year. The motivation for that is two-fold: first because they are very likely cases of mistyped CPFs, and second because we wanted to have a single full name associated with each 11-digit CPF.

Although the removal reduces the likelihood of finding some CSF candidates in RAIS, it represents a small fraction of the observations corresponding to each year: 0.01%, on average. In RAIS 2018, for example, there were 55,648,275 distinct 11-digit CPFs, out of which we removed 10,498 (0.0189%). For the CPFs that were linked to more than one worker's full name that we considered as being related to the same person, we defined the longest full name as the single full name associated to it[28]. The idea was that the longest full name in a given RAIS year is more likely to be the true full name of the worker[29].

The criteria were defined using thresholds for the similarity scores. Two specific situations for a given 11-digit CPF were considered as not referring to the same person: i) similarity score in the range [0, 0.6], and ii) similarity score in the range (0.6, 0.9) such that the first letter of at least two full names associated to that CPF were different and the similarity score of the first name was in the range [0, 0.6].

---

[26] For each year of RAIS, we removed all observations for which the 11-digit CPF was composed only of zeros (there are no missings in the variable for CPF in RAIS data). These cases represent a tiny percentage of the observations ($< 0.001\%$). Just for reference, in RAIS 2018, we found 1,389 cases out of more than 55 million observations.

[27] We use the *matchit* Stata module with the bigram algorithm, which decomposes text strings into elements of two characters (grams) in a moving-window fashion. In particular, its logic differs from both phonetic algorithms such as Soundex and edit-distance ones such as Levenshtein.

[28] The tiebreak rule for the cases when the distinct full names had the same length was to randomly select one of them to be the single full name associated to that CPF.

[29] The longest full name is not necessarily the closest to the true full name, as can be seen in the example of Table F12 ('ANTONIO CARLOS DE SOUZA BASRTOS', which corresponds to observation number 5, is the longest one, but probably the true full name is 'ANTONIO CARLOS DE SOUZA BASTOS', observation number 6). Nonetheless, given that the bigram algorithm considers whether the text strings being compared contain the characters of one another, we decided to be conservative by considering the probably most comprehensive set of characters that stem from the longest full name.

As a complementary step to enhance the quality of the procedure, we calculated similarity scores for the first name[30] of both the longest and shortest full names associated with each CPF of the separate data set. In addition, we checked whether the first character of any full names associated with a given CPF was ever different. These two additional checks were used to define the situation (ii) in which we consider the 11-digit CPFs to be excluded from the analysis.

We then conclude the most critical step of the procedure that deals with within-year variation. In the case shown in Table F12, after the step, we are left with a single combination of full name (i.e, the longest one, 'ANTONIO CARLOS DE SOUZA BASRTOS') and some intermediary digits of CPF (extracted from the 11-digit CPF informed by the employer) for RAIS 2018[31]. For each RAIS year, we replicated this process so that we could create the RAIS 2013-2020 database containing the unique combinations of full names and intermediary digits of CPF and the corresponding RAIS year.

F2 Matching RAIS with cross-university data sets using a probabilistic record linkage

The next and most important step was to combine the cross-university data, in which each CSF candidate corresponds to an unique row, with the panel RAIS 2013-2020 containing all the unique combinations of workers' full names and their intermediary digits of CPF year-by-year. To do that, we used a different matching algorithm[32] and kept only the closest observation at RAIS database for each CSF candidate. The procedure generates matches whose score varies from 0 to 1 (exact match). To ensure the quality of the matches, we set a threshold for the similarity score of 0.95 as a minimum to accept a given match as possibly referring to the same person. This was used to reduce the pool of matches so that we could perform a case-by-case eyeballing of all matches in the data set that resulted from this record linkage. This means that the across-years variation of workers' full names is mitigated by both by the fact that the matching algorithm only keeps the closest match and for which the similarity score is greater or equal to 0.95 and that we carried out a visual inspection of all pairs produced by the algorithm.

Finally, we present a hypothetical example based on Table F12 to illustrate how this last step worked. However, it is important to highlight that this step was not needed for exact matches. Suppose that, on one hand, we had the full name 'ANTONIO CARLOS DE SOUZA' and the 5-digit CPF 20392 for a given CSF candidate in the cross-university database and, on the other hand, three distinct full names with the same 5-digit CPF, 20392, one for each RAIS year, all of them from the pooled RAIS 2013-2020: 'ANTONIO CARLOS BASTOS DE SOUSA' for RAIS 2015, 'ANTONIO BASTOS' for RAIS 2016 and 'ANTONIO CARLOS DE SOUZA BASRTOS' for RAIS 2018. The similarity scores between the candidate's full name from the cross-university database and each of the aforementioned full names from the pooled RAIS 2013-2020 are 0.9799, 0.8285 and 0.9903, respectively. Due to the threshold (0.95), we would discard the third match and, therefore, be left with the first two matches to visually inspect and determine on their appropriateness in terms of referring to the CSF candidate named 'ANTONIO CARLOS DE SOUZA' in the cross-university database.

---

[30] We use blank spaces to split the content of the text strings.

[31] Just as an example, consider the set of observations shown in Table F12, for which the longest full name is 'ANTONIO CARLOS DE SOUZA BASRTOS' and the shortest is 'ANTO NIO CARLOS BASTOS'. For the bigram algorithm, the similarity score between the longest and shortest full names is 0.7591 and 0.7071 for the first names ('ANTONIO' versus 'ANTO'). In particular, the first character of the full names is always the same ('A'). It means that such observations satisfy neither criterion (i) nor criterion (ii) and are therefore preserved in the RAIS data sets that we use to find CSF candidates.

[32] We took advantage of the *reclink4* Stata package (Borusyak and Jaravel 2018) with the option that keeps only closest match (e.g., *npairs(1)* and set a requirement for a exact matching composing of the six intermediary CPF digits (e.g., *required(cpf_digits)*) for each observation from the master data set.

F3 SUCUPIRA

To use the SUCUPIRA data, we pooled the data sets referring to each year between 2013 and 2020 that contain information about postgraduate students in Brazil. Each data set contains the student's full name and, in the case of Brazilians, the six intermediary digits of CPF and, in the case of foreigners, the passport number. Since only Brazilians were eligible for the CSF, we removed all the observations for which only the passport number was available. The data sets have no issues with duplicated full names associated with the same individual because there is an identifier (ID) variable that permits us to precisely tag observations linked to the same person. It means that pooling the data sets from 2013 to 2020 and removing duplicates provides us the unique combinations of students' full name and six intermediary digits of CPF.

F4 RFB

The Brazilian Internal Revenue Service (RFB) makes publicly available three different data sets relative to firms and individuals connected with formal businesses in Brazil. The first gathers information at the establishment level and contain variables such as the economic activity, municipality in which it is located, address and ZIP code, while the second refers to the company level, and contains information about the legal nature and firm size according to a Brazilian classification[33]. Finally, the third is about individuals and other firms that are registered as partners of companies to which the second data sets refer.

We combined these data sets into a new database in which each row refers to a different establishment so that to have information about the economic activity, municipality, address, legal nature, business start date, whether the firm is active, and the list of all owners and partners that are entitled to a share of the profits. When the owner is an individual with a tax identification number (CPF) in Brazil, both the full name and the six intermediary digits of CPF are available. A particularly important feature of the public version of RFB data is that the third data set only contains information about individuals that are formally considered business partners of societies or partnerships. It means that for firms that are not registered as partnerships (or societies), which is the case of firms formed by a single owner (sole proprietorship), there is no information about the legal person (individual) attached to it. More than 60% of the business entities in Brazil are registered under a sole proprietorship regime, and thus this is an important limitation of the public data sets.

To overcome this limitation, we requested RFB a special data extraction to obtain three variables related to the individuals linked to firms registered under the legal nature *Empresário Individual*, CONCLA code 213-5, under which more than 90% of the sole-proprietorship firms in Brazil are registered. Such variables are: i) full name, ii) six intermediary digits of CPF, and iii) the tax identification number (CNPJ) of firms to which the individuals are connected. We then incorporated these variables and reshaped the data set to ensure that each row represented a given individual, as indicated by the person's full name and six intermediary digits of CPF. There are no typos in the full name of the individuals in RFB data, which means that it was not necessary to unify the full name of each individual, as we did in the RAIS data sets.

[33] Every firm in Brazil is registered under a specific legal nature and each of the different types is appropriate for a given purpose. The publicly available information about firm size is limited because it only informs whether the firm is a microenterprise (*Microempresa*, ME), small company (*Empresa de Pequeno Porte*, EPP) or of other type. This classification is important, among other reasons, because MEs and EPPs benefit from favored treatment in federal public procurement, especially in low-value contracts.

G Matching technique

The matching technique we used is based on a modified version of the bigram algorithm[34]. Both the matching technique and the bigram algorithm return a numeric variable called similarity score, which ranges from 0 to 1 and is such that the higher the score, the more likely the strings are similar. We describe the bigram algorithm and its modified version in detail in the following subsection, and then we present the matching technique that takes the similarity score from the modified version of the bigram algorithm as an input to calculate its own similarity score.

In general, the results of the bigram string comparator fare better than alternative string comparators, such those that use phonetic encoding or edit-distance calculations. For a comprehensive review of alternative algorithms, see Christen (2012). The bigram typically produces a higher rate of successful linkages, but also a higher rate of false positive matches, which means that best results can be achieved after a thorough clerical review of linked pairs. Particularly in the case of administrative records for which typing is free of spelling checks such as RAIS, we noted in our data that edit-distance algorithms are not good at producing a similarity score compatible with a one that a human visual inspection would suggest.

G1 Bigram algorithm

For the sake of illustration, consider the two strings: 'ANTONIO' and 'ANTO'. The first step of bigram algorithm is to decompose each string into a set of tokens, called bigrams, each of them composed of two consecutive letters. The first string produces the following set of bigrams: 'AN', 'NT', 'TO', 'ON', 'NI' and 'IO' while the second one generates the tokens 'AN', 'NT' and 'TO'. The similarity score associated with the bigram algorithm is calculated by the following ratio:

$$\text{bigram score } = \frac{\text{Number of bigrams common to the two strings}}{[((L_1-1)+(L_2-1))/2]}$$

where L_1 is the length of the first string and L_2 is the length of the second string. In our example, $L_1 = 7$, $L_2 = 4$ and the number of common bigrams is three (i.e., 'AN', 'NT' and 'TO'). Thus, the similarity score of the bigram is $3/4.5 = 0.666667$.

One modified version of the score associated with the bigram algorithm is called Winkler adjustment, which increases the similarity score related to strings that share the same first four characters. The score resultant from the Winkler adjustment is calculated by the following formula:

$$\text{Winkler score} = \text{bigram score} + J * \frac{(1 - \text{bigram score})}{10}$$

where $J \in \{0,1,2,3,4\}$ is the number of times that a letter from the first string is equal, in the same position, to that from the second string considering only the first four characters of both strings. In particular, the first four characters of 'ANTONIO' have the same letters and in the same position of 'ANTO', which means that in this example $J = 4$. The score is therefore given by $0.666667 + 4 * (1 - 0.666667)/10 = 0.8$.

G2 Similarity score of the matching technique

The similarity score of the matching technique is calculated by the following ratio:

$$\text{similarity score} = \frac{M}{M+N}$$

[34] The matching technique was implemented using Stata package named 'reclink2' (avaiable from SSC) (Blasnik 2010).

where

$$M = \begin{cases} \frac{(\text{Winkler score})^2}{2} & \text{if} \quad \text{Winkler score} \leq \text{minbigram} \\ \frac{(\text{Winkler score})^2}{2} + \frac{0.5\sqrt[3]{\text{Winkler score} - \text{minbigram}}}{\sqrt[3]{1-\text{minbigram}}} & \text{if} \quad \text{Winkler score} > \text{minbigram} \end{cases}$$

and

$$N = \begin{cases} 1 & \text{if} \quad \text{Winkler score} < \text{(minbigram - 0.2)} \\ 1 - (\text{Winkler score})^2 & \text{if} \quad \text{Winkler score} \geq \text{(minbigram - 0.2) and} \\ & \quad \text{Winkler score} < \text{minbigram} \\ (1 - \text{Winkler score})^2 & \text{if} \quad \text{Winkler score} \geq \text{(minbigram - 0.2) and} \\ & \quad \text{Winkler score} \geq \text{minbigram} \end{cases}$$

where minbigram is the minimum threshold that we defeined as acceptable to keep a given match in the data set. We set the minimum threshold to be 0.6, which implies that the similarity score of the matching technique for our example that compares 'ANTONIO' and 'ANTO' is calculated as follows. In this case, the Winkler score (0.8) is greater than the minimum threshold and therefore M is calculated as $\frac{(\text{Winkler score})^2}{2} + \frac{0.5\sqrt[3]{\text{Winkler score} - \text{minbigram}}}{\sqrt[3]{1-\text{minbigram}}} = \frac{(0.8)^2}{2} + \frac{0.5\sqrt[3]{0.8-0.6}}{\sqrt[3]{1-0.6}} = 0.7168$. We also conclude that the Winkler score is greater than the minimum threshold minus 0.2 and, as a consequence, N is calculated as $(1 - \text{Winkler score})^2 = (1-0.8)^2 = 0.04$. We then compute the ratio $\frac{M}{M+N} = \frac{0.7168}{0.7168+0.04}$ to obtain the similarity score: 0.9471. We consider this value as reasonable because'ANTONIO' and 'ANTO' are very similar strings. Just to put into perspective, if we have used the bigram algorithm in its original version, we would end up with a score of 0.6667, which does not seem to properly reflect the similarity score that a human visual inspection would suggest.