# Higher Education and Economic Development: Evidence from College Expansion in India

Bipul Verma*

February 2026

CLICK HERE FOR LATEST VERSION

**Abstract**

I study the role of private higher education in driving economic development. Using newly digitalized data on college and university locations in India and a large-scale policy reform in the private higher education sector, I estimate the causal effect of college expansion on macroeconomic outcomes. My identification strategy exploits heterogeneity in college entry generated by institutional restriction that required new private colleges to obtain licenses from pre-determined state public universities. I find that a one-percentage-point increase in the share of college graduates over a decade reduces agricultural employment by 0.8 percentage points while increasing services employment by 0.9 percentage points. It also raises real GDP per worker by 2.3 percent and nightlight intensity by 6.2 percent. The estimates imply that social returns to human capital are 30 percent larger than private returns. These results are consistent across instrumental variables, event studies, and border discontinuity designs. I then use the reduced-form estimates to calibrate a spatial general equilibrium model with endogenous education choices. The model shows that for the average district, India's higher education reform accounts for 20 percent of output growth between 2001 and 2011. Counterfactual simulations reveal that removing spatial barriers to college opening would have increased the average district's 2011 output by 3 percent. These findings demonstrate that higher education expansion is a powerful driver of economic development.

**Keywords:** Human Capital; Higher Education; Structural Transformation; Economic Growth.

*Address: Department of Economics, University of Minnesota. Email: verma176@umn.edu. I am greatly indebted to my advisors Kjetil Storesletten, Todd Schoellman, Hannes Malmberg, and Alessandra Fogli for their invaluable advice and support. I also wish to thank Andrew Goodman Bacon, Anmol Bhandari, Honey Batra, Jason Hall, Agustin Hurtado, Abhiroop Mukhopadhyay, Jo Mullins, and Mariacristina De Nardi for helpful comments and discussion. I am also grateful to Patrick Behrer, Rishabh Choudhary, and Dhruv Sharma for providing me with district level GDP data for India.

# 1 Introduction

Economic development is associated with systematic changes in the structure of the economy. As countries grow, workers move out of agriculture into manufacturing and services, and employment shifts toward larger and more formal firms. Education has long been regarded as one of the fundamental engines of this process. A more educated workforce accelerates innovation, enhances the capacity to adopt frontier technologies, and facilitates knowledge diffusion across firms and workers. Despite the close correlation between education and development, establishing whether education causes economic development remains a central empirical challenge.

The lack of quantitative estimates for how higher education contributes to growth makes it difficult for policymakers in both developed and developing economies to assess how much to invest in higher education. Concerns regarding institutional quality, skill mismatch, queuing for government jobs, migration, and the lower ability of marginal students further reduce confidence in higher education investment. The main challenge in providing a quantitative estimate of education's causal effect on economic development has been the difficulty of isolating education from other drivers of growth. Economies can grow for many reasons (such as trade reforms, technology transfer, or institutional improvements) and this growth can then drive increases in education, making it hard to isolate the effect of education on development. Identifying the causal role of education therefore requires a setting where educational attainment varies for reasons unrelated to factors that directly affect economic development.

I provide new causal evidence on the role of higher education in driving economic development. I exploit quasi-random variation in the expansion of college-educated workers across Indian districts to address two questions: (i) does higher education cause economic development, and (ii) how much of output growth due to education arises from human-capital externalities?

India provides a unique setting to study these questions. Between 1991 and 2000, the Government of India undertook a series of reforms that reduced barriers to entry for private colleges, making it substantially easier to establish new institutions. The reforms relaxed restrictions on private college operation, streamlined approval processes, and encouraged state governments to facilitate entry. The resulting wave of entry led to a threefold increase in both the number of degree-granting colleges and total enrollment over ten years.

However, the impact of these reforms varied substantially across space due to a key institutional constraint: India's affiliation system. Under this system, every college must be affiliated with a specific public university responsible for curriculum design, examinations, and degree conferral. By law, each district is assigned to a single affiliating university based on jurisdictional boundaries established decades earlier.[1][2] Colleges cannot affiliate with a geographically

[1] A district is a sub-state administrative unit in India.

[2] Affiliation works like a franchise or licensing arrangement. The private college operates the campus and enrolls students, but the university oversees academic standards: it sets the curriculum, conducts examinations, and issues the degree under its name. Each district is legally assigned to one public university, and colleges must affiliate with that university even if another university is geographically closer. For example, in Andhra Pradesh, both Srikakulam and East Godavari districts are required to affiliate with Andhra University in Visakhapatnam.

closer university if it lies outside their legally defined jurisdiction. This means that districts assigned to more distant universities faced higher costs of establishing new colleges, even after the reforms removed other barriers to entry. Because the jurisdictional map was fixed before the reform period and determined by historical factors unrelated to future economic conditions, distance to the affiliating university provides a plausibly exogenous source of variation in post-reform college expansion.

I develop three complementary identification strategies to address the endogeneity of education. First, I use distance to the affiliating university as an instrument for the change in the local share of college graduates between 2001 and 2011 in a two-stage least squares framework. Second, I implement an event-study design that exploits the 2001 liberalization as a common policy shock, allowing the treatment effect to vary continuously with pre-reform distance to the affiliating university. Third, I estimate a border discontinuity design comparing adjacent districts on opposite sides of university jurisdiction boundaries. All three approaches yield consistent point estimates of the causal effect of higher education on economic development across outcomes. [3]

The results reveal large and economically meaningful effects. A one-percentage-point increase in the district-level share of college graduates over a decade leads to a 2.3 percent increase in GDP per worker and a 6.2 percent increase in nightlight intensity. The expansion of higher education also drives substantial labor reallocation: agricultural employment declines by 0.8 percentage points while service employment rises by 0.9 percentage points. Consistent with a strong local supply response, the skill wage premium falls by 5 percent. These findings demonstrate that higher education plays a causal role in accelerating labor reallocation and output growth at the regional level.

The magnitude of the total effect on GDP per worker substantially exceeds what would be predicted by private Mincerian returns alone, pointing to important human-capital externalities. I quantify these externalities by decomposing the total effect into two components: the direct effect operating through private returns and labor market substitution, and the residual effect that I attribute to human capital externalities.[4] Of the total 2.3 percent effect on GDP per worker, Mincerian returns and substitution account for 1.0 percentage points (44 percent), while the remaining 1.3 percentage points (56 percent) reflect human-capital externalities. This decomposition implies that social returns to higher education are 30 percent larger than private returns. The magnitude of the estimated human-capital externality can reconcile the observed

Srikakulam is about 120 km away, while East Godavari is about 250 km away. Both face the same academic rules, but the farther district incurs higher recurring travel and coordination costs for inspections, exam administration, and transcript processing.

[3] Statistical significance varies across identification strategies. The instrumental variable approach produces the strongest set of results, with all estimates statistically significant and high first-stage F-statistics avoiding weak instrument concerns. However, in the border discontinuity specification, the smaller number of effective districts sharing borders and the possibility of inter-district trade attenuating local effects result in insignificant coefficients for output, though point estimates remain similar in magnitude and all other outcomes are statistically significant.

[4] It is possible that part of the difference between Mincerian returns and total returns is internalized by firms or workers. In such case, the estimates for externalities presented here represent an upper bound.

high correlation between GDP and schooling in cross-section data with the relatively low Mincerian returns found in micro studies [Banerjee and Duflo, 2005].

There are two complementary mechanisms driving these results. First, the increase in college graduates reduces the skill wage premium, making skilled workers cheaper to hire and inducing firms to substitute toward skill-intensive production. I document this by disaggregating employment changes: within services, high-skill employment growth concentrates in large establishments while low-skill employment growth concentrates in small establishments, suggesting that higher-education expansion fosters modernization in skill-intensive activities and generates complementary demand for low-skill services. Second, beyond substitution effects, more educated people can have a direct effect on TFP through knowledge spillovers, and innovation.

To formalize this interpretation, I calibrate a spatial general-equilibrium model with endogenous education choices and localized human-capital spillovers. Two counterfactual exercises illustrate the quantitative importance of these findings. First, the model implies that for the average district, India's higher-education reform accounts for approximately 20 percent of output growth between 2001 and 2011. Second, a policy that removes spatial constraints on university affiliation by allowing districts to affiliate with the nearest university would increase the average district's output by 3 percent relative to the 2011 baseline. This demonstrates substantial potential gains from further reducing geographic barriers to higher education access.

The paper proceeds as follows. Section 2 describes the data sources and construction of key variables. Section 3 presents an illustrative model that clarifies why OLS estimates are biased and motivates the identification strategy. Section 4 provides institutional details on India's higher-education system and the affiliation requirement. Section 5 outlines the empirical specification and identification strategies. Section 6 presents the main results. Section 7 discusses robustness checks. Section 8 develops and calibrates the spatial equilibrium model. Section 9 concludes.

## Literature Review

This study contributes to several branches of research on higher education, economic growth, and structural transformation. Its primary contribution is to the literature examining the causal effect of education on economic development. Establishing causality is challenging because schooling is endogenous to income, technological change, and institutional quality [Bils and Klenow, 2000, Krueger and Lindahl, 2001]. Early causal studies focused on private returns to education [Duflo, 2001, Khanna, 2023]. Evidence on aggregate effects of education on economic development is limited, particularly in developing economies. In advanced-economy settings, several studies find no effect of increased schooling on regional productivity [Acemoglu and Angrist, 2000, Ciccone and Peri, 2006], while others document small positive spillovers [Moretti, 2004, Gennaioli et al., 2013]. Valero and Van Reenen [2019] find that university open-

ings raise regional GDP growth in a cross-country panel. In a developing-economy context, Castelló-Climent et al. [2018] use historical missionary institutions to show that early access to higher education predicts long-run regional prosperity in India. Causal evidence on contemporaneous large-scale higher-education expansion in developing economies remains limited. This paper exploits a nationwide policy reform that expanded private colleges across India, providing causal estimates of how increases in tertiary attainment affect productivity growth and structural transformation.

The paper also contributes to the literature on structural transformation and human capital. Theoretical and empirical work links rising schooling to the expansion of high-skill service activities and systematic shifts toward skill-intensive industries [Buera et al., 2022]. Cohort-based evidence indicates that increases in educational attainment accelerate the exit from agriculture [Porzio et al., 2022]. Country-specific evidence further supports this view in Brazil [Cox, 2024], China [Coelli et al., 2023], and Vietnam [Vu, 2023]. This paper provides complementary evidence from India, showing that a rise in educational attainment compresses the skill premium and reallocates employment toward sectors intensive in educated workers—particularly modern services.

The paper also contributes to research on India's post-liberalization growth. The prevailing view emphasizes productivity improvements and demand-side forces, particularly globalization and the rise of skill-intensive, export-oriented services [Shastry, 2012, Ghose, 2021], as well as sector-specific TFP gains [Gordon and Gupta, 2005, Eichengreen and Gupta, 2013, Fan et al., 2023]. This paper introduces a complementary supply-side perspective, demonstrating that the liberalization of higher education played a pivotal role in sustaining structural transformation.

Finally, the paper relates to research on how geographic access shapes college enrollment. Foundational studies show that proximity to colleges strongly influences educational attainment in advanced economies [Card, 1993, Currie and Moretti, 2003, Goodman et al., 2017]. Evidence from developing economies is more limited. Sekhri et al. [2024] find that expanding college access in India increases human-capital accumulation and women's empowerment. This paper provides causal evidence that increasing local college density raises tertiary enrollment, extending predominantly high-income evidence to a setting where educational supply constraints have historically been severe.

# 2 Data

The empirical analysis uses a balanced panel of Indian districts observed in 1991, 2001, and 2011. The district serves as the unit of analysis because it represents the primary administrative subdivision below the state level and provides a natural geography for capturing local labor markets and educational access. The sample covers all districts in mainland India for which consistent data are available across the three census years. To ensure temporal consistency despite administrative boundary changes and district splits, all variables are harmonized to a

common 1991 district classification using geographic concordance procedures detailed in the Appendix.

**University Location and Affiliation Structure** The analysis requires detailed information on the spatial organization of India's higher education system. These data are constructed from multiple editions of the *University Handbook* published by the Association of Indian Universities, a comprehensive directory of all recognized universities in India. Each edition reports the name, year of establishment, location, and jurisdictional boundaries of every university, along with information on affiliated colleges, including their locations and enrollment figures.

The 1991 edition serves as the baseline for measuring the pre-liberalization structure of the university affiliation system. For each district, I identify: (i) the affiliating university to which the district is assigned, (ii) the geographic distance from the district centroid to the university headquarters, and (iii) the number and location of affiliated colleges within the district. This institutional mapping provides the foundation for the instrumental variable strategy, which exploits exogenous variation in access to university resources based on historical affiliation boundaries.

**Educational Attainment and Employment** Data on educational attainment and employment come from two complementary sources. The decennial Population Census of India provides comprehensive population counts disaggregated by education level (primary, secondary, tertiary) and sector of employment (agriculture, manufacturing, services). These data are available at the district level for 1991, 2001, and 2011 and form the basis for measuring changes in the educational composition of the workforce and sectoral employment shares.

Individual-level wage data are obtained from the Employment and Unemployment Rounds of the National Sample Survey (NSS), India's principal household survey. I use several NSS survey rounds to obtain detailed information on wages, education, occupation, industry, and demographic characteristics. The NSS data are aggregated to construct district-level measures of mean real wages by education group.

**Firm Size Distribution and Industrial Composition** Information on the organization of production and the distribution of employment across establishment sizes comes from the Economic Census of India. Conducted in 1990, 1998, 2005, and 2013, the Economic Census provides a complete enumeration of all non-farm establishments in the country. For each district, the census reports the number of establishments and total employment disaggregated by industry (at the 2-digit NIC level) and establishment size class. These data permit analysis of structural transformation not only across sectors but also within sectors, distinguishing between small enterprises and larger establishments.

**District GDP** District-level measures of real GDP are obtained from Behrer et al. [2023], who compile real district GDP from officially published state government reports and CEIC, a data platform that aggregates data from various official sources. The GDP series is available for three different base years (1999-00, 2004-05, and 2011-12) starting from fiscal year 1999-2000. To construct a consistent time series, Behrer et al. [2023] create a spliced growth rate series by using the real GDP growth rate implied by the latest available base year; for instance, if the 2011 growth rate is available under both the 2004-05 and 2011-12 base years, they use the growth rate based on the 2011-12 base year.

**Nighttime Lights** I use nighttime light intensity data from the Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS) for the period 1992-2013 [NOAA National Geophysical Data Center, 2014]. The DMSP/OLS sensors detect low-light imaging of the earth's surface, capturing anthropogenic light sources such as cities, towns, and industrial sites. The data are provided as annual composites with cloud-free coverage at approximately 1 $km^2$ spatial resolution. Each pixel is assigned a digital number ranging from 0 to 63, representing the intensity of stable lights. I aggregate pixel-level light intensity to the district level by calculating the total brightness across all pixels within each district's boundaries and then normalize by district area to obtain light intensity per square kilometer. Nighttime lights serve as a complementary measure of economic activity that is particularly useful in settings where official GDP data may be limited or measured with error [Henderson et al., 2012].

**Sample Selection** The main analysis excludes all Union Territories due to frequent administrative reorganizations and substantial gaps in census data.[5] The sample is further restricted to districts that did not have a state public university in 1991. This restriction ensures that the estimated effects capture the regulatory impact of distance to the affiliating university on college expansion, rather than conflating this with the direct economic benefits of hosting a major educational institution locally. Importantly, this sample restriction implies that the presence of large universities is not driving the results, since the comparison is conducted only among districts without a state public university. An additional advantage of this restriction is that it excludes major metropolitan areas such as Kolkata, Bangalore, and Chennai, which hosted universities from the outset. Institutional restrictions on college entry are likely less binding in such large cities, which can experience growth through multiple channels unrelated to the affiliation system. By focusing on non-university districts at varying distances from their affiliating institutions, the sample of districts becomes more homogeneous, adding credibility to the identification strategy described in Section 5. For comparison, the Appendix presents results without imposing this sample restriction.

[5]Union Territories are administrative divisions in India that are directly governed by the central government, unlike states which have their own elected governments and greater administrative autonomy. During the sample period, Union Territories included Delhi, Chandigarh, Puducherry, Andaman and Nicobar Islands, Dadra and Nagar Haveli, Daman and Diu, and Lakshadweep.

**Summary Statistics** Table 1 presents summary statistics for key variables used in the analysis, pooled across the years 1991, 2001, and 2011. The sample comprises 370 districts after applying sample selection criteria. With three census years, the maximum potential sample size is 1,110 district-year observations. However, the actual sample size varies across variables due to data availability. Real GDP per worker is not available for 1991, and some districts have missing data in certain years. Other key variables such as college supply, educational attainment, and sectoral employment shares have more complete coverage, with sample sizes ranging from 1,084 to 1,115 observations. The average district has 32 colleges and a college graduate share of 9 percent of the working-age population. Agricultural employment accounts for 68 percent of employment on average, while manufacturing and services represent 8 percent and 20 percent, respectively.

Table 1: Summary Statistics

| Variable | Mean | Median | SD | Min | Max | N |
|---|---|---|---|---|---|---|
| Total colleges | 32 | 13 | 54 | 0 | 1048 | 1115 |
| College graduate share | 0.09 | 0.07 | 0.06 | 0.01 | 0.48 | 1110 |
| Agriculture share | 0.68 | 0.72 | 0.15 | 0.13 | 0.93 | 1110 |
| Manufacturing share | 0.08 | 0.07 | 0.06 | 0.00 | 0.40 | 1110 |
| Service share | 0.20 | 0.18 | 0.09 | 0.05 | 0.66 | 1110 |
| Real GDP per worker | 72164 | 53132 | 60452 | 11000 | 490640 | 748 |
| Nightlight intensity | 4.49 | 2.76 | 5.05 | 0.00 | 46.07 | 1084 |
| College Wage Premia | 3.21 | 2.82 | 1.69 | 0.50 | 16.36 | 1102 |

# 3 Illustrative Model and Identification Strategy

The objective of the paper is to understand how expansion of skill affects macroeconomic variables in the economy. The effect can be captured by estimating the coefficients (semi-elasticities) from the following regression:

$$\Delta y \;=\; \beta_0 \;+\; \beta_1 \, \Delta h \;+\; \varepsilon,$$

where $\Delta y$ is the change in a macroeconomic variable such as employment share, output, or wages, $\Delta h$ is the change in the share of skilled individuals, and $\beta_1$ captures the effect of skill expansion on outcomes. However, OLS estimates of $\beta_1$ will be biased since there are unobserved factors in $\varepsilon$ such as productivity growth that are omitted from the regression. An unbiased estimate of $\beta_1$ requires an instrument that shifts $\Delta h$ but is uncorrelated with other factors affecting both outcomes and skill changes. The following illustrative model makes this point more clearly.

Consider a stylized environment in which aggregate output is produced with skilled labor $h$

and unskilled labor $\ell = 1 - h$:

$$y \;=\; A(h)\, F(\theta_H h,\, \theta_L \ell), \qquad F_{\cdot} > 0.$$

Productivity features a skill externality,

$$A(h) \;=\; \bar{A}\, h^{\Omega}, \qquad \Omega > 0,$$

so a higher skilled share raises total factor productivity.

Taking a first-order finite-difference approximation around an interior baseline with skilled share $h_0$ and factor cost shares $s_H, s_L$, the change in output satisfies

$$\Delta \log y \;=\; \psi_{yh}\, \Delta \log h \;+\; \Delta \log \bar{A},$$

where the composite elasticity

$$\psi_{yh} \;\equiv\; \Omega \;+\; s_H \;-\; s_L \,\frac{h_0}{1 - h_0}$$

captures (i) the direct externality of skills on productivity and (ii) the induced compositional effect from reallocating labor between $h$ and $\ell$.

On the supply side, the skill share responds to incentives via a finite-difference logit equation,

$$\Delta \log\!\Big(\tfrac{h}{1-h}\Big) \;=\; \beta_w\, \Delta \log\!\Big(\tfrac{w_H}{w_L}\Big) \;+\; \beta_y\, \Delta \log y \;+\; \beta_c\, \Delta \log C,$$

with $C$ denoting the cost of skill acquisition (tuition, access costs, etc.). The cost shifts move the skilled share directly through $\beta_c\, \Delta \log C$.

The key empirical object is the causal effect of skills on output embedded in $\psi_{yh}$. A naïve regression,

$$\Delta \log y_r \;=\; \alpha \;+\; \beta\, \Delta \log h_r \;+\; \varepsilon_r,$$

is biased because unobserved technology shocks $\Delta \log \bar{A}_r$ raise both $\Delta \log y_r$ (by construction) and $\Delta \log h_r$ (via the supply equation through $\beta_y$), implying $\mathrm{Cov}(\Delta \log h_r, \varepsilon_r) \neq 0$.

Cost-shift instruments address this endogeneity. If changes in education cost are orthogonal to contemporaneous technology shocks—formally,

$$\mathrm{Cov}\big(\Delta \log C_r,\; \Delta \log \bar{A}_r\big) = 0,$$

then $\Delta \log C_r$ provides a valid instrument for $\Delta \log h_r$, shifting skill supply without directly entering the output equation. The resulting IV estimator consistently recovers the causal elasticity $\psi_{yh}$.

The following section outlines how a cost-shift instrument is constructed. India instituted a

series of higher education reforms between 1991 and 2000 by making it easier for new private colleges to open. However, the effect of reform was heterogeneous across districts due to the institutional requirement that new colleges be affiliated with an existing state public university. Because the cost of opening and operating new colleges is proportional to the distance from the affiliating university, I use the district's distance to its affiliating university as an instrument for the shift in the supply of college graduates. The next section provides the institutional details of higher education and reform in India.

# 4 Institutional Details of Higher Education Reform

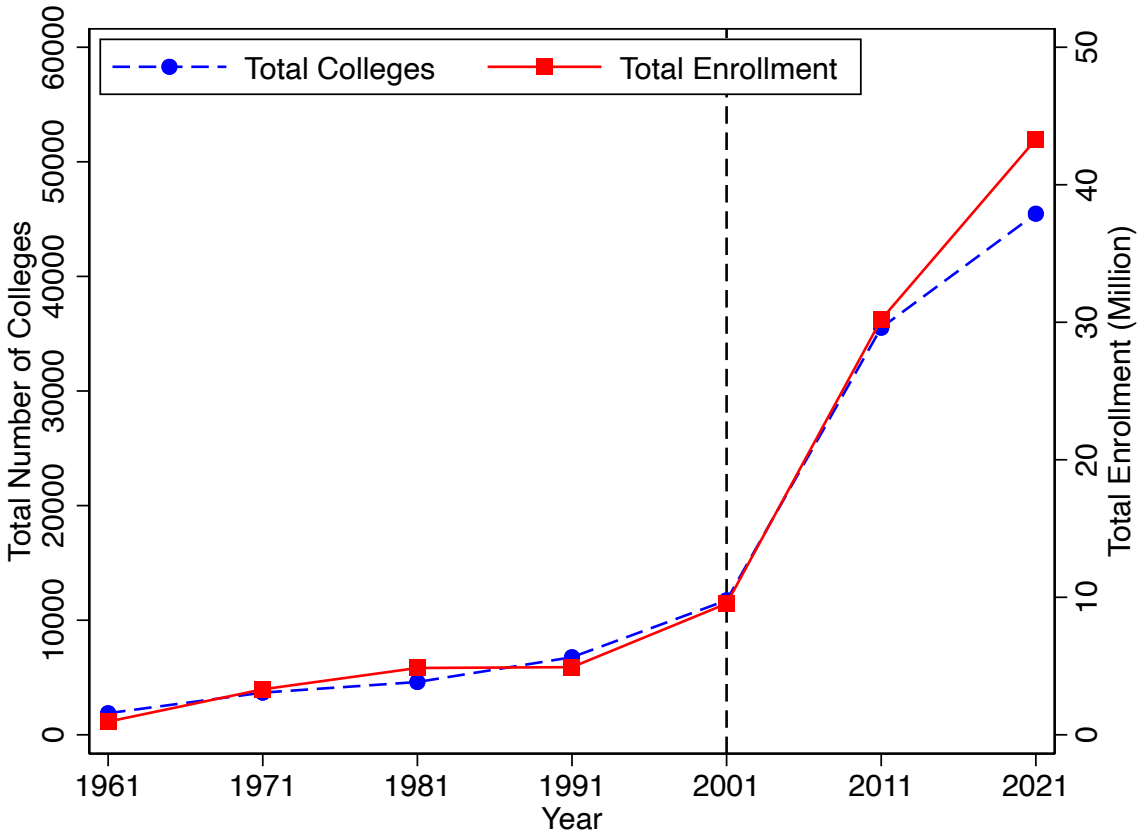



Figure 1: Growth of colleges and enrollment

The institutional structure of higher education in India is characterized by a dual-tier system, in which private colleges operate under the academic and administrative oversight of public universities. Such distinction between colleges and universities are absent in the United States where most enrollment in higher education is in the universities. In contrast, private colleges accounts for close to 70% of total enrollment in higher education with universities accounting for less than 20% of total enrollment in India.

Historically, higher education in India was dominated by a heavily regulated, state-controlled

system. The government exercised significant control over nearly all aspects of college operations, including faculty salaries, course offerings, teacher appointments, and infrastructure requirements, leaving little room for private sector participation. However, beginning in the late 1980s and accelerating through the 1990s, a significant policy shift began to unfold. This was not a single, discrete event but rather a series of legislative and policy signals that progressively dismantled the state-centric model and encouraged private investment in the sector.

The intellectual groundwork for this shift was laid by the National Policy on Education (1986) and its subsequent Programme of Action (1992), which were among the first official documents to formally encourage private sector participation in higher education as a means to expand access and capacity. This policy orientation gained further momentum with the Eighth Five-Year Plan (1992–1997), which explicitly recommended and planned for the involvement of the private sector to supplement public efforts in higher education. These foundational recommendations were translated into more concrete action by the end of the decade. The Special Action Plan for Education (1998) called for a comprehensive review of the specific laws and regulatory procedures that constrained private entry and expansion. It emphasized the need to remove bureaucratic hurdles and create a more enabling environment for non-state actors. This policy direction was reinforced at the highest levels of government when, in a 1999 joint session of Parliament, the President of India explicitly stated the government's commitment to encouraging private investment in higher and technical education. This address signaled a decisive departure from the earlier policy stance that had prioritized direct public provision and tight regulation. The culmination of this reform process was formalized in the Tenth Five-Year Plan (2002-2007). It consolidated the goals of the previous decade by recommending the simplification of laws, rules, and procedures governing the establishment and operation of private, cooperative, and not-for-profit educational institutions. The Plan also argued for dismantling restrictive controls on faculty salaries, infrastructure requirements, and staffing norms—policies that had historically constrained the growth and autonomy of private colleges.

Given this history, it is clear that the liberalization of India's higher education sector was a gradual process spanning more than a decade. For the purpose of the empirical analysis in this paper, which focuses on the subsequent expansion, I define the period up to 2001 as the "pre-reform" era, as this is when the critical mass of these foundational policy shifts was enacted, setting the stage for the rapid growth in private college entry that followed.

Together, these policy shifts laid the institutional foundation for a dramatic expansion of private higher education in India during the 2000s. Figure 1 shows the growth in number of college in India over time. As a consequence of the reforms, the number of colleges increased by approximately 4 times in the 20 years time period between 2001 and 2021.

## College Affiliation and Role of Distance to Affiliating University

The expansion of private colleges in India following the 1990s reforms was not uniform across regions. While the 1990s reforms relaxed many entry restrictions, the affiliation requirement remained intact as a key regulatory constraint. The affiliation system is a regulatory framework where every private college must obtain formal affiliation from a designated state public university, which acts as the sole degree-granting authority. This affiliation entails a binding relationship: the reputation of a college is tied to its university, degrees are issued in the university's name, curricular expansion requires university approval, and affiliated colleges pay recurring fees to the university.

Figure 2 illustrates how this system operates using the jurisdictional map for state public universities in Andhra Pradesh, a state in India with 23 district boundaries. In 1991, Andhra Pradesh had a population of 66 million people—about the size of France—with each district averaging around 3 million people, similar to Connecticut. On this map, each colored area shows the districts legally assigned to a single university, marked by a black dot. Districts shaded in the same color fall under the jurisdiction of the same university. For example, the pink region denotes the jurisdiction of Andhra University, which is headquartered in Visakhapatnam. Five districts—Visakhapatnam, Vizianagaram, Srikakulam, East Godavari, and West Godavari—are assigned to Andhra University, even though they are located at varying distances from the university.

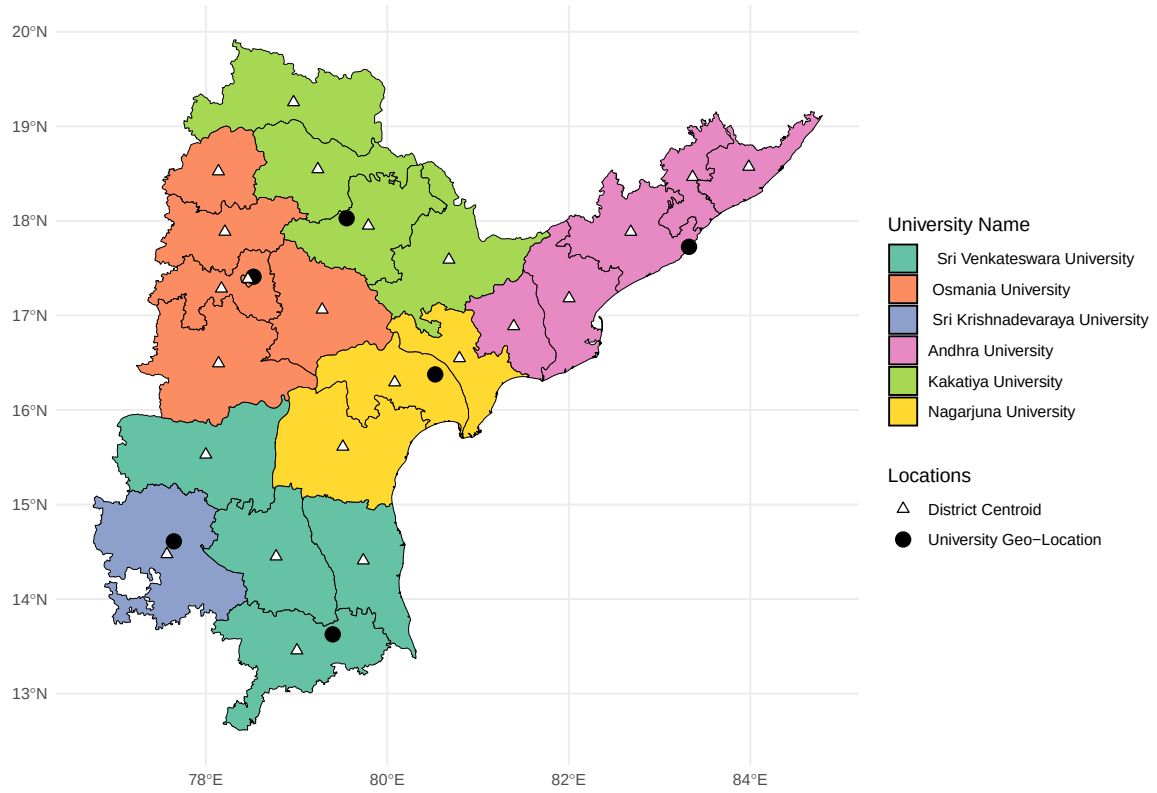



Figure 2: Territorial coverage of state public universities in Andhra Pradesh.

This rigid affiliation system imposed significant operational and administrative costs on colleges, and these costs were strongly proportional to the geographic distance from the affiliating university. By law, each district was assigned to a specific university, regardless of geographic proximity. Colleges could choose their district but had no choice over which university to affiliate with—they could not even apply to the geographically nearest university if it was not their designated affiliating institution. There are several reasons why distance to the affiliating university mattered for opening and operating costs. First, the centralized, paper-based examination system required the physical distribution of question papers from the university to all affiliated colleges, followed by the secure collection and transport of answer scripts back to centralized evaluation centers—logistical burdens that scaled directly with distance.[6] Second, throughout the sample period, affiliation workflows—No Objection Certificates (NOCs), inspections, approvals, and periodic renewals—required repeated, in-person submissions at university offices, and in an era of limited digital connectivity, greater distance magnified these transaction costs substantially. Third, the burden extended to students, who frequently had to travel to the affiliating university for administrative tasks such as filing examination forms, collecting transcripts, and resolving disputes, with excessive travel costs potentially depressing enrollment demand.[7] Fourth, from the university's perspective, monitoring a geographically dispersed network of colleges is costly, which often led them to confine new affiliations to a manageable radius to ensure better governance and safeguard academic standards. A detailed discussion of why distance to the affiliating university mattered for college operations and entry is provided in Appendix B.1.

The tangible impact of these distance-related costs is evident in college entry patterns. Figure 3 plots the histogram of the number of new colleges established after 2001 against their distance to the 1991 affiliating university. The histogram reveals a stark negative relationship: a vast majority of new colleges opened in close proximity to their affiliating university, with the number of entries dropping off sharply as distance increases.

Given this strong empirical relationship, I use the pre-determined distance to the affiliating university in the pre-reform era (1991) as a source of quasi-random variation in the entry of colleges and, consequently, in the growth of college graduates. Specifically, the instrument $D_r$ is defined as the log geographic distance between the entroid of district $r$ and the geolocation of its affiliating university as of 1991. A shorter $D_r$ lowers the administrative and monitoring costs of opening and running a college, predicting a larger increase in local educational capacity and, hence, a greater supply of graduates. To fix ideas, within Andhra University's jurisdiction (headquartered in Visakhapatnam), the corresponding centroid–university distances are $82.22$ km for Vizianagaram, $117.05$ km for Srikakulam, $152.70$ km for East Godavari, and $225.34$ km for West Godavari, illustrating substantial within-jurisdiction variation in cost-relevant dis-

[6] UGC (2019), *Evaluation Reforms in Higher Education Institutions in India*.

[7] Illustrative reports: *The Telegraph*, Ranchi, 2 Aug. 2004; *Hindustan Times*, 11 Jan. 2021; *Times of India*, Ahmedabad, 15 Jul. 2014.

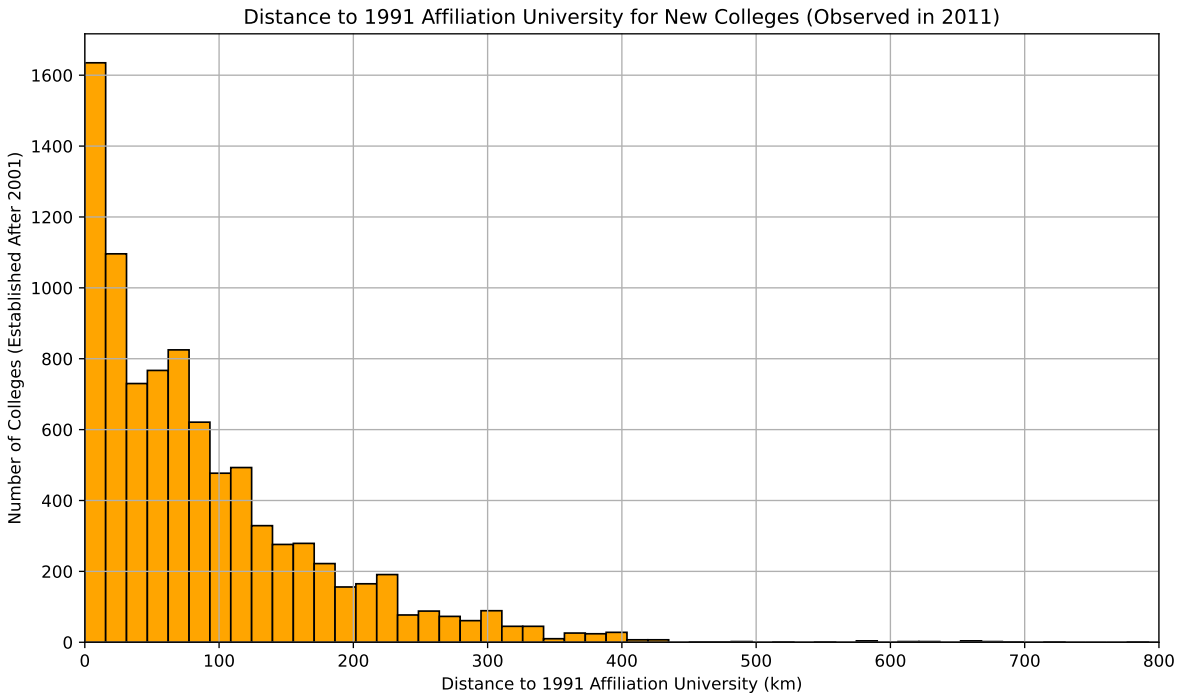



Figure 3: Number of New Colleges (Established After 2001) by Distance to 1991 Affiliation University.

tance.

In the next section, I formally test the hypothesis that pre-reform affiliating-university distance predicts post-reform changes in the district share of college graduates, and I report a series of validity checks that establish the exogeneity of $D_r$ with respect to non-education channels.

# 5 Empirical Specification and Identifying Assumptions

The main empirical specification relates changes in district-level outcomes to changes in human capital:

$$\Delta Y_d = \beta_0 + \beta \, \Delta h_d + \varepsilon_d \tag{1}$$

where $\Delta Y_d$ denotes the change in district $d$'s outcome between 2001 and 2011, and $\Delta h_d$ is the change in the share of college-educated workers. The outcomes of interest span three dimensions of economic development: *employment structure* (reallocation of labor out of agriculture and into services), *wage premium* (compression of the college wage premium), and *output* (increase in GDP per worker). The coefficient $\beta$ captures the causal effect of expanding higher education on local economic development. As argued in Section 3, ordinary least squares (OLS) estimation of equation (1) yields biased estimates. The fundamental problem is that $\Delta h_d$ is not randomly assigned across districts. Districts that experience larger increases in college graduates likely differ along multiple unobserved dimensions—industrial composition, political economy, local demand for skills—that independently affect economic outcomes. The

presence of omitted variables, reverse causality, and simultaneity bias makes it impossible to disentangle the causal effect of education from confounding factors using OLS. To address this endogeneity, I employ an instrumental variables (IV) strategy, using the log geographic distance from a district's centroid to its legally mandated affiliating university (as of 1991) as an instrument for $\Delta h_d$. The first-stage regression takes the form:

$$\Delta h_d = \alpha_0 + \alpha \log D_d + \nu_d \tag{2}$$

where $\log D_d$ is the instrument. The validity of this IV strategy rests on two conditions: relevance and exogeneity. The relevance condition requires that the instrument be sufficiently correlated with the endogenous variable, formally requiring that $\text{Cov}(\Delta h_d, \log D_d) \neq 0$. As demonstrated in Figure 3, there is a strong negative relationship between distance to the affiliating university and the number of new colleges established after 2001, suggesting that the instrument has substantial predictive power for changes in human capital. I formally test this condition in the first-stage regression presented in the following section, where I show that the F-statistic far exceeds conventional thresholds for weak instruments.

The exogeneity condition requires that the instrument satisfy the exclusion restriction: distance to the affiliating university should affect economic outcomes *only* through its effect on educational attainment, with no direct effect on output, employment, or wages. Formally, this requires that $\text{Cov}(\log D_d, \varepsilon_d | \Delta h_d) = 0$. This is the critical—and untestable—assumption underlying the IV strategy. While the exclusion restriction cannot be verified directly, I provide extensive empirical evidence to establish its plausibility.

To formalize the relevance and exogeneity conditions above in the potential-outcomes framework commonly used in applied IV work, let $\Delta Y_d(\Delta h)$ denote district $d$'s potential outcome change between 2001 and 2011 as a function of the change in human capital $\Delta h$, and let $\Delta h_d(z)$ denote the potential change in human capital when the instrument takes value $z = \log D_d$. Under the maintained assumptions below, the 2SLS estimand in (1)–(2) admits a causal interpretation.

**Assumption IV1 (Instrument relevance).** $\log D_d$ shifts $\Delta h_d$ on average: $\mathbf{E}[\Delta h_d(z') - \Delta h_d(z)] \neq 0$ for some $z' \neq z$. In the linear first stage (2), this corresponds to $\alpha \neq 0$ and a sufficiently large first-stage $F$-statistic.

**Assumption IV2 (Independence / as-good-as-random assignment).** The instrument is independent of potential outcomes and potential treatments:

$$\log D_d \;\perp\; \Big(\Delta Y_d(\cdot),\; \{\Delta h_d(z) : z \in \mathcal{Z}\}\Big),$$

possibly conditional on baseline covariates and fixed effects included in the specification. In the present setting, this assumption rules out the possibility that the affiliation-based

distance measure is systematically related to unobserved determinants of outcome growth between 2001 and 2011.

**Assumption IV3 (Exclusion restriction).** The instrument affects outcomes only through human capital. Formally, for any $z$ and any $\Delta h$,

$$\Delta Y_d(\Delta h, z) = \Delta Y_d(\Delta h),$$

so that $\log D_d$ has no direct effect on $\Delta Y_d$ other than via its impact on $\Delta h_d$.

**Assumption IV4 (Monotonicity / no defiers).** Higher distance to the affiliating university does not increase educational expansion for some districts while decreasing it for others. In potential-outcomes terms, for any $z' > z$,

$$\Delta h_d(z') \leq \Delta h_d(z) \quad \text{for all } d,$$

so the first stage is weakly monotone in $\log D_d$. This condition is standard for interpreting IV estimates as weighted averages of local causal responses when treatment intensity varies.

**Assumption IV5 (SUTVA / no interference).** A district's outcome change depends on its own change in human capital and not directly on other districts' instrument values (or, more generally, cross-district spillovers are not systematically related to $\log D_d$ in a way that violates IV2–IV3).

Consistent with Assumption IV1 (relevance), Figure 3 provides a transparent visual summary of the first-stage relationship. Districts located farther from their 1991 affiliating university experienced systematically smaller increases in post-2001 college entry. This also meant a smaller increase in share of college graduates. I formally test instrument relevance in the first-stage regression reported in Table 2, which presents the estimated coefficient on $\log D_d$ and the first-stage $F$-statistic. Figure 3 also supports Assumption IV4 (monotonicity). The number of new colleges established after 2001 declines monotonically with distance to the affiliating university, consistent with the interpretation that greater distance weakly reduces reform-induced educational expansion for all districts.

Given the evidence for relevance and monotonicity, the central identifying concern becomes instrument exogeneity, specifically Assumptions IV2 (independence) and IV3 (exclusion). The primary threat is that $\log D_d$ may be correlated with unobserved determinants of district-level growth between 2001 and 2011. For example, if universities were historically located in economically dynamic regions, then distance to the affiliating university could proxy for remoteness from markets, infrastructure, or state capacity, violating the independence condition. Alternatively, if districts farther from their affiliating university were already on different trajectories prior to the reform, then the post-2001 divergence could reflect pre-existing differential

trends rather than the causal effect of education, undermining the exclusion restriction as implemented in (1). I address these concerns through three complementary empirical strategies designed to probe the most salient violations of Assumptions IV2 and IV3.

The first strategy examines pre-reform parallel trends using an event-study specification. If distance to the affiliating university were capturing unobserved growth determinants, its effect should be visible not only after the reform but also before it. To test this, I estimate an event-study specification that examines whether districts located farther from their affiliating university experienced differential changes in college graduates, employment structure, and output during the pre-reform period. The event-study results show no significant relationship between distance and any outcome variable in the pre-reform period. Districts near and far from their affiliating universities exhibited parallel trends in educational attainment, sectoral employment shares, and economic output prior to liberalization. This rules out the possibility that distance was systematically correlated with pre-existing differential growth trajectories. The divergence in outcomes emerges sharply only after 2001, coinciding precisely with the post-reform surge in private college entry, which strongly suggests that distance affected development solely through its impact on post-reform educational expansion.

The second strategy implements a border discontinuity analysis that compares districts located on opposite sides of university jurisdiction boundaries. A threat to validity is that districts differ along many dimensions, and distance might be picking up some omitted characteristic correlated with both remoteness and economic potential. The border discontinuity design addresses this concern by exploiting sharp discontinuities in regulatory assignment at jurisdictional boundaries. The logic is straightforward: consider two districts separated by a jurisdictional boundary. Both lie in similar economic environments, but one is assigned to a nearby university while the other is assigned to a distant one. If distance matters only through its regulatory effect on college entry, then the district closer to its affiliating university should experience greater college expansion and, consequently, larger gains in human capital and economic outcomes. Importantly, in the pre-reform period, these border villages should exhibit no systematic differences, since the affiliation system was not binding when private college entry was heavily restricted.

The results strongly support the exclusion restriction. In the pre-reform period, districts on different sides of jurisdictional boundaries, despite facing different distances to their affiliating universities, show no significant differences in college graduates, agricultural employment share, service employment share, or wages. In sharp contrast, during the post-reform period (2001–2011), districts located closer to their affiliating university experienced significantly larger increases in the share of college graduates, greater reallocation of labor out of agriculture and into services, and faster growth in output per worker. The border discontinuity design thus provides compelling evidence that distance affects economic outcomes solely through its impact on educational expansion, not through any direct channel.

The third strategy implements a placebo test that exploits the institutional rigidity of the

affiliation system. If distance were merely proxying for general remoteness or lack of connectivity, then distance to the nearest university, regardless of affiliation, should also predict college growth. However, if the instrument captures the regulatory channel, then only distance to the legally mandated affiliating university should matter. I test this by including both distance to the affiliating university and distance to the nearest university in the regressions. The results show that distance to the affiliating university remains a strong and significant predictor of college entry, while distance to the nearest university has no predictive power whatsoever. This finding is particularly compelling because many districts are located closer to a university other than their designated affiliating institution, yet this geographic proximity confers no advantage. The stark difference between the two distance measures confirms that the instrument operates through the regulatory channel, via the binding affiliation requirement, rather than through general accessibility or economic geography. This placebo test significantly strengthens the case for exogeneity by ruling out alternative interpretations based on remoteness or connectivity.

Taken together, these three empirical strategies provide robust support for the validity of the distance instrument. The pre-reform event study rules out differential pre-trends. The border discontinuity design isolates the regulatory effect of distance by comparing otherwise similar districts. Finally, the placebo test confirms that it is distance to the *affiliating* university, not general remoteness, that matters. While no single test can definitively establish the exclusion restriction, the convergence of evidence across multiple identification strategies lends credibility to the claim that distance to the affiliating university affects economic outcomes exclusively through its impact on post-reform educational expansion. More detailed results from these robustness checks are presented in Section 7.

# 6 Main Results

This section presents the main empirical findings. I begin by formally testing the first stage of the instrumental variables strategy, as specified in equation (2). The first-stage analysis examines whether distance to the affiliating university is a strong predictor of changes in the share of college-educated workers across districts. I then report the second-stage results, which estimate the causal effect of changes in the share of college graduates on measures of economic development, including employment structure, wages, and output per worker.

## First Stage: Effect of Distance on Graduate Share

Table 2 reports the estimates of this first-stage equation. The analysis is conducted for both the full working-age population (Column 1) and, more specifically, for the young adult cohort aged 20–34 (Column 2), who would be the primary beneficiaries of increased college access during the 2001–2011 period. The results reveal a robustly negative and statistically significant relationship across both specifications. The coefficient on log-distance is $-0.017$ for the full

Table 2: First–Stage Regression: Effect of Distance on the Change in District Graduate Share

| | (1) All ages | (2) Young (20–34) |
|---|---|---|
| log-Dist | -0.017*** | -0.028*** |
| | (0.003) | (0.005) |
| Constant | 0.124*** | 0.206*** |
| | (0.014) | (0.027) |
| Obs. | 374 | 374 |
| F-Stat | 35.39 | 27.37 |

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Notes:* Each column reports the first-stage regression of the change in a district's share of college graduates on the logarithm of its distance to the affiliating university.

population (Column 1) and $-0.028$ for the young cohort (Column 2). As expected, the effect is larger in magnitude for the younger demographic most likely to be on the margin of college entry. To interpret the economic significance, the coefficient in Column (1) implies that a doubling of distance from the affiliating university (e.g., from 100 km to 200 km) reduces the growth in the graduate share by 1.2 percentage points ($-0.017 \times \ln(2)$). For the younger cohort, the effect is a more pronounced 1.9 percentage point reduction.

Crucially, the first-stage F-statistics are large across both specifications, 35.39 and 27.37. These values are well above the conventional Staiger–Stock–Yogo rule-of-thumb threshold of 10, indicating that the log-distance instrument is a strong predictor of changes in the graduate share and is not susceptible to weak instrument bias.

For the main second-stage analysis throughout the remainder of the paper, I adopt the specification from Column (1) as the preferred first stage. This choice is motivated by two considerations. First, not all aggregate outcome variables used in the second stage are available in a form that can be disaggregated by age, so focusing on the full working age population ensures consistency and comparability across specifications. Second, I am able to run the same first stage regression inside my structural model, which ensures consistency between the reduced-form and structural estimation approaches.

Taken together, these results confirm that distance to affiliating university provides a powerful and plausibly exogenous source of variation in the supply of college graduates, which I exploit in the subsequent causal analysis.

## Second Stage: Effect on sectoral employment shares

Table 3 presents the main second-stage instrumental variables estimates of the effect of increasing the share of college graduates on economic outcomes. Each column reports the coefficient on $\Delta h$, the instrumented change in the share of college-educated workers, from a separate

regression. Columns (1)-(3) examine the reallocation of labor across sectors, measuring the change in employment shares in agriculture, manufacturing, and services, respectively. Column (4) estimates the effect on output, measured as the change in log GDP per worker. Column (5) examines the effect on the college wage premium, measured as the change in the log ratio of wages of college graduates to non-graduates.

Table 3: IV Estimates: Effect of Graduate Share on Economic Outcomes

| | (1) $\Delta s$ (Agri.) | (2) $\Delta s$ (Manu.) | (3) $\Delta s$ (Serv.) | (4) $\Delta$ log(GDPpw) | (5) $\Delta$ log(NL) | (6) $\Delta\, log(w_H/w_L)$ |
|---|---|---|---|---|---|---|
| $\Delta h$ | -0.774*** | -0.115 | 0.894*** | 2.262* | 6.173** | -4.938* |
| | (0.196) | (0.108) | (0.175) | (1.200) | (2.583) | (2.821) |
| Constant | 0.008 | -0.009** | 0.002 | 0.462*** | 0.179 | 0.110 |
| | (0.009) | (0.005) | (0.008) | (0.054) | (0.118) | (0.134) |
| Obs. | 374 | 374 | 374 | 374 | 358 | 357 |
| Eff. OP F | 35.39 | 35.39 | 35.39 | 35.39 | 28.67 | 34.38 |

*Notes:* Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Effective F-stat is the Montiel-Olea and Pflueger (2013) effective first-stage $F$-statistic.

The results reveal substantial and economically meaningful effects of educational expansion on structural transformation and economic development. The first three columns examine the reallocation of labor across sectors. Beginning with sectoral employment, agriculture exhibits a large and statistically significant negative response: a one-percentage-point increase in the graduate share reduces the agricultural employment share by 0.77 percentage points. Manufacturing shows a negative point estimate of $-0.12$ that is not statistically distinguishable from zero. In contrast, services absorb almost exactly these outflows: the coefficient of 0.89 is positive, highly significant. These results indicate that the dominant reallocation triggered by higher graduate supply is a shift of labor out of agriculture and into services, with little systematic impact on manufacturing.

Turning to output effects, Column (4) shows that a one-percentage-point increase in the graduate share raises GDP per worker by approximately 2.3%. To address potential concerns about GDP mismeasurement, Column (5) presents results using nightlight intensity from DMSP-OLS satellite data as an alternative proxy for economic activity. The coefficient of 6.46 indicates that a one-percentage-point increase in the graduate share raises nightlight intensity by approximately 6.5%. To benchmark this estimate against the GDP per worker results, I draw on elasticities from the nightlight-GDP literature. Henderson et al. [2012] estimate that a 1% increase in nightlight intensity is associated with a 0.275% increase in GDP. Using this elasticity, the nightlight coefficient implies an effect on GDP per worker of approximately 1.8% ($6.17 \times 0.275 = 1.78$), reassuringly close to the direct estimate of 2.3%. More recent estimates by Hu and Yao [2022] find a higher elasticity of 0.77, which would imply an effect on GDP per worker of approximately 5.0% ($6.46 \times 0.77 = 4.97$). The nightlight results thus provide corroborating evidence for substantial output gains from educational expansion, with implied

magnitudes ranging from close alignment with the direct GDP estimates to potentially larger effects.

Finally, Column (6) examines the effect on the college wage premium. The coefficient of $-4.9$ indicates that a one-percentage-point increase in the graduate share reduces the college wage premium by approximately 5%. The negative sign is consistent with a standard supply-side interpretation: as the supply of college-educated workers increases, their relative wage falls, compressing the skill premium.

The effective F-statistics, reported at the bottom of Table 3, range from 28.89 to 35.39, comfortably exceeding the conventional threshold of 10 and indicating that weak instrument concerns are not an issue. Taken together, these results provide robust evidence that the expansion of higher education following the 1990s reforms generated substantial effects on structural transformation and output growth. Districts that experienced larger increases in the supply of college graduates—driven by proximity to their affiliating university—reallocated labor away from agriculture and toward services, experienced faster growth in output per worker and night-light intensity, and saw a compression of the college wage premium.

Table 4: Mean Characteristics by GDP per Worker Quartile (2011)

| | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| GDP per worker | 37081 | 62662 | 98866 | 218922 |
| Graduate share | 0.09 | 0.11 | 0.14 | 0.22 |
| Agri. share | 0.76 | 0.71 | 0.65 | 0.51 |
| Service share | 0.18 | 0.21 | 0.27 | 0.37 |
| $\log(w_H/w_L)$ | 1.26 | 1.15 | 1.09 | 1.03 |

*Notes:* Quartiles based on GDP per worker in 2011.

To provide context for the magnitudes of these effects, Table 4 presents the cross-sectional distribution of key variables across districts sorted by GDP per worker in 2011. The table reveals strong correlations between development levels and both educational attainment and sectoral structure. Moving from the poorest quartile (Q1) to the richest quartile (Q4), the graduate share more than doubles from 9% to 22%. This increase in educational attainment is accompanied by substantial structural transformation: the agricultural employment share declines from 76% to 51%, while the service share rises from 18% to 37%. The college wage premium also exhibits a strong negative gradient, falling from 1.26 log points in Q1 to 1.03 log points in Q4. These cross-sectional patterns are consistent with the causal effects identified in the IV estimates, but raise the question of how much of the observed variation in development outcomes can be attributed to differences in educational attainment. To address this question, I use the IV estimates to calculate the predicted effects of observed differences in graduate shares and compare them to the actual gaps between quartiles.

To interpret the economic magnitudes of these effects, I compare the predicted impacts from the IV estimates to the observed cross-sectional differences between districts. Table 5

Table 5: Predicted vs. Observed Differences: Q1 to Q2 Districts

| Outcome | Q1 | Q2 | Observed Difference | Predicted Effect | Share Explained |
|---|---|---|---|---|---|
| *Panel A: Employment Shares (pp)* | | | | | |
| Agriculture | 0.76 | 0.71 | $-5.0$ | $-1.5$ | 31% |
| Services | 0.18 | 0.21 | 3.0 | 1.8 | 60% |
| *Panel B: Output and Wages (%)* | | | | | |
| GDP per worker | 37,081 | 62,662 | 52.5 | 4.5 | 8.6% |
| College wage premium (log) | 1.26 | 1.15 | $-11.0$ | $-9.9$ | 90% |

*Notes:* Q1 and Q2 show mean values from Table 4. Graduate share increases from 0.09 (Q1) to 0.11 (Q2). Observed difference is Q2 minus Q1. Predicted effect calculated by multiplying IV coefficients from Table 3 by the 2 percentage-point increase in graduate share. Employment share differences measured in percentage points (pp). GDP per worker and college wage premium differences measured in percent (%). Share explained is predicted effect divided by observed difference.

presents this comparison using districts in the first and second quartiles of the GDP per worker distribution (from Table 4). Q1 districts have an average graduate share of 9%, while Q2 districts have 11%—a 2 percentage-point difference.

The results show that educational expansion accounts for a substantial portion of cross-sectional variation in sectoral structure. The predicted decline in agricultural employment (1.5 percentage points) explains 31% of the observed 5 percentage-point gap between Q1 and Q2 districts. For services, the predicted increase (1.8 percentage points) accounts for 60% of the observed 3 percentage-point gap, indicating that differences in educational attainment play a major role in explaining why more developed districts have larger service sectors.

For output, the predicted 4.5% increase in GDP per worker explains 8.6% of the 52.5% difference between Q1 and Q2 districts. While this may appear modest, it suggests that educational expansion is one of several important factors driving cross-sectional income differences across Indian districts.

The wage premium results are particularly striking: the predicted 9.0% compression accounts for 90% the observed 11% gap in the college wage premium between Q1 and Q2 districts. This suggests that differences in the supply of college graduates are the dominant factor explaining cross-sectional variation in wage inequality, highlighting the powerful equilibrium effects of educational expansion on local labor markets.

## Comparision with OLS Estimates

Table 6 presents ordinary least squares (OLS) estimates of the relationship between changes in college graduate share and changes in economic outcomes. These estimates do not account for the endogeneity of educational expansion and therefore should be interpreted as correlations rather than causal effects. The results show that a one-percentage-point increase in the graduate

share is associated with a 0.57 percentage point decline in agricultural employment share, a 0.60 percentage point increase in service sector employment share, and a 1.5 percent increase in GDP per worker. However, the coefficient on manufacturing employment is essentially zero and insignificant, and the estimates for nightlights and skill premium are imprecisely estimated.

Notably, these OLS estimates are smaller in magnitude than the IV estimates presented later. This pattern where OLS coefficients are attenuated relative to IV estimates is not unusual in the returns to schooling literature. Card [1999] documents this regularity across approximately ten studies using various identification strategies, finding that IV estimates consistently exceed OLS estimates by 20-40 percent. More recently, Gethin [2025] compiles a comprehensive meta-analysis of 33 studies covering 23 countries, confirming that IV estimates of returns to schooling systematically exceed OLS estimates across diverse contexts and identification strategies.

This pattern can be explained by the nature of IV estimates as Local Average Treatment Effects (LATE). Our instrument based on distance to the nearest university identifies returns for the marginal individuals whose college attendance decisions are affected by proximity to higher education institutions. These compliers are likely individuals who face binding constraints and stand to gain substantially from college access. In contrast, OLS estimates capture average returns including for those that have lower marginal returns. The IV estimates thus represent returns for a policy-relevant subpopulation: those whose educational attainment can be shifted by changes in college accessibility. This interpretation is consistent with the fact that distance-based instruments affect college attendance primarily among students from less advantaged backgrounds, who typically have higher returns to education [Card, 1995, Carneiro et al., 2011].

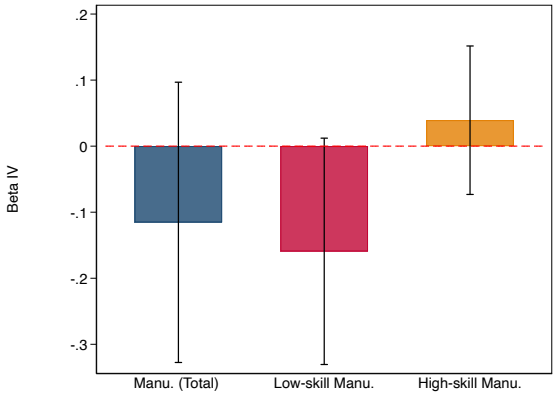
Table 6: OLS Estimates: Effect of Graduate Share on Economic Outcomes

| | (1) $\Delta s$ (Agri.) | (2) $\Delta s$ (Manu.) | (3) $\Delta s$ (Serv.) | (4) $\Delta$ log(GDPpw) | (5) $\Delta$ log(NL) | (6) $\Delta \log(w_H/w_L)$ |
|---|---|---|---|---|---|---|
| $\Delta h$ | -0.568*** | -0.001 | 0.595*** | 1.519*** | 0.854 | -0.978 |
| | (0.085) | (0.044) | (0.064) | (0.325) | (0.525) | (0.872) |
| Constant | -0.001 | -0.015*** | 0.016*** | 0.496*** | 0.417*** | 0.022 |
| | (0.004) | (0.002) | (0.003) | (0.019) | (0.037) | (0.052) |
| Observations | 374 | 374 | 374 | 374 | 358 | 347 |
| R-squared | 0.23 | 0.00 | 0.29 | 0.06 | 0.00 | 0.00 |

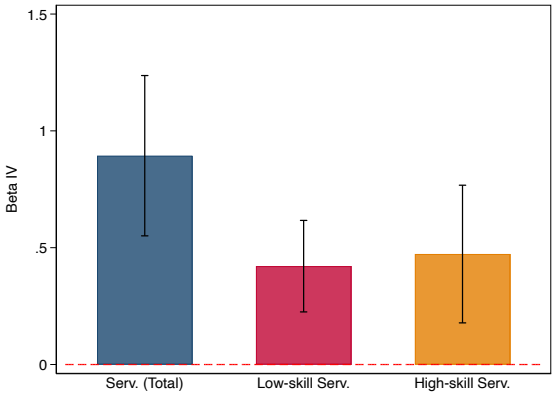
*Notes:* Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

## Additional Results

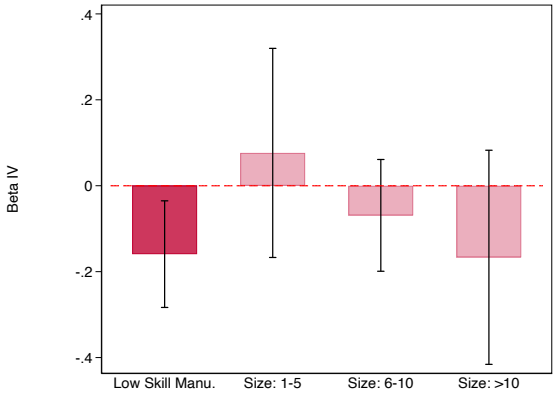
The baseline results document substantial effects of higher-education expansion on sectoral employment shares: a one-percentage-point increase in college attainment shifts employment by 0.89 percentage points from agriculture into services. However, these aggregate shifts obscure important questions about the nature of structural transformation. In particular, does higher



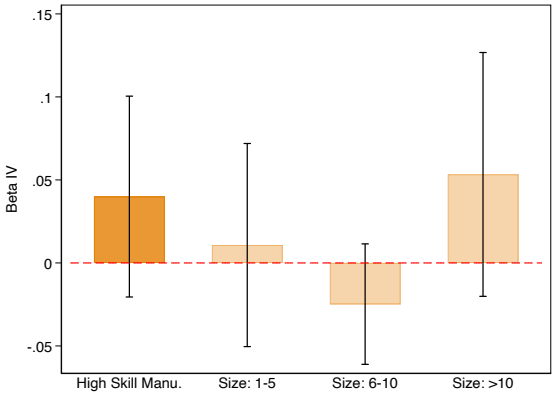

(a) Manufacturing: Overall and by Skill



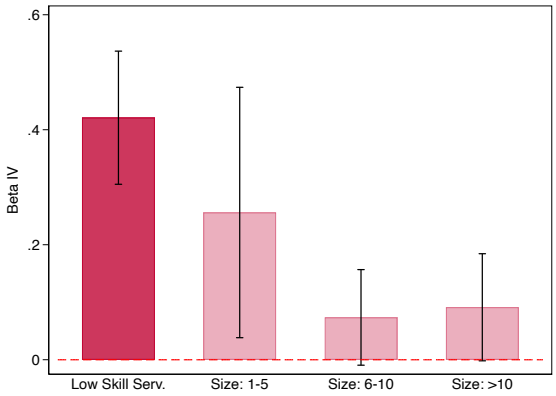

(b) Services: Overall and by Skill



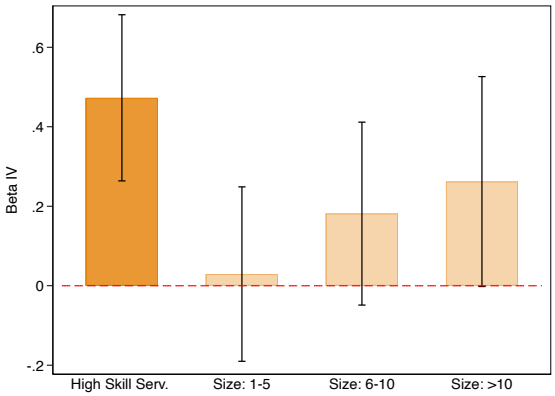

(c) Low-Skill Manufacturing by Size




(d) High-Skill Manufacturing by Size




(e) Low-Skill Services by Size




(f) High-Skill Services by Size

Figure 4: Employment Share Semi-Elasticities: Heterogeneity by Sector, Skill, and Establishment Size

*Notes:* Each panel displays IV estimates of employment share semi-elasticities with respect to distance to the affiliating university. Panels (a) and (b) decompose total effects by skill level. Panels (c)-(f) further disaggregate by establishment size categories: small (1-5 workers), medium (6-10 workers), and large (¿10 workers). Error bars represent 95% confidence intervals using robust standard errors. The dashed red line indicates zero. All regressions use log distance to the nearest university as an instrument for change in education share.

education primarily facilitate entry into skill-intensive activities within expanding sectors, or does it generate more diffuse employment growth across the skill distribution? Similarly, does labor reallocation occur predominantly through expansion of large, formal establishments or through proliferation of smaller enterprises?

These questions are central to understanding the mechanisms through which human capital drives development. Recent evidence suggests that economic development is closely associated with shifts in the establishment size distribution, with employment reallocating toward larger and more productive firms [Engbom et al., 2024]. If higher-education expansion generates employment primarily in large establishments within skill-intensive sectors, it would indicate that human capital facilitates modernization and formalization of economic activity.

To investigate these mechanisms, I decompose the employment share responses by sectoral skill intensity and establishment size. Figure 4 presents IV estimates using the same specification as the baseline results, with separate regressions for high-skill and low-skill employment within manufacturing and services, and further disaggregation by establishment size categories.

The aggregate service sector expansion masks considerable heterogeneity by skill level. The total service employment share increases by 0.89 percentage points, with 0.42 percentage points attributable to low-skill service employment and 0.47 percentage points to high-skill services. Both effects are statistically significant at the 1 percent level, with high-skill services accounting for a slightly larger share (53 percent) of the total expansion. This balanced distribution suggests that service sector growth encompasses both skill-intensive modern services and traditional activities.

In contrast, manufacturing exhibits a modest overall decline in employment share of 0.12 percentage points, driven primarily by a large and statistically significant contraction in low-skill manufacturing (-0.16 percentage points, significant at the 5 percent level) that is only partially offset by a small and statistically insignificant increase in high-skill manufacturing (0.04 percentage points).

Panels (c)-(f) reveal striking differences in adjustment patterns across establishment size categories. For high-skill services, the expansion is strongly concentrated in large establishments ($> 10$ workers), which account for 0.26 percentage points of the 0.47 percentage point total effect—approximately 55 percent of the high-skill service response (panel f, significant at 10 percent level). Medium-sized establishments (6-10 workers) contribute 0.18 percentage points (not statistically significant), while small establishments (1-5 workers) show a negligible and insignificant effect of 0.03 percentage points. This pronounced size gradient suggests that larger firms possess substantially greater capacity to absorb increases in skilled labor supply, potentially reflecting scale economies, better access to credit markets, or organizational capabilities that facilitate expansion.

The establishment size patterns differ markedly for low-skill services (panel e). Here, the expansion is more evenly distributed across establishment sizes, though small establishments show the largest response at 0.26 percentage points (significant at the 5 percent level). Large

establishments contribute 0.09 percentage points and medium establishments 0.07 percentage points (both marginally significant at the 10 percent level). This more balanced pattern—with small establishments still playing an important role—suggests that low-skill service activities expand through both proliferation of small-scale enterprises and growth of existing establishments. For manufacturing (panels c and d), the estimates across size categories are uniformly small and statistically insignificant, though point estimates suggest the low-skill manufacturing decline is most pronounced in large establishments.

These findings have important implications for understanding how local labor markets adjust to increases in the supply of skill. The stark contrast between high-skill and low-skill services is particularly revealing: high-skill service employment expands primarily through large establishments, suggesting that skilled labor supply increases generate opportunities in firm types characterized by higher productivity and formality. In contrast, low-skill service expansion occurs more diffusely across the establishment size distribution, with small firms playing a substantial role. This dual pattern of service sector growth likely reflects complementary mechanisms operating through both labor supply and demand channels. The expansion of high-skill service employment in large, productive establishments generates income growth among skilled workers, which in turn stimulates local demand for low-skill services—from retail and personal services to food preparation and domestic work. This demand-side channel provides a natural explanation for the observed low-skill service expansion: as educated workers move into higher-paying occupations, their increased purchasing power supports a parallel expansion of traditional service activities that cater to local consumption. This pattern points to the dual nature of service sector growth in developing economies, encompassing both formal sector expansion in skill-intensive activities and continued proliferation of small-scale, potentially informal enterprises in traditional services.

# 7 Robustness and Instrument Validity

The validity of the instrumental variables estimates rests critically on the assumption that distance to the affiliating university affects economic outcomes exclusively through its impact on educational expansion. While this exclusion restriction cannot be tested directly, I provide extensive empirical evidence to establish its plausibility through four complementary strategies. First, I estimate an event-study specification to verify that distance had no effect on outcomes during the pre-reform period, ruling out pre-existing differential trends. Second, I implement a border discontinuity design comparing districts on opposite sides of university jurisdiction boundaries to isolate the regulatory effect of distance from general remoteness. Third, I test the stability of the estimates to the inclusion of a rich set of baseline controls. Fourth, I conduct a placebo test showing that distance to the nearest university (as opposed to the legally mandated affiliating university) has no predictive power for college growth. Taken together, these checks provide robust support for the validity of the distance instrument.

## 7.1 Event Study: Pre-Reform Parallel Trends

The first validation strategy examines whether districts located at different distances from their affiliating university exhibited parallel trends in outcomes prior to the 1991 reforms. If distance were correlated with unobserved determinants of economic growth—such as proximity to markets, infrastructure quality, or local institutional capacity—then its effect should be evident not only after the reform but also before it. The event-study specification tests this directly by estimating the relationship between distance and outcomes separately for each decade.

The specification is:

$$Y_{dt} = \sum_{\tau \neq 2001} \beta_\tau (\log D_d \times \mathbf{1}[t = \tau]) + \alpha_d + \lambda_t + \varepsilon_{dt} \tag{3}$$

where $Y_{dt}$ is the outcome variable (share of sectoral employment shares, or log GDP per worker) in district $d$ at time $t$, $\log D_d$ is the log distance to the affiliating university (fixed as of 1991), $\mathbf{1}[t = \tau]$ is an indicator for year $\tau$, and $\alpha_d$ and $\lambda_t$ are district and year fixed effects, respectively. The year 2001 serves as the omitted baseline period, so all coefficients are estimated relative to 2001.

The coefficient $\beta_\tau$ identifies how the effect of log distance on outcomes changes over time, comparing the trajectories of districts located at different distances from their affiliating university, while partialling out the common time component across all districts (via $\lambda_t$) and the constant component of each district over time (via $\alpha_d$). In other words, $\beta_\tau$ captures the differential evolution of outcomes across districts with varying distances to their affiliating university, net of aggregate time trends and time-invariant district characteristics.

The inclusion of district fixed effects $\alpha_d$ absorbs all time-invariant characteristics of districts—such as geographic features, historical institutions, cultural factors, and persistent economic advantages. The year fixed effects $\lambda_t$ control for common aggregate shocks that affect all districts uniformly in a given year, such as national economic reforms, macroeconomic conditions, changes in trade policy, or nationwide technological progress.

Interpreting $\beta_{2011}$ as the causal effect of the reform operating through differential exposure by baseline distance relies on the following assumptions, which are standard in the event-study and continuous-treatment DiD literature.

**Assumption 1 (Predetermined exposure).** The exposure measure $\log D_d$ is predetermined with respect to the reform and is not itself affected by the reform. In practice, distance is constructed using the pre-reform affiliating-university network and treated as time-invariant.

**Assumption 2 (Generalized parallel trends in exposure).** In the absence of the reform, counterfactual changes in outcomes would not vary systematically with baseline distance:

$$\mathbf{E}[Y_{dt}(0) - Y_{ds}(0) \mid \log D_d] \;=\; \mathbf{E}[Y_{dt}(0) - Y_{ds}(0)] \quad \text{for } t, s \in \{1981, 1991, 2001, 2011\}, \tag{4}$$

where $Y_{dt}(0)$ denotes the potential outcome absent the reform. This restriction is the continuous-exposure analogue of DiD parallel trends: after netting out additive district and year components (absorbed by $\alpha_d$ and $\lambda_t$), untreated outcome changes are mean-independent of $\log D_d$.

**Assumption 3 (No anticipation).** The reform does not affect outcomes prior to 2001. In the event-study parameterization, this implies that the pre-period coefficients (1981 and 1991, relative to 2001) should be close to zero.

**Assumption 4 (Stable exposure mapping / limited interference).** A district's potential outcomes depend on the reform only through its own exposure $\log D_d$ (or, more generally, any spillovers do not generate systematic differential trends correlated with $\log D_d$ that would violate Assumption (4)).

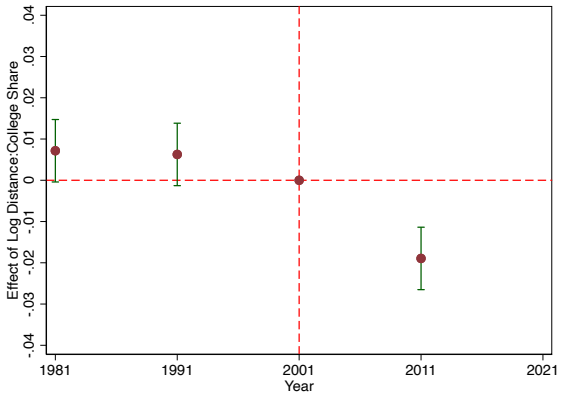
The key test is whether the pre-reform coefficients $\beta_\tau$ and $\tau < 2001$ are statistically indistinguishable from zero. Non-zero pre-reform coefficients would indicate that districts at different distances from their affiliating university were already on divergent trajectories before the major college expansion period, undermining the validity of distance as an instrument. In contrast, the post-reform coefficient $\beta_{2011}$ should be significant and of the expected sign if distance affected outcomes through the educational expansion channel that intensified after 2001.

Figure 5 presents the event-study estimates for six key outcomes: graduate share, agricultural employment share, service employment share, GDP per worker, nightlight intensity, and the college wage premium. Each panel plots the estimated coefficients $\beta_\tau$ with 95% confidence intervals. For most outcomes, 2001 is normalized to zero (indicated by the vertical dashed red line). For nightlight intensity, 1997 serves as the normalization year. To reduce yearly noise in the data, I use five-year forward means for GDP per worker and nightlight intensity.

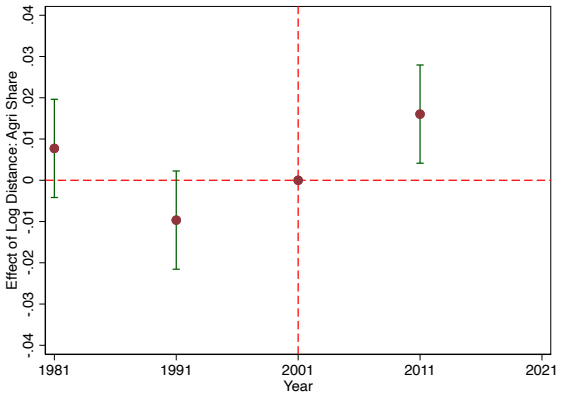
The results provide strong support for the parallel trends assumption. Across all six outcomes, the pre-reform coefficients are small in magnitude and statistically indistinguishable from zero where data is available. This pattern is evident in each panel of Figure 5.

Panel (a) presents the results for graduate share. The pre-reform coefficients for 1981 and 1991 are economically small and statistically insignificant, confirming that districts at different distances from their affiliating university had similar levels of educational attainment before the reform. The coefficient turns significantly negative in 2011, indicating that districts farther from their affiliating university experienced slower growth in graduate share during the post-reform expansion period.

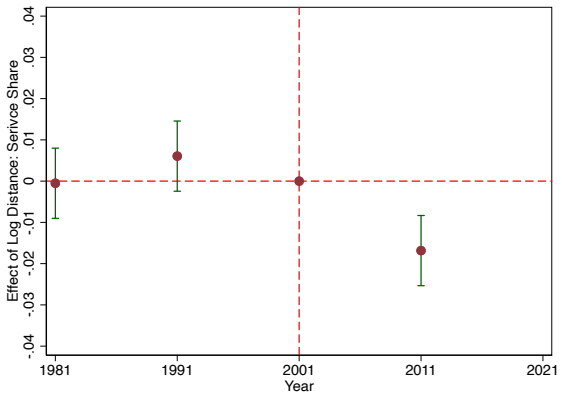
Panels (b) and (c) show the results for sectoral employment shares. For agriculture (panel b), all pre-reform coefficients—spanning two decades from 1981 to 1991—are tightly clustered around zero and statistically insignificant. Similarly, for services (panel c), the pre-reform coefficients show no systematic relationship with distance. These patterns rule out the possibility



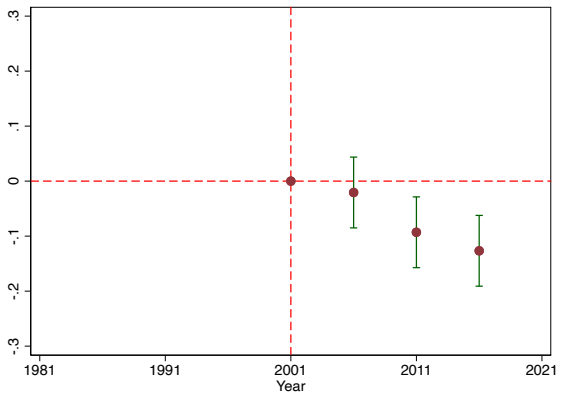

(a) Graduate Share



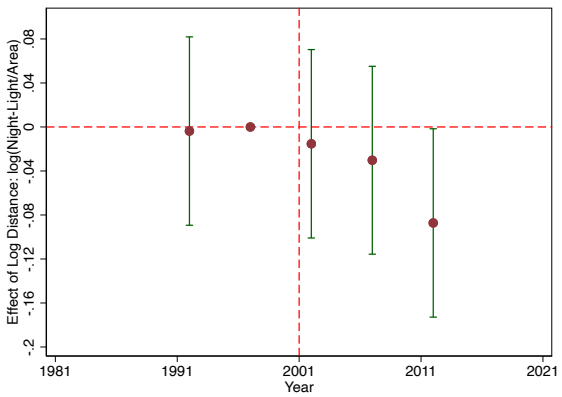

(b) Agricultural Employment Share



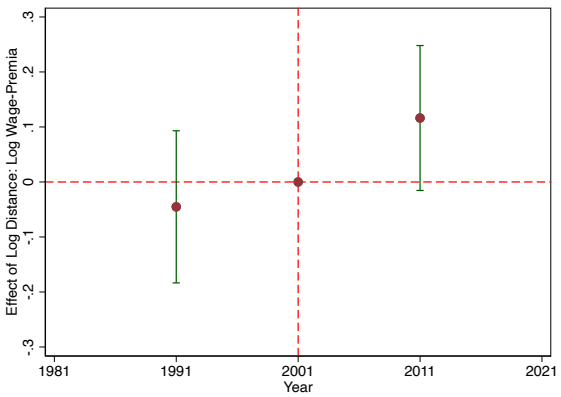

(c) Service Employment Share




(d) Log GDP per Worker




(e) Log Nightlight per Area




(f) Log College Wage Premium

Figure 5: Event Study: Effect of Log Distance to Affiliating University

*Notes:* Each panel plots coefficients from equation (3) with 95% confidence intervals. The year 2001 is normalized to zero (vertical dashed line). The outcome variables are: (a) share of college graduates, (b) agricultural employment share, (c) service employment share, (d) log GDP per worker, (e) log nightlight intensity per area, and (f) log college wage premium. All specifications include district and year fixed effects.

that distance was correlated with pre-existing trends in structural transformation. The coefficients only become economically meaningful and statistically significant in 2011, coinciding with the post-reform period when educational expansion accelerated.

Panel (d) examines GDP per worker, for which data is available starting from 2001. The coefficient is normalized to zero in 2001 and shows a clear negative trend in subsequent periods. By 2011, the coefficient is negative and statistically significant, indicating that districts farther from their affiliating university experienced slower GDP growth. This effect persists through later periods, demonstrating the sustained impact of educational expansion on productivity.

Panel (e) presents results for nightlight intensity per area, an alternative measure of economic activity that addresses potential concerns about GDP mismeasurement. With 1997 as the normalization year, the pre-reform coefficient for 1991 is close to zero and statistically insignificant. The coefficients turn increasingly negative in the post-reform period, with significant effects emerging by 2011. This pattern corroborates the GDP results and provides independent evidence of the economic gains from educational expansion.

Panel (f) shows the effect on the college wage premium. The pre-reform coefficient for 1991 is close to zero and statistically insignificant, suggesting no relationship between distance and wage inequality before the reforms. However, the coefficient turns positive and significant in 2011, indicating that districts farther from their affiliating university (which received fewer college graduates) experienced smaller reductions in the college wage premium, consistent with supply-driven wage compression in areas with greater educational expansion.

If distance were proxying for unobserved factors affecting economic development—such as remoteness from markets, poor infrastructure, or weak institutions—we would expect to see some evidence of differential trends even before the reform. The fact that we observe no such patterns across two decades and six distinct outcomes strongly suggests that distance to the affiliating university was uncorrelated with pre-existing growth trajectories.

Moreover, the timing of the divergence is precisely what we would expect if the instrument operates through the educational expansion channel. The coefficients remain near zero throughout the pre-reform period and shift sharply only after 2001, when private college entry surged following the relaxation of licensing restrictions. This temporal pattern—parallel trends before the reform, followed by sharp divergence afterward—provides evidence that distance affected economic outcomes solely through its impact on post-reform educational expansion, not through any direct effect on development.

The event-study specification also provides an alternative method for estimating the semi-elasticities of economic outcomes with respect to changes in educational attainment. The relationship between the event-study coefficients and the instrumental variables estimates is straightforward. Recall that the IV coefficient is given by:

$$\beta^{IV} = \frac{\text{Cov}(\Delta y, \log D)}{\text{Cov}(\Delta h, \log D)} \tag{5}$$

where $\Delta y$ is the change in the outcome variable and $\Delta h$ is the change in graduate share. The event-study specification in equation (3) provides separate estimates of $\text{Cov}(\Delta y, \log D)/\text{Var}(\log D)$ for the outcome variable and $\text{Cov}(\Delta h, \log D)/\text{Var}(\log D)$ for educational attainment. Taking the ratio of these two coefficients recovers the IV estimate:

$$\frac{\text{Cov}(\Delta y, \log D)/\text{Var}(\log D)}{\text{Cov}(\Delta h, \log D)/\text{Var}(\log D)} = \frac{\text{Cov}(\Delta y, \log D)}{\text{Cov}(\Delta h, \log D)} = \beta^{IV} \tag{6}$$

This implies that dividing the 2011 event-study coefficient for any outcome variable by the 2011 event-study coefficient for graduate share yields estimates that is algebraically equivalent to the two-stage least squares estimate. While the event-study-based estimated elasticities are mathematically equivalent to the IV estimates, I adopt the IV specification as my main approach for two reasons. First, it allows me to obtain standard errors on the estimated coefficients directly through the two-stage least squares procedure, rather than requiring additional calculations to derive standard errors from the ratio of event-study coefficients. Second, several outcome variables examined in the additional results—such as detailed occupational categories, firm-level outcomes, and various measures of economic activity—lack historical data necessary for event study analysis, making the IV approach the only feasible method for these variables. This consistency between the two methods reinforces the interpretation that the estimated effects capture the causal impact of educational expansion on structural transformation, output growth, and wage compression, rather than spurious correlations driven by omitted variables or measurement error.

## 7.2 Border Discontinuity Analysis

The second validation strategy implements a border discontinuity design that compares pairs of districts located on opposite sides of university jurisdiction boundaries. This approach addresses the concern that districts may differ along many unobserved dimensions, and that distance to the affiliating university might be correlated with omitted characteristics affecting economic development. By focusing on contiguous districts that share a border but are assigned to different universities, I compare observationally similar districts that face different regulatory distances to their respective affiliating institutions.

The unit of observation in this analysis is a pair of contiguous districts. I construct the sample by identifying all pairs of districts that share a geographic boundary but belong to different university jurisdictions. For each pair, I designate the district farther from its affiliating university as *Far* and the district closer to its affiliating university as *Near*. The key identifying variation comes from the difference in log distance to the affiliating university within each pair, denoted $\Delta_p D = D_{p,\text{far}} - D_{p,\text{near}}$.

A critical assumption of the border discontinuity design is that districts on either side of jurisdictional boundaries are similar in their observable and unobservable characteristics. Table

7 tests this assumption by comparing mean pre-reform characteristics between *Far* and *Near* districts in 2001. The table reports the mean values for each group, the difference (Near - Far), and the p-value from a test of equality of means.

Table 7: Balance of Pre-Reform Characteristics: Bordering District Pairs (2001)

| | Far | Near | Near - Far | p-value |
|---|---|---|---|---|
| Graduate share | 0.075 | 0.078 | 0.003 | 0.090 |
| Graduate share (Young) | 0.111 | 0.115 | 0.004 | 0.117 |
| Agricultural share | 0.723 | 0.715 | -0.007 | 0.118 |
| Service share | 0.187 | 0.190 | 0.003 | 0.294 |
| Mean wage (Rs./Year) | 49,056 | 48,182 | -874 | 0.410 |
| GDP per worker (Rs./Year) | 63,573 | 62,168 | -1,406 | 0.470 |
| Log Night Lights | 9.475 | 9.456 | -0.019 | 0.831 |
| Observations | | 780 district pairs | | |

*Notes:* Sample consists of district pairs sharing a geographic boundary but assigned to different university jurisdictions. *Far* indicates the district farther from its affiliating university; *Near* indicates the district closer to its affiliating university. P-values are from two-sample t-tests of equality of means.

The results in Table 7 confirm that bordering districts are well balanced across key characteristics in the pre-reform period. The mean share of college graduates is statistically indistinguishable between *Far* and *Near* districts (p = 0.09), and an alternative measure based on the younger population yields a similar result (p = 0.12). GDP per worker is nearly identical across the boundary, differing by only 1,406 rupees per year (p = 0.47). Mean wages are similarly balanced (p = 0.41). Differences in sectoral composition are economically negligible: the agricultural employment share differs by only 0.7 percentage points (p = 0.12), and the service share by 0.3 percentage points (p = 0.29). Nightlight intensity shows virtually no difference across the boundary (p = 0.83). Overall, there is no evidence of systematic differences in education, income, or sectoral structure across jurisdictional boundaries, providing strong support for the identifying assumption of the border discontinuity design.

The border discontinuity analysis estimates the causal effect of educational expansion on economic outcomes using variation in distance gaps across bordering district pairs. The key insight is that if distance affects outcomes only through its impact on educational attainment, then differences in outcomes between paired districts should be proportional to differences in their distance to affiliating universities, mediated through differences in educational expansion.

I implement this using a two-stage least squares (2SLS) regression on differenced data. All variables are defined as differences within pairs: $\Delta_p X = X_{p,\text{far}} - X_{p,\text{near}}$. The first stage estimates the effect of the distance gap on the change in graduate share:

$$\Delta(\Delta_p h) = \alpha_0 + \alpha_1 \Delta_p \log Dist_{1991} + \nu_p \tag{7}$$

where $\Delta(\Delta_p h) = (h_{p,\text{far},2011} - h_{\text{far},2001}) - (h_{p,\text{near},2011} - h_{p,\text{near},2001})$ is the difference-in-differences

in graduate share across the pair, and $\Delta_p \log Dist_{1991}$ is the pre-determined distance gap as of 1991.

The second stage estimates the causal effect of educational expansion on outcomes:

$$\Delta(\Delta_p Y) = \beta_0 + \beta_1 \Delta(\widehat{\Delta_p h}) + v_p \tag{8}$$

where $\Delta(\Delta_p Y)$ is the difference-in-differences in the outcome variable, and $\Delta(\widehat{\Delta_p h_p})$ is the predicted value from the first stage. This approach identifies the causal effect by comparing how outcomes evolved differentially across border pairs as a function of their differential exposure to educational expansion, which in turn is driven by differential distance to affiliating universities.

Tables 8–10 present the empirical results from the border discontinuity design. Each observation corresponds to a district pair $p$ that shares a geographic boundary but belongs to different university jurisdictions. All variables are expressed as $\Delta(\Delta_p \cdot)$—that is, the change over time (2011–2001) of the Far–Near difference within each pair. The instrument $\Delta_p \log Dist_{1991}$ captures the baseline jurisdictional distance gap that shaped how the two districts were differentially exposed to the higher-education liberalization.

Table 8: Border Discontinuity: First Stage ($\Delta(\Delta_p \cdot)$ Specification)

| | (1)<br>$\Delta(\Delta_p h)$ (All Ages) | (2)<br>$\Delta(\Delta_p h)$ (Young, 20-34) |
|---|---|---|
| $\Delta \log(Dist)$ | -0.011*** | -0.016*** |
| | (0.003) | (0.005) |
| Constant | 0.003 | 0.005* |
| | (0.002) | (0.003) |
| Obs. | 780 | 780 |
| F-Stat | 14.16 | 11.53 |

*Notes:* Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Each observation is the 2011–2001 change in the Far–Near difference within a district pair.

The first-stage estimates confirm that pre-reform distance to affiliating universities strongly predicts the differential expansion of college graduates across bordering districts. A one-unit increase in the log distance gap between the Far and Near districts reduces the relative change in graduate share by 1.1 percentage points for the full population and 1.6 percentage points for the younger population aged 20-34. Both coefficients are statistically significant at the one-percent level. These results validate that the baseline jurisdictional geography of universities generated powerful, exogenous variation in educational expansion across borders.

The second-stage estimates using the $\Delta(\Delta_p \cdot)$ specification indicate large and economically meaningful effects of higher-education expansion on sectoral structure. Interpreting coefficients per one-percentage-point increase in the change-in-difference of graduate share, $\Delta(\Delta_p h)$, the point estimate for agriculture is $-0.902$ (significant at the 5 percent level). This

Table 9: Border Discontinuity: Second Stage ($\Delta(\Delta_p \cdot)$ IV Estimates)

| | (1) $\Delta(\Delta_p s)$ (Agri.) | (2) $\Delta(\Delta_p s)$ (Manu.) | (3) $\Delta(\Delta_p s)$ (Serv.) | (4) $\Delta(\Delta_p \log y)$ (GDPpw.) | (5) $\Delta(\Delta_p \log \text{NL})$ | (6) $\Delta(\Delta_p \log(w_H/w_L))$ |
|---|---|---|---|---|---|---|
| $\Delta(\Delta_p h)$ | -0.902** | -0.100 | 0.763*** | 2.051 | 6.523* | -10.407* |
| | (0.387) | (0.185) | (0.249) | (3.508) | (3.949) | (5.519) |
| Constant | -0.002 | 0.001 | -0.001 | -.010 | -0.033 | 0.035 |
| | (0.002) | (0.001) | (0.001) | (.020) | (0.021) | (0.036) |
| Observations | 780 | 780 | 780 | 780 | 732 | 721 |
| Effective F-stat | 14.16 | 14.16 | 14.16 | 14.16 | 14.13 | 15.35 |

*Notes:* Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
$\Delta(\Delta_p h)$ is instrumented with $\Delta_p \log Dist_{1991}$. Effective F-statistics follow Montiel-Olea & Pflueger (2013).

implies that a 1 pp greater rise in graduate share leads to agricultural employment share fall by about 0.90 pp more than its neighbor. For services, the coefficient is $0.763$ (significant at the 1 percent level), implying a 0.76 pp larger increase in the service-sector employment share per 1 pp larger increase in graduates. Manufacturing, by contrast, shows a small and statistically insignificant coefficient of $-0.100$.

Turning to macroeconomic aggregates, the estimated effect on output per worker is positive but imprecisely estimated: $\Delta(\Delta_p \log \text{GDPpw}) = 2.051$ with a standard error of $3.508$, not statistically significant at conventional levels. However, the nightlight-based measure provides stronger evidence of output effects: $\Delta(\Delta_p \log \text{NL}) = 6.523$ (significant at the 10 percent level), indicating that a 1 pp larger rise in graduates corresponds to a 6.5 percent higher growth in nightlight intensity on the Near side relative to the Far side. The wage-premium coefficient is $-10.407$ (significant at the 10 percent level), implying a strong local compression of the skill premium in places where graduate attainment expanded more rapidly.

Table 10: Comparison of Main IV and Border Discontinuity IV Estimates

| | Agri. | Manu. | Serv. | $\log(GDPpw)$ | $\log(NL)$ | $\log(w_H/w_L)$ |
|---|---|---|---|---|---|---|
| $\beta$ (BDD-IV) | -0.902 | -0.100 | 0.763 | 2.051 | 6.523 | -10.407 |
| $\beta$ (Main IV) | -0.774 | -0.115 | 0.894 | 2.262 | 6.173 | -4.938 |

*Notes:* BDD estimates use 2011-2001 changes in the Far-Near difference within border pairs, instrumented by $\Delta_p \log Dist_{1991}$. Main IV estimates rely on district-level variation in log distance to affiliating universities.

A direct comparison with the baseline district-level IV estimates in Table 10 shows striking concordance in both signs and magnitudes. For agriculture, the BDD coefficient is $-0.902$ while the Main IV estimate is $-0.774$. The difference in point estimates is modest (about $0.13$ pp per 1 pp in graduates) and well within the uncertainty implied by the BDD standard error (0.387), indicating that both designs agree that higher-education expansion induces a near one-for-one reduction in agricultural employment share at the margin. For services, BDD yields $0.763$ versus $0.894$ in the Main IV, with a difference of only $0.13$ pp. Manufacturing effects are

small and statistically indistinguishable from zero in both designs ($-0.100$ in BDD vs. $-0.115$ in Main IV).

For output effects, the comparison is most compelling when examining nightlight intensity: the BDD semi-elasticity of $6.523$ percent is nearly identical to the Main IV estimate of $6.173$ percent. The GDP per worker estimates show relatively close alignment as well (BDD: 2.05 vs. Main IV: 2.262), with both positive and of similar magnitude, though the BDD estimate is imprecisely estimated. The nightlight and GDP per worker concordance provides reassuring evidence that both designs capture similar underlying output effects. For the wage premium, BDD yields $-10.407$ compared to $-4.938$ in Main IV, both indicating substantial skill premium compression. In short, across sectoral shares, output measures, and wage inequality, the two designs—one exploiting continuous distance variation, the other exploiting border differences—deliver a coherent set of estimates.

The close correspondence between the border discontinuity estimates and the main IV estimates is particularly compelling. The border discontinuity design uses a fundamentally different source of variation—comparing adjacent districts with similar characteristics but different regulatory assignments—yet yields nearly identical estimates, especially for sectoral reallocation and nightlight-based output measures. This consistency across identification strategies provides strong evidence that the estimated effects capture the causal impact of educational expansion rather than spurious correlations driven by omitted variables. The similarity of estimates suggests that the exclusion restriction is satisfied: distance affects economic outcomes solely through its impact on educational attainment, not through any direct channel related to remoteness, connectivity, or unobserved local characteristics.

## 7.3 Placebo Test: Affiliating vs. Nearest University.

A key feature of the institutional setting is that each district is administratively linked to one affiliating university, which is the relevant margin for college entry because new colleges must obtain affiliation rights before opening. A potential concern is that $\log D_r^{\text{affil}}$ might proxy for generic geographic remoteness rather than the specific regulatory channel of affiliation that governs college entry. To address this, I exploit the distinction between the *designated* affiliating university (which grants affiliation rights) and the *geographically* nearest university. Among the 308 districts without a university headquarters during the reform period, 205 districts had their nearest university coincide with their designated affiliating university, in which case the two distances are mechanically identical. However, for the remaining 103 districts, the nearest university differed from the assigned affiliating university. This split allows for a clean placebo test restricted to the latter set, where the two distances are not collinear.

For these 103 districts, I jointly include *both* $\log D_r^{\text{affil}}$ and $\log D_r^{\text{near}}$ in a single *horse-race* regression. The analysis examines the impact on the change in the graduate share ($\Delta h$), changes

Table 11: Placebo Test: Affiliating vs. Nearest University Distance (2001–2011)

| | (1) $\Delta h$ | (2) $\Delta s$ (Agri.) | (3) $\Delta s$ (Serv.) | (4) $\Delta s$ (Manu.) | (5) $\Delta \log(\text{GDPpw})$ |
|---|---|---|---|---|---|
| Log-Dist-Affil | -0.021** | 0.018* | -0.019** | -0.075 | 0.174 |
| | (0.009) | (0.010) | (0.008) | (0.117) | (0.185) |
| Log-Dist-Near | -0.014 | 0.001 | -0.010 | -0.016 | -0.199 |
| | (0.009) | (0.010) | (0.008) | (0.114) | (0.201) |
| Constant | 0.219*** | -0.122*** | 0.180*** | 0.873** | 0.058 |
| | (0.049) | (0.044) | (0.032) | (0.397) | (0.613) |
| Observations | 103 | 103 | 103 | 103 | 99 |
| F-Stat | 6.87 | 2.86 | 10.62 | 0.66 | 0.53 |

*Notes:* Sample restricted to districts without a university headquarters for which the geographically nearest university differs from the designated affiliating university (103 districts). Robust standard errors in parentheses. The F-statistic is for the joint significance of both distance variables. Observations for GDP growth are lower due to data availability. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in sectoral employment shares ($\Delta s_k$ for agriculture, manufacturing, and services), and growth in log GDP per worker ($\Delta \log(\text{GDPpw})$). The estimating equation is:

$$\Delta y_r \;=\; \beta_{\text{A}} \log D_r^{\text{affil}} \;+\; \beta_{\text{N}} \log D_r^{\text{near}} \;+\; \gamma \;+\; \varepsilon_r,$$

where $\Delta y_r \in \{\Delta h,\, \Delta s_{\text{agri}},\, \Delta s_{\text{manu}},\, \Delta s_{\text{serv}},\, \Delta \log(\text{GDPpw})\}$, and $\gamma$ is a constant. By conditioning on both distances simultaneously, $\beta_{\text{A}}$ isolates the affiliation-based regulatory channel while $\beta_{\text{N}}$ captures residual effects of generic proximity. If affiliation is the operative margin for college entry and subsequent economic change, we should observe statistically significant estimates for $\beta_{\text{A}}$ (with signs indicating that greater distance hinders development), while $\beta_{\text{N}}$ should be close to zero.

Table 11 reports the results. The findings strongly support the proposed institutional mechanism. Across multiple specifications, the coefficient on the affiliation distance, $\log D_r^{\text{affil}}$, is statistically significant and consistently larger in magnitude than the coefficient on the nearest-distance placebo. In column (1), a greater distance to the affiliating university significantly reduces the growth in the graduate share ($\beta_A = -0.021$). This educational effect is linked to structural transformation: columns (2) and (3) show that affiliation distance is associated with a slower transition out of agriculture ($\beta_A = 0.018$, ) and a reduced expansion of services ($\beta_A = -0.019$). In stark contrast, the coefficient on the geographically nearest university, $\log D_r^{\text{near}}$, is statistically indistinguishable from zero and economically negligible in every single specification.

These placebo results convincingly rule out the concern that the instrument is simply capturing generic remoteness. Instead, what matters for human capital accumulation and its conse-

quent impact on economic structure is proximity to the administratively *designated* affiliating university, not just any nearby university. The fact that the joint specification assigns all the explanatory power exclusively to $\log D_r^{\text{affil}}$ powerfully reinforces the identification logic: the reform relaxed affiliation-related supply frictions, and districts closer to their affiliating university were better positioned to expand their stock of college-educated workers, in turn driving structural transformation.

**Summary.** Across a number of tests—absence of pre-trends in college expansion, null effects on pre-reform employment shares, spatial discontinuities in college entry, and negligible pre-reform correlations with outcomes—the evidence consistently supports the validity of the distance instrument. Distance to the affiliating university became salient only after liberalization reduced barriers to college entry, validating its use as a cost-shifting instrument for human capital supply.

# 8 Model

In this section, I develop the theoretical framework to identify human capital externalities based on empirical analysis. I then use the structural model to conduct two counterfactuals: (i) the evolution of the economy in the absence of education reform, and (ii) the effect of a subsidy for higher education. I begin by presenting a simple, illustrative model to build intuition and to clarify my strategy for identifying the human capital externality. This preliminary model demonstrates how the overall impact of education on productivity can be decomposed into a direct private return and an external spillover effect. Following this, I introduce the full structural model of a spatial economy. This more comprehensive framework incorporates endogenous skill acquisition decisions by individuals and production by firms, allowing for a richer analysis of the general equilibrium effects of educational policies. I subsequently use this structural model for estimation, calibration, and to perform counterfactual policy simulations.

## Identification of the Externality: An Illustrative Model

To fix ideas, I first consider a simple aggregate production function that clearly illustrates how I conceptualize and measure the human capital externality. I adopt a standard Cobb-Douglas production function in which output per worker, $y$, depends on the capital-output ratio, $k/y$, total factor productivity (TFP), $A$, and the level of human capital, $h$. Following the seminal work of Hall and Jones [1999], the production function is given by:

$$y = \left(\frac{k}{y}\right)^{\frac{\alpha}{1-\alpha}} Ah$$

where $\alpha$ is the capital share in the economy.

Human capital, $h$, is formed through formal education. I adopt the standard specification from the labor economics literature where the logarithm of human capital is proportional to the average years of schooling, $S$. Specifically,

$$\log h = \phi S$$

Here, $\phi$ represents the Mincerian return to an additional year of schooling. The average years of schooling in the economy, $S$, is a weighted average of the schooling levels of high-skilled workers ($S_H$) and low-skilled workers ($S_L$). Letting $s_c$ denote the share of the workforce that is college-educated (i.e., high-skilled), I can express the average years of schooling as:

$$S = s_c S_H + (1 - s_c) S_L$$

My primary goal is to understand how a change in the skill composition of the workforce—specifically, an increase in the share of college graduates, $s_c$—affects aggregate output per worker. To analyze this, I take the natural logarithm of the production function and differentiate it with respect to $s_c$:

$$\frac{d \log y}{d s_c} = \frac{\alpha}{1 - \alpha} \frac{d}{d s_c} \log\left(\frac{k}{y}\right) + \frac{d \log A}{d s_c} + \phi \left(S_H - S_L\right)$$

This expression shows that a change in the college share affects output through three distinct channels: its effect on the capital-output ratio, its effect on TFP, and its direct effect on the aggregate stock of human capital.

To isolate the key mechanisms of interest, I make two simplifying assumptions for this illustrative exercise. First, I assume that the economy is in a steady state where the capital-output ratio, $k/y$, is constant. This implies that the first term in the expression above is zero. Second, and most critically for my analysis, I posit that TFP is not exogenous but instead depends on the concentration of skilled workers in the economy. Specifically, I model TFP as $A = \bar{A} \exp\{\Omega s_c\}$, where $\bar{A}$ is a baseline productivity level and $\Omega$ captures the intensity of the human capital externality. This specification formalizes the idea that a higher share of college-educated workers may increase economy-wide productivity through knowledge spillovers, innovation, or other positive externalities.

Under these assumptions, the relationship simplifies considerably:

$$\frac{d \log y}{d s_c} = \phi \left(S_H - S_L\right) + \Omega$$

This equation provides a clear and intuitive decomposition of the total effect of an increase in the college share on output. The term $\phi \left(S_H - S_L\right)$ represents the direct, private effect of increasing the average human capital stock. The parameter $\Omega$ captures the external effect on total factor productivity. My identification strategy, therefore, is to estimate the total effect,

$\frac{d \log y}{ds_c}$, and subtract the direct effect to recover an estimate of the externality, $\Omega$.

To quantify this, I use established parameter values from the literature. The schooling gap between high-skilled and low-skilled workers, $S_H - S_L$, is approximately 8 years. The private Mincerian return to schooling, $\phi$, is typically estimated to be around 0.10 (10%). Suppose empirical observation suggests that the total elasticity of output with respect to the college share is 2.3. I can then back out the magnitude of the externality:

$$\begin{aligned} \frac{d \log y}{ds_c} &= \phi \left( S_H - S_L \right) + \Omega \\ 2.3 &= (0.10 \times 8) + \Omega \\ 2.3 &= 0.8 + \Omega \\ \Omega &= 1.5 \end{aligned}$$

This simple calculation suggests a substantial externality effect, where $\Omega = 1.5$. This value implies that a one percentage point increase in the share of college graduates in the workforce leads to a 1.5% increase in GDP per worker, purely through the externality channel. While this illustrative model provides a clear intuition for my approach, it abstracts from several important economic forces, such as endogenous skill choice and factor substitution. To address these limitations, I next present a more detailed structural model.

## A Structural Model of a Spatial Economy

I now develop a structural model of a spatial economy with $R$ regions, indexed by $r$, over two time periods, $t \in \{2001, 2011\}$. The model features a production side with firms that combine high-skilled and low-skilled labor, and a household side where a continuum of workers endogenously chooses their skill level. The equilibrium is characterized by wages, skill distributions, and output levels across all regions and time periods.

### Production

In each region $r$ and time period $t$, a representative firm produces a single final good. I assume a Constant Elasticity of Substitution (CES) production function that combines high-skilled labor ($h_{rt}$) and low-skilled labor ($l_{rt}$):

$$y_{rt} \;=\; A_{rt} \left[ \alpha_t h_{rt}^{\frac{\sigma-1}{\sigma}} + (1-\alpha_t) l_{rt}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

Here, $y_{rt}$ is output, $\sigma$ is the elasticity of substitution between high-skilled and low-skilled labor, and $\alpha_t$ is a time-varying parameter that governs the relative productivity of high-skilled labor.

The key element for my analysis is the total factor productivity term, $A_{rt}$. I model TFP as being determined by three components: a region-specific fixed effect, an aggregate time

trend, and an endogenous component driven by the local human capital stock. Specifically, the logarithmic specification for TFP is:

$$\log(A_{rt}) = \log(\bar{A}_r) + g_A \cdot t + \Omega h_{rt} + \varepsilon_{rt}$$

where $\varepsilon_{rt} \sim N(0, \sigma_\varepsilon)$ is a random shock. The term $\bar{A}_r$ captures time-invariant regional productivity differences, such as those arising from geography or institutions. The term $g_A \cdot t$ represents exogenous aggregate productivity growth that is common to all regions. The central component is $\Omega h_{rt}$, which explicitly links a region's productivity to its share of high-skilled workers, $h_{rt}$. The parameter $\Omega$ is the structural counterpart to the externality identified in the illustrative model and is the primary object of interest in my estimation.

**Skill Choice**

The supply of skilled and unskilled labor in each region is determined by the endogenous education choices of a continuum of workers. Each worker chooses an education level, $e \in \{L, H\}$, to maximize their utility. The utility from choosing a particular skill level depends on the logarithm of the corresponding wage, $w_{e,rt}$, the cost of acquiring education if choosing to become high-skilled, and an idiosyncratic taste shock. The optimization problem for a worker is:

$$\max_{e \in \{L,H\}} \log w_{e,rt} \; - \; \mathbf{1}\{e = H\}\, C_{rt} + \zeta\epsilon$$

where $\mathbf{1}\{e = H\}$ is an indicator function equal to one if the worker chooses to become high-skilled, and $\zeta\epsilon$ is an individual-specific preference shock drawn from a Type-I extreme value distribution, which yields a logit formulation for the skill choice probability.

The cost of acquiring a high-skilled education, $C_{rt}$, varies across regions and over time. I model this cost as having three distinct components:

$$C_{rt} = \mu_r + C_{f,t} + C_{d,t} \log D_r$$

The first component, $\mu_r$, is a region-specific, time-invariant factor that represents persistent differences in preferences for education or access to local educational infrastructure. The other components are the nationwide fixed cost, $C_{f,t}$, and the distance-dependent cost, $C_{d,t}$. I specify the cost in this manner as a reduced-form approach to capture the idea that the cost of obtaining a college degree in a region depends on the density and proximity of colleges. Indeed, it has been argued in several recent papers that proximity to a college is a significant determinant of enrollment [Acton et al., 2025]. This formulation directly relates to the empirical context of my study. As I show in the empirical section, a nationwide reform made it easier to open new colleges, leading to an overall reduction in the indirect costs of accessing higher education, a phenomenon captured by the evolution of $C_{f,t}$. Furthermore, as the empirical exercise demon-

strates, these costs varied with the distance to an affiliating university; this geographic variation is captured by the $C_{d,t}$ term.

### Equilibrium

For each region $r$ and time period $t$, a competitive equilibrium is a set of wages $\{w_{L,rt}, w_{H,rt}\}$, a skill share of high-skilled workers $h_{rt}$, and an aggregate output level $y_{rt}$ such that the following conditions are met:

1. **Optimal Skill Choice:** Given wages and costs, the skill share $h_{rt}$ results from the utility-maximizing decisions of all workers in the region.
2. **Profit Maximization:** The representative firm chooses its inputs of high-skilled and low-skilled labor, $\{l_{rt}, h_{rt}\}$, to maximize profits, taking wages and TFP as given. This implies that wages are equal to the marginal products of each labor type.
3. **Labor Market Clearing:** The demand for each type of labor by the representative firm equals the supply determined by workers' skill choices.
4. **Goods Market Clearing:** Regional output $y_{rt}$ is determined by the production function evaluated at the equilibrium levels of labor inputs and TFP.

I solve this model for its equilibrium outcomes, which then form the basis of my structural estimation and subsequent counterfactual analyses. The estimated parameters from this model will allow me to quantify the human capital externality, $\Omega$, and evaluate its contribution to economic growth.

## Quantification and Calibration

To proceed with the quantification of the model, I first characterize the equilibrium relationship between wages and skill supply. The demand for skills is derived from the firm's profit-maximization problem. The first-order condition for optimal labor inputs equates the wage premium to the relative marginal product of high-skilled and low-skilled labor. This condition yields a log-linear demand curve for the skill premium:

$$\log\left(\frac{w_{H,rt}}{w_{L,rt}}\right) \;=\; \log\left(\frac{\alpha_t}{1-\alpha_t}\right) \;-\; \frac{1}{\sigma}\,\log\left(\frac{h_{rt}}{l_{rt}}\right)$$

The supply of skills is determined by the aggregation of individual workers' utility-maximizing education choices. As specified in the model, the assumption of Type-I extreme value idiosyncratic taste shocks, with a dispersion parameter $\zeta > 0$, results in a logit formulation for the aggregate skill share. This gives a log-linear supply curve for skills:

$$\log\left(\frac{h_{rt}}{l_{rt}}\right) \;=\; \frac{1}{\zeta}\left[\log\left(\frac{w_{H,rt}}{w_{L,rt}}\right) \;-\; C_{rt}\right]$$

where the cost index, $C_{rt}$, is a composite of a region-specific fixed component, $\mu_r$, a national time-varying component, $C_{f,t}$, and a distance-dependent shifter, $C_{d,t} \log D_r$:

$$C_{rt} \;=\; \mu_r \;+\; C_{f,t} \;+\; C_{d,t} \log D_r$$

Combining the skill demand and skill supply equations provides a two-equation system for the two primary endogenous variables in each market $(r, t)$: the wage premium and the skill ratio. Solving this system simultaneously yields closed-form solutions for the equilibrium log wage premium and the log skill ratio as functions of the model's structural parameters and the cost shifters:

$$\log\left(\tfrac{w_{H,rt}}{w_{L,rt}}\right) \;=\; \frac{\sigma\zeta \,\log\left(\tfrac{\alpha_t}{1-\alpha_t}\right) + C_{rt}}{\sigma\zeta + 1}$$

$$\log\left(\tfrac{h_{rt}}{l_{rt}}\right) \;=\; \frac{\sigma \,\log\left(\tfrac{\alpha_t}{1-\alpha_t}\right) - \sigma\, C_{rt}}{\sigma\zeta + 1}$$

These equations form the basis of my estimation strategy, allowing me to recover the structural parameters of the model from the data on regional wages and skill shares.

I calibrate the model's parameters jointly using a Method of Simulated Moments (MSM) approach. The parameters are chosen to minimize the weighted squared distance between a set of moments predicted by the model and their empirical counterparts. While the estimation is joint, it is instructive to discuss the intuition for how specific moments provide identification for key parameters.

The elasticity of substitution, $\sigma$, is identified by both the cross-sectional relationship between the wage premium and the relative supply of skills, and by the IV response of wages to changes in skill composition; a higher elasticity implies that wages are less sensitive to changes in the skill ratio. The skill-share parameters, $\alpha_t$, are identified by the level of the wage premium for a given skill ratio in each period. The region-specific cost shifters, $\mu_r$, are treated as fixed effects and are set to perfectly match the observed skill ratio and wage premium in each region in the initial period ($t = 2001$), absorbing all time-invariant regional heterogeneity. The national time-varying cost components, $C_{f,t}$, are pinned down by the average skill ratio across all regions in each period, while the distance-dependent cost parameters, $C_{d,t}$, are identified by the empirical relationship between the change in skill shares and the distance measure, $\log D_r$. This slope captures the spatially heterogeneous impact of the nationwide education reform. The dispersion of idiosyncratic taste shocks, $\zeta$, which governs the elasticity of skill supply, is calibrated externally to 0.26, following [Khanna and Morales, 2025]. The key parameter of interest, the human capital externality $\Omega$, is identified using an instrumental variable-style moment. This moment is equivalent to the IV coefficient from a regression of the change in log output per worker on the change in the high-skill share, where the distance measure serves as the instrument. Analytically, this coefficient is calculated as the ratio of the covariance between regional output growth and the distance measure to the covariance between regional human

capital growth and the distance measure. This strategy uses the exogenous variation in skill accumulation driven by the reform's differential geographic impact to isolate the causal effect of human capital on productivity. Finally, the exogenous aggregate productivity growth rate, $g_A$, and the standard deviation of productivity shocks, $\sigma_\varepsilon$, are chosen to match the aggregate mean and variance of log output per worker in the final period ($t = 2011$).

# Estimation Results

Table 12: Structural Parameter Estimates

| Parameter | Description | Value |
|---|---|---|
| $\Omega$ | Human Capital Externality/Spillover | 1.31 |
| $\sigma$ | Elasticity of Substitution | 2.27 |
| $\alpha_0$ | High-Skill Share Parameter (t=0) | 0.52 |
| $\alpha_1$ | High-Skill Share Parameter (t=1) | 0.55 |
| $C_{f,0}$ | National Fixed Cost (t=0) | 0.84 |
| $C_{f,1}$ | National Fixed Cost (t=1) | 0.42 |
| $C_{d,0}$ | Distance Cost Component (t=0) | 0.23 |
| $C_{d,1}$ | Distance Cost Component (t=1) | 0.27 |
| $g_A$ | Exogenous Productivity Growth | 0.49 |
| $\sigma_\varepsilon$ | Std. Dev. of Productivity Shocks | 0.23 |

Table 12 presents the key parameter estimates from this procedure. The estimate for the human capital externality, $\Omega$, is 1.31, which represents the central finding of this study. This magnitude implies substantial spillover effects from skilled labor accumulation: a 1 percentage point increase in a region's share of college graduates leads to a 1.31% increase in local total factor productivity over the 10-year period. This finding is economically significant and suggests that the social returns to education substantially exceed the private returns captured by individual wage premia. The estimated spillover effect is consistent with theories of knowledge diffusion, agglomeration economies, and learning-by-interacting in labor markets. It implies that regions with higher concentrations of skilled workers benefit not only from the direct productivity of those workers but also from positive externalities that raise the productivity of all workers in the region.

The estimated elasticity of substitution between high- and low-skilled labor, $\sigma$, is 2.27. This value falls within the range commonly found in the labor economics literature and indicates that skilled and unskilled labor are substitutes, though not perfect ones. An elasticity above 2 suggests that firms can relatively easily adjust their production processes in response to changes in the skill composition of the workforce. This has important implications for wage inequality: as the supply of skilled workers increases, the wage premium falls, but the decline is moderated by the fact that firms can reorganize production to utilize the newly abundant skilled labor.

The share parameters, $\alpha_t$, which govern the relative productivity weight on high-skilled labor in the production function, increase modestly from 0.52 in 2001 to 0.55 in 2011. This 5.8% increase captures skill-biased technological change over the decade, reflecting the growing importance of skilled labor in production. While modest, this shift indicates that technological progress favored skilled workers, partially offsetting the downward pressure on wage premia from increased skill supply.

The cost parameters reveal a striking pattern of policy-driven educational expansion. The national fixed component of education cost, $C_{f,t}$, falls dramatically from 0.84 to 0.42—a 50% reduction. This substantial decline captures the impact of major national reforms during the 2000s that lowered the fixed costs of establishing new colleges. The magnitude of this reduction suggests that policy reforms removed approximately half of the non-distance-related obstacles to college entry, making it substantially easier to open and operate institutions.

In contrast, the distance-dependent cost component, $C_{d,t}$, increases slightly from 0.23 to 0.27, a 17% rise. This increase reflects a fundamental shift in how the spatial distribution of colleges was determined. In the pre-reform era, college supply was limited and dominated by the public sector, where state and central governments made location decisions based largely on political considerations. Public colleges operated with direct government funding and oversight, making affiliation distance economically irrelevant.

The post-reform expansion, however, was driven primarily by private sector growth. Private colleges must affiliate with existing university, and the distance between a district and its affiliating university directly affects operating costs—including administrative overhead, compliance and monitoring costs, coordination expenses, and the costs of obtaining affiliation. As private institutions came to dominate new college openings, these distance-related costs became increasingly binding constraints on where colleges could viably operate. The divergent trends in cost components reveal that while national reforms successfully reduced fixed barriers to entry, the market-based expansion through private affiliation networks made spatial frictions—specifically, the distance to affiliating universities—increasingly important in determining the geographic availability of higher education.

The exogenous productivity growth parameter, $g_A$, is estimated at 0.49, indicating annual TFP growth of approximately 4.9% per year over the period, which is consistent with the rapid economic expansion observed in the data. The standard deviation of idiosyncratic productivity shocks, $\sigma_\varepsilon$, is 0.23, capturing the dispersion in productivity across regions beyond what is explained by observable factors.

# Counterfactual Experiments

The estimated structural model provides a laboratory for policy analysis. I conduct two counterfactual experiments that use the estimated parameters to simulate alternative policy regimes. These exercises quantify how different educational policies shape skill formation, wage in-

equality, and aggregate economic growth through both direct effects on enrollment decisions and indirect effects through the human capital externality.

## Counterfactual 1: No Education Reform

The first counterfactual assesses the economic importance of the observed reduction in education costs between 2001 and 2011. To isolate the causal impact of these reforms, I simulate the 2011 equilibrium under the assumption that the supply-side cost parameters ($C_{f,t}$ and $C_{d,t}$) remained fixed at their higher 2001 levels. All other parameters, including the human capital externality $\Omega$ and the elasticity of substitution $\sigma$, retain their estimated values. Importantly, I assume that the skill-biased technology parameter $\alpha_t$ and the exogenous TFP growth component $g_A$ would have evolved as estimated—these represent the sources of economic growth that are independent of education policy in the model. This counterfactual therefore isolates the contribution of endogenous skill accumulation (driven by lower education costs) from exogenous technological progress. The experiment answers the question: what would the economy have looked like in 2011 if it had experienced the same technological change but without the education reforms of the 2000s?

Table 13: Counterfactual (i): No Education Reform ($C_f$ and $C_d$ at t=0 levels)

| Variable | Data t=0 | Data t=1 | Counterfactual (i) t=1 |
|---|---|---|---|
| Skill Share ($h_{rt}$) | 0.0691 | 0.1102 | 0.0795 |
| (%$\Delta$ from t=0) | | +59.4% | +15.1% |
| Wage Premium ($wp_{rt}$) | 3.44 | 3.06 | 3.58 |
| (%$\Delta$ from t=0) | | -11.2% | +4.0% |
| Output ($y_{rt}$) | 42,963 | 75,829 | 69,229 |
| (%$\Delta$ from t=0) | | +76.5% | +61.1% |

The results, presented in Table 13, reveal that the education reforms had profound effects on the economy. Without the cost reductions, the skill share would have grown by only 15.1% over the decade, compared to the 59.4% growth observed in the data. This represents a suppression of nearly three-quarters of the actual skill accumulation.

This suppressed skill accumulation would have fundamentally altered wage dynamics. Instead of the 11.2% decline in the wage premium observed in the data—reflecting increased supply of skilled labor—the wage premium would have increased by 4.0% in the counterfactual. This reversal occurs because, without sufficient growth in skilled labor supply, skill-biased technological change (captured by the increase in $\alpha_t$) would have dominated, driving up the relative price of scarce skilled labor.

Most significantly, the counterfactual reveals substantial aggregate output losses from foregone education reform. Average output growth in 2011 would have been of only 61.1% rather

than 76.5%. The education reforms are therefore responsible for 15.4 percentage points of aggregate output growth over the decade—approximately 20% of total growth. This output gain operates through two channels: the direct effect of additional skilled workers contributing to production, and the indirect effect of human capital externalities raising overall regional productivity.

## Counterfactual 2: Removing Spatial Constraints on University Affiliation

The second counterfactual examines a policy intervention aimed at reducing spatial frictions in access to higher education. Specifically, this experiment simulates the removal of geographic constraints on university affiliation, allowing each district to affiliate with the nearest university rather than *assigned* affiliating university. In the model, this intervention is implemented by setting the effective distance to the nearest university for all districts, which mechanically reduces the distance-related component of education costs. While the national fixed cost $C_{f,1}$ remains at its 2011 level of 0.42, the spatial friction component is substantially reduced. This counterfactual asks: how much additional economic growth could be achieved by further reducing spatial barriers to open and operate colleges?

Table 14: Counterfactual (ii): Removing Spatial Constraints (Nearest University Affiliation)

| Variable | Data t=0 | Data t=1 | Counterfactual (ii) t=1 |
|---|---|---|---|
| Skill Share ($h_{rt}$) | 0.0691 | 0.1102 | 0.1173 |
| (%$\Delta$ from t=0) | | +59.4% | +69.8% |
| Wage Premium ($wp_{rt}$) | 3.44 | 3.06 | 2.96 |
| (%$\Delta$ from t=0) | | -11.2% | -13.9% |
| Output ($y_{rt}$) | 42,963 | 75,829 | 78,098 |
| (%$\Delta$ from t=0) | | +76.5% | +81.8% |

Table 14 shows that this targeted spatial reform would have economically significant effects beyond what was already achieved by the historical reforms. The skill share would rise to 0.1173, which is 6.4% higher than the 2011 baseline. This additional skill accumulation would come primarily from districts that were historically far from affiliating universities and therefore faced high affiliation-related costs for establishing colleges. This highlights the remaining importance of spatial frictions: even with substantial reductions in regulatory barriers and fixed costs, the geographic distribution of affiliating universities continues to constrain where colleges can viably operate and thus limits educational access.

The expansion in skilled labor supply would exert further downward pressure on the wage premium, which would fall by 13.9% rather than the observed 11.2%—an additional 2.7 percentage point decline relative to 2011. Most importantly, the combination of higher skill attainment and the human capital externality would boost aggregate output growth to 81.8%, which

is 5.3 percentage points higher than the 76.5% growth observed in the data. This represents approximately a 3% additional gain in GDP relative to the 2011 baseline. This substantial incremental gain, despite the major reforms already implemented, suggests that spatial barriers remain an important binding constraint on economic growth.

The policy implications are striking. While the historical education reforms of the 2000s generated large economic gains by reducing fixed costs and expanding capacity, significant additional gains remain achievable by targeting the spatial organization of university affiliation networks. Policies that reduce affiliation-related frictions—such as establishing new affiliating universities in underserved regions, or implementing flexible multi-university affiliation systems—could yield substantial returns. Given that such spatial policies may be less costly than further reductions in entry barriers or massive additional capacity expansion at existing universities, this counterfactual suggests a potentially high return on investment for geographically-targeted higher education policy.

# 9 Conclusion

This paper provides evidence that despite concerns about institutional quality, skill mismatch, and other potential limitations, the returns to higher-education expansion are substantial. Exploiting India's liberalization of private college entry combined with spatial variation generated by university affiliation requirements, I show that a one-percentage-point increase in the share of college graduates leads to a 2.3 percent increase in GDP per worker and drives substantial labor reallocation: agricultural employment declines by 0.8 percentage points while service employment rises by 0.9 percentage points. The estimates imply that the social returns to education exceed private returns by 30 percent. A calibrated spatial equilibrium model implies that for the average district in India, the reform accounts for 20 percent of output per worker growth between 2001 and 2011.

These findings carry several implications for development policy. The results demonstrate that higher-education expansion can be achieved through private provision when regulatory barriers are reduced, offering a feasible path for resource-constrained governments. The large externalities I document provide an economic rationale for continued government intervention despite private sector involvement. While concerns about institutional quality, skill mismatch, queuing for government jobs, and migration are relevant, the evidence shows these factors do not eliminate the substantial productivity gains from expanding college access. In the Indian context, counterfactual simulations reveal that removing spatial constraints on university affiliation would increase output by an additional 3 percent, suggesting that policy reforms targeting geographic barriers to educational access can yield substantial additional gains.

The evidence positions human-capital investment as a central mechanism for economic development and structural transformation in developing countries. For economies seeking to transition workers from traditional agriculture to modern industry and services, expanding

higher education builds the skills necessary for sustained growth and modernization.

# References


Daron Acemoglu and Joshua Angrist. How large are human-capital externalities? evidence from compulsory schooling laws. *NBER macroeconomics annual*, 15:9–59, 2000.

Riley Acton, Kalena E Cortes, Lois Miller, and Camila Morales. Distance to degrees: How college proximity shapes students' enrollment choices and attainment across race-ethnicity and socioeconomic status. *Economics of Education Review*, page 102724, 2025.

Abhijit V Banerjee and Esther Duflo. Growth theory through the lens of development economics. *Handbook of economic growth*, 1:473–552, 2005.

A Patrick Behrer, Rishabh Choudhary, and Dhruv Sharma. Air pollution reduces economic activity: Evidence from india. Technical report, The World Bank, 2023.

Mark Bils and Peter J Klenow. Does schooling cause growth? *American economic review*, 90 (5):1160–1183, 2000.

Francisco J Buera, Joseph P Kaboski, Richard Rogerson, and Juan I Vizcaino. Skill-biased structural change. *The Review of Economic Studies*, 89(2):592–625, 2022.

David Card. Using geographic variation in college proximity to estimate the return to schooling, 1993.

David Card. Using geographic variation in college proximity to estimate the return to schooling. *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pages 201–222, 1995.

David Card. The causal effect of education on earnings. In Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics*, volume 3A, pages 1801–1863. Elsevier, Amsterdam, 1999.

Pedro Carneiro, James J. Heckman, and Edward J. Vytlacil. Estimating marginal returns to education. In *American Economic Review*, volume 101, pages 2754–2781. 2011.

Amparo Castelló-Climent, Latika Chaudhary, and Abhiroop Mukhopadhyay. Higher education and prosperity: From catholic missionaries to luminosity in india. *The Economic Journal*, 128(616):3039–3075, 2018.

Antonio Ciccone and Giovanni Peri. Identifying human-capital externalities: Theory with applications. *The Review of Economic Studies*, 73(2):381–412, 2006.

Federica Coelli, Difei Oouyang, Weidi Yuan, and Yuan Zi. Educating like china. Technical report, mimeo, University of Zurich, 2023.

Alvaro Esteban Lescano Cox. The role of human capital and institutions in economic growth. 2024.

Janet Currie and Enrico Moretti. Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly journal of economics*, 118 (4):1495–1532, 2003.

Esther Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review*, 91(4):795–813, 2001.

Barry Eichengreen and Poonam Gupta. The two waves of service-sector growth. *Oxford Economic Papers*, 65(1):96–123, 2013.

Niklas Engbom, Hannes Malmberg, Tommaso Porzio, Federico Rossi, and Todd Schoellman. Economic development according to chandler. Technical report, Tech. rep., mimeo, New York University, 2024.

Tianyu Fan, Michael Peters, and Fabrizio Zilibotti. Growing like india—the unequal effects of service-led growth. *Econometrica*, 91(4):1457–1494, 2023.

Nicola Gennaioli, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. Human capital and regional development. *The Quarterly journal of economics*, 128(1):105–164, 2013.

Amory Gethin. Distributional growth accounting: Education and the reduction of global poverty, 1980–2019. *Quarterly Journal of Economics*, 140(4):2571–2618, 2025. doi: 10.1093/qje/qjaf033.

Devaki Ghose. *Trade, Internal Migration, and Human Capital: Who Gains from India's IT Boom?* World Bank, 2021.

Joshua Goodman, Michael Hurwitz, and Jonathan Smith. Access to 4-year public colleges and degree completion. *Journal of Labor Economics*, 35(3):829–867, 2017.

James Gordon and Poonam Gupta. Understanding india's services revolution. In *India's and China's recent experience with reform and growth*, pages 229–263. Springer, 2005.

Robert E Hall and Charles I Jones. Why do some countries produce so much more output per worker than others? *The quarterly journal of economics*, 114(1):83–116, 1999.

J Vernon Henderson, Adam Storeygard, and David N Weil. Measuring economic growth from outer space. *American economic review*, 102(2):994–1028, 2012.

Yingyao Hu and Jiaxiong Yao. Illuminating economic growth. *Journal of Econometrics*, 228 (2):359–378, 2022.

Gaurav Khanna. Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. *Journal of Political Economy*, 131(2):549–591, 2023.

Gaurav Khanna and Nicolas Morales. The it boom and other unintended consequences of chasing the american dream. 2025.

Peter J Klenow and Andres Rodriguez-Clare. The neoclassical revival in growth economics: Has it gone too far? *NBER macroeconomics annual*, 12:73–103, 1997.

Zovanga L Kone, Maggie Y Liu, Aaditya Mattoo, Çağlar Ozden, and Siddharth Sharma. Internal borders and migration in India. *Journal of Economic Geography*, 18(4):729–759, 2018.

Alan B Krueger and Mikael Lindahl. Education for growth: Why and for whom? *Journal of economic literature*, 39(4):1101–1136, 2001.

Sang Min Lee and Bipul Verma. Higher education policies and spatial inequality across US states. Unpublished manuscript, 2024.

Enrico Moretti. Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review*, 94(3):656–690, 2004.

NOAA National Geophysical Data Center. Version 4 dmsp-ols nighttime lights time series. Technical report, NOAA National Centers for Environmental Information, 2014. URL `https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`.

Tommaso Porzio, Federico Rossi, and Gabriella Santangelo. The human side of structural transformation. *American Economic Review*, 112(8):2774–2814, 2022.

Sheetal Sekhri, Md Amzad Hossain, and Pooja Khosla. Access to colleges, human capital, and empowerment of women. *Journal of Human Resources*, 59(2):502–544, 2024.

Gauri Kartini Shastry. Human capital response to globalization: Education and information technology in india. *Journal of Human Resources*, 47(2):287–330, 2012.

Anna Valero and John Van Reenen. The economic impact of universities: Evidence from across the globe. *Economics of education review*, 68:53–67, 2019.

Khoa Vu. Higher education expansion and the rise of the skill-intensive service sector. 2023.

Jacob Wright. On the spatial distribution of colleges. Unpublished manuscript, 2024.

# A Data Appendix

## A.1 Sample Selection

The main analysis excludes all Union Territories and the seven North Eastern states (Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, and Tripura). These restrictions are necessary to maintain data consistency across time periods. Union Territories experienced frequent administrative reorganizations during the sample period, making it difficult to construct consistent district-level panels. The North Eastern states exhibit substantial gaps and inconsistencies in economic census data going back in time. Excluding these regions ensures that the estimated effects are not confounded by measurement error or structural breaks arising from data collection issues.

The main analysis further restricts the sample to districts that did not host a university in 1991. This restriction is imposed for two reasons. First, districts that host universities represent a qualitatively different economic environment. Universities generate direct local economic effects—through employment of faculty and staff, demand for local services, knowledge spillovers to nearby firms, and attraction of complementary institutions—that are independent of their role as affiliating authorities for the broader network of colleges. These agglomeration effects are likely to be strongest in the immediate vicinity of the university and decline rapidly with distance. Including university-hosting districts in the analysis would therefore conflate two distinct channels: the regulatory effect of distance on college entry throughout the affiliated network, and the direct economic impact of hosting a major educational institution locally.

Second, excluding university districts enhances the comparability of the sample. The remaining districts are homogeneous in a crucial respect—none possess a university of their own—and differ primarily in their regulatory distance to the affiliating university. This ensures that the instrument captures variation in access to higher education through the affiliation channel, rather than reflecting the differential economic benefits of hosting versus not hosting a university. By comparing only non-university districts located at varying distances from their designated affiliating institution, the sample restriction makes the exclusion restriction more credible and strengthens the interpretation that distance affects economic outcomes exclusively through its impact on educational expansion.

## A.2 Construction of Crosswalk

India has undergone several changes in the boundaries of its districts. There have been several instances of district merges, splits, as well as redrawing of boundaries. In such situation it becomes essential to have consistent set of boundaries for any time series analysis. I map all the districts in different years to consistent set of boundaries in 1991 census. To do this I

overlay the district boundary map of a year on the district boundary map of 1991. I then use the overlayed map to construct the area under interseaction for each of new district with the disrict in 1991. This area is then used to construct weights to assign fraction of districts in a given year that belongs to a distrct in 1991.

## A.3 Classification of Industries

| Manufacturing | |
|---|---|
| Low-skill | High-skill |
| Mining | Machinery and equipment |
| Food and beverages | Petroleum and chemicals |
| Metal products | |
| Textiles | |
| **Services** | |
| Low-skill | High-skill |
| Personal services | Education |
| Recreation | Finance and real estate |
| Repair | Health services |
| Transport | ICT |
| Wholesale and retail | Professional services |
| Utilities | Public administration |
| Other services | |

Table 15: Skill Classification of Manufacturing and Service Industries

*Notes:* Industries are classified by sector (manufacturing vs. services) and skill intensity (high vs. low).

# B Additional Exercises

## B.1 Why Distance to the Affiliating University Matters

**Exam Logistics.** The affiliation system imposes substantial logistical burdens on universities and colleges alike, particularly in the conduct of end-term examinations. Throughout the sample period (1991-2011) and indeed well beyond, examination administration remained heavily centralized and paper-based. Question papers are set at the university level, printed in secure facilities, and then physically distributed to examination centers scattered across the affiliated network—often spanning hundreds of kilometers. The UGC's 2019 report on *Evaluation Reforms in Higher Education Institutions in India* highlights the scale of this challenge: "The machinery of conduct of end-term examination exerts tremendous pressure on affiliating universities due to the large number of examinees. The question papers, which are set, have to be printed and sent to various examination centers, which are at a distance requiring huge logistics support to ensure that the examinations are conducted simultaneously. The answer scripts of the students are then transported to designated centralized assessment centers."[8] This physical transport requirement creates a direct cost that scales with distance. Moreover, examination security protocols impose strict personal accountability on college administrators. For instance, Bengaluru City University's *Examination Manual* explicitly states: "(The Chief Superintendent) shall arrange to collect question papers on a daily basis from the University or by any other means as directed by the Registrar (Evaluation). Safe custody of the question papers shall be the responsibility of the Chief Superintendent and any lapse in this regard will be viewed very seriously."[9] Such requirements mean that college staff must either make repeated trips to the university campus or arrange costly secure courier services. After examinations conclude, answer scripts must be transported back to centralized evaluation centers, again requiring secure logistics and imposing delays that lengthen with distance. These operational costs—both monetary and in terms of administrative time—rise sharply as colleges are located farther from their affiliating university, creating a significant deterrent to establishing or maintaining colleges in remote districts.

**Administrative Burden.** Beyond examination logistics, the affiliation system generates continuous administrative interactions between colleges and universities. During the period studied, virtually the entire affiliation workflow remained paper-based and required in-person attendance at university offices. Colleges seeking initial recognition or periodic renewal of affiliation had to submit extensive hard-copy dossiers documenting infrastructure, faculty qualifications, library holdings, and financial viability. Each stage of the approval process—from obtaining a No-Objection Certificate (NOC) to securing statutory inspection reports—necessitated multiple follow-up visits to various university departments. As Vice-Chancellor Prof. Sabiha of

[8] UGC (2019), *Evaluation Reforms in Higher Education Institutions in India*.
[9] Bengaluru City University, *Examination Manual*, Clause 15.

Karnataka State Women's University observed when sixteen districts' women's colleges were instructed to shift their affiliation: "One of the major challenges faced by prospective affiliated colleges is connectivity and distance of the university. Many colleges in other districts found it more convenient to affiliate themselves with general universities closer to them."[10] This connectivity constraint was not merely a matter of convenience but reflected genuine transaction costs. College principals and administrative staff were expected to attend periodic meetings, faculty-development workshops, and academic-council sessions convened on the university campus. Curriculum revisions, affiliation renewals (which trigger mandatory physical inspections), and routine degree verifications similarly demanded in-person interaction. The offline nature of these processes persisted well into the 21st century. Digital reform arrived late to most states: as Uttar Pradesh's Deputy Chief Minister announced in December 2020, "Till now the NOC and related proposals were disposed of through offline processes. Affiliations will only be given online from the academic session 2021-22."[11] Since this digitalization occurred after the end of the sample period, physical remoteness remained a binding constraint throughout the 1990s and 2000s. Each additional kilometer between a college and its affiliating university translated into higher travel costs, longer delays in securing approvals, and greater administrative friction—factors that discouraged both new college entry and sustained operation in distant locations.

**Student Costs.** The administrative burden of distance extends beyond college staff to students themselves. Under the affiliation system, students frequently need to travel to the university campus for various administrative tasks: filing examination forms, collecting admit cards, obtaining transcripts and degree certificates, resolving grade disputes, and securing migration or character certificates. When the affiliating university is located far away, these requirements impose substantial monetary and time costs on students, particularly those from economically disadvantaged backgrounds. Student voices from across India consistently highlight this hardship. In Jharkhand, student-union leader Manoj Choubey noted: "We would try to make the officials aware of the problems faced by the students due to the location of the university in Ranchi, a distance of over 200 kilometers from some places."[12] Similarly, when Maharashtra's government created a separate university for the Konkan region, Higher-Education Minister Uday Samant explained: "For students studying in Sawantwadi in Konkan, Mumbai University is almost 650 kilometers away and it is difficult for them to travel that far for any work."[13] The problem is equally acute in other regions. A legislative brief on Gujarat's creation of Kutch University observed: "Over 16 colleges in the Kutch region are affiliated to Gujarat University in Ahmedabad, which is 500 km away from Bhuj. The decision [to create a new university] will put an end to the hardship faced by students who have to rush for various administrative

[10] Times of India, Bengaluru edition, 25 October 2016.
[11] Deputy Chief Minister, quoted in NDTV, 23 December 2020.
[12] *The Telegraph*, Ranchi, 2 August 2004.
[13] *Hindustan Times*, 11 January 2021.

approvals."[14] These travel requirements are particularly burdensome for students from rural areas, for whom a single round trip to a distant university can consume several days and represent a significant fraction of household income. Moreover, the opportunity cost of time spent traveling—time that could otherwise be devoted to coursework, part-time employment, or family responsibilities—further depresses the net returns to enrollment. Consequently, excessive travel costs not only reduce the attractiveness of distant colleges to prospective students but also increase dropout rates among enrolled students who find the cumulative burden unsustainable.

**University Approval Constraints.** Geographic distance also constrains the supply side of the affiliation market by limiting universities' willingness to grant new affiliations. Under India's regulatory framework, each university bears legal responsibility for safeguarding academic standards at every college within its jurisdiction. This responsibility requires routine inspections, monitoring of syllabus implementation, verification of faculty credentials, and oversight of examination administration. As the affiliated network expands geographically, these supervisory tasks become progressively more costly and eventually infeasible. Recognizing this constraint, successive policy committees have urged universities to tighten affiliation criteria. The *National Education Policy* (NEP) 2020 explicitly stipulates that no university should affiliate more than 300 colleges, linking this cap directly to the need for regular monitoring.[15] State governments have responded to these monitoring challenges by creating new universities specifically to reduce geographic span. In Bihar, the Times of India reported in 2023 that the government established new universities in Patliputra, Munger, and Purnea "with a view to ensuring better governance of the colleges under their jurisdiction," thereby "lightening the academic and infrastructural load" of the parent institution.[16] Similarly, when Odisha created a new university for Koraput district, officials conceded that the district "needed an affiliating university due to the distance between the colleges in the district and Berhampur University in Ganjam."[17] A confidential internal review by Rashtrasant Tukadoji Maharaj Nagpur University likewise observed that the wide geographical dispersion of its affiliated colleges "creates serious management challenges."[18] The practical implication is that universities, facing binding resource constraints, systematically favor granting affiliations to colleges located within a manageable radius. Colleges located beyond this threshold face higher rejection rates, longer approval delays, and greater risk of affiliation withdrawal—all of which depress college entry and reduce higher-education capacity in geographically remote areas.

[14]Times of India, Ahmedabad edition, 15 July 2014.
[15]Hindustan Times, 31 July 2020.
[16]Times of India, Patna edition, 7 March 2023.
[17]Times of India, Bhubaneswar edition, 16 February 2023.
[18]Rashtrasant Tukadoji Maharaj Nagpur University internal review.

## B.2 Effect on access to colleges

I begin by examining how the reform altered access to colleges across Indian districts. Since the reform specifically relaxed entry restrictions while maintaining the affiliation requirement, the key prediction is that districts located closer to their affiliating universities should have experienced a greater expansion in college supply relative to more distant districts. To test this hypothesis, I regress the growth in different measures of college access on the log distance to the affiliating university. Table 16 presents the results.

Column (1) uses the total change in the number of colleges in each district between 2001 and 2011 as the outcome. The coefficient on log distance is negative and statistically significant, implying that districts located farther from their affiliating university saw markedly slower growth in the number of colleges. The magnitude of the effect is economically large: an increase in distance of 100 km from the affiliating university (relative to the sample median of about 100 km) is associated with roughly 8 fewer colleges being established during the decade.

Columns (2)–(4) present results for density-based measures of college availability, which scale the number of new colleges by district area, population, and employment, respectively. Column (2) uses the number of colleges per 10 square kilometers. Because land area is fixed by construction, this measure is immune to confounding from demographic changes and isolates how the spatial distribution of colleges responded to distance. The estimated coefficient indicates that a 100 km increase in distance corresponds to a decline of about 0.44 colleges per 10 square kilometers, highlighting the sharp spatial gradient in college entry.

Column (3) scales college expansion by district population. Here the coefficient implies that districts 100 km farther from their affiliating university experienced 4.7 fewer colleges per million residents. Similarly, column (4) normalizes by employment, showing that such districts registered 19 fewer colleges per million employed persons. These results confirm that the distance penalty is not an artifact of differences in local demographic dynamics such as migration or population growth; rather, it reflects genuine differences in the capacity to establish new colleges.

Across all specifications, the constant term is positive and significant, capturing the nationwide expansion in college supply over the 2001–2011 period. However, the consistently negative and precisely estimated coefficients on log distance demonstrate that this nationwide trend was sharply stratified by proximity to affiliating universities. In short, the reform generated a substantial supply response, but the benefits were disproportionately concentrated in districts located closer to their legally mandated affiliating university, where the fixed costs of establishing and operating a new college were lower.

These findings provide the first piece of evidence validating the mechanism underlying my identification strategy: distance to the affiliating university is a powerful predictor of local college expansion after liberalization, consistent with the institutional frictions described earlier. In the next subsection, I show how this expansion in local college capacity translated into higher

educational attainment, thereby establishing the first-stage link between the distance instrument and the change in the share of college graduates.

Table 16: OLS Estimates: Growth in number of colleges (2001-2011)

| | (1) $\Delta Nc$ | (2) $\Delta(Nc/Area)$ | (3) $\Delta(Nc/Pop)$ | (4) $\Delta(Nc/Emp)$ |
|---|---|---|---|---|
| Log-Dist | -12.711* | -0.640*** | -6.778** | -19.352*** |
| | (7.513) | (0.127) | (2.721) | (6.893) |
| Constant | 108.368*** | 3.874*** | 50.761*** | 136.607*** |
| | (36.662) | (0.631) | (13.144) | (33.411) |
| Obs. | 367 | 367 | 367 | 367 |
| F-Stat | 2.86 | 25.35 | 6.21 | 7.88 |

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## B.3 Effect of College Expansion on Graduate Share

Having established that proximity to the affiliating university strongly predicts the expansion of local colleges, I now examine how this expansion translated into changes in educational attainment. In particular, I study how greater access to colleges affected the skill composition of the workforce. This step is crucial for establishing the relevance of the instrument: distance affects graduate supply only insofar as it altered the local availability of colleges.

Table 17 reports instrumental variables estimates of the effect of college expansion on different measures of skill composition. The endogenous regressor is the change in the number of colleges per 10 square kilometers, instrumented with log distance to the affiliating university. I focus on this density measure because it is fixed with respect to land area and hence not confounded by differences in demographic growth or migration across districts.

Column (1) uses as the outcome the change in the share of college graduates in district employment between 2001 and 2011. The coefficient is positive and highly significant: a one-unit increase in the number of colleges per 10 square kilometers raises the graduate share of employment by 2.7 percentage points.

Columns (2) and (3) turn to log skill ratios, which provide alternative measures of compositional change. In column (2) I define the denominator as all non-graduates (college versus no-college), while in column (3) I measure low skilled workers as middle-school equivalent education units. In both cases the coefficients remain positive and significant, with magnitudes implying that a one-unit increase in colleges per 10 square kilometers raised the graduate-to-non-graduate ratio by about 8.6 percent and the graduate-to-middle-school ratio by about 7 percent, respectively. These alternative measures confirm that the effect is not sensitive to the precise definition of the comparison group.

Across all specifications, the constants are positive and significant, reflecting the nationwide rise in college attainment during the 2000s. However, the IV results demonstrate that the extent of this increase varied systematically with local access to colleges induced by affiliation distance. In other words, the reforms did not merely coincide with a uniform nationwide upskilling; rather, the spatial distribution of new colleges directly shaped where the gains in graduate attainment occurred.

These results establish the core first-stage relationship: distance to the affiliating university predicts changes in the graduate share precisely via its effect on the entry of new colleges. This validates the use of distance as a cost-shift instrument for skill composition in the subsequent analysis. In what follows, I focus on the change in the share of college graduates as the main measure of educational attainment, while showing in the appendix that results are robust to alternative measures such as log skill ratios.

Table 17: IV Estimates: Changes in skill composition (2001-2011)

| | (1) $\Delta h$ | (2) $\Delta log(h/1-h)$ | (3) $\Delta(H/L)$ |
|---|---|---|---|
| $\Delta(Nc/Area)$ | 0.027*** | 0.086** | 0.070** |
| | (0.005) | (0.038) | (0.035) |
| Constant | 0.021*** | 0.431*** | 0.397*** |
| | (0.004) | (0.034) | (0.032) |
| Obs. | 367 | 367 | 367 |
| Eff. OP F | 25.35 | 25.35 | 25.35 |

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## B.4 Migration Rates across Districts

| Resident Type | Total | | No College | | College | |
|---|---|---|---|---|---|---|
| | ('000) | (%) | ('000) | (%) | ('000) | (%) |
| Non-migrants | 920,232 | 89.5 | 892,843 | 86.9 | 27,389 | 2.7 |
| Intra-state inter-district | 71,083 | 6.9 | 64,924 | 6.3 | 6,159 | 0.6 |
| Inter-state | 36,634 | 3.6 | 32,919 | 3.2 | 3,715 | 0.4 |
| Total population | 1,027,949 | 100.0 | 990,686 | 96.4 | 37,263 | 3.6 |

Table 18: Migration Patterns in India by Education Level

*Source:* Authors' tabulation based on India Census 2001 migration data [Kone et al., 2018].

The estimates presented in this paper on the effect of increases in college graduates on macroeconomic outcomes capture impacts after individuals have migrated both for education

and for work. However, understanding how gains from expanding education depend on the degree of migration may be of independent interest to policymakers. If a large fraction of individuals migrate out for education and work, the gains from expanding education access locally might be limited. Conversely, in the presence of migration frictions, gains from education expansion could be substantial. Recent evidence from the United States shows that in-state policies matter significantly for educational attainment: Lee and Verma [2024] demonstrate that federal and state higher education policies have contributed to growing cross-state wage inequality, while Wright [2024] shows how the spatial distribution of colleges affects educational access and aggregate welfare.

The Census of India 2001 conducted a detailed enumeration of migration patterns. Kone et al. [2018] used confidential microdata from this census to construct aggregate measures of migration disaggregated by education level. The migration rate is defined as the share of migrants out of the total stock of population currently residing in a district. Table 18 presents migration patterns for the Indian population by educational attainment. The table reveals that 89.5 percent of the population are non-migrants i.e they have never moved out of their native district. Intra-state inter-district migration accounts for 6.9 percent of the population, while inter-state migration represents only 3.6 percent.

Notably, migration rates differ substantially by education level. Among those without college education, 86.9 percent remain in their birth district, while 6.3 percent were from a different district, and 3.2 percent migrate in from different state. For college graduates, mobility rates are even lower: only 0.6 percent were from a different district, and just 0.4 percent from different state. These low inter-district migration rates in India suggest substantial spatial frictions in labor mobility. The limited migration among educated workers indicates that local education policies may have particularly pronounced effects on regional economic outcomes, as educated individuals are far less likely to relocate compared to their counterparts in more mobile economies such as the United States.

## B.5 Development Accounting with Externality

A long-standing puzzle in the growth literature concerns the tension between the strong cross-country correlation of education and income and the modest role of human capital in explaining income differences. Acemoglu and Angrist [2000] document that average schooling levels across countries are highly correlated with GDP per capita and growth rates, yet reconciling this strong observed correlation with a causal interpretation would require human capital externalities substantially larger than the private Mincerian returns to education. Klenow and Rodriguez-Clare [1997] conduct development accounting exercises decomposing income differences into contributions from physical capital, human capital, and total factor productivity (TFP), and find that human capital accounts for only 20-30 percent of cross-country income variation despite the strong bivariate correlation.

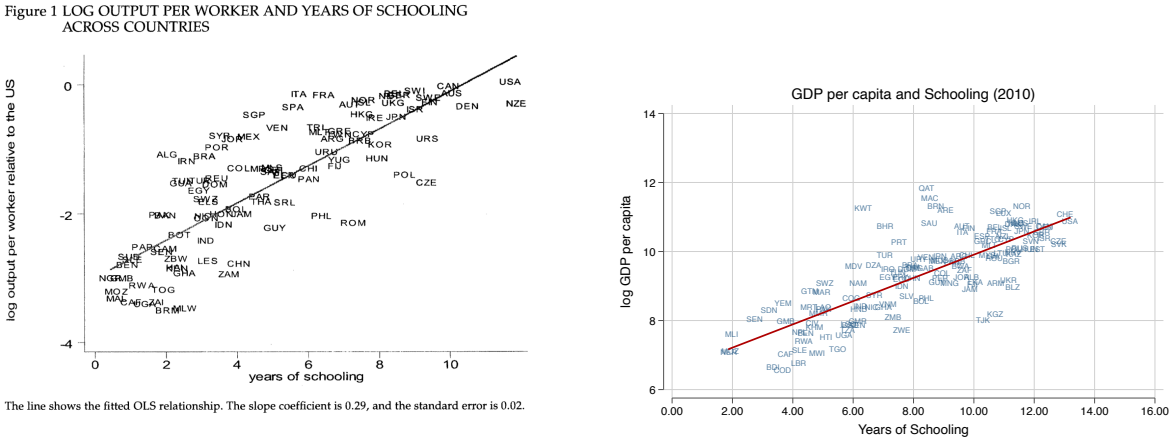
Bils and Klenow [2000] investigate this puzzle and find that using Mincerian returns to calibrate the effect of schooling on growth, the direct channel from schooling to growth explains less than one-third of the empirical cross-country relationship. Instead, they show that reverse causality, where faster expected growth induces greater educational investment by raising the effective return to schooling, and schooling having luxury consumption value is can explain the observed correlation. Acemoglu and Angrist [2000] provide direct empirical evidence using variation in compulsory schooling laws and find external returns to education are not significantly different from zero. Krueger and Lindahl [2001] attempt to reconcile these findings, arguing that measurement error in education data may partly explain the weak relationship in growth regressions.

This paper contributes to resolving this puzzle by providing credibly causal estimates of human capital externalities using within-country variation in India. Unlike cross-country studies that face challenges with omitted variables, reverse causality, and measurement error, this paper exploits quasi-experimental variation in college expansion driven by regulatory distance to affiliating universities. This identification strategy allows me to isolate the causal effect of increases in college graduate share on local economic outcomes, providing direct evidence on the magnitude of human capital externalities. The causal estimates obtained in this paper suggest substantially larger externalities than found in previous cross-country studies, helping to reconcile the strong correlation between education and income.

Figure A1 illustrates the strong bivariate relationship between education and income across countries that motivates the puzzle. Panel (a), reproduced from Acemoglu and Angrist [2000], shows that average schooling explains approximately 65 percent of the cross-country variation in income levels. However, as Acemoglu and Angrist [2000] demonstrate, a causal interpretation of this relationship would require human capital externalities on the order of 25-30 percent—approximately three times the private returns to schooling. Panel (b) presents updated evidence using 2010 data, plotting log GDP per capita against the years of schooling across countries. The strong positive relationship is evident, with countries in the 90th percentile of college attainment having GDP per capita levels approximately 3.2 log points (24 times) higher than countries in the 10th percentile.

Table 19 presents simple OLS regressions quantifying this relationship. Columns (1) and (2) show that a one-year increase in average schooling is associated with a 34 percent increase in GDP per capita and a 30 percent increase in GDP per worker. Columns (3) and (4) show that a one-percentage-point increase in the college graduate share is associated with a 6.9 percent increase in GDP per capita and a 6.0 percent increase in GDP per worker. Also note that string R2 means that variation years of schooling explains 60% of variation in GDP per capita. These strong correlations are consistent across different measures of education and income, but as the preceding literature demonstrates, it remains unclear how much of this relationship is causal.

The causal estimates obtained in this paper allow us to revisit the development accounting exercise with credibly identified externality parameters. The exercise follows the standard




(a) Education and Income (Acemoglu and Angrist, 2001)

(b) GDP and College Graduate Share (2010)

Figure A1: Cross-Country Relationship Between Education and Income

*Notes:* Panel (a) shows the bivariate regression of log GDP per capita relative to US on average years of schooling. Source: Acemoglu and Angrist [2000]. Panel (b) shows the relationship between log GDP per capita and the year of schooling across countries in 2010. Each point represents a country.

Table 19: OLS Estimates: GDP and Schooling (2010)

| | (1) log GDP pc | (2) log GDP pw | (3) log GDP pc | (4) log GDP pw |
|---|---|---|---|---|
| Years of schooling | 0.337*** | 0.299*** | | |
| | (0.023) | (0.023) | | |
| College graduate share | | | 6.906*** | 6.023*** |
| | | | (0.662) | (0.638) |
| Constant | 6.542*** | 7.767*** | 8.397*** | 9.432*** |
| | (0.206) | (0.201) | (0.120) | (0.116) |
| Observations | 137 | 137 | 137 | 137 |
| R-squared | 0.60 | 0.56 | 0.45 | 0.40 |

*Notes:* This table presents OLS regressions of log GDP per capita (columns 1 and 3) and log GDP per worker (columns 2 and 4) on measures of educational attainment across countries in 2010. Years of schooling measures average years of completed education. College graduate share is the fraction of the population with a college degree. Robust standard errors in parentheses.

approach in the growth literature. Consider a production function of the form:

$$y = \left(\frac{k}{y}\right)^{\frac{\alpha}{1-\alpha}} Ah, \tag{9}$$

where $y$ is output per worker, $k/y$ is the capital-output ratio, $A$ is total factor productivity, and $h$ represents human capital per worker. Taking logs and holding the capital-output ratio constant, income differences across countries can be decomposed as:

$$\log y = \log A + \log h. \tag{10}$$

In the absence of externalities, human capital is typically modeled as $\log h = \phi S$, where $S$ denotes years of schooling and $\phi$ represents the private Mincerian return to education, typically estimated around 0.10. However, if human capital generates positive externalities, aggregate productivity also depends on the education level of the workforce: $A = \bar{A}\exp\{\Omega S\}$, where $\Omega$ captures the externality effect. In this case, the income gap between rich and poor countries can be expressed as:

$$\log\left(\frac{y_{\text{rich}}}{y_{\text{poor}}}\right) = (\phi + \Omega)(S_{\text{rich}} - S_{\text{poor}}). \tag{11}$$

To quantify the role of externalities, I compare countries at the 10th and 90th percentiles of the income distribution in 2010. The income gap between these countries is $\log(y_{90}/y_{10}) = 3.19$ log points, while the schooling gap is $\Delta S = 7.04$ years. Using a standard Mincerian return of $\phi = 0.10$, schooling differences alone can account for $0.10 \times 7.04 = 0.704$ log points, or 22 percent of the observed income gap. This limited explanatory power exemplifies the development accounting puzzle.

I now incorporate the externality estimates from this paper to assess whether accounting for spillovers can reconcile the cross-country patterns. The reduced-form estimates indicate that a one-percentage-point increase in the college graduate share raises real GDP per worker by 1.3 percent beyong the internal returns to education. To translate this into a schooling externality parameter $\Omega$, I note that an increase in the college graduate share of one percentage point corresponds to an increase in average years of schooling of approximately 8 years (the difference between a college graduate with 16 years of schooling and a non-graduate with 8 years). This implies an externality effect of $\Omega = 1.3/8 = 0.1625$ per year of schooling.

Table 20 presents the development accounting results with and without externalities. In the baseline case with no externalities ($\Omega = 0$), schooling differences account for only 22 percent of the income gap between the 10th and 90th percentile countries. However, once externalities are incorporated ($\Omega = 0.1625$), the total contribution of schooling increases to $(\phi + \Omega) \times \Delta S = 0.2625 \times 7.04 = 1.848$ log points, which accounts for 58 percent of the observed income gap. Thus, incorporating credibly causal estimates of human capital externalities substantially increases the explanatory power of education in cross-country income differences, helping to

resolve the development accounting puzzle.

Table 20: Development Accounting: The Role of Externalities

| | Income gap $\log(y_{90}/y_{10})$ | Schooling gap $\Delta S$ | Percent of gap accounted No externality ($\Omega = 0$) | With externality ($\Omega = 0.1625$) |
|---|---|---|---|---|
| 10th vs 90th percentile | 3.19 | 7.04 | 22% | 58% |

*Notes:* This table presents development accounting results for countries at the 10th and 90th percentiles of the income distribution in 2010. Without externalities, schooling contributes $\phi \times \Delta S = 0.10 \times 7.04 = 0.704$ log points, accounting for 22 percent of the income gap. With externalities of $\Omega = 0.1625$, the total contribution is $(\phi + \Omega) \times \Delta S = 1.848$ log points, accounting for 58 percent of the income gap.

These results suggest that human capital externalities are quantitatively important for understanding cross-country income differences. The key advantage of this paper's approach is that it provides causal identification of these externalities using quasi-experimental variation within India, addressing the endogeneity concerns that plague cross-country regressions. While reverse causality from growth to schooling likely explains part of the strong cross-country correlation, the presence of substantial causal externalities suggests that education expansion can contribute to economic development.

# C Identification of Parameters

Although all model parameters are calibrated jointly to match a set of empirical moments, this section provides the economic intuition for how each parameter is separately identified. The identification proceeds in two logical stages. First, the parameters governing skill supply and demand are discussed. Second, given these parameters, the identification of the aggregate productivity process is outlined.

## Supply and Demand Parameters ($\sigma, \alpha_t, C_{d,t}, C_{f,t}, \mu_r$)

The identification of the supply and demand parameters proceeds sequentially, anchored by an estimate of the elasticity of substitution, $\sigma$. The cost of skill acquisition, $C_{rt}$, is assumed to have a structure that includes a region-specific fixed effect, a national time-varying component, and a component that depends on the regional characteristic $D_r$:

$$C_{rt} \;=\; \mu_r \;+\; C_{f,t} \;+\; C_{d,t}\,\log D_r.$$

The taste parameter for skill acquisition, $\zeta$, is set externally to 0.26, following the evidence on educational choices in Khanna and Morales [2025].

### 1. Elasticity of Substitution ($\sigma$)

The parameter $\sigma$ governs the slope of the labor demand curve. The model's first-order condition equates the wage premium to the relative marginal product of skill, yielding the log-linear demand relationship:

$$\log\left(\tfrac{w_{H,rt}}{w_{L,rt}}\right) \;=\; \log\left(\tfrac{\alpha_t}{1-\alpha_t}\right) \;-\; \frac{1}{\sigma}\,\log\left(\tfrac{h_{rt}}{l_{rt}}\right).$$

Taking first differences over time gives the relationship in changes:

$$\Delta\log\left(\tfrac{w_{H,rt}}{w_{L,rt}}\right) \;=\; \Delta\log\left(\tfrac{\alpha_t}{1-\alpha_t}\right) \;-\; \frac{1}{\sigma}\,\Delta\log\left(\tfrac{h_{rt}}{l_{rt}}\right).$$

The term $\Delta\log(h_{rt}/l_{rt})$ can be approximated for small changes in $h$ around a mean value $\bar{h}$ as $\Delta\log(h/(1-h)) \approx [1/(\bar{h}(1-\bar{h}))]\cdot\Delta h_{rt}$. This yields:

$$\Delta\log\left(\tfrac{w_{H,rt}}{w_{L,rt}}\right) \approx \Delta\log\left(\tfrac{\alpha_t}{1-\alpha_t}\right) - \frac{1}{\sigma\bar{h}(1-\bar{h})}\Delta h_{rt}.$$

This equation shows that the change in the wage premium is a linear function of the change in the skill share. The regional characteristic $\log D_r$ is used as an instrument for the change in skill share, $\Delta h_{rt}$, assuming it is uncorrelated with the aggregate technology shift, $\Delta\log(\alpha_t/(1-\alpha_t))$. The resulting IV estimate, $\hat{\beta}_{IV} = \text{Cov}(\Delta\log(wp), \log D)/\text{Cov}(\Delta h, \log D)$, identifies the

coefficient on $\Delta h_{rt}$. The parameter $\sigma$ can then be recovered directly from this estimate:

$$\sigma = -\frac{1}{\hat{\beta}_{IV} \cdot \bar{h}(1-\bar{h})}.$$

Given the empirical estimate of $\hat{\beta}_{IV} \approx -5.8$ and a mean skill share of $\bar{h} \approx 0.1$, this strategy implies a value for $\sigma \approx 2.2$.

**2. Cost and Skill-Intensity Parameters ($C_{d,t}, \alpha_t, C_{f,t}$)**

With $\sigma$ and $\zeta$ known, the remaining parameters are identified by targeting cross-sectional moments. The model's closed-form solution for the log skill-ratio is:

$$\ell_{rt} \equiv \log(h_{rt}/l_{rt}) \;=\; \frac{\sigma \,\log\left(\frac{\alpha_t}{1-\alpha_t}\right) - \sigma\, c_{rt}}{\sigma\zeta + 1}.$$

First, the distance-dependent cost parameters, $C_{d,t}$, are identified from the cross-sectional relationship between skill ratios and distance. The model implies a slope of $\frac{\partial \ell_{rt}}{\partial \log D_r} = -\frac{\sigma C_{d,t}}{\sigma\zeta+1}$. This theoretical slope is matched to the empirical regression coefficient of $\log(h_{rt}/l_{rt})$ on $\log D_r$ for each period $t \in \{0, 1\}$ to solve for $C_{d,0}$ and $C_{d,1}$.

Second, the skill-intensity parameters, $\alpha_t$, are identified by taking national expectations over the demand curve equation. This yields a direct mapping from the observed national average wage premia and skill ratios to the technology parameters:

$$\log\left(\tfrac{\alpha_t}{1-\alpha_t}\right) \;=\; \mathbf{E}\Big[\log\left(\tfrac{w_{H,rt}}{w_{L,rt}}\right)\Big] \;+\; \frac{1}{\sigma}\,\mathbf{E}\Big[\log\left(\tfrac{h_{rt}}{l_{rt}}\right)\Big].$$

Third, the national fixed-cost components, $C_{f,t}$, are identified by taking national expectations over the equilibrium skill-ratio equation. Assuming the regional fixed effects have a mean of zero, $\mathbf{E}[\mu_r] = 0$, taking expectations and rearranging the equilibrium equation for $\ell_{rt}$ gives:

$$C_{f,t} = \log\left(\tfrac{\alpha_t}{1-\alpha_t}\right) - C_{d,t}\mathbf{E}[\log D_r] - \frac{\mathbf{E}[\ell_{rt}](\sigma\zeta+1)}{\sigma}.$$

All terms on the right-hand side are either data moments or parameters that have been previously identified, thus allowing for a direct calculation of $C_{f,t}$.

Finally, the regional fixed effects, $\mu_r$, are backed out as the region-specific residuals that ensure the model's supply equation for period $t = 0$ perfectly matches the observed data on wage premia and skill shares for each region $r$.

## Productivity Parameters ($\Omega, g_A, \sigma_\varepsilon$)

The parameters of the productivity process are identified from the evolution of regional output per worker, $y_{rt}$. The model's production function implies that output growth can be decom-

posed as:

$$\Delta \ln y_{rt} \approx g_A + \Omega\, \Delta h_{rt} + \Delta \ln(\text{CES}_{rt}) + \Delta \varepsilon_{rt},$$

where $\Delta \ln(\text{CES}_{rt})$ captures the direct effect of skill accumulation through the production function aggregator, which is fully determined by the parameters identified in the first stage.

The identification of the productivity parameters proceeds in three steps. First, the exogenous growth rate, $g_A$, is calibrated to match the average growth in log output, $\mathbf{E}[\Delta \ln y_{rt}]$, after accounting for the growth attributable to the average change in human capital and the CES term.

Second, the human capital externality parameter, $\Omega$, is identified using an IV-style moment. Differentiating the growth equation with respect to our instrument, $\log D_r$, yields:

$$\frac{\text{Cov}(\Delta \ln y_r, \log D_r)}{\text{Var}(\log D_r)} = \Omega \frac{\text{Cov}(\Delta h_r, \log D_r)}{\text{Var}(\log D_r)} + \frac{\text{Cov}(\Delta \ln(\text{CES}_r), \log D_r)}{\text{Var}(\log D_r)}.$$

Rearranging gives an expression for $\Omega$ that is identified from ratios of empirical covariances:

$$\Omega = \frac{\text{Cov}(\Delta \ln y_r, \log D_r)}{\text{Cov}(\Delta h_r, \log D_r)} - \frac{\text{Cov}(\Delta \ln(\text{CES}_r), \log D_r)}{\text{Cov}(\Delta h_r, \log D_r)}.$$

All terms on the right-hand side are either directly observable from the data or can be computed using the previously calibrated parameters.

Third, the standard deviation of the productivity shock, $\sigma_\varepsilon$, is calibrated to match the residual variance of log output growth, $\text{Var}(\Delta \ln y_{rt} - g_A - \Omega \Delta h_{rt} - \Delta \ln(\text{CES}_{rt}))$.

# D Additional Figures

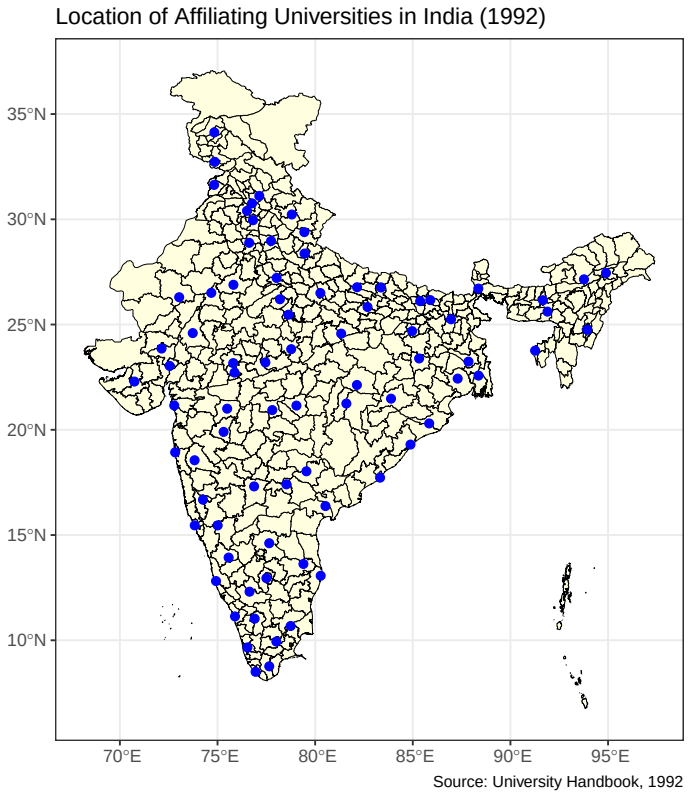
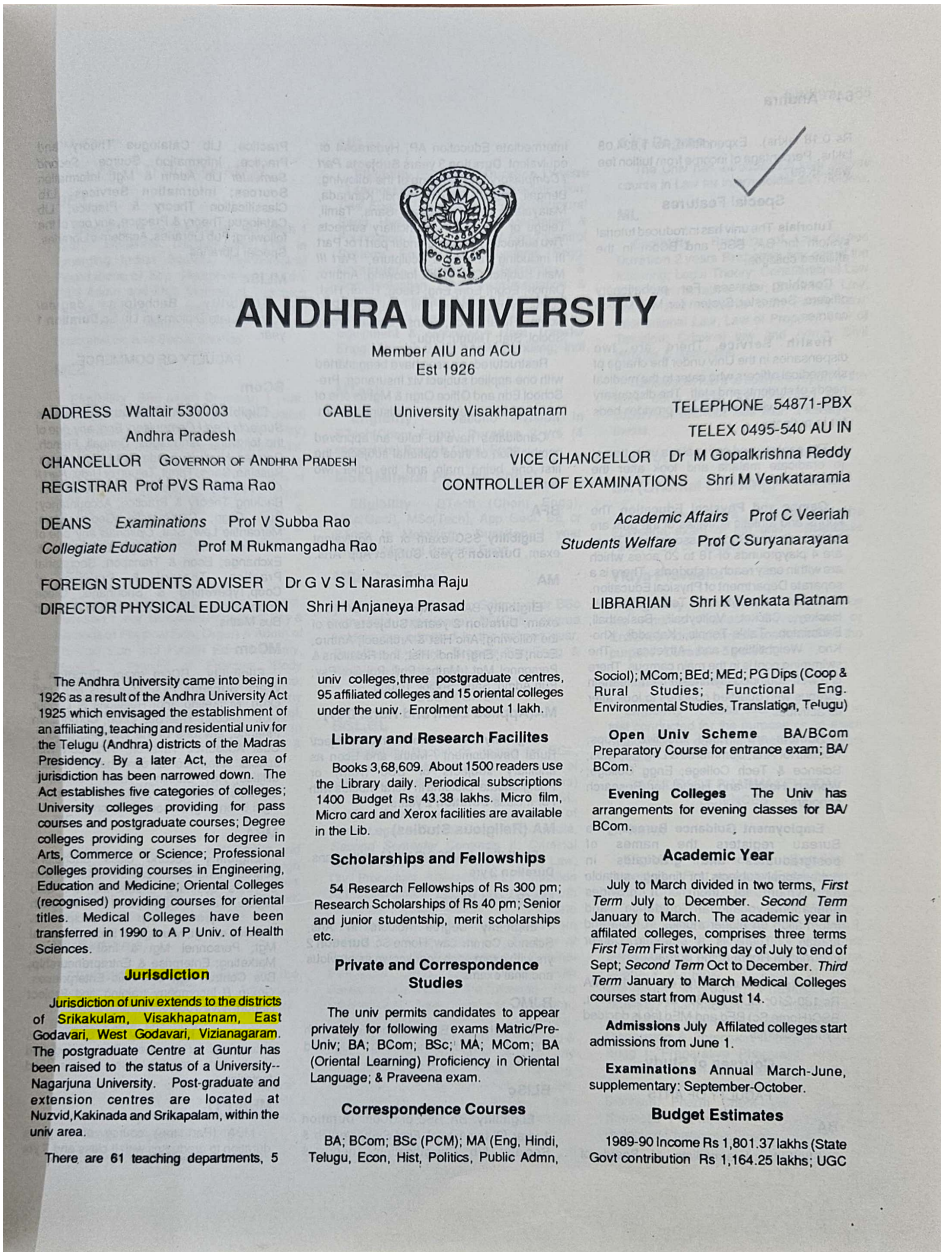



Figure A2: District Map Showing University Presence, 1991

*Notes:* This figure shows the geolocation of state public universities in 1991 that could grant affilaition to new colleges.



# ANDHRA UNIVERSITY

Member AIU and ACU
Est 1926

ADDRESS Waltair 530003
Andhra Pradesh
CABLE University Visakhapatnam
TELEPHONE 54871-PBX
TELEX 0495-540 AU IN

CHANCELLOR GOVERNOR OF ANDHRA PRADESH
VICE CHANCELLOR Dr M Gopalkrishna Reddy
REGISTRAR Prof PVS Rama Rao
CONTROLLER OF EXAMINATIONS Shri M Venkataramia

DEANS *Examinations* Prof V Subba Rao
*Academic Affairs* Prof C Veeriah
*Collegiate Education* Prof M Rukmangadha Rao
*Students Welfare* Prof C Suryanarayana

FOREIGN STUDENTS ADVISER Dr G V S L Narasimha Raju
DIRECTOR PHYSICAL EDUCATION Shri H Anjaneya Prasad
LIBRARIAN Shri K Venkata Ratnam

The Andhra University came into being in 1926 as a result of the Andhra University Act 1925 which envisaged the establishment of an affiliating, teaching and residential univ for the Telugu (Andhra) districts of the Madras Presidency. By a later Act, the area of jurisdiction has been narrowed down. The Act establishes five categories of colleges; University colleges providing for pass courses and postgraduate courses; Degree colleges providing courses for degree in Arts, Commerce or Science; Professional Colleges providing courses in Engineering, Education and Medicine; Oriental Colleges (recognised) providing courses for oriental titles. Medical Colleges have been transferred in 1990 to A P Univ. of Health Sciences.

### Jurisdiction

Jurisdiction of univ extends to the districts of Srikakulam, Visakhapatnam, East Godavari, West Godavari, Vizianagaram. The postgraduate Centre at Guntur has been raised to the status of a University--Nagarjuna University. Post-graduate and extension centres are located at Nuzvid,Kakinada and Srikapalam, within the univ area.

There are 61 teaching departments, 5 univ colleges,three postgraduate centres, 95 affiliated colleges and 15 oriental colleges under the univ. Enrolment about 1 lakh.

### Library and Research Facilites

Books 3,68,609. About 1500 readers use the Library daily. Periodical subscriptions 1400 Budget Rs 43.38 lakhs. Micro film, Micro card and Xerox facilities are available in the Lib.

### Scholarships and Fellowships

54 Research Fellowships of Rs 300 pm; Research Scholarships of Rs 40 pm; Senior and junior studentship, merit scholarships etc.

### Private and Correspondence Studies

The univ permits candidates to appear privately for following exams Matric/Pre-Univ; BA; BCom; BSc; MA; MCom; BA (Oriental Learning) Proficiency in Oriental Language; & Praveena exam.

### Correspondence Courses

BA; BCom; BSc (PCM); MA (Eng, Hindi, Telugu, Econ, Hist, Politics, Public Admn, Sociol); MCom; BEd; MEd; PG Dips (Coop & Rural Studies; Functional Eng. Environmental Studies, Translation, Telugu)

**Open Univ Scheme** BA/BCom Preparatory Course for entrance exam; BA/BCom.

**Evening Colleges** The Univ has arrangements for evening classes for BA/BCom.

### Academic Year

July to March divided in two terms, *First Term* July to December. *Second Term* January to March. The academic year in affilated colleges, comprises three terms *First Term* First working day of July to end of Sept; *Second Term* Oct to December. *Third Term* January to March Medical Colleges courses start from August 14.

**Admissions** July Affilated colleges start admissions from June 1.

**Examinations** Annual March-June, supplementary: September-October.

### Budget Estimates

1989-90 Income Rs 1,801.37 lakhs (State Govt contribution Rs 1,164.25 lakhs; UGC

Figure A3: Example Page from 1992 Handbook of Indian Universities

*Notes:* This figure reproduces a page from the 1992 edition of the *Handbook of Indian Universities*, which was used to extract detailed data on the location and jurisdiction of universities across India.

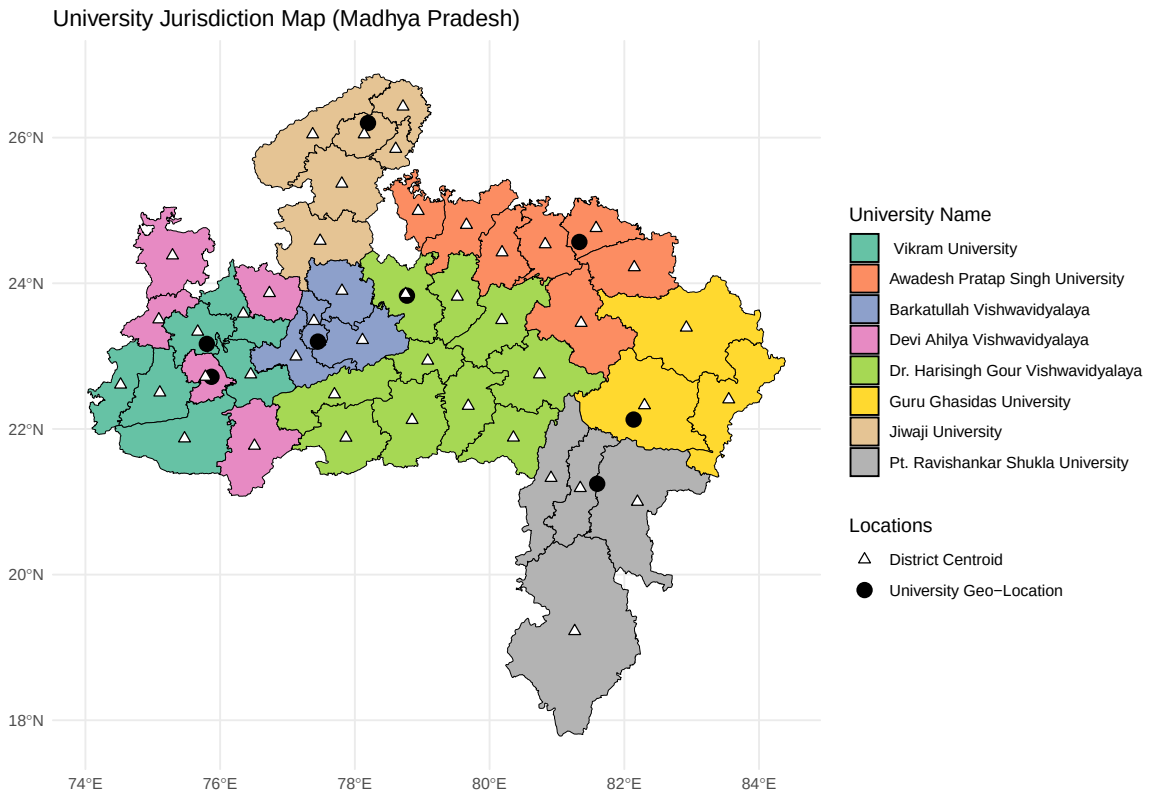



Figure A4: Location and Jurisdiction of Universities in Madhya Pradesh

*Notes:* This figure shows an example of the mapping of universities in the state of Madhya Pradesh, illustrating both their location and their jurisdictional coverage based on historical administrative boundaries.

With reference to the subject cited above, I am herewith informing you that the information you are seeking is as below:

| Sl. No. | Information seeking | Information given |
|---|---|---|
| 1 | Year of establishment of the Univesity. The exact year when University was established. | 1949 |
| 2 | Competing districts considered for the location at the time of establishment. A list of districts that were considered as potential locations for establishing the University during its inception. | Bidar, Bijapur, Gulburga, Belgaum, Dharwad, Uttar Kannada, Ballary, Raichur |
| 3 | Reason behind establishing the University in the current district and rejecting other competing districts. Detailed reasons and justifications for selecting the current district over other competing districts for the University's location. | Karnataka State has two major parts North and South. South had already one University(Mysore University). Northern districts are settled for Karnatak University |
| 4 | List of districts in which the University grants affiliation currently. A comprehensive list of all districts where the University currently grants affiliations to colleges. | Gadag, Dharwad, Uttar Kannada. Other districts are in newly established Univesities. |

Figure A5: Example of Right to Information (RTI) inquiry on university location

*Notes:* This figure shows an example of response to a question on what factors influenced the location of an universty before the reform.

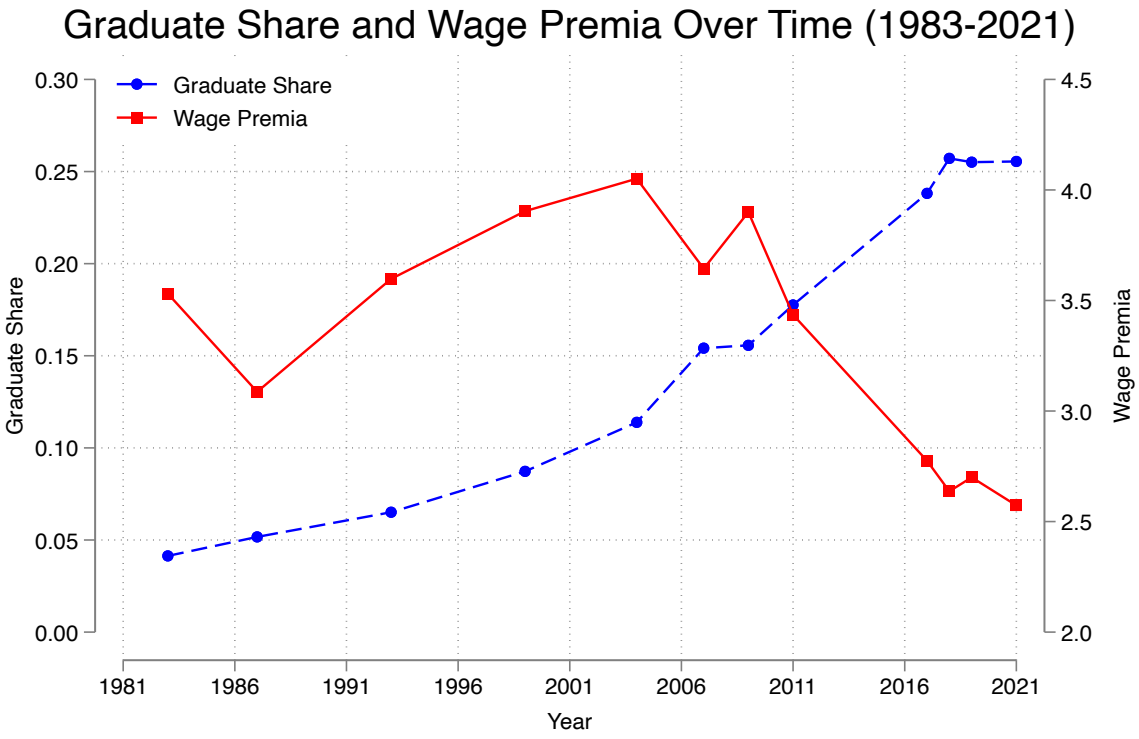



Figure A6: Time Series of graduate share and wage premia over time

*Notes:* This figure shows how increase in supply of college graduates in India has coincided with a decline in the wage premia. This is consistent with a supply driven explanation of the evolution of wage premia and graduate shares.

# E Additional Tables

This section presents additional tables that provide supporting evidence for the main results and describe key features of the data.

## E.1 Composition of Higher Education Enrollment

Table 21 presents the distribution of higher education enrollment across academic disciplines for 2001 and 2011. The table reveals several important trends in the composition of Indian higher education during this period. Arts and humanities remains the largest field, though its share declined from 46.1 percent to 37.1 percent. Engineering and technology experienced the most dramatic expansion, growing from 6.9 percent to 16.1 percent of total enrollment—a reflection of India's IT boom and the associated demand for technical skills. Science enrollment remained stable at around 19 percent, while commerce and management held steady at approximately 18 percent. Education saw notable growth, increasing from 1.3 percent to 3.6 percent of enrollment. These shifts in disciplinary composition are relevant for interpreting the economic effects of college expansion, as different fields may generate different externalities and labor market outcomes.

Table 21: Discipline-wise Enrollment (2001 vs. 2011)

| | 2001 (2001–02) | | 2011 (2011–12) | |
|---|---|---|---|---|
| Discipline | Number | Share | Number | Share |
| Science | 1,739,100 | 19.4% | 3,789,900 | 18.6% |
| Engineering/Technology | 618,600 | 6.9% | 3,261,600 | 16.1% |
| Medicine | 277,900 | 3.1% | 715,700 | 3.5% |
| Agriculture | 89,600 | 1.0% | 97,300 | 0.5% |
| Veterinary Science | 17,900 | 0.2% | 28,500 | 0.1% |
| Arts (incl. Oriental) | 4,132,700 | 46.1% | 7,539,500 | 37.1% |
| Commerce/Management | 1,604,700 | 17.9% | 3,571,100 | 17.6% |
| Law | 286,900 | 3.2% | 373,200 | 1.8% |
| Education | 116,500 | 1.3% | 732,600 | 3.6% |
| Others | 80,700 | 0.9% | 217,900 | 1.1% |
| Total | 8,964,700 | 100% | 20,327,400 | 100% |

*Source:* DST/NSTMIS, *S&T Indicators*, Table 9 "Faculty-wise Enrollment in Higher Education from 2001–02 to 2015–16".
*Notes:* The faculty-wise table pertains only to UG, PG, M.Phil and Ph.D disciplines and is based on actual survey responses; it excludes courses conducted by stand-alone institutions. Hence, the sum of faculty-wise figures may not match overall enrollment totals reported elsewhere.

## E.2 Results for Young Cohort

Table 22 presents instrumental variables estimates focusing on the employment outcomes of workers aged 21-34 years. This age group is of particular interest because younger workers are

more likely to have been directly affected by the college expansion during the study period, as they were of college-going age during 2001-2011. The results show that increases in the graduate share lead to substantial sectoral reallocation among young workers: a one-percentage-point increase in graduate share reduces agricultural employment by 0.66 percentage points and increases service sector employment by 0.70 percentage points. The magnitude of these effects is slightly larger than for the full working-age population, consistent with younger workers being more responsive to changes in educational opportunities and having lower switching costs between sectors. The effects on manufacturing and mining employment are small and statistically insignificant, mirroring the patterns observed for the full sample.

Table 22: IV Estimates: Graduate Share and Economic Outcomes (Young Cohort: 21-34)

| | (1) $\Delta s$ (Agri.) | (2) $\Delta s$ (Manuf.) | (3) $\Delta s$ (Mining) | (4) $\Delta s$ (Service) |
|---|---|---|---|---|
| $\Delta h$ | -0.655*** | -0.036 | -0.006 | 0.698*** |
| | (0.159) | (0.084) | (0.014) | (0.143) |
| Constant | 0.017 | -0.011* | -0.001 | -0.005 |
| | (0.012) | (0.006) | (0.001) | (0.011) |
| Observations | 374 | 374 | 374 | 374 |
| Effective F-statistic | 27.37 | 27.37 | 27.37 | 27.37 |

*Notes:* This table presents IV estimates of the effect of changes in overall college graduate share on changes in sectoral employment shares for young workers (aged 21-34 years) between 2001 and 2011. The instrument is log distance to the nearest affiliating university in 1991. The Effective F-statistic (Olea and Pflueger, 2013) tests for weak instruments. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

## E.3 Full Sample Results

Table 23 presents the first-stage relationship between log distance to the nearest affiliating university and changes in college graduate share. Column (1) shows results for the graduate share of the entire working-age population, while column (2) focuses on young adults aged 20-24 years. The negative and highly significant coefficients confirm that districts farther from universities experienced smaller increases in graduate share during 2001-2011. Specifically, a one percent increase in distance is associated with a 0.011 percentage point smaller increase in graduate share for all ages, and a 0.016 percentage point smaller increase for young adults. The F-statistics of 32.79 and 18.81 substantially exceed conventional thresholds for weak instruments (Stock and Yogo, 2005), indicating that distance provides strong predictive power for educational expansion. The stronger effect for young adults reflects the fact that this cohort was of college-going age during the expansion period.

Table 24 presents the main instrumental variables estimates of the effect of college graduate share on economic outcomes. These results replicate the estimates from the main text and serve

Table 23: IV First Stage Estimates: Changes in Graduate Share (2001-2011)

| | (1) $\Delta h$ (All Ages) | (2) $\Delta h$ (Young: 20-24 years) |
|---|---|---|
| log Distance | -0.011*** | -0.016*** |
| | (0.002) | (0.004) |
| Constant | 0.099*** | 0.148*** |
| | (0.009) | (0.017) |
| Observations | 454 | 454 |
| F-statistic | 32.79 | 18.81 |

*Notes:* This table presents first-stage regressions of changes in college graduate share on log distance to the nearest affiliating university in 1991. Column (1) uses the graduate share of the entire working-age population (15-64 years), while column (2) uses the graduate share of young adults aged 20-24 years. The F-statistic tests the null hypothesis that the instrument is weak. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

as a convenient reference. The table shows that a one-percentage-point increase in graduate share causes a 0.37 percentage point decrease in agricultural employment, a 0.22 percentage point decrease in manufacturing employment, and a 0.61 percentage point increase in service sector employment. GDP per worker increases by 2.3 percent, while the effect on nightlights is positive but imprecisely estimated. The skill premium estimate is negative but also imprecise, suggesting that the increased supply of educated workers does not significantly compress wage differentials.

Table 24: IV Estimates: Graduate Share and Economic Outcomes

| | (1) $\Delta s$ (Agri.) | (2) $\Delta s$ (Manuf.) | (3) $\Delta s$ (Service) | (4) $\Delta$ log(GDPpw) | (5) $\Delta$ log(NL) | (6) $\Delta \log(w_H/w_L)$ |
|---|---|---|---|---|---|---|
| $\Delta h$ | -0.366** | -0.219* | 0.606*** | 2.308** | 1.043 | -3.103 |
| | (0.182) | (0.117) | (0.158) | (0.913) | (1.896) | (2.522) |
| Constant | -0.009 | -0.004 | 0.014* | 0.463*** | 0.396*** | 0.034 |
| | (0.009) | (0.006) | (0.008) | (0.046) | (0.094) | (0.128) |
| Observations | 454 | 454 | 454 | 454 | 436 | 436 |
| Effective F-statistic | 32.79 | 32.79 | 32.79 | 32.79 | 37.15 | 34.45 |

*Notes:* This table presents IV estimates of the effect of changes in college graduate share ($\Delta h$) on changes in sectoral employment shares, GDP per worker, nightlights, and the skill premium between 2001 and 2011. The instrument is log distance to the nearest affiliating university in 1991. The Effective F-statistic (Olea and Pflueger, 2013) tests for weak instruments. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.