Why is the College Premium Falling? The Role of Composition

Tomás Guanziroli[†] Ariadna Jou[‡] Beatriz Rache[§]

July 22, 2024
Link to most recent version

[PRELIMINARY AND INCOMPLETE]

Abstract

Recent studies suggest that the increasing supply of college-educated workers in Latin American countries has negatively impacted returns to skill over the last two decades, as evidenced by the decreasing college premium. In this paper, we show that changes in the college premium do not accurately represent shifts in returns to skill, particularly in the context of a significant expansion in the number of college graduates and institutions. Using novel data with approximately one million college graduates from 20 different cohorts in Brazil, we find that returns to skill have not decreased; in fact, they increased by 24% over 16 years. The supply of college-educated workers has grown, primarily from newer, lower-ranked, and lower-wage-premium universities. Changes in the composition of college workers seem to have driven a decrease in the college premium, even though returns to skill are increasing. Using a simple supply and demand framework (Katz and Murphy, 1992) we show that skill-biased technical change (SBTC) increased by 3% per year over a 12-year period, in contrast to the apparent -0.1% yearly decrease indicated by the unadjusted data.

[†] PUC-Rio; [‡] University of Chile; [§] UCLA

1. Introduction

The college premium in many Latin American countries has flattened or even decreased in the past two decades (Fernández and Messina, 2018). This period coincided with two major structural changes in the higher education sector: (i) a significant increase in the supply of college graduates throughout the region and (ii) the regulation and entry of new private institutions. Due to a lack of data, scholars have primarily focused on the first factor—the expansion in the supply of college graduates —to explain the decrease in the college premium. This has led economists to believe that the demand for skilled labor has been very constrained in developing countries, potentially discouraging both public and private investments in higher education.

In this paper, we study how changes in the composition of college graduates and college institutions have shaped the trajectory of the college premium in Brazil. To do this, we follow multiple cohorts of college graduates from a fixed set of universities in the labor market. By comparing these workers to the overall college population, we decompose price and composition effects. We use these result to estimate a quality-adjusted labor supply and demand model, providing revised estimates of Skill Biased Technical Change (SBTC) in Brazil.

To conduct this analysis, we construct a longitudinal panel dataset by merging college data with labor market data. Initially, we utilize unique data sourced from one million students who graduated from 42 Brazilian universities between 1990 and 2020. These data were obtained through Freedom of Information Law (FOIL) requests and include students' names, their respective universities, majors, and graduation years. Additionally, we leverage the Brazilian employer-employee matched dataset (RAIS) for information on students' wages, age, and other relevant characteristics. We link both datasets using individuals' names from 2002 to 2018 and employ a machine learning algorithm to identify and select the most accurate matches. As a result, we obtain annual wage data pertaining to workers of varying ages and cohorts who graduated from a constant set of universities.

Our main findings indicate that the college premium, when limited to workers from a subset of high-quality universities, has increased. Specifically, we observe a 24% increase in the college wage premium between 2003 and 2018, representing an average annual growth rate of 1.3%. When we broaden our analysis to include the entire sample of college workers—accounting for shifts in institutional composition—we observe a decrease of 16% in the college premium. From this difference, we infer that the apparent decline in the overall college premium is primarily attributed to changes in the university mix.

We provide evidence suggesting that the composition of universities plays a pivotal role in shaping the dynamics of the college premium. Firstly, we demonstrate that students from older and more established institutions exhibit superior performance in end-of-degree standardized exams, implying higher educational quality in these schools. Secondly, we show that the college premium has risen when tracking graduates exclusively from this stable group of universities. Combining these two findings, we infer that a shift in composition is the primary driver behind the flattening of the college premium. Specifically, there is a growing presence of workers holding college degrees from lower-quality institutions, leading to a decrease in the average wages among graduates. Consequently, despite the apparent stabilization of the college premium, the returns to skill are still increasing.

In light of these results, we revise the estimates of skill-biased technical change by incorporating compositional changes into the standard labor supply and demand model (Katz and Murphy, 1992; Card and Lemieux, 2001; Acemoglu and Autor, 2011). In the original model, workers of low and high skill provide one unit of labor and are imperfect substitutes in the production of a homogeneous good. The primary modification we introduce is that the quantity of efficiency units of labor for each skill type can change independently from population changes.[1] We identify changes in the skill premium under the assumption that the wages observed in our fixed sample of universities are not affected by changes in composition or by variations in the number of efficiency units of labor per worker. Notably, this assumption is more likely to hold true than the commonly used (and implicit) assumption that there are no changes in composition within each skill group.

The results from the model estimation show that the last two decades have seen a consistent rise in the relative demand for college-educated workers. The revised estimate for the annual growth in demand for high skill relative to low skill is 3.07%, leading to a positive growth rate of 43.7% over the period of analysis. Notably, the strength of skill-biased technical change is substantially underestimated when changes in the composition of the workforce are not taken into consideration, with estimates predicting a negative annual growth of -0.1%.

We argue that extent of skill-biased technical change may have be underestimated in several other countries around the world. By analyzing harmonized data spanning multiple countries over the past 20 to 40 years, we observe a prevalent pattern: many nations have experienced considerable increases in tertiary enrollment alongside a reduction in educational quality, as approximated by the teacher-student ratio. This negative correlation between enrollment growth and educational quality is particularly pronounced in low and middle-

[1] This is a straightforward modification to the model, as already suggested by Acemoglu and Autor (2011). A similar approach was introduced by Carneiro and Lee (2011).

income countries, where rapid expansion often coincides with declining standards.[2] Such observations suggest that the extent of skill-biased technical change may be underestimated if changes in composition are not adequately considered.

This paper contributes to two strands of literature. First, we give empirical foundation to the "degraded tertiary hypothesis"—i.e., that the average quality of graduate students is decreasing—in explaining the recent changes in the wage structure in Latin America. While this hypothesis has been proposed by several studies, the lack of consensus within the literature is primarily attributed to data limitations (Rodriguez et al., 2016; Camacho et al., 2017; Acosta et al., 2019).[3] Jaume (2021) finds that the Brazilian education expansion led to a reduction in wage gaps between educational groups and other measures of wage inequality. Barros et al. (2010) argue that half of the decline in inequality was driven by an acceleration of educational progress. Ferreira et al. (2017), Alvarez et al. (2018), and Fernández and Messina (2018) contest this assertion, attributing the decrease in earnings inequality to a compression of returns to firm and worker characteristics, such as experience and education. We argue that decomposition of certain measures of inequality often incorrectly attribute decreasing returns to education to a group that actually experienced increasing returns, potentially biasing the results.

Secondly, we contribute to the literature exploring the effects of education expansions on the evolution of the skill price. Building on Katz and Murphy (1992), Card and Lemieux (2001), Acemoglu and Autor (2011), Autor (2014), Blundell et al. (2021), and many others, we are able to account for the effects of school quality on the trends in skill premiums. The conceptual framework of our paper is most similar to the work by Carneiro and Lee (2011). Similar to our case in Brazil, the authors argue that increases in college enrollment lead to a decline in the average quality of U.S. college graduates. However, due to a lack of data, the authors use an empirical strategy that requires strong assumptions about the characteristics of internal migrants in the U.S. In contrast, we have detailed college information for a large sample of college workers, which we observe in a long panel of labor market data. This allows us to identify changes in quality composition with fewer strict assumptions.

The findings of this paper carry significant implications for public policy decisions concerning investments in higher education. Our analysis demonstrates that, despite the extensive expansion in tertiary education over the past two decades, the price of skill has continued to rise. This trend aligns with the concept of skill-biased technological change. Historically,

[2]Interestingly, developed countries present a contrasting scenario, characterized by a positive correlation between enrollment growth and educational quality. In these contexts, the expansion of enrollment is accompanied by an enhancement in educational standards.

[3]For a comprehensive review on the topic, see Messina and Silva (2017)

investments in higher education have yielded increasing returns, a pattern that seems likely to persist. These results align with the findings of Carneiro et al. (2022), who argue that the abundance of skilled workers may prompt firms to adopt skill-complementary technologies. In such settings, both governments and individuals may find it beneficial to invest in higher education.

The structure of this paper is organized as follows: Section 2 introduces the data utilized in our analysis and the institutional setting. Section 3 discusses trends in the college premium, adjusting for changes in composition. In Section 4, we provide evidence suggesting a decline in the skill level of the median college graduate over time. Section 5 elaborates on the supply and demand model, incorporating adjustments for skill composition. Section 6 extends the analysis by presenting evidence of similar phenomena in other countries around the world. Section 7 concludes.

2. Institutional Setting

The Higher Education system in Brazil is characterized by Public and Private universities. Public universities in Brazil have historically been considered as institutions of the highest quality and consist of 13% of institutions (25% of enrolled students in 2017). They provide tuition-free higher education, with admission processes governed by highly selective entrance exams, while private universities charge tuition and offer generally lower quality education. The Brazilian higher education system has undergone significant expansion since the late 1990s. On the supply side, the 1996 Education Law induced an expansion in private universities by reducing the entry requirements for establishing new institutions, thereby leading to a rise in the number and share of private universities (Cox, 2024). Throughout the 2000s, the federal government implemented a series of policies aimed at facilitating access and increasing diversity through student scholarships, subsidized loans, affirmative action programs, and a centralized admissions system known as SISU. In this section, we provide an overview of the institutional setting and summarize the existing empirical evidence on some of these policies.

Until 1996, the number of universities was relatively stagnant, as a federal government agency regulated the sector to ensure quality standards; most consisted of public or private not-for-profit institutions and were centered in large urban centers (Cox, 2024). The 1996 and 1997 Education laws allowed for the first time the operation of for-profit universities, as well as the creation of new courses and majors. This led to an increase in the number of institutions providing higher education, especially private, as presented in Appendix Figure C3.[4]

The Brazilian government continued to foment access to higher education through different initiatives. In 1999 they implemented the higher education credit program (FIES), which provided credit for students to enroll in private undergraduate programs, with the condition they would repay the tuition with subsidized interest rates upon finding employment.[5]. In 2005, government-sponsored scholarships were implemented at private universities for students from lower-income households (PROUNI). The scholarships cover students' tuition completely or partially, depending on the household income. In 2007, there was an expansion of the federal universities together with a restructuring (REUNI) with the goal of increasing the number of seats in public universities. In 2010, the Ministry of Education launched a centralized university entrance system (SISU) to simplify the application process

[4] *Lei de Diretrizes e Bases da Educação Nacional* (*Lei* 9.394) and *Decreto* 2.306

[5] FIES was massively expanded in 2010, with beneficiaries increasing 35-fold from 2009 to 2014, after which the program was restructured and substantially curtailed (Dobbin et al., 2022)

to public universities. Though adherence to SISU was only required from federal and state universities, other universities optionally adhered to it, some offering admission dual application processes (through their own exam, *vestibular*) or through SISU. To continue expanding and diversifying access to higher education in 2012 the Federal law mandating affirmative action was approved (*Lei de Cotas*). With this law, 50% of enrollments per course and shift in the 59 federal universities and 38 federal institutes of education, science, and technology became reserved for students who have completed their entire high school education in public schools.

3. Data

We link the Brazilian employer-employee dataset (RAIS) with information on college graduates from a sample of universities to compute the college premium. Below we describe the data sources and variables, the linkage process, and establish some descriptive statistics about our sample of universities.

3.1 Data Sources

Employer-Employee matched Dataset (RAIS). In our main analysis, we use earnings and schooling data from the *Relação Anual de Informações Sociais* (RAIS) from 2003 to 2018. RAIS is an administrative dataset to which firms are required to report, thus covering the universe of formal employees and firms in the private and public sectors in Brazil. It contains restricted information on individuals including their full name and tax-identifier (*Cadastro de Pessoa Física*, or CPF), as well as sociodemographic information, such as age, gender, race, and schooling.

College Graduates Sample. We obtained data on college graduates from 42 public universities in Brazil through FOIL requests (Freedom of Information Law). The data consists of the full name, university, major, year of admission, and year of graduation, for students of 42 universities who responded to the FOIL request. We have information on 1.2 million students who graduated between 1990 and 2020. Appendix Table **??** presents the sample of universities.

Higher Education Census and Ranks. We use four additional datasets to complement the analysis. First, we use the Brazilian census of higher education to document the growth in enrollment by type of institution. The census is available for every year between 2000 and 2019 and covers all universities and majors in Brazil. It includes information such as total enrollment, number of graduates, and each major's date of foundation. In 1995, there were 884 higher education institutions and around 7,000 majors in Brazil. There was strong growth in the sector, such that in 2019 there were 2,608 higher education institutions and more than 40,000 majors.

Secondly, we use national examination data from Exame Nacional de Desempenho dos Estudantes (ENADE), and two rankings of higher institutions that are published online (RUF, from Folha de Sao Paulo newspaper; and Web ranking) to rank universities. Using ENADE's data from 2014, we create a university score by aggregating scores from all students and majors from each university. In Section 5.2, we combine these data to describe the changes in the composition of college graduates over time in terms of school age and ranking.

3.2 Sample Universities

Using the higher education census and RUF, we show that the 42 sample universities are older and better ranked. Table 1 presents the age distribution of sample universities and out-of-sample universities, i.e., all other universities. The data comes from the RUF ranking and is limited to 194 universities. The table shows that most universities in our sample were founded more than 50 years ago (59.5%). In comparison, only 37.7% of out-of-sample universities were founded more than 50 years ago.

Table 2 presents the RUF score and ranking, and the web ranking for the two samples. Universities in our sample have better scores and as a consequence are better ranked, according to the RUF ranking. The Web ranking includes more universities (1,285) and shows an even larger discrepancy between the sample universities and all other universities. The median ranking in the sample is 37, and 663 among other universities. In summary, our sample includes many of the best and oldest universities in the country.

Table 1: University age

University Age	Sample	All other universities
<30 years	10.8%	31.2%
30 to 50 years	29.7%	33.1%
>50 years	59.5%	35.7%
N	37	157

Note: The dependent variable is the fixed effects from the estimation of Equation 1, with the exception that we include university fixed effects and do not include major by university fixed effects. Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

Table 2: Institutions ranking and score

	Mean	S.D.	Median	Min	Max	N
RUF score						
Sample	69.6	21.7	42.1	4.8	98.0	37
All other universities	43.0	21.0	74.0	4.2	97.0	157
RUF ranking						
Sample	48.3	48.7	33	3	197	37
All other universities	110.2	52.6	111	1	196	157
Web ranking						
Sample	52.7	50.4	37	3	219	41
All other universities	662.5	361.0	663.5	1	1285	1244

3.3 Data Linkage

To link the sample of college graduates to RAIS, we first use individuals' full names, then use a machine learning algorithm to select the best match in cases of multiple matches. Third, we impose some sample restrictions to eliminate implausible matches.

The raw college graduates sample includes information of 1,217,440 students who graduated from 42 universities. After removing special characters, we are able to match 74% of these students to one or more workers with the exact same full name in the RAIS dataset. As a result, 2,093,069 workers are matched to 906,420 students. Out of these students, 78% are matched to a single worker and 22% are matched to at most 20 workers with the exact same name. We drop students with very common names that are matched to more than 20 workers.

Among students with multiple matches, we select the best match using a machine learning procedure. One university provided us with students' identification numbers (CPF) such that, for this university, we can match students and workers by both name and CPF. We proceed by matching students by name, and, for a training sample, we estimate a model that uses students' and workers' characteristics to predict whether the match is correct — as defined by the match using the identification number.

We estimate two different models — a logit regression, and a random forest model — using the training sample. Using the model's estimates, we calculate a score for each match. We define the correct match based on 3 rules: (i) The match has the highest score of all matches; (ii) The score of the match is sufficiently large (greater than 5%); and (iii) The score of the top match is sufficiently large relative to the second-best match (the ratio of scores is greater than 1.1).

Appendix Table B1 compares the results of each model using two metrics: the positive predictive value (PPV or accuracy) and the true positive rate (TPR or efficiency). While the random forest model has better PPV and TPR rates in the training sample (96.6% and 97.2%, respectively), these rates are lower in the test sample. Therefore, we decide to use the logit approach due to the consistency of the training sample and test sample metrics (PPV of 87.1% and 87.2% and TPR of 92% and (92.2%). Appendix Table B2 presents the sample sizes used for the in-sample PPV and TPR calculations.

The machine learning algorithm finds a match for 806,893 students/workers. We further restrict the sample to students who graduated between 2000 and 2017. The final dataset includes 17,455,296 employment observations from 545,478 students/workers, that graduated from 42 universities. In the rest of the paper, we refer to students in this sample as the "college graduates' sample" and the universities as "sample universities". All the other universities in Brazil are referred to as "out-of-sample universities".

3.4 Age-adjusted college premium

We define age-adjusted college premium as the weighted ratio of earnings for workers with a college degree and workers with a high school degree. Similar to Fernández and Messina (2018), premiums are constructed using a fixed-weight average of every age subgroup, for workers of ages between 21 and 65 years old. The weights are equal to the mean employment share of each subgroup across all years. We present the weighted average by aggregating all groups.

Figure 1 presents the evolution of the college premium over time using the nationally representative household survey (PNAD, in Panel A) and the employer-employee matched dataset (RAIS, in Panel B). Panel A shows that, on average, college workers' wages were 123% higher than the wages of workers with only a high school degree (2.23 times higher). The college premium increased by around 20p.p. between 1997 and 2004, a 3p.p. annual growth rate. Between 2004 and 2015, the college premium decreased by 19p.p., or -2p.p. a year. Fernández and Messina (2018), describe a similar picture for Brazil and other countries in Latin America.

Panel B of Figure 1 shows that in 1995, conditional on having formal employment, workers with a college degree earned 86% higher wages than workers with only a high school degree. There was a strong increase in the college premium between 1994 and 2002 when the difference in wages was 137%. This represents a 6p.p. annual growth. The college premium has a more moderate growth between 2002 and 2012 when it reaches the peak of a 150% difference in average wages (1p.p. per year). Finally, the college premium decreased to a 129% difference between 2012 and 2019 (-3p.p. per year).

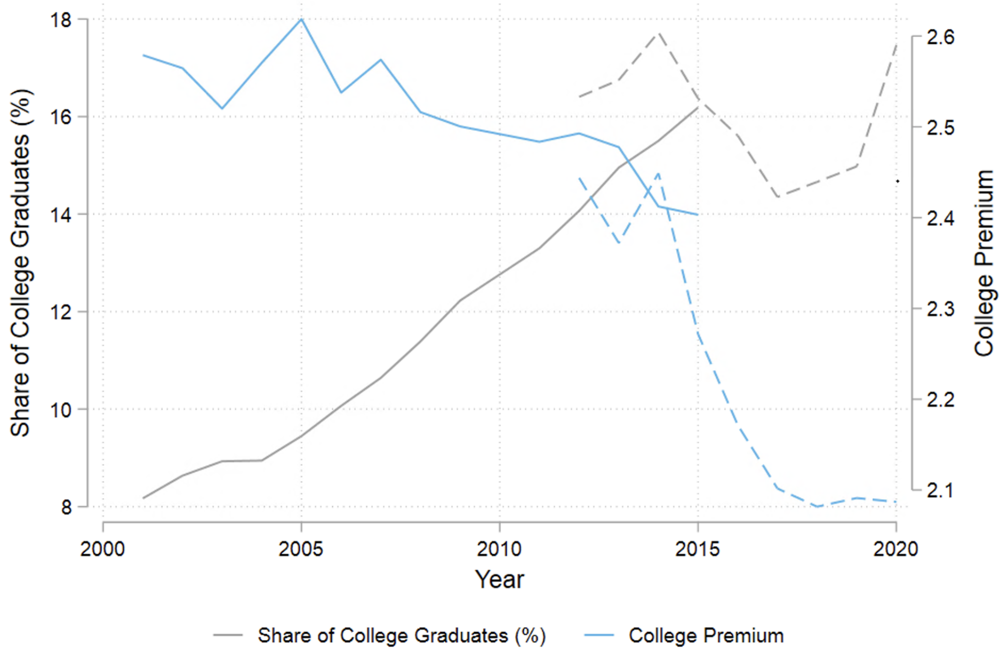
In summary, the age-adjusted college premium has been declining from its peak, a trend that is present in both the nationally representative household survey and the universe of formal employees.

3.5 Changes in labor supply

The share of college graduates in Brazil doubled from 8% to 16% between 1995 and 2020, as a share of total workers, as shown in Figure 1. According to the Higher Education Census, the number of people graduating from college increased from 400 thousand to 1.3 million people per year (Appendix Figure C1). These trends are not specific to Brazil. In fact, college enrollment in upper-middle-income countries—a group of 54 countries as defined by the World Bank—has increased by a similar magnitude during the same period as in Brazil.

Despite initial disparities, the increase in college access was felt across all racial groups and in all regions of the country. Appendix Figure C2 shows the share of college graduates in 1995 and 2000 by race and macro-region. The share of college graduates among brown and black workers was 9,7% and 9,0% in 2020, respectively, less than half of that among whites (19,4%). From the same figure, we can see the unconditional college wage premia on the right-hand panels: university graduates went from earning 2.4 times more than the average worker with a high school education, to 2.7 times more in 2003-2005, and gradually less so since then. By racial group, it is also possible to see a similar pattern, with college wage premia decreasing since the early- to mid-2000s.

Figure 1: Share of College Graduates and the Unadjusted College Premium



Note: The figure shows demographic-adjusted share of college graduates (left axis) and college premium (right axis), adjusted to remove differences in age groups, gender, and geographical region from 2001 to 2020, for working-age employees working at least 40 hours a week using PNAD data. The dashed lines represent data from *PNAD Contínua*, which replaced the annual PNAD survey in 2015.

4. The Adjusted College Premium Trends

In this section, we examine the evolution of the College Premium over the past two decades, considering essential individual characteristics, such as age, the type of higher education institution attended, major, and year of graduation. Our objective is to investigate whether the decline in the college premium is universal or if there are specific heterogeneous effects that merit consideration.

To conduct this analysis, we partition our data on college graduates into two distinct samples. Sample A comprises all college graduate workers identified from the 42 universities obtained through FOIL requests, linked to the RAIS dataset. Sample B includes other workers in the RAIS dataset possessing a college degree but graduating from universities beyond our designated sample. This separation enables us to maintain a constant set of universities, facilitating a comparative examination of wage evolution between the designated sample and others. By doing so, we aim to account for university fixed effects, gaining insight into the underlying factors contributing to the theoretically proposed decline in the college premium.

We present wage estimates from both samples in relation to the wages of all workers in the RAIS dataset holding a high school degree. Consequently, our analysis focuses on estimating the wage gap evolution between college and high school graduates for workers graduating from our designated sample universities versus those from other institutions. The study encompasses all workers in the RAIS dataset aged 21 to 65, working 40 hours per week, employed on December 31st, receiving positive wages, and possessing either a complete high school or college degree. The resulting sample comprises approximately 395 million worker-wage observations. [6]

To compute and adjust the college premium, we conduct a regression on the logarithm of wages for individual i and year t. This regression includes a set of indicators and fixed effects to account for three distinct groups: college graduates from our designated sample universities, college graduates from other universities, and high school graduates, as illustrated in the following equation:

$$\begin{aligned} ln(wage)_{it} = & \sum_{t=1}^{T} \beta_t^U [Sample\ University \times year_t] \\ & + \sum_{t=1}^{T} \beta_t^C [College \times year_t] \\ & + \sum_{t=1}^{T} \beta_t^{HS} [High\ School \times year_t] \\ & + \alpha_{a,s} + \psi_c + \eta_{u,m} + \varepsilon_{it} \end{aligned} \tag{1}$$

where $\alpha_{a,s}$ represents age-by-schooling fixed effects, ψ_c denotes cohort-of-graduation fixed effects, and $\eta_{u,m}$ stands for university-by-major fixed effects. [7] We set values to zero for both ψ_c and η_{um} for high school workers and for college workers who graduated from universities outside our designated sample.[8]

There are several reasons to adjust the analysis of the college premium. Firstly, older workers often receive higher wages, potentially due to an experience premium. Changes in the demographic composition of the labor force or shifts in the age distribution can impact the accuracy of the college premium measurement. To address this, we introduce age-by-

[6] This study excludes individuals with some college education who did not complete their college degree.

[7] We can separately identify age, cohort, and year effects as we observe workers of different ages in the same cohort of graduation and over time. In other words, age, cohort of graduation, and year do not form a colinear relation.

[8] Unfortunately, we only observe the cohort of graduation, university and major for individuals in our designed sample universities

sample fixed effects ($\alpha_{a,s}$) in our specification. Secondly, within our designated sample of universities, the number of observations by cohort varies over time. For example, in 2007, we observe cohorts graduating between 2000 and 2007, while in 2018, our dataset includes individuals graduating between 2000 and 2018. To account for this variation, we incorporate adjustments for the cohort of completion (ψ_c). Lastly, even though sample (a) comprises observations from a fixed set of universities, the composition of majors and universities may change over time due to factors such as universities opening or closing new majors. To control for these variations, we include adjustments for university and major ($\eta_{u,m}$).

We are primarily focused on examining the estimates of β_t^U and β_t^C from the provided equation. These estimates reflect the changes in log wages for each sample of college graduate workers—specifically, the designated sample of universities and everyone else—in comparison to the wages of high school workers. Notably, due to the omission of indicator variables for the first year in the data to address collinearity, all estimates are presented relative to the 2003 college premium.

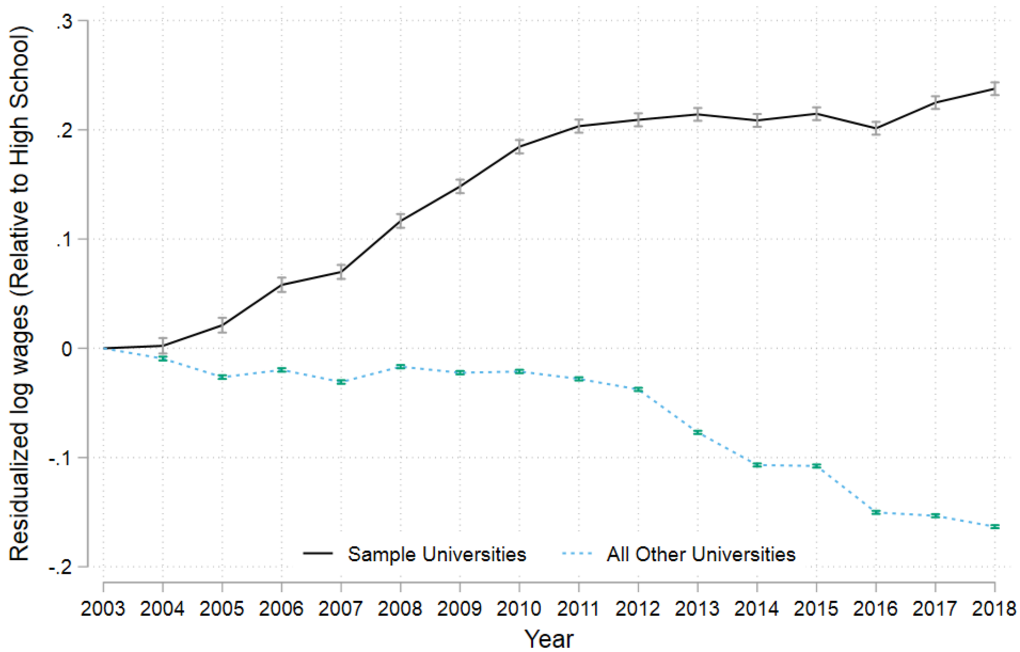
Figure 2 illustrates a significant increase in the college premium for the designated sample of universities. Specifically, it remained relatively stable until 2007, experiencing a subsequent 19% rise between 2007 and 2011, equivalent to an annual increase of 4%. The growth rate of the college premium moderated between 2011 and 2018, resulting in a total increase of 5% or an annual growth rate of 0.6%. It's important to highlight that this observed increase in the college premium contrasts with the unconditional college premium presented in Figure 1, which depicted a decline.

However, the estimates for workers in sample b — those in the RAIS dataset with a college degree, excluding workers from our designated universities — show a decrease during this period. The dashed line in Figure 2 illustrates a decline in the college premium for this sample between 2003 and 2006, stability between 2006 and 2011, and a subsequent notable 14.5% decrease between 2011 and 2018—an annual decline of 1.8%.[9] [10]

[9]Additionally, in Appendix Figure C5, we depict the wage trajectories for individuals with a college degree from our designated universities, those with a college degree from other universities, and individuals with a high school degree, respectively.

[10]As a robustness check, in Appendix Figure C6, we replicate this analysis using a specification that incorporates university and municipality fixed effects instead of university x major fixed effects. The outcomes closely align with those presented in Figure 2.

Figure 2: Trends in residualized log wages (college premium)



Note: The figure illustrates the progression of the college premium, relative to 2003 values, for two distinct groups: the college graduate sample from FOIL requests (Sample Universities) and the sample of all workers in the RAIS dataset with a college degree, excluding workers in the Sample Universities (All Other Universities). Both curves are benchmarked against the same trends in high school residualized wages. The estimates are derived from the estimation of Equation 1 using a sample comprising 298,135,481 observations. Additionally, the figure includes 95% confidence intervals for each point estimate.

4.1 Cohort Heterogeneity in College Premium Trends

We present evidence indicating that the trends in the college premium vary across different demographic groups, particularly in response to the education expansion in Brazil. Given the recent increase in educational opportunities, it is logical to expect disparities in the college premium trends between our two sample groups, particularly among individuals differing in age.

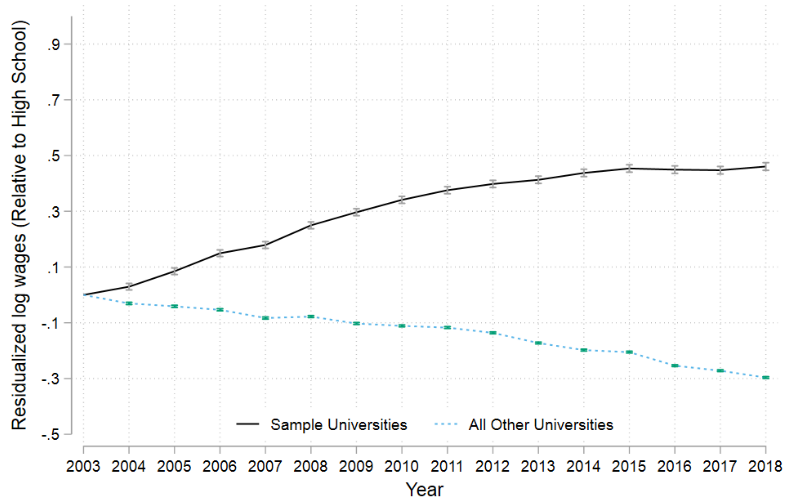
Our forthcoming Figure 3 illustrates two main take-aways. First, older college graduates exhibit similar college premium trends irrespective of the quality of the university they attended. Second, the college premium for older college graduate workers has not declined during the last two decades.

The literature suggests that expansions in education, or the increase in supply of college graduate workers, can lead to varying impacts on individuals based on their age or level of experience. This observation arises from the understanding that workers of different ages are not perfect substitutes for each other. Consequently, older workers appear less affected by educational expansions and subsequent changes in workforce composition.

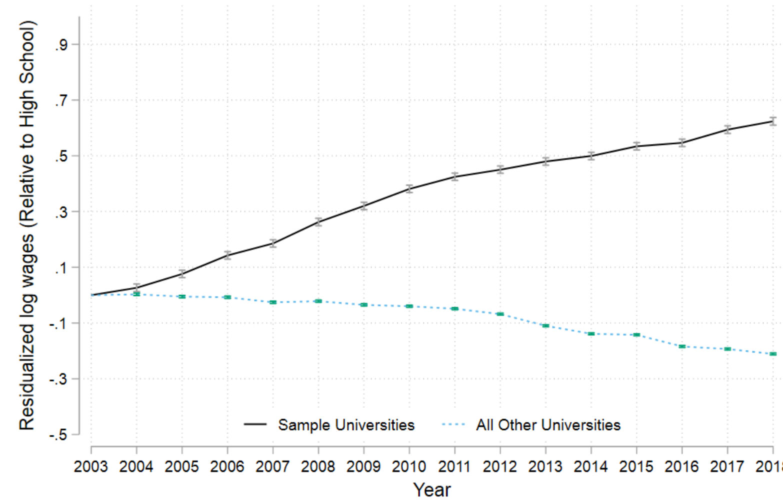
In Figure 3, we observe that the disparities in college premium diminish with the age of graduate workers. Notably, there is a significant gap between graduates from Sample universities and those from other institutions among individuals aged 25 to 34 (Figure 3 a and b). However, this gap gradually narrows for those aged 35 to 44 (Figure 3 c-e) and nearly disappears among workers aged over 45 (Figure 3 f-h). Furthermore, for these latter cohorts, there is no decline in the college premium for either group. This observation underscores the nuanced impact of education expansion on different age cohorts within the labor market and suggests that changes in composition among young college graduate workers do not affect the salaries of older college graduate workers.

Figure 3: Trends in the residualized college premium by age

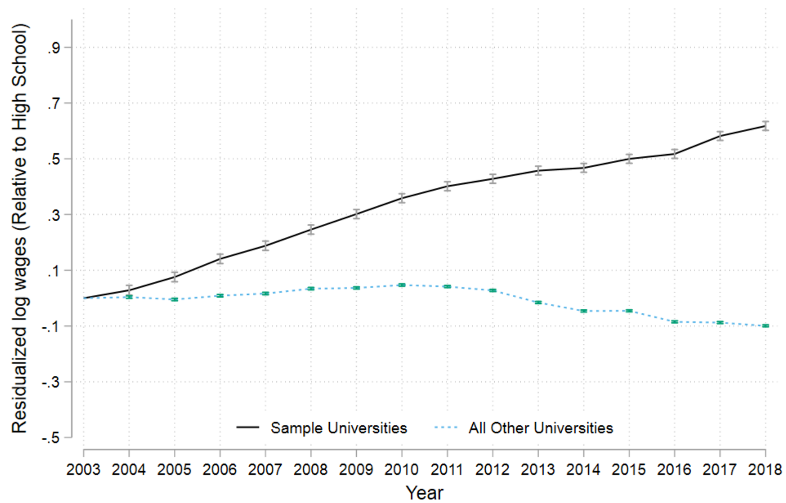
(a) Age group 25-29



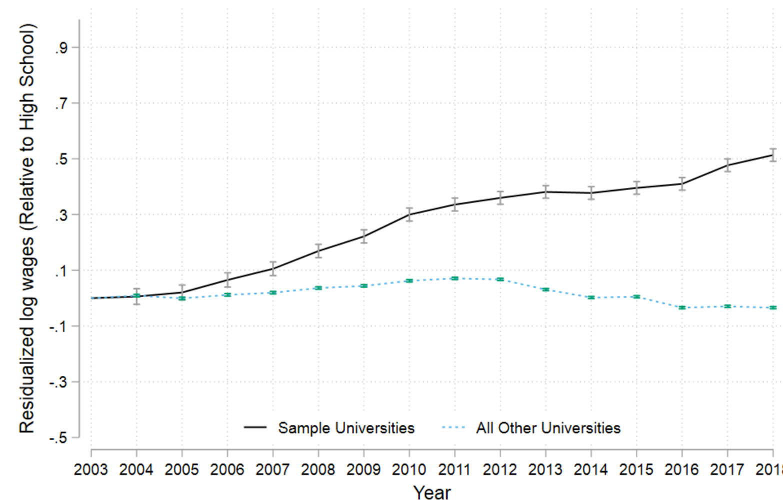
(b) Age group 30-34

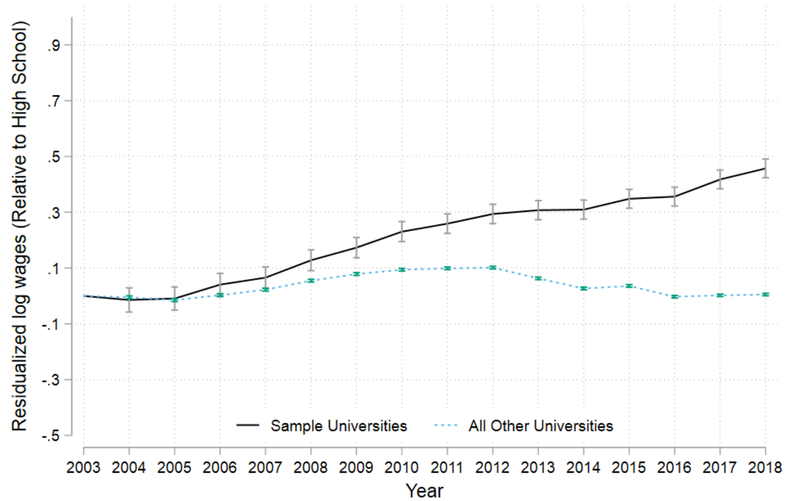


(c) Age group 35-39

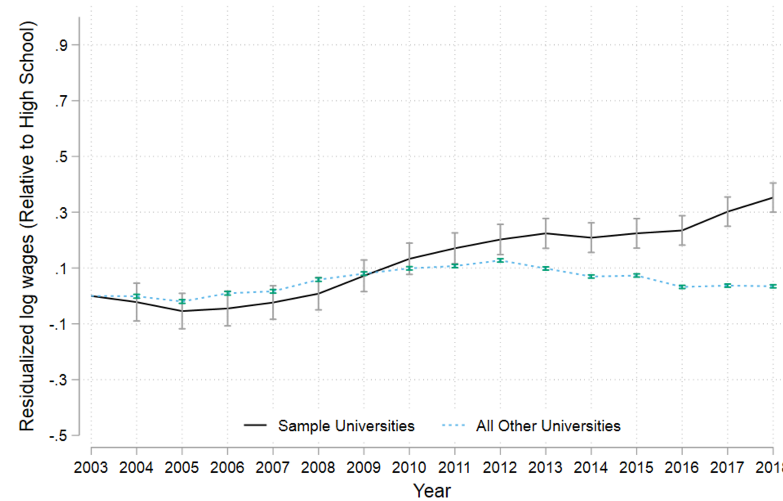


(d) Age group 40-44

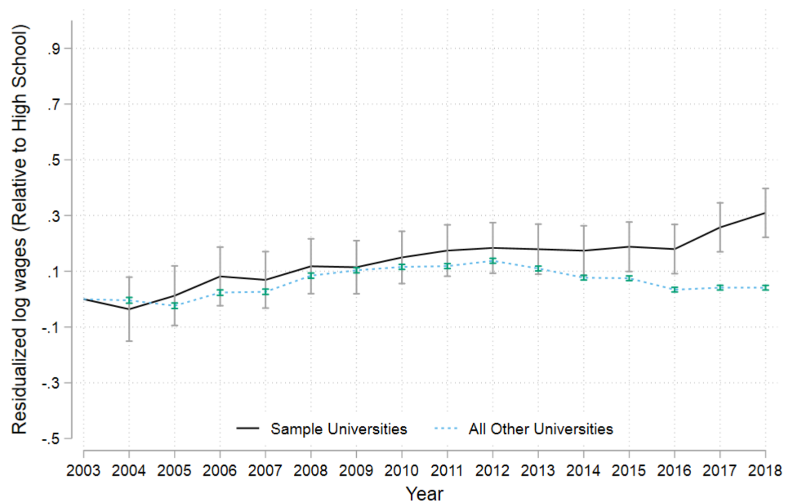




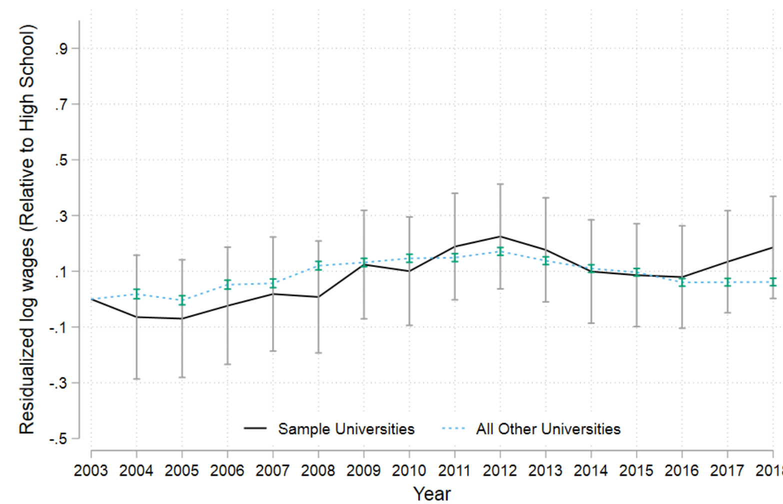
(e) Age group 45-49



(f) Age group 50-54



(g) Age group 55-59



(h) Age group 60-64

Note: The figure illustrates the progression of the college premium, relative to 2003 values, for two distinct groups: the college graduate sample from FOIL requests (Sample Universities) and the sample of all workers in the RAIS dataset with a college degree, excluding workers in the Sample Universities (All Other Universities) by age. Both curves are benchmarked against the same trends in high school residualized wages. The estimates are derived from the estimation of Equation 1 using university × major fixed effects. Additionally, the figure includes 95% confidence intervals for each point estimate.

5. Evidence of Changes in Graduates' Skill Composition

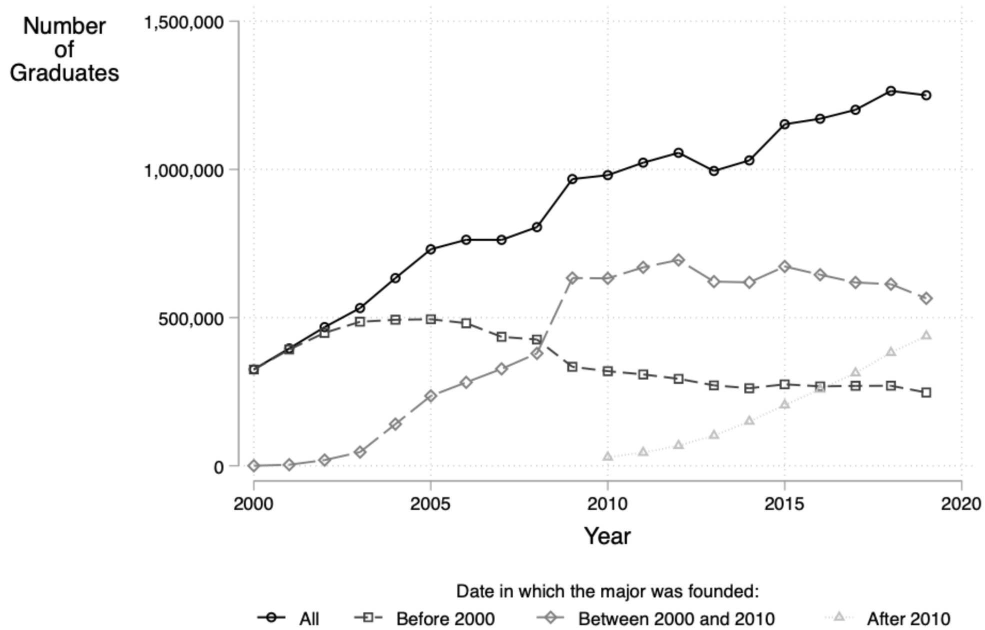
After the analysis presented in the previous section, where we demonstrated the increasing trend in the college premium for our fixed set of universities, this section aims to provide additional evidence pointing to a shift in the composition of Brazilian universities and graduates. This shift may help explain a portion of the overall decrease in the average college premium and the observed patterns discussed earlier. To substantiate this claim, we present three stylized facts.

5.1 The role of new universities and majors

First, a crucial aspect under scrutiny is the impact of new institutions and majors on defining wages for diverse worker categories. Over the last two decades, there has been a significant surge in individuals obtaining a college degree, closely associated with the expansion of higher education programs. Data from the higher education census reveals a notable increase in annual graduates from 400 to 1,300,000 between 2000 and 2020 (Appendix Figure C1).

Furthermore, Figure 4 presents intriguing dynamics in the evolution of annual graduates based on the founding years of universities and majors. Our analysis categorizes these into three groups: those founded before 2000, between 2000 and 2010, and after 2010. The figure reveals distinct trends—graduates from the oldest group of universities have declined over time, those from majors founded between 2000 and 2010 increased until 2010 and stabilized, even slightly decreasing in recent years. In contrast, the number of students graduating from majors founded after 2010 is experiencing an exponential increase. This observation suggests that the rise in college-educated workers is predominantly from recently established majors, which may differ significantly from older institutions in terms of quality standards and the majors offered. Such differences could potentially impact the wages of workers with college degrees, thereby influencing the average college premium. Additionally, Appendix Figures C7 and C8, provide evidence of the significant increase in new majors over the last two decades. Notably, this increase is even more pronounced for private universities.

Figure 4: Number of graduates by major's date of foundation



Note: The figure presents the number of students graduating each year. Categories are defined by the year in which the major was founded. Source: Higher Education Census (2000-2018)

The shift in the composition of universities and graduates, as discussed earlier, may account for the variations in the evolution of the college premium between our designated sample of universities and all other universities. A potential explanatory framework comes from the "Degraded Tertiary Hypothesis" (Camacho et al., 2017). This hypothesis posits that the average skill of college workers has declined due to an increase in graduates with lower-quality degrees, rather than a reduction in the quality of each degree individually.

Two legislative changes in Brazil, detailed in the setting section, could have influenced the quality of new institutions. Firstly, Law number 9394 enacted in 1996 authorized and promoted distance learning, a mode that constituted 20% of all student graduations by 2019. Secondly, Executive Order 2207 from 1997 allowed private higher education institutions to operate as for-profit entities, a departure from the previous norm where private higher education institutions were exclusively non-profit. The subsequent discussion will delve deeper into these differences in quality.

5.2 Change in the quality of higher education institutions

As mentioned earlier, there is a potential link between the increased number of graduates with lower-quality degrees and the observed trends. The surge in graduates from recently established universities and majors prompts an investigation into the quality of the programs they offer. If these recently established institutions and majors are associated with lower-quality programs, such an association could shed light on the decline in the college premium for out-of-sample universities. To explore this hypothesis, we examine the relationship between a university's year of foundation and the quality of its programs.

We provide evidence that students from new universities demonstrate lower performance in standardized tests. Table 3, presents compelling evidence showing that students from universities that introduced their majors more recently perform comparatively worse in standardized exams administered post-completion of their college degrees, both in specific and general knowledge.[11] A similar pattern emerges when considering the year of foundation and the ranking positions and scores from the RUF national ranking.[12] In both analyses, older institutions outperform newer counterparts, securing higher positions in the ranking and achieving higher scores. This observation, combined with the declining number of graduates from older, higher-quality universities, contributes to an explanation for the decline in the average college premium.[13]

[11] Additionally, Appendix Figure C9 presents evidence, using a non-parametric regression, of a negative relationship between ENADE scores and the university's foundation year.

[12] RUF is the Folha University Ranking, which is a national ranking that includes a smaller sample of universities compared to ENADE.

[13] Appendix Figure C10 supplements this finding by illustrating a substantial decrease in the fraction of students graduating from the top 25% of universities over the last two decades, according to the ENADE ranking.

Table 3: Association between universities' year of foundation and quality

Dependent variable:	ENADE score			RUF	
	All	Specific knowledge	General knowledge	Ranking	Score
	(1)	(2)	(3)	(4)	(5)
Year in which the institution's first major was founded:					
$1940 < year < 1960$	-2.948***	-3.211**	-2.160*	46.12***	-19.85***
	(1.075)	(1.261)	(1.175)	(10.21)	(3.953)
$1960 < year < 1980$	-5.138***	-5.294***	-4.671***	72.66***	-31.13***
	(0.857)	(1.006)	(0.937)	(8.735)	(3.336)
$1980 < year < 2000$	-6.299***	-6.248***	-6.452***	97.31***	-37.90***
	(0.867)	(1.017)	(0.948)	(13.86)	(5.815)
$2000 < year < 2020$	-5.863***	-5.654***	-6.490***	112.2***	-43.10***
	(0.818)	(0.959)	(0.894)	(13.54)	(6.301)
Constant	49.32***	46.04***	59.13***	41.76***	74.76***
	(0.796)	(0.933)	(0.870)	(7.174)	(2.700)
Observations	1,469	1,469	1,469	197	171
R-squared	0.046	0.030	0.058	0.364	0.402

Note: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

5.3 Universities' quality and wages

Finally, if workers who graduated from lower-ranked universities receive lower wages, coupled with the last two stylized facts, this could help unravel the puzzle of a decreasing trend in the overall college premium, yet an increasing trend for our designated sample of universities.

To support these hypotheses, we estimate the effects of university quality on average wages. We conduct regressions on the effects of the ENADE scores and university rankings on wages using data on average wages by major from 2019 to 2021 for the universe of workers in RAIS that graduated between 2010 and 2015.[14] Table 4 demonstrates that universities with higher ENADE scores or in higher ranking positions exhibit larger average wages.

[14] For this analysis, we use aggregated data at the major-university level coming from the linkage of the Higher Education Census and RAIS provided by Feinmann and Rocha (2024).

Specifically, it reveals that a one-standard deviation increase in ENADE scores is related to an 8.7% increase in average wages. Students from universities ranked 1st receive wages that are 30% higher than students from universities ranked 100th. In Appendix Table B3 we examine whether the effects of school quality on wages vary based on the years since graduation. We do not observe any significant differences, indicating that school quality does not influence the impact of years of experience on wages.

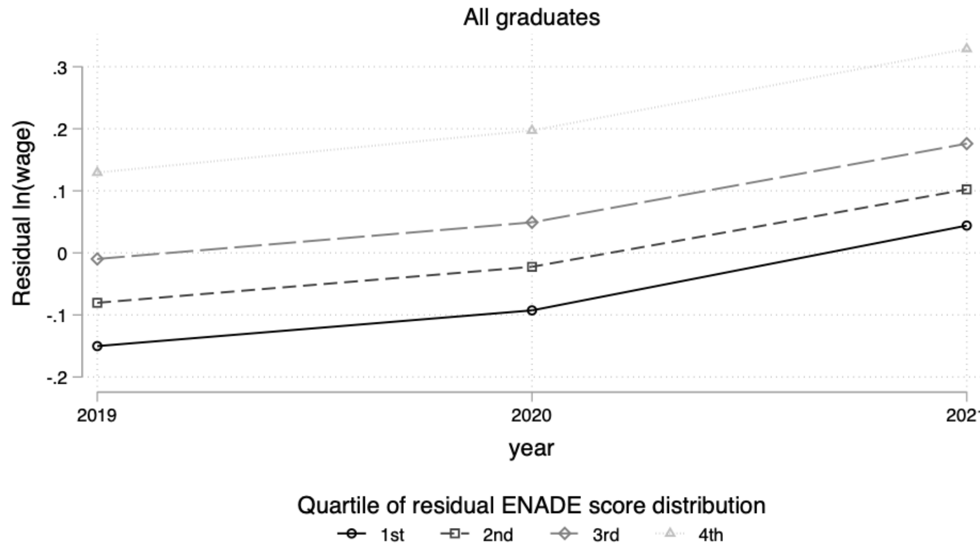
Table 4: Correlation between universities' quality and wages

Dep. Var: log(wages)	(1)	(2)	(3)	(4)
ENADE	0.086***	0.087***		
	(0.001)	(0.001)		
Ranking RUF			-0.003***	-0.003***
			(0.000)	(0.000)
Constant	8.444***	8.611***	8.720***	8.894***
	(0.005)	(0.005)	(0.002)	(0.003)
Cohort Fixed effects	No	Yes	No	Yes
Observations	353,470	353,470	209,905	209,905
Number of universities	2338	2338	196	196
Number of majors	66	66	78	78
R-Squared	0.07	0.14	0.11	0.19

Notes: All speifications inclue year fixed effects. The ENADE score is the starndardized weighted average of all the major-level ENADE scores for an institution between 2010 and 2020.

Additionally, we perform a similar exercise to understand the evolution of wages for graduates from universities of different quality. Figure 5 displays the evolution of the average wages from 2019 to 2021 for universities with different quality levels based on the ENADE Score quartiles. We can observe important wage differences across quartiles, however, wages are increasing for universities at all levels of quality, showing very similar trends.

Figure 5: Residualized log wages by ENADE (2019-2021)



Note: The graph displays the average residuals from a regression of log wages with respect to graduation cohort fixed effects for each quartile of the major´s score distribution. We include all students-workers in the formal labor market. More specifically, the data for each year comes from 2010 to 2015 graduates. Quartiles of ENADE score are derived from the residual of a regression between ENADE score and international major code fixed effects. Data only includes students matched to RAIS. Wages are weighted by the number of students in each major.

Considering these three stylized facts collectively, one could infer that, despite the decreasing trend in the average college premium, it may be the case that all universities exhibit increasing trends in the college premium at different levels. However, the substantial increase in graduates from newer, lower-ranked universities—with lower fixed effects and thus lower wages—translates into a decreasing trend for the average college premium. This shift is attributed to the overall college premium placing greater emphasis on workers from lower-ranked universities by the end of the period, owing to the resulting change in composition.

6. The Composition-Adjusted Labor Supply and Demand Model

We follow the labor supply and demand model developed in Katz and Murphy (1992), Card and Lemieux (2001), Acemoglu and Autor (2011), Fernández and Messina (2018), among others. The primary objectives of this exercise are twofold: Firstly, we aim to quantify the skill-biased technical change in Brazil between 2002 and 2015. Secondly, and more importantly, we intend to demonstrate that relying on the conventional measure of the college premium can result in substantial biases when estimating the growth parameter of skill-biased technical change.

In our model, workers are categorized as either low-skill (L) or high-skill (H). We introduce a modification wherein workers within each skill type $K = \{L, H\}$ are differentiated by the extent of their skill, quantified as efficiency units of skill, denoted as e_i^K. Similar to Carneiro and Lee (2011), we interpret e_i^K as capturing the quality differences among workers who possess identical degrees. Differences in quality may arise from factors such as individual ability, the educational institution's quality, or work experience in the labor market. The production function for the aggregate economy is represented using the Constant Elasticity of Substitution (CES) form, and expresses aggregate output Y as follows:

$$Y_t = \left[(A_{L,t}\mathcal{L}_t)^{\frac{\sigma-1}{\sigma}} + (A_{H,t}\mathcal{H}_t)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{2}$$

Here, $\mathcal{L}_t$ and $\mathcal{H}_t$ denote the aggregate levels of low and high skill in the economy at period t, respectively. It is important to note that $\mathcal{L}_t$ and $\mathcal{H}_t$ do not simply represent the count of low and high-skill workers. Instead, they are the summation of the efficiency units of each worker in their respective categories, i.e., $\mathcal{L}_t = \sum_i e_{it}^L$ and $\mathcal{H}_t = \sum_i e_{it}^H$. The terms $A_{L,t}$ and $A_{H,t}$ are the factor-augmenting technology terms for low and high skills, respectively. The parameter σ denotes the elasticity of substitution. Within this framework, workers offering low and high-skill types are considered imperfect substitutes for each other. Conversely, workers providing the same type of skill are viewed as perfect substitutes, differentiated only by the scaling factor e_i^K.[15]

In an economy with perfect competition where skill of each type is offered inelastically, the wage per efficiency unit of skill equals its marginal revenue product of labor, $\omega_K = MRPL_K$,

[15] In Appendix A, we estimate a model with a nested CES production function in which workers of different experience groups are imperfect substitutes, as in Fernández and Messina (2018).

which gives us the two equilibrium conditions below.

$$\omega_{L,t} = \frac{\partial Y_t}{\partial \mathcal{L}_t} = A_{L,t}^{\frac{\sigma}{\sigma-1}} \left[(A_{L,t})^{\frac{\sigma-1}{\sigma}} + (A_{H,t})^{\frac{\sigma-1}{\sigma}} (\mathcal{H}_t/\mathcal{L}_t)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}} \tag{3}$$

$$\omega_{H,t} = \frac{\partial Y_t}{\partial \mathcal{H}_t} = A_{H,t}^{\frac{\sigma}{\sigma-1}} \left[(A_{H,t})^{\frac{\sigma-1}{\sigma}} + (A_{L,t})^{\frac{\sigma-1}{\sigma}} (\mathcal{L}_t/\mathcal{H}_t)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{1}{\sigma-1}} \tag{4}$$

Combining the equilibrium conditions above and taking logs, the skill premium per efficiency unit ($\omega_t = \omega_{H,t}/\omega_{L,t}$) is defined by the equation below.

$$\ln \omega_t = \frac{\sigma - 1}{\sigma} \ln \left(\frac{A_{H,t}}{A_{L,t}} \right) - \frac{1}{\sigma} \ln \left(\frac{\mathcal{H}_t}{\mathcal{L}_t} \right) \tag{5}$$

For simplicity, assume that there is a log-linear change in the relative demand for skills. We are interested in estimating the parameter γ_1, i.e., the growth rate of relative demand for high-skill.

$$\ln \left(\frac{A_{H,t}}{A_{L,t}} \right) = \gamma_0 + \gamma_1 t \tag{6}$$

Combining the Equations 5 and 6, we get the reduced-form equation bellow, where u_t accounts for measurement error:

$$\ln \omega_t = \frac{\sigma - 1}{\sigma} \gamma_0 + \frac{\sigma - 1}{\sigma} \gamma_1 t - \frac{1}{\sigma} \ln \left(\frac{\mathcal{H}_t}{\mathcal{L}_t} \right) + u_t \tag{7}$$

It is common to use data from household surveys to estimate the equation above. However, we do not directly observe the skill premium (ω_t) and the ratio of labor inputs ($\frac{\mathcal{H}_t}{\mathcal{L}_t}$) in the data. Instead, we have access to average wages (W_t^L and W_t^H) and the number of workers with high school and college degrees (N^L and N^H). To utilize this data effectively, we must introduce certain assumptions to the model.

6.1 Identification

Using few assumptions, we can derive a reduced-form model based on the observable variables. Take $\bar{e}_t^K$ to be the average number of efficiency units across workers of type K at time t. With simple algebra, we get that:

$$\frac{\mathcal{H}_t}{\mathcal{L}_t} = \frac{N_t^H}{N_t^L} \frac{\bar{e}_t^H}{\bar{e}_t^L} \quad , \quad \frac{W_t^H}{W_t^L} = \omega_t \frac{\bar{e}_t^H}{\bar{e}_t^L} \tag{8}$$

By substituting (8) in (7) we get an equilibrium condition in terms of observable variables and the ratio between average efficiency units across skill types ($\frac{\bar{e}_t^H}{\bar{e}_t^L}$):

$$\ln\left(\frac{W_t^H}{W_t^L}\right) = \frac{\sigma-1}{\sigma}\gamma_0 + \frac{\sigma-1}{\sigma}\gamma_1 t - \frac{1}{\sigma}\ln\left(\frac{N_t^H}{N_t^L}\right) + \left(1-\frac{1}{\sigma}\right)\ln\left(\frac{\bar{e}_t^H}{\bar{e}_t^L}\right) + u_t \tag{9}$$

We get the quality-adjusted reduced-form equation by taking differences of (24) with respect to $t = 0$:

$$\ln\left(\frac{W_t^H}{W_t^L}\right) = \beta_0 + \frac{\sigma-1}{\sigma}\gamma_1 t - \frac{1}{\sigma}\ln\left(\frac{N_t^H}{N_t^L}\right) + \underbrace{\left(1-\frac{1}{\sigma}\right)\left[\ln\left(\frac{\bar{e}_t^H}{\bar{e}_t^L}\right) - \ln\left(\frac{\bar{e}_0^H}{\bar{e}_0^L}\right)\right]}_{Quality-adjustment} + v_t \tag{10}$$

where $\beta_0 = \ln\left(\frac{W_0^H}{W_0^L}\right) + \frac{1}{\sigma}\ln\left(\frac{N_0^H}{N_0^L}\right)$. The "quality-adjustment" term above is inadvertently omitted when using wage and employment data to estimate Equation 7, which can result in biased estimates. Our objective is to identify a data equivalent for this term. Identification in this model relies on two key assumptions:

Assumption 1: The average number of efficiency units supplied by low-skill workers remained constant over the period: $\bar{e}_t^L = \bar{e}^L$.

Assumption 2: The average number of efficiency units per high-skill worker in our *University Sample* is constant over the period: $\bar{e}_t^{HU} = \bar{e}^{HU}$

We proceed as follows. From the model definition of average wages in Equation 8, we have that:

$$\ln\left(\frac{W_t^H}{W_t^L}\right) = \ln\omega_t + \ln\left(\frac{\bar{e}_t^H}{\bar{e}_t^L}\right) \tag{11}$$

Taking differences with respect to $t = 0$:

$$\underbrace{\ln\left(\frac{W_t^H}{W_t^L}\right) - \ln\left(\frac{W_0^H}{W_0^L}\right)}_{\Delta \text{ college premium}} = \underbrace{\ln\omega_t - \ln\omega_0}_{\Delta \text{ skill premium}} + \underbrace{\ln\left(\frac{\bar{e}_t^H}{\bar{e}_t^L}\right) - \ln\left(\frac{\bar{e}_0^H}{\bar{e}_0^L}\right)}_{\Delta \text{ skill composition}} \tag{12}$$

Assumptions 1 and 2 imply that the observed change in wages within the *University Sample* does not account for changes in skill composition. Consequently, the change in wages identifies the changes in the skill premium.

$$\ln\left(\frac{W_t^{HU}}{W_t^L}\right) - \ln\left(\frac{W_0^{HU}}{W_0^L}\right) = \ln\omega_t - \ln\omega_0 \tag{13}$$

We identify changes in skill composition from the difference between Equations 12 and 13.

$$f(\mathbf{W}) = \underbrace{\ln\left(\frac{\bar{e}_t^H}{\bar{e}^L}\right) - \ln\left(\frac{\bar{e}_0^H}{\bar{e}^L}\right)}_{\Delta \text{ skill composition}} = \underbrace{\left[\ln\left(\frac{W_t^H}{W_t^L}\right) - \ln\left(\frac{W_0^H}{W_0^L}\right)\right]}_{\Delta \text{ college premium}} - \underbrace{\left[\ln\left(\frac{W_t^{HU}}{W_t^L}\right) - \ln\left(\frac{W_0^{HU}}{W_0^L}\right)\right]}_{\substack{\Delta \text{ college premium in the} \\ \textit{University Sample}}} \tag{14}$$

Using Equations 10 and 14, we establish our estimation specification in which all terms are observable in the data.

$$\ln\left(\frac{W_t^H}{W_t^L}\right) = \beta_0 + \frac{\sigma-1}{\sigma}\gamma_1 t - \frac{1}{\sigma}\ln\left(\frac{N_t^H}{N_t^L}\right) + \left(1-\frac{1}{\sigma}\right) f(\mathbf{W}) + v_t \tag{15}$$

where $\beta_0 = \ln\left(\frac{W_0^H}{W_0^L}\right) + \frac{1}{\sigma}\ln\left(\frac{N_0^H}{N_0^L}\right)$. It's important to note that, in contrast to the estimation of Equation 7, the estimation of the equation above imposes parameter restrictions.

6.2 Discussion: Identification assumptions

The arguments and evidence concerning changes in the average quality of low-skill workers are ambiguous. On one hand, it is possible that average quality is decreasing because (i) increasing access to secondary education may have attracted lower-quality students, and (ii) increasing access to college education may have attracted the best high-school students. On the other hand, average quality may be increasing due to (i) the decrease in the pupil/teacher ratio and (ii) the increase in federal funds for secondary education, resulting in higher expenditure.[16]

Additionally, test scores from PISA, an international evaluation of 15-year-old students, indicate a stagnation in student's proficiency (OECD, 2023). The results are presented in Appendix Figure C13. We interpret this combination of facts as evidence supporting the validity of Assumption 1.

[16] The arguments and evidence regarding changes in the average quality of high-skill workers unambiguously suggest a decrease in quality: (i) increasing access to college education may have attracted lower-quality students; (ii) the opening of the education market in 1997 may have attracted lower-quality institutions; and (iii) Federal and state governments have reduced expenditure per student during the period of analysis (see Appendix C11).

Appendix Figure C14 presents evidence in support of Assumption 2. Panel A demonstrates that the scores of students in the *University Sample* at end-of-graduation standardized exams remain constant during the period of analysis. Panel B illustrates that the proportion of students with more than a high school education does not undergo significant changes over time.[17] Institutional changes, such as the increased utilization of ENEM as an entrance exam in public universities and the REUNI program, which augmented federal investment in public universities, may have enhanced the quality of admitted students and the quality of education at institutions within the University Sample. Other programs, such as affirmative action, could have had the opposite effect. However, it's worth noting that these events occurred towards the end of the analysis period and would primarily impact students graduating at the conclusion of our study period, constituting a relatively small portion of our sample.

6.3 Estimation results

In this section, we present the estimates obtained from Equations 7 and 15. To do so, we combine data from the Brazilian household survey (PNAD) with the estimates of the change in skill composition derived from RAIS and University data.

In our estimation, we utilize PNAD data from 2003 to 2015. There are two reasons for that. First, after 2015, PNAD transitioned to PNAD Contínua, which involves a different methodology. Therefore, changes observed after 2015 could potentially be attributed to differences in methodology rather than underlying economic trends. Secondly, Brazil experienced a severe recession between 2015 and 2017, with GDP declining by 8% during this period. It would be unreasonable to assume that our model could adequately capture such significant structural changes in the economy.

The estimation results are presented in Appendix Table B4. In column 1, we estimate the Naive OLS approach, as in Equation 7. Columns 2 and 3 estimate our main specification with parameter restrictions, as indicated by Equation 15. Appendix Figure C12 shows that the the quality-adjustment time series used in the estimation ($f(\mathbf{W})$) is a decreasing function of time. We calibrate the elasticity of substitution to two different values: one based on the OLS estimate ($\sigma = 4.30$) and another assuming a more inelastic labor substitution technology ($\sigma = 2$).

The results, when adjusted for changes in composition, suggest that the relative demand

[17] Appendix Figure C15 further reveals that the percentage of females in tertiary education experiences a slight increase (3%), while the percentage of whites decreases by 10%.

for college-educated workers is increasing rapidly. Table 5 displays the estimated parameters for skill-biased technical change growth (γ_1) and the elasticity of substitution (σ). Column 1, using traditional college premium data, indicates a negative growth in the relative demand for college-educated workers at -0.1% per year. This contradicts the common belief that technological advancements have generally complemented higher skill levels in recent decades. Columns 2 and 3 reveal that OLS results are significantly biased due to changes in the average skill composition of college workers. The revised estimate for γ_1 is 3.07%, leading to a positive growth rate of 43.7% over the same period. Importantly, the estimates for γ_1 remain relatively stable across a range of plausible values for σ.

In conclusion, the last two decades have seen a consistent rise in the relative demand for college-educated workers. Notably, this increase has occurred alongside a significant expansion in tertiary education. It's important to highlight that the impact of skill-biased technical change is substantially underestimated when changes in the composition of the workforce are not taken into consideration.

Table 5: Implied growth in relative demand for skill and elasticity of substitution

	Naive	With efficiency adjustment	
	OLS	$\sigma = \sigma_{OLS}$	$\sigma = 2$
	(1)	(2)	(3)
γ_1	-0.10%	3.07%	3.26%
	(3.5%)	(0.13%)	(0.30%)
	[-7.94% , 7.73%]	[2.77% , 3.35%]	[2.60% , 3.93%]
σ	4.30	4.30	2.00
	(1.77)	-	-
	[0.34 , 8.25]	-	-

Note: Column 1 presents OLS estimates of Equation 24. Columns 2 and 3 presents estimates of Equation 15 for calibrated values of σ. Data: 2003 to 2015 Brazilian Household Survey (PNAD).

7. International Evidence of Changes in Skill Composition

So far, we have demonstrated that in Brazil, during a period of significant growth in college enrollment, there was also a reduction in the average quality of college workers. This shift can be partly attributed to the growing number of graduates from lower-quality institutions and students with lesser skills. As a result, the extent of skill-biased technical change is significantly underestimated if we do not account for changes in composition. In this section, we present evidence suggesting that this phenomenon may be occurring in other countries as well.

To accomplish this, we utilize harmonized data spanning an extended period for most countries worldwide. The primary objectives of this analysis are to demonstrate that (i) several countries have experienced significant growth in tertiary education over the past 20 to 40 years, and (ii) many of these nations have observed a decline in the average quality of education offered. In this analysis, we approximate the quality of education using the teacher-student ratio in tertiary education.[18]

Figure 6 illustrates a negative correlation between enrolment in tertiary education and the teacher-student ratio in Brazil. From 1970 to 2020, enrolment in tertiary education increased nearly tenfold, growing from less than one million students to almost nine million students. However, the number of teachers in tertiary education did not increase at the same rate, resulting in a 50% reduction in the teacher-student ratio during this period. This leads to a correlation between the two time series of -0.93. The correlation is more pronounced during the period of rapid growth, spanning from 1997 to 2018. This data supports the hypothesis that the expansion of tertiary education occurred concurrently with a decline in average quality.

Figure 7 presents the same correlation (shown on the vertical axis) that we discussed earlier for multiple countries and regions around the world. On the horizontal axis, we display the annual growth rate of enrollment in tertiary education. The figure reveals three key observations. First, a majority of countries exhibit a negative correlation, with several showing a particularly strong negative correlation. Secondly, most countries with a strong negative correlation experienced relatively high annual growth rates in tertiary enrolment, exceeding 3%. These two facts combined suggest that, for many countries, during periods of increasing tertiary enrolment, the average quality of tertiary education tends to decline.

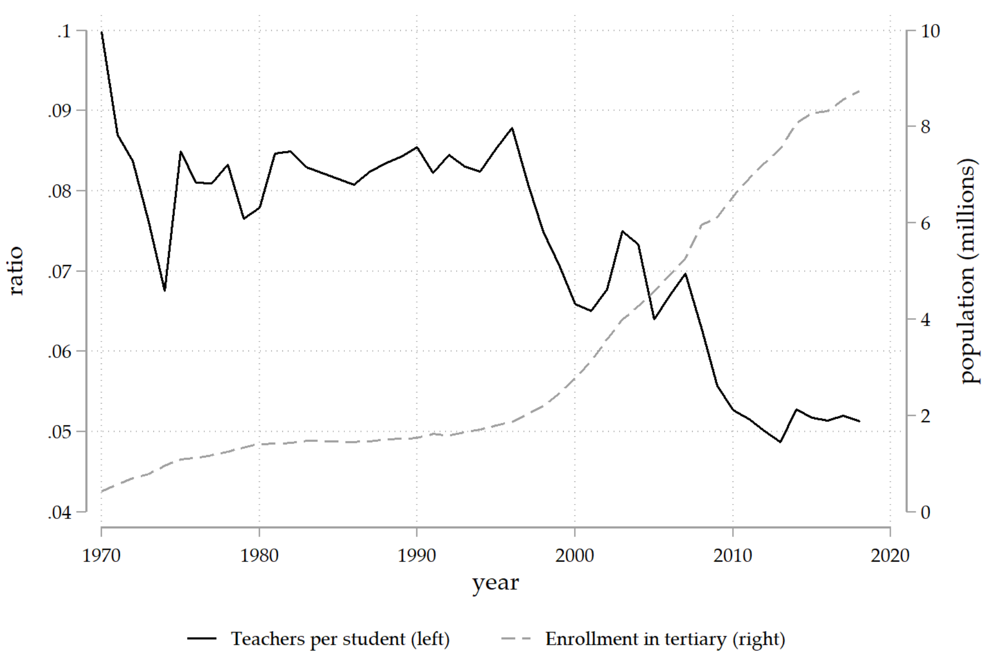
[18]While the student-to-teacher ratio is a widely accepted measure of quality for primary and secondary education, we acknowledge that it may not fully capture quality in tertiary education. We employ this variable as a proxy for quality and investment in tertiary education due to the absence of better quality metrics harmonized over long periods and across numerous countries.

Third, the group characterized by high growth and a negative correlation predominantly consists of low and middle-income countries. Appendix Tables B7 and B8 list these countries and regions, ranking them by their serial correlations between tertiary enrolment and the teacher-student ratio.

Similar to the situation documented for Brazil, the growth rate of relative demand for college-educated workers may be underestimated in various countries worldwide. The mechanism highlighted in this paper is particularly pertinent in nations that have witnessed substantial growth in tertiary enrollment and where there exists a strong negative correlation between growth and quality over time. As depicted in Figure 7, a considerable number of countries and regions fall into this category.

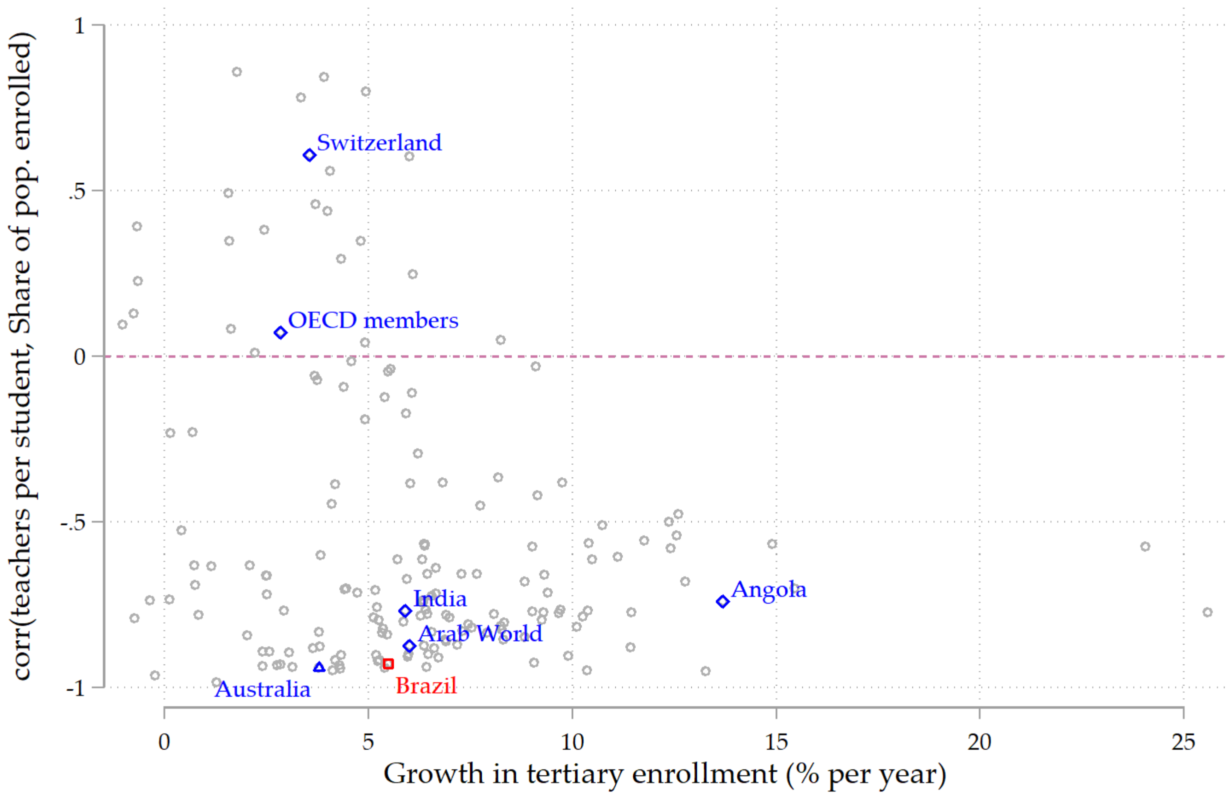
Interestingly, in developed countries, there is a notable positive correlation between the quality of tertiary education and growth in enrollment. In these nations, a rise in enrollment typically coincides with an enhancement in educational quality, as exemplified by Switzerland in Figure 7. In such scenarios, relying solely on the standard college premium to estimate the growth rate of skill-biased technical change could potentially result in an overestimation of both the increase in the price of skill and the extent of technical change itself.

Figure 6: Tertiary expansion and tertiary quality in Brazil



Note: The figure presents the time series of the number of teachers per student in tertiary education (left axis) and the number of people enrolled in tertiary education (right axis). The data is extracted from the UNESCO Institute for Statistics and accessed through the World Bank Open Data website.

Figure 7: Correlation between expansion and quality for different countries



Note: The figure presents the correlation between the number of teachers per student and the share of the population enrolled in tertiary education (vertical axis) and the annual growth rate of enrolment in tertiary education (horizontal axis). Each circle represents a country. The data is extracted from the UNESCO Institute for Statistics and accessed through the World Bank Open Data website. The time series varies across countries due to data constraints.

8. Conclusion

The decreasing college premium in Latin America over the past few last decades could be a cold shower for proponents of economic growth through investments in higher education. Over the last twenty years, the share of workers in Brazil with college degrees increased by 10 percentage points, while the college premium decreased by 40 percentage points, from 2.6 to 2.1. If we were to extrapolate this trend, an expansion in college education reaching the U.S. share of graduate workers could equalize the wages of college and high-school workers. While this could be seen as positive news in terms of reducing inequality, it would represent a failure of the higher education system in increasing productivity and generating economic growth.

In this paper, we provide evidence against the previous argument. We show that the price of skill is actually increasing and that compositional changes in the higher education sector, like decreasing average quality, are responsible for the decline in the college premium. To conduct this analysis, we gather unique data matching several cohorts of college graduates—from 42 Brazilian universities —to multiple years of labor market outcomes. The wage premium rose by 24% among graduates from a stable set of universities for which there were no changes in composition.

Using these results, we estimate a labor supply and demand model as in Katz and Murphy (1992) to verify the extent of skill-biased technical change in Brazil during this period. Our model estimation indicates a consistent rise in the relative demand for college-educated workers. The strength of skill-biased technical change is underestimated when changes in the composition are not accounted for.

In light of these results, one can conclude that the college premium is not a precise indicator for the price of skill during periods of education expansion. This is an important consideration for models of wage inequality that use trends in the college premium trends as target moments (Alvarez et al., 2018; Haanwinckel, 2020). Furthermore, our finding of strong growth in demand for skills during a period of tertiary expansion hints at an interesting path for future research: investigating endogenous technical change models, as in Carneiro et al. (2022), where the increase in college-educated workers encourages firms to adopt skill-complementary technologies.

References

Acemoglu, D. and D. Autor (2011). Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings. Volume 4 of *Handbook of Labor Economics*, pp. 1043–1171. Elsevier.

Acosta, P., G. Cruces, S. Galiani, and L. Gasparini (2019). Educational upgrading and returns to skills in latin america: evidence from a supply–demand framework. *Latin American Economic Review 28*, 1–20.

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018). Firms and the Decline in Earnings Inequality in Brazil. *American Economic Journal: Macroeconomics 10*(1), 149–89.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science 344*(6186), 843–851.

Barros, R., M. de Carvalho, S. Franco, and R. Mendonça (2010). Markets, the state, and the dynamics of inequality in Brazil. In L. F. López-Calva and N. Lustig (Eds.), *Declining inequality in Latin America: A decade of progress?*, pp. 74–134. Brookings Institution Press.

Blundell, R., D. A. Green, and W. Jin (2021). The U.K. as a Technological Follower: Higher Education Expansion and the College Wage Premium. *The Review of Economic Studies 89*(1), 142–180.

Camacho, A., J. Messina, and J. P. Uribe (2017). The Expansion of Higher Education in Colombia: Bad Students or Bad Programs? Documentos CEDE 15352, Universidad de los Andes, Facultad de Economía, CEDE.

Card, D. and T. Lemieux (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis*. *The Quarterly Journal of Economics 116*(2), 705–746.

Carneiro, P. and S. Lee (2011). Trends in Quality-Adjusted Skill Premia in the United States, 1960-2000. *American Economic Review 101*(6), 2309–49.

Carneiro, P., K. Liu, and K. G. Salvanes (2022). The Supply of Skill and Endogenous Technical Change: Evidence from a College Expansion Reform. *Journal of the European Economic Association 21*(1), 48–92.

Cox, A. (2024, January). From classroom to prosperity: Fostering development through higher education.

Dobbin, C., N. Barahona, and S. Oteroz (2022). The Equilibrium Effects of Subsidized Student Loans. Job Market Paper.

Feinmann, J. and R. H. Rocha (2024). Social mobility and higher education in brazil. *Available at SSRN 4702937*.

Fernández, M. and J. Messina (2018). Skill premium, labor supply, and changes in the structure of wages in Latin America. *Journal of Development Economics 135*, 555–573.

Ferreira, F., S. Firpo, and J. Messina (2017). Ageing Poorly? Accounting for the Decline in Earnings Inequality in Brazil, 1995-2012. IZA Discussion Papers 10656, Institute of Labor Economics (IZA).

Haanwinckel, D. (2020). Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution. Unpublished.

Jaume, D. (2021). The labor market effects of an educational expansion. *Journal of Development Economics 149*, 102619.

Katz, L. F. and K. M. Murphy (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Messina, J. and J. Silva (2017). *Wage inequality in Latin America: Understanding the past to prepare for the future*. The World Bank Group.

OECD (2023). *PISA 2022 Results (Volume I)*.

Rodriguez, J., S. Urzua, and L. Reyes (2016). Heterogeneous Economic Returns to Post-Secondary Degrees: Evidence from Chile. *Journal of Human Resources 51*(2), 416–460.

Appendices

A The Composition-Adjusted Labor Supply and Demand model (with Age)

The results presented in Figure 3 suggest that changes in age premiums are influenced by shifts in composition, especially as younger cohorts enter new institutions. Therefore, we follow Card and Lemieux (2001), Fernández and Messina (2018) and many others and extend the model presented in Section 6. by incorporating imperfect substitution among age groups of the same skill level. This extension is also prompted by evidence in the literature indicating changes in returns to potential experience over recent decades (Fernández and Messina, 2018).

In the model, low- and high-skill labor are combined to produce output, as presented by the production function below:

$$Y_t = \left[(A_{L,t}\mathcal{L}_t)^{\frac{\sigma-1}{\sigma}} + (A_{H,t}\mathcal{H}_t)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

where σ represents the elasticity of substitution across skill types, and $A_{L,t}$ and $A_{H,t}$ denote the factor augmenting technology terms for low and high-skill, respectively. In contrast to Section 6., workers of the same skill type but different ages are not perfect substitutes. The aggregate levels of low and high skill are represented by the equations below:

$$\mathcal{L}_t = \left[\sum_g \left(A_{L,g,t}\mathcal{L}_{g,t}\right)^{\frac{\theta_L-1}{\theta_L}}\right]^{\frac{\theta_L}{\theta_L-1}} \quad \text{and} \quad \mathcal{H}_t = \left[\sum_g \left(A_{H,g,t}\mathcal{H}_{g,t}\right)^{\frac{\theta_H-1}{\theta_H}}\right]^{\frac{\theta_H}{\theta_H-1}} \tag{16}$$

where $\mathcal{L}_{g,t}$ ($\mathcal{H}_{g,t}$) represents the number of low-skill (high-skill) efficiency units provided by workers of age group g at period t. The parameter θ_L (θ_H) represents the elasticity of substitution across age groups within low-skill (high-skill) workers. The terms $A_{H,g,t}$ and $A_{L,g,t}$ denote age-specific factor augmenting technologies.

The efficient use of resources implies that equilibrium prices (wages per efficiency unit of labor) are set equal to the marginal products of labor for each age group:

$$w_{K,g,t} = \frac{\partial Y_t}{\partial K_t} \times \frac{\partial K_t}{\partial K_{g,t}} \qquad ,\forall K \in \{\mathcal{L}, \mathcal{H}\},\ \forall g,\ \forall t \tag{17}$$

The goal of this approach is to properly estimate the evolution of the skill-biased technological change over time. The next sections will use the following two ratios to identify the parameters of interest: $\dfrac{w_{K,g,t}}{w_{K,j,t}}$ and $\dfrac{w_{H,g,t}}{w_{L,g,t}}$.

1.1 Naive estimation

First, we introduce the naive estimation of this model, which does not consider changes in composition. In the subsequent section, we introduce the composition adjustment to the model and outline how to estimate the model using our university sample.

Naive estimation of this model proceeds in two steps. First, we compare the evolution of wages across different age groups (g and j) of the same schooling level ($K \in \{\mathcal{L}, \mathcal{H}\}$) using Equation 17. This is expressed by the equation below.

$$ln\left(\frac{w_{K,g,t}}{w_{K,j,t}}\right) = \frac{\theta_K - 1}{\theta_K} ln\left(\frac{A_{K,g,t}}{A_{K,j,t}}\right) - \frac{1}{\theta_K} ln\left(\frac{K_{g,t}}{K_{j,t}}\right) \qquad , \forall K \in \{\mathcal{L}, \mathcal{H}\},\ \forall g > j,\ \forall t \quad (18)$$

We assume that the log ratio of age-specific factor augmenting technologies follow a log linear trend over time, relative to age group j:

$$ln\left(\frac{A_{K,g,t}}{A_{K,j,t}}\right) = \gamma^K_{0,g} + \gamma^K_{1,g} \times t + u_{K,g,t}$$

Combining the two equations above we have the following reduce-form equation:

$$ln\left(\frac{w_{K,g,t}}{w_{K,j,t}}\right) = \frac{\theta_K - 1}{\theta_K}\gamma^K_{0,g} + \frac{\theta_K - 1}{\theta_K}\gamma^K_{1,g} \times t - \frac{1}{\theta_K} ln\left(\frac{K_{g,t}}{K_{j,t}}\right) + u_{K,g,t} \quad (19)$$

Identification comes from the assumption that movements in $K_{g,t}$ represent exogenous supply shocks. We analyze seven age groups (30-34, 35-39, 40-44, 45-49, 50-54, 55-59, and 60-64) relative to the younger age group (25-29). By naively using average wages as skill prices and the number of workers in each age group as the aggregate efficiency for that age and skill group, the procedure estimates two elasticities of substitution across age groups (θ_L and θ_H), 14 constants ($\gamma^K_{0,g}, \forall g \neq j, \forall K \in \{\mathcal{L}, \mathcal{H}\}$), and 14 parameters for relative technology growth ($\gamma^K_{1,g}, \forall g \neq j, \forall K \in \{\mathcal{L}, \mathcal{H}\}$).

In the second step, also utilizing Equation 17, we compare the residual wages of the same age group g between skill groups H and L, as depicted in the equation below.[19]

$$ln\left(\widetilde{\frac{w_{H,g,t}}{w_{L,g,t}}}\right) = \frac{\sigma-1}{\sigma}ln\left(\frac{A_{H,t}}{A_{L,t}}\right) - \frac{1}{\sigma}ln\left(\frac{\mathcal{H}_t}{\mathcal{L}_t}\right) - \frac{1}{\theta_H}ln\left(\frac{\mathcal{H}_{g,t}}{\mathcal{H}_t}\right) + \frac{1}{\theta_L}ln\left(\frac{\mathcal{L}_{g,t}}{\mathcal{L}_t}\right) \qquad ,\forall g,\ \forall t \tag{20}$$

The evolution of relative factor-augmenting technologies is assumed to follow a log linear trend [20], which generates the following reduce-form equation:

$$ln\left(\widetilde{\frac{w_{H,g,t}}{w_{L,g,t}}}\right) = \alpha_0 + \frac{\sigma-1}{\sigma}\gamma_1 \times t - \frac{1}{\sigma}ln\left(\frac{\mathcal{H}_t}{\mathcal{L}_t}\right) - \frac{1}{\theta_H}ln\left(\frac{\mathcal{H}_{g,t}}{\mathcal{H}_t}\right) + \frac{1}{\theta_L}ln\left(\frac{\mathcal{L}_{g,t}}{\mathcal{L}_t}\right) + u_t \tag{21}$$

The equation above can be estimated by using average wages as skill prices and by the constructing aggregates H_t and L_t from the (naively) estimated parameters from step 1, according to Equation 16. Estimating the equation above provides estimates for σ and γ_1. Results are presented in section 1.4.

Next, we demonstrate how to incorporate the composition adjustment into the model and the estimation process.

1.2 Composition adjustment

The estimation in the previous section is biased due to differential changes in composition across different cohorts. To illustrate this point, we define $\bar{e}^K_{g,t}$ as the average number of efficiency units that workers of skill type $K \in \{L, H\}$ in age group g provide at time t. The issue arises when there is a differential evolution of the number of efficiency units ($N^K_{g,t}\bar{e}^K_{g,t}$) across age groups and skill groups. For instance, younger cohorts of college workers might have received lower-quality education or attended less prestigious institutions, resulting in a reduction of $\bar{e}^K_{g,t}$ over time. In this scenario, the number of workers in each group ($N^K_{g,t}$) and the average wage ($W_{K,g,t}$) do not correspond perfectly to the total number of efficiency units per group ($\mathcal{K}_{g,t}$) and to the wage per efficiency unit ($\omega_{K,g,t}$), respectively. To incorporate these characteristics into the model, and more specifically, into the step 1 of the estimation approach, we standardize these objects relative to age group j:

[19]where $ln\left(\widetilde{\frac{w_{H,g,t}}{w_{L,g,t}}}\right) = ln\left(\frac{w_{H,g,t}}{w_{L,g,t}}\right) - \frac{\theta_H-1}{\theta_H}ln(A_{H,g,t}) + \frac{\theta_L-1}{\theta_L}ln(A_{L,g,t})$.
We assume that $A_{K,g,0} = 1, \forall g, \forall K \in \{\mathcal{L}, \mathcal{H}\}$ and $A_{K,j,t} = 1, \forall t, \forall K \in \{\mathcal{L}, \mathcal{H}\}$,

[20]$ln(\frac{A_{H,t}}{A_{L,t}}) = \gamma_0 + \gamma_1 \times t + u_t$

$$ln\left(\frac{W_{K,g,t}}{W_{K,j,t}}\right) = ln\left(\frac{w_{K,g,t}}{w_{K,j,t}}\right) + ln\left(\frac{\bar{e}^K_{g,t}}{\bar{e}^K_{j,t}}\right) \qquad ,\forall K \in \{\mathcal{L}, \mathcal{H}\},\ \forall g > j,\ \forall t \tag{22}$$

$$ln\left(\frac{\mathcal{K}_{g,t}}{\mathcal{K}_{j,t}}\right) = ln\left(\frac{N^K_{g,t}}{N^K_{j,t}}\right) + ln\left(\frac{\bar{e}^K_{g,t}}{\bar{e}^K_{j,t}}\right) \qquad ,\forall K \in \{\mathcal{L}, \mathcal{H}\},\ \forall g > j,\ \forall t \tag{23}$$

Combining Equations 19, 22 and 23, and introducing the log linear framework for the evolution of relative factor-augmenting technologies, we formulate the composition-adjusted wage premium equation below:

$$\ln\left(\frac{W_{K,g,t}}{W_{K,j,t}}\right) = \frac{\theta_K - 1}{\theta_K}\gamma^K_{g,0} + \frac{\theta_K - 1}{\theta_K}\gamma^K_{g,1} \times t - \frac{1}{\theta_K}\ln\left(\frac{N^K_{g,t}}{N^K_{j,t}}\right) + \left(1 - \frac{1}{\theta_K}\right)\ln\left(\frac{\bar{e}^K_{g,t}}{\bar{e}^K_{j,t}}\right) + u_{K,g,t} \tag{24}$$

where $u_{K,g,t}$ account for measurement error. To obtain the composition-adjusted reduced-form equation, we derive the differences of the equation above with respect to $t = 0$.

$$\ln\left(\frac{W_{K,g,t}}{W_{K,j,t}}\right) = \beta^K_{g,0} + \frac{\theta_K - 1}{\theta_K}\gamma^K_{g,1}t - \frac{1}{\theta_K}\ln\left(\frac{N^K_{g,t}}{N^K_{j,t}}\right) + \underbrace{\left(1 - \frac{1}{\theta_K}\right)\left[\ln\left(\frac{\bar{e}^K_{g,t}}{\bar{e}^K_{j,t}}\right) - \ln\left(\frac{\bar{e}^K_{g,0}}{\bar{e}^K_{j,0}}\right)\right]}_{\Delta Skill-composition\ (Step\ 1)} + v_{K,g,t} \tag{25}$$

where $\beta^K_{g,0} = \ln\left(\frac{W_{K,g,0}}{W_{K,j,0}}\right) + \frac{1}{\theta_K}\ln\left(\frac{N^K_{g,0}}{N^K_{j,0}}\right)$. The "$\Delta$ Skill-composition" term above is inadvertently omitted when using wage and employment data to estimate Equation 19, which can result in biased estimates. Our first objective is to identify a data equivalent for this term.

Bias in estimation of Equation 21 (step 2) comes from two sources. First, we need to redefine the terms $H_{g,t}$ and H_t using unbiased estimates from step 1. Secondly, in order to use observed wages, we need to recognize that the log ratio of observed wages is a function of skill prices and average efficiency units of skill, as in the equation below:

$$ln\left(\frac{W_{H,g,t}}{W_{L,g,t}}\right) = ln\left(\frac{w_{H,g,t}}{w_{L,g,t}}\right) + ln\left(\frac{\bar{e}^H_{g,t}}{\bar{e}^L_{g,t}}\right) \qquad ,\forall g,\ \forall t \tag{26}$$

From Equations 21 and 26, we get:

$$ln\left(\frac{W_{H,g,t}}{W_{L,g,t}}\right) - ln\left(\frac{e^H_{g,t}}{e^L_{g,t}}\right) - \frac{\theta_H - 1}{\theta_H} ln(A_{H,g,t}) + \frac{\theta_L - 1}{\theta_L} ln(A_{L,g,t}) =$$

$$= \beta_0 + \frac{\sigma - 1}{\sigma}\gamma_1 \times t - \frac{1}{\sigma} ln\left(\frac{\mathcal{H}_t}{\mathcal{L}_t}\right) + -\frac{1}{\theta_H} ln\left(\frac{\mathcal{H}_{g,t}}{\mathcal{H}_t}\right) + \frac{1}{\theta_L} ln\left(\frac{\mathcal{L}_{g,t}}{\mathcal{L}_t}\right) + \mu_{g,t}$$

We derive the differences of the equation above with respect to $t = 0$.

$$ln\left(\widetilde{\frac{W_{H,g,t}}{W_{L,g,t}}}\right) = \delta_0 + \frac{\sigma - 1}{\sigma}\gamma_1 \times t - \frac{1}{\sigma} ln\left(\frac{\mathcal{H}_t}{\mathcal{L}_t}\right) - \frac{1}{\theta_H} ln\left(\frac{\mathcal{H}_{g,t}}{\mathcal{H}_t}\right) + \frac{1}{\theta_L} ln\left(\frac{\mathcal{L}_{g,t}}{\mathcal{L}_t}\right) + \underbrace{\left[ln\left(\frac{e^H_{g,t}}{e^L_{g,t}}\right) - ln\left(\frac{e^H_{g,0}}{e^L_{g,0}}\right)\right]}_{\Delta Skill-composition\ (Step\ 2)} + \nu_{g,t} \tag{27}$$

where

$$ln\left(\widetilde{\frac{W_{H,g,t}}{W_{L,g,t}}}\right) = ln\left(\frac{W_{H,g,t}}{W_{L,g,t}}\right) - \frac{\theta_H - 1}{\theta_H} ln\left(\frac{A_{H,g,t}}{A_{H,g,0}}\right) + \frac{\theta_L - 1}{\theta_L} ln\left(\frac{A_{L,g,t}}{A_{L,g,0}}\right)$$

$$\delta_0 = ln\left(\frac{W_{H,g,0}}{W_{L,g,0}}\right) + \frac{1}{\sigma} ln\left(\frac{\mathcal{H}_0}{\mathcal{L}_0}\right) + \frac{1}{\theta_H} ln\left(\frac{\mathcal{H}_{g,0}}{\mathcal{H}_0}\right) + -\frac{1}{\theta_L} ln\left(\frac{\mathcal{L}_{g,0}}{\mathcal{L}_0}\right)$$

1.3 Identification and estimation

Identification in this model relies on two key assumptions:

Assumption 1A: The average number of efficiency units supplied by low-skill workers remained constant over the period, for all age groups: $\bar{e}^L_{g,t} = \bar{e}^L_g$.

Assumption 2A: The average number of efficiency units per high-skill worker in our *University Sample* is constant over the period, for all age groups: $\bar{e}^{HU}_{g,t} = \bar{e}^{HU}_g$

The assumptions above are equivalent to those in section 6.. The main difference is the requirement they hold true for every age group g. Studies that ignore changes in efficiency composition are implicitly adopting Assumption 1A and a modified version of assumption 2A, in which the average number of efficiency units per high-skill worker in the full college sample is constant over the period. The current study relaxes the latter assumption. Instead, we impose Assumption 2A, which posits that different cohorts of graduates from the same university supply the same number of efficiency units.

We use the assumptions above to identify the composition adjustment terms from Equations 25 and 33, from estimation steps 1 and 2, respectively.

To identify the step-1 Δ Skill-composition, we differentiate Equation 22 with respect to $t = 0$:

$$\underbrace{\ln\left(\frac{W_{H,g,t}}{W_{H,j,t}}\right) - \ln\left(\frac{W_{H,g,0}}{W_{H,j,0}}\right)}_{\Delta \text{ wage premium}_g} = \underbrace{\ln\left(\frac{\omega_{H,g,t}}{\omega_{H,j,t}}\right) - \ln\left(\frac{\omega_{H,g,0}}{\omega_{H,j,0}}\right)}_{\Delta \text{ skill premium}_g} + \underbrace{\ln\left(\frac{\bar{e}^H_{g,t}}{\bar{e}^H_{j,t}}\right) - \ln\left(\frac{\bar{e}^H_{g,0}}{\bar{e}^H_{j,0}}\right)}_{\Delta \text{ Skill-composition (Step-1)}_g} \tag{28}$$

Assumption 2A implies that the observed change in wages of age group g relative to age group j within the *University Sample* does not account for changes in skill composition. Consequently, the change in wages identifies the changes in the skill premium of age group g relative to age group j.

$$\underbrace{\ln\left(\frac{W_{\mathbf{HU},g,t}}{W_{\mathbf{HU},j,t}}\right) - \ln\left(\frac{W_{\mathbf{HU},g,0}}{W_{\mathbf{HU},j,0}}\right)}_{\substack{\Delta \text{ wage premium}_g \text{in the} \\ \textit{University Sample}}} = \underbrace{\ln\left(\frac{\omega_{H,g,t}}{\omega_{H,j,t}}\right) - \ln\left(\frac{\omega_{H,g,0}}{\omega_{H,j,0}}\right)}_{\Delta \text{ skill premium}_g} \tag{29}$$

The difference between Equations 28 and 29 identifies changes in skill composition.

$$f(\mathbf{W}_g) = \Delta \text{ Skill-composition (Step-1)}_g = \substack{\Delta \text{ wage premium}_g \\ \text{in the } \textit{Full Sample}} - \substack{\Delta \text{ wage premium}_g \text{ in the} \\ \textit{University Sample}} \tag{30}$$

Using Equations 25 and 30, we write the step-1 composition adjusted reduced-form equation. This equation allows the unbiased estimation of the elasticity of substitution (θ_H) and the relative growth of age-specific factor-augmenting technologies ($\gamma^H_{g,1}$). The resulting equation is presented below, in which all terms are observable in the data.

$$\ln\left(\frac{W_{H,g,t}}{W_{H,j,t}}\right) = \beta^H_{g,0} + \frac{\theta_H - 1}{\theta_H}\gamma^H_{g,1}t - \frac{1}{\theta_H}\ln\left(\frac{N^H_{g,t}}{N^H_{j,t}}\right) + \left(1 - \frac{1}{\theta_H}\right) f(\mathbf{W}_g) + v_{H,g,t} \tag{31}$$

We analyze eight age groups: 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, and 60-64. Estimation is done relative to the 25-29 year age group and identifies seven relative growth parameters. Given Assumption 1A, the estimation of the low-skill elasticity of substitution (θ_L) and the relative growth of age-specific factor-augmenting technologies ($\gamma^L_{g,1}$) is identical to that in Equation 19. This is because the composition-adjustment term in Equation 25 equals zero.

Unbiased estimation of Equation 33 (step-2) requires: (i) construction of the parameters and variables δ_0, H_t, L_t, $H_{g,t}$, $L_{g,t}$, $ln(\frac{A_{H,g,t}}{A_{H,g,0}})$ and $ln(\frac{A_{L,g,t}}{A_{L,g,0}})$, and (ii) identification of the Δ Skill-composition (Step-2) term.

First, we normalize to one the following terms: $A_{H,g,t=0}$, $A_{L,g,t=0}$, $e^H_{g,t=0}$, and $e^L_{g,t=0}$ $\forall g$; $A_{H,j,t}$, $\forall t$ and $j = 25$ *to* 29. Assumption A1 guarantees the identification of average efficiency terms associated to low-skill workers ($\bar{e}^L_{g,t} = 1$). Assumption A2 guarantees the identification of average efficiency terms associated to high-skill workers ($\bar{e}^L_{g,t}$) as shown below:

$$\ln\left(W_{H,g,t}\right) - \ln\left(W_{H,g,0}\right) = \ln(\omega_{H,g,t}) - \ln(\omega_{H,g,0}) + \ln(\bar{e}^H_{g,t}) - \ln(\bar{e}^H_{g,0})$$

$$\ln\left(W_{\mathbf{HU},g,t}\right) - \ln\left(W_{\mathbf{HU},g,0}\right) = \ln(\omega_{H,g,t}) - \ln(\omega_{H,g,0})$$

The equation below identifies the average number of efficiency units of high-skill labor for all periods and age groups (including age group 25 to 29), relative to the normalization $e^H_{g,t=0} = 1$.

$$g(\mathbf{W}_g) = \ln\left(\frac{W_{H,g,t}}{W_{H,g,0}}\right) - \ln\left(\frac{W_{\mathbf{HU},g,t}}{W_{\mathbf{HU},g,0}}\right) = \ln\left(\bar{e}^H_{g,t}\right) \tag{32}$$

Given our assumptions, the term $\ln\left(\bar{e}^H_{g,t}\right)$ matches the Δ Skill-composition (Step-2) component in Equation 33. With this information we construct: (i) the supply of high-skill work for each age group, using $\mathcal{H}_{g,t} = N^H_{g,t}e^H_{g,t}$; (ii) the factor augmenting technological components $A^L_{g,t}$ and $A^H_{g,t}$, using the assumptions and estimates above; and (iii) the aggregates $\mathcal{L}_t$ and $\mathcal{H}_t$, using their definitions in Equation 16, estimated elasticities, and the construction of their components. The step 2 estimating equation that derives from Equation 33 is presented below.

$$ln\left(\widetilde{\frac{W_{H,g,t}}{W_{L,g,t}}}\right) = \delta_0 + \frac{\sigma-1}{\sigma}\gamma_1 \times t - \frac{1}{\sigma}ln\left(\frac{\mathcal{H}_t}{\mathcal{L}_t}\right) - \frac{1}{\theta_H}ln\left(\frac{\mathcal{H}_{g,t}}{\mathcal{H}_t}\right) + \frac{1}{\theta_L}ln\left(\frac{\mathcal{L}_{g,t}}{\mathcal{L}_t}\right) + g(\mathbf{W}_g) + \nu_{g,t} \tag{33}$$

where

$$ln\left(\widetilde{\frac{W_{H,g,t}}{W_{L,g,t}}}\right) = ln\left(\frac{W_{H,g,t}}{W_{L,g,t}}\right) - \frac{\theta_H - 1}{\theta_H}ln\left(\frac{A_{H,g,t}}{A_{H,g,0}}\right) + \frac{\theta_L - 1}{\theta_L}ln\left(\frac{A_{L,g,t}}{A_{L,g,0}}\right)$$

$$\delta_0 = ln\left(\frac{W_{H,g,0}}{W_{L,g,0}}\right) + \frac{1}{\sigma}ln\left(\frac{\mathcal{H}_0}{\mathcal{L}_0}\right) + \frac{1}{\theta_H}ln\left(\frac{\mathcal{H}_{g,0}}{\mathcal{H}_0}\right) + -\frac{1}{\theta_L}ln\left(\frac{\mathcal{L}_{g,0}}{\mathcal{L}_0}\right)$$

1.4 Estimation results

Appendix Table A1 presents the estimates for the main parameters of interest from the naive and the composition-adjusted estimations. The first panel presents the growth in demand for each age group of high-skill workers relative to the 25 to 29-year-old college graduates. The second panel presents the growth in demand for high-skill workers relative to low-skill workers. The regression coefficients can be found in Appendix Tables B5 and B6.

The results from the naive estimation, presented in column 1 of Appendix Table A1, indicate a surprising finding: the demand for high-skill workers shows a constant growth rate of 0.2% annually over the period from 2003 to 2015. This is unexpected, as we anticipated a stronger increase in skill demand during this time-frame. Even more notably, demand for older cohorts of high-skill workers is increasing relative to younger cohorts. For instance, workers aged 60 to 64 experience a 1.9% growth in demand relative to those aged 30 to 34, amounting to a cumulative 25% increase over the 12-year period. For simplicity, we calibrate all elasticities of substitution to five in this analysis.

Column 3 of Appendix Table A1 presents a markedly different perspective when adjusting for changes in the composition of college workers. The demand for high-skill workers shows an annual growth of 7.2% relative to low-skill workers. This growth is particularly pronounced among younger cohorts, who typically receive lower wages. Combining the estimates of $\gamma_{1,c}^{H}$ and γ_1, we observe that the relative demand for high-skill workers aged 60 to 64 remains unchanged over the analyzed period. In column 5, we re-calibrate the model parameters using elasticities derived from Fernández and Messina (2018). Interestingly, these estimates are found to be insensitive to such adjustments.

Table A1: Estimated and calibrated parameters (PNAD 2003-2015)

	Naive		With Efficiency Adjustment			
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
	(1)	(2)	(3)	(4)	(5)	(6)
Estimates from Step 1						
$\gamma^H_{1,g}$ (relative to age 25-29)						
30-34 y.o.	-0.0063	(0.0018)	-0.0002	(0.0046)	-0.0022	(0.0039)
35-39 y.o.	-0.0069	(0.0027)	-0.0227	(0.0092)	-0.0230	(0.0075)
40-44 y.o.	-0.0168	(0.0031)	-0.0520	(0.0098)	-0.0475	(0.0085)
45-49 y.o.	-0.0165	(0.0022)	-0.0776	(0.0054)	-0.0732	(0.0049)
50-54 y.o.	-0.0064	(0.0021)	-0.0826	(0.0061)	-0.0799	(0.0052)
55-59 y.o.	0.0088	(0.0031)	-0.0708	(0.0065)	-0.0730	(0.0059)
60-64 y.o.	0.0125	(0.0028)	-0.0713	(0.0070)	-0.0752	(0.0064)
Estimates from Step 2						
γ_1 (relative to low-skill)	0.0023	(0.0016)	0.0726	(0.0032)	0.0778	(0.0098)
Calibrated elasticities						
θ_L	5		5		3.6	
θ_H	5		5		10.8	
σ	5		5		2.3	

Note: The table presents the estimated relative growth parameters ($\gamma_{1,g}$ and γ_1) in three different specifications. Column 1 presents OLS estimates of Equations 19 and 21. Columns 3 and 5 presents estimates of Equations 31 and 33 for calibrated values of θ_L, θ_H and σ. The elasticities used in Column 5 are derived from Fernández and Messina (2018). In this analysis, we use data from the Brazilian Household Survey (PNAD) between 2003 and 2015 combined with the estimates presented in Figure 3.

B Appendix Tables

Table B1: Comparing Matching Algorithms

Algorithm	Hyper Parameters		Algorithm Quality: Training sample (50%)		Algorithm Quality: Test sample (50%)	
	b1	b2	PPV	TPR	PPV	TPR
Logit	0.1	1.25	87.10%	92.00%	87.20%	92.20%
	0.05	1	86.70%	92.50%	87.10%	93.00%
Random Forest	0.1	1.25	96.60%	97.20%	88.80%	90.30%
	0.05	1	96.00%	98.70%	86.70%	92.70%

Note: Hyper parameters b1 and b2 are the threshold for whether the match's score is sufficiently large and the threshold for whether the ratio between the best and the second-best scores is sufficiently large, respectively. Positive Predictive Value (PPV or Accuracy): number of true positives over total number of positives. True Positive Rate (TPR or Efficiency): number of true positives over total number of correct cases.

Table B2: Confusion Matrix—Out of Sample Predictions

2[4]*Algorithm Prediction	True Status		2[4]*Total
	False	Correct	
Not Matched	13,613	691	14,304
Matched	1,362	8,900	10,262
Total	14,975	9,591	24,566

Note: The table presents the sample sizes from the logit model with b1=0.05 and b2=1.1. Positive Predictive Value (PPV or Accuracy): number of true positives over total number of positives = 8900/(8900+1362) = 86.7%. True Positive Rate (TPR or Efficiency): number of true positives over total number of correct cases = 8900/(8900+691)= 92.8%.

Table B3: Effects of School Quality on Wages by Years Since Graduation

Dep. Var	Log (Wage) after "X" years of graduation							
	4 years	5 years	6 years	7 years	8 years	9 years	10 years	11 years
ENADE	0.077***	0.084***	0.088***	0.088***	0.086***	0.086***	0.088***	0.093***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Observations	22,327	43,488	63,853	61,151	58,102	53,984	34,346	16,219
Cohorts	2,015	2014-2015	2013-2015	2012-2014	2011-2013	2010-2012	2010-2011	2010
R-Squared	0.03	0.04	0.05	0.05	0.04	0.04	0.05	0.05
Ranking RUF	-0.002***	-0.002***	-0.002***	-0.002***	-0.003***	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	12,637	25,017	36,753	35,683	34,455	33,243	21,656	10,461
Cohorts	2,015	2014-2015	2013-2015	2012-2014	2011-2013	2010-2012	2010-2011	2010
R-Squared	0.05	0.06	0.07	0.09	0.11	0.12	0.13	0.14

Notes: All specifications inclue year fixed effects. The ENADE score is the starndardized weighted average of all the major-level ENADE scores for an institution between 2010 and 2020.

Table B4: Reduced-form regressions

Dependent variable: ln(*college premium*)

	Naive OLS	With efficiency adjustment $\sigma = \sigma_{OLS}$	With efficiency adjustment $\sigma = 2$
	(1)	(2)	(3)
t	-0.001	0.024	0.0163
	(0.002)	(0.001)	(0.002)
$\ln\left(\frac{N_t^H}{N_t^L}\right)$	-0.232	-0.233	-0.500
	(0.096)	-	-
Efficiency Adjustment		0.767	0.500
		-	-
Constant	0.711	0.731	0.487
	(0.099)	-	-
Observations	13	13	13

Note: Column 1 presents OLS estimates of Equation 24. Column 2 presents the constrained linear regression estimates of Equation 15. Data: 2001 to 2015 Brazilian Household Survey (PNAD).

Table B5: Reduced-form regressions (Step 1)

	Naive		Adjusted	Naive		Adjusted
	HS	College	College	HS	College	College
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(N^H_{g,t}/N^H_{j,t})$	-0.2000	-0.2000	-0.2000	-0.2800	-0.0930	-0.0930
	-	-	-	-	-	-
$f(\mathbf{W}_g)$			0.8000			0.9070
			-			-
$t\times$ Age group						
30-34 y.o.	-0.0073	-0.0050	-0.0001	-0.0072	-0.0038	-0.0020
	(0.0007)	(0.0015)	(0.0037)	(0.0010)	(0.0013)	(0.0036)
35-39 y.o.	-0.0106	-0.0055	-0.0181	-0.0109	-0.0025	-0.0208
	(0.0013)	(0.0022)	(0.0074)	(0.0014)	(0.0018)	(0.0069)
40-44 y.o.	-0.0125	-0.0134	-0.0416	-0.0121	-0.0098	-0.0431
	(0.0017)	(0.0025)	(0.0078)	(0.0017)	(0.0024)	(0.0077)
45-49 y.o.	-0.0114	-0.0132	-0.0621	-0.0094	-0.0114	-0.0664
	(0.0020)	(0.0018)	(0.0044)	(0.0019)	(0.0018)	(0.0045)
50-54 y.o.	-0.0025	-0.0051	-0.0660	0.0010	-0.0051	-0.0725
	(0.0027)	(0.0018)	(0.0049)	(0.0027)	(0.0017)	(0.0048)
55-59 y.o.	-0.0029	0.0071	-0.0566	0.0013	0.0046	-0.0662
	(0.0027)	(0.0025)	(0.0052)	(0.0025)	(0.0026)	(0.0054)
60-64 y.o.	-0.0010	0.0101	-0.0570	0.0023	0.0072	-0.0681
	(0.0054)	(0.0023)	(0.0056)	(0.0052)	(0.0023)	(0.0058)
Constant						
30-34 y.o.	-0.0666	-0.1173	-0.1588	0.1568	-0.3410	-0.3582
	(0.0386)	(0.0185)	(0.0514)	(0.0375)	(0.0179)	(0.0524)
35-39 y.o.	0.0592	-0.0044	-0.0437	0.2687	-0.2247	-0.2358
	(0.0393)	(0.0207)	(0.0719)	(0.0380)	(0.0200)	(0.0700)
40-44 y.o.	0.1502	0.1058	0.0786	0.3430	-0.1133	-0.1305
	(0.0399)	(0.0222)	(0.0696)	(0.0387)	(0.0225)	(0.0717)
45-49 y.o.	0.1931	0.1602	0.2100	0.3550	-0.0377	0.0207
	(0.0402)	(0.0207)	(0.0520)	(0.0388)	(0.0210)	(0.0535)
50-54 y.o.	0.1445	0.1853	0.2222	0.2633	0.0329	0.0672
	(0.0426)	(0.0205)	(0.0572)	(0.0415)	(0.0198)	(0.0583)
55-59 y.o.	0.0854	0.1049	0.1565	0.1472	0.0258	0.0796
	(0.0436)	(0.0235)	(0.0653)	(0.0420)	(0.0236)	(0.0677)
Constant	0.2579	0.4156	0.3636	0.0239	0.6240	0.5616
	(0.0380)	(0.0159)	(0.0414)	(0.0368)	(0.0161)	(0.0435)
Observations	133	133	84	133	133	84

Note: Columns 1, 2, 4 and 5 present constrained linear regression estimates of Equation 19. Columns 3 and 6 present constrained linear regression estimates of Equation 31. Data: 2003 to 2015 Brazilian Household Survey (PNAD).

Table B6: Reduced-form regressions (Step 2)

	Naive		With Efficiency Adjustment			
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
	(1)	(2)	(3)	(4)	(5)	(6)
t	0.0019	(0.0013)	0.0581	(0.0026)	0.0444	(0.0056)
$ln\left(\frac{\mathcal{H}_t}{\mathcal{L}_t}\right)$	-0.2	-	-0.2	-	-0.43	-
$ln\left(\frac{\mathcal{L}_{g,t}}{\mathcal{L}_t}\right)$	0.2	-	0.2	-	0.28	-
$ln\left(\frac{\mathcal{H}_{g,t}}{\mathcal{H}_t}\right)$	-0.2	-	-0.2	-	-0.093	-
$g(\mathbf{W}_g)$	-	-	1	-	1	-
Constant	0.732	(0.0088)	0.724	(0.0182)	0.918	(0.0394)
Calibrated elasticities						
θ_L	5		5		3.6	
θ_H	5		5		10.8	
σ	5		5		2.3	

Note: Column 1 presents estimates of Equation 21. Columns 3 and 5 present the constrained linear regression estimates of Equation 33. Data: 2003 to 2015 Brazilian Household Survey (PNAD).

Table B7: List of countries by serial correlation between teacher-student ratio and enrollment in tertiary education, and tertiary enrollment growth

Group 1: positive correlation		
Country - Region	Correlation	Growth
Japan	0.86	1.8
Uruguay	0.84	3.9
Panama	0.80	4.9
Croatia	0.78	3.4
Switzerland	0.61	3.6
Korea, Rep.	0.60	6.0
Lebanon	0.56	4.1
Italy	0.49	1.6
Spain	0.46	3.7
Austria	0.44	4.0
Puerto Rico	0.39	-0.7
Mexico	0.35	4.8
Cuba	0.29	4.3
Congo, Rep.	0.25	6.1
Georgia	0.23	-0.6
Aruba	0.13	-0.8
Bosnia and Herzegovina	0.10	-1.0
Germany	0.08	1.6
Cote d'Ivoire	0.05	8.3
Albania	0.04	4.9
American Samoa	0.01	2.2

Group 2: weak negative correlation		
Country - Region	Correlation	Growth
Portugal	-0.02	4.6
Eritrea	-0.03	9.1
Samoa	-0.04	5.6
Hong Kong SAR, China	-0.05	5.5
Norway	-0.06	3.7
Philippines	-0.07	3.7
Andorra	-0.09	4.4
Nicaragua	-0.11	6.1
Malta	-0.12	5.4
Iraq	-0.17	5.9
Madagascar	-0.19	4.9
France	-0.23	0.7
Estonia	-0.23	0.1
Guatemala	-0.29	6.2
Cambodia	-0.37	8.2
Venezuela, RB	-0.38	6.8
Namibia	-0.38	9.8
Peru	-0.38	6.0
Kyrgyz Republic	-0.39	4.2
Togo	-0.42	9.1
Ecuador	-0.45	4.1
Morocco	-0.45	7.7
Lao PDR	-0.48	12.6

Group 3: medium negative correlation		
Country - Region	Correlation	Growth
China	-0.50	12.4
Ethiopia	-0.51	10.7
Ukraine	-0.53	0.4
Burkina Faso	-0.54	12.6
Chad	-0.56	11.8
Bhutan	-0.56	10.4
Myanmar	-0.57	6.4
New Caledonia	-0.57	14.9
Singapore	-0.57	6.4
Iran, Islamic Rep.	-0.57	9.0
Oman	-0.58	24.0
United Arab Emirates	-0.58	12.4
Malawi	-0.60	3.8
Benin	-0.61	11.1
Niger	-0.61	10.5
Iceland	-0.61	5.7
Afghanistan	-0.61	6.3
Bermuda	-0.63	0.7
Russian Federation	-0.63	1.2
Qatar	-0.64	6.7
Kuwait	-0.66	7.3
Bangladesh	-0.66	6.5
Mauritius	-0.66	7.7
Tunisia	-0.66	9.3
Netherlands	-0.66	2.5
Bulgaria	-0.66	2.5
Mozambique	-0.68	12.8
Jordan	-0.68	8.8
Azerbaijan	-0.69	0.8
Cabo Verde	-0.70	15.4
Guyana	-0.70	4.5
Slovenia	-0.70	4.4
Ireland	-0.71	5.2
Tanzania	-0.71	9.4
El Salvador	-0.71	4.7
Sierra Leone	-0.72	6.7
Kazakhstan	-0.72	2.5
Moldova	-0.73	0.1
Canada	-0.74	-0.4
Honduras	-0.74	6.4
Angola	-0.74	13.7
Macao SAR, China	-0.74	6.3

Group 4: strong negative correlation-low growth		
Country - Region	Correlation	Growth
United Kingdom	-0.77	2.9
Serbia	-0.78	0.8
Uzbekistan	-0.79	-0.7
Belarus	-0.84	2.0
Belgium	-0.89	2.4
Sweden	-0.89	2.6
Lithuania	-0.93	2.8
Armenia	-0.96	-0.2
Slovak Republic	-0.98	1.3

Group 5: strong negative correlation		
Country - Region	Correlation	Growth
Greece	-0.76	5.2
Sri Lanka	-0.77	6.4
Burundi	-0.77	9.7
Saudi Arabia	-0.77	10.4
India	-0.77	5.9
Malaysia	-0.77	9.0
Botswana	-0.77	11.5
Cyprus	-0.77	9.3
Djibouti	-0.77	25.6
Vietnam	-0.77	9.7
Turkiye	-0.78	8.1
Luxembourg	-0.78	6.9
Mauritania	-0.78	6.3
Cameroon	-0.78	10.3
Algeria	-0.79	7.0
Brunei Darussalam	-0.80	9.2
Eswatini	-0.80	5.9
Congo, Dem. Rep.	-0.80	8.3
Thailand	-0.81	7.4
Bahrain	-0.82	10.1
Senegal	-0.82	7.5
Mali	-0.82	8.3
West Bank and Gaza	-0.83	7.3
North Macedonia	-0.83	3.8
Argentina	-0.83	5.3
Indonesia	-0.84	7.9
Egypt, Arab Rep.	-0.84	5.5
Lesotho	-0.85	8.8
Guinea	-0.86	8.3
Nepal	-0.87	7.2
Mongolia	-0.87	6.4
Finland	-0.87	3.8
Rwanda	-0.88	11.4
New Zealand	-0.88	3.6
Latvia	-0.89	3.1
St. Lucia	-0.90	6.0
Czechia	-0.90	4.3
Tajikistan	-0.90	5.2
Ghana	-0.90	9.9
Hungary	-0.92	4.2
Uganda	-0.92	9.1
Brazil	-0.93	5.5
Poland	-0.93	4.3
Colombia	-0.94	6.4
Australia	-0.94	3.8
Romania	-0.94	4.3
Belize	-0.95	10.4
Libya	-0.95	13.3

Note:

Table B8: List of regions by serial correlation between teacher-student ratio and enrollment in tertiary education, and tertiary enrollment growth

Group 1: positive correlation		
Country - Region	Correlation	Growth
High income	0.38	2.45
North America	0.35	1.59
OECD members	0.07	2.86

Group 2: weak negative correlation		
Country - Region	Correlation	Growth

Group 3: medium negative correlation		
Country - Region	Correlation	Growth
Middle East & North Africa (excluding high income)	-0.57	6.39
Euro area	-0.63	2.09
South Asia	-0.67	5.94
Middle East & North Africa	-0.72	6.55

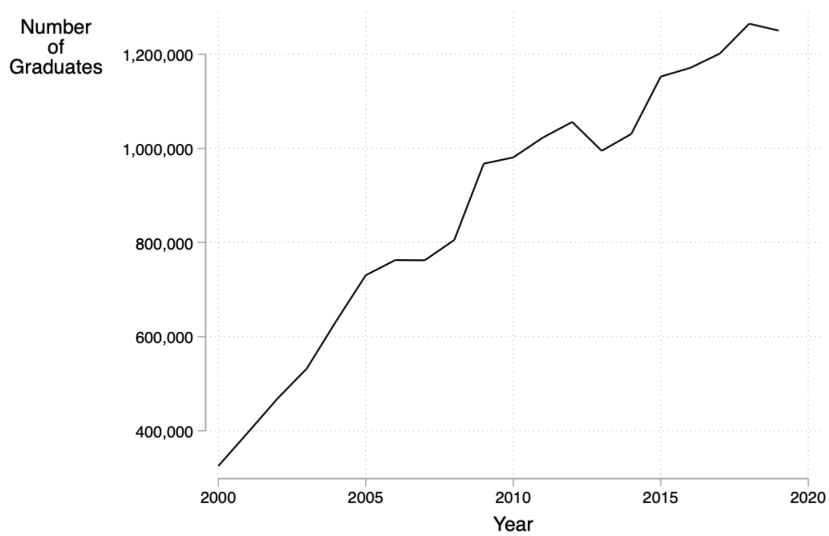
Group 4: strong negative correlation-low growth		
Country - Region	Correlation	Growth
Europe & Central Asia (excluding high income)	-0.93	2.84
Europe & Central Asia	-0.93	2.41

Group 5: strong negative correlation		
Country - Region	Correlation	Growth
Least developed countries: UN classification	-0.78	6.44
Lower middle income	-0.79	5.13
Latin America & Caribbean	-0.79	5.26
East Asia & Pacific (excluding high income)	-0.81	8.25
Latin America & Caribbean (excluding high income)	-0.82	5.36
Central African Republic	-0.83	6.56
Sub-Saharan Africa (excluding high income)	-0.85	6.90
Sub-Saharan Africa	-0.86	6.91
Arab World	-0.87	6.01
East Asia & Pacific	-0.88	6.63
Low income	-0.90	6.48
Global Partnership for Education	-0.91	5.96
Heavily indebted poor countries (HIPC)	-0.91	6.72
Middle income	-0.92	5.27
Low & middle income	-0.92	5.23
European Union	-0.94	3.14
Upper middle income	-0.94	5.41
World	-0.95	4.13

Note:

C Appendix Figures

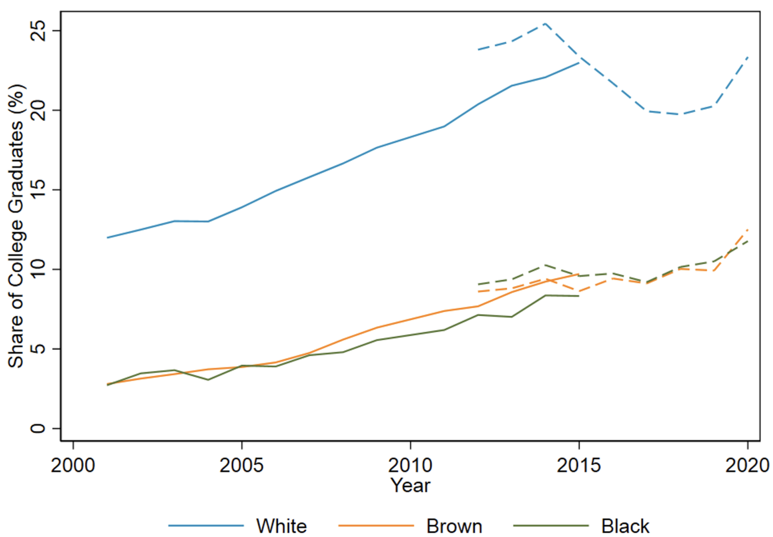
Figure C1: Number of College Graduates



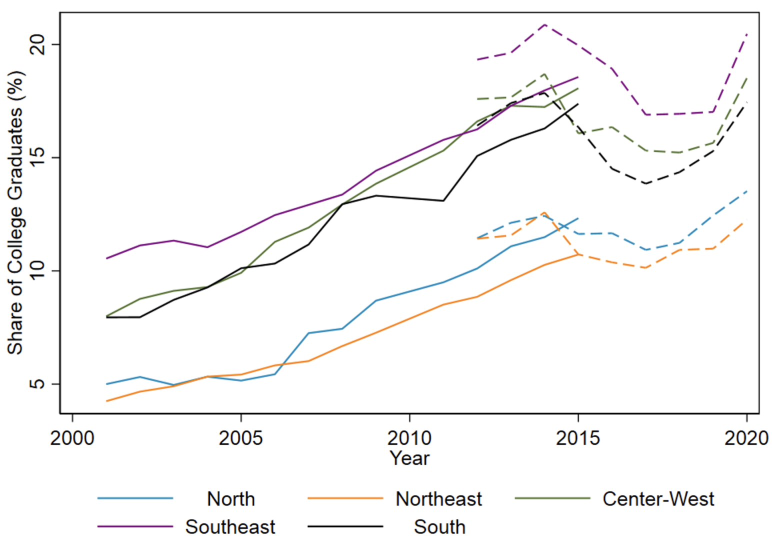
Note: The figure presents the number of students graduating with a college degree each year. Source: Higher Education Census (2000-2018)

Figure C2: Share of Graduates by Race and Region

(a) Share by Race



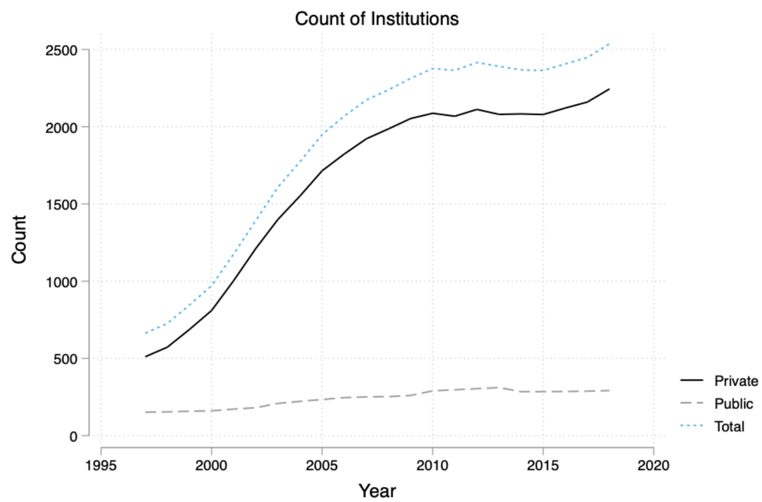
(b) Share by Geographical Region



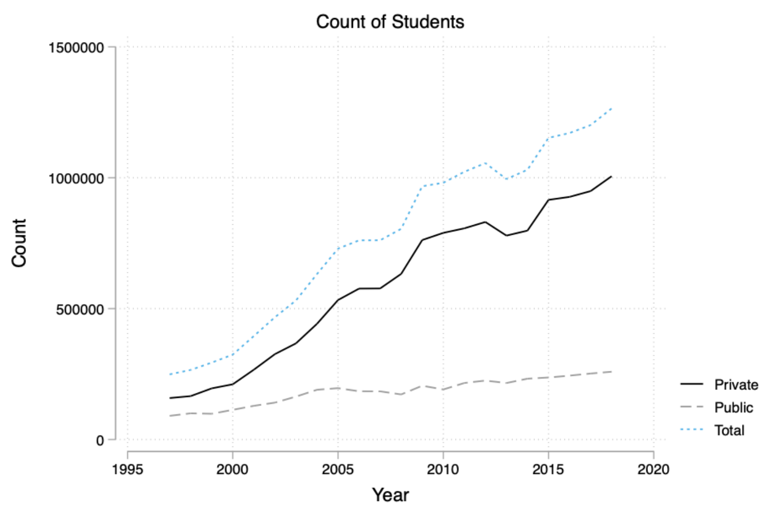
Note: The figure shows the share of working-age employees who graduated college, using PNAD data. The same filters as described in the caption of Figure 1 apply. The dashed lines represent data from *PNAD Contínua*, which replaced the annual PNAD survey in 2015.

Figure C3: Number of Higher Education Institutions and Students by University Type

(a) Institutions



(b) Graduating Students

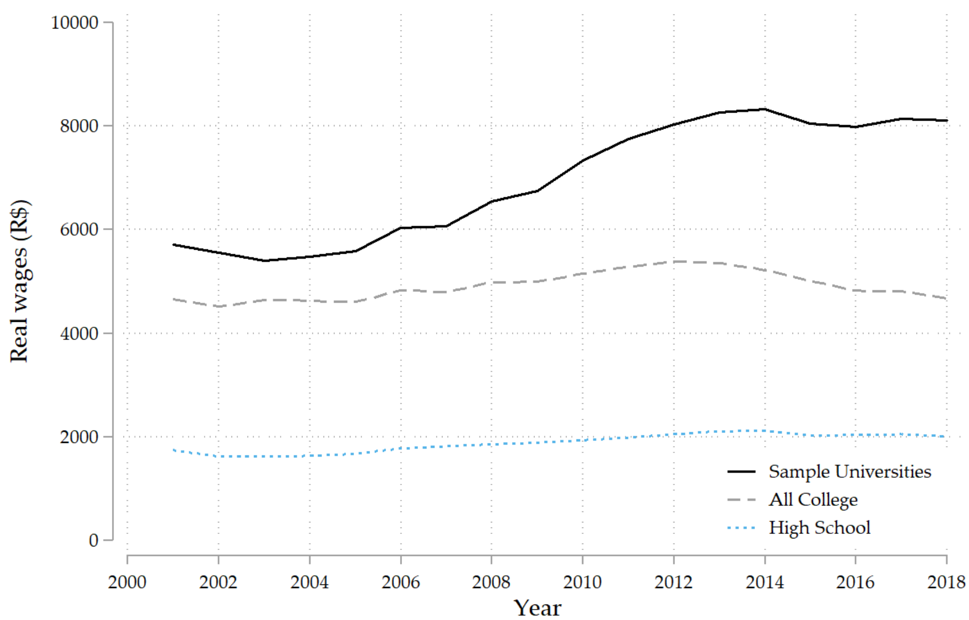


Note: The figures present the number of higher education institutions and graduating students each year by the type of university, private and public respectively. Source: Higher Education Census

Figure C4: Evolution of the creation of new majors

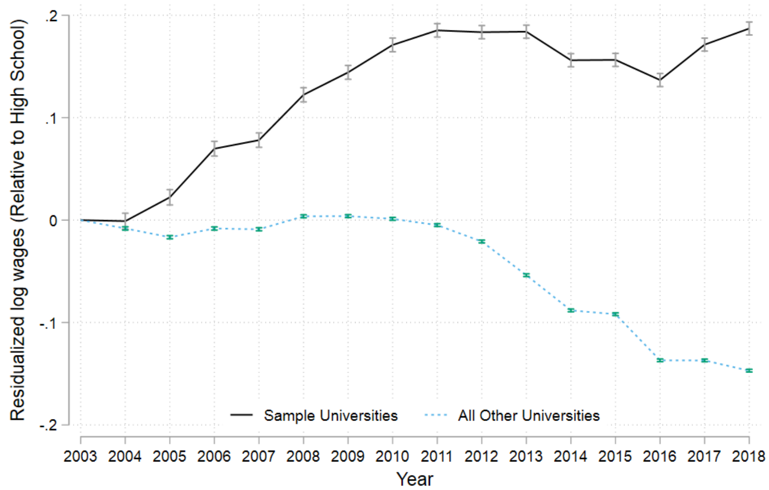
Note: The figure presents the number of new majors created each year. Source: Higher Education Census

Figure C5: Trends in real wages by education



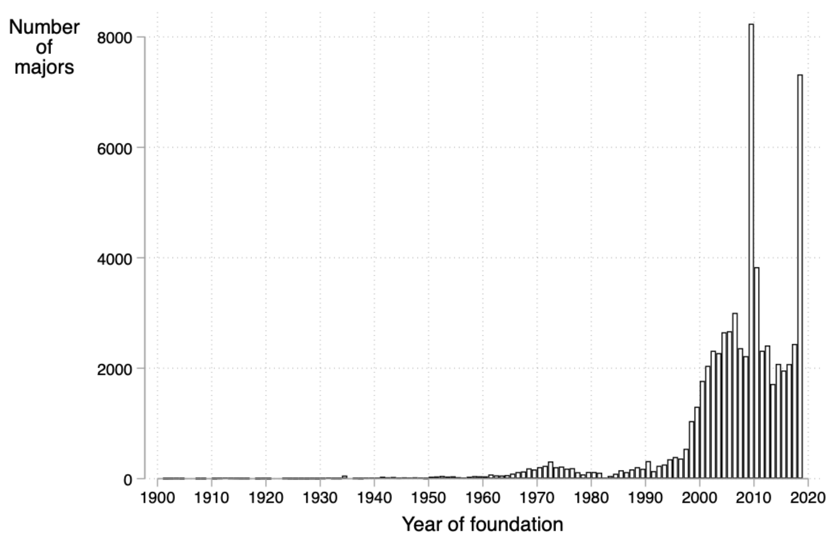
Note: The figure presents wage trends for three distinct groups: the college graduate sample from FOIL requests (Sample Universities), the sample of all workers in the RAIS dataset with a college degree (excluding those in the Sample Universities, referred to as All Other Universities), and workers with only a high school degree. The sample comprises 295,809,393 observations.

Figure C6: Trends in residualized log wages (college premium)



Note: The figure illustrates the progression of the college premium, relative to 2003 values, for two distinct groups: the college graduate sample from FOIL requests (Sample Universities) and the sample of all workers in the RAIS dataset with a college degree, excluding workers in the Sample Universities (All Other Universities). Both curves are benchmarked against the same trends in high school residualized wages. The estimates are derived from modeling an equation similar to Equation 1, but incorporating university and municipality fixed effects instead of university x major fixed effects. The sample comprises 295,809,393 observations. Additionally, the figure includes 95% confidence intervals for each point estimate.

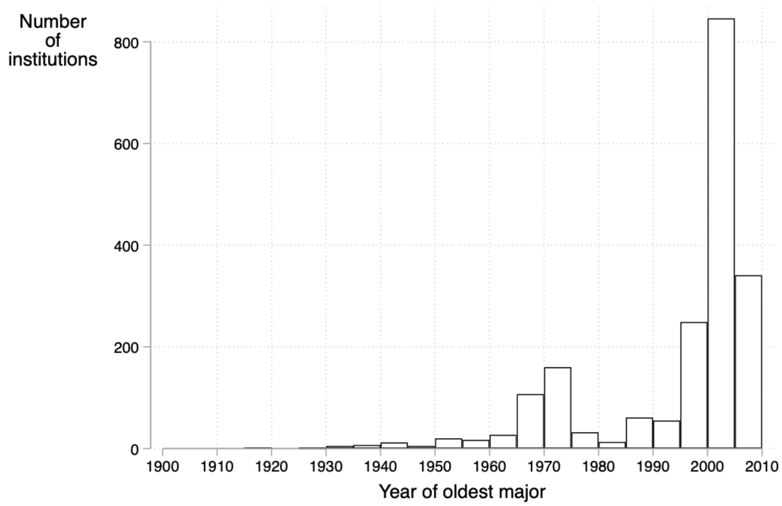
Figure C7: Evolution of the creation of new majors



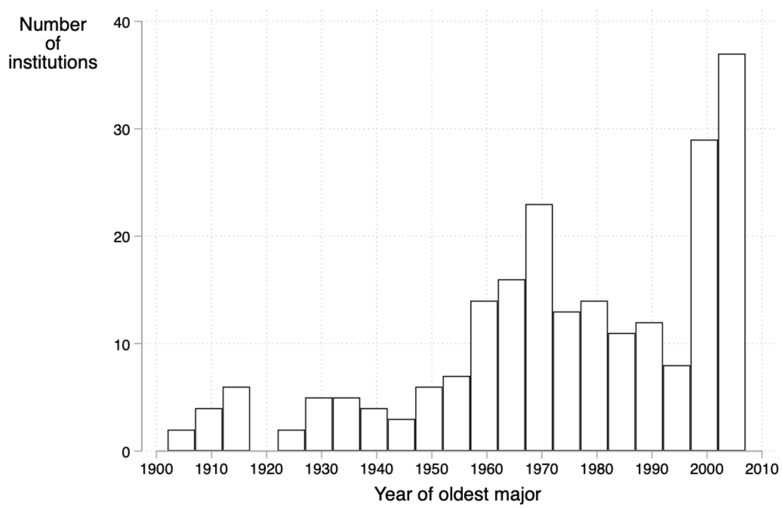
Note: The figure presents the number of new majors created each year. Source: Higher Education Census

Figure C8: Evolution of the creation of new majors by university type

(a) Private universities

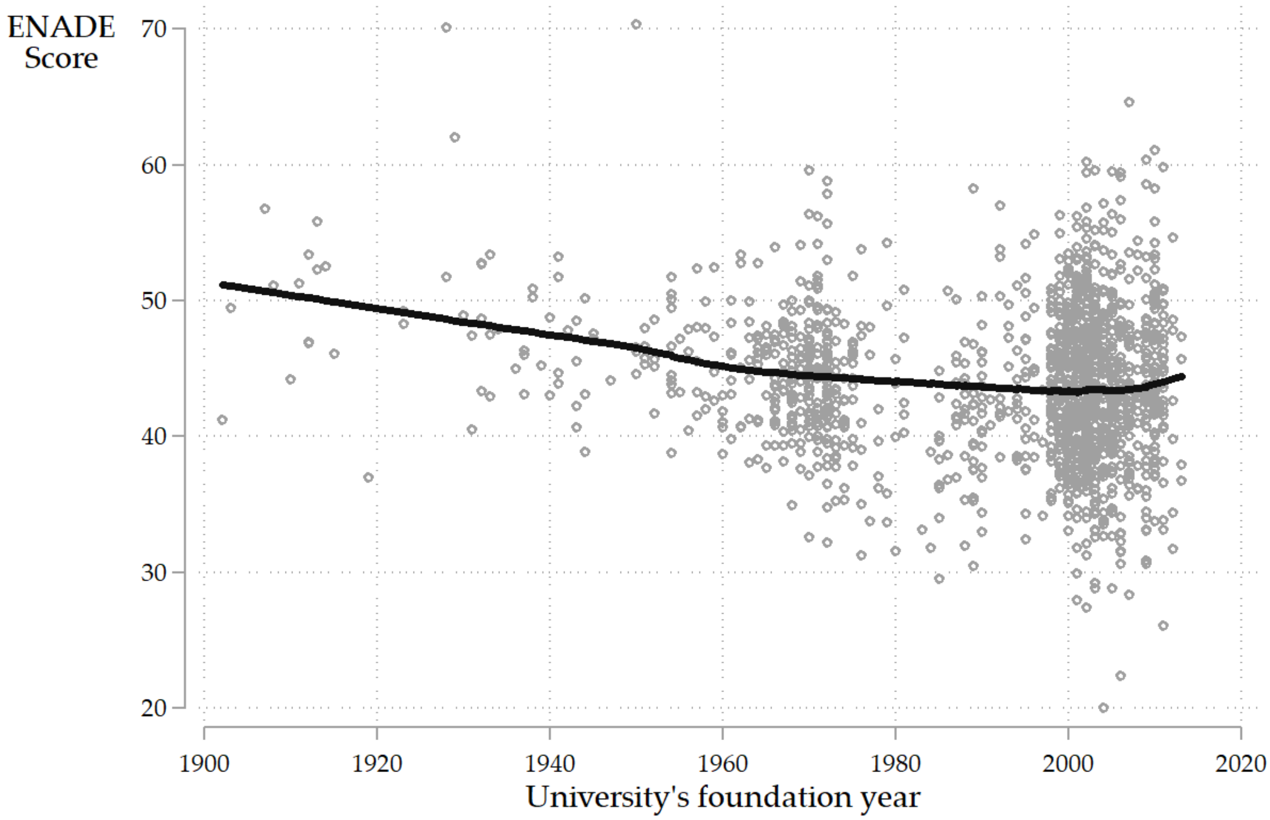


(b) Public universities



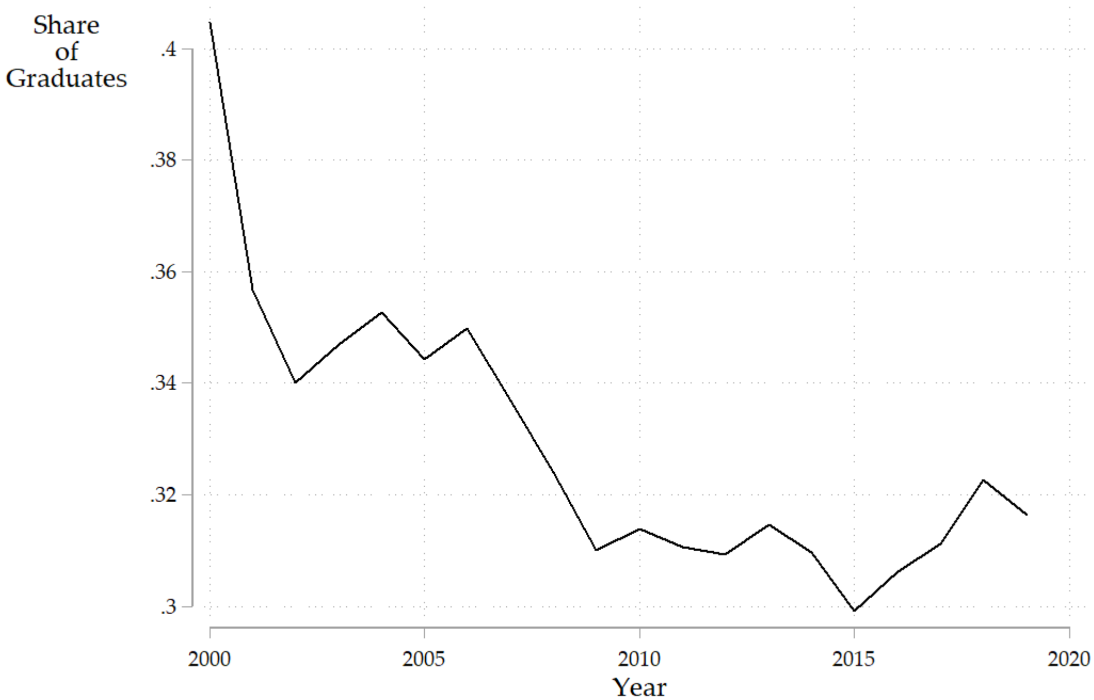
Note: The figures present the number of new majors created each year by the type of university, private and public respectively. Source: Higher Education Census

Figure C9: ENADE Score and University's foundation year. Nonparametric regression



Note: ENADE scores per university are constructed as the average across all students in all majors that had an ENADE exam in 2014. University's foundation year is the year of the foundation of the earliest major of the university. The bandwidth for the nonparametric regression is 0.8.

Figure C10: Share of graduates from Top 25 Universities according to ENADE ranking



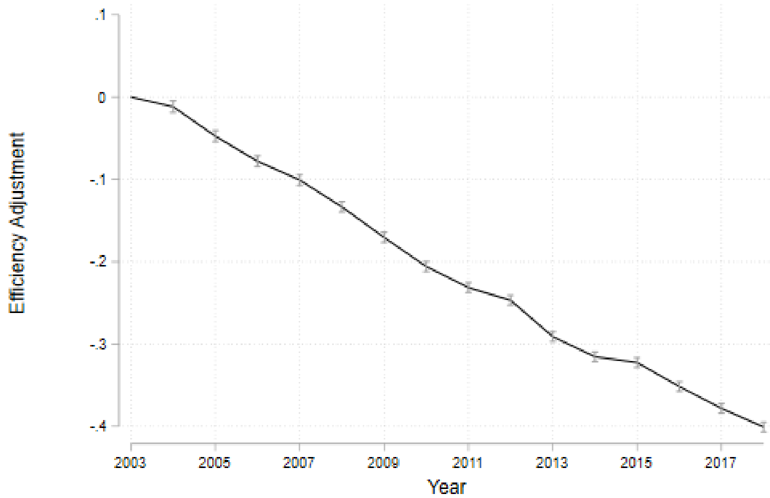
Note: The number of graduates from each university comes from the higher education census.

Figure C11: Government funding per tertiary student - Brazil

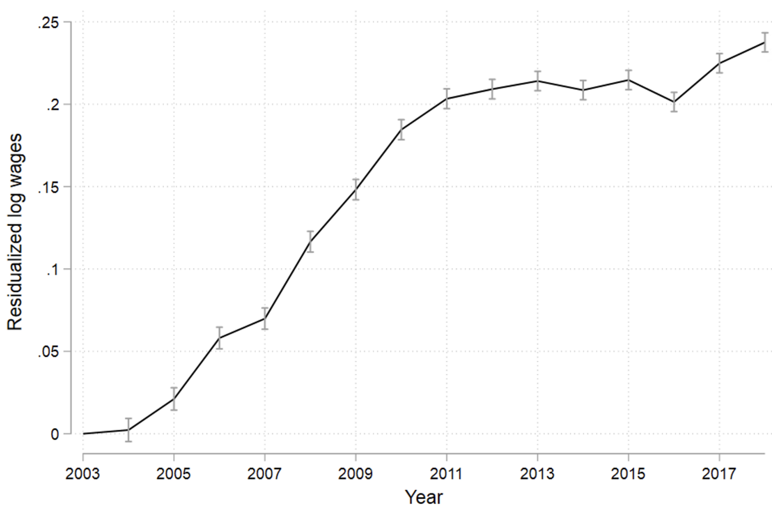
Source: World Bank

Figure C12: Estimated changes in composition and skill premium

(a) Δ $skill\ \widehat{composition}$



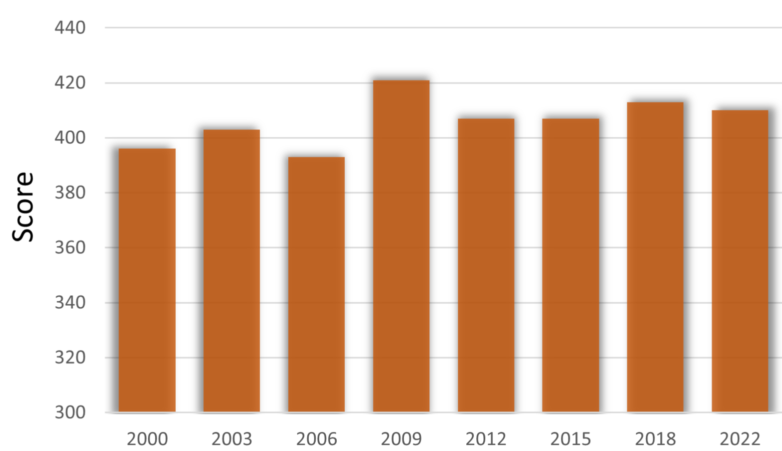
(b) Δ $skill\ \widehat{premium}$



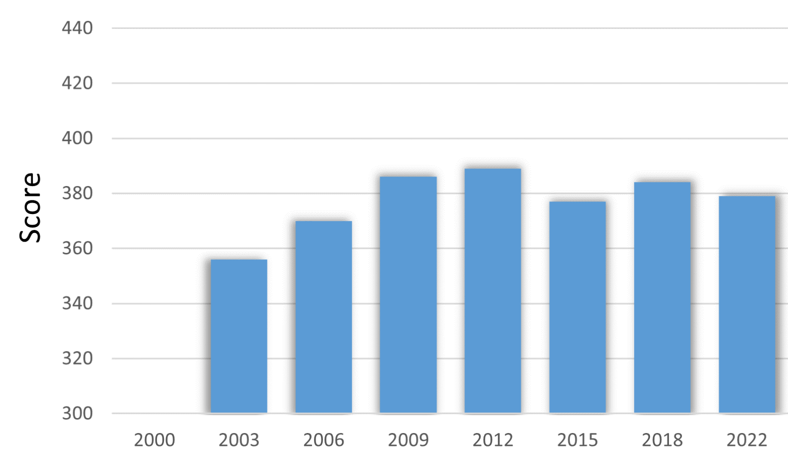
Note: The figure plots the efficiency adjustment, as described by Equation 14.

Figure C13: PISA Scores - Brazil

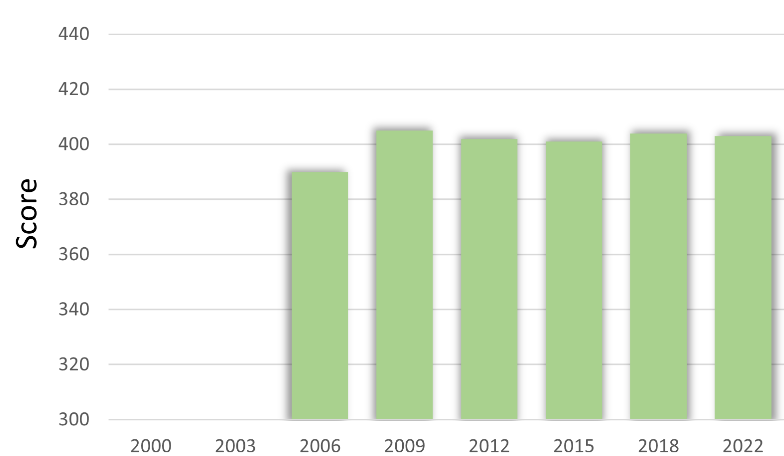
(a) Reading



(b) Math



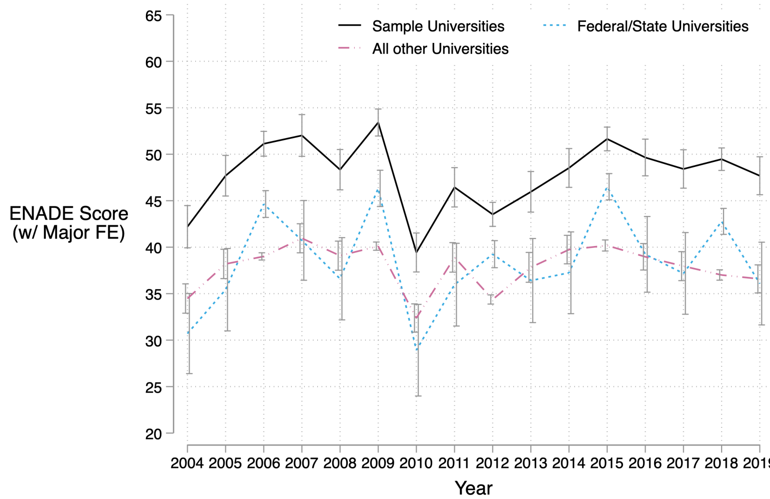
(c) Science



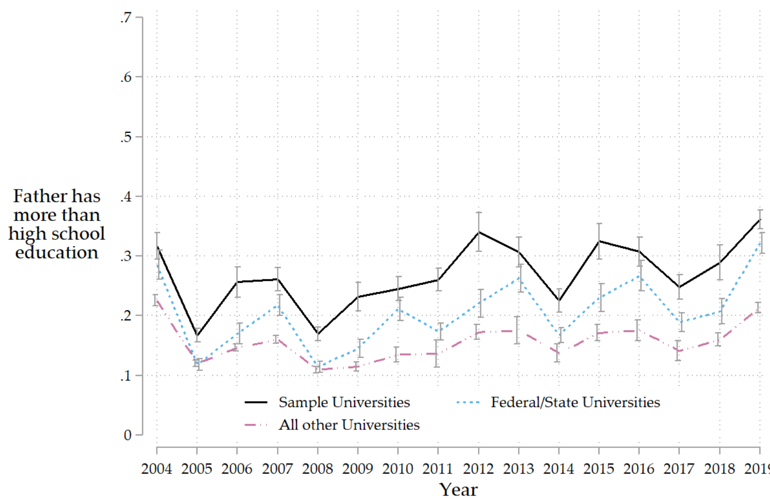
Note: The Programme for International Student Assessment (PISA) assesses the knowledge and skills of 15-year-old students in mathematics, reading and science. Source: OECD (2023)

Figure C14: ENADE scores and father's education

(a) ENADE Scores (with Major fixed effects)



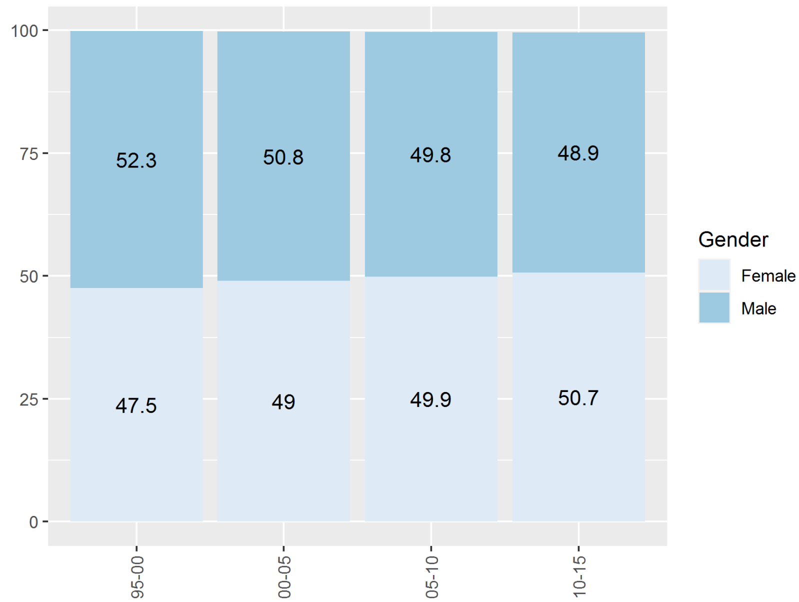
(b) Father Education: Share with more than High-School



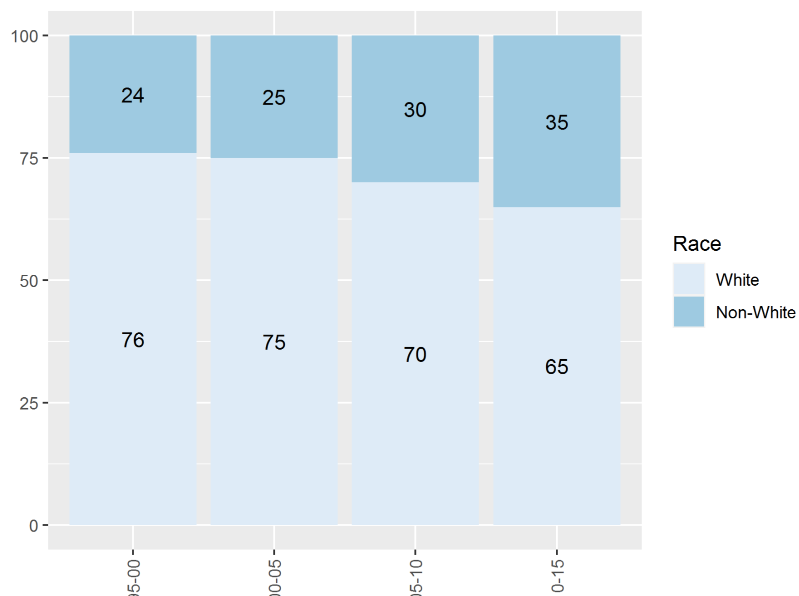
Note: ENADE (Exame Nacional de Desempenho dos Estudantes) is a standardized national assessment in Brazil that evaluates the academic performance and quality of undergraduate courses in higher education institutions. ENADE assesses students' knowledge and skills in specific majors or fields of study and is typically taken by students nearing the completion of their undergraduate degrees. Majors are selected for evaluation on a cyclical basis, usually every three years. Source: ENADE

Figure C15: Race and Gender composition by year of graduation for the *University Sample*

(a) Gender



(b) Race



Source: Sample Universities and RAIS