Fostering Development Through Higher Education: College Attainment, Firms and Economic Growth

Alvaro Cox*

University of Oslo

March 20, 2025

Click here for the latest version.

Abstract

This paper examines the role of higher education in economic development by exploiting a 1996 education reform in Brazil that lowered entry barriers for private colleges. The reform's impact on educational attainment varied across local labor markets. I estimate this heterogeneous effect through an IV strategy exploiting the idea that private colleges targeted locations under-served before the reform. Rising college attainment causes income growth, a decline in agricultural employment, and an increase in the prevalence and employment share of large firms. To quantify the macroeconomic effects of rising educational attainment, I propose a multi-region structural model featuring an educational choice and a sectoral specialization of educated workers in non-agricultural activities. A key implication of the theory is that increasing college attainment facilitates firm expansion and improves productivity in the non-agricultural sector. The model successfully replicates the IV estimates from the empirical analysis and shows that lower college costs account for 22% of Brazil's GDP growth and 46% of the decline in agricultural employment between 2000 and 2010. The gains from higher education expansion were largest in regions more exposed to skill-biased technologies.

*alvaro.coxlescano@yale.edu. I'm extremely grateful to Michael Peters, Fabrizio Zilibotti, and Mark Rosenzweig for their constructive comments and suggestions. I thank Costas Arkolakis, Orazio Attanasio, Francesco Beraldi, Paula Bustos, Lorenzo Caliendo, Kevin Donovan, José-Antonio Espín-Sánchez, Mayara Felix, Joel Flynn, Karl Harmenberg, Tim Hersey, Jonas Hjort, John Eric Humphries, Ilse Lindenlaub, Costas Meghir, Giuseppe Moscarini, Mushfiq Mobarak, Rohini Pande, Tommaso Porzio, Nick Ryan, Todd Schoellman, Matthew Schwartzman, Petr Sedláček, Rodrigo Soares, Kjetil Storesletten, Lindsey Uniat, John Van Reenen and seminar participants at the CESifo/ifo Junior Workshop on the Economics of Education 2024, EEA-ESEM Congress 2024, Lacea Lames 2024, Nordic Conference in Development Economics 2024, Nordic Summer Symposium in Macroeconomics 2024, RIDGE 2023 and 2024, FGV-EPGE, INSPER, PUC-Rio, Tilburg University, Universidad Carlos III de Madrid, University of New South Wales, University of Oslo and Yale University for their comments. Access to administrative data was granted by NEREUS (Regional and Urban Economics Lab from University of Sao Paulo) and SEDAP/INEP.

1 Introduction

Structural transformation—the reallocation of labor from agriculture and self-employment into wage employment in modern firms—has been a defining feature of successful development. Yet, in many countries, this transition remains elusive.

A longstanding hypothesis in economics posits that expanding access to education can spur this transformation by fostering human capital accumulation, which in turn drives economic development. For instance, Lucas Jr (1988) and Galor and Moav (2004) emphasize the role of expanding access to education as an engine of income growth, while Caselli and Coleman II (2001) and Porzio, Rossi, and Santangelo (2022) highlight its role in structural transformation. Yet, this view is not uncontroversial. Bils and Klenow (2000), for instance, argue that education is largely a byproduct of technological progress and rising incomes rather than an independent driver of development.[1] This debate has motivated a research agenda exploiting within-country variation to estimate the causal effects of education on development. Most studies focus on pre-college education and consistently find positive effects on income growth. While this research has advanced our understanding of education's role, it has not fully resolved the debate for two reasons. First, regional comparisons cannot capture the aggregate impact of rising educational attainment on economic growth when general equilibrium forces are at play. Second, by concentrating on primary and secondary schooling, this literature provides an incomplete picture of education's broader role, leaving the potential benefits of expanding college education—where cross-country disparities remain largest—largely unexplored.[2]

This paper quantifies the contribution of expanding college access to income growth and structural transformation. I approach this question empirically and theoretically, focusing on the case of Brazil. Brazil provides an ideal setting, as an education reform allows me to identify exogenous variation in the supply of college-educated workers. This variation enables me to estimate the effects of rising college attainment and examine the mechanisms through which it fosters economic development and structural transformation. Building on the empirical findings, I develop a structural model to quantify the aggregate effects of expanding college access, accounting for the potential general equilibrium forces triggered by this large-scale policy.

On the empirical front, this paper exploits a 1996 education reform that lowered entry barriers for private for-profit colleges. Prior to the reform, the Federal Council of Education (FCE) exercised control over the establishment of new private colleges, restricting capacity growth despite rising demand. Institutional constraints and opposition from incumbents

[1] A related view considers education a luxury good, with its consumption increasing as incomes rise. MacLeod and Urquiola (2019) provide evidence that households often select schools based on reputation and absolute achievement levels rather than the value added by the institution.

[2] For instance, Barro and Lee (2013) document substantial gaps in higher education attainment between the U.S. and Brazil. By 2010, the share of individuals with a college degree in the U.S. was nearly four times higher than in Brazil, while the difference in secondary education attainment was only 1.7 times larger.

created a mismatch between the potential demand for college education and the available supply. The reform deregulated entry for private institutions and introduced a more transparent accreditation and quality assessment process. Guided by this institutional setting, I implement an instrumental variable approach that exploits variation in higher education capacity scarcity relative to potential future demand. Specifically, I use the pre-reform ratio of colleges to the number of 6-to-15-year-olds in 1991. By 2010, this cohort had entered the 25-to-34 age group and became the primary beneficiaries of the reform. Intuitively, once regulatory barriers were lifted, private colleges targeted underserved regions, lowering access costs and increasing college attainment more rapidly in these areas. The evidence supports this hypothesis: conditional on pre-reform development, previously underserved regions experienced significantly larger increases in higher education capacity and college attainment.

The empirical strategy estimates the impact of rising college attainment on local economic growth and structural transformation. Comparing regions with similar initial conditions and pre-trends but differential post-reform increases in college attainment due to the instrument, this paper documents the following findings. First, rising college attainment increased annual wage growth by 0.7 percentage points and accelerated the decline in the agricultural employment share by six percentage points over the decade. Most of the labor reallocation occurred toward the service sector. To put these magnitudes into perspective, between 2000 and 2010, the average annual income growth was 6.3%, and the agricultural employment share declined by an average of 11 percentage points across labor markets. Second, results indicate a significant increase in the employment share of white-collar occupations, particularly among younger cohorts who directly benefited from the expansion of higher education.

To understand the mechanisms behind these effects, this paper examines the role of firms using administrative firm-level data and documents four novel findings. Rising college attainment (i) increases the prevalence of large firms, (ii) raises the share of firms that employ skilled labor and have a large workforce, (iii) increases the growth rate of formal job creation, and (iv) increases the share of formal workers employed in large firms. Furthermore, employers report that college graduates are predominantly employed in managerial and professional roles—unlike workers without a college degree. These findings suggest that expanding access to higher education, by increasing the supply of skilled workers—particularly in managerial roles—facilitates firm expansion and stimulates job creation.

The exclusion restriction assumption requires that, conditional on pre-reform development, the instrument affects economic development only through its impact on the rise in college attainment. Three pieces of evidence support this assumption. First, a difference-in-differences specification estimates no effect of the instrument on economic outcomes before the reform. Second, annual firm-level data show that the observed economic effects emerge only after college graduates enter the labor market. Third, migration data reveal no correlation between instrumental variation in college attainment expansion and either college graduate migration or overall migration, addressing a key identification concern.

Although the empirical estimates are informative about the causal effect of rising col-

lege attainment on local growth, they do not capture the aggregate impact in the presence of general equilibrium effects and cross-regional spillovers. Therefore, I develop a structural model where college attainment, income, agricultural employment, and the firm size distribution are jointly determined. The model features a multi-region economy with intra-national trade, where locations differ in productivity fundamentals and access costs to education. The model has three key features. First, individuals differ in ability and decide whether to pursue a college education based on expected returns and costs, allowing both demand and supply forces to shape college attainment. Second, the non-agricultural sector consists of a collection of heterogeneous firms that use skilled workers to expand. To capture the differential effects of rising college attainment across firms of different sizes, I introduce a non-homothetic production function, which links the demand for skilled labor to firm size. One interpretation of this mechanism is that expanding firms face managerial bottlenecks and benefit disproportionately from hiring college workers.[3] Third, consistent with the structural transformation literature, the model adopts PIGL preferences to capture income effects that drive labor reallocation. In equilibrium, when college costs are high, skilled workers are scarce, firms remain small, and aggregate productivity is constrained. Lowering education costs expands the skilled workforce, reduces the skill premium, enables firms to scale, and enhances productivity, ultimately driving income growth and structural transformation.

The model's structural parameters and region-specific fundamentals are estimated using Brazilian microdata. These parameters include preferences, production technology, and the ability distribution. First, household expenditure data are used to estimate the income elasticity of demand. Second, production function parameters are inferred from moments in the data motivated by the theory. Third, administrative records on pre-college test scores, college enrollment, and labor market earnings inform the estimation of the ability distribution. Finally, regional technology and cost fundamentals are identified based on equilibrium conditions, firm-level data, and labor market outcomes.

A key aspect of the structural estimation is recovering the implicit cost of attending college. College education costs can be inferred from observed college attainment, conditional on the prevailing skill premium and the estimated ability distribution. Intuitively: holding skill prices and ability distribution constant, regions with higher college attainment exhibit lower implicit education costs. Moreover, I formally establish that college costs and skill-biased technology can be uniquely identified for each region. The model's quantitative performance is validated by replicating the empirical IV estimates as non-targeted moments, underscoring the complementarity between the empirical and structural approaches. The empirical analysis informs the model's structure and provides well-identified moments that serve as a benchmark for validation. Supported by the model's ability to replicate the IV estimates, one can proceed with quantifying the aggregate effects of expanding college access.

To quantify the macroeconomic benefits of expanding college access, I construct a coun-

[3]For instance, in small firms, the entrepreneur handles all administrative activities. As firms expand, time constraints and operational complexity make delegation to skilled workers essential for scaling up and improving productivity.

terfactual Brazil using the estimated model. Between 2000 and 2010, Brazil's GDP per capita grew by 29%, while the agricultural employment share declined by 6%. The counterfactual analysis shows that lower college costs increased annual GDP per capita growth by 0.96 percentage points. Compared to the 0.7 percentage points estimated in the empirical analysis, general equilibrium forces amplified the benefits of education expansion by 0.26 percentage points—accounting for roughly 25% of the total effect. Regarding structural transformation, reduced college costs explain 46% (2 percentage points) of the overall decline in agricultural employment share. This magnitude is consistent with the model's inference that college costs fell more sharply in poorer regions, where agricultural employment shares were initially higher, reinforcing the role of education expansion in driving labor reallocation.

The model's multi-region structure enables me to examine why some regions benefit more from expanding access to college. Economic gains were largest in regions with greater exposure to skill-complementary technologies (e.g., such as computers and the internet). Moreover, I document a strong interaction between education expansion and reductions in firms' fixed set-up costs, which likely reflect broader frictions-such as credit constraints and regulatory barriers-that limit firm growth. By combining empirical evidence with structural modeling, these findings demonstrate that lowering barriers to higher education plays a transformative role in economic development—not only by expanding human capital but also by reshaping firm structure and labor markets. These results contribute to ongoing policy discussions on higher education deregulation.[4]

Related Literature This paper contributes to three strands of literature. First, it relates to the macro-development literature on human capital and economic growth, inspired by the seminal work of Lucas Jr (1988).[5] Related studies emphasize human capital's role in structural transformation, linking education to the reallocation of labor across sectors.[6] However, the extent to which education is an engine of development remains debated. Bils and Klenow (2000), for instance, argue that education investment is largely a byproduct of technological progress rather than a driver of growth. This paper makes a contribution to this debate by quantifying the impact of expanding access to college education on income growth and structural transformation, for the case of Brazil. The paper achieves this by exploiting a higher education reform in Brazil that created within-country variation in college attainment, allowing me to estimate the causal effects of rising skilled labor supply on economic outcomes. Guided by these findings, this paper develops and validates a structural model to quantify the aggregate impact of college expansion, considering general equilibrium

[4]For instance, Greece faced widespread protests in early 2024 over a proposal to allow private universities. See the article titled "Students and police clash in Greece as debate rages over legalizing private universities" in The Associate Press.

[5]See Galor and Moav (2004), Aghion et al. (2009), Kaboski (2009), Jones (2014), Gennaioli et al. (2013), Manuelli and Seshadri (2014), Hendricks and Schoellman (2018), and Engbom, Malmberg, et al. (2024).

[6]See Caselli and Coleman II (2001), Buera and Kaboski (2012), Buera, Kaboski, et al. (2022) and Porzio, Rossi, and Santangelo (2022).

effects.

Second, this paper builds on empirical research that leverages within-country variation to estimate the causal effects of education on economic development. Duflo (2001), Hsiao (2022), and Akresh, Halim, and Kleemans (2023) examine a large-scale school construction program in Indonesia and find that it increased educational attainment and wages at productive ages. In India, Khanna (2023) shows that rising public education spending leads to higher incomes. However, evidence on higher education expansion remains scarce.[7] The closest related work is Hsiao (2022), which combines empirical and structural methods to quantify the benefits of an elementary school construction program. This paper, by contrast, focuses on expanding access to higher education—where cross-country gaps remain largest—and quantifies its role in the development process. Moreover, it emphasizes firms as a key mechanism through which rising college attainment drives economic transformation, an aspect that remains under-explored in the literature.

This paper also contributes to the literature on human capital and firms. Classic work by Chandler (1977) and Lucas Jr (1978) highlights the productivity gains from organizational structures that facilitate managerial delegation. A growing body of research documents the role of management in explaining firm productivity differences.[8] To the best of my knowledge, this paper is the first to provide causal evidence on how an exogenous expansion in college-educated labor affects firms, showing that rising college attainment shifts employment toward larger firms. Related work by Engbom, Malmberg, et al. (2024) develops a structural framework where skill scarcity constrains firm growth and productivity and use my empirical evidence to validate their model. Their analysis, which relies on cross-country data, complements my findings by illustrating the relevance of these mechanisms at the global scale. Together, these studies underscore the role of higher education in firm dynamics and economic development.

Road Map The paper is structured as follows. Section 2 details the institutional context of the education reform. Section 3 presents the empirical analysis. Section 4 introduces the model used to quantify the contribution of higher education to economic development. Section 5 outlines the estimation and validation of the model using empirical evidence. Section 6 quantifies the economic importance of facilitating access to college for development. Section 7 concludes.

[7] See Rocha et al. (2017), Valero and Van Reenen (2019), Nimier-David (2022), and Vu and Vu-Thanh (2022).

[8] See Bloom, Eifert, et al. (2013), Bloom, Lemos, et al. (2014), Bender et al. (2018), Akcigit, Alp, and Peters (2021), and Hjort, Malmberg, and Schoellman (2022).

2 Institutional Context

This section describes the institutional context around the 1996 education reform, which substantially expanded higher education capacity, enrollment, and graduation.

Before 1996, public and private not-for-profit colleges were the primary providers of higher education. Public colleges were prestigious, offering high-quality education free of charge, with a wide range of majors and advanced courses. These institutions operated autonomously, organizing and structuring their curricula. Private providers, on the other hand, were non-profit institutions catering to students who could not secure admission to public colleges, often focusing on low-cost programs like law and administration.[9] Established in 1961, the Federal Council of Education (FCE) was responsible for regulating the private segment of higher education.

The FCE authorized and supervised the establishment of new private colleges and courses across Brazil.[10] This regulatory role exposed the FCE to lobbying from two opposing groups. On one side, private institutions pushed for a more permissive approach to expansion. On the other, professional organizations and public colleges resisted, arguing that the FCE could not adequately supervise the quality of instruction nationwide.[11] These pressures influenced the evolution of higher education capacity. Between 1961 and 1975, the FCE adopted a lenient approach, increasing the number of colleges and courses. However, starting in 1975, resistance from professional organizations prompted stricter regulatory oversight, causing private college expansion to stagnate throughout the 1980s and early 1990s.[12]

Before the reform, prospective students faced two key constraints. First, access to colleges varied significantly across regions. The left panel of Figure 1 illustrates the geographic distribution of colleges in Brazil. In this context, students often had to travel long distances to take entrance exams and, if admitted, relocate to pursue their education.[13] Second, capacity constraints were binding. Using data from the 1995 Higher Education Census, one can calculate the ratio of new enrollments to available slots at each college. The right panel of Figure 1 shows the distribution of this ratio, which serves as a proxy for capacity utilization, revealing a system operating near full capacity.

By the mid-1990s, growing demand for higher education had outpaced the system's ca-

[9] While I use "college" as a general term, Brazilian higher education institutions are legally classified as universities, campuses, faculties, and others, which differ in scope and enrollment capacity. All these institutions grant college degrees.

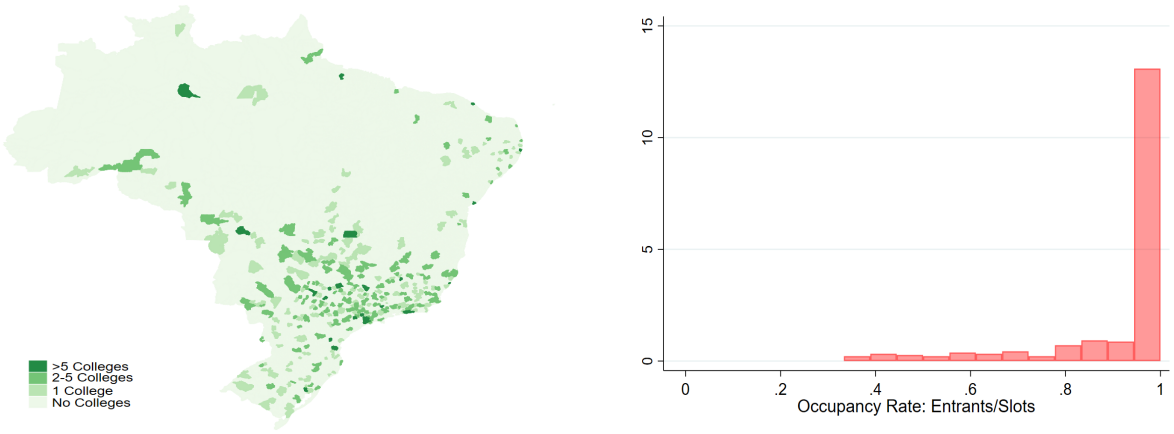
[10] This process involved approving programs, curricula, and enrollment capacity. See Sampaio (2000) for details on the higher education system and the role of the FCE before the 1996 reform.

[11] S. Schwartzman (1998) describes the political pressures the FCE faced in its regulatory role.

[12] During this period, some private institutions consolidated into universities to escape FCE control (S. Schwartzman 1998).

[13] Appendix Table A.1 reports statistics on the development levels of Commuting Zones (CZ) with and without colleges. While average incomes are similar, areas with colleges tend to have larger populations, higher educational attainment, and lower employment shares in agriculture. CZs, as defined in Dix-Carneiro and Kovak (2017), represent labor markets with stable geographic boundaries.

Figure 1: Higher Education System Before the Reform - 1995



Notes: The left panel shows the spatial distribution of Colleges across Brazil in 1995. The region's color depends on the number of colleges it hosts. There are four categories: 0 colleges (lightest green), one college, between two and five colleges, and more than five colleges (darkest green). Using the Census of Higher Education in 1995, occupancy rate is defined as the ratio between new enrollments and the number of slots offered. The right panel plots the distribution of the occupancy rate at the college level. The value of 1 means that the college was operating at full capacity.

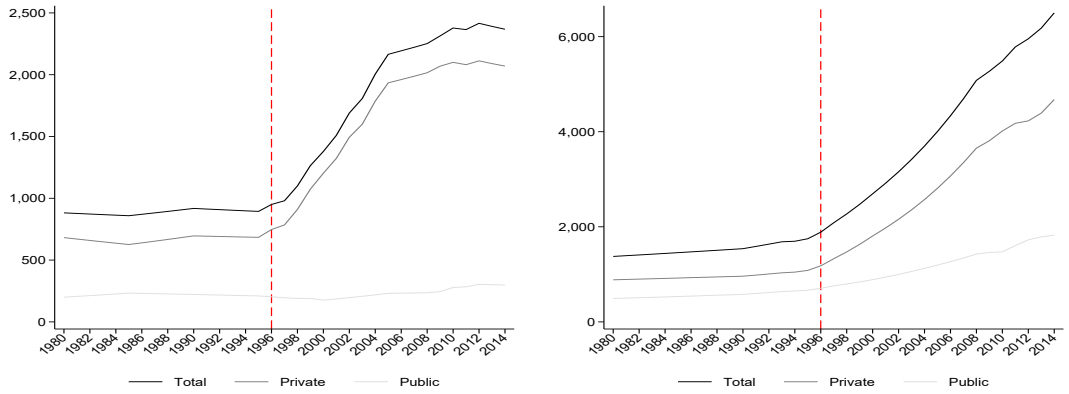
pacity, prompting the government to enact the 1996 Education Law to expand access.[14] The reform replaced the FCE with the National Institute for Educational Studies and Research ($INEP$) as the primary regulatory authority and introduced three key changes. First, private providers were allowed to operate as for-profit institutions. Second, the entry requirements for establishing new colleges were simplified. Third, $INEP$ implemented a more transparent system for quality assessment and accreditation, which is described later in this section. These legal changes attracted new investments and entrepreneurs, leading to a significant expansion in the number of colleges. Unlike in many countries, where public investment led higher education expansion, Brazil relied on market forces to expand capacity while maintaining regulatory oversight for quality assurance.[15]

Under the new framework, private colleges gained greater academic autonomy and the legal ability to operate for profit, leading to a significant expansion in the sector. As shown in Figure 2 panel (a), the number of colleges increased substantially after 1996. The rise in private providers explains the observed growth. Panel (b) of Figure 2 highlights a corresponding surge in higher education enrollment. To contextualize this growth, Figure 3 extends the analysis of Figure 1 to 2005, showing the new spatial distribution of colleges across Brazil. The expansion increased the availability of institutions and slots and reduced access costs for students who previously had to travel long distances.

[14]The Education Law of 1996, along with decrees 2.207 and 2.306 enacted in 1997, formed the legal framework for privatizing the higher education sector. These reforms are collectively referred to as the education reform.

[15]For a detailed account of the institutional details and expansion of the Brazilian Higher Education sector after the reform, some references are Neves (2015), Pinto (2004), Magalhães Castro (2015) and S. Schwartzman (2015).

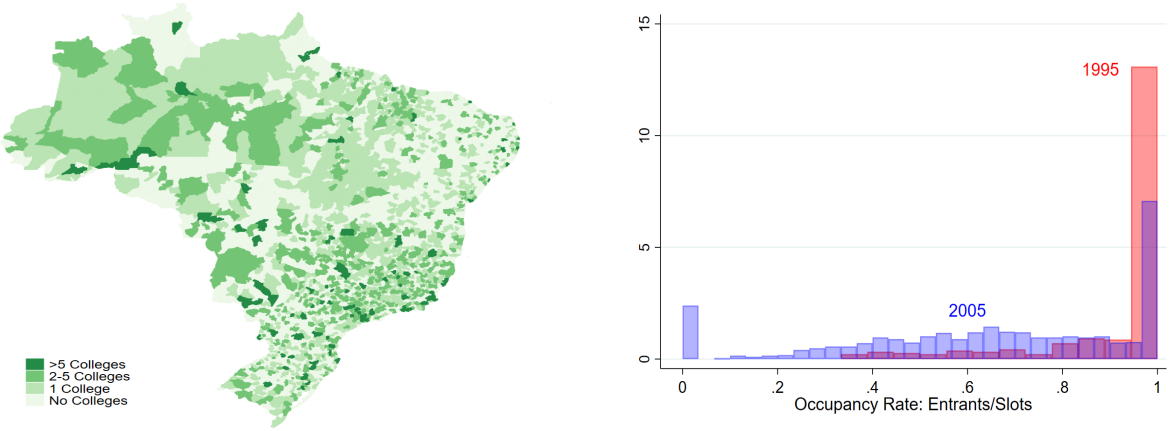
Figure 2: Higher Education Reform



(a) Number of Colleges (b) Enrollment (1000s)

Notes: Panel (a) plots the time series of the total number of colleges and the number of colleges by type of institution, public or private. Panel (b) reproduces the same information using college enrollment measured in thousands. In both panels, the red line highlights 1996, when the education law was enacted.

Figure 3: Higher Education System Before the Reform - 2005

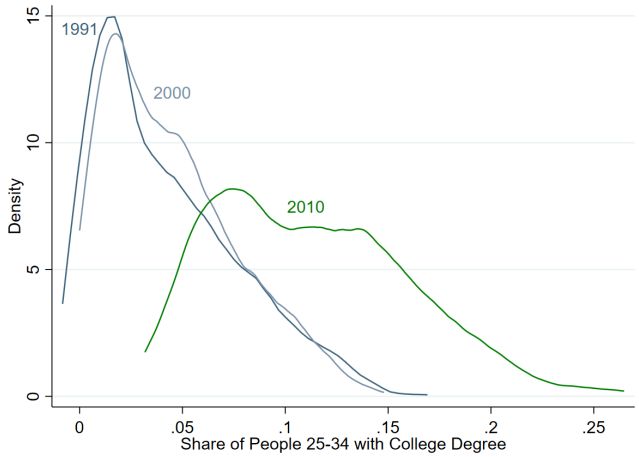


Notes: The left panel illustrates the spatial distribution of colleges across Brazil in 2005, with regions color-coded based on the number of institutions they host. The four categories are: no colleges (lightest green), one college, two to five colleges, and more than five colleges (darkest green). The right panel depicts the distribution of college occupancy rates, defined as the ratio of entrants to available slots. The figure presents this distribution using data from the Census of Higher Education for 1995 and 2005, with the 1995 distribution in red and the 2005 distribution in blue. An occupancy rate of 1 indicates that the college was operating at full capacity.

The increase in enrollment following the reform led to a corresponding rise in college graduates. To quantify this, I calculate the share of individuals aged 25 to 34 who completed higher education in each CZ. Figure 4 presents the distribution of this share across three census years. In 1991, five years before the reform, the median CZ had 3.3% of its youth population with a college degree. By 2000, three years after the reform, this figure remained largely unchanged, as most students had not yet graduated. However, by 2010, the distribution shifted significantly to the right, reflecting a substantial increase in college grad-

uates. The median share of youth with a college degree rose to 10.5%, though the expansion was heterogeneous across CZs.

Figure 4: Expansion in College Attainment



Notes: The figure shows the distribution across CZs (labor markets) of the share of people in the 25 to 34 age group with a college degree. The figure shows the distribution of this share for 1991, 2000, and 2010. Appendix A.2 depicts the rise in college attainment at the municipality level.

Three additional aspects of the reform merit discussion. First, the rapid growth of private colleges raised concerns about educational quality. In response, the Ministry of Education implemented an accreditation system requiring colleges to undergo evaluation every five years. A key component was the introduction of a compulsory national standardized exit exam, ENADE, administered to graduating cohorts to assess learning outcomes. Exam results were publicly disclosed and became central to the accreditation process.[16] Appendix A.3 provides suggestive evidence on how instructional quality evolved post-reform. Using administrative data, I estimate a value-added measure at the institution-program level, indicating that for-profit colleges established after the reform delivered quality comparable to the non-profit institutions that were prominent before the reform.

Second, Appendix A.3 presents descriptive statistics on college majors and the job profiles of new graduates. The data show no significant differences in major selection or occupational outcomes between graduates in 2010, who enrolled post-reform, and those in 2000, who enrolled before the reform. Most college graduates studied careers in business administration, law, and other professional services. They are employed in white-collar occupations in the service sector. Most of them are formally employed. In other words, the reform increased the availability of skilled workers but did not change the choice of majors and job profiles. The increase in college enrollment, however, raises the possibility that the ability of the marginal student declined. The appendix also examines the potential implications of changes in the ability composition of college graduates for the analysis.

[16]For details on the Provao (later ENADE), see S. Schwartzman (2010).

Third, administrative data linking students to their pre-enrollment home addresses enables the analysis of migration patterns related to college enrollment. The data show that around 95% of private-sector students enrolled in colleges within an 100km radius of their residence, while 90% of public-sector students enrolled within a 220km radius. These patterns suggest that the rise in college attainment was predominantly driven by local enrollment rather than migration. This finding is consistent with Figure 3, which shows that new colleges entered previously underserved markets rather than clustering in existing educational hubs.

3 Empirical Analysis

3.1 Data

This paper utilizes three administrative and public datasets to assess the impact of an increased supply of skilled workers, defined here as college-educated workers.

First, this paper uses Brazil's decennial censuses from 1980, 1991, 2000, and 2010. These censuses allow me to calculate the share of individuals aged 25 to 34 with a college degree in each year and Commuting Zone (CZ). To assess the implications of the reform for structural transformation, employment shares by sector and occupation are computed for both the younger cohort and the broader local economy. Income growth is measured using average earnings, and changes in the college skill premium are analyzed by estimating the income gains associated with holding a college degree for the three most recent census years. The methodology for estimating the college premium is detailed in Appendix B.3.

Second, I incorporate data from INEP's Census of Higher Education, to quantify the expansion in higher education capacity following the 1996 Education Law. This dataset includes the location of college campuses, the programs offered, and the number of available slots per program. It allows me to measure the stock of colleges in each CZ in 1995, one year before the reform, and to track capacity expansion from 1995 to 2005. Descriptive statistics for variables derived from the Higher Education and Population Censuses are presented in Appendix B.2.

Third, the analysis utilizes the Relação Anual de Informações Sociais (RAIS), which provides annual administrative data covering the universe of formal firms and employees in Brazil from 1985 to 2018. This dataset includes information on firm location, two-digit sector classification, hours worked, earnings, education level, and occupations for each employee as reported by employers. RAIS enables me to measure the firm size distribution in each local labor market and analyze the occupations of college graduates within firms. This allows for tracking how an increase in skilled labor supply influences firms' size distribution over time.

3.2 Empirical Strategy

This section outlines the empirical strategy to assess the consequences of rising college attainment on economic development. The rise in college attainment is measured as the change in the share of individuals aged 25 to 34 with a college degree between 2000 and 2010. Since these individuals enrolled at least five years earlier, the reform enacted in 1996 would have influenced college attainment by 2010, when its effects became fully observable.

$$\Delta Y_r = \delta_0 + \beta \Delta h_r + X_r \gamma + \theta_r^s + \epsilon_r \tag{1}$$

Equation (1) presents the baseline specification. The dependent variable ΔY_r represents various measures of local economic development, such as income growth, the change in employment shares in white-collar occupations or high-skill services, the growth in the number of firms, and the increase of the employment share in large firms. The main explanatory variable, Δh_r, denotes the change in the share of people aged 25 to 34 with a college degree between 2000 and 2010. The coefficient β captures the effect of increasing college attainment on economic development outcomes. All specifications include state fixed effects, θ_r^s, to account for state-specific trends between 2000 and 2010.[17] Control variables X_r are included to measure initial levels of development in 1991 and development trends between 1991 and 2000. Finally, ϵ_r represents unobserved factors affecting the development of a CZ.

Estimating β using Ordinary Least Square would yield the causal effect of a change in college attainment (Δh) on the outcome variable if Δh were uncorrelated with the error term. However, there are reasons why this condition could fail in this case. For instance, human capital investments may respond to economic development or expectations of future growth, creating a reverse causation issue. Individuals might have the incentive to attend college in regions where future job opportunities for college graduates are plentiful.[18] Additionally, suppose a college degree is a luxury good. In that case, any shocks that increase earnings in the local economy will also increase demand for college, inducing a spurious correlation between education and development.[19]

The following procedure addresses identification concerns. First, I include controls (X_r) that capture pre-reform levels of human capital and development, as well as development pretrends. These controls serve as proxies for a region's growth prospects that might influence future demand for college education. Summary statistics for these controls are provided in Table B.1. Second, and most importantly, I implement an instrumental variable approach

[17] There are 26 States plus one Federal District in Brazil. Each State has, on average, 18 CZ. There are 486 CZs in Brazil, with an average population of 302 (393) thousand in 1991 (2010). Each CZ has, on average, 11 municipalities.

[18] While skill-biased technical change is often considered a key driver of growth that could generate this reverse causation issue, technological progress does not necessarily have to be skill-biased to increase demand for skilled workers. For instance, Buera, Kaboski, et al. (2022) show that exogenous neutral technological change can raise the demand for skilled labor when economic development shifts the composition of value added toward more skill-intensive sectors.

[19] See MacLeod and Urquiola (2019).

to address bias from unobserved regional growth determinants not captured by X_r but relevant to the decision to attend college. The strategy builds on pre-reform higher education scarcity to predict the entry of new colleges and the subsequent rise in college attainment. Specifically, I measure college availability before the reform relative to the population aged 6 to 15 in 1991, who constituted the 25–34 age group by 2010. The underlying hypothesis is that, following deregulation, private colleges expanded primarily in areas where colleges were relatively *scarce* compared to potential future demand, leading to a faster increase in college attainment. Instrument Z_r^1 is defined as the ratio of pre-reform higher education capacity ($HEC_{r,1995}$) to the population aged 6 to 15 in 1991.

$$Z_r^1 = \frac{HEC_{r,1995}}{\#\text{6-15yo youth}_{r,1991}}$$

Here, $HEC_{r,1995} = \sum_{m \in \Omega_r} \omega_m Colleges_{m,1995}^{100km}$. Where Ω_r represents the set of municipalities in r, ω_m is the 1991 population weight of municipality m, and $Colleges_{m,1995}^{100km}$ is the number of colleges available within $100km$ of municipality m. The $100km$ threshold is motivated by evidence that 95% of private college students enroll within this distance from their residence (see Section 2). HEC can be measured using the number of colleges, programs, or available slots in a Commuting Zone (CZ). Appendix B.4.1 shows that all alternative measures yield similar results. The baseline results use the measure with the highest $F-$statistic.

Some CZs had no colleges in 1995, as depicted in Figure 1. Figure B.1 displays the density of Z_r^1 for the subsample of CZs prioritized by the FCE to have a college compared to those CZs that were not. Among non-prioritized CZs, Z_r^1 shows low variation, with most values concentrated around zero, which weakens the instrument's predictive power on college attainment.[20] To address this, I measure higher education scarcity for this subsample by exploiting variation in potential future demand for college. Conditional on controls, CZs with similar population size and development patterns but a higher share of youth under 15 in 1991 were more likely to attract colleges post-reform, as a larger pool of potential enrollees increased incentives for expanding capacity, leading to a faster rise in college attainment. However, a larger share of individuals under 15 in 1991 (aged 25 to 34 in 2010) also implies a smaller pool of older individuals (aged 35 to 50 in 2010), who typically possess more labor market experience and higher earnings, potentially violating the exclusion restriction. If labor market experience and higher incomes at older ages contribute to economic development, the estimated effect of rising college attainment based on this variation would likely represent a lower bound of its true impact.

$$Z_r^2 = \text{Youth Pop.Share}_{r,1991}$$

The identification strategy is motivated by the institutional context preceding the reform. Before 1996, college planning was centralized by the FCE, which faced challenges in

[20] The instrument has a F-statistic > 10, but fails Montiel Olea and Pflueger (2013) weak instrument test.

forecasting regional demand for higher education. This misalignment between supply and future demand forms the core of my identification strategy.[21] The first stage specification takes the form:[22]

$$\Delta h_r = \delta_0 + \alpha_1 Z_r^1 + \alpha_2 Z_r^2 + X_r\gamma + \theta_r^s + \epsilon_r \tag{2}$$

Appendix B.2 provides the full list of controls, X_r. It includes a dummy variable for whether the CZ was prioritized to host college institutions. Thus, conditional on controls, Z_r^1 creates variation in the rise in college attainment for CZs that were prioritized to host colleges, while Z_r^2 creates variation for CZs that were not prioritized. The identification assumption for this strategy is that, conditional on controls, the pre-reform relative scarcity of higher education capacity is unrelated to economic development prospects, except through the increase in college attainment following the reform. The main results, estimated using Z_r^1 and Z_r^2 as instruments, are presented in Section 3.3. For robustness, Appendix B.5 replicates results using only Z_r^1 as the instrument. The validity of the identification assumption is discussed in detail, with supporting evidence, at the end of Section 3.3.

3.3 College Expansion and Economic Development: Results

This section introduces the main results of the empirical analysis. First, the results of the first stage of the IV regression are discussed. Next, I move to the second stage and estimate the effects of the expansion in college attainment on local development. The analysis focuses on income growth, structural transformation, and firms' outcomes. The section concludes by (i) providing evidence supporting the exclusion restriction assumption and (ii) discussing the implications of migration after graduation.

3.3.1 First Stage Analysis: Implications for College Attainment

The instrumental variable strategy exploits pre-reform scarcity of higher education capacity (HEC). When the reform eased entry restrictions and provided legal permission to seek profits, private colleges had incentives to establish in CZs where centralized planning had constrained higher education capacity relative to potential demand. This entry reduced access costs in underserved areas, leading to a faster increase in college attainment. I use an instrumental variable approach rather than directly relying on observed college entry, as entry decisions may be correlated with regional economic prospects, potentially inducing bias.

[21] Professional organizations and public colleges highlight FCE's limited capacity to oversee the higher education sector as a justification for curbing the accreditation of new colleges during the pre-reform period.

[22] An alternative approach would directly use observed entry of new colleges to generate variation in college attainment. However, this strategy raises endogeneity concerns, as private colleges may have selectively entered labor markets with stronger growth prospects, potentially biasing the estimates.

To test this hypothesis, one can estimate specification (2), where Δh_r measures the increase in college attainment among the youth between 2000 and 2010 or the growth in higher education capacity between 1995 and 2005. The analysis focuses on higher education capacity built before 2006, as later expansions would not affect college attainment by 2010. The instruments (Z_r^1, Z_r^2) represent the pre-reform college scarcity, while θ_r^s are state fixed-effects. Control variables X_r capture initial human capital and development levels in 1991 and pre-reform trends from 1991 to 2000. The design compares CZs similar in controls before the reform but differed in the relative scarcity of college capacity because of $FCE's$ allocation of colleges.

Table 1: First Stage Analysis

Dep.Variable	ΔHE Capacity		Δh, 2000-2010	
Specification	(1)	(2)	(3)	(4)
Z_r^1	-230.85***	-171.64***	-3.30**	-4.10**
	(40.08)	(50.46)	(1.06)	(1.57)
Z_r^2	1.08***	1.53	-0.04***	0.14**
	(0.14)	(1.30)	(0.00)	(0.05)
R^2	0.16	0.47	0.15	0.63
F-test	41.96	7.46	41.42	12.77
Obs.	486	486	486	486
Controls	No	Yes	No	Yes

Notes: Columns (1) and (2) measure the effect of the instrument on the increase in higher education capacity measured as the entry of new higher education programs between 1995 and 2005. Columns (3) and (4) measured the effect of the increase in college attainment of the 25 to 34 age group between 2000 and 2010. Columns (1) and (3) estimate specification (2) without controls. Columns (2) and (4) estimate specification (2) with controls. The complete list of controls is in Table B.1 in Appendix B.2.

The hypothesis states that lower (higher) values of Z_r^1 (Z_r^2) induced entry of new higher education programs, which would increase college attainment faster. Formally, I test whether $\hat{\alpha}_1 < 0, \hat{\alpha}_2 > 0$. Table 1 presents the OLS estimates of these coefficients. Columns 1 and 2 assess the effect on the growth in higher education capacity. Columns 3 and 4 introduce the results of the increase in college attainment. The analysis focuses on Columns 2 and 4, which include controls. The patterns in the data are consistent with the hypothesis described above. CZs where colleges were scarce relative to their potential demand saw more new higher education programs opened after the reform, and consequently, college attainment increased faster ($\hat{\alpha}_1 < 0$). At the same time, those places with no colleges pre-reform and with a relatively high share of youth experienced a not precisely estimated increase in higher education capacity and faster increase in college attainment ($\hat{\alpha}_2 > 0$).[23] The F-

[23]In Column 3, the effect of Z_r^2 on college attainment is negative in the specification with no controls. Regions with high youth population share (presumably because of higher fertility rates) might also be poorer, where parents might have fewer resources to invest in each kid's education, reducing prospects for a future

statistic exceeds 12, raising potential concerns about weak instruments. To address this, the second-stage results will report the efficient F-statistic from the robust weak instrument test developed by Montiel Olea and Pflueger (2013), along with a test for overidentifying restrictions. Appendix B.4.1 presents additional specifications using different measures of HEC_r, finding results consistent with Table 1. Figure B.2 in appendix B.4.2 depicts the variation in the rise in college attainment explained by the instruments.

3.3.2 Second Stage Analysis: Implications for Economic Development

This section examines the consequences of the rise in college attainment on different local development measures by estimating specification 1. First, Table 2 reports the estimated effects of rising college attainment on income growth and the college premium.[24] Second, Table 3 explores the implications for the allocation of labor across occupations and sectors. Expanding college education may generate benefits beyond the young cohorts directly affected by the reform, potentially creating spillovers for the local economy. To consider the effects on the young cohort and the potential local spillovers, I run the regressions separately for the 25-34 age group and for all workers.[25]

Table 2: Expansion in College Attainment and Income Growth

Dep.Variable	Income growth						Δ College premium		
Cohort	All cohorts			Young Cohort			Young Cohort		
Specification	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\Delta h_{(25-34)}$	-0.03	1.43***	1.64*	0.34	1.59***	2.40**	1.11	-3.42*	-13.37*
	(0.32)	(0.19)	(0.73)	(0.30)	(0.23)	(0.90)	(1.41)	(1.41)	(6.48)
R^2	0.62	0.88	0.88	0.57	0.82	0.81	0.27	0.45	0.40
Eff. F-test			11.11			11.11			11.11
CV 5%(10%)			8.84(6.33)			8.91(6.37)			10.09(7.05)
Overid. Stat(p-val)			3.85(0.05)			1.69(0.19)			1.18(0.28)
Obs.	486	486	486	486	486	486	486	486	486
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: I estimate specification (1) to assess the implications of rising college attainment on average income growth for all cohorts and the young cohort. The last three columns assess the implications on the returns to college. The unit of observation is the commuting zone, CZ. I consider three specifications for each sample. All specifications include state dummies. Columns (1) do not include controls. Columns (2) and (3) include controls aimed to capture pre-reform development patterns. Columns (3) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013), the TSLS critical values at the 5% and 10% levels, and statistics from the over-identification test.

college education. The effect aligns with the hypothesis when comparing locations with similar development patterns.

[24] Using cross-section variation across individuals within each CZ, one can obtain estimates of the income gains from having a college degree (see Appendix B.3). These estimates reflect both market forces and changes in the unobserved ability composition of college graduates.

[25] A similar approach is followed by Porzio, Rossi, and Santangelo (2022), who assess the contribution of human capital to changes in the agricultural employment share following a cross-cohort analysis.

Table 2 presents the effects of rising college attainment on income growth and the college premium. Columns (1) present the OLS results with no controls, Columns (2) introduce controls, and Columns (3) report the TSLS IV results. The IV results include the efficient F-statistic (Eff. F-test) from Montiel Olea and Pflueger (2013), which provides a robust test for weak instruments under general error structures, along with the reported critical values at 5% and 10% (CV 5%(10%)).[26] Over-identification statistics are also reported.[27] In Column (1), there is no significant association between rising college attainment and income growth or the college premium. In Column (2), once controls are included, a positive association emerges between rising college attainment and income growth and a significant reduction in the college premium. These results suggest that the effect of college attainment is heterogeneous and depends on the pre-reform *observable* development patterns. Even though the inclusion of these controls aims to capture important factors influencing the growth prospect of a region, the presence of *unobservable* determinants of growth would prevent us from interpreting the OLS coefficients as causal. To address this concern, Column (3) reports the IV estimates. The results indicate that college attainment leads to income growth in the local economy, with an effect similar in magnitude to the OLS estimates. For the young cohort—who are directly affected by the reform—the IV estimates show faster income growth and a stronger reduction in the college premium. This stronger decline in the college premium is reassuring, as it suggests that the IV strategy accounts for demand-side factors that increase the need for college-educated workers.

Table 3 introduces the effect of rising college attainment on the allocation of labor across occupations and sectors for all cohorts (Panel A) and for the young cohort in the 25 to 34 age group (Panel B). The analysis focuses on one occupation, white-collar, and three sectors: high-skill services, low-skill services, and agriculture. The full matrix of transitions across occupations and sectors is available in Table B.4 in Appendix B.5.[28] For each category, I estimate three specifications: Column (1) reports OLS estimates with no controls, Column (2) shows OLS estimates with controls, and Column (3) presents TSLS IV estimates. The analysis focuses on Columns (2) and (3). In Column (2), higher college attainment is associated with a reallocation toward white-collar occupations and both high- and low-skill services, with a stronger shift toward white-collar and high-skill services among younger workers, as well as a reallocation away from the agriculture sector. In Column (3), IV coefficients indicate that rising college attainment leads to reallocation toward white-collar occupations, high- and low-skill services, and away from agriculture across all workers. Reassuringly, the young cohort experiences a stronger reallocation toward white-collar occupations and high-

[26]The reported critical values correspond to the thresholds beyond which we reject the null hypothesis that the estimator's approximate asymptotic bias exceeds 5% or 10% of a worst-case benchmark.

[27]A p-value higher than 0.05 suggests that the instruments are likely valid.

[28]Occupations are categorized as follows: white-collar (e.g., professionals, technicians, and clerical workers), blue-collar (e.g., painters, mechanics, and artisans), service-related (e.g., merchants, cooks, and tourist guides), and elementary (e.g., agricultural laborers, guards, and watchmen). Sectors are grouped into agriculture, manufacturing (including mining), high-skill services (e.g., business, logistics, health, and education), and low-skill services (e.g., wholesale and retail trade, hospitality, food service, and personal services).

Table 3: Expansion in College Attainment and Structural Transformation

Dep.Variable	Change in employment share in											
Category	Occupation			Sector								
	White-Collar			High-Skill Services			Low-Skill Services			Agriculture		
Specification	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: All Cohorts												
$\Delta h_{(25-34)}$	0.21**	0.45***	0.68***	0.04	0.39***	0.64**	-0.05	0.34***	0.83**	0.21	-0.53***	-1.09**
	(0.06)	(0.05)	(0.20)	(0.08)	(0.06)	(0.20)	(0.11)	(0.09)	(0.28)	(0.16)	(0.10)	(0.36)
R^2	0.32	0.61	0.59	0.40	0.75	0.73	0.50	0.75	0.73	0.46	0.83	0.82
Eff. F-test			11.11			11.11			11.11			11.11
CV 5%(10%)			8.86(6.34)			9.23(6.56)			10.38(7.22)			11.05(7.61)
Overid. Stat(p-val)			0.11(0.74)			2.04(0.15)			6.02(0.01)			2.33(0.13)
Panel B: Young Cohort (25-34)												
$\Delta h_{(25-34)}$	0.64***	0.80***	1.50***	0.30***	0.60***	1.27***	-0.24	0.21	0.69	0.19	-0.55***	-1.36***
	(0.06)	(0.07)	(0.28)	(0.08)	(0.07)	(0.30)	(0.13)	(0.11)	(0.42)	(0.17)	(0.10)	(0.41)
R^2	0.36	0.52	0.38	0.34	0.61	0.52	0.47	0.73	0.71	0.40	0.81	0.78
Eff. F-test			11.11			11.11			11.11			11.11
CV 5%(10%)			9.30(6.60)			9.01(6.43)			10.72(7.42)			9.97(6.98)
Overid. Stat(p-val)			0.71(0.40)			1.31(0.25)			1.05(0.31)			0.29(0.59)
Obs.	486	486	486	486	486	486	486	486	486	486	486	486
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: I estimate specification (1) to assess the implications of rising college attainment on the process of structural transformation. Panel A reports results for all cohorts. Panel B focuses on the younger cohort. The analysis focuses on White-collar occupations and three economic sectors. The three economic sectors are High-Skill and Low-Skill Services and Agriculture sectors. The unit of observation is the commuting zone, CZ. I consider three specifications for each sample. All specifications include state dummies. Columns (1) do not include controls. Columns (2) and (3) include controls. Columns (3) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013), the TSLS critical values at the 5% and 10% levels, and statistics from the over-identification test.

skill services. Interestingly, the increase in college attainment among younger workers also induces a shift toward low-skill services in the local economy.

Table B.6 in the Appendix further examines labor market reallocation between self-employment, informal employment, and formal employment. The results indicate a substantial decline in self-employment. While workers transition into informal and formal employment, most of the reallocation occurs toward informal employment. Table B.3 and Table B.5 in the Appendix replicate these results using Z_r^1 as the sole instrument. The estimates obtained with Z_r^1 alone closely align with those presented in the main analysis, reinforcing the robustness of the findings. Finally, Appendix A.3 discusses how changes in instructional quality or the ability composition of new college graduates might influence these findings.

Nationwide, between 2000 and 2010, the average annual income growth rate across CZs was around 6.3%, and the college premium fell on average by 9.5% annually. The employment share of white-collar occupations grew by 4%, while high-skill and low-skill service employment share increased by 3% and 7%. In contrast, the agriculture sector's employment share decreased by 11%. To contextualize the results, the analysis focuses on the IV coefficient, which captures the impact of an exogenous expansion in the supply of skilled labor. During this period, college attainment among youth increased by around 6%. This rise in college attainment is associated with a 0.7 percentage point increase in income growth—equivalent to

about 11% of the average national income growth rate. Additionally, the increase in college attainment contributed to a 15% annual reduction in the college premium, a decline that outpaced the national average. This sharper decrease likely reflects the role of demand-side forces shaping the national trend, which are accounted for in the IV estimates. Regarding structural transformation, a 6% rise in college attainment led to a 4% increase in the white-collar employment share and a 3.8% (5%) increase in high-skill (low-skill) service employment shares. Finally, college attainment growth decreases the agricultural employment share by 6%— roughly half the pace of the national decline.

Results from Tables 2 and 3 reveal similar patterns in both OLS and IV estimates, with some differences. First, the IV estimates suggest faster income growth for younger cohorts and a sharper decline in the college premium. Second, these estimates indicate a more rapid increase in the employment shares of white-collar occupations and high-skill services, accompanied by a decline in the employment share of the agricultural sector. Third, the IV estimates suggest that rising college attainment also leads to a reallocation of employment shares toward low-skill services. These differences between OLS and IV estimates may stem from unobserved demand factors. Autor and Dorn (2013) discuss how skill-biased technical change (SBTC) can simultaneously increase demand for college graduates while displacing low-skill workers. If such forces are present in the OLS estimates, they could attenuate the effects on both income growth and the reallocation of labor across occupations and sectors. Specifically, in the OLS specification, rising college attainment may be associated with faster income growth for college-educated workers but slower growth for displaced low-skill workers, which could dampen the estimated effects on both average income growth and employment reallocation. Appendix B.7 provides suggestive evidence in support of this hypothesis. First, the estimated effect on income growth for "no-college" workers is lower than the average income growth, particularly in the OLS estimates. Second, the estimated effect on the difference in income growth between unskilled and skilled workers is smaller in magnitude under the IV specification compared to the OLS. Third, employment growth for workers without a college degree has a negative point estimate in the OLS specification but turns positive in the IV specification. While some coefficients are imprecisely estimated, these patterns suggest the presence of demand forces that differentially affect skilled and unskilled workers in the OLS specification.

Assuming the exclusion restriction holds, the IV specification accounts for demand forces that might be correlated with the rise in college attainment in the OLS specification. In this case, how does an expansion in the supply of college workers generate income growth? This paper proposes that increasing college attainment provides firms with skilled labor, enabling them to scale up and become more productive. While skilled workers may drive firm growth through various mechanisms, I focus on one to illustrate this relationship. Drawing on Chandler (1977), one interpretation suggests that college-educated workers provide managerial talent that facilitates delegation—particularly when limits on the span of control

constrain firm size and productivity.[29] A rise in college attainment increases the supply of skilled workers, reducing their relative cost and allowing more firms to hire them. As firms integrate these workers, they expand and become more productive. These productivity gains, in turn, translate into higher incomes for the local economy.[30]

The IV results indicate that increasing college attainment promotes income growth and structural transformation. My hypothesis suggests that college-educated workers enable firms to expand and increase productivity. The following section tests this hypothesis empirically.

3.3.3 Second Stage Analysis: Implications for Firms

The hypothesis posits that college-educated workers help firms expand and become more productive.[31] To fix ideas, this paper focuses on one possible mechanism through which college-educated workers enable firm expansion: by providing managerial skills. In a scenario where (i) entrepreneurs have limited time to oversee a firm's productive activities and (ii) managing these activities requires knowledge acquired in college, the availability of skilled workers fosters productivity growth. When college-educated workers are scarce and expensive to hire, firms remain small, and entrepreneurs handle all productive activities themselves. As college attainment rises, skilled workers become more abundant and affordable, allowing more firms to hire them, expand in size—creating additional job opportunities—and improve productivity. Following this hypothesis, both the prevalence of large firms and the employment share in large firms should increase.

To test this empirically, the analysis utilizes the matched employer-employee database, which provides annual data on firms spanning the pre- and post-reform periods. Using specification (1), I estimate the effects of rising college attainment on the growth in the total

[29] For the importance of managerial skills in delegation and firm expansion, see Bloom, Eifert, et al. (2013), Bloom, Brynjolfsson, et al. (2019), Akcigit, Alp, and Peters (2021), and Engbom, Malmberg, et al. (2024).

[30] Other mechanisms may also underlie this relationship. For example, Aghion et al. (2009) emphasizes the role of education in promoting technology adoption and innovation. Unlike the *SBTC* hypothesis, which views rising college attainment as a response to technological progress, their paper suggests that an increase in skilled workers actively drives technological advancement. Alternatively, as highlighted by Arrow (1973) and Spence (1973), a college degree can serve as a signal of worker ability when such information is otherwise unobservable to employers. If ability is difficult to observe and attending college is relatively easier for high-ability individuals, an expansion in college attainment could improve employers' ability to distinguish high-ability workers from low-ability ones. Specifically, if only higher-ability individuals find it worthwhile to obtain a college degree following the reform (i.e., a separating equilibrium emerges), then expanding college access would enhance firms' ability to identify and hire more capable workers, improving productivity and fostering growth. Identifying the precise mechanisms through which higher education fosters income growth is beyond the scope of this paper and remains an important avenue for future research.

[31] Data from enterprise surveys align with this hypothesis. The World Bank surveyed firms of different sizes in Brazil in 2008. The survey's objective was to elicit firms' perceptions of the main constraints they face to grow. Table B.9 in the Appendix shows the results of the Survey. The low education of the workforce ranks as the second biggest constraint to growth below taxes and is chosen by 75% of Firms in Brazil as one major constraint to growth.

number of firms, growth in small and large firms employing skilled workers, the change in the share of firms that are large and use skilled workers, formal employment growth, and the change in the share of formal employment in large firms.[32] To ensure consistency with the analysis in Section 3.3.2 and to leverage the annual data frequency, outcome variables measure changes between 1996 and 2010. Section 3.3.4 further exploits the annual frequency of the data.

Table 4 presents the results. Panel A reports the effects on the growth in the total number of firms, as well as the growth in the number of small and large firms employing skilled workers. Panel B shows the effect on the change in the share of firms using skilled labor, local formal employment growth, and the change on the employment share of large firms. Column (1) reports OLS results without controls, Column (2) includes controls, and Column (3) presents the TSLS IV estimates.

Table 4: Expansion in College Attainment and Firms

Panel A									
Dep.Variable	Growth # of firms			Growth # small firms w/ skill			Growth # large firms w/ skill		
Specification	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\Delta h_{(25-34)}$	-1.27*	0.34	1.84	1.45*	2.71**	-0.60	0.96	2.20***	7.71**
	(0.49)	(0.43)	(2.08)	(0.65)	(0.83)	(4.39)	(0.60)	(0.65)	(2.44)
R^2	0.47	0.73	0.72	0.37	0.46	0.44	0.41	0.57	0.48
Eff. F-test			10.70			10.69			10.70
CV 5%(10%)			9.29 (6.59)			9.02 (6.43)			8.36 (6.05)
Overid. Stat(p-val)			11.50 (0.00)			0.59 (0.44)			0.81 (0.37)
Panel B									
Dep.Variable	Δ Sh.Large firms w/ skill			Formal employment growth			ΔEmp. share in large firms		
$\Delta h_{(25-34)}$	0.49***	0.35***	0.83**	-0.44	2.01*	8.94*	0.15*	0.16**	0.90*
	(0.09)	(0.09)	(0.30)	(0.97)	(0.88)	(4.07)	(0.06)	(0.06)	(0.40)
R^2	0.48	0.55	0.52	0.45	0.66	0.61	0.30	0.41	0.26
Eff. F-test			10.70			10.70			10.70
CV 5%(10%)			8.36 (6.05)			8.99 (6.41)			8.51 (6.13)
Overid. Stat(p-val)			1.70 (0.19)			1.20 (0.27)			4.15 (0.04)
Obs.	486	486	486	486	486	486	486	486	486
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: I estimate specification (1) to assess the implications of college attainment for firms' outcomes. The unit of observation is the commuting zone, CZ. I consider three specifications for each sample. All specifications include state dummies. Columns (1) do not include controls. Columns (2) and (3) include controls aimed to capture pre-reform development patterns. Columns (3) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013) and the TSLS critical values at the 5% and 10% levels. I also include statistics from the over-identification test.

[32]For firms with establishments in multiple CZs, I count them as separate firms at the CZ level. Small firms have at most five workers, while large firms have six or more. A large firm hiring skilled workers is defined as one with at least six employees, including at least one college graduate. Firm growth is measured using the DHS growth rate from Davis, Haltiwanger, and Schuh (1998), which accommodates cases where a labor market initially has zero formal firms. For instance, in 1996, two CZs had no large firms employing skilled workers. Results remain consistent when using log differences to measure growth rates.

The analysis focuses on the IV results in Column (3) and highlights key differences compared to the OLS results in Column (2) below. Panel A shows that rising college attainment increases the growth in the total number of firms, but the effect is not significant. Notably, however, rising college attainment increases the prevalence of large firms hiring skilled workers rather than small firms. Panel B shows that rising college attainment significantly increases the share of firms that are large and employ college graduates, accelerates formal employment growth, and consequently raises the employment share of large firms. While Columns (2) and (3) shared qualitative patterns, some differences stand out. First, rising college attainment increases the prevalence of large firms hiring skilled workers more rapidly in the IV specification than in the OLS specification. The OLS estimates suggest a modest rise in the number of both small and large firms employing skilled workers.[33] Second, formal employment growth is faster in the IV specification, consistent with the stronger labor reallocation observed in Table 3. As large firms employing skilled workers become more prevalent and create jobs at an accelerated rate, results show a corresponding increase in the employment share of large firms. These results align with the hypothesis presented at the beginning of this section: rising college attainment supplies skills that enable firms to expand in size and generate more job opportunities within the local economy.[34] Appendix B.7.2 investigates whether the rise in the prevalence of large firms is driven by new entrants or the expansion of existing firms. The evidence indicates that firms already established by 2003—when the first cohorts of post-reform college graduates entered the labor market—account for most of the increase in the prevalence of large firms.

Nationwide, between 1996 and 2010, the average growth rate in the number of firms was 88%, with the share of firms that are large and employ skilled workers increasing by 3.3pp. During this period, the growth in the number of small and large firms employing skilled workers reached 126% and 112%, respectively. Formal employment grew by 98%, with no significant increase observed in the employment share of large firms. A 6% increase in college attainment over this period increased the prevalence of large firms hiring skilled workers by approximately 46%, increasing the share of firms that are large and employ skilled workers by 5pp, boosting formal employment growth by 54%, and increasing the employment share of large firms by 5pp. These results are significant relative to the national averages across CZs and highlight the role of human capital as a key driver of firm growth and formal employment.

This section concludes by emphasizing two key aspects of the role of skilled workers in

[33] Results are similar in magnitude if one measure growth rates using log differences.

[34] One concern is whether the observed firm-level effects stem directly from the establishment of new colleges rather than the increased supply of skilled workers. First, colleges themselves are large employers of college-educated workers and could mechanically influence firm dynamics. Second, they may generate local demand effects by increasing consumption of goods and services. The next section addresses these concerns by excluding colleges from the analysis and leveraging the annual frequency of the data to show that the estimated effects become significant and grow in magnitude only after 2003, when most new colleges had been established and graduates entered the labor market.

firms. First, my hypothesis suggests that skilled workers represent a distinct labor input in firms' production processes. The matched employer-employee data includes information on each worker's education level and occupation. Notably, the information on occupation is reported by the employer. Figure B.3 in the Appendix shows that young college-educated workers are primarily employed in managerial and professional roles, contrasting sharply with the roles of workers without college degrees. These findings provide evidence that skilled workers contribute unique and valuable skills to firms. Second, my hypothesis suggests that firms increasingly rely on skilled workers to support expansion. If this holds, firm size should be positively associated with skill intensity, measured as the share of college-educated workers in a firm's workforce. Consistent with this prediction, Figure B.4 documents a positive and monotonic relationship between skill intensity and firm size. Moreover, the concave nature of this relationship suggests that hiring skilled workers is particularly critical when firms grow from very small to medium size, reinforcing the role of skilled labor in enabling firm growth.

3.3.4 Difference in Difference Analysis

The empirical analysis exploits pre-reform relative scarcity of higher education capacity to generate exogenous variation in the rise of college attainment. Following the reform, for-profit colleges prioritized entry into areas where higher education was scarce relative to future demand, leading to a faster increase in college attainment. Evidence supporting this hypothesis is presented in Section 3.3.1. The identification assumption requires that the instruments affect economic development in 2010 solely through their impact on college attainment expansion.

A potential concern is that the instruments, which measure the relative scarcity of colleges, may be correlated with unobserved factors—distinct from college attainment expansion—that independently influence long-run economic development. If such unobserved factors promoted growth in regions initially underserved by colleges, the exclusion restriction would be violated. Since the empirical design relies on the interaction between the reform and pre-reform college scarcity, the instruments should have no effect on outcomes before the reform unless they capture unobserved drivers of long-run growth. To evaluate this possibility, I extend the analysis to the pre-reform period using census data from 1980, 1991, 2000, and 2010. Specifically, I test whether the instruments predict the outcomes analyzed in Section 3.3.2 in both the pre- and post-reform periods using the following specification:

$$\Delta y_{r,t} = \mathbf{Z_r}\beta_{2000} + 1'_{t=2010} \times \mathbf{Z_r}\beta_{2010} + X_{r,t-1}\Gamma + \theta_r^s + \theta_t + u_{rt} \tag{3}$$

Where θ_r^s are state-fixed effects, θ_t are decade fixed effects, and $X_{r,t-1}$ include controls that account for pre-period development patterns. $\mathbf{Z_r}$ is the vector of instruments and $1_{t=2010}$ is a vector of dummy variables that equal one for the 2000-2010 decade. Thus, β_{2000} measures the effect of the instruments on outcomes before the reform, while β_{2010} captures

their reform-driven impact on local development. Table 5 presents the results, focusing on the outcomes analyzed in Section 3.3.2. Column (1) reports the estimates for income growth, while Columns (2)-(5) examines the process of structural transformation. All specifications include controls.[35]

Table 5: Diff-in-Diff Analysis

Dep. Variable	Income	ΔEmployment share in			
	growth	White-Collar Occ.	High-Skill Serv.	Low-Skill Serv.	Agriculture
Specification	(1)	(2)	(3)	(4)	(5)
Z_r^1	7.04	0.15	0.82	12.49***	-8.72**
	(8.27)	(1.49)	(1.24)	(2.63)	(2.65)
$Z_r^1 * 1_{2010}$	-59.73***	-4.71**	-4.51**	-23.27***	19.72***
	(10.00)	(1.75)	(1.50)	(3.79)	(2.91)
Z_r^2	-0.17	0.08*	0.15***	0.05	-0.08
	(0.21)	(0.03)	(0.04)	(0.05)	(0.07)
$Z_r^2 * 1_{2010}$	0.24***	0.03***	0.05***	0.08***	-0.07***
	(0.05)	(0.01)	(0.01)	(0.01)	(0.02)
Obs.	958	958	958	958	958
R^2	0.98	0.31	0.51	0.53	0.69

Notes: This table presents the results from the estimation of specification (3). Column (1) reports estimates for income growth, while Columns (2)–(5) examine structural transformation, as characterized in Table 3. Z_r^1 represents the ratio of higher education capacity to the youth population 20 years before the observed decadal change in outcomes. Z_r^2 captures the share of individuals under 15, 20 years before the observed decadal change in outcomes. 1_{2010} is a dummy variable indicating the 2000–2010 decade. All specifications include controls consistent with the baseline analysis. The complete list of controls, detailed in Table B.1 in Appendix B.2, excludes income-related variables from 1980 due to data limitations.

Results confirm that the instruments had no statistically significant effect on income growth before the reform-driven increase in skill supply, with significant effects emerging only in the post-reform period. A similar pattern holds for structural transformation: prior to the reform, Z_r^1 does not correlate with shifts in employment shares toward white-collar occupations and high-skill services. Some pre-reform effects appear for low-skill services and agriculture, but in the opposite direction of the post-reform period—regions with a relative abundance of colleges experienced a faster decline in agriculture's employment share and a faster rise in the employment share of low-skill services. For the subset of CZs not prioritized by the FCE, I use the 1991 youth share as a pre-reform shifter to predict the rise in college

[35]Two data limitations merit discussion. First, income data in the 1980 Census is poorly measured, with over 95% of municipalities lacking income information. This prevents a precise replication of the analysis in Section 3.3.2, as initial average income in 1980 and income growth between 1980 and 1991 cannot be consistently measured. Second, 1995 is the earliest year for which higher education capacity can be observed at the CZ level. The instruments for the pre-reform period are therefore constructed using 1980 youth population data and higher education capacity in 1995. As documented in Section 2, the number of colleges in Brazil remained largely unchanged between 1980 and 1996, suggesting that this limitation is unlikely to introduce significant bias.

attainment two decades later. Consistent with earlier findings, Z_r^2 shows no significant pre-reform effect on income or changes in employment shares for agriculture or low-skill services. However, it is positively associated with an increase in the employment share of white-collar and high-skill services in the pre-period. Table B.8 suggests that this increase is driven by the education sector, likely reflecting investments in higher education capacity between 1996 and 2000, such as hiring professors and administrative staff. Reassuringly, this shift in employment shares had no effect on average income growth in the pre-reform period.

To further assess the plausibility of the exclusion restriction, the analysis leverages the annual frequency of matched employer-employee data from 1985 to 2018, focusing on key moments of the firm size distribution examined in Section 3.3.3. Since the reform was enacted in 1996 and implemented in 1997, and college students typically take five years to graduate, the share of workers with a college degree is expected to increase starting in 2003. Using the IV specification, I estimate the effect of rising college attainment on firm outcomes for each year. The analysis examines two key outcomes: the growth in the prevalence of large firms hiring skilled workers and changes in the employment share of large firms. The empirical model is specified as follows:

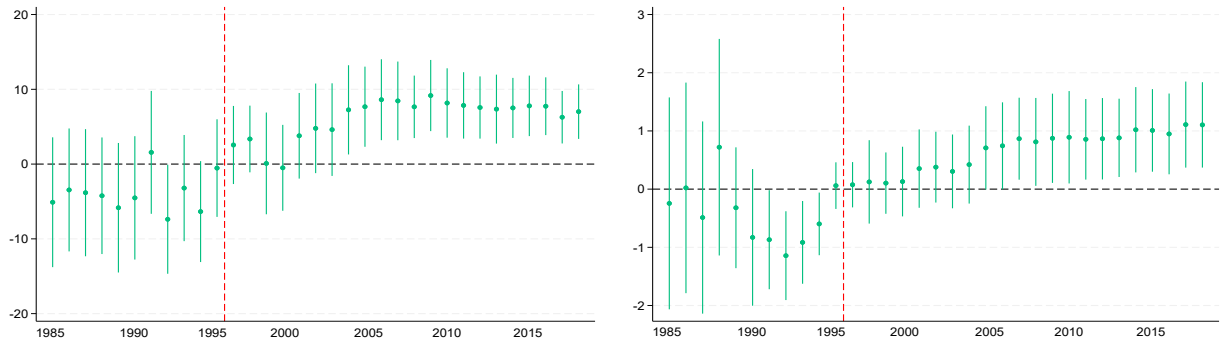
$$\begin{aligned} y_{rt} - y_{r1996} &= \theta_t + \beta_t \Delta h_r + X_r^{t-20}\Gamma_t + \theta_{rt}^s + u_{rt} \qquad \text{if } t > 1996 \\ y_{r1996} - y_{rt} &= \theta_t + \beta_t \Delta h_r + X_r^{t-20}\Gamma_t + \theta_{rt}^s + u_{rt} \qquad \text{if } t < 1996 \end{aligned} \tag{4}$$

Using an instrumental variable approach to generate exogenous variation in college attainment, equation (4) is estimated to assess its impact on firm outcomes.[36] Firm outcomes are measured relative to the year of the reform, $(y_{rt} - y_{r1996})$. Panel A examines the effect of college attainment on the growth in the number of large firms employing skilled workers, while Panel B focuses on changes in the employment share of large firms. All specifications include the full set of controls detailed in Table B.1. Following the approach of Dix-Carneiro and Kovak (2017), the empirical model allows the effects of college attainment, control variables, and state fixed effects $(\beta_t, \Gamma_t, \theta_{rt}^s)$ to vary flexibly over time. Figure 5 plots the sequence of β_t, with the y-axis showing coefficient values and 95% confidence intervals, the x-axis representing time in years, and a vertical red dotted line marking the year of the reform. For example, in Panel A, β_{2005} captures the effect of rising college attainment on the growth in the number of large firms employing skilled workers between 1996 and 2005, while β_{1985} measures the correlation between post-reform college attainment and the growth in large firms with skilled workers between 1985 and 1996. A positive and significant β_{1985} would indicate that regions experiencing post-reform growth in college attainment already exhibited faster growth in large firms hiring skilled workers before the reform, raising concerns about potential pre-trends.

Panel (a) shows no statistically significant correlation between increased college attainment and the growth in the number of firms employing skilled workers during the pre-reform

[36] College attainment is measured using Census data, available roughly every 10 years. The reform induced an expansion in college attainment between 2000 and 2010, the period used in specifications (4).

Figure 5: Results on Formal Firms



(a) Growth # large firms w/ skill

(b) Large firms' employment share

Notes: The Figure plots the IV coefficient estimates $\{\hat{\beta}_t\}$ from specification (4) with 95% CI. The $y-axis$ measures the coefficient values. The $x-axis$ measures time in years. Panel (a) depicts the implications for the growth in the number of large firms with skill, defined as those with at least six workers and hiring at least one worker with a college degree. Panel (b) shows the implications for the change in the employment share in large firms. In both cases, I measure the change with respect to its level in 1996, the year of the reform. The red vertical dotted line identifies the year of the reform, 1996.

period. This effect becomes positive and statistically significant after 2003. Panel (b) reveals a similar pattern for the employment share of large firms, which only begins to rise after 2004. While there is no statistically significant difference in the change in the employment share of large firms between 1985 (or 1990) and 1996 as a function of instrument-induced variation in the rise in college attainment, some heterogeneity emerges between 1991 and 1996, when regions with greater pre-reform college availability experienced an increase in the employment share of large firms.

One potential concern is that the establishment of colleges might directly contribute to the rise in large firms hiring college workers, as these institutions typically meet the criteria to be classified as large firms employing skilled workers. To address this, Appendix B.7.5 replicates Panels (a) and (b) while excluding firms in the higher education sector, and the results remain unchanged. Although the data do not entirely rule out the possibility that colleges influenced firm dynamics through local demand effects, the fact that the estimated effects become strong and significant only after 2003 suggests that the expansion in skilled labor supply is a key driver.

The results in this section provide evidence supporting the exclusion restriction assumption. First, the instruments exhibit no effect before the reform, indicating that their relevance emerges only in the post-reform period. Second, annual firm-level data reveal no evidence of pre-trends, reinforcing the validity of the identification strategy. Third, the timing of the estimated effects aligns with the entry of college graduates into the labor market, consistent with the hypothesis that the expansion in skilled labor supply is the primary driver of the results.

3.3.5 The Role of Migration

This section addresses the potential influence of migration across CZs on my findings. Recently Hsiao (2022) provided evidence that education investments facilitate migration to labor markets with better growth prospects. The systematic association between college graduate migration and the instrumental variation could be problematic. Specifically, while the instrument is designed to identify areas where colleges were most needed prior to the reform, migration patterns may instead reflect more favorable growth prospects in destination CZs. If the instrumental variation is correlated to migration patterns (reflecting unobserved growth prospect), the exclusion restriction assumption would not hold.

To analyze migration patterns among college graduates, I construct eight migration measures using data from the 2010 Population Census, which reports individuals' CZ of residence in both 2005 and 2010. First, I calculate the emigration rate as the share of college graduates who moved out of their 2005 CZ. Second, immigration rates are defined as the proportion of college graduates who moved into a CZ relative to the college graduate population residing in that CZ in 2005. Third, net migration is calculated as the difference between the number of immigrants and emigrants relative to the 2005 college graduate population in each CZ. Fourth, *within* CZ immigration rates are defined as the proportion of college graduates who moved across municipalities within the same CZ from 2005 to 2010. Finally, I replicate these metrics for all individuals, regardless of education level, to assess overall migration flows during the five years preceding 2010.

I estimate specification (2) to examine whether the instrumental variation systematically correlates with migration flows. Two potential scenarios could bias the IV coefficient. First, in regions where the instrument increased college attainment, *net emigration* of college graduates would "export" the human capital benefits to control regions, likely attenuating the observed impact of increased college attainment on development. In this scenario, the estimated effect of the rise in college attainment would be a lower bound. Second, *net immigration* of college graduates into regions where the instrument raised college attainment could introduce bias if these migration flows reflect better unobserved growth prospects. In this second case, the validity of the exclusion restriction would be compromised.

Appendix B.8 presents the results of this analysis in Table B.11, revealing three notable findings. First, increased college attainment has no significant impact on net migration across CZs, whether considering all cohorts or just young college graduates. It is important to emphasize that these results hold conditional on controls aimed to capture observable drivers of growth. Second, while the effect on net migration is close to zero, college attainment expansion does significantly affect both immigration and emigration among college graduates. Rising college attainment by 6% corresponds with faster immigration(emigration) that accounts for 7%(6%) of native college graduates. As a result, the net migration of college graduates is close to 0. This pattern is consistent across all cohorts, irrespective of educational status, with net migration remaining close to zero and non-significant. Third, consistent with the findings of Hsiao (2022), rising college attainment prompts migration

within the CZ (i.e. across municipalities within the same CZ). A 6% increase in college attainment induces migration flows that increase the share of migrants in the CZ by 13% among both young college graduates and broader cohorts.[37]

These findings provide reassuring evidence that migration patterns—whether among college graduates or the broader population—do not systematically correlate with the instrument, suggesting that cross-CZ migration does not threaten the identification strategy. Instead, the expansion in college attainment primarily induces *within*-CZ migration for both groups, an interesting result that warrants further exploration in future research. The analysis also corroborates that CZs are a suitable unit of analysis in this context.

3.3.6 Robustness

During the period of rising college attainment, other forces influenced economic development across regions. First, Brazil's early 1990s trade liberalization, examined by Dix-Carneiro and Kovak (2017) and Dix-Carneiro, Goldberg, et al. (2021), had long-term effects on labor markets and economic development. Second, the adoption of genetically engineered soybeans boosted agricultural productivity, with implications for structural transformation and development, as analyzed by Bustos, Caprettini, and Ponticelli (2016) and Bustos, Garber, and Ponticelli (2020). Third, Engbom and Moser (2022) and Haanwinckel (2023) investigate how minimum wage policies shaped labor market outcomes, including earnings inequality. These factors could be correlated with the rise in college attainment, potentially biasing the findings discussed. Appendix B.9 assesses the robustness of the results after controlling for these influences and finds the estimates remain stable.

3.3.7 Taking Stock

The empirical analysis provides novel evidence on the role of human capital in fostering local development. Rising college attainment promotes income growth. The benefits are not only captured by young college graduates but reach workers who do not hold a college degree. Specifically, increased college attainment drives a substantial reallocation of labor away from agriculture and elementary occupations towards higher-paying jobs in the service sector. Results shed light on a mechanism whereby college expansion facilitates firms' expansion, increasing the prevalence of large firms in the labor market and reallocating labor toward these large firms.

These results are informative of the "relative" effect of an exogenous increase in college attainment for local development. In the presence of general equilibrium effects and cross-CZ spillovers, the empirical estimates would not be fully informative about the aggregate growth consequences of increasing the supply of skilled workers in the economy. To quantify these aggregate effects, I propose a multi-region structural model, where each region corresponds to a CZ in the data.

[37] This within-CZ migration effect deserves further exploration in future research.

The empirical findings inform the structural analysis in two key ways. First, they motivate the core elements and assumptions of the model in Section 4. Second, they provide well-identified empirical moments—specifically, the $\hat{\beta}$ coefficients—which serve as *non-target* moments for validating the quantitative model. Since each region in the model corresponds to a CZ in the data, the model can generate a counterfactual distribution of CZs, allowing for the estimation of empirical models comparable to those in Section 3.2 using model/generated data. Finally, the quantitative model enables a counterfactual analysis in which we keep only the access costs to higher education unchanged to quantify the role of reduced college costs in Brazil's development between 2000 and 2010.

4 Theory

This section develops a static general equilibrium model in which college attainment, income, the sectoral allocation of labor, and the firm size distribution are endogenously determined. The elements and assumptions of the model are motivated by the empirical evidence.

The model features R regions and two tradable goods: Agriculture (A) and Non- Agriculture (NA).[38] Each region r produces a variety of both goods. Consumers derive utility from consuming regional varieties of these goods. The production factors are skilled and unskilled labor, where skilled workers are individuals with a college degree. Workers do not migrate across regions, and markets are frictionless and competitive.

The model is built on three key components. First, individuals heterogeneous in ability decide whether to invest in college education based on local labor market returns and regional costs of accessing higher education. This framework captures how demand and supply forces shape education decisions. Second, the NA sector comprises heterogeneous firms employing skilled and unskilled workers. Firms operate using a non-homothetic production technology, which establishes a positive relationship between skilled labor and firm size. This relationship can be interpreted as reflecting the efficiency gains achieved through improved production organization, particularly through the delegation to skilled workers. Third, the demand side of the economy is modeled using non-homothetic preferences of the PIGL class, which accommodate income effects critical to explain the process of structural transformation.

4.1 College Choice

I introduce a static model of college choice influenced by the approach in Kaboski (2009). Individuals are heterogeneous in ability and decide whether to pursue college education based on their prospective income (e_r^i). The model assumes perfect capital markets, abstracting from financial constraints. Prospective income is specified as follows:

[38] I group the manufacturing and service sectors together as Non-Agriculture (NA). This assumption allowsto remain neutral regarding the productive linkages between these two sectors. It also implies that the productivity measure for the NA sector captures combined productivity in both services and manufacturing.

$$e_r^i = \begin{cases} w_r^h \epsilon_{ir} - w_r^h \mathbf{C_r}, & \text{if College Graduate} \\ w_r^l, & \text{otherwise} \end{cases}$$

Here, w_r^h and w_r^l denote skilled and unskilled wages in region r, respectively, while $\mathbf{C_r}$ captures regional barriers that increase the effective cost of pursuing higher education. Individual ability, ϵ_{ir}, is distributed according to $F(\epsilon)$, assumed to be identical across regions. The individual $\bar{\epsilon}_r$, who is indifferent between pursuing or not pursuing a college education, determines regional college attainment:

$$h(\bar{\omega}_r, \mathbf{C_r}) = 1 - F\left(\bar{\epsilon}_r\right), \quad \text{where} \quad \bar{\epsilon}_r = \frac{1}{\bar{\omega}_r} + \mathbf{C_r}$$

Here, $\bar{\omega}_r = \frac{w_r^h}{w_r^l}$ is the skill premium in r. Market forces that increase the skill premium ($\bar{\omega}_r$) or policies that reduce access costs ($\mathbf{C_r}$) incentivize higher education investments, thereby increasing college attainment and the supply of skill to the labor market. The efficiency units of skilled labor supplied in region r, H_r^s, are given by:

$$H_r^s = E\left[\epsilon_{ir} | \epsilon_{ir} \geq \bar{\epsilon}_r\right] h(\bar{\omega}_r, \mathbf{C_r}) N_r$$

Here, $E\left[\epsilon_{ir} | \epsilon_{ir} \geq \bar{\epsilon}_r\right]$ denotes the average efficiency units of skill supplied by individuals who choose to invest in education. Regional education investments involve a cost, $\mathbf{C_r} h(\bar{\omega}_r, \mathbf{C_r}) N_r$, measured in efficiency units of skilled labor.[39] This framework accounts for changes in the ability composition of skilled workers as college attainment rises, which influences the average ability of the skilled labor pool. The remaining population consists of unskilled workers, L_r, who allocate their time to various roles in the labor market, as detailed below.

4.2 Technology

Each location has two tradable sectors: Agriculture (A) and Non-Agriculture (NA). The A sector employs only unskilled workers (L), while the NA sector employs both skilled (H) and unskilled (L) workers. The technology for the agricultural sector is linear:

$$Y_r^a = A_r^a L_{ar} \tag{5}$$

Here, A_r^a is the exogenous productivity fundamental in the agricultural sector. For the Non-Agricultural sector, I extend the framework of Melitz (2003) to establish a connection between firm size and skilled labor. I assume each location has a continuum of heterogeneous firms that maximize profits using a Stone Geary (non-homothetic) technology:

[39] For instance, college professors devote their time to educating students rather than participating in private-sector management.

$$\pi(z_{ir}) = P_r^{na} y_{ir} - w_r^l \left(\bar{\omega}_r h_{ir} + l_{mir} + f_r^o\right)$$
$$\text{s.t.} \quad y_{ir} = z_{ir}^{1-\alpha-\beta} l_{mir}^{\alpha} \left(\Gamma_r h_{ir} + \kappa\right)^{\beta} \tag{6}$$

In this setup, z_{ir} represents firm productivity, and I assume $\alpha + \beta < 1$, reflecting diminishing returns to scale in managerial control. Firms employ unskilled labor (l_{mir}) and skilled labor (h_{ir}), with the skill-biased component of technology captured by the exogenous parameter Γ_r. The parameter κ introduces non-homotheticity into the technology, representing managerial resources the entrepreneurs contribute. I assume κ is uniform across all firms and regions. When $\kappa > 0$, firm size and skill intensity are positively associated, consistent with the empirical relationship in Figure B.4. Given productivity z_{ir}, a firm planning to produce $\bar{y}$ minimizes costs by choosing the following skill intensity:

$$\frac{h_{ir}}{l_{mir}} = \max\left\{\Phi_1(\bar{\omega}_r) - \kappa\Phi_2(\Gamma_r, \bar{\omega}_r)\frac{z_{ir}^{\phi_1}}{\bar{y}^{\phi_2}}, 0\right\} \tag{7}$$

Here $\Phi_1(\bar{\omega}_r) = \frac{\beta}{\alpha\bar{\omega}_r}$, $\Phi_2(\Gamma_r, \bar{\omega}_r) = \frac{1}{\Gamma_r}\left(\frac{\beta\Gamma_r}{\alpha\bar{\omega}_r}\right)^{\frac{\beta}{\beta+\alpha}}$, $\phi_1 = \frac{1-\alpha-\beta}{\alpha+\beta}$ and $\phi_2 = \frac{1}{\alpha+\beta}$. When $\kappa = 0$, the model reverts to the Cobb-Douglas case, where optimal skill intensity is independent of both production scale, $\bar{y}$, and firm productivity, z_{ir}. However, when $\kappa > 0$, skill intensity becomes scale-dependent. Specifically, for a given productivity level z_{ir}, small-scale production does not require skilled workers, while large-scale production (i.e., high $\bar{y}$) increasingly relies on employing skilled workers. One interpretation of this relationship centers on managerial delegation: smaller firms often rely on entrepreneurs to handle administrative tasks such as management, accounting, and marketing. As firm size grows, entrepreneur's time constraints and the complexity of these tasks require delegation to skilled workers. The interaction between κ, the efficiency and relative costs of skilled workers shapes the firm size distribution, as discussed below.

The number of firms is determined by free entry, governed by two fixed costs. Firms enter the market whenever the expected value of net profits exceeds the entry cost. Specifically, each firm must pay an entry cost of $w_r^l f_r^e$ to draw productivity from a distribution $G_r(.)$, where I assume $G_r(z) = 1 - \left(\frac{B_r}{z}\right)^{\lambda}$. After paying the entry cost, firms observe their productivity level z_{ir} and proceed with operations only if net profits, after covering overhead costs $w_r^l f_r^o$, are positive. Both entry and overhead costs are measured in units of unskilled labor (L), reflecting different roles unskilled workers play in the production process.[40]

Let z_r^* denote the lowest productivity at which production is profitable, and $\bar{z}_r$ the threshold beyond which firms find it optimal to hire skilled labor. Define $\zeta_r := \frac{\bar{\omega}_r\kappa}{\Gamma_r f_r^o}$, which is proportional to $\bar{\omega}_r$, the only endogenous variable in the definition. Proposition 1 formalizes firm behavior, productivity gains from hiring skilled labor, and the role of the skill premium in shaping the firm size distribution.

[40]For instance, f_r^e represents unskilled workers involved in entrepreneurial activities or supporting firm entry, while f_r^o captures unskilled workers engaged in overhead-related activities.

Proposition 1 *Profit maximization yields the following results:*

(i) Define $\tilde{\beta}_1(\bar{\omega}_r) = \frac{\beta}{\alpha\bar{\omega}_r}$. *Skill composition:*

$$\frac{h_{ir}}{l_{mir}} = \begin{cases} 0 & \text{for } z_{ir} \in [z_r^*, \bar{z}_r) \\ \tilde{\beta}_1(\bar{\omega}_r)\left(1 - \frac{\bar{z}_r}{z_{ir}}\right) & \text{for } z_{ir} \geq \bar{z}_r \end{cases}$$

(ii) Define $\tilde{\beta}_2\left(\frac{w_r^l}{P_r^{na}}, \zeta_r\right) = \frac{w_r^l}{P_r^{na}}\frac{\zeta_r f_r^o}{\beta}$. *Production:*

$$y(z_{ir}) = \begin{cases} \tilde{\beta}_2\left(\frac{w_r^l}{P_r^{na}}, \zeta_r\right)\left(\frac{z_{ir}}{\bar{z}_r}\right)^{\frac{1-\alpha-\beta}{1-\alpha}} & \text{for } z_{ir} \in [z_r^*, \bar{z}_r) \\ \tilde{\beta}_2\left(\frac{w_r^l}{P_r^{na}}, \zeta_r\right)\frac{z_{ir}}{\bar{z}_r} & \text{for } z_{ir} \geq \bar{z}_r \end{cases}$$

(iii) Role of skill premium:

$$\frac{\partial z_r^*(\bar{\omega}_r)}{\partial \bar{\omega}_r} < 0 \qquad \frac{\partial \bar{z}_r(\bar{\omega}_r)}{\partial \bar{\omega}_r} > 0$$

Proof of Proposition 1 is in Appendix C.1-C.2.

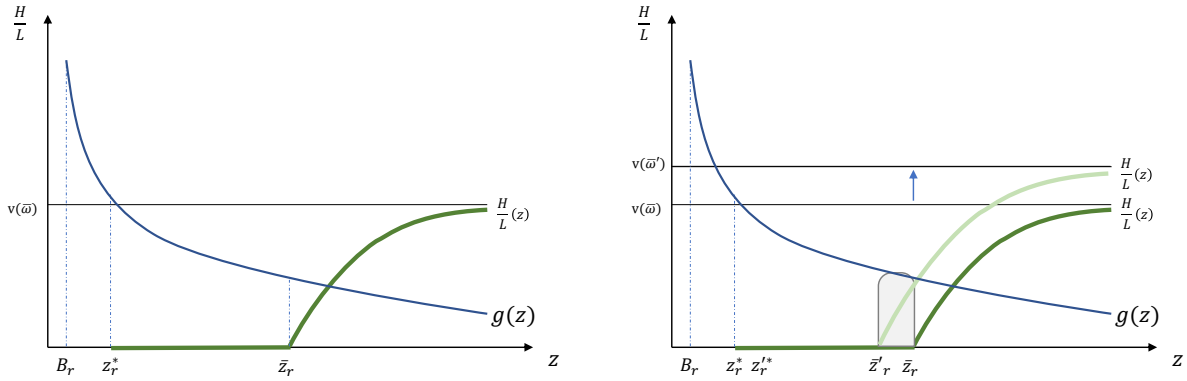
Proposition 1 characterizes the optimal skill composition and production. Firms with productivity between z_r^* and $\bar{z}_r$ rely solely on the entrepreneur's managerial skills. In contrast, firms with productivity above $\bar{z}_r$ hire skilled workers and skill intensity increases with productivity. Production $y(z_{ir})$ falls into two cases: firms with productivity $z_{ir} \in [z_r^*, \bar{z}_r)$ experience diminishing returns to productivity due to the limited span of control assumption, while firms with productivity $z_{ir} \geq \bar{z}_r$ hire skilled workers, allowing production to scale with productivity. This transition from decreasing to constant returns to z_{ir} represent the extensive margin of the productivity gains accrue to employing skilled workers.

When $\mathbf{C_r}$ is high, skilled labor is scarce and costly, shaping the firm size distribution. Figure 6 illustrates this relationship. In Panel (a), the blue curve depicts the Pareto density function of productivity (i.e. $g_r(z_{ir}) = G_r'(z_{ir})$), while the green line shows skill intensity as a function of firm productivity, z_{ir}. The black line indicates the limit of skill intensity as $z_{ir} \to \infty$, corresponding to a model with Cobb-Douglas production ($\kappa = 0$). Panel (b) highlights the impact of increased college attainment driven by policies that reduce college costs ($\mathbf{C_r}$). A larger supply of skilled workers lowers the skill premium, allowing more firms to hire skilled labor. This shift results in two key outcomes: (i) an increase in the number of firms at the margin that employ skilled labor (shaded region) and (ii) a rise in skill intensity among firms already employing skilled workers. Together, these dynamics lead to a greater prevalence of large, skill-intensive firms, and an increase in their employment share.

Next, I derive the aggregate behavior of sector NA in each region. In this competitive environment, some entrepreneurs earn positive profits, while others incur losses and remain inactive even after incurring the entry cost.[41] In equilibrium, expected profits are equal to

[41] This outcome holds under Assumption B. Appendix C.3 provides a detailed derivation of this condition.

Figure 6: Labor Composition and Firm's Productivity



Notes: In Panel (a) the $x-axis$ measures the productivity of the firm z_{ir}. There are three thresholds: B_r is the productivity floor in the region r, z_r^* is the productivity at which firms find it profitable to start operations, $\bar{z}_r$ is the productivity at which firms find it profitable to hire skilled workers H. The $y-axis$ measures the firm's optimal skill intensity (dark green line) as a function of z_{ir}, the Pareto probability distribution of productivity z_{ir} (blue line), the limit of skill intensity as $z_{ir} \to \infty$ (black line). Panel (b) represents the consequences of a fall in the skill premium ($\bar{\omega}_r$) for the optimal skill composition. When the $\bar{\omega}_r$ falls, the new optimal skill composition shifts to the left (light green line). The shaded region represents the extensive margin of firms that hire skilled workers after the fall in the skill premium.

the entry cost.[42]

$$E[\pi(z_{ir})] = w_r^l f_r^e$$

Proposition 2 derives expressions for local productivity and production in sector NA:

Proposition 2 *(i) In equilibrium, productivity and employment in sector NA is given by:*

$$A_r^{na} = A^{na}(\underbrace{\bar{\omega}_r}_{(-)}, \underbrace{\Gamma_r}_{(+)}, \underbrace{B_r}_{(+)})$$

$$\mathbf{L_{na,r}} = L_{mr} + L_{or} + L_{er} + \bar{\omega}_r H_r$$

(ii) Local production in sector NA can be written as:

$$Y_r^{na} = A_r^{na} \mathbf{L_{na,r}}$$

Proof of Proposition 2 is in Appendix C.2.

Proposition 2 provides a set of analytical expressions describing local productivity and employment in sector NA by integrating optimal choices across heterogeneous firms. Productivity in this sector (A_r^{na}) is negatively associated with the skill premium ($\bar{\omega}_r$), reflecting

[42] I assume that profits are pooled into a mutual fund managed by households. Local profits thereby cover local entry costs $w_r^l f_r^e M_r$, where M_r is the total number of firms including those inactive in equilibrium.

the adverse consequences of the scarcity of skilled labor, and positively associated with technology fundamentals, Γ_r and B_r. Local production function in sector NA takes a closed-form solution, linear in both the productivity fundamental A_r^{na} and the labor index $\mathbf{L_{na,r}}$, similar to the agricultural sector. Notably, A_r^{na} is endogenous and shaped by the scarcity of skill in the labor market.

4.3 Preferences

Following Boppart (2014), the model incorporates income effects by assuming that individuals' preferences belong to the PIGL class. These non-homothetic preferences are particularly well-suited to the analysis for three reasons. First, they allow for the aggregation of individual demand. Second, they align individual and regional demand parameters, enabling the estimation of preference parameters directly from household expenditure data. Third, as proved by Fan, Peters, and Zilibotti (2023), these preferences establish a natural mapping between preferences for final goods and value-added under certain assumptions. This mapping is especially important because the production function in the model is expressed in value-added terms, while the estimation of preference parameters relies on expenditure data on final goods.[43] A detailed derivation of local demand for each sector and variety is provided in Appendix C.4. Using these preferences, local demand for sector NA in region r:

$$P_r^{na} C_r^{na} = \underbrace{\left[(1-\gamma) - v\Psi_r \left[\frac{\mathrm{E}_r}{\left(P_r^a\right)^{\gamma} \left(P_r^{na}\right)^{1-\gamma}} \right]^{-\varepsilon} \right]}_{\vartheta_r^{NA}} \mathrm{E}_r N_r \tag{8}$$

Here, $\mathrm{E}_r = w_r^l \left[F(\bar{\epsilon}_r) + \bar{\omega}_r \left(E[\epsilon^i | \epsilon^i > \bar{\epsilon}_r] - \mathbf{C}_r \right) h(\bar{\omega}_r, \mathbf{C}_r) \right]$ represents the average income in region r, N_r represents the workforce of r, and ϑ_r^{NA} is the expenditure share of region r on sector NA. Expression (8) shows that individual and local demand share a similar structure, adjusted for inequality, $\Psi_r = \frac{E_r\left[\left(e^i\right)^{1-\varepsilon}\right]}{\mathrm{E}_r^{1-\varepsilon}}$.[44] Note that when $v = 0$, the model reverts to the homothetic case, nesting this more common specification within the PIGL framework. After determining expenditure shares for each sector aggregate, I assume that individuals exhibit CES preferences over the regional varieties within each sector, with an elasticity of substitution across varieties denoted by σ.

[43] Herrendorf, Rogerson, and Valentinyi (2013) emphasizes the distinction between using value-added and final consumption measures in analyzing the process of structural transformation.

[44] By Jensen's inequality, Ψ_r is less than 1, and:

$$E_r\left[\left(e^i\right)^{1-\varepsilon}\right] = \left(w_r^l\right)^{1-\varepsilon} \left[F(\bar{\epsilon}_r) + \left(\bar{\omega}_r\right)^{1-\varepsilon} E\left[\left(\epsilon^i - \mathbf{C}_r\right)^{1-\varepsilon} | \epsilon^i > \bar{\epsilon}_r \right] h(\bar{\omega}_r, \mathbf{C}_r) \right]$$

4.4 General Equilibrium

Regional varieties are produced in competitive markets, where the price of each regional variety r sold in region k reflects both the marginal production cost and iceberg costs τ. The CES demand structure implies the following equilibrium conditions, which hold across all regions r and sectors s:

$$w_r^l L_{sr} = \sum_{k=1}^{R} \frac{\left(w_r^l \tau_{kr} (A_r^s)^{-1}\right)^{1-\sigma}}{\sum_j \left(w_j^l \tau_{kj} (A_j^s)^{-1}\right)^{1-\sigma}} \vartheta_k^s \mathrm{E}_k N_k \tag{EQ1}$$

In equilibrium, the supply of sector s in region r must equal the total national demand for this variety. National demand depends on the expenditure shares of each region on sector s and the comparative advantage of region r in producing the sector variety s. Each region hosts two distinct labor markets: one for skilled workers and another for unskilled workers. Labor markets equilibrium, for all regions r, are given by:

$$\begin{aligned} E\left[\epsilon_{ir} | \epsilon_{ir} \geq \bar{\epsilon}_r\right] h(\bar{\omega}_r, \mathbf{C_r}) N_r &= H_r^d + \mathbf{C_r} h(\bar{\omega}_r, \mathbf{C_r}) N_r \\ F\left(\bar{\epsilon}_r\right) N_r &= L_r^d \end{aligned} \tag{EQ2}$$

As noted above, the term $\mathbf{C_r} h(\bar{\omega}_r, \mathbf{C_r}) N_r$ represents the local costs of education investments, expressed in efficiency units of skilled labor in region r.[45] The local supply of unskilled labor is distributed between agricultural and non-agricultural activities. In the non-agricultural sector, unskilled labor performs various roles, including entrepreneurial work (L_{er}), overhead labor (L_{or}), and manufacturing duties (L_{mr}). Thus, total unskilled demand is given by $L_r^d = L_{ar} + L_{mr} + L_{or} + L_{er}$. This system determines the equilibrium allocations of labor $\{L_{ar}, \mathbf{L_{na,r}}\}$ and prices $\{w_r^l, \bar{\omega}_r\}$.

4.5 Consequences of reducing access costs to college

Proposition 3 formalizes the implications of reducing $\mathbf{C_r}$ for the local economy.

Proposition 3 *A reduction in* $\mathbf{C_r}$ *has the following theoretical implications:*

1. *Increases the supply of skilled workers.*
2. *Reduces the equilibrium skill premium* $\bar{\omega}_r$.
3. *Increases the share of large firms hiring skilled workers.*

[45] This assumption is not uncontroversial. If the expansion of higher education capacity were driven by the federal government, it could lead to a substantial reallocation of resources across regions to cover the local costs, making the assumption less appropriate. However, in the case of Brazil, this concern is mitigated because the private sector primarily led the expansion of higher education capacity after the deregulation reform. Private universities charged tuition to local students to recover investment costs, meaning that the financial burden of capacity expansion fell largely on local residents. While federal education subsidies existed, they only became significant after 2005. In the structural estimation, I focus on the 2000–2010 period, when federal subsidies were less of a concern for the analysis.

4. Increases the employment share of large firms.
5. Increases average income in the economy.
6. Induces employment reallocation from sector A to NA.
The proof of Proposition 3 is in Appendix C.5.

Reducing access costs to education expands the supply of skilled workers. Predictions 1 through 3 reflect how this increased supply of skilled labor lowers the skill premium $\bar{\omega}_r$, enabling more firms to hire skilled labor, expand their operations, and become more productive, as shown in Proposition 1. Both the extensive and intensive margins contribute to this increase in the use of skilled labor: more firms begin employing skilled labor, and firms already employing skilled labor increase their skill intensity. These dynamics result in firms hiring more labor and growing in size. As a result, the employment share of relatively large firms hiring skilled workers rises, aligning with prediction 4.

Aggregating productivity gains at the firm level across the size distribution leads to a rise in A_r^{na}, as stated in Proposition C.2.The rise in productivity stems from two sources. First, a greater number of firms in the economy employ skilled workers, transitioning from decreasing to constant returns to scale. Second, skilled labor—initially scarce—has the highest marginal productivity of labor. As the supply of skilled labor expands, the labor factor with the highest productivity increases, further enhancing overall productivity. The resulting productivity gains raise average income, which, through non-homothetic preferences, shifts expenditure shares toward the NA sector. Additionally, a higher A_r^{na} strengthens the competitive position of the NA sector from region r in the national market. Together, these forces reallocate employment from sector A to sector NA. Importantly, All these theoretical predictions are consistent with the empirical findings in Section 3.3.[46]

5 Structural Estimation

This section outlines the structural estimation and calibration of the model. The procedure involves identifying a set of structural parameters, which are common across all regions r, and recovering a set of economic fundamentals, which vary by region r and time t:

Structural Parameters: $\Omega^1 = \{\underbrace{F(\,\cdot\,;\delta)}_{\textit{Ability}}; \underbrace{\Pi(\bar{\tau})}_{\text{Trade Costs}}; \underbrace{\gamma, \upsilon, \varepsilon, \sigma}_{\text{Preferences}}, \underbrace{\alpha, \beta, \lambda, \kappa}_{\text{Technology}}\}$

[46]The model makes some abstractions, which I briefly outline and discuss further in Appendix C.6. First, I assume that changes in the supply of skills do not affect technology fundamentals, B_r and Γ_r. However, consistent with Aghion et al. (2009), if education facilitates technological adoption, these fundamentals could plausibly depend on college attainment. Second, the model abstracts from worker mobility across labor markets, a simplification motivated by the empirical results. Finally, I abstract from capital market imperfections that may influence individuals' decisions to pursue college education. Appendix C.6 provides a detailed discussion of these abstractions. I discuss the implications of these assumptions for estimation in Section 5.

Economic Fundamentals: $\Omega^2 = \{A_r^a, B_r, \Gamma_r, f_r^o, f_r^e\}$

The procedure develops in two steps. First, I estimate the structural parameters Ω^1, which consist of four groups: (i) parameters governing the ability distribution, (ii) the matrix of trade costs, (iii) preference parameters, and (iv) technology parameters. Second, given the estimated structural parameters Ω^1, I recover the region-specific fundamentals Ω^2. This step leverages the equilibrium structure of the model, along with observed data on the distribution of firms, labor allocation, wages, and income gains from a college degree.[47]

5.1 Identification of Structural Parameters

In this section, I outline the procedure to estimate Ω^1, with detailed descriptions provided in Appendix D.1.

Ability distribution The estimation leverages model conditions and rich administrative data that link pre-college ability measures, college enrollment, and labor market earnings to estimate the ability distribution, $F(\cdot\,;\delta)$.

Trade costs Recovered using data on travel times across Brazilian labor markets, following the methodology of Alder (2016) to map travel times to trade costs.

Preference parameters The income elasticity parameter, ε, is estimated using household expenditure surveys, while the elasticity of substitution parameter, σ, is set to 5 based on trade literature estimates. The parameter γ is calibrated using the U.S. agricultural employment share, and v is identified by matching the aggregate observed decline in Brazil's agricultural employment between 2000 and 2010.

Technology parameters The parameter α is recovered from manufacturing and non-agricultural employment data, and β is estimated based on the skill intensity in the US. The tail parameter, λ, is identified using moments from Brazil's firm size distribution. Since κ cannot be separately identified from regional fundamentals, its value is normalized to 1.

5.2 Identification of Economic Fundamentals

This section outlines the strategy for calibrating the distribution of economic fundamentals in Ω^2 across CZs and over time. A comprehensive description is provided in Appendix D.2.

First, I use model equations governing college attainment, skill demand, and income gains from earning a college degree. Using these equations and rich administrative data, I

[47] This approach follows methods similar to Gancia, Müller, Zilibotti, et al. (2013) and Fan, Peters, and Zilibotti (2023) for recovering productivity fundamentals.

Table 6: Structural Parameters

Parameters	Target	Value
	Ability Distribution	
$F(\cdot;\delta)$	Returns to ability in administrative data	$\log\mathcal{N}$ (0,0.12)
	Preference Parameters	
ε	Household expenditure survey (POF, 2008)	0.3
γ	Agriculture employment share, US, 2010	0.01
v	Economic Transformation Database (ETD)	0.6
σ	Set exogenously	5
	Production Function Parameters	
α	Manufacture employment share, Brazil, 2000	0.12
β	Skill intensity in the US	0.58
λ	Tail of employment distribution, RAIS, 2000	1.3
κ	Normalization	1

recover the distributions of returns to skill ($\bar{\omega}_{r,t}$), college costs ($\mathbf{C}_{r,t}$), and the product of the skill-biased technology component and overhead fixed costs ($\Gamma_{r,t}f^o_{r,t}$).

Next, I combine administrative data on firms with the model's structure to calibrate the distributions of overhead fixed costs ($f^o_{r,t}$), entry costs ($f^e_{r,t}$), and the skill-biased technology component ($\Gamma_{r,t}$). This step uses data on the share of firms hiring skilled workers and the total number of firms in the NA sector for each region r.

Finally, following Fan, Peters, and Zilibotti (2023), I calibrate the productivity distributions for both sectors, $A^a_{r,t}$ and $A^{na}_{r,t}$, using equilibrium conditions from (EQ1). This procedure relies on data on sectoral labor allocation and observed earnings. The model's equilibrium structure uniquely maps productivity fundamentals and observed data. Once $A^{na}_{r,t}$ is obtained, I apply Proposition 2, which links $A^{na}_{r,t}$ to $B_{r,t}$, to calibrate the distribution of $B_{r,t}$. A detailed derivation of the method is provided in Appendix D.2.

Figure 7 presents access costs to college ($\mathbf{C}_r$) in 2000 and their changes between 2000 and 2010 as a function of per capita GDP in 2000. Since I infer college costs from observed attainment, these values reflect access conditions roughly five years before individuals entered higher education. The results indicate that poorer CZs faced higher initial college costs. However, between 2000 and 2010, these costs declined more sharply in poorer regions. In other words, areas with the highest initial barriers to college access saw the largest reductions over the decade. This pattern aligns with the intuition behind the empirical strategy, suggesting that college attainment expanded more rapidly in regions where access to higher education was initially more restricted. Figure D.1 in Appendix D.2 presents the full set of calibrated technology fundamentals.

Figure 7: Calibrated access costs to college, $\mathbf{C}_r$



Notes: The figure illustrates the relationship between college costs and initial economic development, measured as log GDP per capita in 2000. The x-axis represents 2000 log GDP per capita, while the left y-axis displays the calibrated college cost $\mathbf{C}_r$ in 2000. The right y-axis shows the log change in college costs between 2000 and 2010. Each marker corresponds to a distinct labor market (CZ).

5.3 Model Validation

This section evaluates the quantitative performance of the model by creating a counterfactual Brazil. Specifically, I compute an equilibrium where technology fundamentals remain fixed at their 2000 values ($\hat{A^a}_{r,2000}, \hat{\Gamma}_{r,2000}, \hat{B}_{r,2000}, \hat{f^e}_{r,2000}, \hat{f^o}_{r,2000}$) but college costs are set to their 2010 levels ($\hat{C}_{r,2010}$). This exercise generates equilibrium allocations for wages ($w^l_r, \bar{\omega}_r$), labor ($L_{ar}, L_{mr}, L_{er}, L_{or}, H_r$), and the firm size distribution. Using the equilibrium allocation in 2000 and the counterfactual allocation with reduced college costs, I reproduce key outcomes: the rise in college attainment, income growth, changes in returns to education, employment reallocation, and shifts in the firm size distribution.[48] Similar to Section 3, I evaluate the effects of expanded college attainment—driven by a reduction in education costs—on these outcomes using specification 1:

$$\Delta y_r = \theta_0 + \beta \Delta h_r + X_r \Gamma + v_r \tag{9}$$

Here, X_r includes controls capturing the 2000 level of development, such as population (in logs), average income, college attainment, agricultural employment share, and state fixed effects. This strategy underscores the complementarity between the model and empirical analysis. I simulate a counterfactual Brazil where the only change is a reduction in college costs and estimate specification 9 in this setting. To validate the model's ability to replicate reality, I require an empirical environment where an exogenous increase in college attainment—driven by lower college costs—can be credibly identified. The empirical analysis exploits cross-sectional variation induced by a reform that lowered access costs to education,

[48] Appendix C.5 provides details on constructing firm size distribution outcomes.

resulting in a heterogeneous expansion in college attainment. By accounting for potential demand shifts, the instrumental variable analysis provides key moments to assess the model's quantitative performance.

Two aspects of this strategy warrant emphasis. First, there is no reason to expect the model to replicate the results observed in Section 3 as the model estimation does not rely on the regression results. The model imposes a structure and the counterfactual allocations are generated within that framework. The coefficients from the model-generated data will only coincide with the empirical results if the model is a good representation of reality. Second, the error term in the model-generated data, v_r, captures potential misspecification introduced by assuming a linear relationship between changes in college attainment and outcomes. Under the null hypothesis that the model accurately reflects reality, this same misspecification would also affect the empirical analysis in Section 3, where a linear relationship was similarly assumed.

Table 7 presents the results of this validation exercise. Columns (1) and (2) reproduce the findings from Section 3.3, with Column (1) showing the OLS estimates and Column (2) reporting the IV results. Columns (3) and (4) provide regression coefficients derived from the model-generated data. Column (3) uses a counterfactual where college costs are reduced from their 2000 levels to their 2010 levels, holding all other fundamentals constant. Column (4), in contrast, examines a counterfactual where technology fundamentals are updated to their 2010 levels, while college costs remain fixed at their 2000 levels.

The results indicate that the model-generated coefficients closely approximate the effects of a supply-driven increase in college attainment on key outcomes, including changes in average income, the reduction in the skill premium, the decline in the agricultural employment share, the rise in the share of firms that are large and hire skilled workers, the employment share within those firms, and the growth in the number of large firms. Additionally, when specification 9 is estimated using model-generated data based on changes in technology fundamentals, the resulting coefficients are both qualitatively and quantitatively distinct. The model successfully replicates the IV coefficients only under a simulated reduction in college costs. These findings suggest that the model is well-suited to quantify the aggregate effects of changes in college access.

6 Growth Decomposition

This section examines the role of the decline in access costs to college in Brazil's economic growth and structural transformation between 2000 and 2010. Specifically, it explores how Brazil's development trajectory would have differed had college costs remained at their high levels observed in 2000. The analysis focuses on three outcomes: GDP growth, employment reallocation away from agriculture, and the rise in college attainment.

To address this question, I construct a counterfactual version of Brazil in 2010, in which all technology fundamentals—$\{A^a_{r,t}, B_{r,t}, \Gamma_{r,t}, f^o_{r,t}, f^e_{r,t}\}_{t=2010}$—are held at their respective 2010

Table 7: Comparison between model Regression and Empirical Regression

Results	Section 3.3		Model counterfactuals	
Specification	OLS	IV	$\mathbf{C}_r$	Technology
	(1)	(2)	(3)	(4)
Δ Average Income	1.43	1.64	2.17	1.83
	(0.19)	(0.73)	(0.09)	(0.26)
Δ Skill premium	-3.42	-13.37	-4.74	3.19
	(1.41)	(6.48)	(0.17)	(0.84)
Δ Agricultural employment share	-0.53	-1.09	-1.04	0.30
	(0.10)	(0.36)	(0.05)	(0.15)
Growth # Large Firms	2.20	7.71	12.47	1.41
	(0.65)	(2.44)	(0.77)	(1.23)
Δ Share of firms large w/ skill	0.35	0.83	1.12	0.12
	(0.09)	(0.30)	(0.26)	(0.59)
Δ Employment share in large firms	0.16	0.90	1.14	0.24
	(0.06)	(0.40)	(0.07)	(0.15)

Notes: The table presents regression coefficients with standard deviations reported in parentheses. Column (1) provides the OLS coefficients, while Column (2) reports the IV coefficients from Table 2 and Table 3 in Section 3. Columns (3) and (4) show regression coefficients estimated from specification (9) using model-generated data. Column (3) presents a counterfactual analysis where college access costs are updated to the calibrated 2010 distribution, $\mathbf{C}_{r,2010}$, while holding all other technology fundamentals constant. Using the 2000 baseline and this counterfactual, the variables required to estimate specification (9) are constructed to estimate the effect of rising college attainment—driven by lower college access costs—on economic development within the model. Column (4) replicates the exercise in Column (3) but instead updates technology fundamentals to their calibrated 2010 values. In the model, large firms are approximated as those with productivity $z_{ir} \geq \bar{z}_r$, corresponding to firms that employ college-educated workers.

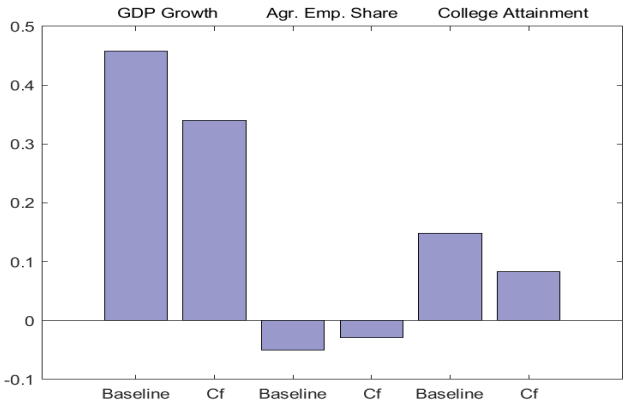
levels, while the distribution of college costs, $\{\mathbf{C}_{r,t}\}_{t=2000}$, is reverted to its 2000 values. I then compute income growth, the decline in agricultural employment, and the rise in college attainment relative to the 2000 baseline. Comparing these counterfactual trajectories to the observed baseline allows me to quantify the relative contributions of the reduction in access costs to college and technological progress to economic development.[49]

Figure 8 presents the results. The first two columns examine GDP growth, the next two track the decline in agricultural employment, and the last two measure the rise in college attainment. For each outcome, the first column reports the observed change, while the second column presents the counterfactual scenario in which college costs remain at their 2000 levels, but technology fundamentals evolve as dictated by the calibration.

The results indicate that, had access costs to college remained at their 2000 levels, Brazil would have experienced 22% lower GDP growth, a 46% slower decline in agricultural employment, and a 43% smaller increase in college attainment. Reducing these costs rose GDP growth by 1 percentage points (pp) per year and GDP per capita growth by 0.96 pp. In my framework, general equilibrium forces operate through two main mechanisms. First, when region experiences income growth, it generates spillovers by increasing demand for goods

[49]This exercise is similar in nature to the approach in Kaboski (2009), which quantifies the role of declining education costs in the U.S. in driving long-run growth.

Figure 8: The Contribution of Higher Education to Economic Development



Notes: The first two columns represent GDP growth, the next two track the decline in agricultural employment, and the last two measure the rise in college attainment. For each outcome, "Baseline" columns depict the observed decadal change for Brasil, while "Cf" columns depict the counterfactual change. In the "Cf" scenario, outcomes are recomputed using a counterfactual 2010 Brazil in which college access costs remain fixed at their calibrated 2000 levels, while all other technology fundamentals evolve according to their recovered calibration for 2000 and 2010.

produced in other regions. Due to the non-homotheticity of preferences, this increase in demand is biased toward non-agricultural goods, further enhancing education investments in other regions. Second, a reduction in access costs to college makes sector NA in a given region more productive (see Proposition 2), strengthening its comparative advantage in the national market while reducing the market share of other regions. The interaction of these opposing forces results in ambiguous predictions about the net influence of general equilibrium dynamics on aggregate growth. Estimates from Section 3.3 show that rising college attainment increases average income growth by 0.7 pp. The structural analysis reveals that reducing access costs to college accounts for 0.96 pp of aggregate GDP per capita growth, suggesting that general equilibrium forces amplify income growth by 0.26 pp.

My findings rank at the upper end of previous studies. For instance, Bils and Klenow (2000) finds that a one-year increase in average educational attainment raises annual GDP growth by approximately 0.3 pp. Similarly, Kaboski (2009) finds that declining education costs explain an average of 0.5 pp of annual wage growth in the U.S. More recently, Hsiao (2022) show that Indonesia's INPRES school construction program increased aggregate output by 6%. More closely related to my study, Aghion et al. (2009) finds that increasing US state-level funding for four-year colleges raises annual economic growth between 0.3 and 0.7 pp, with stronger effects in states closer to the technological frontier. Likewise, Valero and Van Reenen (2019) show that a 10% increase in the number of universities raises annual regional GDP per capita growth by 0.4 pp.

However, direct comparisons between my analysis and these studies present challenges. First, previous studies infer access costs to education from measures such as average years of education, the number of universities, or per-student funding, whereas my structural approach infers the reduction in access costs to college based on the expansion of college

attainment, conditional on the income gains from holding a college degree and the prevailing distribution of ability. Second, during the period of analysis, Brazil experienced robust economic growth of 4% per year, suggesting that the benefits of expanding educational attainment may be amplified in periods of strong economic growth.[50]

The reduction in access costs to college, by rising college attainment, also influences labor allocation, accounting for 46% of the decline in the agricultural employment share. This strong effect aligns with the quantitative calibration, as the recovered distribution of college costs, $\mathbf{C}_r$, suggests that poorer regions—where agricultural employment shares were higher—experienced a faster decline in education access costs, as shown in Figure 7. College-educated workers predominantly enter urban labor markets and, as demonstrated in Section 3.3, contribute to job creation in the non-agricultural sector. My findings align with the existing literature. For instance, Porzio, Rossi, and Santangelo (2022) estimate that rising human capital levels accounted for 40% of the global labor reallocation away from agriculture over the 20th century. While my study provides novel evidence on the role of college education on structural transformation focusing on Brazil, their analysis demonstrates that similar patterns hold across a broader set of economies worldwide.[51]

A recent related study by Engbom, Malmberg, et al. (2024) develops a structural framework where skill scarcity plays a central role in firm organization. Similar to my model, their framework shows that hiring skilled workers to perform a larger share of tasks brings the firm closer to constant returns to scale, leading to larger firm sizes. Their counterfactual simulations suggest that the rise in skilled labor can explain the reorganization of production in large firms who rely more intensively on white-collar occupations. Importantly, they demonstrate that standard theories of structural transformation—centered on aggregate and sectoral productivity growth and non-homothetic preferences—fail to explain these patterns.

The two studies offer complementary perspectives. My work establishes a causal link between rising college attainment, structural transformation, and the reorganization of production into larger firms in Brazil. The authors use these estimates to validate their quantitative model. At the same time, their cross-country analysis provides evidence that these mechanisms operate globally, not just in Brazil. Together, these studies underscore the central role of higher education in driving structural change and shaping the organization of production in large firms.

[50] Other studies emphasize education's role in explaining differences in *income levels*, providing further context for these results. Klenow and Rodriguez-Clare (1997) and Caselli (2005) suggest that human capital accounts for only a limited share of cross-country income differences, while Gennaioli et al. (2013), Manuelli and Seshadri (2014), Jones (2014), and Hendricks and Schoellman (2018) find a more substantial role for human capital in shaping income disparities.

[51] When accounting for general equilibrium effects, the rise in human capital accounts for 20% of the global decline in the agricultural employment share. Similarly, Caselli and Coleman II (2001) underscore the role of expanding educational access in reducing agricultural employment and fostering regional convergence in earnings across regions in the U.S.

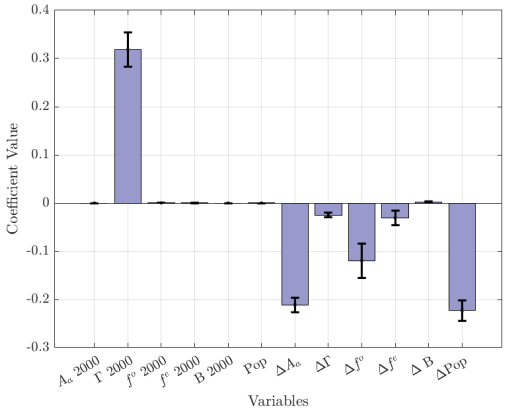
Regional heterogeneity The model features a multi-region economy with trade linkages, allowing me to analyze both the local gains from reducing access costs to college and the heterogeneity of these effects. Specifically, I examine why some labor markets (CZs) benefit more than others from lower college costs. To quantify this variation, the counterfactual exercise estimates the share of GDP growth attributable to the reduction in college access costs in each CZ. Figure E.1 presents a scatter plot of these shares against baseline GDP per capita in 2000 (panel (a)) and the percentage reduction in college costs (panel (b)). Two key patterns emerge. First, the contribution of reduced college costs to GDP growth varies substantially across regions. In particular, this contribution is smaller in poorer areas, suggesting that other constraints may have limited the gains from expanding college attainment, or that alternative growth drivers played a more dominant role during this period. Second, I observe substantial heterogeneity in how GDP growth responds to reductions in college costs, emphasizing the importance of region-specific economic conditions in shaping the impact of educational expansion.

What explains this heterogeneity? I examine the role of technology fundamentals in shaping the economic gains from lowering college access costs. Specifically, I estimate the following specification at the CZ level:

$$\Delta y_r = \theta_0 + \mathbf{A}_r\theta_1 + \Delta\mathbf{A}_r\theta_2 + u_r \tag{10}$$

where Δy_r represents the share of GDP per capita growth attributable to lowering college access costs, $\mathbf{A}_r$ denotes the vector of technology fundamentals and population levels in 2000 $(A_r^a, \Gamma_r, B_r, f_r^e, f_r^o, Pop_r)$, and $\Delta\mathbf{A}_r$ captures the calibrated changes in fundamentals and population growth between 2000 and 2010. Figure 9 presents the estimated coefficients.

Figure 9: Heterogeneity in income growth gains



Notes: The figure presents the estimated coefficients from specification (10) with 95% confidence intervals. Technology fundamentals include agricultural productivity (A_r^a), the Pareto productivity floor (B_r), overhead fixed costs (f_r^o), fixed entry costs (f_r^e), and population growth (Pop_r). The terms x 2000 represent the initial value of each technology fundamental in 2000, while Δx captures its change between 2000 and 2010.

The empirical estimates reveal important patterns. The contribution of lower college

costs to GDP growth is larger in labor markets with higher initial skill-biased technology (Γ_r) and where reductions in fixed overhead costs (Δf_r^o) and fixed entry costs (Δf_r^e) are greater. Conversely, when technological progress in agriculture (ΔA_r^a) is faster, skill-biased technical change ($\Delta\Gamma_r$) is more pronounced, or population growth (ΔPop) is higher, the role of lowering college costs in driving local economic growth diminishes.

These results highlight key complementarities. First, education policies and skill-biased technologies reinforce one another. Lowering barriers to higher education yields larger economic benefits in regions where access to productivity-enhancing technologies is more widespread.[52] Second, the expansion of college attainment has a stronger impact on growth when firms face lower set-up costs. Specifically, reductions in overhead labor and entry costs amplify the positive effects of an increasingly skilled workforce. These costs often reflect broader institutional and financial frictions, such as credit constraints or regulatory barriers, which limit firms' ability to expand. For instance, Ponticelli and Alencar (2016) and Bustos, Garber, and Ponticelli (2020) provide evidence that improvements in credit access in Brazil lead to significant increases in firm borrowing, investment, and employment. Similarly, Ponczek and Ulyssea (2022) provide evidence that stringent labor regulations can hinder firms' ability to adapt to economic shocks. Moreover, the World Bank's Brazilian Enterprise Survey indicates that 45% of firms identify access to finance as a severe constraint, while 48% cite regulatory bottlenecks—such as business licensing and permits—as significant obstacles.

Collectively, these findings suggest promising avenues for future research. Notably, they underscore the potential efficiency gains from coordinating education policies with reforms that foster technological adoption and mitigate firm-level constraints.

7 Conclusions

The role of human capital in economic growth and structural transformation remains at the center of a long-standing debate. While Lucas Jr (1988) and Galor and Moav (2004) view human capital accumulation as the primary engine of modern economic growth, Bils and Klenow (2000) suggest that education investments respond to productivity-driven economic expansion. Establishing a causal link between education and growth is often challenging due to concerns about reverse causality. However, the case of Brazil provides compelling new evidence on the role of higher education in fostering development.

I leverage a 1996 education reform in Brazil that lowered entry barriers for private for-profit colleges. I devise an IV strategy exploiting the idea that private colleges targeted locations under-served before the reform to create exogenous variation in the rise of college attainment. This approach identifies the *relative effects* of rising college attainment, which I find leads to income growth and a reallocation of labor from agriculture to services. Moreover, I document a novel mechanism through which higher education expansion operates: it

[52] For instance, previous research has linked the diffusion of technologies such as computers and the internet to skill-biased technical change. See Caselli and Coleman (2001) and Goldin and Katz (2009).

increases the prevalence of large firms and shifts employment toward them. I interpret these results as evidence that higher education supplies firms with skilled workers—particularly in managerial roles—who enhance productivity, enable firms to scale, and contribute to job creation.

To assess the *aggregate effects* of reducing college access costs, I develop and estimate a model where college attainment, the firm size distribution, income, and sectoral labor allocation are jointly determined. The model highlights a key mechanism: college-educated workers enable firm expansion, generating productivity gains that raise incomes and accelerate the shift of labor out of agriculture. I validate the model by comparing its predictions to the empirical evidence from the reform and find that declining college costs accounted for 22% of Brazil's GDP growth and 46% of the decline in agricultural employment between 2000 and 2010.

Additionally, I exploit the model's regional structure to uncover new evidence on the complementarity between education and skill-biased technological change. Expanding access to higher education yields the largest benefits in regions with greater exposure to skill-complementary technologies, such as computers and the internet. Moreover, I document a strong interaction between education expansion and reductions in firms' fixed set-up costs, which reflect broader frictions such as credit constraints and regulatory barriers that limit firm growth. These findings suggest that education policies can be particularly effective when coupled with reforms that alleviate firm-level constraints and foster the adoption of productivity-enhancing technologies. By combining empirical evidence with structural modeling, these results demonstrate that lowering barriers to college education plays a transformative role in economic development—not only by increasing human capital but also by reshaping firm structure and labor markets.

These findings open several promising avenues for future research. First, a recent paper by M. Schwartzman (2024) highlights how productivity growth, amplified by income effects, drives the modernization of the service sector. Understanding how rising skilled labor contributes to productivity growth in services remains an open question. Second, my results suggest that education expansion is more growth-enhancing when skill-biased technologies are present. The recent expansion of internet access provides a unique opportunity to study the complementarity between digital infrastructure and education policies in boosting firm productivity. Third, while I document that rising college attainment increases the prevalence of large firms in the service sector, an open question is whether these firms represent local entrepreneurs scaling up or productive firms from major cities expanding into new labor markets. Hsieh and Rossi-Hansberg (2023) document the spatial expansion of productive firms in the U.S. service sector, and my findings suggest that expanding the supply of skilled workers may facilitate a similar pattern in developing economies. Fourth, further research is needed on the interaction between education, credit markets, and regulatory constraints. If credit access and business regulations shape the productivity gains from education, then coordinated policy efforts could yield greater economic benefits than education policies alone.

References

Aghion, Philippe et al. (2009). "The causal impact of education on economic growth: evidence from US". *Brookings papers on economic activity* 1.1, pp. 1–73.

Akcigit, Ufuk, Harun Alp, and Michael Peters (2021). "Lack of selection and limits to delegation: firm dynamics in developing countries". *American Economic Review* 111.1, pp. 231–75.

Akresh, Richard, Daniel Halim, and Marieke Kleemans (2023). "Long-term and intergenerational effects of education: Evidence from school construction in Indonesia". *The Economic Journal* 133.650, pp. 582–612.

Alder, Simon (2016). "Chinese roads in India: The effect of transport infrastructure on economic development". *Available at SSRN 2856050*.

Arrow, Kenneth J (1973). "Higher education as a filter". *Journal of public economics* 2.3, pp. 193–216.

Autor, David H and David Dorn (2013). "The growth of low-skill service jobs and the polarization of the US labor market". *American economic review* 103.5, pp. 1553–1597.

Barro, Robert J and Jong Wha Lee (2013). "A new data set of educational attainment in the world, 1950–2010". *Journal of development economics* 104, pp. 184–198.

Bender, Stefan et al. (2018). "Management practices, workforce selection, and productivity". *Journal of Labor Economics* 36.S1, S371–S409.

Bils, Mark and Peter J Klenow (2000). "Does schooling cause growth?" *American economic review* 90.5, pp. 1160–1183.

Bloom, Nicholas, Erik Brynjolfsson, et al. (2019). "What drives differences in management practices?" *American Economic Review* 109.5, pp. 1648–1683.

Bloom, Nicholas, Benn Eifert, et al. (2013). "Does management matter? Evidence from India". *The Quarterly journal of economics* 128.1, pp. 1–51.

Bloom, Nicholas, Renata Lemos, et al. (2014). "JEEA-FBBVA Lecture 2013: The new empirical economics of management". *Journal of the European Economic Association* 12.4, pp. 835–876.

Boppart, Timo (2014). "Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences". *Econometrica* 82.6, pp. 2167–2196.

Buera, Francisco J and Joseph P Kaboski (2012). "The rise of the service economy". *American Economic Review* 102.6, pp. 2540–69.

Buera, Francisco J, Joseph P Kaboski, et al. (2022). "Skill-biased structural change". *The Review of Economic Studies* 89.2, pp. 592–625.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli (2016). "Agricultural productivity and structural transformation: Evidence from Brazil". *American Economic Review* 106.6, pp. 1320–65.

Bustos, Paula, Gabriel Garber, and Jacopo Ponticelli (2020). "Capital accumulation and structural transformation". *The Quarterly Journal of Economics* 135.2, pp. 1037–1094.

Caselli, Francesco (2005). "Chapter 9 Accounting for cross-country income differences". *Handbook of Economic Growth*, pp. 679–741.

Caselli, Francesco and Wilbur John Coleman (2001). "Cross-country technology diffusion: The case of computers". *American Economic Review* 91.2, pp. 328–335.

Caselli, Francesco and Wilbur John Coleman II (2001). "The US structural transformation and regional convergence: A reinterpretation". *Journal of political Economy* 109.3, pp. 584–616.

Chandler, Alfred D (1977). *The Visible Hand, The Managerial Revolution in American Business*. Harvard University Press.

Cordeiro, Fernando and Alvaro Cox (2024). "College Quality and Tuition Subsidies in Equilibrium".

Davis, Steven J, John C Haltiwanger, and Scott Schuh (1998). "Job creation and destruction". *MIT Press Books* 1.

Dix-Carneiro, Rafael, Pinelopi K Goldberg, et al. (2021). *Trade and Domestic Distortions: The Case of Informality*. Tech. rep. National Bureau of Economic Research.

Dix-Carneiro, Rafael and Brian K Kovak (2017). "Trade liberalization and regional dynamics". *American Economic Review* 107.10, pp. 2908–46.

Duflo, Esther (2001). "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment". *American economic review* 91.4, pp. 795–813.

Engbom, Niklas, Hannes Malmberg, et al. (2024). "Economic Development According to Chandler". *Technical report, mimeo, New York University*.

Engbom, Niklas and Christian Moser (2022). "Earnings inequality and the minimum wage: Evidence from Brazil". *American Economic Review* 112.12, pp. 3803–47.

Fan, Tianyu, Michael Peters, and Fabrizio Zilibotti (2023). "Growing Like India—the Unequal Effects of Service-Led Growth". *Econometrica* 91.4, pp. 1457–1494.

Foster, Andrew D and Mark R Rosenzweig (1996). "Technical change and human-capital returns and investments: evidence from the green revolution". *The American economic review*, pp. 931–953.

Galor, Oded and Omer Moav (2004). "From physical to human capital accumulation: Inequality and the process of development". *The Review of Economic Studies* 71.4, pp. 1001–1026.

Gancia, Gino, Andreas Müller, Fabrizio Zilibotti, et al. (2013). "Structural development accounting". In: *Advances in Economics and Econometrics: Tenth World Congress*. Vol. 2. Cambridge University Press Chicago, pp. 373–418.

Gennaioli, Nicola et al. (2013). "Human capital and regional development". *The Quarterly journal of economics* 128.1, pp. 105–164.

Goldin, Claudia and Lawrence F Katz (2009). *The race between education and technology*. harvard university press.

Haanwinckel, Daniel (2023). *Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution*. Tech. rep. National Bureau of Economic Research.

Hanushek, Eric A et al. (2015). "Returns to skills around the world: Evidence from PIAAC". *European Economic Review* 73, pp. 103–130.

Heckman, James J, Jora Stixrud, and Sergio Urzua (2006). "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". *Journal of Labor economics* 24.3, pp. 411–482.

Hendricks, Lutz and Todd Schoellman (2018). "Human capital and development accounting: New evidence from wage gains at migration". *The Quarterly Journal of Economics* 133.2, pp. 665–700.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi (2013). "Two perspectives on preferences and structural transformation". *American Economic Review* 103.7, pp. 2752–89.

Hjort, Jonas, Hannes Malmberg, and Todd Schoellman (2022). *The missing middle managers: Labor costs, firm structure, and development*. Tech. rep. National Bureau of Economic Research.

Hsiao, Allan (2022). *Educational Investment in Spatial Equilibrium: Evidence from Indonesia*. Tech. rep. Working Paper.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg (2023). "The industrial revolution in services". *Journal of Political Economy Macroeconomics* 1.1, pp. 3–42.

Jones, Benjamin F (2014). "The human capital stock: a generalized approach". *American Economic Review* 104.11, pp. 3752–3777.

Kaboski, Joseph P (2009). "Education, sectoral composition, and growth". *Review of Economic Dynamics* 12.1, pp. 168–182.

Khanna, Gaurav (2023). "Large-scale education reform in general equilibrium: Regression discontinuity evidence from india". *Journal of Political Economy* 131.2, pp. 000–000.

Klenow, Peter J and Andres Rodriguez-Clare (1997). "Economic growth: A review essay". *Journal of monetary economics* 40.3, pp. 597–617.

Lin, Dajun, Randall Lutter, and Christopher J Ruhm (2018). "Cognitive performance and labour market outcomes". *Labour Economics* 51, pp. 121–135.

Lucas Jr, Robert E (1978). "On the size distribution of business firms". *The Bell Journal of Economics*, pp. 508–523.

— (1988). "On the mechanics of economic development". *Journal of monetary economics* 22.1, pp. 3–42.

MacLeod, W Bentley and Miguel Urquiola (2019). "Is education consumption or investment? Implications for school competition". *Annual Review of Economics* 11.1, pp. 563–589.

Magalhães Castro, Maria H de (2015). "Higher education policies in Brazil: A case of failure in market regulation". *Higher education in the BRICS countries: Investigating the pact between higher education and society*, pp. 271–289.

Manuelli, Rodolfo E and Ananth Seshadri (2014). "Human capital and the wealth of nations". *American economic review* 104.9, pp. 2736–62.

Melitz, Marc J (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity". *econometrica* 71.6, pp. 1695–1725.

Montiel Olea, José Luis and Carolin Pflueger (2013). "A robust test for weak instruments". *Journal of Business & Economic Statistics* 31.3, pp. 358–369.

Nelson, Richard R and Edmund S Phelps (1966). "Investment in humans, technological diffusion, and economic growth". *The American economic review* 56.1/2, pp. 69–75.

Neves, Clarissa Eckert B (2015). "Demand and supply for higher education in Brazil". *Higher education in the Brics countries: investigating the pact between higher education and society*, pp. 73–96.

Nimier-David, Elio (2022). *Local human capital and firm creation: Evidence from the massification of higher education in France*. Tech. rep. Working Paper.[1537].

Pinto, José Marcelino de Rezende (2004). "O acesso à educação superior no Brasil". *Educação & Sociedade* 25, pp. 727–756.

Ponczek, Vladimir and Gabriel Ulyssea (2022). "Enforcement of labour regulation and the labour market effects of trade: Evidence from Brazil". *The Economic Journal* 132.641, pp. 361–390.

Ponticelli, Jacopo and Leonardo S Alencar (2016). "Court enforcement, bank loans, and firm investment: evidence from a bankruptcy reform in Brazil". *The Quarterly Journal of Economics* 131.3, pp. 1365–1413.

Porzio, Tommaso, Federico Rossi, and Gabriella Santangelo (2022). "The human side of structural transformation". *American Economic Review* 112.8, pp. 2774–2814.

Rocha, Roberto Hsu et al. (2017). "A relação entre o Ensino Superior público e privado e a renda e emprego nos municípios brasileiros".

Saldanha, Raphael (2024). *Road distances and trip duration matrix for Brazilian municipalities*. Tech. rep. Zenodo.

Sampaio, Helena (2000). "Ensino superior no Brasil: O setor privado". *Estudos Brasileiros*.

Schwartzman, Matthew (2024). "From Street Markets to Shopping Malls: The Transformation of the Service Sector". *Mimeo*.

Schwartzman, Simon (1998). *Higher education in Brazil: the stakeholders*. Vol. 28. World Bank, Latin America and the Caribbean Regional Office Santiago.

— (2004). "Equity, quality and relevance in higher education in Brazil". *Anais da Academia Brasileira de Ciências* 76, pp. 173–188.

— (2010). "The national assessment of courses in Brazil". *Public policy for academic quality: Analyses of innovative policy instruments*, pp. 293–312.

— (2015). "Demands and policies for higher education". In: *Higher Education in the BRICS Countries*. Springer, pp. 13–41.

Spence, M (1973). "Job Market Signaling". *The Quarterly Journal of Economics* LXXXVII, pp. 355–374.

Valero, Anna and John Van Reenen (2019). "The economic impact of universities: Evidence from across the globe". *Economics of Education Review* 68, pp. 53–67.

Vu, Khoa and Tu-Anh Vu-Thanh (2022). *Higher education expansion, labor market, and firm productivity in Vietnam*. Tech. rep. Working Paper.

World Bank, The (2012). *Brazil: How to Decrease Freight Logistics Costs in Brazil*.

A Institutional Context

A.1 Allocation of colleges before the reform

Starting in 1961, the allocation of colleges across the territory was supervised by the Federal Council of Education FCE. Figure 1 shows the allocation of colleges across the space. A significant share of Commuting Zones (CZs) was not prioritized to have a college. Table A.1 shows statistics for the subset of CZs prioritized and not prioritized to have a college.

Table A.1: Baseline characteristics of CZs with and without colleges, 1991

	With Colleges		With No Colleges	
Characteristic	Mean	Std. Dev.	Mean	Std. Dev.
Population	514	112	133	111
Average Income (log)	8.44	0.29	8.40	0.54
Agriculture Employment Share	0.25	0.15	0.48	0.15
High Skilled Services Employment Share	0.20	0.05	0.16	0.05
College Attainment	0.07	0.05	0.02	0.02
High School Attainment	0.12	0.04	0.07	0.03
N	215		271	

Notes: Labor markets (CZ) are categorized based on the stock of colleges, as reported in the 1995 Higher Education Census. A CZ is classified as "With Colleges" if it contains at least one college within its boundaries. Population data, measured in thousands, is obtained from the 1991 Demographic Census. Average income is reported in logs. Skilled services comprise employment in business, financial, health, public, and education services. College attainment represents the share of individuals in the CZ with a college degree, while high school attainment reflects the share of individuals who have completed high school.

A.2 Rise in college attainment at the municipality level

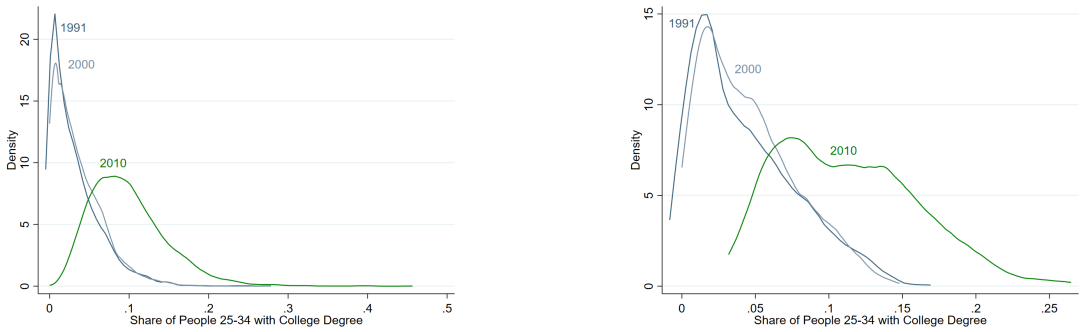
Figure A.1 depicts the distribution of college attainment among the youth at the municipality and CZ levels. In both cases, I observe a substantial rise in college attainment by 2010.

A.3 The 1996 education reform: quality of instruction, selection on ability and job profile of college graduates.

In this section, I discuss three additional features of the reform: (i) quality of instruction, (ii) type of skills and job profile of new college graduates, and (iii) implications of changes in ability composition.

The expansion of higher education capacity was largely driven by the entry of private providers, raising concerns about whether regulators could keep pace with market growth and ensure instructional quality (S. Schwartzman 2004). To assess the extent of these concerns, I analyze rich administrative data to compare instructional quality across private for-profit,

Figure A.1: Expansion in College Attainment



(a) Municipality (b) CZ

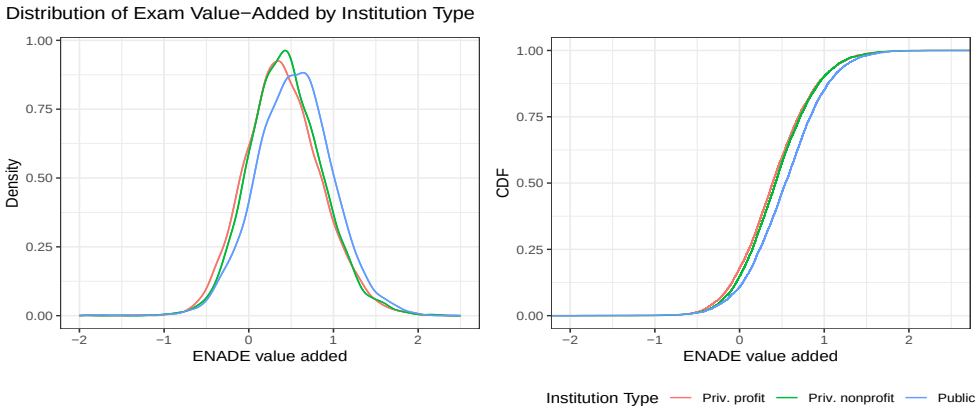
Notes: College attainment among the youth is calculated as the share of individuals aged 25 to 34 with a college degree for each municipality and labor market (CZ). The figure presents the distribution of college attainment across municipalities (Panel (a)) and CZs (Panel (b)). In both cases, the distributions are shown for 1991, 2000, and 2010, illustrating changes in the distribution over time.

private non-profit, and public institutions. Private non-profit and public institutions dominated the higher education landscape before the reform, while private for-profit institutions were legally permitted only afterward. The analysis leverages four key datasets: (1) ENEM scores, a standardized national exam widely used to allocate college slots and financial aid, to measure baseline ability; (2) program and college enrollment and graduation records; (3) ENADE scores, a compulsory standardized exam for graduating cohorts, to evaluate acquired knowledge; and (4) formal labor market wage data for college graduates. By linking these datasets, I estimate and recover college-program fixed effects, controlling for baseline ability ($ENEM$ scores) and household socioeconomic characteristics at enrollment (collected via the ENEM questionnaire). These fixed effects are interpreted as value-added measures of instructional quality. A detailed description of the data and the estimation method can be found in Cordeiro and Cox (2024).

Figures A.2 and A.3 plot the distribution of these fixed effects for public, private non-profit, and private for-profit programs. While private non-profit and for-profit institutions exhibit overlapping distributions, public colleges show higher average quality, as their distributions are shifted to the right. These patterns hold whether value-added is measured using ENADE scores or formal labor market earnings, consistent with the perception of public colleges as more prestigious in Brazil. However, the similarity between private for-profit and non-profit institutions suggests that the legal shift allowing colleges to operate for profit did not lead to an average decline in instructional quality, as indicated by the value-added estimates. Cordeiro and Cox (2024) validate these fixed effects as measures of instructional quality by linking them to faculty characteristics, including the share of professors with postgraduate degrees, full-time status, and average faculty earnings—all of which are positively and significantly associated with instructional quality. These findings highlight the importance of regulation and public disclosure of quality measures, which incentivized pri-

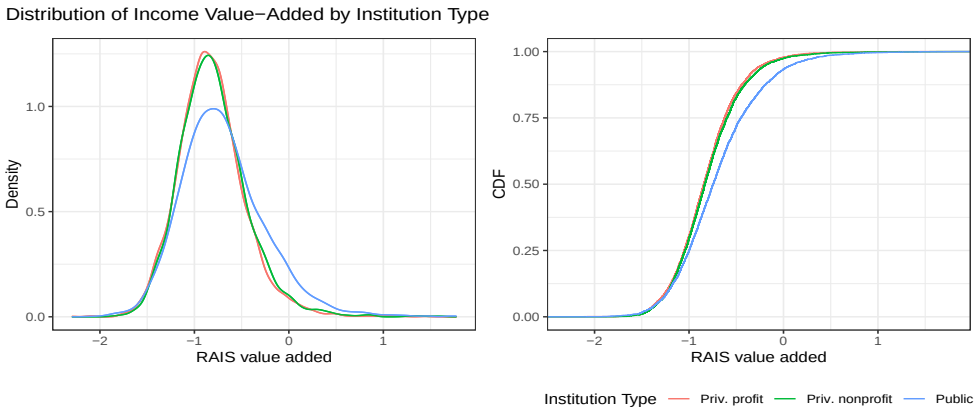
vate colleges to invest in quality, perform well in ENADE, and market their programs based on exam results.

Figure A.2: ENADE Performance: College level Fixed Effects



Notes: I regress college graduates' performance in the ENADE exit exam on pre-college ability measures, household background, and college-major fixed effects. Approximately 80% of all college graduates can be linked across these datasets. I interpret these fixed effects as measures of instructional quality and classify these fixed effects into three categories based on their administrative status: private for-profit (red), private non-profit (green), and public colleges (blue). The left panel presents the density distributions for each institution type, while the right panel displays the corresponding cumulative distribution functions.

Figure A.3: Labor Market Performance: College level Fixed Effects



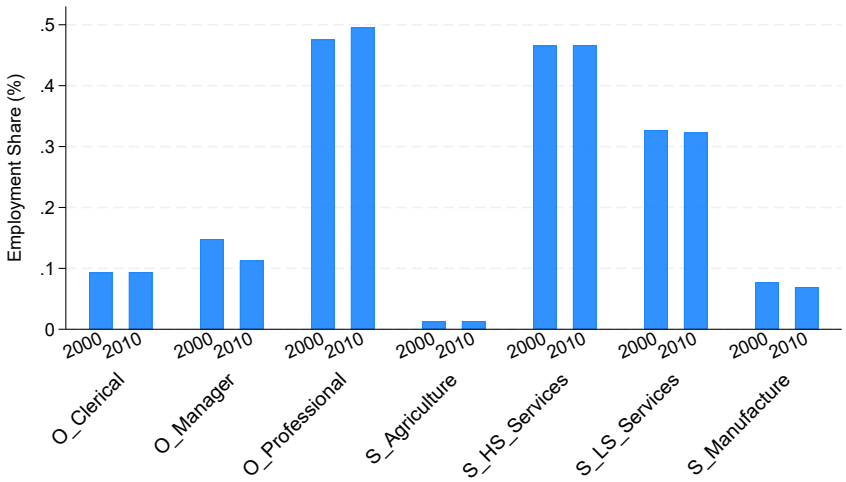
Notes: I regress college graduates' earnings in their formal employment on pre-college ability measures, household background, and college-major fixed effects. Approximately 50% of all college graduates can be linked across these datasets. I interpret these fixed effects as measures of instructional quality and classify these fixed effects into three categories based on their administrative status: private for-profit (red), private non-profit (green), and public colleges (blue). The left panel presents the density distributions for each institution type, while the right panel displays the corresponding cumulative distribution functions.

To conclude point (i), I assess the implications of a potential decline in instructional quality for the analysis in Section 3.2, which examines the impact of rising college attainment on the local economy. If new private colleges provided lower-quality instruction, their graduates would acquire less human capital and contribute less to the labor market than

cohorts from pre-reform institutions, where instructional quality was higher. Consider a benchmark scenario in which all new colleges maintained pre-reform instructional quality. Relative to this benchmark, an expansion in college attainment driven by lower-quality colleges would likely understate the true effect of increasing college attainment while preserving instructional quality.

Regarding point (ii), I analyze descriptive statistics on college graduates' fields of study and labor market outcomes before and after the reform, using data from the Higher Education and Population Censuses. First, I compute the distribution of college graduates aged 25–34 across fields of study. Panel (a) of Figure A.5 shows that the composition of majors remained largely stable, with notable increases in enrollment in administration and education-related fields at the expense of medicine, humanities, and social sciences. Second, Figures A.4 indicate that sectoral and occupational outcomes for college graduates in 2010 closely resembled those in 2000 (i.e., those who enrolled before the reform), with both cohorts primarily employed in managerial and professional roles in the service sector. Third, Panel (b) of Figure A.5 shows that job formality rates remained comparable across cohorts, suggesting that the reform did not significantly alter the labor market profiles of college graduates.[53]

Figure A.4: Sector/Occupation profile of old and new graduates

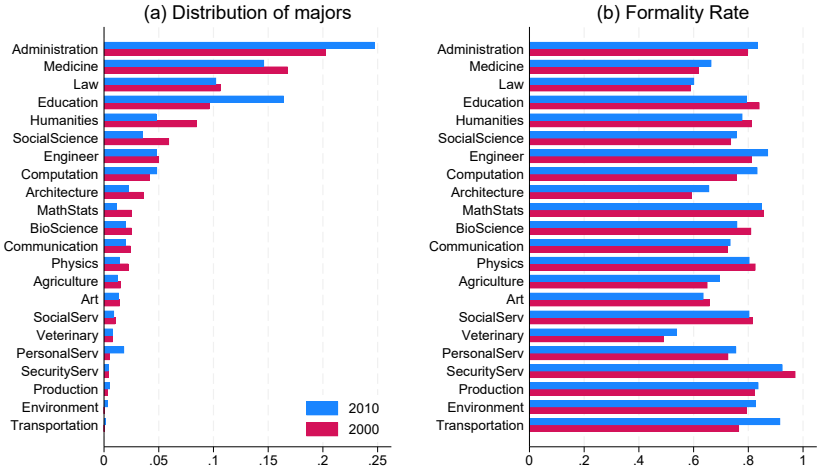


Notes: I present the occupational ($O_$) and sectoral ($S_$) distribution of college graduates aged 25–34 in the 2000 and 2010 census years. "HS_Services" refers to high-skill services, encompassing education, health, business and public services, while "LS_Services" refers to low-skill services, including retail, hospitality, transportation and utilities. The excluded occupational categories are blue-collar and other elementary occupations, while the excluded sectoral category is "other general services".

Finally, I address point (iii), which concerns potential changes in the ability composition of college graduates. As higher education expands, new enrollees may have lower baseline

[53]Formality includes employment in the formal sector as well as individuals reporting themselves as employers or business owners, excluding self-employment.

Figure A.5: Formal Job Status of College Graduates



Notes: Panel (a) displays the distribution of all college degrees across majors for college graduates aged 25 to 34 in the years 2000 and 2010. Panel (b) shows the formality status within majors for the same age group of college graduates in 2000 and 2010.

ability than pre-reform cohorts, assuming meritocratic selection before the reform. Since higher-ability graduates contribute more to the labor market, they supply more efficiency units of skill.[54] If college attainment expansion primarily comes from lower-ability graduates, the growth in efficiency units of skill will slow down. This mirrors the discussion on instructional quality: relative to a scenario where new enrollees maintain pre-reform ability levels, an increase in college attainment with declining ability has a weaker impact on local development. In other words, a decline in the marginal ability of new college graduates attenuates the developmental effects of rising college attainment.

The model in Section 4 explicitly accounts for shifts in the ability composition as college attainment increases. In its structural estimation, I use data from the 2009 ENEM exam to estimate the ability distribution in the model.

B Empirical Analysis

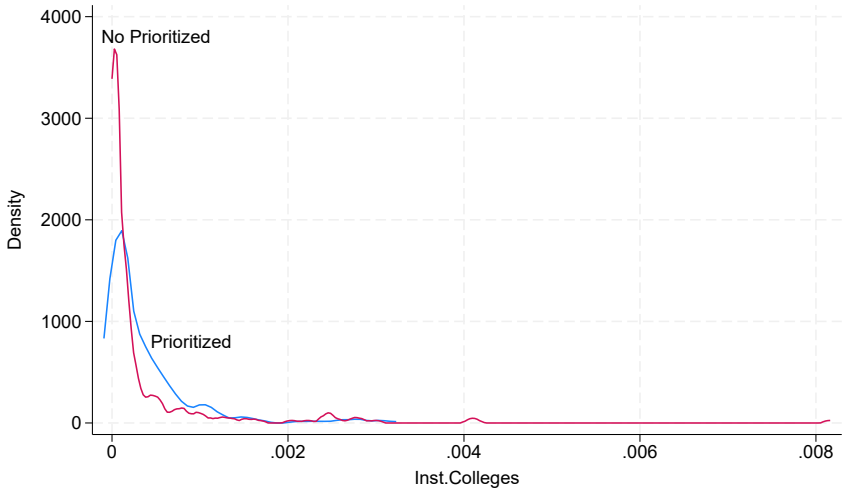
B.1 Variation of instrument across CZs with and without Colleges

B.2 Baseline Controls

Table B.1 lists the controls used in the empirical analysis outlined in Section 3.2, capturing initial development levels and pre-reform trends from 1991 to 2000. Economic development

[54] Conditional on earning a college degree, higher ability corresponds to higher wages and can be interpreted as supplying more efficiency units of skill.

Figure B.1: Density of Z^1 for prioritized and no prioritized CZs



Notes: The Figure displays the density of instrument Z_r^1 measuring the access to higher education capacity relative to the youth population for two sub-samples. The red density displays the distribution of Z_r^1 for the subsample of CZs that were not prioritized by the FCE to have a college. The blue density displays the distribution of Z_r^1 for the subsample of CZs that were prioritized by the FCE to have a college. Instrument Z_r^1 is defined in Section 3.2.

proxies include population size, average earnings, sectoral labor allocation, and the formality rate. Human capital proxies include the share of young adults with completed secondary and college education (ages 25–34). To account for regional trajectories, the trends of these variables between 1991 and 2000 are also included. Additionally, I control for the size and average income of the adult cohort (35 and older) as a proxy for financial frictions, under the assumption that a wealthier adult population faces fewer constraints in financing investments through savings or credit backed by higher collateral. Finally, the variable "College pre-reform" is a dummy indicating whether the CZ had a college prior to the reform, meaning it was among the regions prioritized by the FCE to host college institutions.

The inclusion of these controls ensures that regions with similar pre-reform development levels and trends are compared, while differences in the rise of college attainment are driven by the unexplained variation in the instruments. The instruments, defined in Section 3.2, exploit pre-reform scarcity of higher education capacity relative to potential demand. Conditional on controls, including the "College pre-reform" dummy, Z_r^1 generates variation for CZs prioritized to host college institutions, while Z_r^2 does so for CZs that were not. Both instruments capture differences in the pre-reform scarcity of higher education capacity relative to the potential demand for college education.

B.3 Estimation of College Premium

I obtain estimates of the average earnings gain from holding a college degree using the population census. For each CZ, I focus on the sample of individuals between 25 and 34 years old who have at least completed high school. I estimate the following specification:

Table B.1: Baseline Controls

Variables	Mean	St.Dev.	Min	Max	N
College pre-reform	0.44	0.50	0.00	1.00	486
Population (logs), 1991	11.99	0.97	7.43	16.27	486
Population growth, 1991-2000	0.12	0.13	-0.28	1.13	486
Average income (logs), 1991	12.25	0.47	10.86	14.40	486
Average income (logs), 2000	5.94	0.49	4.66	7.04	486
Sh.Employed in Agriculture, 1991	0.38	0.19	0.01	0.79	486
Δ Sh.Employed in Agriculture, 1991-2000	-0.07	0.05	-0.31	0.11	486
Sh.Employed in Manufacture, 1991	0.09	0.07	0.01	0.47	486
Δ Sh.Employed in Manufacture, 1991-2000	0.01	0.03	-0.11	0.14	486
Sh.High school (25-34), 1991	0.16	0.05	0.02	0.45	486
Δ Sh.High school (25-34), 1991-2000	0.03	0.03	-0.04	0.14	486
Sh.College (25-34), 1991	0.05	0.04	0.00	0.19	486
Δ Sh.College (25-34), 1991-2000	-0.01	0.02	-0.07	0.05	486
Formality rate, 1991	0.36	0.16	0.08	0.78	486
Δ Formality rate, 1991-2000	-0.01	0.05	-0.44	0.18	486
Adult population (logs)	10.72	1.02	6.19	15.24	486
Average income - adult cohort, 1991	12.31	0.49	10.82	14.45	486
Average income - adult cohort, 2000	6.01	0.51	4.72	7.21	486

Notes: The unit of analysis is the CZ. The Table lists the variables used as controls in the empirical analysis. For instance, "College pre-reform" is a dummy indicating whether the CZ had a college in 1995, and the variable "Sh. High school (25-34), 1991" measures the share of people in the 25 to 34 age range who have completed high school (secondary) education in 1991. I defined the adult population as people older than 31 in 1991 (and 50 years old in 2010). The average income is measured in *Cruzeiros* in 1991 and *Reais* in 2000. "Average income" calculates the average income of people older than 25 and measures it in logs. "Average income - adult cohort" calculates the average income of people older than 35 and measures it in logs.

$$y_{ir} = \theta_{0r} + \theta_{1r} D_{ir}^{HE} + X_{ir}\Gamma_r + \epsilon_{ir}$$

Where y_{ir} is the log earned income of individual i, D_i^{HE} is a dummy equal to 1 if the individual has a college degree. I add a set of predetermined controls like age, sex, race, and labor experience. This strategy allows me to obtain a distribution $\{\hat{\theta}_{1,r}\}_t$ of the "college premium" for each census year, 1991, 2000, and 2010. Because I estimate the "college premium" from cross-sectional data, changes in this estimate are influenced by market-level forces and changes in the average ability of college graduates. Section 5 describes how I use rich administrative data to estimate the ability distribution of college graduates to recover the variation in the "college premium" driven by market forces.

B.4 First-stage analysis

B.4.1 Alternative measures of higher education access, college entry, and college attainment

The empirical strategy described in Section 3.2 uses the interaction between the reform and a pre-reform measure of the scarcity of higher education capacity (HEC) relative to the potential demand for college education. These heterogeneous effects of the reform across CZs, conditional on controls, allow me to identify the impact of rising college attainment of the workforce on local economic development. There are alternative ways to measure the HEC in the pre-reform period. In this Subsection, I provide the results from the estimation of the model (2) using nine alternative measures of HEC_r, the numerator of Z_r^1.

First, I use three measures of HEC: colleges, programs, and slots.[55] Second, I use three methods to aggregate college capacity from municipality to CZ. The first type, A, is described in Section 3.2. The second measure, B, weights colleges within $100km$ by the distance traveled to enroll.[56] The third measure, C, is the raw number of colleges in the CZ. Alternatives B and C of defining HEC are both formally specified in (11):

$$HEC^B_{r,1995} = \sum_{m \in \Omega_r} \omega_m \sum_{k \in \Omega_m^{100km}} \frac{Colleges_{k,1995}}{(1 + dist_{km})^{0.4}} \;\; ; \;\; HEC^C_{r,1995} = \sum_{m \in \Omega_r} Colleges_{m,1995} \tag{11}$$

While these three measures are highly correlated, measure C may differ from A and B when individuals access colleges located within 100km of a CZ border. In other words, the 100km radius offers a more flexible measure of HEC, as it accounts for higher education capacity beyond administrative boundaries—particularly relevant for municipalities near HEC in neighboring CZs.

Table B.2 provides results for alternative measures of HEC. Independent of whether I measure capacity as colleges, programs, or slots, I find that places with the low relative scarcity of HEC pre-reform experienced a faster increase in college attainment after the reform, conditional on controls. Similarly, for those places with no colleges pre-reform, a higher share of youth in their population is associated with a faster increase in college attainment after the reform. These results are similar for all different types of aggregation. However, the results are not significant in Panel C if I measure higher education capacity in r as the total number of programs or slots in that CZ. This situation could arise if method C were under-measuring the relative scarcity of HEC, reducing the predictive power of the instrument. Overall, results provide a very consistent and robust indication that the mismatch between the government allocation of colleges and the future demand for colleges

[55] I define a Program as a major-campus combination. If a college has two campuses offering one major or one campus offering two majors, then both cases will count as two Programs.

[56] The functional form and elasticity 0.4 were obtained from a regression between observed college enrollment from municipality m in municipality k and the distance between both.

Table B.2: First Stage Analysis

Panel A: Higher education capacity within 100km									
Dep.Variable	Δh, 2000-2010								
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Z_r^1(Colleges)	-3.30**	-4.10**	-5.28***						
	(1.06)	(1.57)	(1.49)						
Z_r^1(Programs)				-0.46*	-0.61**	-0.77***			
				(0.22)	(0.23)	(0.22)			
Z_r^1(Slots)							-0.00*	-0.00*	-0.01**
							(0.00)	(0.00)	(0.00)
Z_r^2 :Youth Pop.Share	-0.04***	0.14**		-0.04***	0.14**		-0.04***	0.15***	
	(0.00)	(0.05)		(0.00)	(0.04)		(0.00)	(0.04)	
R^2	0.15	0.63	0.62	0.14	0.63	0.62	0.14	0.63	0.62
F-test	41.42	12.77	12.51	40.00	12.20	11.77	40.24	11.38	9.86
Panel B: Higher education capacity within 100km distance weighted									
Z_r^1(Colleges)	-3.92**	-4.87**	-6.26***						
	(1.26)	(1.86)	(1.77)						
Z_r^1(Programs)				-0.54*	-0.73**	-0.91***			
				(0.26)	(0.27)	(0.26)			
Z_r^1(Slots)							-0.00*	-0.00*	-0.01**
							(0.00)	(0.00)	(0.00)
Z_r^2 :Youth Pop.Share	-0.04***	0.14**		-0.04***	0.14**		-0.04***	0.15***	
	(0.00)	(0.05)		(0.00)	(0.04)		(0.00)	(0.04)	
R^2	0.15	0.63	0.62	0.14	0.63	0.62	0.14	0.63	0.62
F-test	41.42	12.79	12.56	40.01	12.21	11.78	40.24	11.37	9.82
Panel C: Higher education capacity in the CZ									
Z_r^1(Colleges)	-54.95	-68.35*	-46.73						
	(33.83)	(26.65)	(27.41)						
Z_r^1(Programs)				19.93	-2.76	1.12			
				(10.19)	(8.09)	(8.16)			
Z_r^1(Slots)							0.08	-0.10	-0.05
							(0.10)	(0.08)	(0.08)
Z_r^2 :Youth Pop.Share	-0.05***	0.18***		-0.03***	0.17***		-0.04***	0.18***	
	(0.01)	(0.04)		(0.01)	(0.04)		(0.01)	(0.04)	
R^2	0.14	0.63	0.61	0.15	0.62	0.61	0.14	0.62	0.61
F-test	38.89	9.69	2.91	41.64	7.77	0.02	39.91	8.34	0.31
Obs.	486	486	486	486	486	486	486	486	486
Control	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: Panel A measures HEC as specified in Section 3.2. Panel B measures HEC as HEC^B in definition 11. Panel C measures HEC as HEC^C in definition 11. Columns (1), (2) and (3) use colleges to measure HEC. Columns (4), (5) and (6) use programs to measure HEC. Columns (7), (8) and (9) use slots to measure HEC. Columns (1), (4), (7) provide the results with no controls. Columns (2), (5), (8) provide the results with controls. Additionally, columns (3), (6) and (9) provide the first stage results using Z_r^1 as the only instrument and include controls. The full list of controls is listed in B.1.

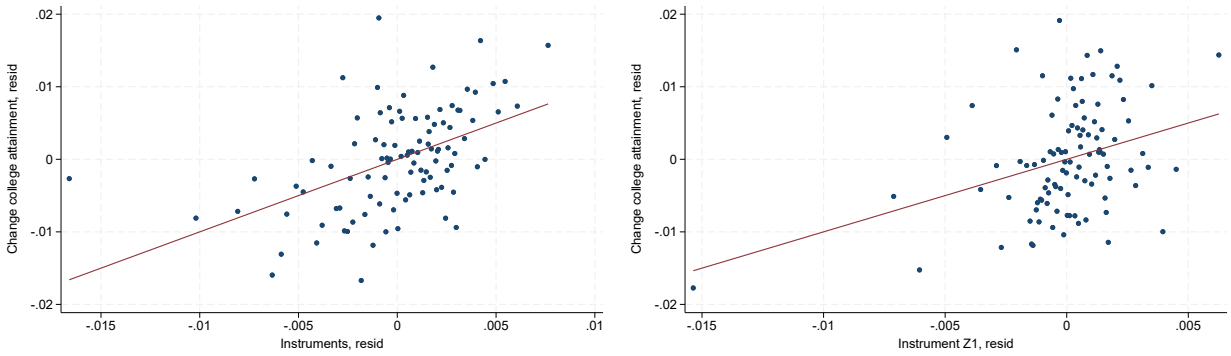
measured before the reform are good predictors of the rise in college attainment after I condition on relevant measures of the initial level of development and pre-trends of a CZ.

B.4.2 Variation in the rise of college attainment explained by instruments

Figure B.2 illustrates how variation in the instruments translates into variation in the endogenous regressor, the rise in college attainment. The y-axis displays the residuals from a

regression of Δh on state fixed effects and pre-reform development controls, listed in Appendix B.2. The x-axis represents the variation in the rise in college attainment explained by the residual variation in the instruments, after accounting for state fixed effects and controls.

Figure B.2: Residual variation in Δ college attainment explained by instruments



(a) Instruments Z_r^1 and Z_r^2 (b) Instrument Z_r^1

Notes: Panel (a) shows a scatter plot of the residual variation in the rise in college attainment (y-axis) and how much of this residual variation in the rise in college attainment is explained by the instruments Z_r^1 and Z_r^2 (x-axis), after I account for the variation in state fixed effects and controls. Panel (b) provides similar information when I use Z_r^1 as the instrument. Table B.1 presents the list of controls.

B.5 Main results

Table B.3: Expansion in College Attainment and Income Growth

Dep.Variable	Income growth						Δ College premium		
Cohort	All cohorts			Young Cohort			Young Cohort		
Specification	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\Delta h_{(25-34)}$	-0.03	1.43***	3.04**	0.34	1.59***	3.58*	1.11	-3.42*	-7.84
	(0.32)	(0.19)	(1.08)	(0.30)	(0.23)	(1.41)	(1.41)	(1.41)	(7.85)
R^2	0.62	0.88	0.86	0.57	0.82	0.78	0.27	0.45	0.44
Eff. F-test			12.51			12.51			12.51
CV 5%(10%)			37.42(23.11)			37.42(23.11)			37.42(23.11)
Obs.	486	486	486	486	486	486	486	486	486
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: I estimate specification (1) to assess the implications of rising college attainment on average income growth for all cohorts and the young cohort. These specifications use only Z_r^1 as the instrument. In the last three columns I assess the implications on the college premium. The unit of observation is the commuting zone, CZ. I consider three specifications for each sample. All specifications include state dummies. Columns (1) do not include controls. Columns (2) and (3) include controls aimed to capture pre-reform development patterns. Columns (3) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013) and the TSLS critical values at the 5% and 10% levels.

Table B.4: Expansion in College Attainment and Structural Transformation

Panel A: Occupations - All cohorts												
Dep.Variable	Change in employment share in											
Occupation	White-Collar			Blue-Collar			General Services			Elementary		
Specification	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\Delta h_{(25-34)}$	0.21**	0.45***	0.68***	-0.30***	-0.05	0.10	-0.11	0.08	0.35	0.20	-0.47***	-1.13**
	(0.06)	(0.05)	(0.20)	(0.08)	(0.07)	(0.29)	(0.06)	(0.05)	(0.20)	(0.15)	(0.09)	(0.40)
R^2	0.32	0.61	0.59	0.33	0.64	0.64	0.59	0.76	0.75	0.50	0.83	0.81
Eff. F-test			11.11			11.11			11.11			11.11
CV 5%(10%)			8.86(6.34)			9.71(6.84)			11.20(7.69)			8.81(6.31)
Overid. Stat(p-val)			0.11(0.74)			0.92(0.34)			2.74(0.10)			2.04(0.15)
Panel B: Occupations - Young cohort (25-34)												
$\Delta h_{(25-34)}$	0.64***	0.80***	1.50***	-0.42***	-0.17	0.18	-0.27***	-0.04	0.13	0.05	-0.60***	-1.82***
	(0.06)	(0.07)	(0.28)	(0.10)	(0.09)	(0.38)	(0.07)	(0.06)	(0.25)	(0.17)	(0.11)	(0.47)
R^2	0.36	0.52	0.38	0.27	0.61	0.60	0.57	0.76	0.76	0.43	0.79	0.72
Eff. F-test			11.11			11.11			11.11			11.11
CV 5%(10%)			9.30(6.60)			9.39(6.65)			11.26(7.73)			8.97(6.40)
Overid. Stat(p-val)			0.71(0.40)			1.61(0.20)			1.16(0.28)			1.16(0.28)
Panel C: Sectors - All cohorts												
Sector	Agriculture			Manufacture			High Skill Services			Low Skill Services		
$\Delta h_{(25-34)}$	0.21	-0.53***	-1.09**	-0.20*	-0.21*	-0.39	0.04	0.39***	0.64**	-0.05	0.34***	0.83**
	(0.16)	(0.10)	(0.36)	(0.08)	(0.08)	(0.29)	(0.08)	(0.06)	(0.20)	(0.11)	(0.09)	(0.28)
R^2	0.46	0.83	0.82	0.18	0.37	0.36	0.40	0.75	0.73	0.50	0.75	0.73
Eff. F-test			11.11			11.11			11.11			11.11
CV 5%(10%)			11.05(7.61)			9.93(6.96)			9.23(6.56)			10.38(7.22)
Overid. Stat(p-val)			2.33(0.13)			0.37(0.54)			2.04(0.15)			6.02(0.01)
Panel D: Sectors - Young cohort (25-34)												
$\Delta h_{(25-34)}$	0.19	-0.55***	-1.36***	-0.25*	-0.27**	-0.60	0.30***	0.60***	1.27***	-0.24	0.21	0.69
	(0.17)	(0.10)	(0.41)	(0.10)	(0.10)	(0.39)	(0.08)	(0.07)	(0.30)	(0.13)	(0.11)	(0.42)
R^2	0.40	0.81	0.78	0.16	0.36	0.34	0.34	0.61	0.52	0.47	0.73	0.71
Eff. F-test			11.11			11.11			11.11			11.11
CV 5%(10%)			9.97(6.98)			9.62(6.78)			9.01(6.43)			10.72(7.42)
Overid. Stat(p-val)			0.29(0.59)			0.09(0.76)			1.31(0.25)			1.05(0.31)
Obs.	486	486	486	486	486	486	486	486	486	486	486	486
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: I estimate specification (1) to assess the implications of rising college attainment on the process of structural transformation. In Panels A and C, I report results for all cohorts. In Panels B and D, I focus on the younger cohort. Panels A and B present the results for the reallocation of labor across four occupations. I group occupations into four categories: white-collar, blue-collar, general services, and elementary occupations. I group economic sectors into four categories: agriculture, manufacturing, high-skill, and low-skill service sectors. The unit of observation is the commuting zone, CZ. I consider three specifications. All specifications include state dummies. Columns (1) do not include controls. Columns (2) and (3) include controls to capture pre-reform development patterns. Columns (3) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013) and the TSLS critical values at the 5% and 10% levels. I also include statistics from the over-identification test.

Table B.5: Expansion in College Attainment and Structural Transformation

Panel A: Occupations - All cohorts												
Dep.Variable	Change in employment share in											
Occupation	White-Collar			Blue-Collar			General Services			Elementary		
Specification	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\Delta h_{(25-34)}$	0.21**	0.45***	0.62*	-0.30***	-0.05	0.38	-0.11	0.08	0.67	0.20	-0.47***	-1.67***
	(0.06)	(0.05)	(0.28)	(0.08)	(0.07)	(0.41)	(0.06)	(0.05)	(0.34)	(0.15)	(0.09)	(0.49)
R^2	0.36	0.52	0.46	0.27	0.61	0.54	0.57	0.76	0.73	0.43	0.79	0.66
Panel B: Occupations - Young cohort (25-34)												
$\Delta h_{(25-34)}$	0.64***	0.80***	1.26**	-0.42***	-0.17	0.63	-0.27***	-0.04	0.41	0.05	-0.60***	-2.30***
	(0.06)	(0.07)	(0.42)	(0.10)	(0.09)	(0.50)	(0.07)	(0.06)	(0.44)	(0.17)	(0.11)	(0.60)
R^2	0.36	0.52	0.38	0.27	0.61	0.60	0.57	0.76	0.76	0.43	0.79	0.72
Panel C: Sectors - All cohorts												
Sector	Agriculture			Manufacture			High Skill Services			Low Skill Services		
$\Delta h_{(25-34)}$	0.21	-0.53***	-1.59***	-0.20*	-0.21*	-0.21	0.04	0.39***	0.39	-0.05	0.34***	1.41***
	(0.16)	(0.10)	(0.39)	(0.08)	(0.08)	(0.36)	(0.08)	(0.06)	(0.24)	(0.11)	(0.09)	(0.43)
R^2	0.46	0.83	0.79	0.18	0.37	0.37	0.40	0.75	0.75	0.50	0.75	0.65
Panel D: Sectors - Young cohort (25-34)												
$\Delta h_{(25-34)}$	0.19	-0.55***	-1.55***	-0.25*	-0.27**	-0.48	0.30***	0.60***	0.98**	-0.24	0.21	1.05
	(0.17)	(0.10)	(0.42)	(0.10)	(0.10)	(0.51)	(0.08)	(0.07)	(0.38)	(0.13)	(0.11)	(0.65)
R^2	0.40	0.81	0.77	0.16	0.36	0.35	0.34	0.61	0.58	0.47	0.73	0.68
Eff. F-test			12.51			12.51			12.51			12.51
CV 5%(10%)			37.42(23.11)			37.42(23.11)			37.42(23.11)			37.42(23.11)
Obs.	486	486	486	486	486	486	486	486	486	486	486	486
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: I estimate specification (1) to assess the implications of rising college attainment on the process of structural transformation. These specifications use only Z_r^1 as the instrument. In Panels A and C, I report results for all cohorts. In Panels B and D, I focus on the younger cohort. Panels A and B present the results for the reallocation of labor across four occupations. I group occupations into four categories: white-collar, blue-collar, general services, and elementary occupations. I group economic sectors into four categories: agriculture, manufacturing, high-skill, and low-skill service sectors. The unit of observation is the commuting zone, CZ. I consider three specifications. All specifications include state dummies. Columns (1) do not include controls. Columns (2) and (3) include controls to capture pre-reform development patterns. Columns (3) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013) and the TSLS critical values at the 5% and 10% levels.

Table B.6: Expansion in College Attainment and Structural Transformation (cont')

Dep. Variable	Δ Employment share in								
	Self-employment			Informality			Formality		
Specification	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ College attainment	-0.16	-0.69***	-1.52***	-0.02	0.50***	1.37**	0.19	0.20*	0.15
	(0.13)	(0.10)	(0.43)	(0.14)	(0.10)	(0.47)	(0.10)	(0.10)	(0.45)
R^2	0.54	0.75	0.72	0.59	0.82	0.79	0.28	0.37	0.37
Eff. F-test			11.13			11.13			11.13
CV 5%(10%)			9.86(6.93)			11.25(7.73)			10.16(7.10)
Overid. Stat(p-val)			0.37(0.54)			1.96(0.16)			0.86(0.35)
Obs.	486	486	486	486	486	486	486	486	486
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Notes: I estimate specification (1) to assess the implications of rising college attainment on the allocation of labor between self-employment, informal employment, and formal employment. Columns (1) to (3) report results on the change in the share of employment that is self-employed. Columns (4) to (6) report results on the change in the share of informal employment. Columns (7) to (9) report results on the change in the share of formal employment. All specifications include state dummies. Columns (1), (4) and (7) do not include controls. Columns (2), (5) and (8) include controls to capture pre-reform development patterns. Columns (3), (6), and (9) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013) and the TSLS critical values at the 5% and 10% levels. I also include statistics from the over-identification test.

Table B.7: Expansion in College Attainment and Income Growth

Dep.Variable	Income growth, unskilled		Relative income growth		Employment growth, unskilled	
Specification	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta h_{(25-34)}$	0.74***	1.17	0.69***	0.46	-0.30	2.06
	(0.20)	(0.74)	(0.08)	(0.40)	(0.33)	(1.20)
Obs.	486	486	486	486	486	486
R^2	0.85	0.85	0.62	0.61	0.54	0.47
F-test		11.11		11.11		11.11
CV 5%(10%)		8.80 (6.31)		10.40 (7.23)		8.80 (6.31)
Overid. Stat(p-val)		0.70 (0.40)		3.25 (0.04)		0.94 (0.33)

Notes: I estimate specification (1) to assess the implications of rising college attainment on average income growth for those with no college education, labeled as unskilled, in Columns (1) and (2). Columns (3) and (4) assess the implications for the difference between the growth in income for the average worker and the growth in income for the average unskilled workers. Columns (5) and (6) assess the growth in employment for unskilled workers. The unit of observation is the commuting zone, CZ. All specifications include state dummies and controls aimed to capture pre-reform development patterns. Columns (2) and (4) introduce the IV estimates. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013), the TSLS critical values at the 5% and 10% levels, and statistics from the over-identification test.

B.6 Diff-in-Diff Analysis

Table B.8: Diff-in-Diff Analysis, cont'

Dep. Variable	ΔEmployment share in		
	High-Skill Serv.	Education/Health/Public Serv.	Banking Serv.
Specification	(1)	(2)	(3)
Z_r^1	0.82	0.52	0.30
	(1.24)	(1.25)	(0.52)
$Z_r^1 * 1_{2010}$	-4.51**	-3.11*	-1.40
	(1.50)	(1.42)	(0.80)
Z_r^2	0.15***	0.12***	0.03
	(0.04)	(0.04)	(0.01)
$Z_r^2 * 1_{2010}$	0.05***	0.05***	0.00
	(0.01)	(0.01)	(0.00)
Obs.	958	958	958
R^2	0.51	0.50	0.39

Notes: This table presents the results from the estimation of specification (3). Columns (1)-(3) focus on the implications on the employment share in high-skilled services and its subsectors categorized as education/health/public services and other high-skilled services. Z_r^1 represents the ratio of higher education capacity to the youth population 20 years before the observed decadal change in outcomes. Z_r^2 captures the share of individuals under 15 20 years before the observed decadal change in outcomes. 1_{2010} is a dummy variable indicating the 2000–2010 decade. All specifications include controls consistent with the baseline analysis. The complete list of controls, detailed in Table B.1 in Appendix B.2, excludes income-related variables from 1980 due to data limitations.

B.7 Firm Level Analysis

B.7.1 World Bank survey of firms

Table B.9 presents data from the World Bank Enterprise Surveys (WBES) for Brazil, based on interviews conducted with business owners and top managers from 1,802 firms between May 2008 and June 2009. Column (1) highlights the share of firms that consider the qualifications of the workforce as the biggest constraint to its growth. Column (4) highlights the share of firms that consider the qualifications of the workforce as a major or severe constraint to its growth.

B.7.2 Implications for the age distribution of firms

Does the increase in the number of large firms stem from new entrants or the expansion of existing firms? To address this question, I analyze the growth in the number of firms within different age groups, classifying firms based on their age in 2010, measured from the date of their first registered labor contract. Table B.7.2 presents the results, with Columns (1) reporting OLS estimates and Columns (2) showing IV estimates, all specifications including controls.

The results indicate that rising college attainment significantly increases the number of large firms that are 6–10 years old and those older than 10 years. The former group comprises

Table B.9: World Bank Survey

Variable	Low Education of Workforce					
	Biggest Constraint			Major/Severe Constraint		
Region	Brazil	LAC	World	Brazil	LAC	World
	(1)	(2)	(3)	(4)	(5)	(6)
All	12.6%	7.8%	9.8%	74.9%	28.6%	20.8%
Small Firms 6 to 19 Workers	13.4%	6.7%	8.6%	71%	28.1%	19.3%
Medium Firms 20 to 99 Workers	12.4%	9.8%	11.5%	86.3%	27.6%	22.6%
Large Firms 100+ Workers	4.6%	11.9%	14.2%	69.9%	36.1%	25%

Notes: The table presents survey results from a representative sample of 1,802 firms in Brazil, assessing the main constraints to firm growth, with a focus on workforce education. Columns (1)–(3) report the share of firms identifying low workforce education as their biggest constraint, while Columns (4)–(6) show the share of firms considering it one of their major constraints. Columns (1) and (4) display results for Brazil, Columns (2) and (5) provide a benchmark for Latin America and the Caribbean, and Columns (3) and (6) present the corresponding results for the World.

Table B.10: Growth number of firms, by age

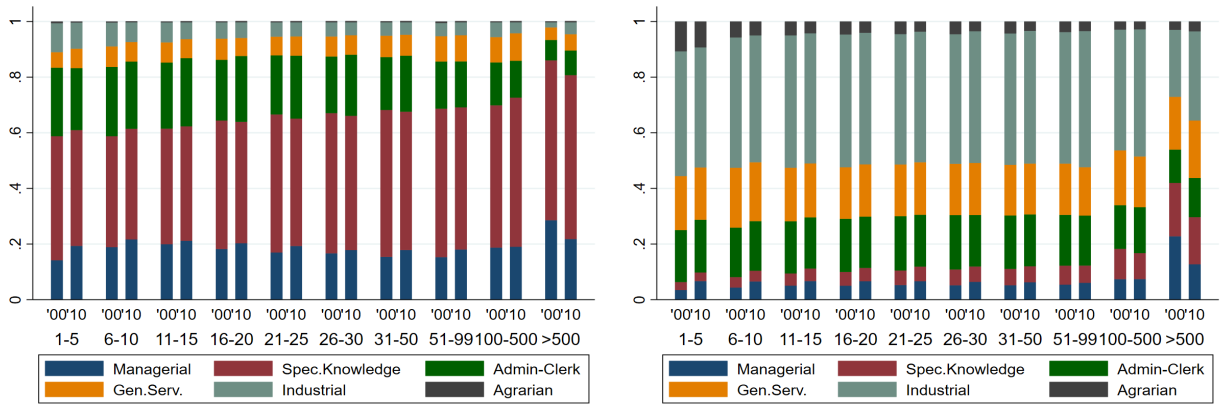
Dep.Variable	Growth # of firms							
Firms' age	≤2yo		3yo-5yo		6yo-10yo		> 10yo	
Specification	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$\Delta h_{(25-34)}$	2.81***	0.64	1.83	4.61	3.21**	9.87*	1.79**	7.94*
	(0.80)	(3.78)	(0.95)	(4.31)	(1.15)	(4.34)	(0.69)	(3.32)
R^2	0.44	0.43	0.36	0.34	0.41	0.33	0.57	0.48
F-test		10.47		11.31		10.75		10.70
CV 5%(10%)		9.30(6.60)		9.08(6.47)		13.75(9.15)		9.65(6.80)
Overid. Stat		0.25(0.62)		2.01(0.16)		4.01(0.05)		2.49(0.11)
Obs.	482	482	481	481	483	483	486	486

Notes: I estimate specification (1) to assess the implications of college attainment for the growth in the number of firms classified in four age bins. The unit of observation is the commuting zone, CZ. I consider two specifications for each age bin. Columns (1) report the OLS results, while columns (2) introduce the IV estimates. All specifications include state dummies and controls. The complete list of controls is in Table B.1 in Appendix B.2. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013) and the TSLS critical values at the 5% and 10% levels. I also include statistics from the over-identification test.

firms established before 2005, while the latter includes those already operating before 2000. These findings suggest that expanding college attainment primarily contributed to the growth of existing firms rather than fostering new firm entry. Notably, there is no significant increase in the number of large firms aged 1–2 or 3–5 years, indicating that higher college attainment did not lead to a sustained rise in the creation of young large firms.

B.7.3 Role of skilled workers in the firm

Figure B.3: Occupation Differences Within Formal Firms

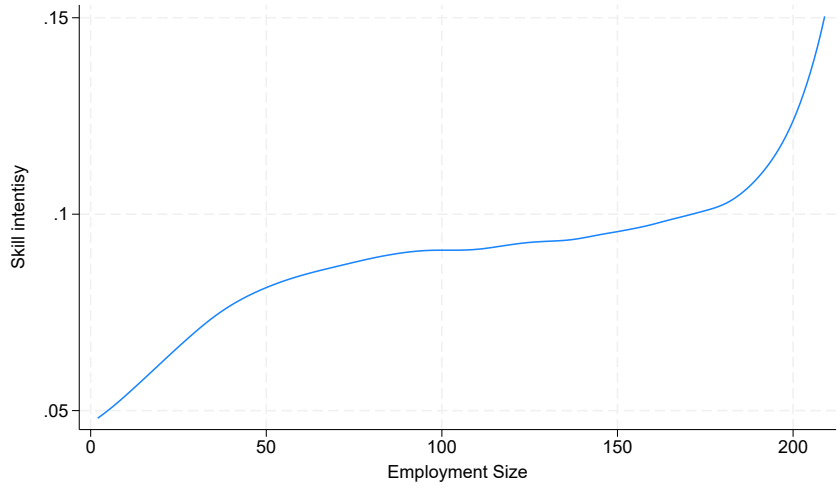


(a) 25-34yo College Workers

(b) No-College Workers

Notes: Panel (a) depicts the occupation distribution of college-educated workers aged 25–34 by firm size for 2000 ('00) and 2010 ('10). Panel (b) reports the occupation distribution of all workers without a college degree. The y-axis measures the share of workers employed in a specific occupations. The occupations are reported by the employer. Occupations are categorized into managerial roles, professional roles (e.g., lawyers, accountants, physicians, nurses, professors), administrative and clerical positions, general services, industrial occupations, and agricultural work. Firms are grouped into size bins, including 1–5 workers, 6–10 workers, 11–15 workers, and larger categories as indicated in the figure.

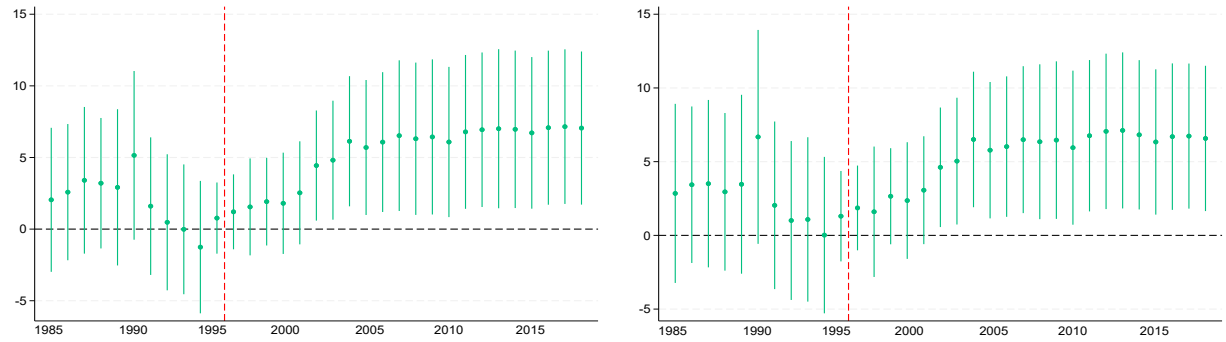
Figure B.4: Skill intensity by employment level



Notes: The y-axis represents average skill intensity by firm size, calculated using the matched employer-employee data from 2010. Skill intensity is defined as the share of a firm's workforce holding a college degree. The x-axis depicts firm size, measured by the number of employees. For firms with more than 200 employees, size intervals are grouped in increments of 100 (e.g., 201–300 employees, 301–400 employees). Firms with over 1,000 employees are grouped into a single category.

B.7.4 Implications for firms using annual frequency

Figure B.5: Results on Formal Firms



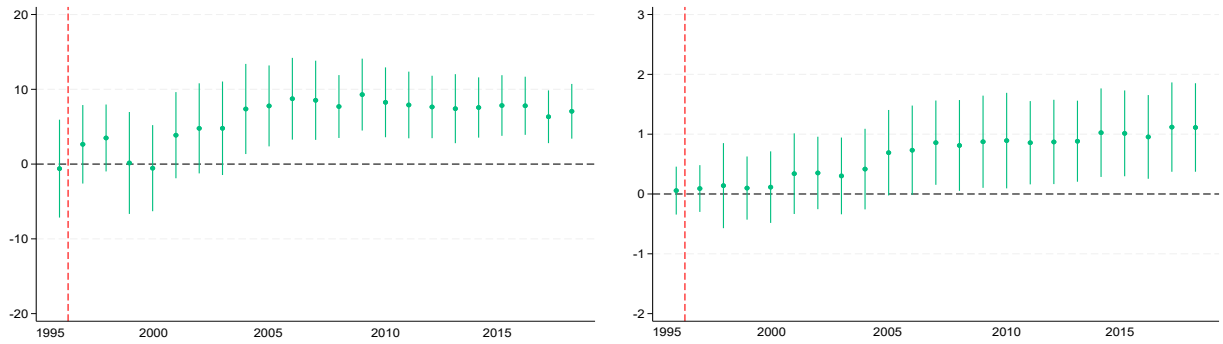
(a) Growth # of large firms

(b) Growth # of small firms

Notes: The Figure plots the IV coefficient estimates $\{\hat{\beta}_t\}$ from specification (4) with 95% CI. The y-axis measures the coefficient values and CI. The x-axis measures time in years. Panel (a) shows the results when I assess the implications of rising college attainment for the growth in the total number of large firms. Panel (b) shows the results when I assess the growth in the number of small firms as the outcome. In both cases, I measure the change with respect to its level in 1996, the year of the reform. The red vertical dotted line identifies the year of the reform, 1996.

B.7.5 Implications for firms using annual frequency, excluding higher education sector

Figure B.6: Results on Formal Firms



(a) Growth # large firms w/ skill

(b) Large firms' employment share

Notes: The Figure plots the IV coefficient estimates $\{\hat{\beta}_t\}$ from specification (4) with 95% confidence intervals (CI). The y-axis measures the coefficient values and CI. The x-axis measures time in years. Outcomes exclude firms in the higher education sector (i.e. colleges). Panel (a) depicts the implications from rising college attainment for the growth in the number of large firms with skill, defined as those with at least six workers and hiring at least one worker with a college degree. Panel (b) shows the implications for the change in the employment share in large firms. In both cases, I measure the change with respect to its level in 1996, the year of the reform. The red vertical dotted line identifies the year of the reform, 1996. The plot begins in 1995, when firms can be consistently classified into more disaggregated industries. Prior to 1995, the classification of firms by sector is subject to substantial measurement error.

B.8 Assessing the association of the instrument with migration

I present the results from Section 3.3.5, which examine the relationship between variation in the rise in college attainment—generated by the instruments—and migration flows. Several migration measures are constructed for this analysis. For instance, immigration flow of college graduates is defined as the ratio of young college graduates residing in r in 2010 who lived in a different CZ in 2005 to the initial population of young college graduates in r in 2005. Similarly, the emigration flow of college graduates is defined as the share of college graduates who resided in r in 2005 but moved to a different CZ by 2010. Net migration of college graduates is calculated as the difference between the number of young college graduates moving into r and those leaving r, divided by the number of college graduates residing in r in 2005. Finally, within-CZ immigration of college graduates is measured as the share of college graduates who moved between municipalities within the same CZ from 2005 to 2010. Table B.11 reports the main results, showing the association between rising college attainment and migration patterns after controlling for observable proxies of a region's growth prospects.

Table B.11: Association Between Instruments and Migration

Dep.Variable	Panel A: Immigration to the CZ				Panel B: Emigration from the CZ			
Sample	All cohorts		College graduates (25-34)		All cohorts		College graduates (25-34)	
Specification	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$\Delta h_{(25-34)}$	0.04	0.60*	0.15	1.26**	0.15***	0.63***	0.16	0.95*
	(0.07)	(0.30)	(0.09)	(0.43)	(0.04)	(0.19)	(0.09)	(0.40)
R^2	0.71	0.66	0.68	0.57	0.66	0.57	0.36	0.24
F-test		11.11		11.11		11.11		11.11
CV 5%(10%)		9.91 (6.95)		11.92 (8.11)		10.87 (7.50)		12.10 (8.21)
Overid. Stat(p-val)		0.01 (0.92)		0.26 (0.61)		0.72 (0.40)		1.23 (0.27)
Dep.Variable	Panel C: Netmigration to the CZ				Panel D: Immigration within the CZ			
Sample	All cohorts		College graduates (25-34)		All cohorts		College graduates (25-34)	
Specification	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$\Delta h_{(25-34)}$	-0.11	-0.01	0.01	0.46	0.59**	2.42**	0.48*	2.48**
	(0.10)	(0.37)	(0.15)	(0.61)	(0.20)	(0.75)	(0.20)	(0.77)
R^2	0.71	0.70	0.49	0.47	0.76	0.72	0.70	0.63
F-test		11.11		11.11		11.11		11.11
CV 5%(10%)		10.89 (7.52)		10.23 (7.14)		9.24 (6.56)		9.00 (6.42)
Overid. Stat(p-val)		0.08 (0.78)		0.97 (0.32)		0.01 (0.94)		0.52 (0.47)
Obs.	486	486	486	486	486	486	486	486
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: I test the association between the rise in college attainment and migration flows. I use four outcomes: immigration flows to the CZ (Panel A), emigration flows from the CZ (Panel B), net-migration flows (Panel C), and immigration within the CZ (Panel D). For each outcome, I measure the importance of general migration flows (people between 25-60) and for the cohort of new college graduates (youth with a college degree). Column (1) shows the OLS results. Column (2) introduces the IV results. In both specifications, controls and state-fixed effects are included. The complete list of controls is in Table B.1 in Appendix B.2.

B.9 Robustness

In this section, I evaluate the robustness of the main results by incorporating controls that capture significant economic forces during the analysis period. Specifically, I assess the sensitivity of these findings to three major factors shaping Brazil's economic growth: (i) the trade liberalization process of the 1990s (Dix-Carneiro and Kovak 2017; Dix-Carneiro, Goldberg, et al. 2021), and (ii) labor market reforms that increased the minimum wage (Engbom and Moser 2022; Haanwinckel 2023).[57]

These forces could potentially bias the estimated effects of rising college attainment on

[57] Technological advancements in agriculture, particularly the adoption of genetically engineered (GE) soybeans, played a parallel role in structural transformation (Bustos, Caprettini, and Ponticelli 2016; Bustos, Garber, and Ponticelli 2020). Bustos, Caprettini, and Ponticelli (2016) finds that regions more suited for GE soybean adoption experienced faster labor reallocation away from agriculture. In contrast, another agricultural innovation that expanded arable land increased labor demand in agriculture, slowing this reallocation. If these technological changes are correlated with my measure of pre-reform relative college scarcity, they could introduce bias in the IV estimates. Since these studies define labor markets differently, there is no straightforward mapping between their definitions and the one used in this paper. Replicating their approach using my labor market definition would allow for a robustness check on the role of agricultural technological change, which I leave for future research.

the main outcomes. For instance, Dix-Carneiro and Kovak (2017) finds that regions more exposed to tariff reductions experienced slower wage growth and deteriorating employment conditions. If exposure to trade shocks correlates with the relative abundance of higher education capacity, the control labor markets may have been disproportionately affected, compromising the validity of my results. Second, Engbom and Moser (2022) shows that rising minimum wages reduced wage inequality. If labor markets with fewer colleges relative to the potential demand were also more exposed to minimum wage increases, my findings on the reduction in college returns and the transition away from agriculture could be affected.

Table B.12 tests whether the baseline results change significantly when I control for these economic forces. I focus on six baseline outcomes: income growth, the change in the college premium, the growth in the share of firms that are large and employ skilled workers, the change in the employment share of large firms, and shifts in employment toward white-collar occupations and high-skill services. For each outcome, I estimate four specifications using the IV model. Column (1) reports the baseline results. Column (2) includes Dix-Carneiro and Kovak (2017)'s measure of regional exposure to trade liberalization. Column (3) adds the share of formal employment as a proxy for exposure to rising minimum wages. Column (4) combines the controls from columns (2) and (3). Reassuringly, the results remain robust, as evidenced by the similarity between the baseline coefficients and those in column (4), which includes all proxies for the economic forces analyzed.

Whether these forces *shape* the implications of rising college attainment—or are themselves influenced by it—remains an important avenue for future research.

Table B.12: Robustness analysis

Dep.Variable	Income Growth				Δ College Premium			
Specification	Baseline	Trade lib.	Min. Wage	All	Baseline	Trade lib.	Min. Wage	All
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\Delta h_{(25-34)}$	1.64*	1.87*	1.67*	1.89*	-12.53	-12.38	-11.47	-11.23
	(0.73)	(0.78)	(0.80)	(0.83)	(6.42)	(6.61)	(6.73)	(6.77)
R^2	0.88	0.88	0.88	0.88	0.40	0.40	0.41	0.42
F-test	11.13	10.16	10.24	9.85	11.13	10.16	10.24	9.85
CV 5%(10%)	10.08(7.06)	8.79(6.30)	10.76(7.45)	9.81(6.90)	10.93(7.54)	10.83(7.47)	11.54(7.90)	11.45(7.84)
Overid. Stat(p-val)	3.47(0.06)	2.32(0.13)	6.74(0.01)	4.74(0.03)	1.25(0.26)	1.80(0.18)	1.03(0.31)	1.46(0.23)
Dep.Variable	Δ Sh. Large Firms w/ Skill				Δ Emp. Share Large Firms			
Specification	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\Delta h_{(25-34)}$	0.77*	0.72*	0.75*	0.71*	1.07*	0.98*	0.90*	0.82*
	(0.30)	(0.30)	(0.30)	(0.30)	(0.42)	(0.41)	(0.41)	(0.40)
R^2	0.53	0.53	0.53	0.53	0.18	0.22	0.26	0.30
F-test	11.13	10.16	10.24	9.85	11.13	10.16	10.24	9.85
CV 5%(10%)	9.56(6.76)	8.20(5.96)	10.33(7.21)	9.32(6.61)	10.12(7.08)	8.58(6.18)	10.57(7.35)	9.62(6.79)
Overid. Stat(p-val)	1.19(0.27)	1.15(0.28)	1.45(0.23)	1.34(0.25)	4.58(0.03)	4.57(0.03)	6.11(0.01)	5.87(0.02)
Dep.Variable	Δ Emp. Share in White-Collar Occ.				Δ Emp. Share in High-Skill Services			
Specification	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
$\Delta h_{(25-34)}$	0.70***	0.74***	0.60**	0.63**	0.66**	0.62**	0.60**	0.58**
	(0.20)	(0.21)	(0.21)	(0.22)	(0.21)	(0.20)	(0.20)	(0.19)
R^2	0.59	0.59	0.61	0.62	0.73	0.73	0.74	0.74
F-test	11.13	10.16	10.24	9.85	11.13	10.16	10.24	9.85
CV 5%(10%)	9.54(6.75)	9.99(6.99)	10.29(7.19)	9.79(6.89)	9.93(6.97)	9.41(6.65)	10.84(7.50)	10.36(7.21)
Overid. Stat(p-val)	0.18(0.67)	1.36(0.24)	0.01(0.93)	0.50(0.48)	2.55(0.11)	2.01(0.16)	1.11(0.29)	0.78(0.38)

Notes: I estimate specification (1) to examine the effects of rising college attainment on key outcomes analyzed in Section 3.3. The analysis focuses on six baseline outcomes: income growth, the change in the college premium, the change in the share of firms that are large and employ skilled workers, the change in the employment share of large firms, and shifts in employment toward white-collar occupations and high-skill services. Column (1) replicates the IV estimates from Section 3.3, excluding pre-reform formality and the trend in employment formality between 1991 and 2000. Column (2) adds Dix-Carneiro and Kovak (2017)'s measure of regional exposure to trade liberalization. Column (3) reintroduces measures of employment formality, aligning with the main results in Section 3.3. Column (4) combines the controls from Columns (2) and (3). All specifications include controls similar to those in the baseline analysis. The complete list of controls is detailed in Table B.1. The IV estimates include the efficient F-test from Montiel Olea and Pflueger (2013), the TSLS critical values at the 5% and 10% levels, and statistics from the over-identification test.

C Theory

C.1 Proof of Proposition 1 (i) and (ii)

Firms are heterogeneous in productivity z and use unskilled and skilled labor to produce. Firms maximize profits by optimally choosing their labor inputs:

$$\max_{l_{mir},h_{ir}\geq 0} \quad p_r^{na} y_{ir} - w_r^l \left(\bar{\omega}_r h_{ir} + l_{mir} + f_r^o\right)$$
$$\text{where,} \quad y_{ir} = z_{ir}^{1-\alpha-\beta} (l_{mir})^{\alpha} \left(\Gamma_r h_{ir} + \kappa\right)^{\beta}$$

When the non-homothetic term κ is greater than 0, some firms might choose not to hire skilled workers (i.e. $h_{ir} = 0$). Consequently, the optimal solution of the firm involves finding two thresholds, z_r^* and $\bar{z}_r$. The first threshold determines the lowest productivity level at which production becomes profitable, $(\pi(z_r^*) = 0)$. The second threshold identifies the productivity level at which it becomes optimal to hire skilled workers $(h(\bar{z}_r) = 0)$. Formally,

$$z_r^* = \left[\frac{w_r^l}{P_r^{na}} \frac{1}{\alpha\kappa^{\beta}} \left(\frac{\alpha f_r^o}{1-\alpha}\right)^{1-\alpha}\right]^{\frac{1}{1-\alpha-\beta}} \qquad \bar{z}_r = \left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{1-\alpha}{1-\alpha-\beta}} z_r^* \tag{T1}$$

Here, $\zeta_r := \frac{\kappa\bar{\omega}_r}{\Gamma_r f_r^o}$. When $\zeta_r > \frac{\beta}{1-\alpha}$ then $\bar{z}_r > z_r^*$, and only high productivity firms hire skilled workers. The optimal demand for inputs, final output, and profits are defined as follows:

$$l_{mir}(z_{ir}) = \begin{cases} \frac{\alpha f_r^o}{\beta}\zeta_r \frac{z_{ir}}{\bar{z}_r} & \text{for } z_{ir} \geq \bar{z}_r \\ \frac{\alpha f_r^o}{\beta}\zeta_r \left(\frac{z_{ir}}{\bar{z}_r}\right)^{\frac{1-\alpha-\beta}{1-\alpha}} & \text{for } z_{ir} \in [z_r^*, \bar{z}_r) \end{cases}$$

$$h(z_{ir}) = \begin{cases} f_r^o \zeta_r \left(\frac{z_{ir}-\bar{z}_r}{\bar{z}_r}\right) & \text{for } z_{ir} \geq \bar{z}_r \\ 0 & \text{for } z_{ir} \in [z_r^*, \bar{z}_r) \end{cases} \tag{D1}$$

$$y(z_{ir}) = \begin{cases} \frac{w_r^l}{p_r^{na}} \frac{f_r^o}{\beta}\zeta_r \frac{z_{ir}}{\bar{z}_r} & \text{for } z_{ir} \geq \bar{z}_r \\ \frac{w_r^l}{p_r^{na}} \frac{f_r^o}{\beta}\zeta_r \left(\frac{z_{ir}}{\bar{z}_r}\right)^{\frac{1-\alpha-\beta}{1-\alpha}} & \text{for } z_{ir} \in [z_r^*, \bar{z}_r) \end{cases}$$

$$\pi(z_{ir}) = \begin{cases} w_r^l f_r^o \zeta_r \left(\frac{1-\alpha-\beta}{\beta}\frac{z_{ir}}{\bar{z}_r} + 1 - \frac{1}{\zeta_r}\right) & \text{for } z_{ir} \geq \bar{z}_r \\ w_r^l f_r^o \left(\frac{1-\alpha}{\beta}\zeta_r \left(\frac{z_{ir}}{\bar{z}_r}\right)^{\frac{1-\alpha-\beta}{1-\alpha}} - 1\right) & \text{for } z_{ir} \in [z_r^*, \bar{z}_r) \end{cases}$$

I define skill intensity as the ratio of the optimal hiring of skilled to unskilled labor:

$$\frac{h(z_{ir})}{l_{mir}(z_{ir})} = \begin{cases} 0 & \text{for } z_{ir} \in [z^*, \bar{z}_r) \\ \frac{\beta}{\alpha\bar{\omega}_r}\left(1 - \frac{\bar{z}_r}{z_{ir}}\right) & \text{for } z_{ir} \geq \bar{z}_r \end{cases}$$

All active firms hire college workers when $\zeta_r \leq \frac{\beta}{1-\alpha}$. In this case, there is only one threshold z_r^*. The optimal demand for inputs is determined by (D1) first case where $z_i \geq \bar{z}_r$ but replacing z_r^* instead of $\bar{z}_r$.

Note that the thresholds derived in (T1), depend on $\bar{\omega}_r$ through $\frac{w_r^l}{p^n a_r}$ and ζ_r. Next, I analyze the aggregate behavior of firms in region r, assuming a free-entry condition to derive

a closed-form solution for $\frac{w_r^l}{p_r^{na}}$ as a function of the returns to skill $\bar{\omega}_r$. This exercise allows me to assess how a reduction in returns to skill, $\bar{\omega}_r$, resulting from an expansion in the supply of skilled workers, affects the thresholds z_r^* and $\bar{z}_r$. Appendix C.2 concludes the proof of Proposition 1 and provides the proof of Proposition 2.

C.2 Proof of Proposition 1 (iii) and Proposition 2

Production stage in the NA sector begins once firms decide to enter the market. Upon paying the entry cost, firms draw their productivity from a Pareto distribution with probability distribution $G_r(z_i) = 1 - \left(\frac{B_r}{z_{ir}}\right)^{\lambda}$ and initiate production only if the productivity draw exceeds a certain threshold. Consequently, firms decide whether to enter the NA sector based on the expected benefits net of entry costs. Specifically, a firm enters when the following condition is met:

$$E[\pi(z_i)] \geq f_r^e w_r^l$$

The free entry condition ensures that the equation holds with equality. In Appendix C.1, I derive profits as a function of productivity z_{ir}. Here, I compute the expected value of profits using the Pareto density function. When $\zeta_r \geq \frac{\beta}{1-\alpha}$, expected profits take the form:

$$E[\pi(z)] = w_r^l f_r^o \zeta_r \left(\int_{z_r^*}^{\bar{z}_r} \left(\frac{1-\alpha}{\beta} \zeta_r \left(\frac{z}{\bar{z}_r} \right)^{\frac{1-\alpha-\beta}{1-\alpha}} - \frac{1}{\zeta_r} \right) dG(z) + \int_{\bar{z}_r} \left(\frac{1-\alpha-\beta}{\beta} \frac{z}{\bar{z}_r} + 1 - \frac{1}{\zeta_r} \right) dG(z) \right)$$

When $\zeta_r < \frac{\beta}{1-\alpha}$, expected profits take the form:

$$E[\pi(z)] = w_r^l f_r^o \int_{z_r^*} \left(\frac{1-\alpha-\beta}{\beta} \frac{z}{\bar{z}_r} + 1 - \frac{1}{\zeta_r} \right) dG(z)$$

Substituting the expected profits into the free entry condition yields the following relationships:

$$(z_r^*)^{\lambda} = \begin{cases} \frac{f_r^o(1-\alpha-\beta)B^{\lambda}}{(\beta+(1-\alpha)(\lambda-1))f_r^e} \left[\frac{\frac{\beta}{(\lambda-1)(1-\alpha)} + \left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}}{\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}} \right] & \zeta_r \geq \frac{\beta}{1-\alpha} \\ \frac{f_r^o(1-\zeta_r)}{(\lambda-1)f_r^e} B^{\lambda} & \zeta_r < \frac{\beta}{1-\alpha} \end{cases}$$

$$\bar{z}_r^{\lambda} = \frac{f_r^o(1-\alpha-\beta)B^{\lambda}}{(\beta+(1-\alpha)(\lambda-1))f_r^e} \frac{1-\alpha}{\beta} \zeta_r \left[\frac{\beta}{(\lambda-1)(1-\alpha)} + \left(\frac{1-\alpha}{\beta}\zeta_r \right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}} \right]$$

$$\frac{w_r^l}{p_r^{na}} = \begin{cases} \alpha\kappa^{\beta} \left(\frac{1-\alpha}{\alpha f_r^o}\right)^{1-\alpha} (z^{*\lambda})^{\frac{1-\alpha-\beta}{\lambda}} & \zeta_r \geq \frac{\beta}{1-\alpha} \\ \alpha^{\alpha} \left(\frac{B_r(1-\alpha-\beta)}{[(\lambda-1)f_r^e]^{\frac{1}{\lambda}}} \right)^{1-\alpha-\beta} \left(\frac{1}{f_r^o(1-\zeta_r)} \right)^{\frac{(\lambda-1)(1-\alpha-\beta)}{\lambda}} \left(\frac{\beta\kappa}{f_r^o\zeta_r} \right)^{\beta} & \zeta_r < \frac{\beta}{1-\alpha} \end{cases}$$

These expressions establish the relationship between the thresholds and ζ_r:

$$\text{A reduction in } \zeta_r \rightarrow \begin{cases} \text{increases } z_r^* \\ \text{decreases } \bar{z}_r \quad ; \text{ when } \zeta_r \geq \frac{\beta}{1-\alpha} \\ \text{increases } \frac{w_r^l}{p_r^{na}} \quad ; \text{ when } \zeta_r \geq \frac{\beta}{1-\alpha} \\ \text{ambiguous } \frac{w_r^l}{p_r^{na}} \quad ; \text{ when } \zeta_r < \frac{\beta}{1-\alpha} \end{cases}$$

Since $\zeta_r = \frac{\kappa \bar{\omega}_r}{f_r^o \Gamma_r}$, a reduction in $\bar{\omega}_r$ leads to a decrease in ζ_r, which completes the proof of Proposition 1.

Firms employ unskilled workers to cover entry costs $w_r^l f_r^e$ and overhead costs $w_r^l f_r^o$. Specifically, each firm hires f_r^e unskilled workers to enter the market and f_r^o for overhead activities. The aggregate labor demand for these activities is determined by multiplying the number of firms M by the number of workers required per firm (f_r^e, f_r^o). Substituting the real wage expressions and integrating across firms gives total demand for unskilled workers in each role. When $\zeta_r \geq \frac{\beta}{1-\alpha}$, the aggregate labor demand in the NA sector is:

$$L_{mr} = \alpha \mathbf{L_{na,r}}$$

$$\bar{\omega}_r H_r = \frac{\beta + (\lambda-1)(1-\alpha)}{\lambda} \frac{\frac{\beta}{(\lambda-1)(1-\alpha)}}{\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(\lambda-1)(1-\alpha)+\beta}{1-\alpha-\beta}} + \frac{\beta}{(\lambda-1)(1-\alpha)}} \mathbf{L_{na,r}}$$

$$L_{or} = \frac{\beta + (\lambda-1)(1-\alpha)}{\lambda} \frac{\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(\lambda-1)(1-\alpha)+\beta}{1-\alpha-\beta}}}{\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(\lambda-1)(1-\alpha)+\beta}{1-\alpha-\beta}} + \frac{\beta}{(\lambda-1)(1-\alpha)}} \mathbf{L_{na,r}}$$

$$L_{er} = \frac{1-\alpha-\beta}{\lambda} \mathbf{L_{na,r}}$$

Here, $\mathbf{L_{na,r}} = L_{mr} + L_{or} + L_{er} + \bar{\omega} H_r$ represents the total labor input in the NA sector. The total number of firms in the economy is given by:

$$M = \frac{1-\alpha-\beta}{\lambda f_r^e} \mathbf{L_{na,r}}$$

I derive aggregate production in the NA sector as follows:

$$Y_r^{na} = \frac{w_r^l}{p_r^{na}} \frac{f_r^o}{\beta} \zeta_r \left(\int_{z_r^*}^{\bar{z}_r} \left(\frac{z}{\bar{z}_r}\right)^{\frac{1-\alpha-\beta}{1-\alpha}} dG(z) + \int_{\bar{z}_r} \frac{z}{\bar{z}_r} dG(z) \right) M_r$$

This results in the following expression:

$$Y_r^{na} = \underbrace{\frac{w_r^l}{p_r^{na}}}_{A_r^{na}} \mathbf{L_{na,r}}$$

When $\zeta_r < \frac{\beta}{1-\alpha}$, the aggregate labor demand in the NA sector is:

$$
\begin{aligned}
L_{mr} &= \alpha \mathbf{L_{na,r}} \\
\bar{\omega}_r H &= \left(\beta - \frac{(\lambda-1)(1-\alpha-\beta)}{\lambda}\frac{\zeta_r}{1-\zeta_r}\right)\mathbf{L_{na,r}} \\
L_{or} &= \frac{(\lambda-1)(1-\alpha-\beta)}{\lambda}\frac{1}{1-\zeta_r}\mathbf{L_{na,r}} \\
L_{er} &= \frac{1-\alpha-\beta}{\lambda}\mathbf{L_{na,r}}
\end{aligned}
$$

Similarly, I find that $\mathbf{L_{na,r}}$ encompasses all labor inputs in the NA sector, and the total number of firms is:

$$
M_r = \frac{1-\alpha-\beta}{\lambda f_r^e}\mathbf{L_{na,r}}
$$

Aggregate production in the NA sector is derived as follows:

$$
Y_r^{na} = \frac{w_r^l}{p_r^{na}}\frac{f_r^o}{\beta}\zeta_r\left(\int_{z_r^*}\frac{z}{\bar{z}_r}dG(z)\right)M_r
$$

As before, it yields the following expression:

$$
Y_r^{na} = \underbrace{\frac{w_r^l}{p_r^{na}}}_{A_r^{na}}\mathbf{L_{na,r}}
$$

This completes the proof of Proposition 2.

C.3 Formal derivation of Assumption B

The assumption ensures that some firms remain inactive in equilibrium in every region r.

$$
\frac{f_r^o}{f_r^e} > \frac{\beta + (1-\alpha)(\lambda-1)}{1-\alpha-\beta} \tag{B}
$$

This condition is necessary to avoid cases that do not contribute meaningful insights to the analysis. The share of firms operating in the economy is given by:

$$
\left(\frac{B_r}{z_r^*}\right)^{\lambda} = \begin{cases} \frac{(\beta+(1-\alpha)(\lambda-1))f_r^e}{f_r^o(1-\alpha-\beta)}\frac{\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}}{\left[\frac{\beta}{(\lambda-1)(1-\alpha)}+\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}\right]} & \zeta_r \geq \frac{\beta}{1-\alpha} \\ \frac{(\lambda-1)f_r^e}{f_r^o(1-\zeta_r)} & \zeta_r < \frac{\beta}{1-\alpha} \end{cases}
$$

In the limiting case where $\zeta_r \to 0$, the following condition ensures the presence of inactive firms in equilibrium:

$$\frac{f_r^o}{f_r^e} > \lambda - 1 \tag{B1}$$

Similarly, when $\zeta_r \to \infty$, the condition ensuring some firms remain inactive in equilibrium is:

$$\frac{f_r^o}{f_r^e} > \frac{\beta + (1-\alpha)(\lambda - 1)}{1 - \alpha - \beta} \tag{B2}$$

Importantly, if condition B2 holds, it also satisfies condition B1, effectively nesting condition B1 within it.

C.4 Preferences

Following Boppart (2014), I incorporate income effects by assuming that individuals' preferences belong to the PIGL class. These non-homothetic preferences are particularly well-suited to my analysis for three key reasons. First, they facilitate the aggregation of individual demand. Second, they align individual and regional demand parameters, enabling the estimation of preference parameters directly from household expenditure data. Third, as demonstrated by Fan, Peters, and Zilibotti (2023), these preferences establish a natural mapping between preferences for final goods and value-added under certain assumptions. This mapping is especially important because the production function in the model is expressed in value-added terms, while the estimation of preference parameters relies on expenditure data on final goods. The indirect utility function is specified as:

$$V(e^i, P_r^a, P_r^{na}) = \frac{1}{\varepsilon}\left[\frac{e^i}{(P_r^a)^\gamma (P_r^{na})^{1-\gamma}}\right]^\varepsilon - v \ln\left[\frac{P_r^a}{P_r^{na}}\right]$$

where e^i denotes individual i's income, with $\varepsilon, \gamma \in (0,1)$. Using Roy's identity, I derive the individual demand for the NA sector:

$$P_r^{na} C_{ir}^{na} = \left[(1-\gamma) - v\left[\frac{e^i}{(P_r^a)^\gamma (P_r^{na})^{1-\gamma}}\right]^{-\varepsilon}\right] e^i$$

When $v = 0$, the model simplifies to the homothetic case. For $v \neq 0$, changes in real income influence the expenditure share allocated to the NA sector, with this share converging to $(1-\gamma)$ as e^i grows large. The parameter ε determines the strength of income effects. When $v > 0$, the NA sector acts as a luxury good, with its expenditure share increasing as income rises. Local demand in region r:

$$P_r^{na} C_r^{na} = \underbrace{\left[(1-\gamma) - v\Psi_r\left[\frac{\mathrm{E}_r}{(P_r^a)^\gamma (P_r^{na})^{1-\gamma}}\right]^{-\varepsilon}\right]}_{\vartheta_r^{NA}} \mathrm{E}_r N_r \tag{12}$$

Here, $\mathrm{E}_r = w_r^l \left[F(\bar{\epsilon}_r) + \bar{\omega}_r \left(E[\epsilon^i | \epsilon^i > \bar{\epsilon}_r] - \mathbf{C_r}\right) h(\bar{\omega}_r, \mathbf{C}_r)\right]$ represents the average income in region r. Expression (8) shows that individual and local demand share a similar structure, adjusted for inequality, $\Psi_r = \frac{E_r\left[\left(e^i\right)^{1-\varepsilon}\right]}{\mathrm{E}_r^{1-\varepsilon}}$.[58] After determining expenditure shares for each sector aggregate, I assume that individuals exhibit CES preferences over the regional varieties within each sector, with an elasticity of substitution across varieties denoted by σ.

C.5 Proof of Proposition 3

The access costs of college $\mathbf{C_r}$ affect the decision to acquire a college degree. The total supply of efficiency units of skilled workers is given by:

$$H_r^s = E\left[\epsilon^i - \mathbf{C_r} | \epsilon_i \geq \bar{\epsilon}_r\right] h(\bar{\omega}_r, \mathbf{C_r}) N_r$$

Here, $h(\bar{\omega}_r, \mathbf{C_r}) = 1 - F\left(\bar{\epsilon}_r\right)$ and $\bar{\epsilon} = \frac{1}{\bar{\omega}_r} + \mathbf{C_r}$. Reducing $\mathbf{C_r}$ expands the supply of skilled workers for any given value of the returns to skill $\bar{\omega}_r$. Given the demand for skilled labor, this increase in supply reduces $\bar{\omega}_r$. Average income in region r is given by:

$$E_r[e^i] = w_r^l \left[F(\bar{\epsilon}_r) + \bar{\omega}_r E_r[\epsilon^i - \mathbf{C}_r | \epsilon^i > \bar{\epsilon}_r] h(\bar{\omega}_r, \mathbf{C}_r)\right]$$

A reduction in $\mathbf{C_r}$ leads to higher college attainment and increased earnings for college graduates. Moreover, as shown in Appendix C.2, the real earnings of unskilled workers, $\frac{w_r^l}{p_r^{na}}$, also increase. While there may be a second-order reduction in incomes for skilled workers who would have pursued college education even before the reduction in $\mathbf{C_r}$, the average income in region r rises overall.

A reduction in $\bar{\omega}_r$ also decreases ζ_r. As described in the proof of Proposition 2 in Appendix C.2, the expression for $\bar{z}_r$ allows me to compute the share of firms hiring skilled workers:

$$\left(\frac{B_r}{\bar{z}_r}\right)^{\lambda} = \frac{(\beta + (1-\alpha)(\lambda-1)) f_r^e}{f_r^o (1-\alpha-\beta)} \left[\frac{1-\alpha}{\beta}\zeta_r \left[\frac{\beta}{(\lambda-1)(1-\alpha)} + \left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}\right]\right]^{-1}$$

The total employment in these firms hiring skilled workers is given by:

$$L_{mir} + L_{eir} + L_{oir} + H_{ir} = \left(\frac{\alpha+\beta}{\beta}\right) \frac{f_r^o \zeta_r}{\bar{z}_r} \; z_{ir} - f_r^o \zeta_r + f_r^o + f_r^e \tag{13}$$

It follows that a reduction in ζ_r increases both the share of firms hiring skilled workers and the total employment in such firms. Consequently, for any definition of large firms in

[58] By Jensen's inequality, Ψ_r is less than 1, and:

$$E_r\left[\left(e^i\right)^{1-\varepsilon}\right] = \left(w_r^l\right)^{1-\varepsilon} \left[F(\bar{\epsilon}_r) + \left(\bar{\omega}_r\right)^{1-\varepsilon} E\left[\left(\epsilon^i - \mathbf{C}_r\right)^{1-\varepsilon} | \epsilon^i > \bar{\epsilon}_r\right] h(\bar{\omega}_r, \mathbf{C}_r)\right]$$

terms of employment, a reduction in ζ_r (weakly) increases the share of large firms hiring skilled workers. Since the total number of workers in region r is fixed, the share of total employment in these firms will also increase, provided that total employment in the NA sector rises when ζ_r decreases.

From Proposition 2, a decrease in ζ_r increases $\frac{w_r^l}{p_r^{na}}$, which is equal to A_r^{na}. From (EQ1):

$$\frac{L_{ar}}{\mathbf{L_{na,r}}} = \left(\frac{A_r^a}{A_r^{na}}\right)^{\sigma-1} \frac{\sum_{k=1}^{R} \frac{\tau_{kr}^{1-\sigma}\vartheta_k^A E_k\left[e^i\right]N_k}{\sum_j\left(w_j^l \tau_{kj}(A_j^a)^{-1}\right)^{1-\sigma}}}{\sum_{k=1}^{R} \frac{\tau_{kr}^{1-\sigma}\vartheta_k^{NA} E_k[e^i]N_k}{\sum_j\left(w_j^l \tau_{kj}(A_j^{na})^{-1}\right)^{1-\sigma}}}$$

When $\sigma - 1 > 0$, given A_r^a, an increase in A_r^{na} reduces the ratio of workers in A to those in NA. This labor transition from A to NA is driven by two forces: (i) an increase in the competitive standing of region r in the national NA market, and (ii) a rise in the expenditure share allocated to the NA sector due to income growth in region r. The reallocation of labor toward the NA sector, combined with the (weakly) increased share of large firms hiring skilled workers, implies a higher labor share in large firms within the NA sector.

C.6 Abstractions from the model

The model abstracts in several dimensions. First, it does not account for the potential effects of an expansion in the supply of skills on technology adoption. Theoretical and empirical work in the growth and development literature has established a link between education and technology adoption.[59] In the model, such technology adoption could influence the skill-biased technology component (Γ_r), neutral technology in the NA sector (B_r), and neutral technology in the A sector (A_r). If these mechanisms are significant in the data, the estimation procedure would infer growth in these technology components as exogenous and unrelated to reductions in $\mathbf{C_r}$. Consequently, omitting these mechanism from the model likely underestimates the contribution of rising college attainment to economic growth resulting from reduced $\mathbf{C_r}$. Incorporating these channels into the model would require assumptions about how college investments influence technology levels. For instance, one might posit:

$$\Gamma_r = \Gamma_r^0 \left(H_r\right)^{\gamma_s}$$

Here, Γ_r^0 represents the exogenous component of skill-biased technology, and $(H_r)^{\gamma_s}$ captures the effect of skilled workers on facilitating technology adoption. Estimating γ_s, the elasticity of technology to skilled labor, would be key. I leave this extension for future research.

Second, the model abstracts from worker migration across labor markets. Recent work by Hsiao (2022) developed a tractable framework in which college and migration choices are

[59] See Nelson and Phelps (1966), Foster and Rosenzweig (1996), Caselli and Coleman (2001) and Aghion et al. (2009).

complementary decisions. Empirical evidence from my analysis indicates that, conditional on controls, the net migration of college graduates is not significantly correlated with expansions in college attainment driven by reduced access costs. While this evidence supports the assumption of no mobility, Section E.1 leverages detailed census migration data to quantify the contribution of cross CZ migration flows to Brazil's observed economic development between 2000 and 2010. A more comprehensive analysis of the spatial sorting of skilled and unskilled workers resulting from increased college attainment is left for future work.

Third, the model assumes perfect capital markets in decisions regarding college enrollment. This assumption implies that individuals can separate education choices from consumption decisions without financial constraints.[60] If capital market imperfections are a binding constraint, the increase in college attainment following a reduction in $\mathbf{C_r}$ would depend on the correlation between wealth and ability.

Consider two extreme cases. In one, ability and wealth are perfectly positively correlated. In this scenario, capital market imperfections would have minimal effect on ability composition and would only affect the level of $\mathbf{C_r}$. In the other extreme, where ability and wealth are perfectly negatively correlated, reducing access costs would have limited impact on college attainment: financially constrained individuals with high ability would remain excluded, while financially unconstrained individuals with low ability would have low benefits from attending college. In intermediate cases, college attainment would rise for individuals who possess both sufficient ability and wealth.

Extending the model to incorporate heterogeneity in both ability and wealth would require detailed data on the wealth of potential college students. While some income information is available in the ENEM sociodemographic survey, it is reported in brackets and contains measurement error. These limitations mean that my measure of $\mathbf{C_r}$ may be influenced by the extent of financial constraints faced by prospective students. Moreover, any improvements in access to financial resources, such as loans or scholarships, would reduce $\mathbf{C}_r$ in the structural estimation. Thus, when financial constraints are binding, $\mathbf{C_r}$ should be interpreted broadly, encompassing any barrier preventing college enrollment for individuals who would benefit if adequate financing were available.

D Structural Estimation

D.1 Estimation of Structural parameters

This subsection describes the structural estimation of the parameters of the model.

[60] During the analysis period, higher education subsidies, such as FIES (2001) and PROUNI (2004), were not yet major contributors to the rise in college attainment. These programs, which became significant after 2005, were open to eligible students based on family income and ENEM scores.

Parameters governing the ability distribution: Using three administrative datasets, I estimate the ability distribution in the model $\epsilon \sim F(\cdot\,;\delta)$. The estimation connects cognitive ability measures, college affiliation, and labor market outcomes for Brazil's population of college graduates. First, I use data from the annual editions of a national standardized pre-college exam, ENEM, conducted between 2009 and 2012, to obtain measures of cognitive ability.[61] I will interpret the score obtain in this exam as representing ability in the model. Through personal identifiers, I link ENEM exam results to the higher education census.[62] This linkage identifies ENEM participants who completed college. I then match this subsample to the matched employer-employee data, RAIS, to obtain post-graduation formal sector income.[63] The linked data allows me to estimate returns to cognitive ability in the subsample of college graduates observed in RAIS.

In the model, the gross income of college workers is expressed as $e^i = w_r^h \epsilon_i$. Taking logarithms, I write $\ln e^i = y_i = \ln(w_r^h) + \ln(\epsilon_i)$. To approximate $\ln(\epsilon_i)$, I use the standardized distribution of ENEM exam scores for each year. Assuming $\epsilon_i = \exp(\delta_0 + \delta_1 S_i)$:

$$y_i = \ln(w_r^h) + \delta_0 + \delta_1 S_i$$

I estimate δ_1 using the following regression specification:

$$y_{irt} = \theta_0 + \delta S_{it} + X_i\beta + \theta_t + \theta_r + e_{irt} \tag{14}$$

Here, θ_r represents labor market fixed effects capturing $\ln(w_r^h)$, θ_t are year fixed effects, and X_i includes individual characteristics such as race, sex, age, and parental income and education. The variable S_{it} represents the standardized ENEM test score for each exam year. Table D.1 presents the estimation results.

Columns (1) to (3) show a positive and significant relationship between post-graduation income and ENEM scores. A one standard deviation increase in the ENEM score corresponds to an average income gain of 12%. This effect remains stable with the addition of pre-college controls and major fixed effects. The coefficient in Column (1) aligns with prior studies, such as Hanushek et al. (2015).[64] Given the estimates obtained from the data, I set $\delta_1 = 0.12$. Because δ_0 cannot be separately identified from w_r^h, I set its value equal to 1 to match the average college premium. Through the lens of the model, this procedure allows me to specify $\epsilon \sim \log\mathcal{N}(1, (0.12)^2)$.

[61] The ENEM is widely used by colleges for admissions and by the government to allocate tuition subsidies. Between 2009 and 2012, an average of five million individuals participated in the exam annually.

[62] Approximately 70% of college graduates were matched to the ENEM exam. The unmatched 30% likely took the exam before 2009 or entered college without ENEM scores.

[63] Around 50% of college graduates were matched to RAIS. Income data are restricted to two years after graduation.

[64] For example, Heckman, Stixrud, and Urzua (2006) estimated returns to cognitive ability for four-year college-educated workers in the US at approximately 0.12. Similarly, Lin, Lutter, and Ruhm (2018) reported values exceeding 0.20 by age 30, while Hanushek et al. (2015) provided a range of 0.12 to 0.24 across countries. My estimate is on the lower end of this range.

Table D.1: Ability distribution parameter

Dep. Variable	Annual income (logs)			Months employed			In RAIS		
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ENEM Score	0.12***	0.11***	0.09***	0.16***	0.15***	0.15***	0.04	0.02	0.02
	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Obs.	680124	645121	642044	680124	645121	642044	3180124	2945121	2939044
R^2	0.15	0.17	0.20	0.18	0.19	0.22	0.13	0.13	0.17
CZ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HH controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Experience	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Major FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: I use administrative data linking the ENEM, the Higher Education Census, and the matched employer-employee database (RAIS) to estimate specification (14) for the sample of college graduates. Columns (1) to (3) examine the relationship between pre-college cognitive ability and post-graduation income. Columns (4) to (6) focus on the share of months employed in a given year as recorded in the RAIS database. Finally, Columns (7) to (9) assess whether cognitive ability influences the probability of being observed in the RAIS database two years after post-graduation. Household controls include age, sex, race, an indicator for pre-college public education, commuting distance to the university, and parental education and income. The variable for experience is constructed as the number of months the graduate was observed employed in the RAIS since graduation. I include in specifications (3), (6), and (9) fixed effects for the major of study. All specifications include year and CZ fixed effects.

I also examine the relationship between cognitive ability and formal employment stability, measured as the share of months in a year the individual was employed in the RAIS. ENEM scores are positively and significantly associated with formal employment stability; a one standard deviation increase in the ENEM score results in a 16% increase in the share of months employed. Finally, using the ENEM linkage to the Higher Education Census, I assess the effect of cognitive ability on securing a formal sector job (i.e., being present in the RAIS). Unlike the previous outcomes, cognitive ability is not systematically related to formal sector employment probability.

Trade costs matrix: I recover the matrix of trade costs across labor markets using two primary sources of information. First, I use the matrix of travel times across Brazilian municipalities from Saldanha (2024). Second, I utilize estimates of average logistic costs in Brazil from World Bank (2012). With these two datasets, I adopt the mapping from travel times to trade costs proposed by Alder (2016). Specifically, I employ the following relationship:

$$1 + \tau_{rk} = 1 + \gamma \left(\text{Travel Time}_{rk}\right)^{0.8}$$

Here, τ_{rk} represents the iceberg trade costs between regions r and k in the model. I calibrate γ to target the average logistic costs as a percentage of value-added production in Brazil. By 2010, Brazil's logistic costs accounted for 22% of value-added production.

This estimate is consistent with an average trade cost of 1.28. The exponent value of 0.8 is adopted from Alder (2016).

Preference parameters: The model features four preference parameters. First, I identify γ. In the model, the agricultural expenditure share in region r is given by:

$$\vartheta_r^A = \gamma + v\Psi_r \left[\frac{\mathrm{E}_r}{(P_r^a)^\gamma (P_r^{na})^{1-\gamma}} \right]^{-\varepsilon}$$

As GDP per capita, E_r, grows large, ϑ_r^A asymptotically approaches γ. To identify γ, I use the agricultural employment share in the United States (0.01) as a proxy for the asymptotic value.

Second, Fan, Peters, and Zilibotti (2023) demonstrate that, under certain conditions, preferences over final consumption goods map directly to preferences over value-added, with the same parameter ε governing income elasticity strength in both cases. These features are critical for structural estimation, as they enable the use of household expenditure data to estimate preference parameters and calibrate local demand. Proposition 1 from Fan, Peters, and Zilibotti (2023) formally proves that, under PIGL preferences, one can estimate the parameter ε in the model, which features preferences over value-added, using household expenditure data. The adoption of PIGL preferences enables me to interpret the model's preference parameters, γ and v, as linear combinations of technology and preference parameters in a framework where final good producers combine outputs from sectors A and NA as inputs, while individuals derive utility from consuming these final goods (see Fan, Peters, and Zilibotti 2023). I use data from the Consumer Expenditure Survey (POF) for 2008-2009.[65] In the model, the individual-level agricultural expenditure share can be expressed as:

$$\frac{P_r^a C_{ir}^a}{e^i} - \hat{\gamma} = v \left[\frac{e^i}{(P_r^a)^\gamma (P_r^{na})^{1-\gamma}} \right]^{-\varepsilon}$$

Taking logs the relationship becomes linear, $y_i^a = \ln v + \varepsilon \ln \left((P_r^a)^\gamma (P_r^{na})^{1-\gamma} \right) - \varepsilon e^i$, which I estimate as:

$$y_{ir}^a = \theta_0 + \theta_r + \alpha e_{ir} + X_{ir}\beta + u_{ir} \tag{15}$$

Specifically, y_{ir}^a represents the food expenditure share, e_{ir} denotes total individual expenditure, and θ_r captures regional price effects. From the regression results, I identify the income elasticity $\varepsilon = -\hat{\alpha}$. The results of this estimation are summarized in Table D.2.

I estimate the relationship between household income and the expenditure share allocated to food using six specifications. Column (1) presents the baseline relationship between the household's monthly log expenditure and the share allocated to food, without any controls.

[65] The data was cleaned and processed using "Data Zoom," a resource provided by the Pontifical Catholic University of Rio de Janeiro.

Table D.2: Income elasticity parameter

Dep. Variable	Expenditure share in food					
Specification	(1)	(2)	(3)	(4)	(5)	(6)
Monthly expenditure (logs)	-0.29***	-0.29***	-0.29***	-0.30***	-0.27***	-0.27***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Obs.	48458	48458	48458	48458	48458	48458
R^2	0.04	0.04	0.06	0.08	0.22	0.23
HH Controls	No	Yes	Yes	Yes	Yes	Yes
HH Occupation FE	No	No	Yes	Yes	No	Yes
State FE	No	No	No	Yes	No	No
Strata FE	No	No	No	No	Yes	Yes

Notes: I estimate specification (15) to assess the association between monthly expenditure and the expenditure share in food. Household-level (HH) controls include household size, the number of earners, urban or rural status, and the age, education, and race of the household head. I also include household occupation fixed effects. Some specifications further incorporate state fixed effects or strata fixed effects. The strata are defined based on geographic features, income levels, and urban-rural status, resulting in 4,093 distinct groups.

Columns (2) and (3) add household-level controls and fixed effects for the occupation of the household head, yielding a stable coefficient of approximately 0.29. In Column (4), I include state fixed effects, resulting in a slightly higher coefficient of 0.3.

To estimate specification (15), the inclusion of CZ fixed effects would be ideal. However, because the POF survey does not identify CZ, I instead use the survey's stratification structure, which groups households based on geographic features, rural-urban classification, and income levels. This stratification is more granular than CZ fixed effects, making the specification more demanding. When I add strata fixed effects, the coefficient decreases slightly to 0.27. Further adding occupation fixed effects does not significantly alter the result, with the coefficient remaining stable around 0.27. Overall, the income elasticity remains consistent across specifications. Based on these results, I use $\varepsilon = 0.3$ for the structural estimation.

Third, I set the trade elasticity $\sigma = 4$ consistent with estimates from the trade literature. Finally, I estimate v using aggregate data from Brazil and the agricultural equilibrium condition in EQ1. Summing across regions r:

$$\sum_{r=1}^{R} w_r^l L_{ar} = \sum_{r=1}^{R} \vartheta_r^A \mathrm{E}_r N_r$$

By substituting the expression for ϑ_r^A, I derive the following relationship in 2000 and in 2010:

$$\frac{\sum_{r=1}^{R} w_r^l L_{ar}}{\sum_{r=1}^{R} \mathrm{E}_r N_r} - \hat{\gamma} = v \frac{\sum_{r=1}^{R} \Psi_r \left[\frac{\mathrm{E}_r}{(P_r^a)^{\gamma}(P_r^{na})^{1-\gamma}}\right]^{-\varepsilon} \mathrm{E}_r N_r}{\sum_{r=1}^{R} \mathrm{E}_r N_r}$$

I normalize $P_t^{na} = \sum_{r=1}^{R} P_{r,t}^{na} = 1$ for both $t = 2000$ and $t = 2010$. I then set the aggregate relative price of agriculture to non-agriculture at unity in 2000, $\frac{P_{2000}^a}{P_{2000}^{na}} = 1$, and use data obtained from the Economic Transformation Database (ETD) to obtain the observed change in the relative price between 2000 and 2010.[66] These information allows me to recover v using aggregate data from Brazil.

Technology parameters: I estimate four technology parameters. First, the demand for manufacturing workers in the model, $L_{mr} = \alpha \mathbf{L_{na,r}}$, provides a data moment to estimate α. Specifically, I compute:

$$\hat{\alpha} = \frac{1}{R} \sum_{r}^{R} \frac{L_{mr}}{\mathbf{L_{na,r}}}$$

Here, L_{mr} represents the number of manufacturing and craft workers, while $\mathbf{L_{na,r}}$ is the labor composite in the NA sector, consistent with the model:

$$\mathbf{L_{na,r}} = L_{mr} + L_{or} + L_{er} + \bar{\omega} H_{na}$$

Second, I estimate β using the ratio $\frac{\bar{\omega} H}{L_{mr}}$ within the NA sector and evaluate this ratio as $\Gamma_r \to \infty$. Under this condition, $\zeta_r < \frac{\beta}{1-\alpha}$ applies (see Proposition 2 in Appendix C.2). For any r:

$$\lim_{\Gamma_r \to \infty} \frac{\bar{\omega}_r H_r}{L_{mr}} = \frac{\beta}{\alpha}$$

I approximate this limit case using U.S. data. In the U.S., the share of workers with a college degree is approximately 0.25, while the share of production workers is around 0.19. These shares imply $\beta = 0.58$.[67]

Third, I assume the productivity distribution z_i follows a Pareto distribution with tail parameter λ. To estimate λ, I leverage the model structure and firm size distribution moments in Brazil (2000). The relationship between employment and productivity z_{ir} in the model is given by:

$$L_{mir} + H_{ir} = \underbrace{\left(\frac{\alpha + \beta}{\beta}\right) \frac{f_r^o \zeta_r}{\bar{z}_r}}_{\Phi_r} z_{ir} - f_r^o \zeta_r \tag{16}$$

[66] See website.

[67] I use the statistics from the US Bureau of Labor Statistics for the year 2012. See website.

This relationship implies a 1-to-1 mapping between employment and productivity. Using the productivity distribution assumption, I derive the following employment distribution condition:

$$\bar{G}(l) = G_r\left(\Phi_r z_{ir} - f_r^o \zeta_r \leq l\right)$$

Rearranging yields:

$$1 - \bar{G}(l) = \left(\frac{B_r \Phi_r}{l + f_r^o \zeta_r}\right)^{\lambda}$$

Taking logs motivates the estimating equation:

$$\ln(1 - \bar{G}(l)) = \theta_r + \lambda \ln\left(l + constant\right) \tag{17}$$

I construct the dataset for this specification using the matched employer-employee database, focusing on São Paulo or the top 5 or 10 labor markets with the highest number of firms. I define the upper tail of the firm size distribution using different employment thresholds, including the top 60%, 65%, 70%, 75%, and 80%. After selecting a sample and a tail definition, I segment firms into 100 grid points and compute $\ln(1 - \bar{G}(l_k))$ and $\ln\left(l_k + \text{constant}\right)$ at each point. I then estimate specification 17 based on this segmented sample. Table D.3 reports the results, from which I select $\lambda = 1.3$.

Table D.3: Tail parameter Lambda - Pareto distribution

Dep.Variable	Cumulative distribution of employment at the grid (logs)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Employment at grid (logs)	-1.37***	-1.34***	-1.31***	-1.29***	-1.26***	-1.31***	-1.28***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Obs.	65245	55409	48024	42712	34718	125879	168849
R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Tail	60	65	70	75	80	70	70
Sample	Sao Paulo	Sao Paulo	Sao Paulo	Sao Paulo	Sao Paulo	Top 5 MMC	Top 10 MMC

Finally, given the estimated values of α, β, and λ, the labor allocation in the NA sector depends on κ only through ζ_r. Setting $\kappa > 0$ preserves the positive relationship between skill intensity and firm size and does not affect labor allocation within the NA sector, as κ cannot be separately identified from Γ_r or f_r^o. I therefore normalize κ to 1.

D.2 Identification of Economic Fundamentals

Recovering college costs and returns to skill This subsection outlines how I recover the distribution of college costs, skill-biased technology, overhead fixed costs, and returns to college, denoted as $\mathbf{C}_r, \bar{\omega}, \Gamma_r f_r^o$. I leverage the structure of the model and the estimates of Ω^1, using labor market-level data and three model-based relationships. These relationships

capture the supply of skilled workers, the demand for skills, and the income gains from earning a college degree. The latter is expressed as:

$$\frac{E[e_r^i|\text{i is skilled}]}{E[e_r^i|\text{i is unskilled}]} = \underbrace{\bar{\omega}_r}_{\text{Skill premium}} \times \underbrace{E\left[\epsilon_{ir}|\epsilon_{ir} \geq \bar{\epsilon}_r\right]}_{\text{Ability composition}}$$

The observed income differences between skilled and unskilled workers reflect both market dynamics and selection. The skill premium measures wage differences, while the ability composition captures changes in the skill pool among college graduates. The three key equations are:

$$h_r = 1 - F\left(\bar{\epsilon}_r\right) \tag{18}$$

$$\theta_{1r} = \bar{\omega}_r E\left[\epsilon_{ir}|\epsilon_{ir} \geq \bar{\epsilon}_r\right] \tag{19}$$

$$\frac{\bar{\omega}_r H_r}{\mathbf{L_{na,r}}} = \Theta_1 \frac{\Theta_2}{\tilde{\zeta}_r(\bar{\omega}_r, \Gamma_r, f_r^o) + \Theta_2} \tag{20}$$

Here, h_r represents college attainment in region r, $\bar{\epsilon}_r$ is defined as $\frac{1}{\bar{\omega}} + \mathbf{C}_r$, and θ_{1r} are the income gains from earning a college degree. The parameters in equation (20) are formally defined as $\Theta_1 = \frac{\beta+(\lambda-1)(1-\alpha)}{\lambda}$, $\Theta_2 = \frac{\beta}{(\lambda-1)(1-\alpha)}$ and $\tilde{\zeta}_r(\bar{\omega}_r, \Gamma_r, f_r^o) = \left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(\lambda-1)(1-\alpha)+\beta}{1-\alpha-\beta}}$. These equations describe the interplay between skill supply, demand, and the distribution of ability. By using data on h_r, θ_{1r}, and my estimate of $F(\cdot\,|\hat{\delta})$, I solve this system to recover the economic fundamentals. The strategy is formalized as follows:

Proposition 4 *There is a unique solution that maps economic fundamentals to moments in the data using the system of equations (18) - (20):*

$$\bar{\omega}_r = \frac{\hat{\theta}_{1r} - 1}{E\left[\epsilon_{ir}|\epsilon_{ir} \geq F^{-1}(1-h_r)\right] - F^{-1}(1-h_r)}$$

$$\mathbf{C}_r = \frac{\hat{\theta}_{1r} F^{-1}(1-h_r) - E\left[\epsilon_{ir}|\epsilon_{ir} \geq F^{-1}(1-h_r)\right]}{\hat{\theta}_{1r} - 1}$$

$$\Gamma_r f_r^o = \left[\Theta_2\left(\frac{\Theta_1}{d_r} - 1\right)\right]^{-\rho_1} \frac{(1-\alpha)\kappa\left(\hat{\theta}_{1r} - 1\right)}{\beta\left(E\left[\epsilon_{ir}|\epsilon_{ir} \geq F^{-1}(1-h_r)\right] - F^{-1}(1-h_r)\right)}$$

Where $\rho_1 = \frac{1-\alpha-\beta}{(\lambda-1)(1-\alpha)+\beta}$, *and* $d_r = \frac{\bar{\omega}_r H_r}{\mathbf{L_{na,r}}}$.

The proof of Proposition 4 involves solving a system of three equations on three unknowns. I solve this system using data on average income gains from a college degree, college attainment, and the estimated ability distribution. Income gains from a college degree (θ_{1r}) are obtained from census data (2000 and 2010). For each labor market, I estimate θ_{1r} using the following regression:

$$y_{ir} = \theta_0 + \theta_{1r} D_{ir}^{HE} + X_{ir}\Gamma + \epsilon_{ir}$$

Here, y_{ir} is the log income of individual i, and $D^{HE}i$ is a dummy equal to 1 if the individual has a college degree. Controls include age, sex, race, and 2-digit sector fixed effects. This strategy yields a distribution $\hat{\theta}_{1r}$ for each census year. Conditional on h_r, larger observed income differences $\hat{\theta}_{1r}$ imply higher college costs $\mathbf{C}_r$, while lower h_r implies higher $\mathbf{C}_r$ for a given $\hat{\theta}_{1r}$.

There are three assumptions underlying this calibration. First, I assume the ability distribution is identical across CZs. This assumption impacts the interpretation of $\mathbf{C}_r$. For example, if region A's ability distribution stochastically dominates region B's and $h_A < h_B$, the model attributes the difference to higher college costs in A. If ability distributions differ significantly, differences in h_r may instead reflect differences in the distribution of ability rather than costs. This assumption may affect level estimates of $\mathbf{C}_r$ and its changes over time. Limited data on ability distributions at the CZ level constrains the analysis, so I interpret some variation in $\mathbf{C}_r$ as reflecting differences in ability distributions.

Second, the method includes both local and immigrant populations in college attainment measures. If migration increases h_r in a CZ, the model will underestimate $\mathbf{C}_r$ relative to a no-migration scenario. Conversely, net emigration of college graduates will overestimate $\mathbf{C}_r$. To address this measurement problem, I use census data to calculate pre-migration college attainment in each CZ. This adjustment provides corrected estimates of local college costs.

Finally, financial conditions influencing college enrollment decisions affect the calibrated $\mathbf{C}_r$. Improved access to loans or scholarships during the analysis period would reduce calibrated college costs. Notably, two major programs—FIES, which provides college loans, and PROUNI, which offers scholarships for high-achieving, low-income students based on ENEM exam scores—became available in the early 2000s but only benefited a small fraction of total enrollment in Brazil during the period analyzed. Thus, their influence on the measured decline in $\mathbf{C}_r$ is likely limited.

Recovering Γ_r, f_r^o and f_r^e I estimate the skill-biased technology component Γ_r and the fixed cost parameters f_r^o and f_r^e in the NA sector. In the model, the share of firms hiring skilled workers is defined by the following expression:

$$\left(\frac{B_r}{\bar{z}_r}\right)^{\lambda} = \frac{(\beta + (1-\alpha)(\lambda-1))f_r^e}{f_r^o(1-\alpha-\beta)} \left[\frac{\zeta_r}{\lambda-1} + \left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)\lambda}{1-\alpha-\beta}}\right]^{-1}$$

Using the estimated values of Ω^1, $\Gamma_r f_r^o$, and $\bar{\omega}$, I combine this equation with data on the share of firms in the NA sector hiring skilled workers in each region to recover the distribution of $\left\{\frac{f_r^e}{f_r^o}\right\}$.

Next, the total number of firms operating in each location is given by the following relationship:

$$\left(\frac{B_r}{z_r^*}\right)^{\lambda} M_r = \begin{cases} \frac{\beta+(1-\alpha)(\lambda-1)}{f_r^o \lambda} \frac{\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}}{\left[\frac{\beta}{(\lambda-1)(1-\alpha)}+\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}\right]} \mathbf{L_{na,r}} & \zeta_r \geq \frac{\beta}{1-\alpha} \\ \frac{(\lambda-1)(1-\alpha-\beta)}{f_r^o(1-\zeta_r)\lambda} \mathbf{L_{na,r}} & \zeta_r < \frac{\beta}{1-\alpha} \end{cases}$$

Focusing on the case where not all firms in the economy hire skilled workers ($\zeta_r \geq \frac{\beta}{1-\alpha}$), as observed in the data, I use the total number of firms in the NA sector to recover the distribution of f_r^o. Finally, with the calibrated values of f_r^o and the calibrated values of $\Gamma_r f_r^o$ and $\left\{\frac{f_r^e}{f_r^o}\right\}$, one can recover the distributions of Γ_r and f_r^e.

Recovering A_r^a and B_r I follow the method outlined in Fan, Peters, and Zilibotti (2023) to recover the productivity fundamentals $A_{r,t}^a, B_{r,t}$. This approach leverages the model's equilibrium structure, along with data on wages (w_r^l) and sectoral labor allocations ($L_{ar}, \mathbf{L_{na,r}}$), to uniquely identify the productivity distributions in the A and NA sectors for 2000 and 2010.

In equilibrium, prices are linked to productivities as $p_r^s = \frac{w_r^l}{A_r^s}$. The equilibrium conditions in (EQ1) can be rewritten in terms of prices, iceberg costs, wages, and labor allocations:

$$p_r^a = \left(w_r^l L_{ar}\right)^{\frac{1}{1-\sigma}} \left(\sum_{k=1}^{R} \frac{\left(\tau_{kr}\right)^{1-\sigma}}{\sum_j \left(p_j^a \tau_{kj}\right)^{1-\sigma}} \vartheta_k^A E_k \left[e^i\right] N_k\right)^{\frac{1}{\sigma-1}} \tag{21}$$

$$p_r^{na} = \left(w_r^l \mathbf{L_{na,r}}\right)^{\frac{1}{1-\sigma}} \left(\sum_{k=1}^{R} \frac{\left(\tau_{kr}\right)^{1-\sigma}}{\sum_j \left(p_j^{na} \tau_{kj}\right)^{1-\sigma}} \vartheta_k^{NA} E_k \left[e^i\right] N_k\right)^{\frac{1}{\sigma-1}}$$

As observed, this system is homogeneous of degree 0. The normalization $\sum_{r=1}^{R} P_{r,t}^{na} = 1$ for both $t = 2000$ and $t = 2010$, $\frac{P_{2000}^a}{P_{2000}^{na}} = 1$, and $\frac{P_{2010}^a}{P_{2010}^{na}} = 0.87$—consistent with the observed relative price change from the Economic Transformation Database (ETD)—ensures a unique set of equilibrium prices p_r^a, p_r^{na}. Using the recovered price distribution and observed wage data, I compute productivity fundamentals as follows:

$$A_r^a = \frac{w_r^l}{p_r^a} \qquad A_r^{na} = \frac{w_r^l}{p_r^{na}} \tag{22}$$

Given the estimates and calibration of Ω^1, A_r^{na}, f_r^e, and ζ_r, I use the model's expression for A_r^{na} to infer B_r. For $\zeta_r \geq \frac{\beta}{1-\alpha}$, I have:

$$A_r^{na} = \alpha \kappa^{\beta} \left(\frac{1-\alpha}{\alpha f_r^o}\right)^{1-\alpha} \left(z_r^*\right)^{1-\alpha-\beta}$$

Where z_r^* satisfies:

$$(z_r^*)^\lambda = \frac{f_r^o(1-\alpha-\beta)}{(\beta+(1-\alpha)(\lambda-1))f_r^e}\left[\frac{\frac{\beta}{(\lambda-1)(1-\alpha)}+\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+beta}{1-\alpha-\beta}}}{\left(\frac{1-\alpha}{\beta}\zeta_r\right)^{\frac{(1-\alpha)(\lambda-1)+\beta}{1-\alpha-\beta}}}\right]B_r^\lambda$$

For $\zeta_r < \frac{\beta}{1-\alpha}$

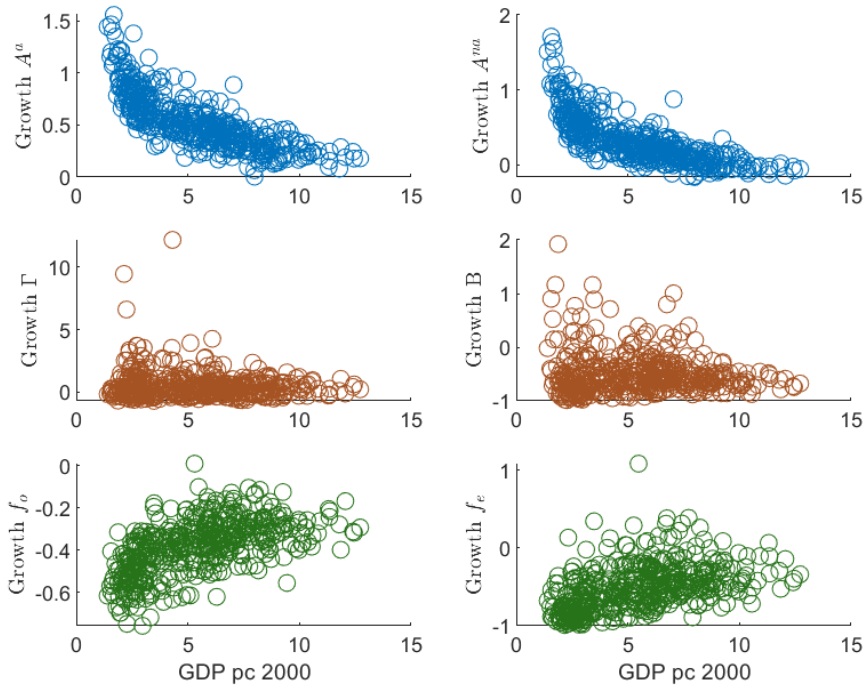
$$A_r^{na} = \alpha^\alpha\left(\frac{1-\alpha-\beta}{f_r^o(1-\zeta_r)}\right)^{1-\alpha-\beta}\left(\frac{\beta\kappa}{f_r^o\zeta_r}\right)^\beta(z^*(B))^{1-\alpha-\beta}$$

with:

$$(z_r^*)^\lambda = \frac{f_r^o(1-\zeta_r)}{(\lambda-1)f_r^e}B^\lambda$$

D.2.1 Calibrated economic fundamentals

Figure D.1: Calibrated technology fundamentals



Notes: Figure D.1 illustrates the relationship between technology fundamentals and initial economic development, measured as log GDP per capita in 2000. The x-axis represents the 2000 log GDP per capita, while the left y-axis displays the calibrated log change in the technology fundamentals between 2000 and 2010. Each marker corresponds to a distinct labor market (CZ).

E Growth Decomposition

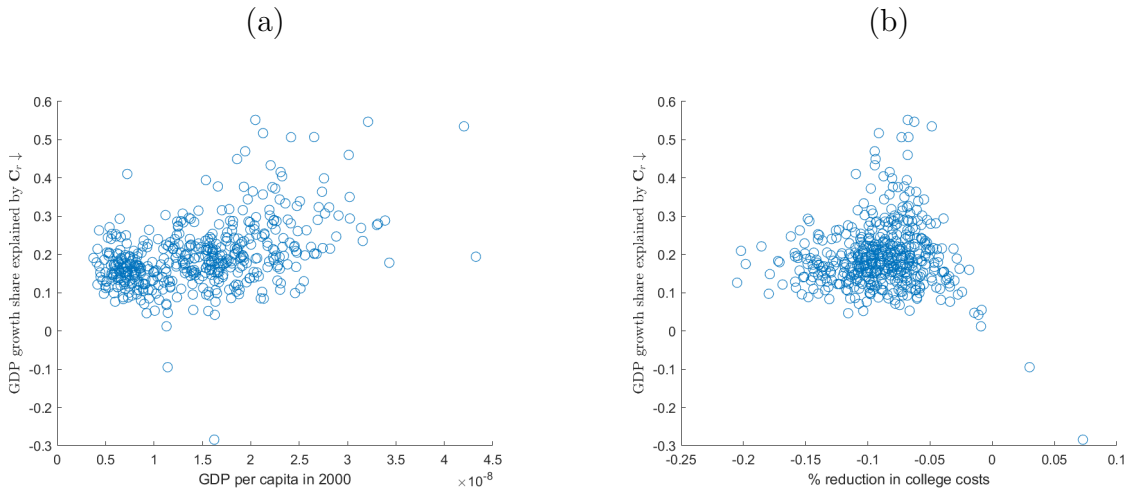
E.1 Considering migration

E.2 Heterogeneity

In Panel (a), the x-axis displays the change in the college costs experienced between between 2000 and 2010. In the y-axis, I show the reduction in GDP per capita growth if one keeps the college costs at the 2000 level. In Panel (b), the y-axis displays the change in the decline of the agricultural employment share if one keeps the college costs at the 2000 level.

First, the results show that the "lost growth" in GDP per capita, or the slower decline in agricultural employment share, increases when the counterfactual represents a larger increase in college costs (i.e. when the reduction in college costs between 2000 and 2010 is larger). These results are in line with the findings for the aggregate economy. Second, there is substantial variation across regions that experienced similar magnitudes in the decline of college costs. I provide some rationale for heterogeneity in the reduction in GDP per capita growth.

Figure E.1: Growth Decomposition at the CZ Level



Notes: In Panel (a), the y-axis shows the loss in GDP growth from the absence of the calibrated decline in college access costs. The x-axis displays the initial level of GDP per capita in 2000. In Panel (b), the y-axis provides the same information as in Panel (a), while the x-axis displays the calibrated decline in college access costs. Each marker represents a CZ.